# Testimony to the Joint Economic Committee of the U.S. Senate

# Providing Health Information to Consumers to Improve the Efficiency and Effectiveness of the U.S. Healthcare System

**By**
**Douglas G. Cave, Ph.D., M.P.H.**
**President of Cave Consulting Group (CCGroup LLC)**
**May 10, 2006**

Thank you for your invitation to present some of the most recent experience with consumer health information from CCGroup LLC and additional observations relating to the experience of the health insurance industry. My name is Douglas Cave, and I am the President of CCGroup. CCGroup is a company focused on improving the efficiency and effectiveness in the healthcare delivery system. We recognize the need to address all components of medical trend (and not just service price discounts), if efficiency and effectiveness are to be improved and medical trend is to be controlled. Today, CCGroup works with some of the largest health insurance companies in the country in the areas of consumer health care transparency, pay for performance (P4P) programs, high performance network (HPN) building, and network tiering based on practitioner efficiency and effectiveness.

My testimony summarizes some current health services literature and "in practice" findings with respect to measuring practitioner efficiency and effectiveness as well as presenting consumers with health information. There are accurate and reliable methodologies available for measuring practitioner efficiency and effectiveness using medical claims data. Understanding the methodologies that work (and do not work) is very important. If practitioner efficiency and effectiveness are not accurately measured, then consumers cannot be provided with meaningful practitioner information. The end result will be that the efficiency and effectiveness of the current U.S. healthcare system may not be improved further through greater levels of consumer health information. This is because many practitioners may be inaccurately measured as efficient and effective (or, conversely, inaccurately measured as inefficient and ineffective), and the consumer may be inadvertently guided to the less efficient and less effective practitioners.

## 1. Vision for Consumer Health Information

### Background

A trend in the health insurance industry today is to build consumer-directed health benefit plans. A definition of consumer-directed benefit plans is health benefit plans that: (1) incentivize consumers to select more affordable and/or higher quality health care options; and (2) provide consumers with cost and/or quality information on practitioners and medical conditions with which consumers can make more informed choices.

A stated objective of consumer-directed health benefit plans is to increase consumer engagement in health care decision making, rather than relying solely on practitioners and hospitals in the decision-making process. The thought is that improved consumerism will result in improved

efficiency and effectiveness (or quality) of care in the U.S. healthcare system, thereby reducing, or at least controlling, current healthcare expenditures. I define here efficiency as using the appropriate level of medical services in an appropriate setting to treat a patient's medical condition and achieve a desired quality of patient care. Thus, efficiency is a function of unit price, volume of service, intensity of service, and quality of service. Effectiveness means coordinating medical services for a patient based on process of care and/or outcome measures that are established by national expert panels.

Whether consumer-directed health benefit plans will succeed depends largely on providing consumers with accurate and reliable physician-level information. Moreover, looking beyond the current debate over the value of consumer-directed health benefit plans, we have to provide consumers with physician-level efficiency and effectiveness information if we desire to improve or reform the current U.S. healthcare system. The reason is that providing this information drives competition, whereby a consumer is expected to choose based on product price, quality, and convenience. This fact is no different than in any other competitive industry.

**Available Methodologies for Accurate and Reliable Efficiency and Effectiveness Scores**

There are several methodologies that will provide accurate and reliable physician "efficiency" information. This is important, as consumers will come to trust and utilize the efficiency information when the data is accurate and stable from year-to-year. Using this information, a "star" rating system (or another easily understood consumer rating system) may be used to rank more efficient and less efficient physicians by specialty type. The star system may apply to a physician's overall efficiency score or to a physician's medical condition-specific efficiency score.

On the other hand, there are many methodologies being employed in the market that will calculate physicians' efficiency scores, but the scores are not accurate or reliable. In essence, this means a physician's efficiency score will be jumping around from year-to-year. For instance, in 2005 the physician may appear efficient (using less resources as compared to a peer group or best practice), but in 2006 the physician may appear inefficient (using more resources as compared to a peer group or best practice). Consumer health information based on non-stable scores will not be meaningful or helpful in reforming the U.S. healthcare system.

There are several methodologies that will provide accurate and reliable physician "effectiveness" information. Using this information, a star rating system (or another easily understood consumer rating system) may be used to rank more effective and less effective physicians by specialty type. The star system may apply to a physician's overall effectiveness score or to a physician's medical condition-specific effectiveness results.

In applying any effectiveness measurement methodology, we need to define a good set of medical condition-specific effectiveness measures. A large amount of work has been performed, or summarized, in this area by the Agency for Healthcare Research and Quality (AHRQ) and other expert organizations with respect to medical condition-specific process and outcomes of care (especially those measures that may be calculated using medical claims data): RAND, National Quality Forum (NQF), National Committee for Quality Assurance (NCQA), and

Ambulatory Care Quality Alliance (AQA). The richness of this effectiveness information should be recognized and utilized.

As with efficiency methodologies, there are effectiveness methodologies being employed in the market that will calculate physicians' effectiveness scores, but the scores are not accurate or reliable. In using these methodologies, the consumer health information will not be meaningful or helpful in improving the quality of care in the U.S. healthcare system.

**Claims Experience Data to Rate Physician Efficiency and Effectiveness**

Medical claims data is the largest single source of information that a health plan has available to rate physician efficiency and effectiveness. Therefore, health plans will continue to use claims data as the primary information source for individual physician and hospital evaluation and measurement.

However, in any geographic region of the country, the membership base for many health plans is not large enough to generate enough claims data experience to measure all physicians in their network. A general rule of thumb is that a health plan's claims data may be able to measure the efficiency and effectiveness of about 60% of the health plan's network physicians (which equates to about 30% of all practicing physicians in the region because not all physicians are under contract with the health plan). Of these 60% of network physicians, about 75% will have a lower volume of assigned episodes of care (typically under 50 episodes) for efficiency and effectiveness measurement.

The question becomes, "How does a health plan obtain enough claims data to accurately and reliability rate the efficiency and effectiveness of all (or most) physicians?" One answer is for the health plan to pool its claims data with competing plans in a geographic region. However, health plans are hesitant to pool their claims data with competing plans out of fear the negotiated unit price discounts with physicians and hospitals may be revealed to a competitor health plan. Then, the competitor health plan may try to replicate these discounts.

Another answer is for the Centers for Medicare and Medicaid Services (CMS) to provide access to the full CMS Part A and Part B claims databases, holding back only to the extent necessary to protect the privacy of individual Medicare beneficiaries. Initial testing shows the CMS databases may be able to measure the efficiency and effectiveness of 80% of "all" (not just the health plan's network physicians) practicing physicians in a geographic region. This percent of physicians measured is more than enough to provide consumers with meaningful, physician-level efficiency and effectiveness information.

It is important to recognize that even using the full CMS claims database, many of the measured physicians will have a lower volume of assigned episodes of care (generally under 70 episodes) for efficiency and effectiveness measurement. Consequently, we need to always consider the Law of Low Episode Numbers when developing and implementing methodologies that will accurately and reliably measure practitioner efficiency and effectiveness.

**Where Is the Market with Respect to Providing Consumer Health Information?**

Most health plans desire to provide physician efficiency and effectiveness information to their plan members. The most frequent way expressed by health plans to disseminate this information is through a secured Internet website (as opposed to a hardcopy network provider directory or other media).

To date, the most commonly offered comparisons have been limited to prescription drug options, surgical procedures, and diagnostic tests. With respect to these services, the information has been on a unit price basis charged by the physician, pharmacy, or hospital. The information has not been presented on a medical condition-specific, longitudinal episode of care basis, which accounts for all components of medical expenditures, which include unit prices, volume of services, and intensity of services. Presenting only unit prices to a consumer is problematic, and may provide misleading signals of relative efficiency. For example, health services researchers have documented that lower negotiated unit prices generally induce physicians to provider a greater volume of services and/or more resource-intensive services.

There are three main reasons as to why health plans have elected to start consumerism efforts with unit price information:

- A key component of consumer-directed health benefit plans has been to provide consumers with information to make more informed healthcare decisions. The most readily available information to share with consumers has been unit prices. Therefore, health plans have started with unit price comparisons for physicians, hospitals, and pharmaceuticals.

- Health plans have been reticent to be the first health plan in a geographic region to present consumers with physician efficiency and effectiveness information out of worry that physicians will threaten to drop out of the network or, alternatively, will negotiate lower price discounts then currently exist. To date, most health plan networks have been built on physician and hospital price discounts alone. Therefore, health plans do not want to upset the balance they presently have with network providers.

- Some health plans do not have enough claims data experience (at least is some geographic regions) to measure enough network physicians to have a successful physician efficiency and effectiveness measurement program. In these instances, the health plan would have little consumer health information to share with their plan members.

The present comparisons aside, many health plans recognize the need to produce more detailed consumer health information. Of these health plans, most are in the process of developing a program that provides physician level (or physician-group level) efficiency and/or effectiveness information. The employer benefit consultant community and larger employers have been vocal to health plans about the urgent need to present this type of meaningful consumer health information. Most health plans are attempting to meet this strong market demand, but each health plan is going about the process in its own unique manner.

## 2. Overview of CCGroup

### About CCGroup

CCGroup believes market efficiency and effectiveness will be improved once practitioner and hospital efficiency and effectiveness are accurately and reliability measured, practitioners are informed of their performance results, and patients have knowledge of — and are directed to — the most efficient and effective practicing providers. Moreover, transparency of practitioner practice patterns to the consumer (patient) will continue to grow. Practitioner pay-for-performance programs will prevail as a key mechanism to improve practitioner performance. Finally, we believe market efficiency and effectiveness will be improved once patients with unstable chronic medical conditions are accurately identified and properly managed.

We began measuring the efficiency and effectiveness of practitioners' practice patterns in 1990 — and published our first article in 1992. Since this time, we have worked with many leading HMOs, insurance companies, employers, physician-hospital organizations, and third party administrators. Today, CCGroup works with some of the largest health insurance companies in the country.

### Efficiency and Effectiveness Measurement Software

The CCGroup Marketbasket System™ compares physician efficiency and effectiveness to a specialty-specific peer group using a standardized set of prevalent medical condition episodes with the intent of minimizing the influence of patient case mix (or health status) differences and methodology statistical errors. The efficiency and effectiveness of physician groups may also be compared.

Derivations of the CCGroup Marketbasket System™ have been validated, published, and tested on over 48 million members and 398,000 physicians in health plans, TPAs, ASOs, Medicare, and Medicaid, with over a decade of research and development.
The CCGroup Marketbasket System™ is made up of four key components:

- **The Cave Grouper™**. The Cave Grouper™ groups over 14,000 unique ICD.9 diagnosis codes into 526 meaningful medical conditions. The 526 medical conditions in the Cave Grouper™ account for 100% of all medical conditions and expenditures as identified by ICD.9 medical condition diagnostic codes.

- **CCGroup EfficiencyCare™ Module (physician efficiency measurement software)**. The CCGroup EfficiencyCare™ Module takes the output from the Cave Grouper™ and develops specialty-specific physician efficiency scores that compare individual physician efficiency (or physician group efficiency) against the efficiency of a peer group of interest.

- **CCGroup EffectivenessCare™ (physician effectiveness or quality measurement software).** The CCGroup EffectivenessCare™ Module takes the output from the Cave Grouper™ and develops specialty-specific physician effectiveness (i.e., process of care quality) scores that compare individual physician effectiveness (or physician group

effectiveness) against the effectiveness of a peer group of interest. There are over 100 effectiveness metrics derived from national expert resources.

- **CCGroup MediScreen™ (high-cost patient predictive model and patient health-risk stratification software).** The CCGroup MediScreen™ Module takes the output from the Cave Grouper™ and develops accurate and reliable scores for predicting the chance of being a high-cost patient next year. Moreover, the CCGroup MediScreen Module™ allows the user to health-risk stratify patients across all medical conditions.

The purpose of the CCGroup Marketbasket System™ methods and software system is to compare physician efficiency and effectiveness to a specialty-specific peer group, using a standardized set of prevalent medical condition episodes, with the intent of minimizing the influence of patient case mix and methodology statistical errors.

## 3. Episodes of Care for Measuring Practitioner Effectiveness and Efficiency

In measuring physician efficiency and effectiveness, the health insurance industry has turned to using medical condition episodes of care. The purpose of episode of care groupers is to form longitudinal episodes of care for a patient using medical claims data. A longitudinal episode of care is defined as all services linked together that are used to treat a patient's medical condition within a specified period of time—including all ambulatory, outpatient, inpatient, and prescription drug experience. This linkage allows examination of a physician's (or several physicians') global patterns of treatment for a specific patient with a specific condition, such as diabetes and arthritis. The longitudinal episode of care may also be used in patient disease management, patient health promotion and wellness, and many other healthcare programs.

For acute conditions (e.g., upper respiratory infections), the patient's episode duration is specified by a time period, or window period, that defines the maximum number of days between contact with a provider for which follow-up care is still reasonable. Each of the acute medical conditions has its unique window period. If the date of service for a patient's episode is separated by a longer period than the window period, the latest date of service is considered the start date for a new condition-specific episode of care.

For example, the window period for upper respiratory infections may be 60 days. Assume that a patient had three treatments in January and two in the following August. Because the treatments in the series were separated by more than 60 days, these would be two episodes of care.

For chronic conditions, once the episode starts, the episode continues on. By definition, a chronic episode does not stop during a defined study period. However, for the practitioner efficiency and effectiveness analysis, we need to ensure a constant and defined number of days for each condition-specific episode. This number of days generally is 180 days or 365 days. For example, we may examine the initial 180 day period after the first diagnosis of the medical condition occurs being the study period.

## 4. Law of Low Episode Numbers

The Law of Low Episode Numbers states that no matter what health plan membership base is examined, about 75% of those practitioners that can be measured in a health plan network will have less than or equal to 50 episodes assigned to each practitioner. The other 25% of network practitioners may have greater than 50 episodes assigned to each practitioner.

Therefore, the majority of practitioners in a health plan's network have a low volume of episodes. The Law of Low Episodes Numbers means that we have to pay very close attention to potential efficiency and effectiveness measurement errors. Otherwise, an employed methodology will provide practitioner efficiency and effectiveness scores, but these scores may not be accurate or reliable (i.e., stable).

The Law of Low Episode Numbers also applies to the Centers for Medicare and Medicaid Services (CMS) Part A and Part B claims databases (the 100% non-sample files). A strong advantage of using the CMS databases is that we can measure the efficiency and effectiveness of a significantly greater number of practitioners:

- For instance, in a given geographic region, a health plan's claims data may be able to measure the efficiency and effectiveness of 60% of the health plan's network physicians (which equates to about 30% of all practicing physicians in the region because not all physicians are under contract with the health plan). Of these 60% of network physicians, about 75% will have a lower volume of assigned episodes of care (typically less than 50 episodes) for efficiency and effectiveness measurement.

- In this same geographic region, the CMS claims data may be able to measure the efficiency and effectiveness of 80% of "all" (not just a health plan's network) practicing physicians. Of these 80% of physicians, about 70% will have a lower volume of assigned episodes of care for efficiency and effectiveness measurement.

These findings show that the Law of Low Episode Numbers also will apply in using the CMS claims databases.

Consequently, we need to always consider the Law of Low Episode Numbers when developing and implementing methodologies to accurately and reliably measure practitioner efficiency and effectiveness. Otherwise, any consumer health information based on practitioner efficiency and effectiveness measurement will not produce healthcare improvements. The Law of Low Episode Numbers does not disappear when using very large claims-based databases.

## 5. Examples of Practitioner Measurement Accuracy and Reliability Issues

One of the most frequent questions being asked by health insurance companies is, "Why do my physician efficiency and effectiveness scores change between time periods, or when small changes are made to my current methodology?" This question brings into light some of the difficulties with accurately and reliably measuring practitioner efficiency and effectiveness.

I present here summarized results from three different studies that illustrate the lack of agreement and reliability within and between measurement systems. Then, I present three of the top methodology reasons that lead to this lack of agreement and reliability. These methodology issues are particularly pronounced with the Law of Low Episode Numbers.

### Health Plan Study on Reliability of Scores Over Time

The first study was performed by a large, national BlueCross BlueShield health plan (results presented at a 2005 managed care conference). This study was designed to examine the reliability of physician efficiency results from time period T1 (2002-2003) to time period T2 (updated 6 months of claims data added to the 2002-2003 T1 claims data), whereby the same exact episode-based physician efficiency methodology was employed in T1 and T2.

The following is a summary of key methodology elements:

- To be included in the analysis, a physician needed to have 30 or more episodes in both T1 and T2. There were 11,951 physicians that met this study criterion.

- T1 was January 1, 2002 – December 31, 2003. Then, six months of new claims data was added to the T1 claims data, and the initial 6 months of claims data in T1 was removed from the study. This became the T2 claims data set. Therefore, the T2 claims dataset had 25% of new claims data added to it; 75% of the T1 claims data remained exactly the same with the T2 claims data.

- In this study, "efficient" physicians received a score of 0.95 or less (whereby the peer group comparator was a 1.00). Physicians practicing of or around the "peer group" efficiency received a score of 0.96 – 1.04. "Inefficient" physicians received a score of 1.05 or more.

The following table summarizes the results from this health plan study:

- In the shaded diagonal area, good correlation of physician efficiency scores would have resulted in 100% of the physicians being in the yellow highlighted areas of the table (or perfect correlation of scores between T1 and T2). Instead, we observe that 78% of the physicians are in the yellow highlighted areas.

- Moreover, we observe that only 67% of the physicians ranked an "inefficient" in T1 were also ranked as "inefficient" in T2. We determine this 67% correlation between T1 to T2 by dividing the 21% (which is the percent of physicians ranked inefficient in both T1 and T2) by the 32% (which is the percent of physicians ranked "inefficient" in T1) in the below table.

- Yet, these inefficient physicians are the very physicians we want to accurately identify for the implementation of and success for a consumer health information initiative.

**T2 Analysis with 6 Months New Data**

| T1 Original Analysis | Score Type | Efficient or Peer Group | Inefficient | Total |
|---|---|---|---|---|
| | Efficient or Peer Group | 57% | 11% | 68% |
| | Inefficient | 11% | 21% | 32% |
| | Total | 25% | 25% | 100% |

Conclusions: The study results showed that the utilized episode-based efficiency methodology resulted in fairly unstable year-to-year practitioner efficiency scores. Furthermore, this instability occurred when 75% of the claims data remained the same between T1 and T2. We expect the results to be substantially worse with no overlap in claims data between T1 and T2.

There appear to be two main methodology issues for this low reliability in physician efficiency scores over time (which will be detailed in Section 6). These issues appear to be more prominent with the Law of Low Episode Numbers:

- Issue #1: Inadequate patient case-mix adjustment
- Issue #2: Standard deviation statistical bias.

### Academic Study on Accuracy of Scores in One Study Time Period

The second study was performed by Thomas, Grazier, and Ward (William Thomas, Kyle Grazier, Kathleen Ward. A Comparative Evaluation of Risk-Adjustment Methodologies for Profiling Physician Practice Efficiency. A report to the Robert Wood Johnson Foundation, September 2002; work recently published in Health Services Research, August 2004, Economic Profiling of PCPs: Consistency Among Risk-Adjusted Measures.) This study was designed to examine the inter-rater agreement between six physician efficiency methodologies in measuring physician efficiency within one time period (T1). Some of the methodologies were not episode-based, but instead, were "overall charges per member" based.

In summary, the study applied about a 100,000 member claims-based database (time T1) to six efficiency measurement systems. The thought was to observe how well these systems agreed with each other in ranking a physician as inefficient:

- The results showed that when one system ranked a physician as inefficient, sometimes only one of the other five systems ranked that same physician as inefficient.

- After some statistical bias corrections, the authors found that sometimes two of the other physician efficiency measurement systems ranked that same physician as inefficient.

- The results showed generally 30% agreement or less across measurement systems in ranking the same physician as inefficient.

- Yet, these inefficient physicians are the very physicians we want to accurately identify for the implementation of and success for a consumer health information initiative.

Conclusions: The study results showed that the employed efficiency methodologies resulted in low agreement between systems in ranking a physician as inefficient. Thomas *et al* concluded that there may be at least two methodology issues for this low agreement in physician efficiency scores over time (which will be detailed in Section 6). These issues appear to be more prominent with the Law of Low Episode Numbers.

- Issue #1: Inadequate patient case-mix adjustment
- Issue #2: Standard deviation statistical bias.

### CCGroup Study on Accuracy and Reliability of Scores

The third study was performed by CCGroup in working with a large health plan's claims data. This study was designed to examine the agreement and reliability (both) of two physician efficiency measurement systems – both using a medical condition, episode-based approach. The first system was the CCGroup Marketbasket System method and software. The second system was the health plan's internal efficiency measurement system.

The CCGroup Marketbasket System adjusts for the two main methodology issues resulting in the low agreement and low reliability of physician efficiency scores: (1) inadequate patient case-mix adjustment; and (2) standard deviation statistical bias. The second system does not adjust for these two issues.

In this study, CCGroup initially examined the inter-rater agreement between the two physician efficiency methodologies in measuring physician efficiency within one time period (T1 = calendar year 2003). The following is a summary of key comparison elements:

- To be included in the analysis, a physician needed to have 35 or more episodes present in the time period T1 under both physician efficiency methodologies. We examined different specialty types. However, only the family and general practitioner (FP/GPs) results are presented here. There were 227 FP/GPs that met the minimum episode criterion.

- After receiving an efficiency score, physicians were separated into quartiles (1, 2, 3, and 4). Quartile 1 physicians used fewer medical resources to treat the patient episodes-of-care as compared to their physician peer group. Quartile 2 and Quartile 3 physicians were the next quartiles of physicians in terms of the amount of resources used to treat the patient episodes-of-care. Quartile 4 physicians used greater medical resources to treat the patient episodes-of-care as compared to the physician peer group.

- If there was perfect agreement between the two efficiency measurement systems, then we would expect results as found in the following table. Note that 100% of physicians should fall along the diagonal because the Quartile 1, Quartile 2, Quartile 3, and Quartile 4 physicians should be exactly the same between the two systems.

**CCGroup Marketbasket System**

| Health Plan's Internal Efficiency Measurement System | Quartile | Q1 | Q2 | Q3 | Q4 | Total |
|---|---|---|---|---|---|---|
| | Q1 | 25% | 0% | 0% | 0% | 25% |
| | Q2 | 0% | 25% | 0% | 0% | 25% |
| | Q3 | 0% | 0% | 25% | 0% | 25% |
| | Q4 | 0% | 0% | 0% | 25% | 25% |
| | Total | 25% | 25% | 25% | 25% | 100% |

**N = 227 Family/General Practitioners with >=35 episodes of care in both efficiency measurement systems.**

The next table summarizes the results from this agreement study:

- We observe that only 41% of the physicians are in the yellow highlighted diagonal area – and not 100%.

- As importantly, we observe that only 11% of the most inefficient physicians (Quartile 4) matched between the two systems – and not the desired 25%.

- Furthermore, 23% of the physicians are two or more quartiles away between the two measurement systems.

| Health Plan's Internal Efficiency Measurement System | CCGroup Marketbasket System | | | | |
|---|---|---|---|---|---|
| **Quartile** | **Q1** | **Q2** | **Q3** | **Q4** | **Total** |
| **Q1** | **12%** | 7% | 3% | 3% | 25% |
| **Q2** | 5% | **9%** | 6% | 5% | 25% |
| **Q3** | 5% | 5% | **9%** | 6% | 25% |
| **Q4** | 3% | 4% | 7% | **11%** | 25% |
| **Total** | 25% | 25% | 25% | 25% | 100% |

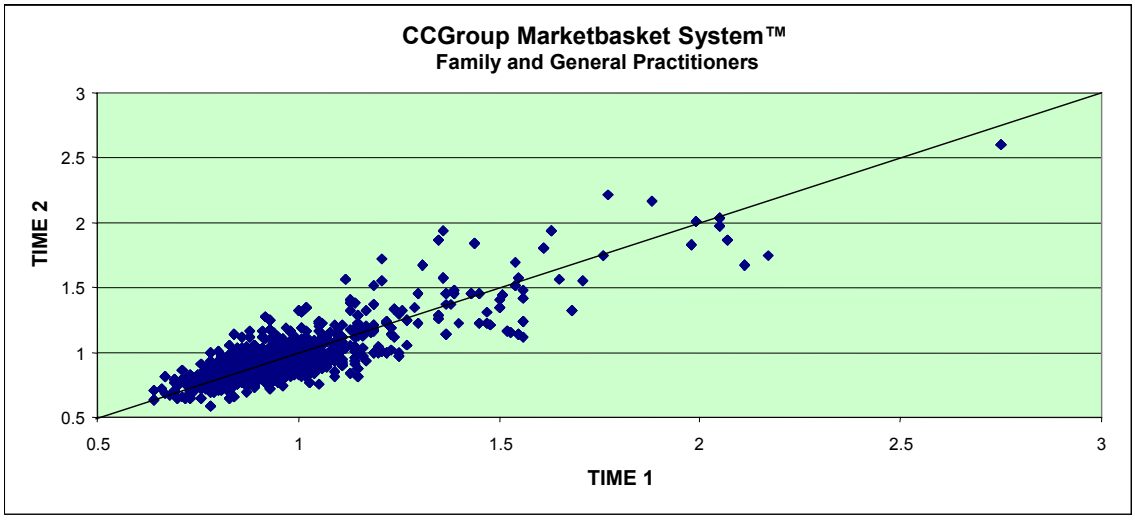
**N = 227 Family/General Practitioners with >=35 episodes of care in both efficiency measurement systems.**

Conclusions: The study results showed that the employed efficiency methodologies resulted in low agreement between the two systems in ranking a physician as inefficient. As with the other two studies defined above, the health plan's internal efficiency measurement system maintained the two main methodology issues: (1) inadequate patient case-mix adjustment; and (2) standard deviation statistical bias. The CCGroup Marketbasket System adjusted for these two main methodology issues.

The next step in this study was to examine the reliability of physician efficiency scores from time period T1 and T2. For the CCGroup Marketbasket System, T1 equaled the 2002 calendar year, and T2 equaled the 2003 calendar year. There was no overlapping claims data between T1 and T2. For the health plan's internal efficiency measurement system, T1 equaled the two year period 2002-2003, and T2 equaled the two year period 2003-2004. Thus, about 50% of the claims data remained the same between T1 and T2.

The reason for performing the reliability analysis is as follows. We wanted to determine if one of the systems has a better reliability of physician efficiency scores between T1 and T2. If yes, then this system may be considered to be more accurate and more reliable then the other system. However, if both systems have the same reliability between T1 and T2, then we cannot judge that correcting for the two main methodology issues has any relevance to improving the accuracy and reliability of efficiency scoring.

CCGroup employed the Pearson's correlation coefficient (r) to determine the agreement between physician efficiency scores for the two measurement systems. The correlation coefficient varies between 0.00 and 1.00. The closer the score is to 1.00, the stronger the agreement between T1 and T2 (or the stronger the correlation).




The results showed the following:

- The T1 and T2 correlation results for the FP/GPs was 0.88 using the CCGroup Marketbasket System methodology. (Refer to the inserted diagram above.) This is a strong indicator that the physician efficiency scores are accurate and reliable. For the other tested specialties, the T1 and T2 results showed a correlation of between 0.70 and 0.95 – with the average specialty-specific correlation being about 0.79.

- The T1 and T2 correlation results for the FP/GPs was 0.42 using the health plan's internal physician efficiency measurement system. These results indicate that the physician efficiency scores are not very accurate or reliable. Many of the other tested specialties had T1 and T2 correlation results in a similar correlation (r) range – with the average specialty-specific correlation being about 0.39.

Conclusions: The results indicated that the CCGroup Marketbasket System methodology resulted in fairly stable year-to-year physician efficiency results (r = 0.88). The CCGroup Marketbasket System did adjust for the two main methodology measurement issues. The health plan's system, which maintained the two main methodology measurement issues, had a significantly lower year-to-year correlation. This finding indicates that correcting for the two main methodology issues does have a positive impact on physician efficiency score accuracy and reliability.

## 6. Main Methodology Errors/Issues with Respect to Practitioner Efficiency and Effectiveness Measurement, and How to Correct for These Errors/Issues

An accurate and reliable practitioner efficiency and effectiveness methodology needs to address the main sources of accuracy and reliability error. This error reduction is important to consider – if market efficiency and effectiveness are to be improved through consumer health information. Otherwise, consumers may be inadvertently guided to the less efficient and less effective practitioners, when the actual objective of consumer health information is to direct consumers to the more efficient and more effective practitioners.

I discuss here three of the most important measurement methodology issues/errors, and provide one example of how to correct for each measurement methodology issue/error. There may be additional ways to correct for each measurement methodology issue/error. The three issues are the following:

- Issue #1: Inadequate patient case-mix adjustment
- Issue #2: Inadequate standard deviation statistical bias adjustment
- Issue #3: Inadequate episode of care severity of illness adjustment.

### Issue #1: Inadequate Patient Case-Mix Adjustment

Many practitioner efficiency and effectiveness methodologies continue to examine "services per 1,000 members" or "all non-outlier episodes of care" treated by a physician. These approaches probably add the most to efficiency measurement error.

These methodologies attempt to adjust "services per 1,000 members" or "all episodes of care" treated by a physician by age and gender—and then compare one practitioner's utilization pattern to a peer group average. However, age and gender explain less than 5% of the variance in a patient's medical expenditures. This means that over 95% of the variance is unexplained, and may be attributed to differences in patient health status (or case mix). Patient health status and patient case mix are used inter-changeably in this Section 6.

Some methodologies adjust "services per 1,000 members" or "all non-outlier episodes of care" based on specific ICD.9 (or diagnosis) algorithms that measure expected resource intensity. The idea is that a patient's diagnosis codes will provide more predictive power than age/gender alone. However, the most predictive of the published and marketed models explain only 20% to 30% of the variance in a patient's medical expenditures. This means that 70% or more of the variance continues to be unexplained, and may be attributed to differences in patient health status.

The best predictive models on the market today explain only 20% to 30% of the variance in a patient's medical expenditures. (This includes DxCGs, ACGs, ERGs, and other adjustment tools.) This means that 70% or more of the variance is unexplained, and may be attributed to differences in patient health status. Consequently, including all—or almost all—patients in practitioner efficiency measurement will result in unstable and inaccurate ratings.

This is a particular problem because of the "Law of Low Episode Numbers" in that each physician generally has only 20-50 assigned episodes (refer to Section 4). With the Law of Low Episode Numbers, we need a case-mix adjustment that explains about 70% to 90% of the variance in a patient's medical expenditures – and claims data is just not sensitive enough to obtain this predictive power.

In addition, practitioners often criticize the "services per 1,000 members" and "all episodes of care" methodologies for not appropriately adjusting for differences in patient health status — rightly stating that their patients may be "sicker."

What does this mean with respect to measuring individual physician efficiency? If all claim line items (CLIs) and/or episodes of care (tracked to a physician) are used in the efficiency or effectiveness analysis, then up to 70% of the observed utilization difference between physicians may be attributed to patient health status differences. Therefore, you are not measuring individual physician efficiency differences, but patient health status differences.

This weakness in current case-mix adjustment tools means that we cannot examine all CLIs or patient episodes of care treated by a physician. Instead, an isolated set of more prevalent medical conditions, stratified by severity of illness (SOI) level, needs to be examined across physicians of a similar specialty type. In effect, the idea is to eliminate patient health status (or case mix) background noise by isolating down what medical condition-specific episodes are examined in a practitioner efficiency and effectiveness analysis.

Yet, many existing physician efficiency methodologies continue to examine all non-outlier episodes of care assigned to a physician:

- An actual weighted average episode charge is calculated.

- Then, an expected weighted average episode charge is calculated. The actual weighted average episode charge is compared to the expected weighted average charge — and an efficiency score is calculated.

- Finally, this efficiency score is adjusted by DxCGs, ACGs, or another case-mix index. The problem is that these indices are all less than 30% predictive. With the Law of Low Episode

Numbers being the norm for a physician, these tools are not adequate to perform the case mix adjustment.

- Consequently, those physicians with a higher case-mix of patients may continue to be rated as inefficient as compared to a peer group.

***One Way to Minimize Patient Case Mix Adjustment Error?***

A key to improving the accuracy and reliability of efficiency and effectiveness measurement is to remove the confounding effect of patient case mix (or health status) differences between practitioners. CCGroup uses marketbaskets of the most common medical conditions for each specialty type (i.e., CCGroup Marketbasket System™). The following is a portion of the cardiology marketbasket. (Not all medical conditions are included in this example. SOI in the table means severity-of-illness.)

**4. CARDIOLOGY**

| Order Number | Medical Condition Number | SOI Level | Medical Condition Short Description | Market-basket Weight |
|---|---|---|---|---|
| 1 | 10.13 | 1 | Ischemic heart disease | 0.150 |
| 2 | 10.13 | 2 | Ischemic heart disease | 0.050 |
| 3 | 17.4 | 1 | Diabetes with circulatory | 0.050 |
| 4 | 17.4 | 2 | Diabetes with circulatory | 0.050 |
| 5 | 10.21 | 1 | Acute myocardial infrct, active | 0.075 |
| 8 | 10.5 | 1 | Supraventricular arrhythmias | 0.050 |
| 9 | 10.4 | 1 | Ventricular arrhythmias | 0.050 |
| 10 | 10.1 | 1 | Abnormal heart beat | 0.050 |
| 12 | 36.19 | 1 | Chest pain | 0.050 |
| 13 | 10.10 | 1 | Conduction disorders | 0.050 |
| 14 | 10.16 | 1 | Congestive heart failure | 0.050 |
| 15 | 10.17 | 1 | Cardiomyopathy | 0.050 |
| 18 | 10.8 | 1 | Angina pectoris | 0.025 |
| 19 | 10.12 | 1 | Rheumatic heart disease | 0.025 |
| 20 | 36.18 | 1 | Dyspnea | 0.025 |
| | | | For all 20 medical conditions | 1.000 |

The Marketbasket System examines only common medical condition episodes for a particular specialty type. This results in a fair apples-to-apples comparison of each practitioner's practice patterns. Under this approach, patient case mix (or health status) differences are significantly controlled as compared to examining all episodes of care assigned to a physician, and then applying a commercially available case-mix adjustment tool. Therefore, the variation in practice patterns is related to actual practitioner efficiency, and not to sicker or healthier patients.

Medical conditions are placed in a specialty-specific marketbasket if they are a prevalent part of the particular specialty type's practice — generally accounting for 75% to 90% of the episodes

treated by that specialty type. The medical conditions are selected for the marketbasket in work effort order—a function of the prevalence rate and average medical condition charges.

> Medical conditions are placed in a specialty-specific marketbasket if they are a prevalent part of the particular specialty type's practice—generally accounting for 75% to 90% of the episodes treated by that specialty type.
>
> Therefore, the specialty-specific marketbaskets comprise a large percent of a PCP's and specialist's practice and are very representative of the patients treated in their given practice area.

***Summary to Patient Case-Mix Adjustment***

A solution to controlling for episode patient case-mix in physician efficiency and effectiveness analysis is <u>not</u> the following. You should not assign all non-outlier episodes to a physician and then apply a case-mix adjustment tool, such as DxCGs or ACGs, to control for the significant remaining differences in episode case mix. With the "Law of Low Episode Numbers" per physician, all available patient case-mix adjustment tools are not sensitive enough, as the health services literature well defines that up to 70% of the patient case-mix differences may remain.

Instead, the Marketbasket System builds a marketbasket of prevalent medical conditions, and examines a consistent set of episodes by severity-of-illness level from these medical conditions. In effect, this is the CCGroup case-mix adjustment tool. This technique removes patient case mix background noise that cannot be adjusted with any case mix tool. Another way to think about this is that CCGroup has developed a more robust outlier episode analysis whereby only a defined large grouping of condition-specific episodes are examined for a physician as compared to the peer group.

**Issue #2: Inadequate Standard Deviation Statistical Bias Adjustment**

Practitioners of a specialty type generally treat similar medical conditions. However, for a given health plan and with the Law of Low Episode Numbers, many specialists (of a given specialty type) will have a significantly different set of condition-specific episodes with a marketbasket of conditions as compared to the other specialists.

For example, one cardiologist's assigned episodes may be a quite homogeneous set of routine hypertension episodes. On the other hand, another cardiologist's assigned episodes may be the severity-of-illness two (SOI-2) ischemic heart disease patients. This second cardiologist has a patient population that is very "heterogeneous" as compared to the first cardiologist. With the Law of Low Episode Numbers, we know the second cardiologist will have a significantly greater standard deviation around this cardiologist's mean episode charge.

Health services research shows us that there may be a significant statistical bias present when two cardiologists have the chance for a different standard deviation around their mean episode charge. As I will demonstrate below, this statistical bias disadvantages those cardiologists (and any other practicing PCP or specialist) that have an assigned episode composition of a higher case mix. These physicians with a higher case mix episode composition have a greater probability of being rated as inefficient, as compared to physicians that have an assigned episode composition of a lower case mix.

***Statistical Bias Results From Direct Standardization***

Many efficiency methodologies examine a practitioner's actual episode composition as compared to a specialty-specific peer group —and then compare the efficiency of that practitioner to another practitioner. This is called "direct standardization." Under direct standardization, each physician's episode distribution is applied to determine that physician's observed and expected efficiency results.

However, in using direct standardization, the differences in each practitioner's patient case-mix composition results in differences in variability (i.e., the standard deviation) around a practitioner's average episode treatment charges. This variability is not due to the efficiency or inefficiency of a practitioner, but instead results because longer and more resource-intensive medical conditions generally require more services and, therefore, have more potential variability around the average (or mean) episode treatment charges. Moreover, more resource-intensive conditions generally have a lower prevalence rate, also contributing to the variability, or heterogeneity, around the average episode treatment charges.

For example, easier-to-treat upper respiratory infection (URI) episodes may have the following mean and standard deviation (with outlier episodes removed): $185 $\pm$ $65. Here, the standard deviation around the mean is not large—and is 0.35 the size of the mean (i.e., 65 / 185 = 0.35).

However, easier-to-treat pediatric asthma episodes may have the following mean and standard deviation (with outlier episodes removed): $1,650 $\pm$ $850. Here, the standard deviation around the mean is larger than for URI episodes—and is 0.52 the size of the mean (i.e., 850 / 1,650 = 0.52).

The variation difference between the two conditions is 49% greater for asthma than URIs [(0.52 - 0.35) / 0.35]. This variation difference occurs for two reasons: (1) more resource-intensive conditions require more services to treat; and (2) there generally are a small number of episodes available to examine in a given practitioner efficiency study as compared to the universe of episodes that could actually be studied—and a smaller number of episodes results in a higher chance for variability around the mean. (On the other hand, this variation is not the result of practitioner practice pattern differences.)

If the statistically based variability around the mean is not corrected, then substantial error may enter into the practitioner efficiency measurement equation. Consequently, the final practitioner efficiency score differences may be attributed to the statistical condition-specific variability around the mean episode charge (due only to the case-mix of episodes treated).

We showed above that the variation difference may be 50% or more (around a condition-specific mean episode value). Logically, then, if we examine all episodes treated by practitioners and calculate efficiency scores, there has to be some statistical bias present. A significant statistical bias generally is present:

- Using a more traditional episode-based efficiency measurement methodology, lower-episode-volume practitioners treating patients with a higher case-mix index score are more likely to receive an inefficient ranking as compared to lower-episode-volume practitioners treating patients with a lower case-mix index score.

- This finding results because a physician with higher case-mix patients treats episodes having more variability (i.e., a greater standard deviation) around average episode treatment charges. With a low volume of episodes (most often the norm, and not the exception), this physician needs only a few higher-cost episodes then the peer group average to make his/her treatment pattern appear significantly higher than the peer group comparator.

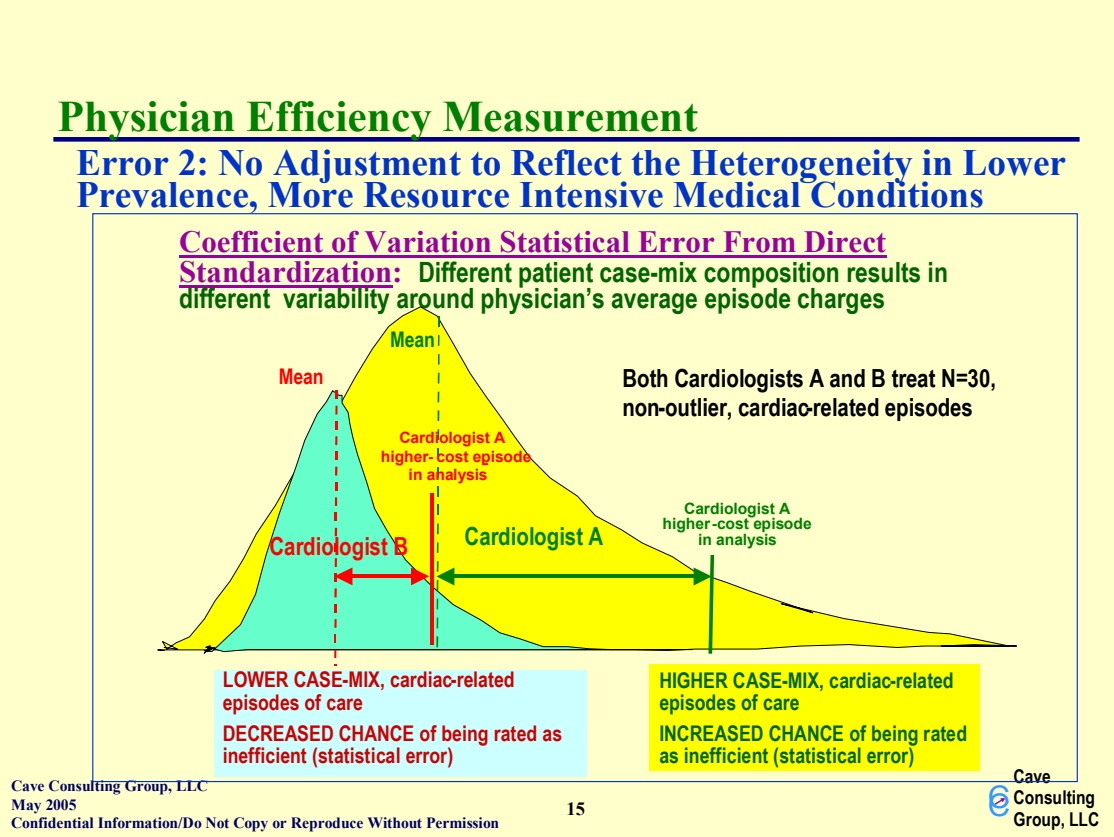
- However, a physician with lower case-mix patients treats episodes having less variability around average episode treatment charges. With a low volume of episodes, this physician's practice pattern will not be as influenced by one or two higher-cost episodes as compared to the peer group average. Consequently, his/her practice pattern does not appear (as often) significantly higher than the peer group comparator.

Using "direct standardization," a correlation analysis shows that lower-volume practitioners with a higher patient case-mix index for episodes treated are more likely to receive an inefficient score as compared to practitioners with a lower patient case-mix index.

***Comparing Results From Direct Standardization and Indirect Standardization***

The following graphical insert shows the medical condition episodes treated by Physician A and Physician B (both cardiologists). Notice that Physician A treats a higher case mix episode composition than Physician B. Also, notice the significant variance around Physician A's mean episode charges — attributed to Physician A's higher case mix patient population.

Next, we examine the above Physician A's actual weighted average episodes of care using "direct standardization and "indirect standardization". Whereby direct standardization uses each physician's episode distribution to comprise observed and expected efficiency results, indirect standardization uses an outside standard set of weights that is applied to comprise observe and expected efficiency results. This set of weights is applied in the same exact manner for all physicians of a given specialty type in determining the weighted average episode charges. Most often, the indirect standardization set of weights is formulated using the peer group's episode distribution.




Physician A's actual results using direct standardization are found in the top table immediately below. Physician A's actual results using indirect standardization are found in the bottom table immediately below. The example is fairly self-evident:

- You will observe that by using direct standardization, the cardiologist receives an efficiency score of 1.29. However, we will determine that this score is not accurate or reliable, and is based on only one SOI-2 episode of ischemic heart disease – which has significant heterogeneity.

- On the other hand, you will observe that by using indirect standardization, this same cardiologist's efficiency score is 1.04. This score is more accurate and more reliable. Indirect standardization controls for the significant heterogeneity found in the one non-outlier episode of SOI-2 ischemic heart disease assigned to the cardiologist.

- Note that only Column 2 in the below tables changes, which is the weights used to comprise the weighted average episode charges across all assigned episodes of care.

**Cardiologist A's Efficiency Results (Patient Case-Mix Index = 1.25)**

**DIRECT STANDARDIZATION: PHYSICIAN A**
**Efficiency Score = 1.29 ($4,491 / $3,473) Non-Outlier Episodes = 32 Episodes**
**One (1) Ischemic Heart Disease SOI-2 Patient = $30,000 *Case Mix Index = 1.25***

| SOI Level (1) | Medical Condition Short Description (2) | Physn A's *Actual* Episode Distribution (3) | Physn A's *Actual* Episode Average Charges (4) | Physn A's *Actual* Weighted Average Charges (5) = (4)x(3) | Peer Group Episode Average Charges (6) | Physn A's *Expected* Weighted Average Charges (7) = (6)x(3) |
|---|---|---|---|---|---|---|
| 1 | Ischemic heart disease | 0.300 | $2,450 | $735 | $2,550 | $765 |
| 2 | Ischemic heart disease | 0.125 | $14,500 | $1,813 | $5,500 | $688 |
| 1 | Diabetes with circulatory | 0.200 | $3,950 | $790 | $4,100 | $820 |
| 2 | Diabetes with circulatory | 0.125 | $8,250 | $1,031 | $8,500 | $1,063 |
| 1 | Hypertension | 0.250 | $490 | $123 | $550 | $138 |
| --- | **Overall Sum** | **1.000** | --- | **$4,491** | --- | **$3,473** |

**INDIRECT STANDARDIZATION: PHYSICIAN A**
**Efficiency Score = 1.04 ($2,962 / $2,846) Non-Outlier Episodes = 32 Episodes**
**One (1) Ischemic Heart Disease SOI-2 Patient = $30,000 *Case Mix Index = 1.25***

| SOI Level (1) | Medical Condition Short Description (2) | Indirect Standardization Episode Distribution (3) | Physn A's *Actual* Episode Average Charges (4) | Physn A's *Actual* Weighted Average Charges (5) = (4)x(3) | Peer Group Episode Average Charges (6) | Physn A's *Expected* Weighted Average Charges (7) = (6)x(3) |
|---|---|---|---|---|---|---|
| 1 | Ischemic heart disease | **0.300** | $2,450 | $735 | $2,550 | $765 |
| 2 | Ischemic heart disease | **0.025** | $14,500 | $363 | $5,500 | $138 |
| 1 | Diabetes with circulatory | **0.275** | $3,950 | $1,086 | $4,100 | $1,128 |
| 2 | Diabetes with circulatory | **0.075** | $8,250 | $619 | $8,500 | $638 |
| 1 | Hypertension | **0.325** | $490 | $159 | $550 | $179 |
| --- | **Overall Sum** | **1.000** | --- | **$2,962** | --- | **$2,846** |

CCGroup's approach to correct for the differences in variability around average episode charges examines a consistent set of medical conditions for each specialty type. The methodology uses indirect standardization for weighting together the episodes within the core group of medical conditions in a consistent fashion — thereby allowing each practitioner's efficiency performance to be more accurately compared to one another. That is, each practitioner now has the same chance for variability around his/her average episode treatment charges. The same standardized weights are applied, regardless of each practitioner's actual episode composition.

In effect, to eliminate, or greatly reduce, the variation statistical bias, each physician must have the same opportunity for the same heterogeneity (or variation) around his/her weighted average episode charges. As we will describe below, the Marketbasket System method with "indirect standardization" allows each physician of a given specialty type to have the opportunity for the same variation around his/her weighted average episode charges:

- The Marketbasket System efficiency measurement method is as follows. Each medical condition in a specialty-specific marketbasket is assigned a weight factor that reflects the importance or relevance of that medical condition to the marketbasket. The weight factors are used to compute the overall marketbasket weighted average value of a charge or utilization service category — across medical conditions — for a peer group or a physician ("indirect standardization"). The sum of the weight factors in a marketbasket equals 1.00.

- Therefore, regardless of a physician's (or physician group's) actual episode work effort, we standardize each physician's (or physician group's) actual work effort to a static set of weight factors – most often the peer group's episode composition. These weight factors represent the work effort that an average specialty-specific physician treats in medical practice — where work effort is a function of the prevalence rate and the average charges to treat an episode of care.

> Indirect standardization allows each physician (or physician group) of a given specialty type to have the "same opportunity" for the "same variation" around his/her weighted average episode charges.

Refer to the standardized weights in the last column of the following table, which shows many of CCGroup's copyrighted and patent-pending Family/General Practitioner Marketbasket medical conditions.

**1. FAMILY AND GENERAL PRACTITIONERS**

| Order Number | Medical Condition Number | SOI Level | Medical Condition Short Description | Market-basket Weight |
|---|---|---|---|---|
| 1 | 10.2 | 1 | Hypertension | 0.100 |
| 2 | 31.9 | 1 | Low back pain | 0.050 |
| 3 | 31.8 | 1 | Cervical spine pain | 0.025 |
| 11 | 9.7 | 1 | Pneumonia | 0.050 |
| 12 | 13.13 | 1 | Noninfect gastroent & colitis | 0.050 |
| 13 | 9.11 | 1 | Asthma | 0.025 |
| 14 | 10.13 | 1 | Ischemic heart disease | 0.025 |
| 15 | 36.19 | 1 | Chest pain | 0.025 |
| 16 | 13.6 | 1 | Gastroesophageal reflux | 0.025 |
| 17 | 13.5 | 1 | Gastritis and duodenitis | 0.025 |
| 18 | 19.4 | 1 | Disorders of lipid metabolism | 0.025 |
| 19 | 16.3 | 1 | Hypothyroidism | 0.025 |
| 20 | 22.3 | 1 | Urinary tract infections | 0.025 |
| 21 | 34.17 | 1 | Nonpsychotic depression | 0.025 |
| 27 | 36.17 | 1 | Abdominal pain | 0.020 |
| 28 | 29.6 | 1 | Skin keratoses | 0.015 |
| 29 | 29.1 | 1 | Ill-defined integument sym | 0.015 |
| 30 | 36.15 | 1 | General presenting symptoms | 0.015 |
| | | | For All 30 Medical Conditions | 1.000 |

### *Summary to Standard Deviation Statistical Bias Adjustment*

A solution to controlling for standard deviation statistical bias is to employ "indirect standardization" when calculating a practitioner's average weighted episode charges. Otherwise, a statistical bias will remain because of the "Law of Low Episode Numbers" and the fact that physicians of a particular specialty type with a higher episode case mix have significantly more heterogeneity (or a larger standard deviation) around their weighted average episode charges. This heterogeneity causes a standard deviation statistical bias, which unfairly scores those physicians as inefficient that have a higher episode case mix.

### Issue #3: Inadequate Episode of Care Severity-of-Illness adjustment

Severity-of-illness is defined as the probability of loss of function due to a specific medical condition. Some claims-based episode groupers and methods do not have a severity-of-illness index by medical condition. This issue is the third most important factor leading to efficiency measurement error because the episodes for a given medical condition have significant heterogeneity due to patient severity-of-illness. Consequently, the end result may be practitioner efficiency differences that are attributed to inaccurate episode severity-of-illness adjustment—and not to practitioner practice patterns variation.

Moreover, some claims-based episode groupers stratify formulated episodes for a medical condition by the presence or absence of a specific surgery or service (e.g., knee derangement with and without surgery; ischemic heart disease with and without heart catheterization). The

reason for performing this stratification is to reduce episode heterogeneity for a medical condition. In effect, the stratification serves as a sort of severity-of-illness adjustment.

However, the presence of surgery or a high-cost service results is not necessarily a reflection of differences in severity-of-illness, but often a difference in practice patterns. Stratification based on the presence of surgery or a procedure may result in at least two practitioner efficiency measurement errors:

- Obscure the practice pattern variation we need to examine to determine practitioner efficiency and effectiveness differences; and

- Divide episodes unnecessarily into smaller episode groups whereby practitioners may not have enough episodes to examine in any one smaller group.

Consequently, the stratified episodes based on surgery, or a high-cost event, need to be recombined for accurate practitioner efficiency and effectiveness measurement.

The following table shows an example of a procedure-based grouper (based on searching for specific CPT-4 and/or HCPCS codes to stratify episodes of care by medical condition):

- The columns on the right-hand side of the table show that this orthopedist had 100 non-outlier episodes of low back pain. Seventy (70) of these episodes had some type of procedure.

- When we examine those 70 episodes with a procedure, the results show that the physician's observed average charge per episode was $2,765. The physician's expected average charge per episode (based on the peer group of orthopedist results) was $3,568. Therefore, this orthopedist's efficiency score is 0.77 (or $2,765 divided by $3,568) – or the orthopedist appears to be a very efficient physician.

- However, this is contradictory to what we expect because the physician performed a procedure on many low back pain patients. The reason is as follows. Once you stratify based on a procedure, you can only study the practice pattern within those low back pain episodes that had a procedure. In this example, the orthopedist performed a procedure on many marginal patients (i.e., those that may not have needed a procedure). Consequently, the healthier patients with a procedure resulted in the orthopedist having a significantly lower average charge than the peer group. The result is that this orthopedist appears efficient.

This result is not accurate or reliable, and shows why an episode-of-care grouper cannot be based on the presence or absence of a procedure. Otherwise, the consumer may be inadvertently guided to the less efficient and less effective practitioners, when the actual objective of consumer health information is to direct consumers to the more efficient and more effective practitioners.

| Area of Interest | Diagnosis Code-Based Grouper (Based on ICD.9 Codes) | | | Procedure Code-Based Grouper (Based on CPT-4 Codes) | |
|---|---|---|---|---|---|
| | Assume an Orthopedist has 100 episodes of Low Back Pain, and 70 of these episodes have a surgical event present (and, therefore, 30 episodes do not have a surgical event present).* | | | | |
| | Epsiode Distribution | | | Episode Distribution | |
| | SOI-1 | SOI-2 | SOI-3 | With Surgery | Without Surgery |
| Episode Distribution | 86 | 12 | 2 | 70 | 30 |
| Physician Observed Weighted Avg Results | $752 | $2,143 | $10,826 | $2,765 | $280 |
| Physician Expected Weighted Avg Results | $308 | $3,750 | $14,598 | $3,568 | $297 |
| Efficiency Score by Cell | 2.44 | 0.57 | Not examine because prevalence too low | 0.77 | 0.94 |
| Overall Efficiency Score | 2.21 = Q4 for Low Back Pain | | | 0.84 = Q1 for Low Back Pain | |

**Note: A surgical event may be a spinal fixation, spinal decompression, or a more routine surgery such as a spinal manipulation. Not all surgeries are significantly invasive and high cost.**

**Note: This orthopedist provides surgeries on many healthier low back pain patients. Consequently, the physician's average charges per episode are lower within the "With Surgery" procedure class.**

### *An Appropriate Method for Episode Severity-of-Illness Adjustment*

Accurate practitioner efficiency and effectiveness measurement requires an episode grouper that has a valid and tested severity-of-illness index by medical condition to reduce the heterogeneity in longitudinal episodes of care.

The grouper, such as the Cave Grouper™, should use a severity-of-illness index based only on ICD.9 diagnosis codes to assign a patient's episode with a severity-of-illness marker. On the other hand, the grouper should not define severity-of-illness by resource utilization within the patient's condition-specific episode (such as whether a surgery or a resource-intensive diagnostic test was present in the patient's episode of care).

Using this diagnosed code-based severity index, each patient's condition-specific episode is labeled with a severity-of-illness marker to reduce the heterogeneity of episodes within a medical condition. There may be three (3) or more severity-of-illness (SOI) levels for a medical condition, with SOI-1 being the least severe (routine, noncomplicated) and SOI-3 being the most severe. Some medical conditions have only one or two severity-of-illness levels, depending on the specificity of the ICD.9 coding for that medical condition of interest.

The above table shows an example of a diagnosis-based grouper (based on searching for specific ICD.9 diagnosis codes to stratify episodes of care by medical condition):

- The columns on the left-hand side of the table show that this orthopedist had 100 non-outlier episodes of low back pain. There were 86 episodes that fell into severity-of-illness Level 1 (SOI-1).

- When we examine those 86 episodes in SOI-1, the results show that the physician's observed average charge per episode was $752. The physician's expected average charge per episode (based on the peer group of orthopedist results) was $308. Therefore, this orthopedist's efficiency score is 2.44 (or $752 divided by $308) – or the orthopedist appears to be a very inefficient physician.

- This result is the appropriate, accurate result. The physician is providing too many procedures as compared to the peer group comparator of orthopedists.

From a consumer health information perspective, employing a grouper based on diagnosis codes is important to guide the consumer to the appropriate efficient and effective practitioners. This more accurate information will help to ensure the consumer's health outcome will be improved.

## 7. Effectiveness Measurement

For accurate and reliable practitioner effectiveness measurement, a methodology also should be employed that adjusts for the three main methodology errors/issues defined above for practitioner efficiency measurement. One technique is to utilize the Marketbasket System approach that employs a specialty-specific standard set of medical condition episodes and indirect standardization.

Moreover, an established set of criteria needs to be developed for selecting effectiveness measures by specialty type (e.g., PCPs, cardiologists, orthopedists, and other specialty types). These criteria need to "fit" within the method used to adjust for the three main methodology error/issues in efficiency and effectiveness measurement. Since I have selected the Marketbasket System as a suggested adjustment approach, the criteria listed below will fit to the Marketbasket System methodology.

The following rules may be used to determine whether a measure may be included, or not included, in the effectiveness analysis:

1. The effectiveness measure needs to apply to a condition-specific, episode of care environment.

2. The effectiveness measure needs to be quantified and analyzed using medical claims data, understanding the limitations posed by claims data such as often missing data. Many potentially sound quality metrics may not be used because they cannot be quantified using claims data.

3. The effectiveness measure, although often process of care oriented, should be tied whenever possible to the outcomes of patient care; this is a National Quality Forum (NQF) criteria as well.

4. The effectiveness measure should be obtained from an expert source to ensure physician acceptance. CCGroup uses four main sources to evaluate potential quality metrics by physician specialty type: (1) RAND Health Quality of Care Reports, a meta analysis of the clinical literature; (2) AHRQ/AHCPR condition-specific guideline measures; (3) NQF and Ambulatory Care Quality Alliance (AQA); and (4) NCQA.

5. The effectiveness measure should apply to a medical condition treated by various specialty types and, therefore, specialty-specific marketbaskets (e.g., quality metrics for hypertension apply to family/general practitioners, general internists, and cardiologists).

6. The effectiveness measure should be from a higher prevalence condition so the individual physicians may be compared to one another. CCGroup generally avoids lower prevalence conditions whereby most physicians cannot be compared on a condition-specific basis.

CCGroup currently maintains 108 different condition-specific effectiveness measures that meet the above six criteria. However, this number may differ based on the specific criteria being implemented to meet the needs of the methodology employed to adjust for the three main efficiency and effectiveness errors/issues. Many of the medical conditions with at least one effectiveness measure may be found in more than one specialty-specific marketbasket (a criterion listed above). Consequently, the 108 measures are found across many different specialty-specific marketbaskets.

CCGroup implements effectiveness and efficiency marketbaskets of medical conditions for each of the following 31 physician specialty types. Of the 31 specialty-specific marketbaskets, CCGroup maintains one or more effectiveness measure for 29 of these specialty-specific marketbaskets; only dermatology and oral maxillary do not have one or more measure(s). Note in the following table that cardiology has 48 measures, cardiothoracic surgery has 36 measures, endocrinology has 20 measures, etc.

**Marketbasket Specialty Types for Efficiency and Effectiveness**

| Market-Basket Number | Marketbasket Specialty Type | Physician Efficiency Measurement | Physician Effectiveness Measurement | Current Process Measures |
|---|---|---|---|---|
| 1 | Family and General Physicians | Yes | Yes | 51 |
| 2 | General Internists | Yes | Yes | 51 |
| 3 | Allergy | Yes | Yes | 10 |
| 4 | Cardiology | Yes | Yes | 48 |
| 5 | Cardiothoracic Surgery | Yes | Yes | 36 |
| 6 | Chiropractic | Yes | Yes | 26 |
| 7 | Dermatology | Yes | No | 0 |
| 8 | Emergency Medicine | Yes | Yes | 37 |
| 9 | Endocrinology | Yes | Yes | 20 |
| 10 | Gastroenterology | Yes | Yes | 8 |
| 11 | General Surgery | Yes | Yes | 11 |
| 12 | Nephrology | Yes | Yes | 16 |
| 13 | Neurology | Yes | Yes | 12 |
| 14 | Neurosurgery | Yes | Yes | 14 |
| 15 | Obstetrics/Gynecology (OB/GYN) | Yes | Yes | 17 |
| 16 | Oncology/Hematology | Yes | Yes | 17 |
| 17 | Ophthalmology | Yes | Yes | 15 |
| 18 | Oral Maxillary | Yes | No | 0 |
| 19 | Orthopedics | Yes | Yes | 32 |
| 20 | Otolaryngology (ENT) | Yes | Yes | 10 |
| 21 | Pediatrics | Yes | Yes | 17 |
| 22 | Plastic Surgery | Yes | Yes | 5 |
| 23 | Podiatry | Yes | Yes | 16 |
| 24 | Psychiatry | Yes | Yes | 10 |
| 25 | Psychology | Yes | Yes | 8 |
| 26 | Pulmonology | Yes | Yes | 21 |
| 27 | Rheumatology | Yes | Yes | 21 |
| 28 | Sports/Physical Medicine | Yes | Yes | 27 |
| 29 | Urology | Yes | Yes | 11 |
| 30 | Vascular Surgery | Yes | Yes | 15 |
| 31 | Critical Care (Intensivist) | Yes | Yes | 21 |

By utilizing the Marketbasket System approach, certain additional consumer health information objectives are achieved:

- First, every physician that receives an efficiency score should also receive an effectiveness score. This is important because a consumer will desire to understand both the practitioner's efficiency and effectiveness of care.

- Second, the same episodes examined in the efficiency analysis may be examined in the effectiveness analysis. Therefore, a consumer may more easily understand the basis for the practitioner's efficiency and effectiveness scores. I will present some practitioner efficiency and effectiveness results in the next section that illustrate how this information may be used to support consumer health information initiatives.

## 8. Examples of Consumer Health Information on Practitioners

One of the most important components of consumer health information is to ensure accurate and reliable efficiency and effectiveness measurement methodologies are employed in the measurement process. Otherwise, consumers may be inadvertently guided to the less efficient and less effective practitioners, when the actual objective of consumer health information is to direct consumers to the more efficient and more effective practitioners. The end result is that we will miss the significant efficiency and effectiveness improvements that may be realized from providing consumers with meaningful health information.

This material presents example results for a cardiologist using the Marketbasket System approach to physician efficiency and effectiveness measurement, which is one method for building accurate and reliable scores. Cardiologist XYZ has a lower effectiveness score and a lower efficiency score as compared to the cardiology peer group. The specific reasons for the lower effectiveness and efficiency scores are stated in the inserted tables below. In summary, Cardiologist XYZ may not be monitoring patients appropriately through effective lab and diagnostic testing, resulting in a higher hospital admission rate.

With respect to consumer health information, the health plan market is attempting to determine the most appropriate way to display this physician level (and physician group level) information to the consumer:

- A common method is to use a "star" system. For instance, Cardiologist XYZ would receive 1 Star for effectiveness, and 1 Star for Efficiency (where 1 Star indicates lower effectiveness and lower efficiency as compared to a peer group or best practice). Generally, feedback from consumer user groups has favored a simple starring system (or similar graphic display) to indicate practitioner efficiency and effectiveness.

- The next step is to determine a drill-down mechanism whereby the consumer can gain more meaningful information at the medical condition level. For example, in the overheads below, we observe that Cardiologist XYZ treats ischemic heart disease SOI-1 with lower efficiency and lower effectiveness than the peer group of cardiologists. By providing this information to the consumer, the consumer may be guided away from Cardiologist XYZ and towards cardiologists with better efficiency and effectiveness results.

This type of information is valuable in terms of improving the consumer's potential short-term and longer-term health status, while simultaneously improving the efficiency of the U.S. healthcare system.

In the following overheads, please observe that the same medical condition-specific episodes are being examined for the effectiveness and efficiency analyses. For example, Cardiologist XYZ has 8 episodes for ischemic heart disease, severity-of-illness one (SOI-1). There are 6 episodes for ischemic heart disease SOI-2. You will observe the results for these 8 episodes and 6 episodes on both the effectiveness and efficiency reports.

The marketbasket for cardiology and the indirect standardization weights (for efficiency and effectiveness measures) are provided in the following slide.

## Physician Effectiveness and Efficiency Results

### Marketbasket for Cardiologists

**4. CARDIOLOGY (10 of 20 Marketbasket Conditions Have Quality Metrics)**

| Order Number | Medical Condition Number | SOI Level | Medical Condition Short Description | Market-basket Weight | ≥1 Quality Process Measure? | Quality Marketbasket Weight |
|---|---|---|---|---|---|---|
| 1 | 10.13 | 1 | Ischemic heart disease | 0.150 | Yes | 0.200 |
| 2 | 10.13 | 2 | Ischemic heart disease | 0.050 | Yes | 0.075 |
| 3 | 17.4 | 1 | Diabetes with circulatory | 0.050 | Yes | 0.100 |
| 4 | 17.4 | 2 | Diabetes with circulatory | 0.050 | Yes | 0.075 |
| 5 | 10.21 | 1 | Acute myocardial infrct, active | 0.075 | Yes | 0.125 |
| 8 | 10.5 | 1 | Supraventricular arrhythmias | 0.050 | No | |
| 9 | 10.4 | 1 | Ventricular arrhythmias | 0.050 | Yes | 0.100 |
| 10 | 10.1 | 1 | Abnormal heart beat | 0.050 | No | |
| 12 | 36.19 | 1 | Chest pain | 0.050 | No | |
| 13 | 10.10 | 1 | Conduction disorders | 0.050 | No | |
| 14 | 10.16 | 1 | Congestive heart failure | 0.050 | Yes | 0.100 |
| 15 | 10.17 | 1 | Cardiomyopathy | 0.050 | No | |
| 18 | 10.8 | 1 | Angina pectoris | 0.025 | Yes | 0.050 |
| 19 | 10.12 | 1 | Rheumatic heart disease | 0.025 | No | |
| 20 | 36.18 | 1 | Dyspnea | 0.025 | No | |
| | | | For all 20 medical conditions | 1.000 | | 1.000 |





Cave Consulting Group, LLC

## Physician Effectiveness and Efficiency Results

### Cardiologist #1: Poor Effectiveness and Poor Efficiency

**[Practitioner ID = XYZ]**

**Effectiveness Score = 0.86 (Quartile 4)** Higher is better
**Efficiency Score = 1.11 (Quartile 4)** Lower is better

- Poor effectiveness results influenced by:
  - Lower serum cholesterol monitoring: ischemic, diabetes
  - Lower IHD labs monitoring: ischemia heart, angina
  - Lower urine protein monitoring: diabetes
  - Lower HTN lab monitoring: hypertension
  - Lower diagnostic test monitoring: CHF, arrhythmias
- Poor efficiency results influenced by:
  - Higher hospital admissions
  - Higher professional inpatient services
  - Higher outpatient facility usage

Note: lower lab services than peers





Cave Consulting Group, LLC

# Physician Effectiveness and Efficiency Results

## Cardiologist #1: Poor Effectiveness Results

**Practitioner Name:**
**Specialty Type:** CARDIOLOGIST
**Practitioner ID:** XYZ
**Agg Group Name:** Region 1
**Marketbasket:** Cardiology

**Quartile:** 4 (Less effective)
**Decile:** 10
**Effectiveness Score:** 0.86 (Higher is better)
**Significant Difference:** Yes (P<0.25)

*Effectiveness Per Episode of Care*

| Medical Condition Name | SOI | Episode Count | Weighted Average Rate | Effectiveness Measure of Interest: Measure | Description of Measure | Protocol Range | Your Rate Results | Peer Group Rate Results | Your Score: [You / Peer] | Your Rate Results in Protocol Range? |
|---|---|---|---|---|---|---|---|---|---|---|
| Peer Group Weighted Avg | – | 4911 | 0.68 | – | – | – | – | – | – | – |
| Practitioner Weighted Avg | – | 45 | * 0.58 | – | – | – | – | – | – | – |
| Ischemic heart disease | 1 | | | Measure 1 | Episodes with semi-annual check-up exam | 70% - 100% | 0.88 | 0.95 | 0.92 | Yes |
| Peer Group Rate: | | 1305 | 0.62 | Measure 2 | Episodes with anemia lab monitoring | 35% - 70% | 0.38 | 0.26 | 1.42 | Yes |
| Practitioner Rate: | | 8 | 0.58 | Measure 3 | Episodes with serum cholesterol monitoring | 35% - 70% | * 0.25 | 0.47 | 0.53 | No |
| Your Score [ You / Peer ]: | | | 0.93 | Measure 4 | Episodes with IHD lab monitoring | 35% - 70% | 0.63 | 0.58 | 1.08 | Yes |
| | | | | Measure 5 | Episodes with IHD/cardiac test monitoring | 35% - 70% | 0.75 | 0.82 | 0.92 | No |
| Ischemic heart disease | 2 | | | Measure 1 | Episodes with semi-annual check-up exam | 70% - 100% | * 0.83 | 0.97 | 0.86 | Yes |
| Peer Group Rate: | | 798 | 0.68 | Measure 2 | Episodes with anemia lab monitoring | 35% - 70% | 0.33 | 0.38 | 0.87 | No |
| Practitioner Rate: | | 6 | 0.53 | Measure 3 | Episodes with serum cholesterol monitoring | 35% - 70% | * 0.17 | 0.56 | 0.30 | No |
| Your Score [ You / Peer ]: | | | 0.78 | Measure 4 | Episodes with IHD lab monitoring | 35% - 70% | * 0.33 | 0.67 | 0.50 | No |
| | | | | Measure 5 | Episodes with IHD/cardiac test monitoring | 35% - 70% | 1.00 | 0.85 | 1.18 | No |





Cave Consulting Group, LLC

# Physician Effectiveness and Efficiency Results

## Cardiologist #1: Poor Efficiency Results

**Practitioner Name:**
**Specialty Type:** CARDIOLOGIST
**Practitioner ID:** XYZ
**Agg Group Name:** Region 1
**Marketbasket:** Cardiology

**Quartile:** 4 (Less efficient)
**Decile:** 8
**Efficiency Score:** 1.11 (Lower is better)
**Significant Difference:** Yes (P<0.25)

***Marketbasket:*** CARDIOLOGY

*Average Utilization Per Episode of Care*

| Medical Condition Name | SOI | Episode Count | Average Episode Duration (Days) | Professional Outpt and Ambulatory: Prof Visits | Diag Tests | Lab/ Path | Med/ Surg | Rx | Prof Inpt (Svcs) | Facility: Outpt (Visits) | Hosp Inpt: Admits | Hosp Inpt: Days | Altern Sites (Svcs) | Other Med (Svcs) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peer Group Weighted Avg | – | 13243 | 160.8 | 2.45 | 4.34 | 2.34 | 0.90 | 8.16 | 3.29 | 0.22 | 0.20 | 0.78 | 0.00 | 2.16 |
| Practitioner Weighted Avg | – | 137 | * 163.7 | 2.46 | 4.24 | * 1.84 | 0.94 | * 8.70 | * 3.73 | * 0.48 | * 0.22 | * 0.88 | 0.00 | 1.90 |
| Ischemic heart disease | 1 | 8 | 180.0 | 3.25 | 7.88 | 1.63 | 0.75 | 8.88 | 2.13 | 1.38 | 0.13 | 0.38 | 0.00 | 1.88 |
| Ischemic heart disease | 2 | 6 | 180.0 | 2.83 | 4.50 | 0.33 | 0.33 | 7.17 | 11.00 | 1.00 | 0.50 | 2.33 | 0.00 | 1.17 |
| Diabetes with circulatory | 1 | 4 | 180.0 | 5.50 | 9.00 | 5.25 | 2.25 | 24.50 | 0.00 | 1.25 | 0.00 | 0.00 | 0.00 | 4.75 |
| Diabetes with circulatory | 2 | 1 | 180.0 | 4.00 | 3.00 | 0.00 | 7.00 | 17.00 | 18.00 | 0.00 | 1.00 | 4.00 | 0.00 | 0.00 |
| Acute myocardial infrct, active | 1 | 0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Acute myocardial infrct, fup | 1 | 0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

NOTE: The efficiency score is calculated by dividing the physician's marketbasket weighted average charges per episode by the peer group's marketbasket weighted average charges per episode. The corresponding Physician Efficiency Charge Report (not presented here) shows the weighted averages as follows: Physician = $4,590 per episode; Peer Group = $4,148 per episode. The "efficiency score" equals 1.11 — shown in the above heading.





Cave Consulting Group, LLC

## 9. Budget-Neutral Opportunity for Congress to Assist in Promoting Consumer Health Information

One of the highest-priority initiatives for Congress to assist in promoting consumer health information is to provide health plans and health insurance companies with access to the full CMS Part A and Part B Medicare claims databases, holding back only to the extent necessary to protect Medicare beneficiary privacy. Initial testing shows the CMS databases may be able to measure the efficiency and effectiveness of 80% of "all" (not just the health plan's network physicians) practicing physicians in a geographic region. This percent of physicians measured is more than enough to provide consumers with meaningful, physician-level efficiency and effectiveness information.

With respect to the full Medicare claims databases, current CMS rules restrict access to research studies that generally benefit the Medicare and Medicaid programs. However, in the view of most external legal experts, wider access to the full CMS claims data is not restricted by the statutory language of HIPAA or the Privacy Act – provided Medicare beneficiary privacy is protected.

Congress should clarify the HIPAA and Privacy Act regulations with CMS, and encourage CMS to revise its regulations to public access to the full CMS Medicare claims databases. The Medicare beneficiary identifiers must be encrypted for full protection of beneficiary privacy; other beneficiary protections can also be addressed to meet HIPAA regulations.

Without release of the full Medicare claims databases, many health plans and health insurance companies will fall short of their objective of providing consumers with physician-level efficiency and effectiveness information. Yet, this information is a requirement to drive competition in the healthcare system. Consumers need physician-level information to choose based on efficiency, effectiveness, and convenience. This fact is no different than in any other competitive industry.