BCBS Louisiana Physician Assessment

A Physician’s Explanation for Physicians

Seeing the big picture to solve the biggest problems in health care.

see more. solve more.
This presentation will review issues surrounding physician performance assessment and the approach Blue Cross Blue Shield of Louisiana uses to address those issues. We will review the building blocks for assessing performance, episode treatment groups (ETGs) and evidence based medicine (EBM) rules, as well as important concepts for defining comparison groups, using standard pricing to adjust for contractual differences in reimbursement, and attribution to physicians. Finally, we will address the physician comparison process and different methodologies for creating validity in that process.
To Be Covered

• Challenges of assessing physician performance
• Building blocks for physician performance assessment
  > ETGs
  > EBM rules
• Important concepts in the BCBSLA approach
  > Comparison groups: peer definitions
  > Standard pricing
  > Attribution
• Physician comparison
  > Creating validity
    • Episode severity adjustment
    • Episode outlier adjustment
    • Episode case-mix adjustment
    • Confidence intervals
Challenges of Assessing Physician Performance

• The availability of the right kind of data to make assessments
• Having enough data from a physician to make an accurate assessment
• Making comparisons that make sense from a clinical point of view
  > How to create clinically meaningful units of comparison
  > Defining who is compared to whom
  > Defining the components of comparison
Physician performance assessment presents a number of challenges. First, does data exist that can be turned into information about physician performance. This is not a trivial problem that affects all work involving understanding healthcare issues. The second challenge is having enough data to provide an accurate assessment for an individual physician, one that is statistically meaningful. The final challenge is to make comparisons that represent real clinical circumstances, of which there are three components; to create units of comparison that are homogenous and clinically meaningful, to define who should be compared to whom, and to define exactly what components one will use to make the comparisons.
How BCBSLA Addresses the Challenges

• Uses claims data, the most complete look at healthcare available in a fractured data environment

• Uses episodes of care as clinically meaningful units of comparison

• Comparisons are to peers defined by BCBSLA
  > Individuals and groups are compared to like specialties
  > Compared episodes of care and evidence based measures of quality are specific for each group of peers
Blue Cross Blue Shield of Louisiana uses a tool and approach that addresses these challenges in the following ways:

1. The “right kind of data” challenge is addressed by using claims data, which, in the fractured data environment that exists in healthcare, is the most complete look at healthcare that is available.

2. Uses episode treatment groups as the units of comparison. These are clinically meaningful units based on a condition.

3. Defines peer comparison groups that are specific to BCBSLA.

4. Compares individuals and groups to peers of like specialties.

5. Uses episodes and quality measures that are specific to each group of peers.
The Building Blocks: Episode Treatment Groups (ETG)

To understand the BCBS physician performance assessment approach, it is important to first understand the underlying components used to make the comparisons, which we will call the building blocks. The first of these is episode treatment groups, or ETGs.
ETG® Overview

• ETG is a basic illness classification methodology based on a clinical condition.

• A software application groups claims data into a medically meaningful statistical unit of analysis representing a patient’s complete episode of care.

• Claims Data Used:
  ➢ Inpatient
  ➢ Outpatient
  ➢ Professional
  ➢ Ancillary
  ➢ Pharmaceutical

• There are 456 basic episode treatment groups which are subdivided into over 1300 individual ETGs
Notes

> The basis for the ETG classification methodology is a clinical condition.

> The software searches through the longitudinal stream of claims to group subsets of claims into clinically meaningful statistical units. This subset of claims represents the complete episode of care for a clinical condition for a patient.

> These are the types of claims data that are used to create the 456 basic episode treatment groups.
ETG Methodology - Creating an Episode

- **Anchor records** (office visit, surgery) start or extend an episode.
- **Ancillary records** (x-rays, labs, pharmaceuticals) do not start or extend an episode.
- An episode is **complete** when an absence of treatment is detected for the condition for a period of time called a clean period.
- Different ETGs have different clean periods
  - Viral pneumonia – 90 days
  - Acute bronchitis – 30 days
- Chronic condition episodes (e.g., diabetes) are complete after 365 days; then another starts
• There are some important concepts to understand around the process of creating episodes. First, is that claims can be divided into two types, anchor records and ancillary records.

• An anchor record is a claim for a face to face visit (office visit, surgery, or hospitalization). This type of claim starts and/or extends an episode.

• All other claims are ancillary records and cannot start or extend an episode.

• Only complete episodes are used to make physician comparisons and an episode is considered complete if the coverage period covers the clean period before the start of the first anchor record and after the last anchor record. These clean periods occur when no evidence of treatment, in the form of an anchor record, for the condition is present.

• There are different clean periods for different episodes. For example, the clean period for viral pneumonia is 90 days and that for acute bronchitis is 30 days.

• Chronic episodes, for example, diabetes, are complete after 365 days and then a new episode starts.
Diagram of a Diabetes Episode

Unrelated services are not part of this episode

An anchor record and any affiliated services form an episode cluster

Diabetes episode Clusters (contain diabetes related services)
Here is how an episode of diabetes is created: An individual patient has a set of claims over time. These are of different types, as demonstrated by the symbols in the key.

For example, an office visit is represented by the sun symbol. An anchor record, along with associated claims that make up what is called an episode cluster, is noted by the ETG software with a diagnosis of diabetes.

This is the start of the episode. The software then identifies several other clusters relating to diabetes in the claims stream. After 365 days this chronic episode is complete. Note that there are also claims that occur throughout this period that are unrelated to diabetes. These claims are grouped into their own specific episodes.
The Quality Surrogate: EBM Rules

True quality outcomes cannot be obtained from claims data, of course, so quality is measured using the best surrogate possible from the claims data, evidence based medicine rules.
EBM Connect®

- EBM Connect combines evidence from published medical research with the observed clinical experience to identify and measure patient compliance with Quality of Care standards
  - Compliance is assessed at the patient level according to the condition and rule criteria for over 500 evidence based rules
  - Includes national standard measures and standards for care established by professional organizations
  - Member compliance results are attributed to physicians to assess Quality of Care aggregated to the physician or group level
  - Physician to peer group compliance comparison results in an Index of Quality for physicians or physician groups
• The Blue Cross Blue Shield of Louisiana Impact Intelligence tool uses a processing engine called EBM Connect. This engine uses algorithms developed to measure patient compliance with quality of care standards.

• This compliance is assessed at the patient level across over 500 evidence based rules, including national standard rules and others developed by professional health care and physician organizations.

• The patient compliance is aggregated to physicians providing the care to patients in order assess quality by physicians or groups or physicians

• Comparison to physician peers is used to develop a quality index.
Evidence Based Medicine Rules

• Over 270 of the EBM rules assigned to 16 specialty peer groups
• EBM attribution is based on attribution of the ETG for all except preventive rules
• Preventive rules are attributed to an imputed primary care provider
• Some specialty peer groups do not have EBMs
  > Dermatology
  > Chiropractors
• For Blue Cross Blue Shield of Louisiana, over 270 of the EBM rules are assigned to 16 different specialty peer groups

• The attribution of EBM rules is based on the attribution of the ETG for all except preventive rules, so that whoever is responsible for the ETG condition is responsible for the associated quality rules

• Preventive rules are attributed to the claims based imputed primary care provider

• Some specialty peer groups, such as dermatology and chiropractors, do not have associated EBM rules
Comparison Groups: Peer Definitions

Another important concept is peer definitions, which are necessary to be able to make comparisons.
Peer Definitions

• Are used to define the basis against which an individual or group of physicians is compared

• Consists of three components
  > Who is a member-For BCBS LA is a specialty, but can be defined with other defining characteristics, e.g. geography or product
  > What is compared—the set of ETGs and EBM rules that clinically make sense for comparing the members of the peer definition
  > How the comparison is done—Is it on the basis of episodes or population

Peer Definition Example: Endocrinology

Who = endocrinologists
What = 28 episodes and 40 EBM rules specific to endocrinology
How = episode based comparisons
Notes

- Peer definitions identify the parameters of who, how and what go into the groups that are used to make comparisons between individual and groups of physicians.

- The first component of a peer definition is who is a member. A major component of this is almost always specialty. There can be peer definitions with more than one specialty, however. An example of this for Blue Cross Blue Shield of Louisiana is the pediatrics peer definition has pediatricians as well as adolescent medicine specialists.

- Each peer definition includes exactly which ETGs and EBM rules are used for comparing members of that particular peer definition.

- Comparisons can be done on the basis of episodes or population, and this has to be delineated for each peer definition.

- Here you see an example of a Blue Cross Blue Shield peer definition, endocrinology, which is defined by having the specialty of endocrinology, 28 ETG types and 40 EBM rules specific to endocrinology, and that the comparisons are made on an episodes basis.
Here is the complete list of peer definitions for Blue Cross Blue Shield of Louisiana. Note that if a specialty is not defined as a part of one of these peer definitions listed, physicians with that specialty are not compared on their performance for quality and cost efficiency.

<table>
<thead>
<tr>
<th>Peer Definition Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP (Family)</td>
<td>Chiropractics</td>
</tr>
<tr>
<td>PCP (Internal Medicine)</td>
<td>General Surgery</td>
</tr>
<tr>
<td>PCP (Pediatric)</td>
<td>Invasive Cardiology</td>
</tr>
<tr>
<td>Cardiology</td>
<td>Obstetrics/Gynecology</td>
</tr>
<tr>
<td>Dermatology</td>
<td>Orthopedics</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>Otolaryngology</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>Urology</td>
</tr>
<tr>
<td>Hematology/Oncology</td>
<td>Facility</td>
</tr>
</tbody>
</table>
Standard Pricing

Standard pricing is used to level the playing field in terms of contractual relationships and their impact on measurements of cost efficiency.
Standard Pricing

• Standard pricing is an Ingenix developed tool that re-prices each claim using specific approaches for different types of claims
  > Facility inpatient-uses a per diem adjusted for aggregated diagnostic service category, presence of major surgery, level of care (acute vs. non-acute), and length of stay
  > Facility outpatient-uses a percentage of billed charges
  > Professional and ancillary services-uses RBRVS from Medicare multiplied by a resources value
  > Pharmacy-average wholesale price

• Approach eliminates contractual relationships from the assessment of cost efficiency

• Cost comparisons then based on 2 things, both within the control of the physician:
  > Units of service
  > Mix of services
Notes

• Standard pricing has long been used to control for differences in reimbursement that are driven by market forces. Ingenix has developed a tool that re-prices each individual claim. Because of a number of factors, a specific approach is needed from re-pricing different types of claims.

• A per diem approach, with the per diem adjusted for service category, the presence of major surgery, level of care, and length of stay is used to re-price inpatient facility claims.

• Facility outpatient claims are re-priced based on a percentage of billed charges.

• The resource based relative value scale, developed by Medicare, multiplied by a consistent resources value is used to re-price professional and ancillary services.

• Pharmacy claims are re-priced using the pharmacy average wholesale price, which is used across the industry.

• The standard pricing approach eliminates contractual relationships from the assessment of cost efficiency.

• This results in 2 things that are within the control of the physician driving any differences in cost comparisons, the number of units of service and the mix of services used.
A key understanding for physicians whose performance is being assessed is how episodes and evidence based medicine rules are assigned or attributed to physicians.
• Episode Responsible Provider—the basis for determining which physician has contributed to an episode to a level to achieve “responsibility”
  > ETG Clusters-based on counts of clusters attributed within an episode; used mostly for non-surgical specialties
  > Cost of Professional Services-based on the percentage of the total of standard priced professional services within the episode; used mostly for surgical specialties

• Threshold Requirements—the minimum level that must be achieved to be attributed the episode
  > 2 types of thresholds
    • For physicians of the same peer definition providing care in the episode
    • For physicians of any peer definition providing care in the episode
• Two important components of episode attribution are episode responsible provider and threshold components.

• There are two approaches to determining which physician has contributed to an episode to a level to achieve responsibility:

• The first is episode clusters, which was discussed in the building blocks ETG section previously. The approach is to use counts of episode clusters, attributed based on the physician providing the face to face visit, to attribute the episode itself. Many acute episodes have only one episode cluster, so this process can be very simple in a number of cases. ETG clusters are used in peer definitions for non-surgical specialties.

• The second approach uses the cost of professional services and is used in peer definitions for some of the surgical specialties. Professional services are CPT codes billed by physicians for services rendered. The percentage of these standard priced costs billed by a physician in the course of an episode are used to determine the Episode Responsible Provider.

• Additionally, there are threshold requirements for each of the two approaches to ensure that a minimum level of involvement is achieved to qualify for episode attribution. There are two thresholds that must be met, one for physicians of the same peer definition providing care in the episode and one for physicians of any peer definition providing care in the episode.
## Attribution Criteria

<table>
<thead>
<tr>
<th>Peer Definition Name</th>
<th>Episode Attribution</th>
<th>Peer Group Threshold (%)</th>
<th>Episode Provider Threshold (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP (Family)</td>
<td>ETG Episode Clusters</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>PCP (Internal Medicine)</td>
<td>ETG Episode Clusters</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>PCP (Pediatric)</td>
<td>ETG Episode Clusters</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Cardiology</td>
<td>ETG Episode Clusters</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Dermatology</td>
<td>ETG Episode Clusters</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>ETG Episode Clusters</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>ETG Episode Clusters</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Hematology/Oncology</td>
<td>ETG Episode Clusters</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Orthopedics</td>
<td>ETG Episode Clusters</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>ETG Episode Clusters</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Urology</td>
<td>ETG Episode Clusters</td>
<td>40</td>
<td>30</td>
</tr>
</tbody>
</table>

- **Most use episode clusters**
- **Some surgical use professional services $**
- **Thresholds for peers can vary**
The illustration on the previous slide shows the two components of attribution for each of the Blue Cross and Blue Shield of Louisiana peer definitions. As you can see, most use episode clusters as the episode responsible provider attribution method. Three of the peer definitions use costs for professional services as the episode responsible provider attribution method. The third and fourth columns represent the minimum thresholds for the peer group and all involved physicians. These percentage thresholds vary for the peer groups and apply either to the number of ETG episode clusters or the cost for professional services, depending on the peer group.
Cluster Approach

• Assigning responsibility for the episode is determined by:
  > Highest number of service clusters within an episode of care
  > Minimum percentage of clusters; If no provider reaches the percentage minimum, the episode is not assigned to a provider in the peer definition.
  > Tie breaker (if needed) is service costs

The ETG episode cluster approach is the most common of the episode responsible provider approaches. The episode is assigned to the physician with the highest number of service clusters within the episode, assuming the physician reaches the percentage minimum threshold both for the peer definition and all involved physicians. An episode is not assigned to any physician for purposes of comparison if a physician does not reach both minimum thresholds. If multiple physicians in a peer definition reach both thresholds and have the same number of episode clusters, the tie breaker is professional service costs. Episodes that are not assigned for lack of thresholds remain in the underlying database, but are not considered for comparison purposes.
Assigning responsibility for the episode is determined by:

> Highest proportion of professional services costs within an episode of care

> Minimum percentage of professional services costs; If no provider reaches the minimum, the episode is not assigned to a provider in the peer definition.

The professional services cost approach is the other episode responsible provider approach used by Blue Cross Blue Shield of Louisiana. The episode is assigned to the physician with the highest proportion of professional services costs within the episode, assuming the physician reaches the percentage minimum threshold both for the peer definition and all involved physicians. An episode is not assigned to any physician for purposes of comparison if a physician does not reach both minimum thresholds. These episodes do, however, remain in the underlying database.
Episode Attribution – Understanding Results

• Only peer definition specific episodes are attributed

• Only the physicians within the peer group who meet both thresholds are in the running for becoming the responsible physician

• Episodes where no physicians meet both thresholds for a peer group will not be assigned a responsible provider

• Episodes that cross peer definitions could have a responsible physician within each peer definition
To recap the important understandings about episode attribution:

- Only episode types that are specifically included in a peer definition are attributed to physicians in that peer definition.
- Eligibility for episode responsible provider is limited to those physicians who meet both minimum thresholds for the approach specific to the peer definition.
- There will be a group of episodes that are not assigned due to no physician meeting both thresholds.
- It is possible, for episode types that are specific to more than one peer definition, to have an episode responsible provider identified within more than one peer definition. If this occurs the episode is attributed to both physicians. Remember, though, that only one physician within a peer definition is attributed an episode no matter how many qualify for episode responsible provider.
Episode Attribution – Quality Rules

• Quality rules are attributed based on who is attributed the episode for the condition to which the quality rules are attached

• Thus, the same quality rules might be attributed to multiple physicians when an episode is attributed to multiple physicians

As a corollary to the potential for the same episode to be assigned to physicians in two different peer definitions, since condition specific quality rules are attributed along with the episode for that condition, the same quality rules could be attributed to physicians in two different peer definitions.
Physician Comparison

The most important consideration for making comparisons is whether there are methodologies to ensure that like comparisons, or apples to apples, are being made. Blue Cross Blue Shield of Louisiana takes advantage of capabilities of the comparison tool to create validity by ensuring that there is adjustment for episode severity, outliers, and case mix. Additionally, confidence intervals are calculated and used to ensure statistical validity to findings.
Creating Validity: Episode Risk Adjustment

- Severity adjustment occurs at the episode level
  - Severity Level: indicator of relative severity from 1 to 4 based upon the patient’s age and gender with the observed mix of complications and comorbidities.

<table>
<thead>
<tr>
<th>ETG Version 7.0</th>
<th>Example of Detail and Summary Information Available for an ETG - CAD I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode</td>
<td>ETG (Base Condition)</td>
</tr>
<tr>
<td>12345</td>
<td>Ischemic Heart Disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co-morbidities</th>
<th>8055</th>
<th>8007</th>
<th>8010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Indicators</td>
<td>9890</td>
<td>9578</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CALCULATION OF RELATIVE EPISODE SEVERITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Complication</td>
</tr>
<tr>
<td>Co-Morbidity</td>
</tr>
<tr>
<td>Co-Morbidity</td>
</tr>
<tr>
<td>Co-Morbidity</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Note: the indicator codes and severity weights used in this example are for demonstration only and do not reflect final coding and weighting approaches.

The Symmetry ETG Episode Severity Level enables analysis of the relative impact of all a member’s conditions, age, and gender to a specific condition.

Each episode has a severity level assigned
Notes

To account for differences in severity within a condition, each episode has a severity level assigned. This severity level is based on the patient’s age and gender along with the observed mix of complications and comorbidities. Note that treatments do not contribute to the level of severity. In this example of a patient with ischemic heart disease, the table contains the components specific to this episode used to calculate the severity score. Each complication and co-morbidity is assigned a severity weight based on a regression analysis for the expectation of relative increase in medical resources for that complication or co-morbidity. This severity weight is specific to these complications and co-morbidities for this specific ETG, as they have different weights for other ETGs. You can see that there is also a severity weight based on the patient’s age and gender. The total severity weight determines the episode severity level for this particular episode. In the lower left hand corner of the table, you can see that this particular ETG has 4 levels of severity, with ranges of severity score for each. In our example, the total severity weight is 6.8, which places this episode into severity level 4. When comparing any physician attributed this episode to his or her peers, the comparison will be only to their Ischemic Heart Disease episodes of severity level 4.
Creating Validity Episode Cost Outliers

• Low Cost Outlier episodes
  > Often driven by miscoding or missing data
  > Excluded from physician measurement to avoid significant understating costs for physicians and peers

• High Cost Outlier episodes
  > High Cost Outlier episodes are included in measurement to capture the relative activity
  > The cost for all service lines within the episode are recalculated to adjust for the truncation for the purpose of physician cost of care measurement
The treatment of episode cost outliers can have a significant impact on comparisons and the methodology of dealing with them is an important component of creating validity. For each episode level of severity, statistical analysis is used to create low and high outlier thresholds. Episodes with costs below the low outlier threshold, often driven by miscoding or incomplete data, are excluded from physician measurement. Episodes with costs that exceed the high cost outlier threshold are truncated or capped at that level. These episodes are included in cost comparisons. Within these episodes, service lines costs are recalculated to account for the truncation for the purpose of cost of care measurement.
Creating Validity: Case Mix Adjustment Level

- Risk Adjustment Level—this accounts for case mix in the comparisons; 2 types are used
  - ETG by severity level
    - Used for peer definitions of non-surgical specialties
  - ETG with surgery, if no surgery, use severity level
    - This adjustment is used for peer definitions for specialties with a significant surgical component
    - Many episodes are subdivided into those with surgery and those without—these episodes are compared at the “with surgery” level if surgery occurred and at the severity level if a surgery has not occurred
A refinement to the comparison paradigm just described occurs in some peer definitions for specialties with significant surgical components. For non-surgical specialties, comparisons are accomplished at the ETG severity level as previously described. For ETGs which are subdivided into “with surgery” and “without surgery” levels, comparisons are made at the ETG with surgery level for those episodes and at the ETG severity level, much like all other ETGs, for the “without surgery” episodes.
Comparisons of cost are made on the basis of comparing a physician’s cost of his mix of cases to the cost of the exact same mix of cases of his peers.

Here is a simple example of how this occurs. Assume there are only 2 conditions for Dr. Jones to treat, HTN and CHF, and there are 2 severity levels of each condition:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Avg Cost/Episode</th>
<th># of Episodes</th>
<th>Total Cost/Condition</th>
<th>Peer Avg Cost/Episode</th>
<th>Total Peer Cost/Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTN-Sever 1</td>
<td>$100</td>
<td>30</td>
<td>$3,000</td>
<td>$125</td>
<td>30 x 125 = $3,750</td>
</tr>
<tr>
<td>HTN-Sever 2</td>
<td>$200</td>
<td>15</td>
<td>$3,000</td>
<td>$250</td>
<td>15 x 250 = $3,750</td>
</tr>
<tr>
<td>CHF-Sever 1</td>
<td>$3,000</td>
<td>10</td>
<td>$30,000</td>
<td>$2,500</td>
<td>10 x 2,800 = $25,000</td>
</tr>
<tr>
<td>CHF-Sever 2</td>
<td>$7,000</td>
<td>5</td>
<td>$35,000</td>
<td>$6,500</td>
<td>5 x 6,500 = $32,500</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td><strong>$71,000</strong></td>
<td></td>
<td><strong>$65,000</strong></td>
</tr>
</tbody>
</table>

The case mix adjusted cost index for Dr. Jones is $71,000/$65,000 = 1.09.
• We will use a specific example, much simplified for illustrative purposes, to demonstrate how validity is created through case mix adjustment. Comparisons of cost are made on the basis of comparing a physician’s cost of his mix of cases to the cost of the exact same mix of cases of his peers. Here is a simple example of how this occurs. In our example for Dr. Jones, an internal medicine specialist, we are going to assume there are only 2 conditions for Dr. Jones to treat, HTN and CHF, and there are only 2 severity levels of each condition. This table contains illustrative information specific to Dr. Jones in the first four columns, and information on the performance of Dr. Jones’ internal medicine peer definition peers in the 5th column. To get the total cost per condition at the severity level, we multiply Dr. Jones’ average cost per episode by his number of episodes at each severity level. To calculate the expected costs for his peers for the same mix of episodes we multiply the peer definition average cost by the same number of episodes for each condition severity level. To give us the ratio of Dr. Jones’ observed costs relative to the expected costs for that same mix of episodes, we divide Dr. Jones’ total observed costs by the total costs expected for the peers. This gives us a case mix adjusted cost index for Dr. Jones of 1.09. For this particular example of Dr. Jones, his observed costs are 9% higher than that expected for his peers for the same exact mix of cases.
Cost Efficiency Comparisons

• Impact Intelligence compares a physician’s costs (resource use) to others of the same specialty by comparing the cost of her mix of ETGs (per the attribution discussion) with the average cost for the specialty for that exact same mix.

• Thus the average index for a specialty (actual/expected costs) is 1.0.

• A physician with an index > 1.0 has actual costs greater than expected for the exact same mix of ETGs.

As just illustrated for Dr. Jones, the Impact Intelligence tool used by Blue Cross Blue Shield of Louisiana creates a cost index for each physician within a peer definition by comparing the cost of her mix of ETGs attributed to the average cost for her specialty peer definition peers for the exact same mix of ETGs. This comparison is to specialty peer definition peers in the BCBS LA claims set. The average across a peer definition is 1.0, so an index of greater than 1 indicates a physician has actual costs greater than expected when compared to peers.
Quality Comparisons

- Impact Intelligence compares a physician’s EBM rule performance to others of the same specialty by comparing her compliance rates on individual measures to the compliance rates of specialty peers for those measures.
- An aggregate of these rates creates an index compared to an index of 1.0 for the specialty/peer group.
- A physician with an index > 1.0 has higher compliance rates on EBM measures than peers.

For quality, the Impact Intelligence tool used by Blue Cross Blue Shield of Louisiana creates a quality index for each physician within a peer definition by comparing her compliance rates on individual evidence based measures to the compliance rates for her specialty peer definition peers for the same measures and then aggregating these compliance comparisons. This comparison is to specialty peer definition peers in the BCBS LA claims set. The average across a peer definition is 1.0, so an index of greater than 1 indicates a physician has compliance with EBM measures greater than expected when compared to peers.
PROFILE SERVICE CATEGORIES – OVERALL COST INDEX

• Classification of services used exclusively to organize services for measuring provider cost efficiency

• Used to weight the service against the physician’s ability to control costs of services

• PSC Classifications
  - Primary Care Core
  - Specialty Care
  - Hospital Services
  - ER
  - Laboratory-includes facility lab
  - Radiology-includes facility imaging
  - Pharmacy
Overall Cost Index, Episode-Based Illustration

<table>
<thead>
<tr>
<th>Table 1: Unweighted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Care Core</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Care Core</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Creating Validity: Confidence Intervals

- BCBSLA uses confidence intervals to guide users in the statistical validity of comparisons
  > A 90% confidence interval around the cost index and quality index is calculated
  > An assessment is then made if the result is different from peers (the index of 1.0) with either a 90% or 95% confidence and that is reported for the cost index and quality index

- Additionally, BCBSLA does not even consider assessing the performance of providers who have fewer than 30 episodes or 30 opportunities for EBM rule compliance attributed

Once overall cost and quality indices are calculated for a physician, confidence intervals for those indices are calculated. Blue Cross Blue Shield of Louisiana uses these confidence intervals to guide users in the statistical validity of comparisons. Two confidence intervals are used. The range of the 90% confidence interval around the cost index and the quality index is calculated. An statistical assessment is then made if the result is different from peers with either a 90 or 95% confidence. That 90 or 95% confidence of difference from peers is reported. While is it possible to have confidence intervals that would seem to indicate differences with small numbers of episodes, such assessments are not reliable, and Blue Cross Blue Shield of Louisiana does not consider assessing the performance of providers who have attributed fewer than 30 episodes or 30 opportunities for EBM rule compliance.
Summary Review

- Overcoming challenges to physician performance assessment
- Building blocks
  - ETGs
  - EBMs
- Concepts
  - Peer definitions
  - Standard pricing
  - Attribution
- Physician comparisons and creating validity
  - Severity adjustment
  - Outlier adjustment
  - Case mix adjustment
  - Confidence intervals
Today’s training has helped you to understand the challenges of physician performance assessment and how Blue Cross Blue Shield of Louisiana responds to those challenges. You should now be familiar with episode treatment groups, how episodes are created, and evidence based medicine measurement rules. We have reviewed the concepts of peer definitions, standard pricing, and attribution to help you understand how these help ensure that the tools are used to assess performance are appropriate. Finally we explored how BCBS LA does physician comparisons and creates validity in the comparisons by using powerful approaches to risk adjustment, outlier adjustment, case mix adjustment, and confidence intervals. Completing this training should have familiarized you with tools important to understanding the meaning of information shared about your or your group’s performance and the attention to ensuring that this information is as accurate and valid as possible.