

Reducing medical costs using a SAS® based system for provider and facility efficiency scoring: a successful strategy for competing on ACA-mandated State Health Exchanges

Stephanie T Poley, PhD, Daryl Wansink, PhD, and Tracy Lewis, MOR.
Blue Cross and Blue Shield of NC, Chapel Hill, NC

ABSTRACT

Healthcare expenditure growth continues to be a prominent healthcare policy issue, and the uncertain impact of the Affordable Care Act (ACA) has put increased pressure on payers to find ways to exercise control over costs. Fueled by provider performance analytics, Blue Cross and Blue Shield of North Carolina (BCBSNC) has developed innovative strategies that recognize, incentivize, and assist providers delivering high-quality and efficient care. A leading BCBSNC strategy has been the introduction of a tiered network product called Blue Select, which was launched in 2013 and is featured in the Health Insurance Marketplace. Blue Select is a PPO plan with differential member cost-sharing for tiered providers. Tier status of providers is determined by comparing providers to their peers on the efficiency and quality of the care they deliver. While the primary goal of provider tiering is cost containment through member steerage, the initiative has also resulted in new and strengthened collaborative relationships between BCBSNC and providers. The strategy offers the opportunity to bend the cost curve and provide meaningful change in the quality of healthcare delivery.

PROVIDER SCORING METHODOLOGY

BCBSNC Blue Select includes a tiered network of physicians from seven specialties including General Surgery, Cardiology, Gastroenterology, Obstetrics and Gynecology, Neurology, Orthopedics, and Endocrinology; other specialties are not tiered and members may choose among them without incurring a difference in cost sharing. Tier 1 provider status is dependent upon meeting performance standards for both quality and efficiency as defined for each specialty. Providers who meet or exceed the standard for quality are then evaluated on their case-mix adjusted costs for total episodic costs. Each practice's performance is compared, through indirect standardization, to expected performance, given the patients and conditions treated within a practice. A ratio of observed to expected performance is calculated for both cost and quality to use in determining the tier status of providers for Blue Select. Scores are computed as a ratio of observed to expected performance based on peer comparisons for each measure.

Quality

In the case of quality, the expected score is a weighted average score based on peer performance for each relevant measure, weighted to reflect the mix of measures the practice has. Quality measures include Healthcare Effectiveness Data and Information

Set (HEDIS®) adherence rates and PROMETHEUS® potentially avoidable complication rates for select procedures. This creates an expected practice level score that assumes the peers of the practice have a similar panel of patients. The peer group for each practice consists of all practices in North Carolina for which there were providers of the given specialty treating members. Quality scores are further weighted to provide more balanced scores across the different measures. This reflects the relative importance, but lower frequency, of procedural complications. The procedural complications were given greater weighting in the final scoring for Cardiology, Orthopedics and OBGYN (see Appendix A). A ratio is computed for actual quality observed compared to expected quality based on peer performance. With this ratio a score of 1 indicates that the practice performed exactly as expected compared to peers. A score below 1 indicates lower quality relative to what was expected.

A 90% confidence interval is computed for the provider quality ratio. The methodology to compute this score follows the methodology laid out by MedVantage for their Composite Quality Measure. It is a proprietary algorithm but involves using indirect standardization with a confidence interval based on a binomial distribution. The basic methodology was adopted from Harvard Pilgrim Health Care which is affiliated with Harvard Medical School.

The 90% confidence interval sets a low and high range. If the high end (upper boundary of the confidence interval) of the range is below 1, then the practice is designated with a Tier 2 status. This threshold can be interpreted to mean that, with 90% certainty, the practice has lower quality than their peers. Otherwise, the provider is further evaluated to determine whether their costs qualify them for Tier 1 status.

An example of the use of indirect standardization is illustrated below.

	Observed Adherence Rate	Expected Adherence Rate	Members Eligible for Measure	Observed N	Expected N
Measure 1	100%	92%	50	50.0	46.2
Measure 2	76%	77%	192	145.0	147.1
Measure 3	50%	62%	85	42.5	52.8
Measure 4	63%	77%	120	75.0	92.4
Measure 5	44%	63%	48	21.3	30.3
Measure 6	57%	70%	35	20.0	24.5
Measure 7		56%	0	-	-
Total N			530	353.8	393.3
Quality Score (Observed N/Expected N)					0.8995
Quality Score - Lower CI Boundary					0.8398
Quality Score - Upper CI Boundary					1.0865
Quality Designation					Tier 1

Cost

Efficiency scores evaluate services that are grouped into an episode of care for the patient. Episodes of care are created using OptumInsight's ETG v7.6 Episode Treatment Grouper (ETG) logic and ETG v7.6 Procedure Episode Grouper (PEG) logic. The ETG software looks across all types of claims to identify services that are related to the treatment of more than 500 specific conditions. All claims that are related to a specific condition and within a clinically defined time window are grouped together into an ETG. The PEG software looks across all claims to identify services related to a surgical procedure. The claims related to that procedure within a clinically defined time window are grouped into the PEG.

Expected costs are calculated as the average cost for each combination of Base ETG, ETG severity level (up to four severity levels), subspecialty and referral or shared episode indicator. In general, the referral indicator flags referrals between providers of the same specialty during an episode that would indicate a more complex condition that requires sub-specialization. Generally, these episodes are more costly due to involvement of multiple specialists and the standardization method adjusts for this issue. For PEGs, the expected cost is a weighted average score based on peer average cost for each combination of PEG and place of service (inpatient vs outpatient). This approach creates an expected practice level total cost that assumes the practice's peers had the same mix of patients. A ratio is computed for actual total costs over expected total costs using indirect standardization. A 90% confidence interval is computed for the provider's ratio. The confidence interval is computed using a weighted mean methodology. Providers whose lower bound of the confidence interval is above 1 and whose ratio score is greater than or equal to 1.1 are considered inefficient and are designated with a Tier 2 status.

A simple example of the use of indirect standardization for cost evaluation is illustrated below.

ETG	Severity	Practice's count of episodes	Peer average episode cost	Practice's expected cost	Practice's actual cost
Inflammatory bowel dis.	1	100	\$3,524	\$352,400	\$300,000
Inflammatory bowel dis.	2	10	\$6,541	\$65,410	\$70,000
Inflammatory bowel dis.	3	1	\$11,074	\$11,074	\$20,000
Sum				\$428,884	\$390,000
Efficiency ratio					0.91

SAS® CODE PRESENTATION

Quality Scoring

After preparing all datasets to determine practice rates, numerators, and denominators for each quality measure, as well as means for the overall group to use as expected rates, we join the datasets and compute expected rates given patient volume,

aggregate data across measures and apply logic to assign Tier 1 or Tier 2 status on basis of cost or quality. The following SAS® code scores patients on basis of quality performance, but can be similarly extended to evaluate cost by replacing the variables. The code is simplified for purpose of illustrating key steps.

STEP 1

Compute expected numerators with relevant weights and calculate scores with upper and lower confidence interval boundaries.

/*JOINING PRACTICE PERFORMANCE (OBSERVED) DATA TO PEER PERFORMANCE (EXPECTED) DATA*/

```
DATA quality_spec1; MERGE practices_spec1
spec1_benchmarks; BY spec1;
practice_id, practice_name,
```

/*CALCULATING EXPECTED NUMERATORS*/

```
exp_measure1_num = measure1_denom*measure1_rate_exp;
exp_measure2_num = measure2_denom*measure2_rate_exp;
exp_measure3_num= measure3_denom*measure3_rate_exp;
exp_measure4_num = measure4_denom*measure4_rate_exp;
exp_measure5_num = measure5_denom*measure5_rate_exp;
exp_measure6_num = measure6_denom*measure6_rate_exp;
```

/*OVERWEIGHTING THE 2 OUTCOME MEASURES TO INCREASE IMPORTANCE OF OUTCOME MEASURES, WHICH ARE MORE RARE AND OF GREATER IMPORTANCE THAN PROCESS MEASURES THAT APPLY TO LARGE POPULATIONS. */

```
wtd_exp_measure5_num = measure5_denom*measure5_rate_exp*20;
wtd_exp_measure6_num = measure6_denom*measure6_rate_exp*20;
```

/*SUMMING EXPECTED AND OBSERVED PERFORMANCE ACROSS ALL MEASURES*/

```
wtd_expected_num = exp_measure1_num + exp_measure2_num + exp_measure3_num +
exp_measure4_num + wtd_exp_measure5_num + wtd_exp_measure6_num;

wtd_actual_num = spec1_measure1_num + spec1_measure2_num + spec1_measure3_num +
spec1_measure4_num + (spec1_measure5_num *20) + (spec1_measure6_num*20);
```

/*CALCULATING SCORE AND UPPER/LOWER CONFIDENCE INTERVALS*/

```
quality_ratio = wtd_actual_num/wtd_expected_num;
lower_ci = quality_ratio - (1.645 * (sqrt(quality_ratio/expected_num)));
upper_ci = quality_ratio + (1.645 * (sqrt(quality_ratio/expected_num)));
run;
```

STEP 2

Suppress scores for practices with insufficient volume and assign tier status to practices based on performance

PROC SQL;

```
create table spec1.spec1_quality_final as select
practice_id, practice_name,
```

```
coalesce(measure1_denom,0) as measure1_N,
coalesce(measure2_denom,0) as measure2_N,
```

```

coalesce(measure3_denom,0) as measure3_N,
coalesce(measure4_denom,0) as measure4_N,
coalesce(measure5_denom,0) as measure5_N,
coalesce(measure6_denom,0) as measure6_N,

case when (coalesce(measure1_denom,0) + coalesce(measure2_denom,0) +
  coalesce(measure3_denom,0) + coalesce(measure4_denom,0) +
  coalesce(measure5_denom,0) + coalesce(measure6_denom,0) ge 30) then quality_ratio else .
end as quality_score,

case when (coalesce(measure1_denom,0) + coalesce(measure2_denom,0) +
  coalesce(measure3_denom,0) + coalesce(measure4_denom,0) +
  coalesce(measure5_denom,0) + coalesce(measure6_denom,0) ge 30) then lower_ci else . end
as quality_score_lci,

case when (coalesce(measure1_denom,0) + coalesce(measure2_denom,0) +
  coalesce(measure3_denom,0) + coalesce(measure4_denom,0) +
  coalesce(measure5_denom,0) + coalesce(measure6_denom,0) ge 30) then upper_ci else .
end as quality_score_uci,

case when (coalesce(measure1_denom,0) + coalesce(measure2_denom,0) +
  coalesce(measure3_denom,0) + coalesce(measure4_denom,0) +
  coalesce(measure5_denom,0) + coalesce(measure6_denom,0) lt 30) then 'NA'
  when upper_ci lt 0.9950 then 'TIER 2'
  else 'TIER 1' end as quality_tier,

(coalesce(measure1_denom,0) + coalesce(measure2_denom,0) + coalesce(measure3_denom,0) +
  coalesce(measure4_denom,0) coalesce(measure5_denom,0) +
  coalesce(measure6_denom,0)) as quality_N

from quality_spec1;
QUIT;

```

APPLICATION AND EXTENSIONS

The original business purpose of provider scoring was to identify the network's highest quality and most efficient specialists for inclusion in a tiered network product, as an option for cost-conscious groups and individuals. This application continues to provide value to the business, but the resulting datasets, with detailed information on performance across each quality measure and type of episode, have provided considerable insight to network management and the providers themselves, as a guide for improving their practice. In many cases, the analyses have drawn attention to relationships among providers that may result in high cost or poor quality for patients, and the data is instructive in identifying opportunities for improvement.

LIMITATIONS

A common constraint to evaluating provider performance is that of limited information. Health plans have access to all claims for members, but administrative claims contain limited clinical information and are often coded for purposes of reimbursement maximization rather than performance evaluation. Consequently, quality measurement

typically reflects process of care rather than outcomes, and the few available outcome measures can stir controversy. As such, it is important to consider using confidence intervals and minimum volume requirements for scoring, and to provide the underlying data to providers so that they have the opportunity to refute and correct the conclusions made from limited data.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Stephanie Poley, PhD

Senior Informatics Scientist, Health Economics
Blue Cross Blue Shield of North Carolina
PO Box 2291
Durham, NC 27707
919.765.1328
Stephanie.Poley@bcbsnc.com

Daryl Wansink, PhD

Director, Health Economics
Blue Cross Blue Shield of North Carolina
PO Box 2291
Durham, NC 27707
919.765.4957
Daryl.Wansink@bcbsnc.com

Tracy Lewis, MOS

Informatics Scientist, Health Economics
Blue Cross Blue Shield of North Carolina
PO Box 2291
Durham, NC 27707
919.765. 3831
Tracy.Lewis@bcbsnc.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.