Focusing Improvement Efforts to Maximize Score

Sean Ramsdale, Management Information Analysis

The Value-Based Purchasing program was put in place to encourage hospitals to improve the quality of patient care. The program requires that 2% of Medicare payments to hospitals be set aside for incentive payments. Essentially, all hospitals put 2% of their Medicare payments into a pool. Hospitals that perform well receive more than 2% back while those hospitals performing worse than average receive less.

For a hospital looking to improve their Value-Based Purchasing total performance score, it can be difficult to determine where to allocate finite resources given the number of measures used to calculate the final score. Assuming it is possible to focus on improving only a limited number measures, the intuitive choice is to address the measures with the lowest scores. In some cases, however, focusing improvement efforts on stronger scores actually maximizes the total performance score. It is the relationship between the measures and national standards that complicates identifying where improvement helps most.

The purpose of this paper is not to provide a detailed example of how total performance scores are calculated, but instead to outline a methodology for identifying measures to be prioritized for improvement.

Total Performance Score Basics and Terminology

Hospitals are rated based on around 22 measures – every year a few measures are added and dropped. The measures are grouped into four domains. Beginning in FY2018, a hospital needs to have scores in three of four domains to be included in the Value-Based Purchasing program.

- The table below lists the four domains used to rate hospitals in FY2018 and the number of measures within each domain. A score of between 0 and 100 is calculated for each domain.

Domain	# Measures
Clinical Care	3
Patient- and Caregiver-Centered Experience of Care/ Care Coordination	9
Efficiency and Cost Reduction	1
Safety	8

- The domains are equally weighed when calculating the total performance score.

- Each measure is given a score of between 0 and 10. The measure score is the better of the achievement score or the improvement score.
    - Achievement score: measures the facility's performance relative to the national threshold and benchmark.
        - Threshold: median of national scores; measures below the threshold are given an achievement score of 0.
        - Benchmark: top decile of national scores; measures at or above the benchmark are given an achievement score of 10.
    - Improvement score: measures the facility's performance relative to the previous year.

Where the strongest measure within a domain is already above the benchmark, no improvement to that measure is made for the purpose of this analysis. Facilities with all measures above the benchmark will have a perfect domain score of 100 and will see no change in domain score. It's important to note, however, that benchmarks and thresholds change from year to year, so a measure that is above the benchmark may be below it at some future point if no improvements are made.

Data for all facilities participating in the Value-Based Purchasing program is available on the Medicare.gov website. The most recent year available is FY2017. For this analysis, FY2017 measures were compared against FY2018 thresholds and benchmarks. The domains changed slightly from FY2017 to FY2018, and measures were shifted where necessary to reflect how they are organized for FY2018. The one measure added for FY2018 is not included in the analysis since data is not available.

Clinical Care Domain

The Clinical Care Domain consists of three 30-day mortality measures, which are actually shown as survival rates. For these measures, a higher score is better. The national benchmarks and thresholds for each measure are shown in the table to the right.

	Benchmark	Threshold
Acute Myocardial Infarction	0.871669	0.851458
Heart Failure	0.903895	0.881794
Pneumonia	0.908124	0.882986

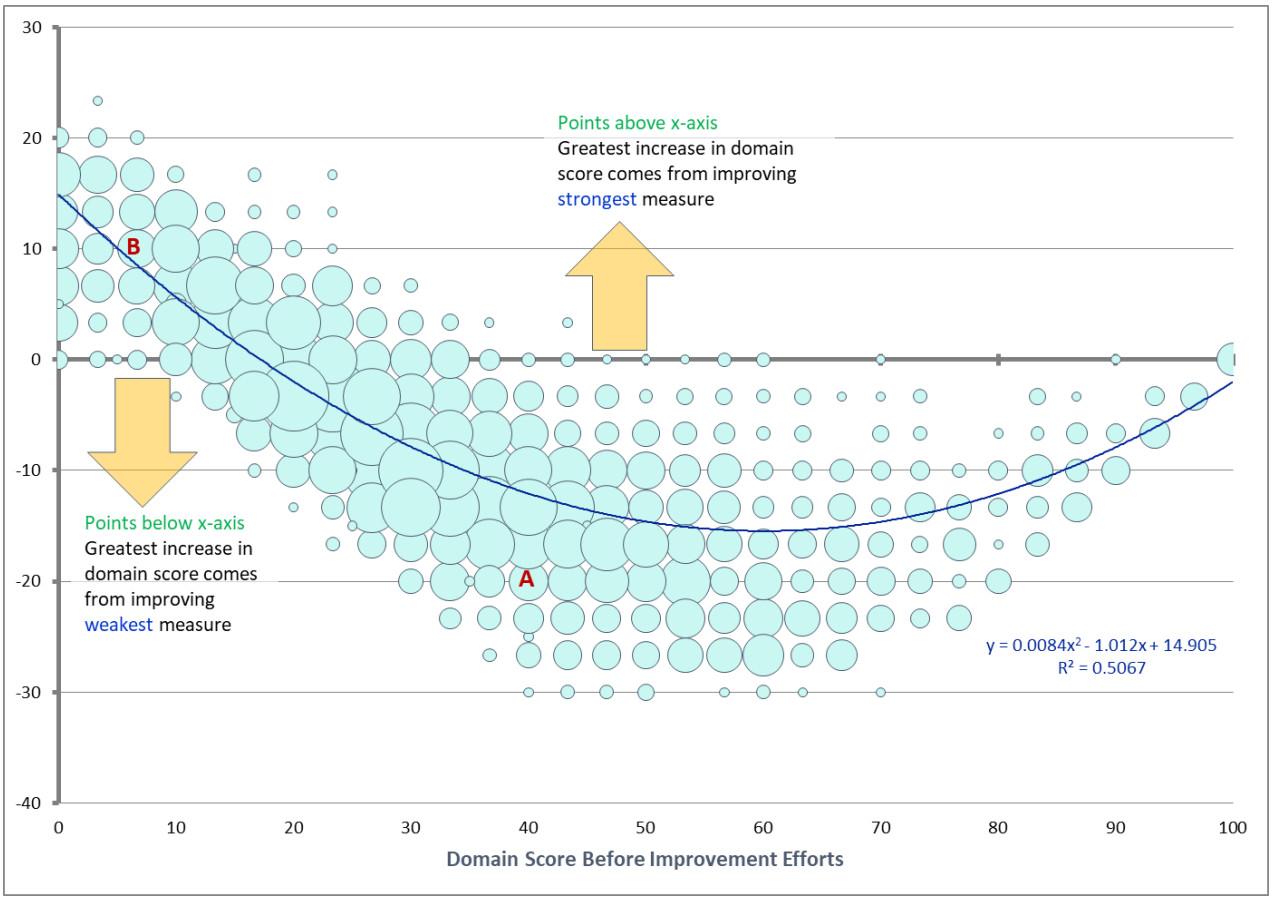
The first step in the analysis is to identify the strongest and weakest measure for each facility. The strongest measure is where performance is closest to the benchmark and the weakest is farthest. Below is an example of how the strongest and weakest measures are identified. The strengths are circled in blue and the weaknesses in red.

	AMI	Heart Failure	Pneu-monia
Provider A			
Performance score	0.8677	0.8852	0.8883
Distance from benchmark	-0.0039	-0.0187	-0.0198
Provider B			
Performance score	0.8563	0.8596	0.8751
Distance from benchmark	-0.0154	-0.0443	-0.0331

It is assumed that each facility has the resources to improve one of the clinical care domain measures by 0.02. First, clinical domain scores are recalculated with 0.02 being added to each facility's strongest measure. Then the clinical domain scores are recalculated with 0.02 added to each facility's weakest measure. The resulting clinical care domain scores for Providers A and B are shown below. Line d shows the difference in domain scores for improving the strongest measure vs. improving the weakest one. A negative number indicates that improving the weakest measure created the most improvement in the domain score. Improving the weakest measure for Provider A resulted in a higher domain score, but the opposite was true for Provider B.

	Domain Score	Provider A	Provider B
a	Before improvement	40.0	10.0
b	After improving strongest measure	46.7	33.3
c	After improving weakest measure	66.7	23.3
d	Line b minus Line c	-20.0	10.0

The same calculations were done for all facilities and the results are shown in the chart below. Plotting line d with the domain score before improvement (line a) shows that facilities with lower scores have generally higher domain scores when they improve the strongest measure (points above the x-axis). The trend line crosses the x-axis at 17. Therefore, facilities with a starting score of 17 or lower should focus improvement efforts on strengths. Providers A and B are marked in red.



Facilities with domain scores approaching 100 almost certainly have at least one measure above the benchmark, so improving the weakest measure will not have as much of an impact on domain score. The trend line shows that improving the weakest measures hits a point of diminishing returns for facilities with a beginning score of 60 or higher since the weakest measure is already close to the benchmark.

Safety Domain

The safety domain combines six health-care associated infection rates, the rate of elective deliveries prior to 39 weeks of gestation and the AHRQ PSI-90, a composite score of eight patient safety and adverse event indicators. For all of these measures, a lower score is better. The national benchmarks and thresholds for each measure are shown in the table below.

	Benchmark	Threshold
Central Line-Associated Bloodstream Infections	0.000	0.369
Catheter-Associated Urinary Tract Infections	0.000	0.906
Surgical Site Infection: Colon	0.000	0.824
Surgical Site Infection: Hysterectomy	0.000	0.710
Methicillin-Resistant Staphylococcus Aureus	0.000	0.767
C. difficile Infections	0.002	0.794
Elective Delivery Prior to 39 Completed Weeks of Gestation (PC-01)	0.000000	0.020408
AHRQ PSI-90	0.397051	0.577321

As with the Clinical Care Domain, the first step is to identify the strongest and weakest measures. A facility may have data for as many eight measures, but as few as three. For facilities with five or more measures, the two strongest and two weakest measures are used to measure the impact of improving strengths versus weakness. In Provider B's case, one of the strongest measures (AHRQ PSI-90) is better than the benchmark and two others (CLABSI and CAUIT) are tied right at the benchmark. As shown in the example below, only the one strongest and one weakest measures are used for facilities with fewer valid measures available.

	CLABSI	CAUTI	SSI Colon	SSI Hyst	MRSA	CDI	PC-01	AHRQ PSI-90
Provider A								
Performance score						0.000	0.065	0.517
Dist from benchmark						-0.002	0.065	0.120
Provider B								
Performance score	0.000	0.000				0.550	0.161	0. 377
Dist from benchmark	0.000	0.000				0.548	0.161	-0.020
Provider C								
Performance score	0.132	0.532	0.945	0.430	0.777	1.126	0.011	0.582
Dist from benchmark	0.132	0.532	0.945	0.430	0.777	1.124	0.011	0.185

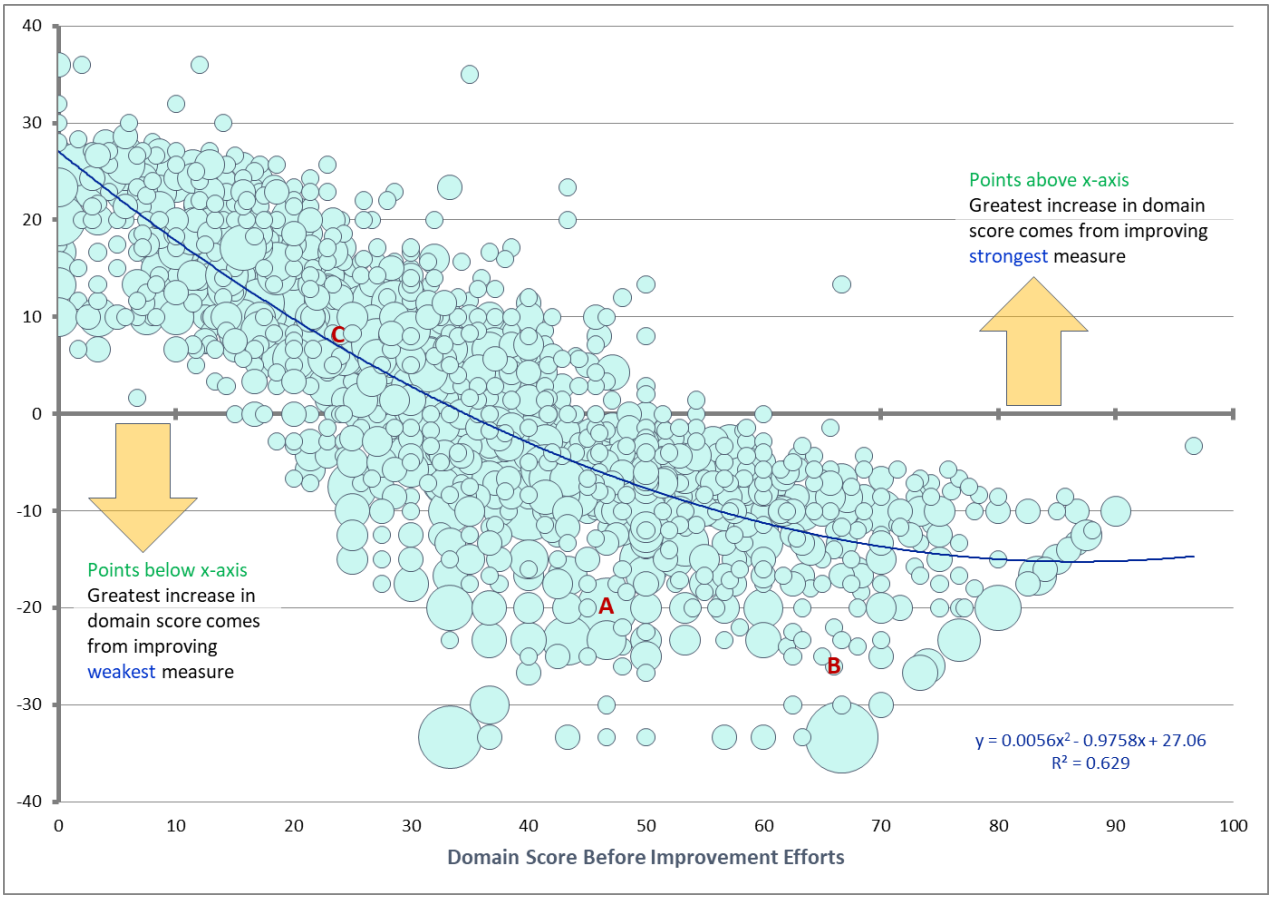
Each facility is able to improve selected measures by a total of -0.59, which may be spread over one or two measures. The example to the right shows how the improvement is split between Provider B's two weakest measures.

	CDI	PC-01
Distance from benchmark	0.548	0.161
Weight based on distance	77%	23%
Improvement applied	-0.456	-0.134

The comparison of improving strengths or weakness for the example facilities is shown in the following table. Strengths for Providers A and B were above the benchmark, so there is no change to domain score if improvement efforts are directed at strengths.

		Provider A	Provider B	Provider C
a	Score before improvement	46.7	66.0	24.3
b	Improve strength	46.7	66.0	37.1
c	Improve weakness	66.7	92.0	28.6
d	Line b minus Line c	-20.0	-26.0	8.6

Line d shows the difference in domain scores for improving the strongest measure versus improving the weakest one. Providers A and B show the most improvement by focusing on weakness while Provider C has a larger increase in the domain score by improving strengths. Generally, facilities with a score above 34-35 will see a greater increase in the domain score if they focus on improving weaknesses. Those with lower beginning scores, like Provider C, will do better to focus on strengths.



Patient- and Caregiver- Centered Experience of Care/Care Coordination Domain

The Patient Experience domain is eight questions from the HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) survey. The performance measures are the percent of respondents who gave a "top box" or most positive reply. In FY2018, the 3-Item Care Transition measure was added to the domain. This is a composite measure from a separate three question survey. It is not included in this analysis since data is not available.

Higher scores are better for Patient Experience measures. The national benchmarks, thresholds and floors (0th percentile) for each measure are shown in the table to the right.

	Benchmark	Threshold	Floor
Communication with nurses	86.68	78.52	55.27
Communication with doctors	88.51	80.44	57.39
Responsiveness of hospital staff	80.35	65.08	38.40
Pain management	78.46	70.20	52.19
Communication about medications	73.66	63.37	43.43
Hospital cleanliness/quietness	79.00	65.60	40.05
Discharge information	91.63	86.60	62.25
Overall rating of hospital	84.58	70.23	37.67

This domain has two components. The first is a performance component that is calculated like scores for the Clinical Care and Safety domains. It is 80% of the overall domain score. The second component measures the consistency of responses to the HCAHPS survey and accounts for 20% of the domain score. Facilities with all measures above the threshold receive a consistency score of 20. The consistency score for the remaining facilities is determined by the lowest measure's distance from the threshold. Facilities with a measure at or below the floor automatically receive a consistency score of zero.

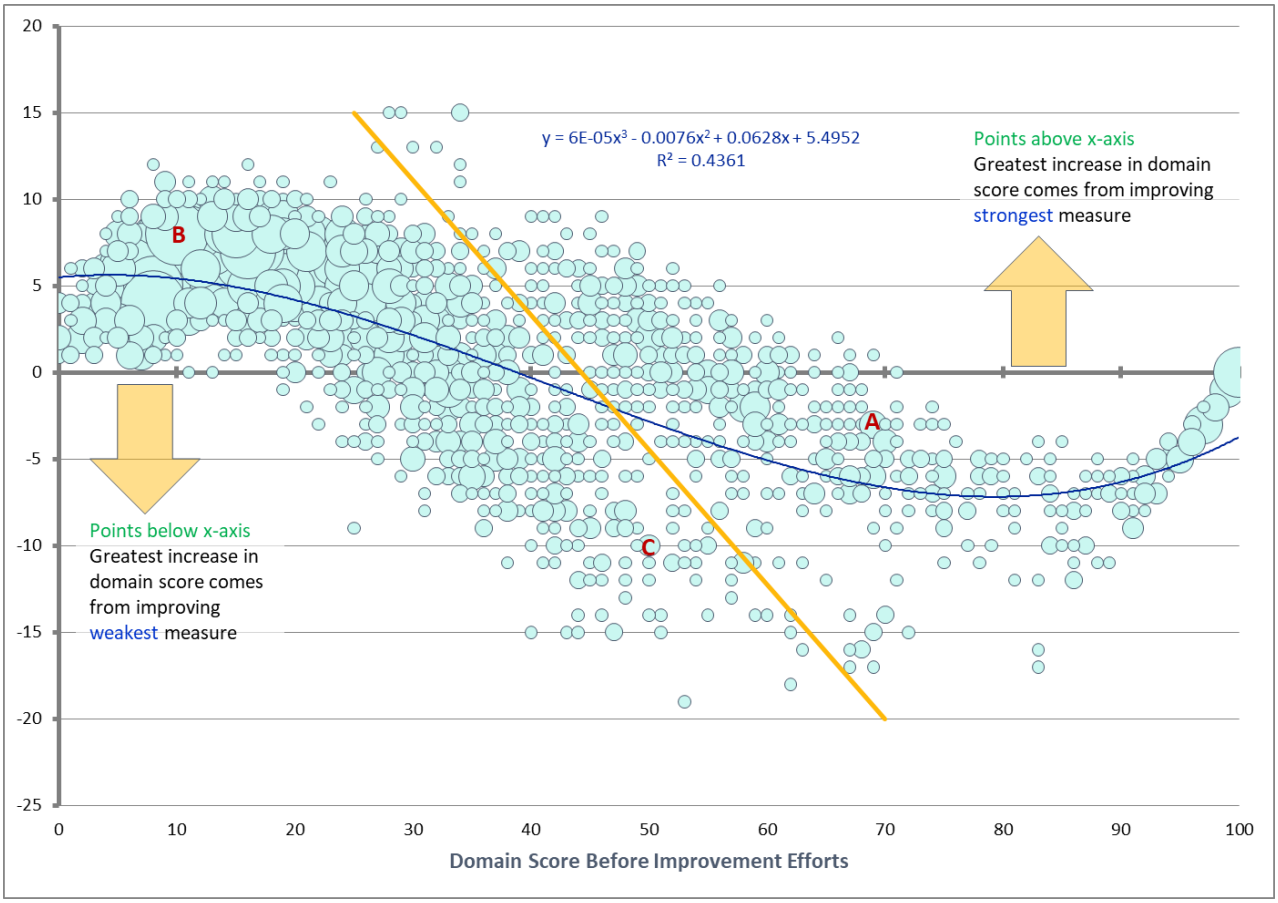
The two strongest and two weakest measures are identified for each facility. Three examples are shown below.

	Comm Nurse	Comm Doctor	Hosp Staff	Pain Mgmt	Comm Meds	Clean Quiet	Dsch Info	Over-all
Provider A								
Performance score	85.66	84.22	75.47	75.59	67.80	72.06	91.24	76.89
Dist from benchmark	-1.02	-4.29	-4.88	-2.87	-5.86	-6.94	-0.39	-7.69
Provider B								
Performance score	77.29	80.15	63.16	67.89	58.75	62.75	84.73	63.45
Dist from benchmark	-9.39	-8.36	-17.19	-10.57	-14.91	-16.25	-6.90	-21.13
Provider C								
Performance score	81.41	85.91	76.27	68.75	66.88	70.77	90.38	77.61
Dist from benchmark	-5.27	-2.60	-4.08	-9.71	-6.78	-8.23	-1.25	-6.97

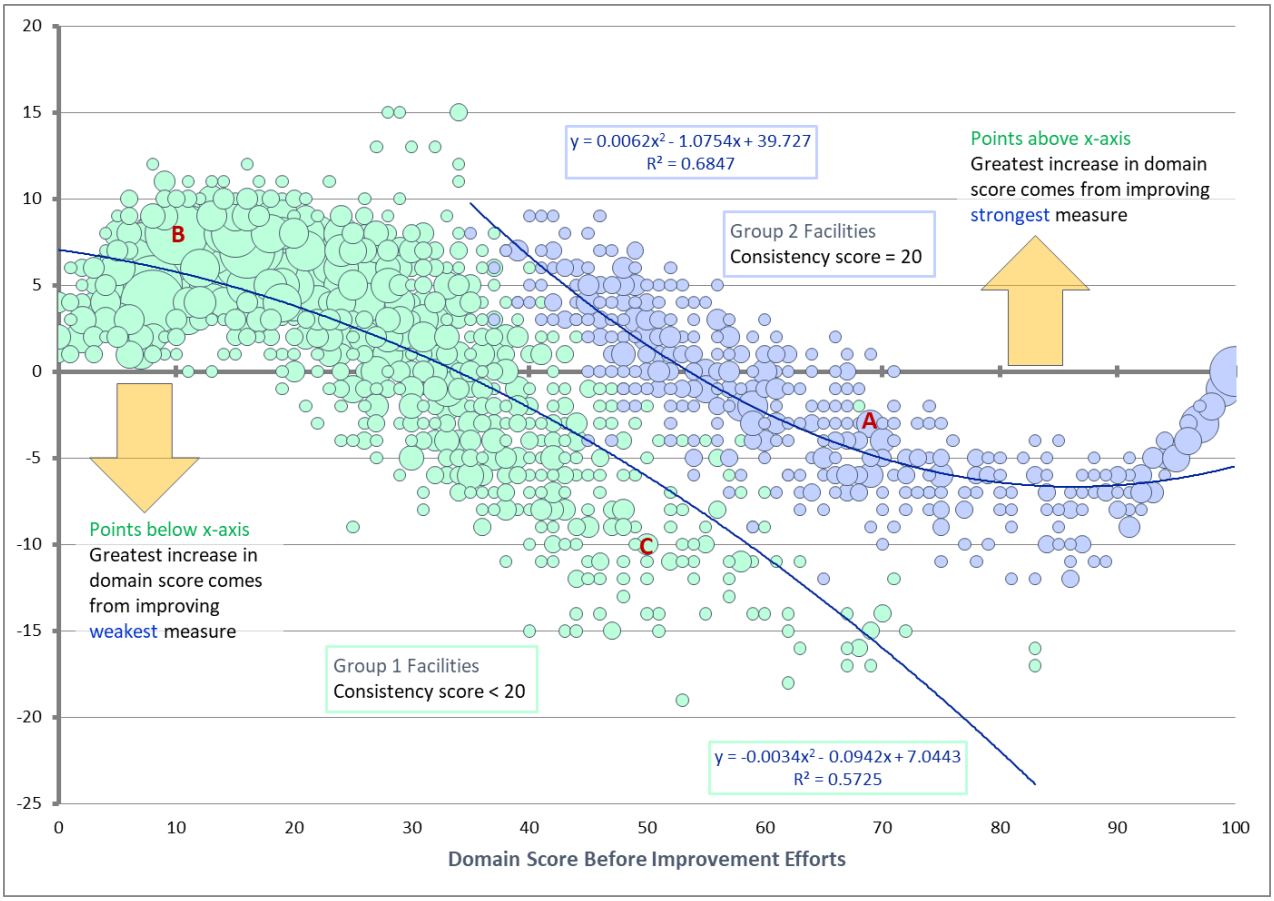
It is assumed that each facility will split 10.26 in improvement over two measures. The comparison of improving strengths or weaknesses is shown in the next table.

		Provider A	Provider B	Provider C
a	Score before improvement	69.0	10.0	50.0
b	Improve strength	72.0	22.0	56.0
c	Improve weakness	75.0	14.0	66.0
d	Line b minus Line c	-3.0	8.0	-10.0

Once again, line d is plotted against the starting domain score. In this instance, however, there appears to be two groups of facilities.



The facilities were split into two groups using the gold line. Upon further examination, it was found that 88% of group to the right of the gold line has a consistency score of 20, but only 0.2% of the group to the left does. The data was analyzed again after splitting facilities with a consistency score of 20 (Group 2) from the others (Group 1).



The trend line for Group 1 facilities (in green) crosses the x-axis between 33 and 34. Group 1 facilities with beginning scores below that point should focus on improving strengths, while others should improve their weakest measures. The exception is the few facilities with a measure below the floor. A measure below the floor should always be a top priority since it results in a consistency score of zero.

Group 2 facilities have, on average, higher starting domain scores than Group 1. Facilities with weaker beginning scores in Group 2 also should focus on strengths, but Group 2's trend line crosses the x-axis at a higher point than Group 1. For Group 2, facilities with a starting domain score below 53 should focus on improving their strongest measures and those above 53 should work on weaknesses. Improving weaknesses reaches a point of diminishing returns once the beginning domain score hits 87 since even the weakest measures are close to or above the benchmark.

Efficiency and Cost Reduction Domain

This domain has only one measure: Medicare spending per beneficiary (MSPB). This measure is risk-adjusted average cost for Medicare inpatient episodes. MSPB includes all Part A and Part B claims that occur 1 to 3 days prior to, during or 1 to 30 days after a Medicare hospital admission. Lower MSPB results in a higher domain score.

Like measures in other domains, MSPB is compared against the national threshold and benchmark to determine a measure score. Cost is considered only as a whole and the types of cost are not evaluated

individually. Therefore, reducing the average cost of any part of Medicare inpatient episodes will improve the domain score.

Conclusion

There is no one domain or measure that holds the key to Value-Based Purchasing success. The domains are all equally weighted, so a high score is worth the same in one domain as another. Within the domains, no one measure or set of measures that predict overall success can be identified.

In fact, which measures to prioritize for improvement efforts varies from facility to facility. With the exception of the Efficiency and Cost Reduction domain, a domain's measures can be sorted into strengths and weaknesses by comparing the facility's measure scores to national standards. Knowing the strengths and weaknesses, along with baseline domain scores is the basis for allocating limited resources for improvement efforts.