Comparison of Approaches for Aggregating Quality Measures in Population-based Payment Models

Alex McDowell, Christina A. Nguyen , Michael E. Chernew, Kevin N. Tran, J. Michael McWilliams, Bruce E. Landon, and Mary Beth Landrum

Objective. To assess the impact of alternative methods of aggregating individual quality measures on Accountable Care Organization (ACO) overall scores.

Data Source. 2014 quality scores for Medicare ACOs.

Study Design. We compare ACO overall scores derived using CMS' aggregation approach to those derived using alternative approaches to grouping and weighting measures.

Principal Findings. Alternative grouping and weighting methods based on statistical criteria produced overall quality scores similar to those produced using CMS' approach ($\kappa = 0.80$ to 0.95). Scores derived from giving specific domains greater weight were less similar ($\kappa = 0.51$ to 0.93).

Conclusions. How measures are grouped into domains and how these domains are weighted to generate overall scores can have important implications for ACO's shared savings payments.

Key Words. Quality of care/patient safety (measurement), health care organizations and systems, health policy/politics/law/regulation, payment systems: FFS/capitation/RBRVS/DRGs/risk adjusted payments etc., evaluation design and research

The American health care system is in the midst of a significant transformation from fee-for-service (FFS) to alternative payment models (APMs). Efforts to tie 50 percent of Medicare payments to quality and value by 2018 have been mirrored by increases in value-based contracting in the private sector (Cassel and Kronick 2015; Muhlestein, Burton, and Win 2017; Health Care Transformation Task Force 2018). Bipartisan support for these new models is likely to continue (U.S. Senate 2015; Muhlestein, Burton, and Win 2017).

Population-based payment models, such as Accountable Care Organization (ACO) programs, are the broadest alternative payment model and hold provider organizations accountable for total spending and quality. Because population-based models put health care organizations on a fixed budget, robust quality incentives are required to counteract incentives to stint on care. There has been much debate over the selection of measures that will best capture quality (Gandhi et al. 2002; Jha et al. 2005; Kessell et al. 2015) and some discussion on the use of empirically driven weighting schemes for determining hospital ratings (Austin et al. 2015). However, there has been less discussion of the methods used to aggregate individual measures into a single overall score required for payment calculations. Further, there has been little investment in examining the sensitivity of different aggregation methods.

Aggregation typically consists of three steps: (1) grouping measures into domains, (2) weighting measures within domains to get domain scores, and (3) weighting the domain scores to compute an overall score. An alternative approach is to weight individual measures into a summary score without grouping. But the intermediate step of creating domain scores is useful to inform providers and outside stakeholders about broad areas of performance because reporting individual measures may be too overwhelming for users. Moreover, in some cases, payers may want to base rewards or interventions on domain-specific scores. Yet, even if the measures are not explicitly aggregated prior to computing payments, any payment based on the measures is a form of aggregation as each measure must be given a weight relative to the others.

In 2014, the Centers for Medicare and Medicaid Services (CMS) grouped 33 quality measures into four domains based on clinical relevance and generated domain-level scores. An ACO's domain score, equally weighted across the six to eight measures in each domain, was used to identify low-performing ACOs in each domain. Overall scores were obtained as equally weighted averages of the four domains, which determined the

Address correspondence to Mary Beth Landrum, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; e-mail: landrum@hcp.med.harvard.edu. Alex McDowell, M.P.H., M.S.N., R.N., is with the Harvard University, Boston, MA. Christina A. Nguyen, A.B., Michael E. Chernew, Ph.D., J. Michael McWilliams, M.D., Ph.D., and Bruce E. Landon, M.D., M.B.A., are with the Department of Health Care Policy, Harvard Medical School, Boston, MA. Kevin N. Tran, M.P.H., is with the IBM Watson Health, Cambridge, MA. J. Michael McWilliams, M.D., Ph.D., is also with the Brigham and Women's Hospital, Boston, MA. Bruce E. Landon, M.D., M.B.A., is also with the Beth Israel Deaconess Medical Center, Boston, MA.

percentage of its shared savings (Centers for Medicare and Medicaid Services, 2011, 2015a,b).

This aggregation process raises three methodological issues: How measures are grouped into domains, how measures are weighted within domains, and how the domains are weighted to produce an overall score. Grouping and weighting of measures could be based, for example, on clinical criteria (combine all diabetes measures and weight based on clinical impact), programmatic criteria (e.g., based on the process for data collection), or statistical criteria (group measures that are correlated and thus seem to capture a similar quality construct and weight based on their variance across organizations to optimize ability to discriminate performance). Different choices may lead to different weighting of the individual measures in the overall score (as well as in domains). Use of grouping based on clinical criteria may lead to more clinically cohesive domains, but may also mask important variation if the clinical measures are weakly correlated within domains (which they often are). On the other hand, use of statistical criteria will lead to more statistically coherent domains, which may better capture latent dimensions of quality but may be harder to interpret and may reflect factors unrelated to quality that induce correlations in performance (e.g., variation in provider coding practices or patient risk). Thus, methodological choices often present a trade-off between clinical coherence and statistical coherence.

In this study, we explore the impact of alternative aggregation approaches on measure groupings and overall scores for Medicare ACOs, the most prominent of the new population-based payment models.

METHODS

Data

We used publicly available CMS files containing 2014 quality scores for Medicare Pioneer and Shared Savings Program (SSP) ACOs on 33 measures (Centers for Medicare and Medicaid Services 2015a,b). We dropped 27 ACOs with at least one missing measure, leaving 326 Pioneer and SSP ACOs.

Aggregation Approaches

We consider three important aspects of aggregation: (1) grouping of individual measures into domains, (2) weighting of individual measures into a domain score, and (3) weighting of domain scores into an overall score.

CMS Aggregation Method. Prior to aggregation, CMS created benchmarks for each measure using 2012 Medicare FFS data. An ACO must meet these benchmarks to earn the quality points for that measure. For example, an ACO that performs below the 30th percentile receives 0 points for that measure; an ACO that performs at or above the 90th percentile receives 100 points.

CMS grouped the 33 measures into four domains based on their clinical similarities. Domain scores, expressed as percentage of possible points, were computed by equally weighing each individual measure within a domain. CMS then averaged all of the domain-specific scores to get an overall score.

Our Alternatives. Following CMS' methodology described above, we used FFS benchmarks to rescale each measure to a 0-to-100 scale. In each alternative (Table 2), in place of using CMS' clinical domains we group measures based on their empirical relationships to one another using exploratory factor analysis (EFA; Samuel et al. 2015; Martsof, Carle, and Scanlon 2017). Measures that are more highly correlated with each other tend to be grouped together. We imposed a PROMAX rotation to allow the empirical domains to be correlated with one another. The number of factors retained was determined using scree plots and Kaiser's criterion. We placed each measure in the domain with the largest factor loading. We examined internal consistency of CMS and empirical domains by calculating Cronbach's alpha coefficients. In sensitivity analyses, we also examined groups of measures generated using principal component analysis (PCA).

This EFA approach differs from the CMS approach of grouping based on common clinical focus. Grouping quality measures into domains based on their clinical relevance to one another may mask important aspects of underlying quality. For example, an organization that effectively controls type II diabetes may not regularly conduct depression screenings. Combining organizations' scores for two different dimensions of quality (e.g., diabetes control and depression screening) in the same domain, we cannot distinguish between an organization that controls diabetes perfectly but fails to screen patients for depression and an organization that does a mediocre job on both measures.

Our seven alternative aggregation approaches use the empirical domains formed by the EFA, varying the weighting within and across domains (Table 2). The first approach (following CMS) equally weights measures within the domain to generate a domain score and equally weights across the domains for the overall score. The other six approaches

use the factor loadings to weight the individual measures into a domain score, giving more weight to measures with higher loadings. They also differ in how they weight across domains to calculate the overall score. Specifically, the second approach equally weights across the domains (following CMS) and the third weights domains based on their variability (measured by their eigenvalue) to get the overall score. This places additional weight on domains with larger observed variation and provides greater statistical power to discriminate performance among organizations, but at the risk of emphasizing domains in which factors unrelated to quality cause additional variation in performance (such as patient risk). Approaches 4–7 examine potential effects of upweighting one of the domains, as policy makers might do if it is determined that some measures are more valid or more meaningful than others.

Analysis

We examined how choices in aggregation methods affect ACO quality scores using several metrics. First, because CMS uses domain scores to identify low-performing ACOs, we examined the impact of alternative grouping methods in detecting low-performing ACOs. Specifically, CMS gives injunctions, ranging from formal warnings to contract termination and withholding of shared savings, if an ACO scores below 70 percent (on benchmark rescaled score described above) in any of the four clinically formed domains (Centers for Medicare and Medicaid Services 2011). We computed the fraction of ACOs that would receive injunctions using CMS versus alternative aggregation methods. Second, we estimated intraclass correlations using a mixed-effects linear model and computed Cohen's kappa to measure the agreement between CMS quality score and various alternative grouping and weighting schemes (Liu et al. 2016). Finally, because an ACO's overall score determined the percentage of its savings (up to the performance payment limit) it was paid by CMS (Centers for Medicare and Medicaid Services 2014), we examined the fraction of ACOs with overall quality scores that change $>|5|$ percent and $>|10|$ percent across alternative methods. Average shared savings in 2014 were \$252 per capita, so a 5 percent change in overall quality score represents a \$12.60 increase or decrease and a 10 percent change represents a \$25.20 increase or decrease in savings per capita on average (Introcaso and Berger 2015). All analyses were conducted in R.

RESULTS

Among CMS domains, ACO performance was highest (relative to benchmarks) and least variable in the Patient/Caregiver Experience domain (mean percent of points earned = 87.6 percent), and all domain-specific scores had medians above 75 percent. ACO performance was highly variable in the At-Risk Population domain (interquartile range of percent of points earned = 71.1 percent to 88.2 percent), with a significant number of ACOs with performance below 40 percent (Figure S1a; summary statistics of ACO performance on the 33 quality measures also presented in Supporting Information).

EFA identified four factors or empirical domains (Table 1; see Table S2 for complete factor loadings). The measures included in the empirical domains differed from those in the CMS domains, although the measures in the CMS Patient/Caregiver Experience domain were similar to those statistically grouped in Domain 2. Individual measures found in all four CMS domains comprised empirical Domain 1. Similarly, Domain 3 included measures drawn from the CMS Care Coordination/Patient Safety and Preventive Health domains. Finally, only some measures from the CMS Care Coordination/Patient Safety domain composed Domain 4.

The internal consistency of each of the CMS domains ranged from $\alpha = 0.13$ for Care Coordination/Patient Safety to $\alpha = 0.83$ for Preventive Health; whereas the internal consistency of the empirical domains from the four factor model were generally higher by construction (ranging from $\alpha = 0.52$ to 0.86; Table S3). Using PCA without rotation, the internal consistencies of resulting components were substantially lower than they were for factors generated through EFA. When we applied PROMAX rotation to the PCA, results were very similar to those from the EFA with PROMAX rotation.

Domain-specific scores using the empirical grouping approach also resulted in larger differences across ACOs, particularly for Domain 4 (Figure S1b). Of 326 ACOs, 46.6 percent scored below 70 percent in at least one CMS domain and 51.2 percent scored below 70 percent in at least one empirical domain. Thirty-eight percent were low-performing in both cases, 8.9 percent were low-performing in the CMS but not empirical domains, and 13.5 percent were low-performing in the empirical but not CMS domains.

We found substantial agreement between the two grouping methods when domains were equally weighted ($\kappa = 0.80$; Figure 1a and Table 2),

Table 1: Factor Analysis of 2014 MSSP and Pioneer Quality Measure Performance

<i>Normative (CMS) Domains</i>	<i>Measure</i>	<i>Empirical Domains</i>			
		<i>Domain 1</i>	<i>Domain 2</i>	<i>Domain 3</i>	<i>Domain 4</i>
Patient/ caregiver experience	CAHPS: Getting timely care, appointments, and information		X		
	CAHPS: How well your providers communicate		X		
	CAHPS: Patients' rating of provider		X		
	CAHPS: Access to specialists		X		
	CAHPS: Health promotion and education		X		
	CAHPS: Shared decision making		X		
	CAHPS: Health status/functional status	X			
Care coordination/ patient Safety	Risk standardized all-condition readmissions				X
	ASC admissions: COPD or asthma in older adults				X
	ASC admissions: heart failure				X
	Percent of PCPs who successfully qualify for an EHR program incentive payment	X			
	Medication reconciliation			X	
	Falls: screening for future fall risk	X			
Preventive health	Preventive care and screening: influenza immunization	X			
	Pneumonia vaccination status for older adults	X			
	Preventive care and screening: BMI screening and follow-up			X	
	Preventive care and screening: tobacco use: screening and cessation intervention	X			
	Preventive care and screening: screening for clinical depression and follow-up plan	X			
	Colorectal cancer screening	X			
	Breast cancer screening	X			
	Preventive care and screening: screening for high blood pressure and follow-up documented			X	

continued

Table 1. *Continued*

Normative (CMS) Domains	Measure	Empirical Domains			
		Domain 1	Domain 2	Domain 3	Domain 4
At-risk population	Diabetes all-or-nothing composite	X			
	Diabetes: Hemoglobin A1c poor control (>9%)	X			
	Controlling high blood pressure (<140/90)	X			
	Ischemic vascular disease: complete lipid panel and LDL control (<100 mg/dl)	X			
	Ischemic vascular disease: use of aspirin of another antithrombotic	X			
	Heart failure: beta-blocker therapy for left ventricular systolic dysfunction	X			
	Coronary artery disease all-or- nothing composite	X			

regardless of how the individual measures were weighted within the domains. However, some ACOs experienced differences in overall scores. When using the empirical domains, equal weights within the domains and equal weights across the domains, 22 percent of ACOs experienced >|5| percent difference in overall quality score and 3 percent of ACOs experienced a >|10| percent difference in quality score (Figure 1b). Using factor loadings to weight within the domains and keeping all else the same, the number of affected ACOs was comparable (21.5 percent experienced >|5| percent difference, 1.8 percent experienced >|10| percent difference).

However, when we used the empirical domains and changed both the within- and across-domain weighting to reflect factor loadings and eigenvalues, respectively, overall scores were very similar to those computed using CMS methods ($\kappa = 0.95$; Table 2).

When we used empirical domains and applied factor loading weights within domains but upweighted one of the four domains, we found moderate-to-high agreement between overall scores and those from CMS ($\kappa = 0.51$ to 0.93). Upweighting Domain 1, the domain containing nearly all measures in the CMS Patient/Caregiver Experience domain, led to almost no differences in overall quality scores ($\kappa = 0.93$). However, upweighting either Domain 3 or 4 led to substantially weaker agreement with the CMS approach.

Figure 1: (a) Scatterplot of 5 percent Difference in Overall Quality Score Calculated with Empirical versus CMS Domains, Equal Weights within and Across Domains. (b) Scatterplot of 10 percent Difference in Overall Quality Score Calculated with Empirical versus CMS Domains, Equal Weights within and Across Domains

Table 2: Percentage of ACOs that would Experience Difference in Overall Quality Score and Correlations between Overall Quality Scores calculated Using CMS and Alternative Methods

CMS method	> 5% Decrease		> 5% Increase		≤5% Difference		> 10% Decrease		> 10% Increase		≤10% Difference		Correlation		Cohen's Kappa (Weighted)*		Intraclass Correlation *	
	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference	Reference	Difference
(1) Empirical domains + Equal weights within + Equal weights across	14.4		7.7		77.9		1.5		1.2		97.2		0.88		0.80 [0.76, 0.84]		0.86 [0.83, 0.89]	
(2) Empirical domains + Factor loadings within + Equal weights across	11.7		9.8		78.5		1.2		0.6		98.2		0.88		0.78 [0.73, 0.82]		0.87 [0.84, 0.89]	
(3) Empirical domains + Factor loadings within + Eigenvalue weights across	0		0.3		99.7		0		0		100		0.99		0.95 [0.94, 0.97]		0.98 [0.98, 0.99]	
(4) Upweight Domain 1 [†]	1.8		0		98.2		0		0		100		0.98		0.93 [0.91, 0.95]		0.98 [0.97, 0.98]	
(5) Upweight Domain 2 [†]	0.6		41.1		58.3		0		13.5		86.5		0.81		0.79 [0.74, 0.83]		0.61 [0.54, 0.68]	
(6) Upweight Domain 3 [†]	34.4		16.9		48.8		12.0		5.2		82.8		0.63		0.48 [0.40, 0.56]		0.61 [0.54, 0.68]	
(7) Upweight Domain 4 [†]	52.5		12.0		35.6		24.2		4.6		71.2		0.56		0.51 [0.43, 0.60]		0.44 [0.36, 0.53]	

Notes. *95% confidence intervals in brackets.
[†]Uses empirical domains and factor loadings (weighting individual measures to produce domain scores). Corresponding upweighted domains are 0.625, and weights for remaining domains are each 0.125, so that total weights sum to 1.

DISCUSSION

In the setting of recent efforts to link quality and payment, the development and selection of quality measures have been widely discussed (Cassel and Kronick 2015; Albright et al. 2016). However, little attention has been paid to the methods used to aggregate measures into a manageable number of domains and an overall score.

We find substantial agreement between the CMS approach and alternative empirical approaches. Specifically, using a fully empirical approach, we find overall scores that were very similar to those from the CMS method. However, we observe substantial differences when we deliberately upweight either of two of the four empirical domains. This upweighting could result in over 50 percent of ACOs experiencing more than a 5 percent change in overall scores and thus savings, and 17–29 percent of ACOs experiencing more than a 10 percent change. Given that measures are not highly correlated, producing an overall score inherently depends on how different measures are weighted. Consequently, there cannot be a gold standard. The best aggregation scheme will depend on what we value and how our aggregation measures reflect those values. The appeal of different domain weighting schemes depends on how payers normatively value the different domains and how much they value statistical versus clinical coherence of domains. Our analyses suggest that different value-based weightings could have substantial implications.

Our analysis has several limitations. First, methods other than those we consider are certainly possible. For example, separate domains are not necessary for payment purposes, and individual measures could be weighted into a summary measure directly (e.g., either using normative weights or according to their association with patient outcomes; Shwartz, Restuccia, and Rosen 2015). Alternatively, data envelopment analysis with cost as an input and the set quality measures as output could be used to derive a summary measure of value (Dowd et al. 2014).

Second, the statistical methods we investigate include *ex post* calculations, while the CMS methods were clearly delineated *ex ante*. Thus, ACOs are able to react to the current measurement approach in a manner that they would not be able to under *ex post* approaches. However, the statistical weights could be derived with lagged data, and thus, an *ex ante* approach is feasible. Third, we cannot assess the impact of alternative approaches in all settings. Our conclusions apply only to the data we analyzed. Specifically, when examining ACO overall scores, grouping and weighting measures based on

their statistical relationships generally produce only modest changes. However, that is not a general property of quality measurement.

Finally, we focus on measuring quality. It is beyond the scope of this work to assess whether financial incentives or other strategies should be used to improve it. Nevertheless, moving forward in a world where quality measurement is of growing importance, it will be crucial to assess the sensitivity of various aggregation methods before rolling out alternative payment policies.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by a grant from the Laura and John Arnold Foundation. The views presented here are those of the author and not necessarily those of the Laura and John Arnold Foundation, its directors, officers, or staff. Dr. McWilliams reports serving as a consultant to Abt Associates, Inc. on an evaluation of the ACO Investment Model.

Disclosure: None.

Disclaimer: None.

REFERENCES

- Albright, B. B., V. A. Lewis, J. S. Ross, and C. H. Colla. 2016. "Preventive Care Quality of Medicare Accountable Care Organizations: Associations of Organizational Characteristics with Performance." *Medical Care* 54 (3): 326–35.
- Austin, J. M., A. K. Jha, P. S. Romano, S. J. Singer, T. J. Vogus, R. M. Wachter, and P. J. Pronovost. 2015. "National Hospital Ratings Systems Share Few Common Scores and May Generate Confusion Instead of Clarity." *Health Affairs* 34 (3): 423–30.
- Cassel, C., and R. Kronick. 2015. "Learning from the Past to Correct the Future." *Journal of the American Medical Association* 314 (9): 875–6.
- Centers for Medicare and Medicaid Services. 2011. "Guide to Quality Performance Scoring Methods for Accountable Care Organizations" [accessed on August 1, 2017]. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2012-11-ACO-quality-scoring-supplement.pdf>.
- Centers for Medicare and Medicaid Services. 2014. "Medicare Shared Savings Program Shared Savings and Losses and Assignment Methodology Specifications" [accessed on June 1, 2017]. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf>.

- Centers for Medicare and Medicaid Services. 2015a. "Medicare Pioneer ACO Performance Year 3 Quality and Financial Results" [accessed on August 1, 2017]: 3. Available at <https://innovation.cms.gov/Files/x/pioneeraco-fncl-py3.pdf>.
- Centers for Medicare and Medicaid Services. 2015b. "Medicare Shared Savings Program Quality Measure Benchmarks for the 2014 and 2015 Reporting Years" [accessed on August 1, 2017]: 1–7. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks.pdf>.
- Centers for Medicare and Medicaid Services. 2015c. "Medicare Shared Savings Program Accountable Care Organizations Performance Year 2014 Results" [accessed on August 1, 2017]. Available at <https://data.cms.gov/ACO/Medicare-Shared-Savings-Program-Accountable-Care-O/ucce-hhpu>.
- Dowd, B., T. Swenson, R. Kane, S. Parashuram, and R. Coulam. 2014. "Can Data Envelopment Analysis Provide a Scalar Index of "Value"?" *Health Economics* 23: 1465–80.
- Gandhi, T. K., E. F. Francis, A. L. Puopolo, H. R. Burstin, J. S. Haas, and T. A. Brennan. 2002. "Inconsistent Report Cards: Assessing the Comparability of Various Measures of the Quality of Ambulatory Care." *Medical Care* 40 (2): 155–65.
- Health Care Transformation Task Force. 2018. "About Us" [accessed on January 10, 2018]. Available at <http://hcttf.org/aboutus/>
- Introcaso, D., and G. Berger. 2015. "MSSP Year Two: Medicare ACOs Show Muted Success." *Health Affairs Blog*. Available at <http://healthaffairs.org/blog/2015/09/24/mssp-year-two-medicare-acos-show-muted-success/>
- Jha, A. K., Z. Li, E. J. Orav, and A. M. Epstein. 2005. "Care in U.S. Hospitals—The Hospital Quality Alliance Program." *New England Journal of Medicine* 353 (3): 265–74.
- Kessell, E., V. Pegany, B. Keolanui, B. D. Fulton, R. M. Scheffler, and S. M. Shortell. 2015. "Review of Medicare, Medicaid, and Commercial Quality of Care Measures: Considerations for Assessing Accountable Care Organizations." *Journal of Health Politics, Policy and Law* 40: 761–96.
- Liu, J., W. Tang, G. Chen, Y. Lu, C. Feng, and X. M. Tu. 2016. "Correlation and Agreement: Overview and Clarification of Competing Concepts and Measures." *Shanghai Archives of Psychiatry* 28 (2): 115–20.
- Martsof, G. R., A. C. Carle, and D. P. Scanlon. 2017. "Creating Unidimensional Global Measures of Physician Practice Quality Based on Health Insurance Claims Data." *Health Services Research* 52 (3): 1061–78.
- Muhlestein, D., N. Burton, and L. Win. 2017. "The Changing Payment Landscape of Current CMS Payment Models Foreshadows Future Plans." *Health Affairs Blog*. Available at <https://www.healthaffairs.org/doi/10.1377/hblog20170203.058589/full/>
- Samuel, C. A., A. M. Zaslavsky, M. B. Landrum, K. Lorenz, and N. L. Keating. 2015. "Developing and Evaluating Composite Measures of Cancer Care Quality." *Medical Care* 53 (1): 54–64.

- Shwartz, M., J. D. Restuccia, and A. K. Rosen. 2015. "Composite Measures of Health Care Provider Performance: A Description of Approaches." *Milbank Quarterly* 93 (4): 788–825.
- U.S. Senate. 2015. "U.S. Senate Roll Call Votes for H.R. 2 Medicare Access and CHIP Reauthorization Act of 2015" [accessed on August 1, 2017]. Washington, D.C.: 114th Congress – 1st Session. Available at https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=114&session=1&vote=00144.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix SA1: Author Matrix.

Figure S1. (a) Boxplot of 2014 ACO Performance in 4 CMS Quality Domains (Domains Equally Weighted to Calculate Overall Quality Score). (b) Boxplot of 2014 ACO Performance in 4 Empirical Domains (Domains Equally Weighted to Calculate Overall Quality Score).

Table S1. Summary Statistics of 2014 MSSP and Pioneer Quality Measure Performance.

Table S2. Factor Analysis of 2014 MSSP and Pioneer Quality Measure Performance.

Table S3. Cronbach's Alphas for Clinical and Statistical Domains.