NATIONAL QUALITY FORUM

Memo

TO: NQF Members

FR: NQF Staff

RE: Voting Draft Report: *NQF-Endorsed Measures for Endocrine Conditions: Cycle 1*

DA: June 6, 2014

Background

Endocrine conditions result from disorders of the endocrine system, most often when either too much or too little of a particular hormone is produced.[1] In the United States, two of the most common endocrine disorders are diabetes and osteoporosis.[2] Diabetes, a group of diseases characterized by high blood glucose levels, affects many as 25.8 million Americans and ranks as the 7th leading cause of death in the United States.[3] Osteoporosis, a bone disease characterized by low bone mass and density, affects an estimated 9% of U.S. adults aged 50 and over.[4] Currently, NQF's Endocrine portfolio includes 39 measures. Many of the diabetes measures in the portfolio are among NQF's longest-standing measures.

NQF selected the Endocrine measure evaluation project to pilot frequent submission and evaluation of measures than what is possible in our current 3-year measure maintenance cycle. This 22-month project will include three full endorsement "cycles," allowing for the submission and review of both new and previously-endorsed measures every six months. In addition, this project is one of the first to transition to the use of Standing Committees. The 20 Standing Committee members recommended 13 out of 17 measures initially submitted for endorsement.

Comments and Revised Voting Report

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF soliciting member and public comments prior to the evaluation of the measures via an online tool located on the project webpage. Third, NQF opens a 30-day comment

---

[1] WebMD. Endocrine Disorders. 2014;March 14 A.D.

[2] Golden SH, Robinson KA, Saldanha I, et al. Clinical review: Prevalence and incidence of endocrine and metabolic disorders in the United States: a comprehensive review. The Journal Of Clinical Endocrinology And Metabolism, 2009;94(6):1853-1878.

[3] Centers for Disease Control and Prevention. National Diabetes Fact Sheet: National estimates and general information on diabetes and prediabetes in the United Staes, 2011. 2014;Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention

[4] Looker AC, Borrud LG, Dawson-Hughes B, et al. Osteoporosis or low bone mass at the femur neck or lumbar spine in older adults: United States, 2005-2008. NCHS Data Brief, 2012;(93):1-8.

period to both members and the public after measures have been evaluated by the full committee and once a report of the proceedings has been drafted.

Pre-evaluation comments

The pre-evaluation comment period was open from January 21-February 7, 2014 for 8 of the 17 measures under review. A total of 76 pre-evaluation comments were received, the majority of which pertained to, and were supportive of, the three newly-submitted osteoporosis measures. All of these pre-evaluation comments were provided to the Committee prior to their initial deliberations held during the workgroups calls.

Post-evaluation comments

The 30-day post-evaluation comment was open from April 03, 2014 to May 2, 2014. During this commenting period, NQF received 83 comments from 10 member organizations:

Consumers – 1	Professional – 3
Purchasers –0	Health Plans – 1
Providers – 1	QMRI – 1
Supplier and Industry – 3	Public & Community Health - 0

A complete table of comments submitted pre- and post-evaluation, along with the responses to each comment and the actions taken by the Standing Committee, is posted to the Endocrine project page on the NQF website, along with the measure submission forms.

The Committee reviewed all comments received and considered the pre-meeting comments prior to making an endorsement recommendation. The Committee also responded to all post-evaluation comments. Revisions to the draft report and the accompanying measure specifications are identified as red-lined changes. (Note: Typographical errors and grammatical changes have not been red-lined, to assist in reading.)

Committee Discussion of Comments

Only one overall theme was identified in the post-evaluation comments: that of support for the recommended measures. Specifically, a total of 48 of the comments received expressed support for (but no additional questions or concerns regarding) the Committee's decisions to recommend 13 of the evaluated measures for endorsement. Several additional comments also expressed support of the Committee's decisions, but also requested clarification regarding measure specifications.

While there were several comments that were not supportive of the Committee's recommendations, most simply explained the reasoning but did not offer additional data to promote additional discussion of the measure.

Measure Specific Comments

#2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus

During the evaluation of the measure at the in-person meeting, the Committee questioned the validity of this measure because it did not exclude patients who switch from oral agents to insulin during the measurement period. The Committee noted that in older adults, transition to insulin (and associated discontinuation of oral medications) is common and that the measure as specified would incorrectly categorize such patients as non-adherent; members also expressed concern that the measure as specified might incentivize physicians to leave patients on oral diabetes agents rather than switch them to insulin when appropriate. The Committee encouraged the developer to quantify the number of patients who transitioned to insulin and, if possible, revise the measure to exclude those patients.

NQF received 3 post-evaluation comments regarding this measure, each of which concurred with the Committee's decision not to recommend the measure for endorsement unless above concerns are addressed.

As requested by the Committee, the measure developer (FMQAI, on behalf of CMS) conducted additional analysis to ascertain how many patients switched from oral diabetes agents to an insulin-only therapy. Results from analyses of a 10-state sample indicate that 13.1% of patients made this switch. Developers subsequently re-specified the measure so as to:

- Limit the number of days in the denominator for those with a switch from oral diabetes agents to insulin-only therapy
- Compute an overall percentage of days covered value for those who switched between oral drug classes

Because these changes in specifications were substantial, the developers also re-tested the newly specified measure for reliability and validity. The changes to the specifications and the new testing results are detailed in the attached response from the developer (***see Appendix***).

Action Taken:

- As requested by the Committee, the measure developer conducted additional analysis and found that 13.1% of patients in their 10-state sample switched from oral diabetes agents to an insulin-only therapy. Based on these results, the developer re-specified the measure to 1) limit the number of days in the denominator for those with a switch from oral diabetes agents to insulin-only therapy and 2) compute an overall percentage of days covered value for those who switched between oral drug classes; they also re-tested the newly specified measure for reliability and validity. After discussion, the committee agreed to re-vote on the measure. Upon re-vote, the Committee agreed that the analysis and re-specification of the measure addressed their initial concerns with the validity of the measure. After additional discussion, the Committee also voted on the Feasbility and Usability and Use criteria, and ultimately recommended the re-specified measure for endorsement.

2418: Discharge Instructions – Emergency Department

During the evaluation of this measure at the in-person meeting, the Committee agreed that this measure did not meet the Evidence subcriterion under Importance to Measure and Report. Specifically, the Committee noted that there is minimal evidence indicating that provision of written discharge instructions improves care for osteoporosis patients or has any impact on outcomes such as prevention of future fractures. Committee members expressed concern that because either provision of discharge instructions or coordination with a Fracture Liaison Service would meet the measure, facilities might focus on discharge instructions instead of FLS use, even though the supporting evidence is weak.

NQF received 4 post-evaluation comments regarding this measure, each of which reflected disagreement with the Committee's decision not to recommend the measure for endorsement. However, none of the comments referenced any additional evidence to show that provision of discharge instructions would help to prevent future fractures.

Action Taken:

- Committee members noted that the 2013 Ganda, et al. article, which found that patient education alone does not seem to increase rates of treatment initiation, had been discussed during the in-person meeting. Members agreed that no additional information was presented to change their evaluation of the measure and therefore declined to revote on the measure.

0055: Comprehensive Diabetes Care: Eye Exam (retinal) performed

NQF received 7 post-evaluation comments regarding this measure. Four of these comments were supportive of the measure and the Committee's decision to recommend the measure for endorsement.

Two of the comments requested clarification as to why women with polycystic ovarian syndrome are excluded from the measure.

> ***Developer response regarding exclusion of PCOS***: Thank you for your comment. This is a long-standing exclusion which was recommended by our joint NCQA-AMA-PCPI expert panel when the diabetes measures were first developed. NCQA will take this comment into consideration during our next re-evaluation of the diabetes care measures.

One commenter noted that the specifications for the measure include CPT codes **92227** (remote imaging for detection of retinal disease) and **92228** (remote imaging for monitoring and management of active retinal disease). The commenter stating that the "*use of these codes to demonstrate compliance with the measure raises significant quality concerns and is contrary to the American Diabetes Association (ADA) and the AOA's own clinical guidelines for patients with diabetes*" and that "*by including the remote retinal imaging codes in the measure specifications NCQA is in effect indicating that remote retinal imaging is sufficient eye care for a patient with diabetes*."

> ***Developer response regarding remote imaging CPTs***: Thank you for your comment. NCQA will review the use of these codes with our expert panels and if appropriate, update the Diabetic Retinal Screening

value set. All changes will be taken through NCQA's established governance structure.

Finally, one commenter suggested that this measure be aligned with the new age specifications agreed to by the developer for measure #0056 (i.e., NCQA removed the upper age restriction so that the measure now applies to diabetes patients ages 18 and older).

Developer response regarding harmonization of age range: Thank you for your comment. NCQA will evaluate appropriate age thresholds during our next re-evaluation of the diabetes care measures.

Actions Taken:

- Committee members noted that the ADA guidelines, as well as other evidence, indicate that retinal photographs are acceptable and therefore did not recommend a change to the specifications of the measure.
- The Committee agreed that because complications of diabetes disproportionately affects older patients, the measure developer should consider changing the specifications to include those aged 18 and older rather than including only those aged 18-75.

2362: Glycemic Control – Hyperglycemia and 2363: Glycemic Control – Hypoglycemia

NQF received 6 post-evaluation comments regarding this measure. One commenter submitted 2 comments, noting low reliability scores for one of the hospitals included in the testing of the measure and questioning the reliability of the measure for smaller facilities. The commenter also expressed the desire that the measures be made consistent. In addition, one commenter questioned the need for these measures while another expressed support for the measures.

Developer response for measure #2362 (hyperglycemia), regarding reliability: Thank you for your comments. We believe these measures are an important step in addressing the recommendations of the National Action Plan for ADE Prevention and will assist hospitals in the identification of both hypoglycemia and hyperglycemia incidence and factors associated with inadequate glycemic control. In addition, these measures are specifically recommended in the National Action Plan for ADE Prevention. Concerning measure score reliability for the hypoglycemia measure, it is correct that the smallest facility tested (i.e., a critical access hospital) had inadequate reliability; however, the other facility had a score of 0.67, which would indicate the measure is closely approaching the reliability threshold of 0.7. If the measure is implemented, we will monitor reliability carefully for small facilities. Regarding moderate hypoglycemia, the specifications originally submitted to NQF included an optional numerator for mild hypoglycemia. After discussion with the steering committee, the decision was made to remove this optional numerator. While we agree that moderate hypoglycemia is an important internal quality indicator, we do not believe it would be appropriate to publicly report, since many cases of moderate hypoglycemia are not preventable. Therefore, the steering committee decided that the endorsed measure should only include the measure numerator that is publicly reported, and the optional numerator was removed. We believe implementation of both hyperglycemia and hypoglycemia measures as a balanced pair will encourage hospitals to put in place interventions and

appropriate protocols to manage blood glucose and thereby improve glycemic control including but not limited to mild to moderate hypoglycemia. Regarding measure consistency, the measures are designed to measure two very different events clinically. Hyperglycemia is usually sustained and can occur in patients that do not have a current diagnosis of diabetes; whereas, severe hypoglycemia is a relatively rare event that typically occurs after the administration of an anti-diabetic agent. We do not feel further alignment of the definitions is feasible without compromising measure validity.

Developer response for measure #2363 (hypoglycemia), regarding reliability: Thank you for your comments. We believe these measures are an important step in addressing the recommendations of the National Action Plan for ADE Prevention and will assist hospitals in the identification of both hypoglycemia and hyperglycemia incidence and factors associated with inadequate glycemic control. In addition, these measures are specifically recommended in the National Action Plan for ADE Prevention. Concerning measure score reliability for the hypoglycemia measure, it is correct that the smallest facility tested (i.e., a critical access hospital) had inadequate reliability; however, the other facility had a score of 0.67, which would indicate the measure is closely approaching the reliability threshold of 0.7. If the measure is implemented, we will monitor reliability carefully for small facilities. Regarding moderate hypoglycemia, the specifications originally submitted to NQF included an optional numerator for mild hypoglycemia. After discussion with the steering committee, the decision was made to remove this optional numerator. While we agree that moderate hypoglycemia is an important internal quality indicator, we do not believe it would be appropriate to publicly report, since many cases of moderate hypoglycemia are not preventable. Therefore, the steering committee decided that the endorsed measure should only include the measure numerator that is publicly reported, and the optional numerator was removed. We believe implementation of both hyperglycemia and hypoglycemia measures as a balanced pair will encourage hospitals to put in place interventions and appropriate protocols to manage blood glucose and thereby improve glycemic control including but not limited to mild to moderate hypoglycemia. Regarding measure consistency, the measures are designed to measure two very different events clinically. Hyperglycemia is usually sustained and can occur in patients that do not have a current diagnosis of diabetes; whereas, severe hypoglycemia is a relatively rare event that typically occurs after the administration of an anti-diabetic agent. We do not feel further alignment of the definitions is feasible without compromising measure validity.

Action Taken:

- After review and discussion of the comments and the response from the developer, the Committee supported the construction of the measures and accepted the explanation of the developer regarding reliability.
- The Committee agreed that these measures meet the three subcriteria under Importance to Measure and Report (i.e., evidence, opportunity for improvement, and high priority).

NQF Member Voting

Information for electronic voting has been sent to NQF Member organization primary contacts. Accompanying comments must be submitted via the online voting tool.

Please note that voting concludes on June 20, 2014 at 6:00 pm ET – no exceptions.

Appendrix

Response to Steering Committee Concerning NQF 2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus

Submitted By: FMQAI on behalf of CMS

May 14, 2014

The NQF Endocrine Steering Committee, which met on February 27, 2014, requested a revision to the measure specifications that would account for patients who switched from oral diabetes agents to insulin-only during the measurement period. In addition, FMQAI received a public comment requesting the measure account for patients using incretin mimetics (i.e., exenatide and liraglutide). This document provides results from additional analyses conducted to evaluate these scenarios and recommendations regarding revision to the measure specifications.

1. **What proportion of patients in the denominator use insulin and incretin mimetics?**
   In the 10-state sample, 24.3% (150,774/620,934) of the denominator population had at least one claim for insulin, and 2.85% (17,690/620,934) had at least one claim for incretin mimetics. Since both insulin and incretin mimetics have the indication to be used as the sole medication therapy for diabetes, the impact of medication switching should be evaluated.

2. **What proportion of individuals switched from oral diabetes agents (ODAs) to insulin- or incretin mimetic-only therapy during the measurement period?**
   In the 10 state sample, among individuals who had at least one claim for insulin (n=150,774), 13.1% switched from ODAs to an insulin-only therapy. Among individuals who had at least one claim for incretin mimetics (n=17,690), 8.8% switched from ODAs to an incretin mimetic-only therapy. This suggests that measure rates would be falsely lowered by not accounting for switching in the measure specification.

3. **How are individuals who switched from ODAs to insulin or incretin mimetics identified?**
   Individuals switching to insulin or incretin mimetics are identified by having at least one claim for any type of insulin or incretin mimetic after the end of the days' supply of the last ODA prescription.

4. **How would adherence to ODAs be calculated for individuals who switched to insulin- or incretin mimetics-only during the measurement period?**
   For these individuals, the ODA measurement period is set to the end date of the days' supply of the last ODA prescription during the measurement year. Therefore, adherence is only calculated while the patient is taking ODAs and there is no disincentive for providers to switch their patients to insulin or incretin mimetics-only.

5. **Should the measure specifications also address switching between ODAs?**
   The current measure specifications calculate an individual's adherence to each class of ODAs separately (e.g., biguanides, sulfonylureas, etc.) and the individual would need to achieve a Proportion of Days Covered (PDC) $\geq$0.8 for at least one of the classes to qualify for the numerator. Since individuals might be switched from one ODA to other and it would be difficult to operationalize all the potential switching that would occur, FMQAI proposes a second revision

of the specifications that would calculate medication adherence to the whole category of ODAs regardless of the class. Therefore, as long as the proportion of days covered across all ODAs was at least 0.8, the individual would qualify for the numerator.

6. **What are the impacts from the proposed specification changes on the measure rates and scientific acceptability?**
   On average, the mean measure rate has increased by approximately 1-3% across each level measured and a substantial gap in performance remains with a mean rate of approximately 76% overall (Appendix A). Variation in performance remains approximately 10-14% between the 10th and 90th percentile (Appendix A). Reliability remains adequate across all levels of measurement and convergent validity is improved (Appendix B).

7. **Based on the review, what are the final recommendations and conclusions for the Steering Committee?**
   FMQAI recommends revising the specifications to account for individuals switching to insulin- or incretin mimetic-only therapy and to calculate adherence across all ODA drug classes collectively. Proposed revisions to the specifications are shown below in red.

Revised Specifications

Numerator Statement: Individuals with diabetes mellitus who have at least two claims for ODAs and have a PDC of at least 0.8 for oral diabetes agents.

Numerator Details:
The numerator is defined as individuals with a PDC of 0.8 or greater.

The PDC is calculated as follows:
• PDC Numerator: The PDC numerator is the sum of the days covered by the days' supply of all drug claims in the ODA class. The period covered by the PDC starts on the day the first prescription is filled (index date) and lasts through the end of the measurement period, or death, whichever comes first. For prescriptions with a days' supply that extends beyond the end of the measurement period, count only the days for which the drug was available to the individual during the measurement period. If there are prescriptions for the same drug (generic name) on the same date of service, keep the prescription with the largest days' supply. If prescriptions for the same drug (generic name) overlap, then adjust the prescription start date to be the day after the previous fill has ended.
• PDC Denominator*: The PDC denominator is the number of days from the first prescription date through the end of the measurement period, or death date, whichever comes first.

*Individuals switching to insulin or incretin mimetics are identified by having at least one claim for any type of insulin or incretin mimetics after the end of the days' supply of the last ODA prescription. For these individuals, the ODA measurement period is set to the end date of the days' supply of the last ODA prescription during the measurement year.

Denominator Statement: Individuals at least 18 years of age as of the beginning of the measurement period with diabetes mellitus and at least two claims for oral diabetes agents during the measurement period (12 consecutive months).

Appendix A – Meaningful Differences in Performance

Table A1. Summary of State Level Performance

	n	Mean	Median	Min	Max	STD	IQR	P10	P25	P50	P75	P90
Original Measure	10	73.9%	75.2%	67.7%	80.8%	4.0%	5.7%	68.2%	70.3%	75.2%	76.0%	78.4%
Revised Measure	10	76.6%	77.9%	70.2%	83.2%	3.9%	5.2%	70.9%	73.3%	77.9%	78.5%	81.0%

Based on the revised measure, four of the 10 states (40.0%) had scores statistically significantly lower than the mean and six states (60.0%) had scores significantly higher than the mean. Measure rates ranged from 70.2% in Mississippi to 83.2% in Iowa, indicating suboptimal performance across all 10 states.

Table A2. Summary of Plan Level Performance

	n	Mean	Median	Min	Max	STD	IQR	P10	P25	P50	P75	P90
Original Measure	40	74.2%	75.0%	60.7%	83.6%	5.7%	6.8%	66.0%	71.2%	75.0%	78.0%	80.8%
Revised Measure	40	76.7%	77.5%	63.2%	86.3%	5.4%	6.4%	69.2%	73.9%	77.5%	80.4%	82.1%

Based on the revised measure at the plan level, 27.5% of providers were statistically significantly lower than the mean, and 50.0% of providers were statistically significantly higher than the mean. For those plans with at least 175 eligible individuals, high- (90th percentile) and low- (10th percentile) performing plans were 12.9% apart, indicating suboptimal performance across all plans and variation between high- and low-performing plans.

Table A3. Summary of Physician Group Level Performance

	n	Mean	Median	Min	Max	STD	IQR	P10	P25	P50	P75	P90
Original Measure	543	72.6%	73.4%	43.6%	88.7%	6.3%	7.6%	64.8%	69.6%	73.4%	77.2%	79.6%
Revised	464	75.9%	76.6%	50.5%	90.5%	5.8%	7.3%	68.2%	72.6%	76.6%	79.9%	82.3%

Measure												

Based on the revised measure at the physician group level, 20.3% of providers were statistically significantly lower than the mean, and 23.9% of providers were statistically significantly higher than the mean, indicating a wide range of scores. For those physician groups with at least 175 eligible individuals, high- (90th percentile) and low- (10th percentile) performing physician groups were 14.1% apart. The results indicate ample room for improvement and meaningful differences in quality of care between the highest and lowest performing physician groups.

Table A4. Summary of ACO Level Performance

	n	Mean	Median	Min	Max	STD	IQR	P10	P25	P50	P75	P90
Original Measure	31	74.6%	74.9%	67.5%	82.5%	3.9%	5.6%	69.0%	71.9%	74.9%	77.5%	79.5%
Revised Measure	31	75.9%	76.5%	69.1%	83.4%	3.9%	5.8%	70.3%	72.6%	76.5%	78.4%	80.8%

Based on the revised measure at the ACO level, 29.0% of providers were statistically significantly lower than the mean, and 38.7% of providers were statistically significantly higher than the mean. Among all 31 ACOs, high- (90th percentile) and low- (10th percentile) performing ACOs were 10.5% apart, indicating suboptimal performance across all ACOs and variation between high- and low-performing ACOs.

Interpretation of the Results

The results indicate that overall performance, calculated using the revised measure, is suboptimal with variation in performance across states, plans, ACOs, and physician groups. Statistically significant differences were identified at the state, plan, ACO, and physician group level when compared to the overall mean.

Appendix B –Reliability and Validity

Table B1. 2011-2012 State Level Measure Rates and Reliability Assessments

State	Original Measure				Revised Measure			
	Num	Denom	Rate	Reliability	Num	Denom	Rate	Reliability
Overall	449,843	620,934	72.5%	--	469,476	623,987	75.2%	--
AZ	19,533	27,773	70.3%	0.994	20,494	27,946	73.3%	0.995
DE	7,706	10,233	75.3%	0.986	8,007	10,286	77.8%	0.988
FL	105,256	144,262	73.0%	0.999	109,918	145,033	75.8%	0.999
IA	30,625	37,915	80.8%	0.997	31,630	38,012	83.2%	0.997
IN	47,862	63,664	75.2%	0.998	49,860	63,946	78.0%	0.998
MO	46,197	60,955	75.8%	0.998	47,976	61,184	78.4%	0.998
MS	32,702	48,289	67.7%	0.996	34,048	48,472	70.2%	0.997
RI	6,146	8,082	76.1%	0.982	6,365	8,107	78.5%	0.985
TX	123,050	179,316	68.6%	0.999	129,167	180,416	71.6%	0.999
WA	30,766	40,445	76.1%	0.996	32,011	40,585	78.9%	0.997

Based on the revised measure, we concluded that the reliability test was adequate, since all state-level reliability scores were greater than 0.7, indicating that the measure would produce reliable scores at the state level.

Table B2. 2011-2012 Plan Level Measure Rates and Reliability Assessments

	Min Denominator	# of Plans	Mean Rate	Reliability Score
Original Measure	150	40	74.2%	0.695
Revised Measure	175	40	76.7%	0.717

Based on the revised measure and using the method of mean denominator and volume categories, a minimum denominator of 175 resulted in an overall reliability score of >0.7, which is within acceptable norms and indicates sufficient signal strength to discriminate performance between plans.

Table B3. 2011-2012 Physician Group Level Measure Rates and Reliability Assessments

	Min Denominator	# of Physician Groups	Mean Rate	Reliability Score
Original Measure	150	543	72.6%	0.697
Revised Measure	175	464	75.9%	0.713

Based on the revised measure and using the method of mean denominator and volume categories, a minimum denominator of 175 resulted in an overall reliability score of >0.7,

which is within acceptable norms and indicates sufficient signal strength to discriminate performance between physician groups.

Table B4. ACO Level Measure Rates and Reliability Assessments

ACO	Original Measure				Revised Measure			
	Num	Denom	Rate	Reliability	Num	Denom	Rate	Reliability
Overall	42,619	57,454	74.2%	--	43,548	57,722	75.4%	--
1	1,327	1,669	79.5%	0.929	1,358	1,675	81.1%	0.932
2	923	1,205	76.6%	0.897	940	1,211	77.6%	0.898
3	1,409	1,854	76.0%	0.929	1,446	1,860	77.7%	0.932
4	760	1,018	74.7%	0.875	777	1,023	76.0%	0.877
5	947	1,276	74.2%	0.897	959	1,279	75.0%	0.897
6	691	892	77.5%	0.868	701	894	78.4%	0.869
7	926	1,199	77.2%	0.898	938	1,206	77.8%	0.898
8	2,013	2,773	72.6%	0.948	2,056	2,778	74.0%	0.948
9	1,984	2,732	72.6%	0.947	2,046	2,753	74.3%	0.949
10	873	1,283	68.0%	0.886	891	1,290	69.1%	0.886
11	1,694	2,244	75.5%	0.940	1,739	2,267	76.7%	0.942
12	528	709	74.5%	0.829	538	709	75.9%	0.831
13	1,465	1,891	77.5%	0.933	1,492	1,894	78.8%	0.935
14	1,035	1,267	81.7%	0.914	1,051	1,272	82.6%	0.916
15	1,470	1,943	75.7%	0.932	1,498	1,952	76.7%	0.933
16	2,284	2,996	76.2%	0.955	2,319	3,000	77.3%	0.956
17	1,677	2,241	74.8%	0.939	1.714	2,248	76.3%	0.940
18	798	1,026	77.8%	0.884	828	1,035	80.0%	0.890
19	659	799	82.5%	0.872	668	801	83.4%	0.874
20	1,112	1,485	74.9%	0.911	1,139	1,488	76.6%	0.913
21	783	982	79.7%	0.885	797	986	80.8%	0.888
22	427	633	67.5%	0.793	448	637	70.3%	0.799
23	2,382	3,148	75.7%	0.957	2,448	3,164	77.4%	0.958
24	2,471	3,436	71.9%	0.957	2,542	3,449	73.7%	0.958
25	1,097	1,589	69.0%	0.907	1,113	1,602	69.5%	0.907
26	750	1,069	70.2%	0.870	777	1,077	72.1%	0.873
27	1,190	1,654	72.0%	0.915	1,207	1,664	72.5%	0.915
28	768	1,129	68.0%	0.872	786	1,136	69.2%	0.873
29	847	1,210	70.0%	0.883	863	1,217	70.9%	0.884
30	1,119	1,425	78.5%	0.916	1,133	1,429	79.3%	0.916
31	6,210	8,677	71.6%	0.982	6,336	8,726	72.6%	0.982

We concluded that the reliability test was adequate, since all ACO-level reliability scores were much greater than 0.7, indicating that the measure would produce reliable scores at the ACO level.

Interpretation of the Results

The results from the reliability assessment indicated that the revised measure was reliable for state and ACO level regardless of the denominator size. For physician groups and plans, the reliable scores (i.e., >0.7) were identified with a minimum denominator sizes of 175.

Convergent Validity

We compared a related NQF-endorsed measure, NQF 0543, which assesses adherence to statin therapy for individuals with coronary artery disease (CAD) at the state, ACO, plan, and physician group levels. We would expect a positive correlation between the two measure scores since both measure medication adherence. We tested the measure distributions for normality at each unit of analysis and then selected the appropriate statistical test for the distribution and assessed the significance of the correlation coefficient.

Table B5. Convergent Validity: Distribution of State Measure Rates

Measure	n	Mean Measure Rate	Standard Deviation	Median	Minimum	Maximum
NQF 2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus	10	76.6%	3.9%	77.9%	70.2%	83.2%
NQF 0543: Adherence to Statin Therapy for Individuals with CAD	10	71.9%	3.7%	72.6%	65.3%	77.8%

The measure rate is positively correlated with NQF 0543 at the state level ($\rho = 0.95$, $p<0.0001$).

Table B6. Convergent Validity: Distribution of Plan Measure Rates

Measure	n	Mean Measure Rate	Standard Deviation	Median	Minimum	Maximum
NQF 2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus	70	75.9%	10.9%	77.1%	40.0%	100%
NQF 0543: Adherence to Statin Therapy for Individuals with CAD	70	71.6%	7.6%	73.0%	50.0%	90.0%

The measure rate is positively correlated with NQF 0543 at the plan level ($\rho = 0.58$, $p<0.0001$).

Table B7. Convergent Validity: Distribution of Physician Group Measure Rates

Measure	n	Mean	Standard	Median	Minimum	Maximum

		Measure Rate	Deviation			
NQF 2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus	6,461	73.4%	17.2%	75.0%	0.0%	100%
NQF 0543: Adherence to Statin Therapy for Individuals with CAD	6,461	67.7%	21.5%	69.4%	0.0%	100%

The measure rate is positively correlated with NQF 0543 at the physician group level (ρ=0.25, p<0.0001).

Table B8. Convergent Validity: Distribution of ACO Measure Rates

Measure	n	Mean Measure Rate	Standard Deviation	Median	Minimum	Maximum
NQF 2468: Adherence to Oral Diabetes Agents for Individuals with Diabetes Mellitus	31	75.9%	3.9%	76.5%	69.1%	83.4%
NQF 0543: Adherence to Statin Therapy for Individuals with CAD	31	70.3%	4.6%	70.8%	59.2%	80.2%

The measure rate is positively correlated with NQF 0543 at the ACO level (ρ= 0.84, p<0.0001).

Interpretation of the Results

The measure was positively correlated with NQF 0543 (Adherence to Statin Therapy for Individuals with CAD) and statistically significant at all reporting levels with the state and ACO levels showing the strongest correlation.