

AHA Measure Testing Report

Measure Name: Lipid Monitoring and Management Measure

The measures tested included:

1. Percentage of patients aged 18 years and older with ASCVD who had a lipid panel completed during the 12-month measurement period. (Monitoring)
2. Percentage of patients aged 18 years and older with ASCVD and who had a lipid panel completed during the 12-month measurement period who achieved an LDL-C of <70 mg/dL on the most recent test during the 12-month measurement period. (Management)

Section 1 Performance Gap

Summarize evidence of performance gap or measurement gap by providing performance scores on the measure as specified at the specified level(s) of analysis. Please include mean, minimum, maximum, and scores by deciles by using the table below or upload an attachment. In the text field here, describe the data source, including number of measured entities, number of patients, and dates of data. If a sample was used, provide characteristics of the entities included.

The eligible population included patients who met the denominator criteria and were seen by providers from the four reporting health systems (Bay Area Community Health (California), M Health Fairview (Minnesota), Franciscan Missionaries of Our Lady Health System (Louisiana), Wellstar Health (Georgia)). The analysis includes data from January 1, 2024 to December 31, 2024. The data criteria (i.e., identification of race from case list and age/sex from patient list) was applied consistently and limited to only qualified NPIs, which resulted in different n for measure denominators.

A total of 286 NPIs and 28690 eligible patient encounters in the monitoring group were analyzed for this measure. The LDL-C monitoring group performance scores ranged from 30.0% to 100.0%, with a mean of 64.9%; decile scores were 1st – 69.9%, 2nd – 76.3%, 3rd – 71.0%, 4th – 66.3%, 5th – 63.9%, 6th – 59.8%, 7th – 59.1%, 8th – 62.4%, 9th – 58.6%, and 10th – 60.9%.

A total of 215 National Provider Identifiers (NPIs) and 11617 eligible patient encounters in the LDL-C management group. The LDL-C management group performance scores ranged from 18.2% to 85.7%, with a mean of 46.8%; decile scores were 1st – 39.1%, 2nd – 41.3%, 3rd – 41.4%, 4th – 45.4%, 5th – 45.3%, 6th – 47.6%, 7th – 45.6%, 8th – 48.7%, 9th – 57.2%, and 10th – 57.3%.

Table 1: Performance Scores by Decile

Monitoring

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Performance Score	64.9	58.0	69.9	76.3	71.0	66.3	63.9	59.8	59.1	62.4	58.6	60.9	60.9
N of Entities	286	5	29	29	29	29	29	29	28	28	28	28	1
N of Persons / Encounters / Episodes	28690	100	643	759	1000	1367	1787	2443	2984	4010	5335	8362	506

Management

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Performance Score	46.8	42.5	39.1	41.3	41.4	45.4	45.3	47.6	45.6	48.7	57.2	57.3	69.4
N of Entities	215	4	22	22	22	22	22	21	21	21	21	21	1
N of Persons / Encounters / Episodes	11617	80	477	555	644	733	834	946	1131	1421	1846	3030	301

Section 2 Summarize Data and Samples

Data Used for Testing

Describe the data used for testing (include dates, sources). The analysis includes data from January 1, 2024 to December 31, 2024 uploaded to GWTG-Outpace Registry. The Outpace registry is a clinical quality improvement tool that pulls in patient-level data from your existing electronic health records and displays your organizations performance against various clinical quality measures tied to mitigating risk factors for cardiovascular disease. This registry experience eliminates the barrier of manual data abstraction and provides real time data results for quality improvement and research strategies.

The two measures were implemented in the Outpace registry in August 2022. Data was received from 11 health systems which represent patients with visits at over 150+ physical locations. As of April 2025, data has been received on over 610,000 patients from all 11 health systems. Four sites were selected from 2024 participating sites. The eligible population included patients met denominator criteria seen by providers from the 4 reporting health systems (Bay Area Community Health (California), M Health Fairview (Minnesota), Franciscan Missionaries of Our Lady Health System (Louisiana), Wellstar Health (Georgia)) between January 1, 2024, and December 31, 2024.

Differences in Data

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), clearly identify which data source/sample is used for each aspect of testing, including the years of data used in each. If there are no differences to report, enter "None."

Patient-level data during the 2024 12-month measurement period were used to report demographics. Provider-level data from the same period were used to assess reliability and conduct validity testing, including comparisons with established provider-level measures to evaluate the validity of the new measure.

Characteristics of Measured Entities

Describe characteristics of measured entities included in the analysis (e.g., number, size, location, type). If you used a sample, describe how you selected measured entities for inclusion in the sample and the representativeness of the sample.

Information from 4 reporting health systems located in California, Minnesota, Louisiana, and Georgia, representing full datasets and geographic diversity across the West, Midwest, South, and Southeast were used for testing. These sites utilize EPIC or Cerner EHR systems. The analysis includes data from January 1, 2024 to December 31, 2024.

Characteristics of Units of the Eligible Population

Describe characteristics of the patients, encounters, episodes, etc., including numbers and percentages by factors such as age, sex, race, or diagnosis. Provide descriptive statistics separately by each specified level of analysis and data source. If you used a sample, describe how you selected the patients for inclusion in the sample and the representativeness of the sample.

The eligible population included patients met denominator criteria seen by providers from the 4 reporting health systems (Bay Area Community Health, M Health Fairview, Franciscan Missionaries of Our Lady Health System, Wellstar Health) between January 1, 2024, and December 31, 2024. Across 15348 patients in the LDL-C monitoring group, the population was 5298 (34.5%) female and 10050 (65.5%) male. Age distribution ranged from a minimum of 18 to a maximum of 99, with a median age of 69, 124 (0.8%) aged 18–39, 4841 (31.5%) aged 40–64, and 10383 (67.7%) aged 65 and older. Race data showed 12330 (80.3%) White, 1665 (10.8%) Black or African American, 334 (2.2%) Asian, 38 (0.2%) American Indian or Alaska Native, 15 (0.1%) Native Hawaiian or Other Pacific Islander, 41 (0.3%) Mixed-race and 925 (6.0%) Unknown. This data set captures all patients seen in an ambulatory setting and common diagnoses include hypertension, cardiovascular disease and accompanying comorbid diagnoses.

Across 8370 patients in the LDL-C management group, the population was 2837 (33.9%) female and 5533 (66.1%) male. Age distribution ranged from a minimum of 18 to a maximum of 98, with a median age of 69, 59 (0.7%) aged 18–39, 2601 (31.1%) aged 40–64, and 5710 (68.2%) aged 65 and older. Race data showed 6810 (81.4%) White, 816 (9.7%) Black or African American, 191 (2.3%) Asian, 22 (0.3%) American Indian or Alaska Native, 8 (0.1%) Native Hawaiian or Other Pacific Islander, 14 (0.2%) Mixed-race and 509 (6.1%) Unknown. This data set captures all patients seen in an ambulatory setting and common diagnoses include hypertension, cardiovascular disease and accompanying comorbid diagnoses.

Patients with missing provider information, those who did not meet the denominator criteria, and encounters at small providers with fewer than 20 eligible patients were excluded. Descriptive statistics are reported at the patient level.

Section 3 Reliability

Level(s) of Reliability Testing Conducted

Choose all that apply.

- ☐ Person or encounter level (i.e., data element) (e.g., inter–abstractor reliability)
- ☒ Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis, Beta-Binomial, Mixed Logistic Regression, or random split-half correlation)

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

Method(s) of Reliability Testing

For each level of reliability testing conducted, describe the method(s) of reliability testing and explain what each tests. Describe the steps and the type of error each method is designed to detect. Provide the type of statistical analysis used. Describe proportion of missing data, how missing data were analyzed and/or excluded, and any sensitivity analysis conducted.

Note: Testing at the person or encounter level requires that all critical data elements be tested (not just agreement of one final overall computation for all persons/encounters/episodes). At a minimum, the numerator, denominator, and exclusions must be assessed and reported separately. Prior evidence of reliability of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements.

Reliability testing at the NPI level was conducted using a signal-to-noise ratio (SNR) approach, where performance scores represented the proportion of eligible patients meeting the measure. The between-entity variance was calculated as the variance of performance scores across NPIs. Within-entity variance was estimated using the binomial approximation formula $\frac{p(1-p)}{n}$, where p is the provider-level performance score and n is the measure denominator (number of eligible patients). Reliability was then calculated for each NPI as $SNR = \frac{\text{between-provider variance}}{(\text{between-provider variance} + \text{within-provider variance})}$. This method quantifies the extent to which observed differences reflect true performance variation rather than measurement error. Providers with missing provider information, and small providers with fewer than 20 eligible patients were excluded. This approach ensures robust assessment of provider-level score stability using available aggregated data.

Reliability Testing Results

Provide the statistical results from reliability testing for each level and type of reliability testing conducted. Where applicable, include results from accountable entity level reliability testing (e.g., signal-to-noise testing) in Table 2.

At the NPI level, SNR reliability testing was conducted using performance scores and measure denominators for each provider.

In the LDL-C monitoring group, the estimated between-NPI variance in performance scores was 0.017. The resulting reliability scores across NPIs ranged from 0.581 to 1.000, with a mean reliability of 0.833. A total of 286 NPIs were evaluated. 249 (87.1%) NPIs had reliability scores above the commonly accepted threshold of 0.7, indicating that observed performance differences are likely to reflect true variation rather than random noise.

In the LDL-C management group, the between-NPI variance in performance scores was 0.017. The resulting reliability scores across NPIs ranged from 0.585 to 0.961, with a mean reliability of 0.759. A total of 215 NPIs were evaluated. 155 (72.1%) NPIs had reliability scores above the commonly accepted threshold of 0.7, indicating that observed performance differences are likely to reflect true variation rather than random noise.

The LDL-C monitoring group, the mean reliability across all NPIs was 0.833, with scores ranging 0.581 to 1.000. Entities in higher deciles consistently showed higher reliability scores, indicating greater stability and confidence in performance measurement. 87.1% of NPIs had reliability above the 0.7 threshold, which supports the inference that the measure can reliably distinguish

performance differences across providers. These findings provide evidence that the measure is sufficiently robust for use in quality improvement and accountability applications.

The LDL-C management group, the mean reliability across all NPIs was 0.759, with scores ranging 0.585 to 0.961. Entities in higher deciles consistently showed higher reliability scores, indicating greater stability and confidence in performance measurement. 72.1% of NPIs had reliability above the 0.7 threshold, which supports the inference that the measure can reliably distinguish performance differences across providers. These findings provide evidence that the measure is sufficiently robust for use in quality improvement and accountability applications

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Enter overall mean, minimum, and maximum performance scores, along with the count of measured entities and persons/encounters/episodes. Organize entities into deciles by the entity number of persons/encounters/episodes (denominator) from 1 (smallest N) to 10 (largest N). Provide mean reliability, performance score, number of entities (total) and number of persons/encounters/episodes (total) for entities assigned to each decile. For minimum reliability, provide reliability value for the entity with the smallest N. For maximum reliability, provide the reliability value for the entity with the largest N.

Monitoring

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.833	0.634	0.688	0.747	0.753	0.787	0.827	0.857	0.884	0.912	0.931	0.954	0.973
Mean Performance Score	64.9	58.0	69.9	76.3	71.0	66.3	63.9	59.8	59.1	62.4	58.6	60.9	60.9
N of Entities	286	5	29	29	29	29	29	29	28	28	28	28	1
N of Persons / Encounters / Episodes	28690	100	643	759	1000	1367	1787	2443	2984	4010	5335	8362	506

Management

Reliability	0.759	0.588	0.624	0.657	0.688	0.716	0.741	0.773	0.802	0.832	0.869	0.908	0.961
Mean Performance Score	46.8	42.5	39.1	41.3	41.4	45.4	45.3	47.6	45.6	48.7	57.2	57.3	69.4
N of Entities	215	4	22	22	22	22	22	21	21	21	21	21	1
N of Persons / Encounters / Episodes	11617	80	477	555	644	733	834	946	1131	1421	1846	3030	301

Interpretation of Reliability Results*

Provide your interpretation of the results in terms of demonstrating reliability for each level and type of reliability testing conducted. Describe how the results support an inference of reliability for the measure or instrument.

Reliability estimates based on SNR demonstrate strong measurement reliability at the NPI level. In monitoring group, the mean reliability across all NPIs was 0.833, with scores ranging 0.581 to 1.000. Entities in higher deciles consistently showed higher reliability scores, indicating greater

stability and confidence in performance measurement. 87.1% of NPIs had reliability above the 0.7 threshold, which supports the inference that the measure can reliably distinguish performance differences across providers. These findings provide evidence that the measure is sufficiently robust for use in quality improvement and accountability applications.

In management group, the mean reliability across all NPIs was 0.759, with scores ranging 0.585 to 0.961. Entities in higher deciles consistently showed higher reliability scores, indicating greater stability and confidence in performance measurement. 72.1% of NPIs had reliability above the 0.7 threshold, which supports the inference that the measure can reliably distinguish performance differences across providers. These findings provide evidence that the measure is sufficiently robust for use in quality improvement and accountability applications.

Section 4 Validity

Level(s) of Validity Testing Conducted

Choose all that apply.

- ☐ Person or encounter level (i.e., data element) (e.g., sensitivity and specificity)
- ☒ Accountable entity level (i.e., measure score) (e.g., criterion validity)

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

Type of Accountable Entity Level Validity Testing Conducted*

Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use (i.e., the score is an accurate reflection of the effect of performance on quality or resource use and can distinguish good from poor performance)

As part of the ACC/AHA joint measure development process, a Technical Expert Panel (TEP) is formed that includes representatives of the clinical team, clinicians from diverse geographical regions and practice settings, clinician(s) with expertise in quality improvement, and/or clinician(s) who have data extraction or registry expertise. During the measure development process, the ACC/AHA conducted informal face validity assessments of the measure with our TEP (e.g., a yes/no vote conducted either via email survey and/or during teleconference meeting(s) to confirm the measure should move to the next level of testing). In 2011, a joint ACCF/AHA/AMA-PCPI TEP identified the need for measurement and approved the LDL-C monitoring measure. [Drozda] In 2025, Williams, et al. reaffirmed the need for measurement in the chronic coronary disease (CCD) population and included a LDL-C management measure. [Williams]

Two measures addressing lipid monitoring and lipid management were implemented in the AHA's OUTPACE registry in August 2022. AHA has received data from 11 practices providing over 610,000 unique patients for measure calculation. In 2024, given a continued gap in performance, the AHA team evaluated combining the concepts into one measure for CMS submission. The combined measure draft was reviewed with a focus group of 23 patients and hosted a public comment period to gather additional clinician and specialty society input. Seventeen patient respondents indicated the measure is meaningful with 15 responding yes to the question "Will a measure confirming your doctor completed a lipid panel once during the year be meaningful to you?" and 16 responding yes

to the question “Will a measure confirming LDL-C control of less than 70 in the year be meaningful to you?”. There was an insufficient number of respondents to face validity survey questions (n =3 with 100% confirming measures what is intended) with respondents opting to provide narrative responses on the draft measure. Comments received supported the measure concept was meaningful and addresses an important clinical gap. Given the support for continued measure development the AHA then focused efforts and resources on full empirical testing of the measure. Empirical testing results are included below.

Empirical validity to be completed by testing team; face validity completed by measure development team.

- ☒ Empirical validity testing at the accountable entity level (e.g., criterion validity, construct validity, known groups analysis)

Note: Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers). If more than one level of analysis is specified, testing must be conducted for each level separately.

Method(s) of Validity Testing*

For each level of testing conducted, describe the method(s) of validity testing and what each tests. Describe the steps and explain what was tested (e.g., accuracy of data elements compared with authoritative source, relationship to another measure as expected). Describe the statistical analysis used. Describe proportion of missing data, how missing data were analyzed and/or excluded, and any sensitivity analysis conducted.

For empirical accountable entity level testing, the following should be included:

- Narrative describing the hypothesized relationships
- Narrative describing why examining these relationships (e.g., correlating measures) would validate the measure
- Expected direction of the association
- Expected strength of the association

Validity testing was conducted at the NPI level using empirical associations between performance scores and related constructs. We hypothesized that higher performance scores would correlate with stronger reliability and larger patient volumes, reflecting more consistent, high-quality care. This hypothesis was tested to establish construct validity, ensuring the new measure captures the intended concept of clinician performance. Additionally, we hypothesized that the performance scores would show a moderate to strong correlation with other established measures of care quality, demonstrating convergent validity. The results confirmed that the new measure aligns with both the theoretical framework of care quality and other validated performance indicators. Providers with missing provider information, and small providers with fewer than 20 eligible patients were excluded.

Validity Testing Results

Provide the statistical results from validity testing for each level and type of validity testing conducted.

In monitoring group:

- Performance Score vs. Reliability: N = 286, r = 0.052
- Performance Score vs. Patient Volume: N = 286, r = -0.269

- LDL-C vs. Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan: N = 243, $r = 0.577$

In management group:

- Performance Score vs. Reliability: $r = 0.337$, $p < 0.001$, 95% confidence interval (CI): 0.213 to 0.450
- Performance Score vs. Patient Volume: N = 215, $r = 0.406$, $p < 0.001$, 95% CI: 0.288 to 0.512
- LDL-C vs. Hemoglobin A1c (HbA1c) Poor Control (>9%): N = 155, $r = 0.568$

Interpretation of Validity Results*

Provide your interpretation of the results in terms of demonstrating validity for each level and type of validity testing conducted. Describe how the results support an inference of validity for the measure and include the precision of the statistical result(s) (e.g., 95% confidence interval) and/or an assessment of statistical significance (e.g., p-value). If the results are not what were expected, explain why.

We conducted clinician score-level validity testing among a sample of Outpace practices by analyzing the correlation of performance scores between the new monitoring measure and QPP128: Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan. We chose to use QPP128 because it is a validated measure that used by CMS in the Quality Payment Program. Furthermore, it aligns with the new monitoring measure in that it is an effective clinical care focused on ensuring an annual evaluation of a clinical process.

Fully validated data were available on performance scores for both measures for 243 clinicians participating in the registry during the 2024 12-month measurement period.

When correlating each clinician's measure performance, we found a correlation coefficient (r) of 0.577. The results indicate a moderate positive correlation between the two measures demonstrating the empiric validity of the new monitoring measure as a meaningful assessment of quality of care.

We conducted clinician score-level validity testing among a sample of Outpace practices by analyzing the correlation of performance scores between the new management measure and QPP122: Hemoglobin A1c (HbA1c) Poor Control (>9%). We chose to use QPP122 because it is a validated measure that is CBE endorsed and used by CMS in the Quality Payment Program. Furthermore, it aligns with the new management measure in that it is an effective clinical care and intermediate outcome measure focused on ensuring a patient's biomarker falls within an acceptable range.

Fully validated data were available on performance scores for both measures for 155 clinicians participating in the registry during the 2024 12-month measurement period.

When correlating each clinician's measure performance, we found a correlation coefficient (r) of 0.568. The results indicate a moderate positive correlation between the two measures demonstrating the empiric validity of the new management as a meaningful assessment of quality of care.

Section 5 Risk Adjustment

Methods Used to Address Risk Factors*

Describe the methods or approaches that were used to explore the effects of risk factors on this measure. Choose all that apply.

- ☒ No risk adjustment or stratification.
- ☐ Statistical risk adjustment model with risk factors
- ☐ Statistical case-mix adjustment model
- ☐ Stratification by risk factor category
- ☐ Other (specify)

Section 6 Equity

Contributions Toward Advancing Health Equity*

Describe how this measure contributes to efforts to advance health equity. Address each of the following elements in your response:

- **Methodology:** Provide a description of your methodology and approach to empirical testing of differences in performance scores across multiple socio-contextual variables (e.g., race, ethnicity, urbanicity/rurality, socioeconomic status, gender, gender identity, sexual orientation, age).
- **Results/Interpretation:** Provide an interpretation of the results, including interpretation of any identified differences and consideration of negative impact or unintended consequences on subgroups.
- **Anticipated Impact:** Describe how accountable entities can use these results to reduce identified disparities.
- **Evidence of Known Disparities:** Evaluate known and existing health disparities specific to the measure focus area and describe how this measure contributes to efforts to advance health equity. This information should be based on relevant literature, internal empirical analyses, and/or qualitative data, including input from technical expert panels.

Please see above disparity gaps identified during testing process. This new measure provides an opportunity to address disparities in care as it has been identified that women have a lower likelihood of attaining LDL-C goals than men. [Gavina]. Safford, et al. found differences in treatment and outcomes based on race and gender. [Safford] Further, differences in lipid control have been found based on insurance status. [Egan]

Egan BM, Li J, Sarasua SM, et al. Cholesterol Control Among Uninsured Adults Did Not Improve from 2001-2004 to 2009-2012 With Both Publicly and Privately Insured Adults Doubled. *Journal of the American Heart Association*. 2017;6(11):e006105.

Gavina C, Araújo F, Teixeira C, et al. Sex differences in LDL-C control in a primary care population: The PORTRAIT-DYS study. *Atherosclerosis*. 2023 Nov;384:117148.

Safford MM, Gamboa CM, Durant RW, et al. Race-Sex Differences in the Management of Hyperlipidemia: The Reasons for Geographic and Racial Differences in Stroke Study. *American Journal of Preventive Medicine*. 2015;48(5):520-7.