Reliability, Encounter Level Testing and Measure Performance Score Results for the Addressing Social Needs (ASN) electronic Clinical Quality Measure (eCQM) Hospital Inpatient Quality Reporting Program, Medicare Promoting Interoperability Program, Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program (MUC2024-069)

Please note due to limitations within the electronic data ecosystem, reliability, encounter level testing, and measure performance score results were not available by the May 2024 Measures Under Consideration (MUC) submission deadline. As noted within the MERIT submission, testing results would be available by August 2024. Therefore, results are included within this attachment for the 2024 MERIT Data Template in relevant sections below. We request reviewers reference this attachment for final testing results.

Data Source

Two datasets were used for testing:

- **Dataset A** used calendar year 2023 with 12 hospitals with 137,721 inpatient encounters for 153,878 patients.
- **Dataset B** used calendar year 2023 with 3 hospitals with 275,443 inpatient encounters for 158,498 patients.

As explained in [Patient/Encounter Level (Data Element Level) Testing section](#), coded screening data were infrequent at these test sites and coded intervention data were not being captured. Prioritization of social needs data elements varied across datasets. Results show the housing domain was most frequently screened of the 4 domains.

Reliability: Measure Score Level (Accountable Entity Level) Testing

Reliability testing was completed for this measure using a Signal-to-Noise calculation at the facility-level.

15 accountable entities (facilities) included in the results. Median in Table 1.

Table. 1. Measure Score Level Reliability Analysis: Signal-to-Noise

Dataset	Reliability	Housing	Food	Transportation	Utilities
Database A **12 hospitals**	**Mean (SD)**	0.811 (0.241)	0.852 (0.216)	N/A	N/A
	Median	0.926	1.000	N/A	N/A
	Min-Max	(0.306-1.000)	(0.375-1.000)	N/A	N/A
Dataset B **3 hospitals**	**Mean (SD)**	0.9998 (0.0001)	N/A	N/A	N/A
	Median	0.9998	N/A	N/A	N/A
	Min-Max	(0.9997-0.9999)	N/A	N/A	N/A

Interpretation of Measure Score Reliability Signal-to-Noise results:

Overall results show high reliability for all measured entities. There are no results on transportation insecurity or utility insecurity domains due to low rates of screening and intervention follow-up captured in structured fields.

Patient/Encounter Level (Data Element Level) Testing

Encounter-level testing of the individual data elements in the final performance measure (i.e., measure of agreement such as kappa or correlation coefficient) were not completed.

There was limited social needs screening captured in structured fields in 2023, shown in Table 2. Our testing partners noted that their hospitals began systematically screening, collecting, and documenting social needs data on more patients beginning in 2024. There were also almost no follow-up interventions being documented in structured and encoded fields in the 2023 testing data.

Table 2. Percentage of encounters with indication of a social needs assessed by domain

Dataset	% of encounters screened or had z-code indicating social need			
	Housing	Food	Transportation	Utilities
Dataset A	1.6%	0.1%	0.5%	0.0%
Dataset B	50.9%	N/A	N/A	N/A

Data element validity requires high rates of data capture and low rates of missing data. Although screening patients for social needs is standard practice according to data from the 2022 American Hospital Association (AHA) Information Technology (IT) survey, our tested entities are not capturing screening data in encoded terminology as required by this measure or even in structure fields (Richwine C, Meklir S. Hospitals' collection and use of data to address social needs and social determinants of health. *Health Serv Res*. 2024; 1-7. doi:10.1111/1475-6773.14341). As an eCQM this measure requires that data are captured in structured fields and using USCDI-aligned standard terminology. To support testing in the absence of aligned terminology with testing partners, we relied on exact match data approximation from structured fields.

Measure Performance

ASN eCQM enables a summary score of quality for each domain. 15 accountable entities (facilities) were included in the analysis of the distribution of performance scores. Dataset A had 12 hospitals with 137,721 inpatient encounters for 153,878 patients. Dataset B had 3 hospitals with 275,443 inpatient encounters for 158,498 patients.

Measure performance scores are the domain-specific summary scores, calculated as the (numbers of encounters screened negative + number of encounters positive with a follow-up) divided by all encounters. Due to the lack of standardized intervention data, summary scores are skewed towards those screened negative divided by all encounters.

Mean, standard deviation, median, minimum, and maximum are included in Table 3 below. 10th and 90th percentiles are not included. Due to the low number of accountable entities, this would not provide more detailed information.

We were unable to test the impact of social risk factors measure scores.

Table 3. Measure Performance: Proportion

Dataset	Summary scores	Housing	Food	Transportation	Utilities
Dataset A 12 Hospitals	**Mean (SD*)**	0.0002 (0.0002)	0.0001 (0.0002)	0.0033 (0.0055)	0.0000 (0.0000)
	Median	0.0001	0	0.0002	0
	Min-Max	(0.0000-0.0008)	(0.0000-0.0007)	(0.0000-0.0135)	(0.0000-0.0001)
Dataset B 3 Hospitals	**Mean (SD*)**	0.439 (0.165)	N/A	N/A	N/A
	Median	0.506	N/A	N/A	N/A
	Min-Max	(0.251 - 0.560)	N/A	N/A	N/A

Interpretation of Measure Performance results:

Meaningful measure performance data requires robust screening and intervention data. The Dataset B summary score for housing domain is meaningful. Other results are based on minimal data; thus, it is difficult to derive meaningful comparisons between measure scores (topics in data availability are detailed above). Even with limited data, all results are presented and demonstrate outcomes for hospitals or providers that do not have consistent processes for documenting social needs screening and interventions. In comparing Dataset B to Dataset A, there are a wide range of measure scores, which indicate differences between hospitals and therefore room for quality improvement. Social needs screening and intervention may be occurring, but the data documentation are incomplete. These findings support the intent of the measure to drive toward interoperable and discrete terminology to support health equity. Data capture uptake continues to expand at measured entities and can be influenced by CMS and other stakeholders.

Measure Performance, Narrative Measure Rates

This measure output uses narrative measure scores for stakeholders to better understand where improvement opportunities lie. Two outputs are shown below, Table 4 shows for Dataset A and Table 5 for Dataset B. Each row is mutually exclusive; each column is a domain, and the columns will total to the number of encounters in each dataset.

Table 4. Narrative Measure Rates, Dataset A (137,721 encounters)

Narrative Food, Transportation, Utilities	Food	Transportation	Utilities	Narrative Housing	Housing
Encounters "No Documented Assessment for Disparity Related Need"	137,622 (99.9%)	137,070 (99.5%)	137,718 (100%)	Encounters "No documented assessment for BOTH housing instability AND homelessness Disparity Related Needs"	135,540 (98.4)
Encounters "Declined Assessment for a Disparity Documented"	0 (0%)	0 (0%)	0 (0%)	Encounters "Declined Assessment for a Disparity Documented"	0 (0%)
Encounters "Assessed, Disparity Related Need identified, but No Intervention Documented"	12 (0.01%)	33 (0.02%)	2 (0%)	Encounters "Assessed, Disparity Related Need identified of EITHER housing instability or homelessness, but No intervention Documented"	2,157 (1.6%)
Encounters "Assessed and Disparity Related Need Identified, and Intervention Documented"	0 (0%)	0 (0%)	0 (0%)	Encounters "Assessed and Disparity Related Need Identified of EITHER housing instability or homelessness and Intervention Documented"	0 (0%)
Encounters "Documented assessment for Disparity Related Need, but none Identified"	23 (0.02%)	618 (0.5%)	1 (0%)	Encounters "Documented assessment for Disparity Related Need, but none Identified for BOTH housing instability and homelessness"	24 (0.01%)

Table 5. Narrative Measure Rates, Dataset B (275,443 encounters)

Narrative Food, Transportation, Utilities	Food	Transportation	Utilities	Narrative Housing	Housing -
Encounters "No Documented Assessment for Disparity Related Need"	N/A	N/A	N/A	Encounters "No documented assessment for BOTH housing instability AND homelessness Disparity Related Needs"	135,243 (49.1%)
Encounters "Declined Assessment for a Disparity Documented"	N/A	N/A	N/A	Encounters "Declined Assessment for a Disparity Documented"	4 (0%)

Narrative Food, Transportation, Utilities	Food	Transportation	Utilities	Narrative Housing	Housing -
Encounters "Assessed, Disparity Related Need identified, but No Intervention Documented"	N/A	N/A	N/A	**Encounters "Assessed, Disparity Related Need identified of EITHER housing instability or homelessness, but No intervention Documented"**	3,381 (1.2%)
Encounters "Assessed and Disparity Related Need Identified, and Intervention Documented"	N/A	N/A	N/A	**Encounters "Assessed and Disparity Related Need Identified of EITHER housing instability or homelessness and Intervention Documented"**	0 (0%)
Encounters "Documented assessment for Disparity Related Need, but none Identified"	N/A	N/A	N/A	**Encounters "Documented assessment for Disparity Related Need, but none Identified for BOTH housing instability and homelessness"**	136,815 (49.7%)

<u>Interpretation Measure Performance, Narrative Measure Rates:</u>

Likely because of the absence of data to support screening and intervention activities, the highest percent is in "No documented assessment for disparity related need." Due to no intervention data stored in structured and encoded fields for either dataset, Encounters Assessed and Disparity Related Need Identified, and Intervention Documented are zero. Drivers, including this measure, are moving the ecosystem toward standardized processes supported by interoperable terminology. As these drivers support best practice, the percent of encounters with measurable findings for assessment and intervention on identified needs will increase, aligned with the mission of the measure.