1. Reliability

Reliability testing was conducted at the **accountable entity level** using signal-to-noise and test-retest reliability methods.

1. Methods
2. Signal-to-noise reliability

Using signal-to-noise reliability, we tested the extent to which a facility’s quality of malnutrition care can be distinguished from that of other facilities using the GMCS. In other words, signal-to-noise reliability tests the precision of measure scores. To compute signal-to-noise reliability, we estimated the proportion of observed variability in the GMCS that is due to differences between facilities in the completeness of malnutrition care (signal variance), as opposed to variability due to differences in care within facilities (noise variance).

For each facility, we computed the facility-specific noise variance as the sample variance of the patient scores on Measure Observation 6 divided by the number of encounters in the facility minus 1. For the signal variance, we used the iterative empirical Bayes method (Morris, 1983) to estimate a single value for all facilities. We then calculated a reliability coefficient for each facility as the ratio of signal variance to the sum of the signal variance and noise variance for that facility. A reliability of 1 indicates perfect reliability, where all variation in the GMCS measure scores reflects between-facility differences rather than within-facility differences.

1. Test-retest reliability

Using test-retest reliability, we assessed the stability of the GMCS across random samples of encounters. In other words, we tested the extent to which the GMCS measure scores are affected by sampling variability in the encounters used to compute the scores.

To do so, we drew 1,000 bootstrap samples (i.e., sampling with replacement) of encounters stratified by facility, where we kept the original number of encounters within each facility. The randomly sampled sets of encounters from a given facility are assumed to reflect an independent set of re-measurement of the GMCS scores for each facility. Adequate reliability is assumed if the GMCS measure scores calculated from the random datasets for the same IPF are similar.

Within each bootstrap sample, we computed the GMCS measure score for each facility and grouped the 1,000 samples into 500 pairs. We then calculated Spearman’s correlation (rho) and the intraclass correlation coefficient (ICC) between those measure scores in each pair of samples to assess how stable the facilities’ measure scores remains as they get computed on a different, randomly sampled set of encounters. Spearman’s correlation quantifies the strength of the rank-order association between the measure scores in each pair, where a value of 1 indicates perfect positive association. The ICC quantifies the strength of absolute agreement between the facility scores in each pair, where a value of 1 indicates perfect reliability. Following the calculations, we examined the distribution of the resulting 500 Spearman’s correlations and ICCs. Adequate reliability is assumed if the GMCS measure rates calculated from the random datasets for the same facility are similar. Note that unlike signal-to-noise, test-retest reliability does not provide a separate reliability coefficient per facility.

1. Results
2. Signal-to-noise reliability

Table IX.1 presents the summary statistics of the signal-to-noise reliability coefficients estimated for each facility. The summary statistics reveal high reliability across the board with coefficients averaging at 0.96 and ranging from 0.69 to 1.00. The median signal-to-noise reliability for the GMCS measure is sufficiently high (above the CBE threshold of 0.6) for all facilities (0.99). Using the CBE threshold of 0.60, all facilities in our sample would be considered to have a sufficiently high reliability.

**Table IX.1.** Distribution of signal-to-noise reliability scores of the GMCS measure across all facilities

| Measure | N (# facilities) | Mean | SD | Min | 25th pctl. | Median | 75th pctl. | Max |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GMCS | 28 | 0.96 | 0.07 | 0.69 | 0.97 | 0.99 | 0.99 | 1.00 |

Source: Mathematica analysis of the test-site data provided by Avalere Health and the Academy of Nutrition and Dietetics (N=28 facilities)

**Table IX.2.** Deciles of signal-to-noise reliability scores of the GMCS measure across all facilities

|   | Reliability | Number of facilities | Number of encounters |
| --- | --- | --- | --- |
| Overall  | -- | 28 | 145,846 |
| Minimum | 0.69 | 1 | 203 |
| Decile 1 | 0.87 | 3 | 733 |
| Decile 2 | 0.95 | 3 | 2,180 |
| Decile 3 | 0.97 | 3 | 5,058 |
| Decile 4 | 0.98 | 2 | 6,463 |
| Decile 5 | 0.99 | 3 | 9,081 |
| Decile 6 | 0.99 | 3 | 16,268 |
| Decile 7 | 0.99 | 2 | 12,435 |
| Decile 8 | 1.00 | 3 | 19,840 |
| Decile 9 | 1.00 | 3 | 26,902 |
| Decile 10 | 1.00 | 3 | 46,886 |
| Maximum | 1.00 | 1 | 12,176 |

Source: Mathematica analysis of the test-site data provided by Avalere Health and the Academy of Nutrition and Dietetics (N=28 facilities)

**Table IX.3.** Mean signal-to-noise reliability scores by deciles of GMCS measure scores

|   | Mean reliability | Mean GMCS score | Number of facilities | Number of encounters |
| --- | --- | --- | --- | --- |
| Overall  | 0.96 | 90.3 | 28 | 145,846 |
| Minimum GMCS score | 0.78 | 83.7 | 1 | 310 |
| Decile 1 of GMCS scores | 0.82 | 84.0 | 3 | 4,945 |
| Decile 2 of GMCS scores | 0.93 | 86.1 | 3 | 8,644 |
| Decile 3 of GMCS scores | 0.95 | 87.9 | 3 | 10,299 |
| Decile 4 of GMCS scores | 0.97 | 88.8 | 2 | 5,724 |
| Decile 5 of GMCS scores | 0.98 | 89.5 | 3 | 11,824 |
| Decile 6 of GMCS scores | 0.99 | 90.5 | 3 | 33,921 |
| Decile 7 of GMCS scores | 1.00 | 91.8 | 2 | 19,748 |
| Decile 8 of GMCS scores | 0.99 | 92.6 | 3 | 13,118 |
| Decile 9 of GMCS scores | 0.97 | 94.2 | 3 | 9,877 |
| Decile 10 of GMCS scores | 1.00 | 98.0 | 3 | 27,746 |
| Maximum GMCS score | 1.00 | 98.2 | 1 | 6,304 |

Source: Mathematica analysis of the test-site data provided by Avalere Health and the Academy of Nutrition and Dietetics (N=28 facilities)

1. Test-retest reliability

Table IX.4 shows the summary statistics of the test-retest reliability coefficients across the 500 bootstrap sample pairs. Spearman’s correlation averaged at 0.97 and ranged from 0.91 to 0.99, and the ICC averaged at 0.96 and ranged from 0.83 to 0.99. Given that a Spearman’s correlation of above 0.90 is often considered very strong, and an ICC of above 0.75 is considered good reliability, these results provide evidence of high reliability for the GMCS.

**Table IX.4.** Distribution of test-retest reliability scores of the GMCS measure across bootstrap samples

| Measure | N (# bootstrap pairs) | Correlation method | Mean | SD | Min | 25th pctl. | Median | 75th pctl. | Max |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GMCS | 500 | Spearman’s rho | 0.97 | 0.02 | 0.91 | 0.96 | 0.97 | 0.98 | 0.99 |
| ICC\* | 0.96 | 0.02 | 0.83 | 0.95 | 0.96 | 0.98 | 0.99 |

Source: Mathematica analysis of the test-site data provided by Avalere Health and the Academy of Nutrition and Dietetics (N=28 facilities)

\*Using ICC(2,1), or the two-way random effects, absolute agreement, single rater/measurement ICC.

1. Interpretation

The results of the reliability analysis indicate that the GMCS is a highly reliable measure, both with respect to signal-to-noise and test-retest reliability. The high reliability coefficients obtained in the signal-to-noise analysis suggest that the GMCS can be reliably used to distinguish between facilities in terms of their completeness of malnutrition care. The high reliability coefficients obtained in the test-retest analysis support that the GMCS is highly stable in the face of random sampling variability in the encounters used to compute the measure scores.