NATIONAL QUALITY FORUM

Reliability

NQF's Current Definition of Reliability and Related Concepts

Reliability refers to the repeatability or precision of measurement. **Reliability of data elements** refers to repeatability and reproducibility of the data elements for the same population in the same time period. **Reliability of the measure score** refers to the proportion of variation in the performance scores due to systematic differences across the measured entities (or signal) in relation to random error (or noise).

Reliability testing – Empirical analysis of the measure as specified that demonstrates repeatability and reproducibility of the data elements in the same population in the same time period and/or the precision of the computed measure scores. Reliability testing focuses on random error in measurement and generally involves testing the agreement between repeated measurements of data elements (often referred to as inter-rater or inter-observer, which also applies to abstractors and coders) or the amount of error associated with the computed measure scores (signal vs. noise).

Reliability, threats – Some aspects of the measure specifications or the specific topic of measurement can affect reliability. Ambiguous measure specifications can result in unreliable measures. Small case volume or sample size, or rare events can affect the precision (reliability) of the measure score.

Brief Summary of Panel Discussion to Date

- Reliability is not an attribute of a healthcare performance measure
- Measure results should be stable (assuming same subjects and timeframe, with no change in underlying processes). This is a facet of reliability not included in NQF's current definition
- General agreement on the signal-to-noise approach as a way to quantify how well a measure can distinguish differences in performance across providers
- Recognition that differences in performance may not be due only to differences in quality of care (this crosses over into a discussion of validity)
- Small sample sizes and rare events can affect reliability
- One reliability estimate may not be adequate: consider asking for mean and variance (or other percentile values), stratification by sample size, standard error of measurement, or other information that demonstrates stability
- Glossary and thresholds are needed