Measure Testing

Measure Testing enables a measure developer to assess the suitability of the quality measure's technical specifications and acquire empirical evidence to help assess the strengths and weaknesses of a measure. Testing provides an opportunity to:

- Refine the draft specifications before they are finalized
- Augment or reevaluate earlier judgments about the measure's importance
- Assess the feasibility, usability, and scientific acceptability of the measure.

Measure Testing can be conducted for a single measure or a set of measures. Information gathered through Measure Testing is a part of full measure development and this information can be used in conjunction with expert judgment to evaluate a measure. Properly conducting Measure Testing and analysis is critical to approval of a measure by CMS and endorsement by the National Quality Forum (NQF).

Measure Testing is an iterative process that includes:

- Developing and implementing a measure testing plan
- Analyzing the test results, and
- Iteratively refining and retesting the measure until it meets testing criteria.

Initial testing during development (sometimes referred to as pilot testing) is generally conducted within the framework of alpha and beta tests. Though both alpha and beta testing are considered part of measure testing, alpha testing may occur as early as information gathering and is repeated iteratively during the development of measure specifications.

Alpha tests (also called formative tests) are of limited scope since they usually occur before detailed specifications are fully developed. The alpha tests include methods to determine if individual data elements are available and if the form in which they exist is consistent with the intent of the measure. The types of testing done in an alpha test vary widely and often depend on the measure's data source or uniqueness of the measure specifications. For example, an alpha test may include a query to a large integrated delivery system database to determine how specific data are captured, where they originate, and how they are currently expressed. The results can impact decisions about what is included in a measure.

Beta testing (also called field testing) generally occurs after the initial technical specifications have been developed and is usually larger in scope than alpha testing. In addition to gathering further information about feasibility, beta tests serve as the primary means to assess scientific acceptability and usability of a measure. They can also be used to evaluate the measure's suitability for risk adjustment or stratification and help expand previous importance and feasibility evaluations. When carefully planned and executed, beta testing helps document measure properties with respect to the evaluation criteria.

Sampling of data from a larger dataset is often used to conduct testing on a smaller, representative dataset. In general, samples used for testing should:

- Represent the full variety of entities whose performance will be measured (e.g., large and small hospitals). This is especially critical if the measured entities volunteer to participate, which limits generalizability to the full population.

- Include adequate numbers of observations to support reliability and validity analyses using the planned statistical methods.
- Be randomly selected.

It is often appropriate to obtain stakeholder inputs at several points during the testing process, which includes obtaining inputs on face validity, feasibility, and burden. These inputs can take many forms, including but not limited to, formal technical expert panels (TEPs), consultation with subject matter experts (SMEs), outreach to professional associations or patient advocacy groups, and public comments. Once obtained at a given step, it is important to follow-up on those communications by providing additional opportunities for stakeholders to comment on the results of their inputs at future stages.

If the measure(s) are submitted to NQF for endorsement, the measure developer helps the CMS Contracting Officer's Representative (COR), as directed, by updating the measure submission documentation with the results of the measure testing. Measure developers also provide additional information as needed and are available to discuss testing results with NQF throughout the endorsement process.

The MMS Blueprint describes the types of testing that may be conducted during measure development, the procedure for planning and testing under the direction of the COR, and key considerations when analyzing and documenting results of testing and analysis, including incorporation of stakeholder inputs after testing is complete.