

## **Response Criteria Attachment 2 - Glossary**

**Evaluative measure** – measures that provide a quantitative or semi-quantitative estimate of disease activity.

**Response criteria** – judgments that relate changes in evaluative measures to clinically meaning improvement.

**Disease activity** – changes that reflect a continuing, but reversible, pathophysiologic disease process; excludes findings known or suspected to be caused by other disease processes.

**Disease damage** – changes that persist after cessation of disease activity; includes scarring, atrophy, and fibrosis.

**Validation of measurements** – a demonstration that measurements perform as expected and intended. Measurements must be repeatable across time, repeatable between evaluators, and show change when change has occurred. Measurements must also have reliability and validity (see below). When a number of individual assessments are combined in to a composite endpoint, each individual component must be validated. The overall composite endpoint must also be validated in order to assess any unforeseen additional variability.

**Reliability** – a gauge of how well the results of a measurement can be replicated, assuming no change in the underlying characteristic being studied. Reliability is also termed as reproducibility, repeatability, or consistency. Reliable measures have low variability, which differs from validity.

**Validity** – a measure of the extent to which an assessment technique measures that which was intended. Measures can be highly reliable but not valid (for a specific assessment) or valid but not reliable. For example, measuring time to walk 50 feet may be highly valid for assessing lower limb and joint function but not useful for assessing joints in the hand.

Four types of validity have been defined.

1. *Face (also termed clinical or biological) validity.* A technique has face validity if it measures the particular attribute of interest as determined by experienced individuals. For example, face validity for the skin assessment indicates agreement that chronic GVHD often involves the skin.
2. *Content validity.* The index measures all relevant aspects of the disease. For example, content validity for the skin means that the proposed measures encompass all major skin manifestations of chronic GVHD.
3. *Construct validity.* Construct validity is concerned with the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses. For example, construct validity for measures of skin

manifestations means that a measure correlates with skin symptoms reported by patients.

4. *Criterion validity.* Criterion validity has two components: concurrent and predictive. Criterion validity assumes a gold standard and measures agreement with the gold standard. Concurrent criterion validity measures agreement with the gold standard at the time of the assessment. The clinician's therapeutic impression at the time of the assessment could serve as the gold standard for concurrent criterion validity. Predictive criterion validity measures agreement with the gold standard at a future time. A major clinical outcome such as disease-free survival to permanent discontinuation of immunosuppressive treatment could serve as the gold standard for predictive criterion validity.