

Statistical Issues and Considerations for TBI Tests

FDA Public Workshop - Advancing the
Development of Biomarkers in
Traumatic Brain Injury

March 3, 2016

Bipasa Biswas, PhD

FDA/CDRH/OSB/DBS

Bipasa.biswas@fda.hhs.gov

301-796-5740

Outline

- Introduction
- Testing versus validation
- Intended Use/Indications for Use
- Measurement and/or analytical validation
- Clinical evaluation of diagnostic tests
- Summary

Introduction

- TBI biomarker tests are diverse
- Examples: EEG, Imaging, Protein assays
- There are challenges for evaluating tests from this diverse group.
 - Evaluation depends on the intended use and the indications for use
 - Types of measurement validation and/or analytical studies

Biomarker and Biomarker Tests

- **Biomarker:** A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions.
 - Molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers.
 - A **biomarker** is not an **assessment** of how an individual feels, functions, or survives.
- **Biomarker test:** a lab test, medical imaging method, or other objective clinical assessment method used to detect or measure an indicator of biologic processes or pharmacologic responses to a treatment (biomarker).

Development and Validation of Biomarker Test

- Measurement validation and clinical evaluation study after finalization of the TBI biomarker test.
 - Once the test is finalized and ‘locked down’ with regard to its design parameters, cutoff selection, the measurement validation and clinical evaluation in a study population independent and separate from that used in the development of the test.
- Independent validation is desired as it objectively assesses the device performance external to the conditions and the data set used in development of the test and thus avoids issues related to “training” bias.

Intended Use(IU)/ Indications for Use (IFU)

- IU/IFU determines the device classification, the clinical purpose, study design, the statistical hypothesis or the estimation goals and the analysis.
- The components of IU/IFU among many other
 - What the device measures, identifies or detects
 - Where the device measures and on what it measures
 - Where is the device used (Environment- ER, outpatient clinic, home?)
 - What the device reports (Quantitative, continuous, categorical?)
 - For what (target condition)
 - On whom (target population)
 - By whom (users of the device)

Test Validation and Evaluation

- Measurement/Analytical validation (Measurement validation involves the characterization of various aspects of a biomarker test's ability to measure the biomarker)
 - Measurement Precision
 - Measurement Bias
 - Analytical studies (limit of blank, limit of detection, limit of quantitation, linearity, bias, precision etc.)
- Clinical validation (two approaches)
 - Clinical Performance Study
 - Clinical Outcome Study

Measurement Validation and Intended Use (IU)/Indications for Use (IFU)

- IU/ IFU impacts the study design and conduct of measurement validation studies.
 - Measurement validation studies for a categorical output test may differ from measurement validation studies for a quantitative output test.
 - Measurement validation studies for a test meant to be used in clinical labs might differ from a test meant to be used at home.
 - For a test with one or more medical decision points, important to assess test's measuring ability near medical decision point, so that users of test can understand what variability exists around the medical decision point(s).

Clinical Evaluation

- Important considerations for clinical evaluation study:
 - Intended Use (IU)/Indications for Use (IFU)
 - Study Design
 - Study Conduct
 - Statistical Analyses of Collected Data
 - Reporting of the results from the clinical studies

Diagnostic Clinical Outcome Study

- Diagnostic device result is used during a treatment or management intervention
- Performance is assessed in part by the intervention's effect on subject outcome (clinical performance may also be evaluated)

Example: evaluate consequences of using the test through a randomized trial (Diagnostic device use in one arm and standard of care in the other arm of the two-arm randomized clinical trial)

Diagnostic Clinical Performance Study

- The basic performance characteristics of a test are to inform how well the test measures, classifies or predicts what it intends to measure compared to a comparative benchmark (the clinical reference standard).
 - For example, the basic performance measures to assess diagnostic accuracy of a qualitative test to distinguish diseased from non-diseased involve sensitivity, that is, the probability that a truly diseased individual will test positive for disease, and specificity, that a truly non-diseased individual will test negative for disease. These measures are usually expressed as a percentage.

Common Clinical Performance Measures for Dichotomous (+/-) Tests

- Sensitivity and specificity pair
- Likelihood ratio of positive test and likelihood ratio of negative test pair
- Positive predictive value and negative predictive value along with prevalence.
- Positive and negative percent agreement pair (if compared against an imperfect reference standard)

Clinical* Performance for Quantitative Test

Almost always involves assessment of measurement:

- Precision (closeness of *repeated* results, e.g. repeatability standard deviation, closeness of results by varying factors, e.g. reproducibility standard deviation)
- Bias (closeness to *right* result, on average, e.g. estimated bias with respect to a reference result)

*Note: precision and bias can be evaluated at the bench/analytically (part of measurement validation) *and* clinically

Common Clinical Performance for Quantitative Tests

- clinical precision (repeatability, reproducibility)
- slope and intercept from a linear regression (many different kinds), scatter plots
- bias (mean difference) between the new test and the reference method
- 95% limits of agreement, Bland-Altman difference plots

Summary

- Intended Use (IU)/Indications for Use (IFU) and the type of test result (Continuous, categorical, qualitative) drive the design, target population, setting of the test and planning of measurement validation and clinical evaluation studies.
- Development of the test and any parameters, cutoffs need to be finalized before validation studies.
- Statistical analysis and reporting of results depend on the IU and IFU and thus it is important to plan the study carefully at the beginning to align the performance with the IU/IFU

Acknowledgements

- Lakshmi Vishnuvajjala
- Kristen Meier
- Arkendra De

References

1. FDA Guidance “Design Considerations for Pivotal Clinical Investigations for Medical Devices” ;issued on November 7, 2013

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM373766.pdf>

2. FDA Guidance “Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests”; issued on: March 13, 2007

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf>

3. Clinical and Laboratory Standards Institute (CLSI) www.clsi.org

4. ISO (International Standards Organization) (standards for terminology, estimating bias and imprecision of test methods)

3. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative, Clin Chem 2003; 49(1): 1-6. (see also <http://www.stard-statement.org>)