ELSEVIER

DRUG DISCOVERY TODAY TECHNOLOGIES

Drug Discovery Today: Technologies

Vol. 4, No. 1 2007

Critical path

Biomarkers for pharmacogenetic and pharmacogenomic studies: Locking down analytical performance

Elizabeth Mansfield*, Zivana Tezak, Sousan Altaie, Kathleen Simon, Steven Gutman

Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, Food and Drug Administration, Rockville, MD 20850, USA

The use of *in vitro* tests to detect and measure biomarkers provides promising avenues for development of new and better drugs, and will be central to the realization of personalized medicine. The importance of proper biomarker test assessment cannot be overemphasized. Whether the test is being used as part of drug development or ultimately used as a companion diagnostic, if the test result is to be meaningful, the test analytical performance must be well characterized. This article will outline important analytical validation issues to consider when developing and assessing an *in vitro* diagnostic test system for use in pharmacogenetic and pharmacogenomic studies.

Section Editor:
Janet Woodcock – Food and Drug Administration, Rockville, MD, USA

Introduction

The analytical performance characteristics of a test will play a large role in how the test will perform in clinical use, and so particular attention should be directed at understanding the analytical, or measurement, issues before using the test as a basis for decision-making in drug development. Conclusions reached based on studies of a system in analytical flux may yield results reflective of testing artifact and whimsy rather than actual pathophysiological phenomenon of medical interest.

*Corresponding author: E. Mansfield (Elizabeth.Mansfield@fda.hhs.gov)

Analytical test parameters

Although *in vitro* biomarker tests will vary in their intended uses, applicable populations, technologies, and use settings, there are several core analytical performance measures that can be generally applied to all types of tests. Test developers should validate their test performance in terms of these core areas by developing test protocols and generating sufficient data to assure that the performance is adequate to inform its use in a trial or in the clinical laboratory.

Accuracy and trueness

Trueness is defined as the closeness of the agreement between the average value obtained from numerous test measurements and a true value. The difference between the average value and the true value is a bias (systematic bias). The ISO (International Organization for Standardization) has introduced the term 'trueness' as a replacement for the term 'accuracy' which now is defined as the closeness of the agreement between the result of a measurement and a true value of what is being measured. Therefore, accuracy includes both bias and imprecision (random error) and is related to the total error of the test system. For *in vitro* tests, it is often not possible to obtain a true value of the analyte or measurand to use for estimating new test accuracy. For many analytes,

1740-6749/$ .Published by Elsevier Ltd. DOI: 10.1016/j.ddtec.2007.10.002

however, it is possible to obtain or generate a reference value using a reference method and materials to which the output, or test measurement, from the new test system can be compared. Thus what is referred to as accuracy is often an assessment of agreement by method comparison. While ideally the reference method being used to study the new diagnostic should be a well-defined traceable reference method, in some cases this is not possible. In these instances, test developers should characterize the comparative method as completely as possible to assure that performance of the new test is understood and that the ability to define bias against the predicate method chosen for comparison is also well understood [1,2].

Statistical measures of comparison will differ depending on whether the measurement is qualitative (i.e. yields a result that is one of two categories, such as yes/no, positive/negative, among others), semi-quantitative (i.e. yields a result that reflects numerically based categories, such as 'two-fold increase', or a ratio) or quantitative (i.e. yields a numerical result, such as 10 mg/dL, or 50 copies). Samples used for establishing accuracy of a method should correspond to the specimen type(s) (also called 'matrices') that will be used, for example, plasma, serum, or formalin-fixed paraffin embedded, in the application of the test. If there are multiple specimen types to which the test is intended to be applied, each should be evaluated in determining test accuracy. When possible, testing should be performed using actual clinical specimens. However, in some cases manipulated specimens (spiked or diluted) can be used provided that measures are taken to ensure comparability of the manipulated specimens to 'real' clinical specimens, in terms of analyte concentration, and other relevant factors.

Precision/reproducibility

Precision is defined as the closeness of agreement between results of measurements obtained under stipulated conditions. The degree of precision is usually expressed as SD (standard deviation) or CV (coefficient of variation). The general concept of precision includes measures of both repeatability and reproducibility. Repeatability is closeness of agreement between results from measurements carried out under the same conditions (i.e. within-run precision), while reproducibility is closeness of agreement between the results of measurements carried under changed conditions, for example, when testing is carried out at different sites (if applicable), using different reagent lots, by different operators, over different instruments, and other sources of variation. A valid statement of precision requires specification of the conditions that were changed. CLSI documents provide information on methodologies for measuring and calculating precision in a simple and standardized manner [2,3]. When studying multiple sites and multiple operators, these should be carefully selected to represent the expected intended users of the test, rather than selecting to obtain 'best case' performance.

Precision measurements ideally should reflect the entire testing process including pre-analytical preparation steps, should be performed using real clinical samples that span the testing range, and should include clear elucidation of both intra and inter-assay reproducibility. If a test is to be performed using multiple different types of instruments or reagents, for example, thermocyclers from several manufacturers, or if the test can be performed both manually and using automated instrumentation, precision studies should be performed according to each relevant permutation available. While it is important to estimate precision across the range of the test, precision estimates made at medical decision points, or cut-offs, are particularly valuable since this information will be most helpful in guiding the clinical use of a particular test.

Linearity/reportable range

Quantitative analytical studies should include information on the linearity of test results and the measurement range over which reliable results can be expected. Linearity of response, or at least monotonic response, may also be important for semi-quantitative tests where there are cut-off values or medical decision points based on particular signal thresholds. Linearity is often assessed using dilutions of a sample with a high (natural or spiked) analyte concentration, for example, two-fold dilutions, with multiple replicates at each dilution. Testing should include samples at the highest and lowest points of the linear range, as well as samples that are near the cut-off or medical decision point(s). A CLSI document provides information on testing and analysis for evaluation of linearity [4].

Traceability, stability, expected values (controls, calibrators, or methods)

Calibrators and controls, and the methods employed in their use, are critical to generating reliable test results, by assuring that the test system is working within expected bounds. Test developers should identify or create appropriate calibrators and controls for new tests, and provide information on how they should be used to monitor the test system. Information should be established on source of materials, how values are assigned, and how the materials and methods used have been evaluated and found acceptable for use in calibrating and controlling the test system [5,6].

When recognized reference materials with traceable values are available for an analyte or test, these should be studied to demonstrate how test performance compares to the established 'metrologically credentialed' values. When recognized reference materials are not available, companies must create or source their own reference materials, credential these using comprehensive scientific studies and maintain these materials over time to assure the test system does not exhibit problematic drifts in performance.

All parts of a test system (e.g. reagents, instrument calibration, analyte in stored specimens) need to be stable over time, both in storage and in use conditions, to assure that test results accurately reflect the state of the analyte that is measured. Stability of all relevant sub-parts of the total testing system should be established using samples with values across the measurement range of the test, and limits of stability, for example, expiry dates, calibration intervals, among others, should be established and observed.

Detection limit

For most *in vitro* tests, it is important to determine the lowest level of analyte that can be detected, as well as the lowest amount that can be measured accurately (i.e. the total error is inside of allowable total error limits), because these measurement levels may determine whether a specimen is called positive or negative, may determine the minimum amount of specimen needed for testing, and may inform the choice of cut-off. When applicable, tests should be developed with information describing minimum sample requirements and limits of detection and quantitation for the analyte, or measurand. Methodologies for determining limits of detection, limits of quantitation and limits of blank are described in a CLSI document [7].

Analytical specificity

The presence in a clinical specimen of substances that either interfere with the generation of an analytical signal, or cross-react to generate spurious analytical signals, can result in false test results, sometimes at high rates. Although assumptions are often made regarding the low likelihood of interference or cross-reactivity, with widespread use of a new test, surprises in the form of specificity problems are one of the most commonly observed problems. Careful up-front evaluation and attention to test design to avoid, eliminate, or minimize such substances can help limit unexpected changes in performance [8,9]. Interfering and cross-reactive substances may be present endogenously in the specimen, such as, human anti-mouse antibodies (HAMA), pseudogenes, among others, or may be introduced during stabilizing or processing steps, such as heparin in blood collection tubes, protein adducts resulting from fixation, among others. Studies to evaluate interference and cross reactivity from both known and potentially problematic substances in specimens should be carried out using analyte concentrations near medical decision points, or cut-offs, and at the upper and lower limits of measurement. Carry-over and contamination studies should also be performed to assess the potential for test reagents and endogenous or introduced substances in a sample to be transferred to another sample and thus to cause false results. Studies of interfering substances should include sufficient levels of the interfering substance and sufficient number of specimens containing those interfering substances to provide reliable estimates on the potential magnitude of the problem.

Assay cut-off

Nearly all *in vitro* tests incorporate a cut-off of some form. There are many ways to establish the analytical cut-off for a test, and analytical cut-offs should be determined using a statistical method that is appropriate to the use of the test. For some quantitative tests, the cut-off might be determined as 3 or more standard deviations above the mean of a negative (or 'normal') sample. If the distribution of measurements of the negative samples is not Gaussian, then the cut-off may be determined as 95th or 99th percentiles. The cut-off for a qualitative test is the threshold above which the observed result is reported as positive and below which the result is reported as negative. For qualitative tests one technique for choosing cut-offs is to determine the measurement level at which repeated tests of a negative sample yield positive results 50% of the time and negative results for the other 50%. This difference in response is a reflection of measurement error and not of the true value of the sample.

Test design should include steps to establish an analytical cut-off, using an appropriately selected statistical method for directing the choice of the cut-off, with a clear rationale for the units used, and the categories of results selected. Test data should be carefully examined to determine if an equivocal zone (sometimes called an indeterminate zone) may be warranted. If such a zone is needed, data should be generated to establish limits of the zone. Analytical testing to select a cut-off should account for various matrices (specimen types) to be used in the test, and how analyte levels are assigned for samples used in the cut-off determination. Once analytical cut-offs have been established using analytical or statistical techniques, they ultimately require validation for their clinical value.

Quality control

In vitro tests require the use of quality control materials to assure that the tests are generating expected results at all levels of analyte to be tested. Ideally quality control materials should be identical (or at the least, similar) in matrix to the specimens intended for use, to ensure that quality control testing provides an adequate challenge to the system being evaluated. Quality control levels should challenge the cut-off points (e.g. controls selected slightly above and below the cut-off(s)), and should be performed frequently enough to ensure that failures in test performance will be identified in a timely manner. The controls should provide information about specimen quality, specimen pre-analytical processing quality, functioning of the instrument/software systems, and overall functioning of the entire testing system from pre-analytical processing through delivery of test result.

In designing a quality control strategy, it is important to determine the following: the nature and function of various

controls that will be needed to assure quality test results; the methods used to assign values to control materials, and to validate the values assigned; and the control parameters that will signal deterioration or failure of the test system, including failure of instrumentation to meet specifications [10]. These decisions should be validated through testing, and materials and instructions for their proper use should be provided with the test system, or their sourcing adequately described to enable users of the test to procure them.

Conclusions

Whether a new biomarker is used for clinical studies, exploratory drug studies, elucidation of basic pathophysiological processes, or basic research, the ability to assure high quality of analytical testing is imperative and should be addressed before generating data which will be used to make drug development or clinical decisions. The process of establishing analytical performance is now well established and described most prominently in a series of voluntary standards developed by the Clinical Laboratory Standards Institute. Key parameters in evaluating a new biomarker within a test system include: accuracy, precision or repeatability, linearity and reportable range, traceability/stability, detection limit, analytical specificity, and test cut-off. The Food and Drug Administration (FDA) has also crafted several guidance documents that may be valuable in understanding how to establish analytical performance for new test systems.

In any test system, an assurance of reliable results must reference a total system quality control of the analytical performance of the test. Quality control materials and methods should be a non-trivial, integral part of the test system, and built into the test design, with careful attention to the nature of the materials that will be required to adequately interrogate the system, as well as to the conditions under which they should be used.

Using standard evaluation techniques and providing transparency in data gathering for the analytical performance of a test should provide more confidence in the use and adoption of new tests. Good study practices, transparency in data generation, and importantly, sharing of data should also allow for data comparison and pooling across sites and across studies. Grounding new biomarkers with rigorous analytical studies and providing transparency in data sets is likely to advance the field of biomarker exploration, promote biomarker use, and lead to improved research, development, and clinical outcomes.

References

1 Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline (EP9-A2), Clinical Laboratory Standards Institute, 2002

2 User Protocol for evaluation of Qualitative Test Performance; Approved Guideline (EP12-A), Clinical Laboratory Standards Institute, 2002

3 Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline (EP5-A2), Clinical Laboratory Standards Institute, 2004

4 Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline (EP6-A), Clinical Laboratory Standards Institute, 2003

5 *In vitro* diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials (ISO 17511), 2003

6 How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline (C28-A2) Clinical Laboratory Standards Institute, 2000

7 Protocols for Determination of Limits, of Detection and Limits of Quantitation; Approved Guideline (EP17-A) Clinical Laboratory Standards Institute, 200

8 Interference Testing in Clinical Chemistry; approved Guideline (EP7-A2) Clinical Laboratory Standards Institute, 2005

9 Evaluation of Matrix Effects; Approved Guideline (EP14-A2), Clinical Laboratory Standards Institute, 2005

10 Statistical Quality Control for Quantitative Measurements; Principles and Definitions; Approved Guideline (C24-A3), Clinical Laboratory Standards Institute, 2006