The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2016

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

REVIEW

Clinical Performance Evaluation of Molecular Diagnostic Tests

Q9 Bipasa Biswas

From the Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland

Accepted for publication June 1, 2016.

Address correspondence to Bipasa Biswas, Ph.D., Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD. E-mail: bipasa.biswas@fda.hhs.gov.

Molecular diagnostic tests with application to clinical diagnostics involve studies in infectious diseases, inherited diseases, oncology, predisposition to disease, or the description of polymorphisms linked to disease states. General considerations in the design of evaluation of diagnostic test trials and statistical principles for reporting the results are discussed. A brief overview of the general statistical considerations related to the intent of use, test development versus validation, different types of biases, and issues with missing data are provided. Furthermore, issues related to commonly used but not necessarily correct methods to characterize the performance in the presence and absence of a clinical reference standard are discussed. These issues are broadly applicable to any molecular diagnostic test with a dichotomous result. This overview may help the clinical molecular diagnostic community to evaluate tests that provide a dichotomous result. *(J Mol Diagn 2016, ■: 1–10; http://dx.doi.org/10.1016/j.jmoldx.2016.06.008)*

Q4 Molecular diagnostic tests encompass a wide area of testing, such as testing for infectious diseases, oncologic tests, genetic tests for inherited diseases, and testing for predisposition to disease or polymorphisms linked to disease states, where the test involves detection of specific molecules, such as DNA, antibodies, or proteins. In the field of oncology, DNA tests have been used for screening for cancer (a multitarget stool DNA test for colorectal cancer screening[1]), microbial assays have been used to diagnose infectious diseases (assay for detection of group B *Streptococcus* in prenatal screening of specimens[2]), qualitative nucleic acid tests have been used for confirmation of hepatitis C virus infection and for screening blood donations,[3] and genetic tests have been used for inherited diseases (next-generation sequencing for cystic fibrosis transmembrane conductance regulator screening[4]). Molecular diagnostic test requires both analytical and clinical evaluations.[5–11]

Molecular diagnostics involve techniques to analyze biomarkers[12] in the genetic code of organisms, the genome, and how the cells express their genes as proteins, the proteomes.[5,12] These techniques apply molecular biology for medical testing to diagnose symptomatic individuals, screen asymptomatic individuals, monitor disease, provide prognosis in diseased patients, detect risk, and select patients for specific therapies. Molecular diagnostic tests use biological assays that detect a molecule, often in low concentrations, using PCR enzyme-linked immunosorbent assay or fluorescence *in situ* hybridization.[7–11,13,14] The detection of the biomarker uses real-time PCR, direct sequencing, or microarrays. Advances in next-generation sequencing will enable high-throughput DNA sequencing at relatively low cost for genomic-based diagnosis.[15]

Biomarker evaluation[12] by molecular diagnostics involves evaluation of both analytical performance and clinical performance. The analytical performance relates to the ability of the molecular diagnostic test to measure the underlying biological quantity under a variety of condition; although an important aspect of the test, it will not be discussed here. Several consensus standards are available to design and evaluate analytical performance of molecular diagnostic tests,[16–26] and useful resources are available from the US Food and Drug Administration's Center for Devices and Radiological Health Standards Program

Disclosures: None declared. Q1 Q2

Portions of this work were presented at the 2006 US Food and Drug Administration/Industry Statistics Workshop held September 27–29, 2006, in Washington, DC.

http://dx.doi.org/10.1016/j.jmoldx.2016.06.008

(US Food and Drug Administration, *http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Standards/default.htm*). This review article focuses on the clinical performance evaluation for molecular diagnostic tests with dichotomous output. Clinical performance assesses the test's ability to detect the clinical or target condition of interest. A test result can be continuous, ordinal, or nominal.[27] A continuous or ordinal test result can be dichotomized[28] to give only two responses or output by using a cutoff. The test output for the molecular diagnostic test with dichotomous output is referred to as positive and negative in this review, which can also be interpreted qualitatively as the presence or absence of a target or clinical condition of interest. This article initially reviews general considerations, such as intent of use, development and validation, study conduct, and biases, then discusses possible performance measures and alludes to certain pitfalls of commonly used measures for performance evaluation, and finally discusses sample size justification and statistical analysis.

General Considerations in the Evaluation of Clinical Performance Trials

Clinical diagnostic performance for molecular diagnostic tests with dichotomous output is best evaluated with proper planning with respect to the intent of use, delineating development from validation, and adhering to appropriate study conduct to avoid potential sources of bias. Reporting of results is appropriately addressed by allowing for understanding of the study methods, the limitations involved, and correct interpretation of results.

Intent of Use

The intent of use of a molecular diagnostic test determines the type of study required to establish its performance. The intent of use describes the clinical purpose, the type of test, the criteria it measures, the specimen it measures (specimen type), the site of measurement, and the population for which the test is intended. Many variables can influence the performance of a test, such as population characteristics, the prevalence of the target condition of interest, the setting, and the type of test, among others. Thus, it is important to design the performance evaluation studies to match the intent of use. In general, it is important to include the following: the clinical purpose (eg, screening, diagnosis, prognosis, risk prediction, therapy or treatment selection for patients), target condition (eg, disease, disease stage, or any other condition of interest), target population, and the environment (eg, clinical laboratory, point of care, home use). Other important things to consider while designing a clinical study are anatomical location (eg, finger stick, venous) or specimen type from which the measurement is taken (eg, whole blood, plasma, serum, tissue), the measurand (which is being measured or detected), type of results (quantitative, continuous, ordinal, or qualitative) from the test, clinical interpretation of the test results, and the need for a trained or skilled user of the test and interpreter or reader of the test result.

Clinical Test Development and Validation

Medical tests often involve a number of technology and design parameters that are established in preclinical studies before conducting validation studies. For example, if the test is intended to be used qualitatively by dichotomizing the test result at a single cutoff or a decision threshold, then this has to be established before the final clinical validation study. This review article focuses on clinical performance of the molecular diagnostic test after finalization of all the design and technologic parameters, and thus considerations during the development are not the focus of further discussion.

A cutoff selection for a molecular diagnostic test with continuous or ordinal output may use the receiver operating characteristic (ROC) curve to select an optimum cutoff based on the clinical needs. The data set used to select an optimum cutoff is a training data set. An independent evaluation of the cutoff requires an assessment in an external data set that is independent and separate from that used in the selection of the cutoff. The ROC curve, for comparing two tests, provides additional support to discern whether a new test is better than a comparator test, although the test is to be used qualitatively by dichotomizing the test output. The ROC curve, which is a plot of 1—specificity and sensitivity on the *xy*-coordinate plane, helps to differentiate whether a new test is indeed on a different ROC curve that is superior to an existent test or whether the new test is just on the same ROC curve but that its operating point (cutoff or decision threshold) has been moved to provide a higher sensitivity at a loss of specificity. Further discussions related to cutoff selection at the development stage and the statistical techniques can be found in previously published articles.[25,26,28–30]

Once the test is finalized with regard to its design parameters and cutoff selection, the clinical performance is evaluated in a study population independent and separate from that used in the development of the test. Independent validation is desired because 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.

Study Conduct

Evaluation studies can be subject to many types of biases,[30–35] and careful consideration is needed at the study design stage and/or during analysis and reporting of performance to avoid potential sources of biases. Commonly observed sources of bias are selections bias, bias attributable to spectrum effect, verification bias, test evaluation bias,

REV 5.4.0 DTD ■ JMDI528_proof ■ 15 September 2016 ■ 5:27 pm ■ EO: JMD15_0078