Statistical Principles for Omics-based Clinical Trials

Michael C Sachs
National Cancer Institute
michael.sachs@nih.gov

Lisa M McShane
National Cancer Institute
McShaneL@CTEP.NCI.NIH.gov

September 2014. Incomplete Draft. Please do not cite without permission.

Abstract

High-throughput techologies enable the measurement of a large number of molecular characteristics from a small tissue sample. High-dimensional molecular information (referred to as omics data) offers the possibility of predicting the future outcome of a patient (prognosis) and predicting the likely response to a specific treatment (prediction). Embedded in the vast amount of data is the hope that there exists some signal that will enable practicioners to deliver therapy personalized to the molecular profile of a tumor, thereby improving health outcomes. The challenges are to determine that the omics assays are valid and reproducible in a clinical setting, to develop a valid and optimal omics-based test that algorithmically determines the optimal treatment regime, to evaluate that test in a powerful and unbiased manner, and finally to demonstrate clinical utility: that the test under study improves clinical outcome as compared to not using the test. We review the statistical con-

siderations involved in each of these stages, specifically dealing with the challenges of high-dimensional, omics data.

Keywords. genomics; personalized medidcine; predictive biomarker; statistics

1 Introduction

Omics technolgies that generate a large amount of molecular data about a cancerous tumor have the potential to provide accurate predictions of a patient's prognosis and their response to a specific treatment regime. It has long been recognized in oncology that tumors of a primary site represent a heterogeneous collection of diseases that may have different characteristics and respond to different therapies [Citation needed]. The idea of omics-based biomarkers is that different tumor types can be identified using the multi-dimensional molecular data, leading to the development of new therapies targeting specific pathways, or leading to treatment decisions personalized to that tumor type.

The prognosis of a patient is their expected clinical outcome. An omics-based test can be used to predict a patient's prognosis. A test that provides accurate predictions of prognosis, regardless of treatment, is referred to as a prognostic biomarker. A predictive omics-test is one that accurately predicts disease outcomes with the application of specific interventions. Predictive markers are therefore useful for the selection among two or more treatment options. Statistically, a prognostic test is strongly associated with clinical outcome and a predictive test modifies the association between treatment and clinical outcome (interaction). The two are not mutually exclusive. It is uncommon for a test to be purely predictive (1), and prognostic tests can be used to inform treatment

decisions.

A patient's prognosis is used to determine the type and intensity of medical treatment. Endopredict (2) is an omics-based test test that is used to determine the likelihood of distant recurrence in ER-positive, HER2-negative breast cancer. The test has been shown to accurately predict prognosis, and is therefore useful for guiding treatment decisions, determining eligibility for trials, and making disease-management decisions. Endopredict as a prognostic test has been rigorously evaluated and shown to be clinically valid, even thought it does not predict response to any specific therapy.

The use of omics technologies for drug development is beyond the scope of this review. High dimensional omics data can be used to identify specific molecular targets as potential mechnisms for drug development. (3) reviews some of the regulatory issues involved in validating genomic markers for drug targets.

Here we review the statistical issues in each of the steps for the evaluation of omics-based tests in clinical trials. The tests can be prognostic, predictive, or both. The long road to implementing a test in a clinical trial starts with analytical validation, that is, demonstrating that the omics-based assay accurately and reproducibly measures the molecular quantities. After the assay performance is established comes the test development and preliminary evaluation. This involves reducing the high-dimensional data into a one-dimensional quantity that will be used to make a decision. This one-dimensional quantity is often a risk score: an estimate of the probability of a speicific clinical outcome. It is necessary to establish the clinical validity of this risk score, that is, demonstrate that the risk score is independently associated with clinical outcome.

Care must be taken to completely separate the development of the risk score from the evaluation, otherwise estimates can be optimistically biased. Finally, the risk score must be translated into a binary decision, often using a threshold. It remains to demonstrate that the use of the test to make this decision improves patient outcomes.

2 Analytical validation

3 Test development and preliminary evaluation

4 Demonstrating clinical utility

5 Concluding remarks

References

1. McShane LM, Polley M-YC. Development of omics-based clinical tests for prognosis and therapy selection: The challenge of achieving statistical robustness and clinical utility. Clinical Trials. SAGE Publications; 2013;10(5):653–65.

2. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in eR-positive, hER2-negative breast cancer adds independent information to conventional clinical risk factors. Clinical Cancer Research. American Association for Cancer Research; 2011;17(18):6012–20.

3. Goodsaid F, Frueh F. Process map proposal for the validation of genomic biomarkers. Pharmacogenomics. 2006;7(5):773–82.