

Statistical Principles for Omics-based Clinical Trials

Michael C Sachs

Lisa M McShane

National Cancer Institute

National Cancer Institute

`maito:michael.sachs@nih.gov`

`maito:McShaneL@CTEP.NCI.NIH.gov`

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 Please do not cite without permission.

Abstract

High-throughput technologies 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 practitioners 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 medicine; predictive biomarker; statistics

1 Introduction

Omics technologies 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. The idea of omics-based biomarkers is that distinct tumor types can be identified using the multi-dimensional molecular data leading to treatment decisions personalized to that tumor type. An omics-based test can guide the decisions to treat or not to treat and help identify the particular therapy most likely to work. The challenge is to identify and demonstrate definitively that the use of an omics-based decision rule improves clinical outcomes in a patient population.

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, however. 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 can be used to determine the type and intensity of medical treatment, or whether to treat at all. Endopredict (2) is an omics-based 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 though it does not predict response to any specific therapy.

The goal of this paper is to review the path to definitively evaluating an omics-based test for prognosis or prediction of treatment response. We assume that the patient population is well-defined, and that there may exist targeted therapies for a subset of that population. High dimensional omics data can be used to identify specific molecular targets as potential mechanisms for drug development, however the use of omics technologies for drug development is beyond the scope of this review.

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 specific 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

Analytical validation of an assay involves evaluating the performance of the measurement in terms of accuracy, bias, and precision under a variety of conditions. Conditions are things like preanalytic factors such as specimen quality, specimen collection, storage, and processing procedures, and technical aspects such as laboratory technician and batch effects from reagent lots or other assay materials. The high-dimensional nature of omics data makes it very difficult to assess each of the hundreds or thousands of outputs from a single assay. In developing a omics-based signature that only uses a subset of the components of a high-dimensional assay, one can analytically validate the final signature alone. However, prior to developing the signature, one must develop detailed standard operating procedures for specimen handling and processing to ensure a baseline level of validity.

Do: develop criteria for the rejection of poor-quality specimens. Percent tumor, necrosis, etc.

Do: filter features based on QC prior to development

Do: assess the impact of sample and specimen handling. (3–7)

Do: assess the impact of lots and batch effects. Bias, accuracy. (8,9)

Do: minimize the impact of technical aspects to greatest extent possible by developing detailed SOP.

3 Test development and preliminary evaluation

Study design: consider retrospective (10)

Don't: confound technical factors with clinical outcomes. (11,12)

Do: maintain strict separation between development and evaluation.

Do: cross validation if you have a data-sparse setting. (1,13–15)

Don't: use convoluted methods leading to overfitting.

Don't: do partial resubstitution

Compare: feature filtering based on association with outcome, regularization. (16,17)

Do: consider all available methods, model averaging. Hard to determine best method in advance.

Don't: rely on clustering to yield good predictions of outcome.

4 Demonstrating clinical utility

Do: define the clinical use (18)

Do: use a valid and interpretable statistical method appropriate to that use (19–21)

Do: Design your study appropriately to answer the clinical question definitively (22–33)

Do: Power your trial appropriately (34,35)

Don't: make these mistakes (36–38)

5 Concluding remarks

Do: follow reporting criteria (39–41)

6 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, Dubsy 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. Werner M, Chott A, Fabiano A, Battifora H. Effect of formalin tissue fixation and processing on immunohistochemistry. *The American journal of surgical pathology*. LWW; 2000;24(7):1016–9.
4. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American journal of pathology*. Elsevier; 2002;161(6):1961–71.
5. Maldegem F van, Wit M de, Morsink F, Musler A, Weegenaar J, Noesel CJ van. Ef-

- fects of processing delay, formalin fixation, and immunohistochemistry on rNA recovery from formalin-fixed paraffin-embedded tissue sections. *Diagnostic Molecular Pathology*. LWW; 2008;17(1):51–8.
6. Specht K, Richter T, Mller U, Walch A, Werner M, Hfler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *The American journal of pathology*. Elsevier; 2001;158(2):419–29.
 7. Iwamoto KS, Mizuno T, Ito T, Akiyama M, Takeichi N, Mabuchi K, et al. Feasibility of using decades-old archival tissues in molecular oncology/epidemiology. *The American journal of pathology*. American Society for Investigative Pathology; 1996;149(2):399.
 8. Pennello GA. Analytical and clinical evaluation of biomarkers assays: When are biomarkers ready for prime time? *Clinical Trials*. SAGE Publications; 2013;1740774513497541.
 9. Isler JA, Vesterqvist OE, Burczynski ME. Analytical validation of genotyping assays in the biomarker laboratory. *Future Medicine Ltd*; 2007;
 10. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute*. Oxford University Press; 2009;101(21):1446–52.
 11. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*. Nature Publishing Group; 2010;11(10):733–9.
 12. Sonesson C, Gerster S, Delorenzi M. Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PloS one*. Public Library of Science; 2014;9(6):e100335.

13. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. development, internal validation, and assessing the incremental value of a new (bio) marker. *Heart*. BMJ Publishing Group Ltd; British Cardiovascular Society; 2012;98(9):683–90.
14. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. external validation, model updating, and impact assessment. *Heart*. BMJ Publishing Group Ltd; British Cardiovascular Society; 2012;heartjnl-l2011.
15. Polley M-YC, Freidlin B, Korn EL, Conley BA, Abrams JS, McShane LM. Statistical and practical considerations for clinical evaluation of predictive biomarkers. *Journal of the National Cancer Institute*. Oxford University Press; 2013;105(22):1677–83.
16. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS biology*. Public Library of Science; 2004;2(4):e108.
17. Hastie T, Tibshirani R, Friedman J, Hastie T, Friedman J, Tibshirani R. *The elements of statistical learning*. Springer; 2009.
18. Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: Lessons from real trials. *Clinical trials*. SAGE Publications; 2010;7(5):567–73.
19. Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. *The international journal of biostatistics*. 2014;10(1):99–121.
20. Pepe MS. Problems with risk reclassification methods for evaluating prediction models. *American journal of epidemiology*. Oxford University Press; 2011;kwr013.
21. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: Do not rely

- on integrated discrimination improvement and net reclassification index. *Statistics in medicine*. John Wiley & Sons, Ltd; 2013;
22. Freidlin B, Korn EL. Biomarker enrichment strategies: Matching trial design to biomarker credentials. *Nature Reviews Clinical Oncology*. Nature Publishing Group; 2014;11(2):81–90.
 23. Baker SG, Sargent DJ. Designing a randomized clinical trial to evaluate personalized medicine: A new approach based on risk prediction. *Journal of the National Cancer Institute*. Oxford University Press; 2010;
 24. Baker SG, Kramer BS, Sargent DJ, Bonetti M. Biomarkers, subgroup evaluation, and clinical trial design. *Discovery medicine*. 2012;13(70):187–92.
 25. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, et al. Confirmatory adaptive designs with bayesian decision tools for a targeted therapy in oncology. *Statistics in medicine*. Wiley Online Library; 2009;28(10):1445–63.
 26. Denne JS, Pennello G, Zhao L, Chang S-C, Althouse S. Identifying a subpopulation for a tailored therapy: Bridging clinical efficacy from a laboratory-developed assay to a validated in vitro diagnostic test kit. *Statistics in Biopharmaceutical Research*. Taylor & Francis; 2014;6(1):78–88.
 27. Eng KH. Randomized reverse marker strategy design for prospective biomarker validation. *Statistics in medicine*. Wiley Online Library; 2014;
 28. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clinical Cancer Research*. AACR; 2010;16(2):691–8.
 29. Freidlin B, McShane LM, Polley M-YC, Korn EL. Randomized phase II trial designs

- with biomarkers. *Journal of Clinical Oncology*. American Society of Clinical Oncology; 2012;30(26):3304–9.
30. Freidlin B, Korn EL, Gray R. Marker sequential test (maST) design. *Clinical Trials*. SAGE Publications; 2013;1740774513503739.
 31. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: A procedure for evaluating treatment with possible biomarker-defined subset effect. *Journal of the National Cancer Institute*. Oxford University Press; 2007;99(13):1036–43.
 32. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. *Journal of Clinical Oncology*. American Society of Clinical Oncology; 2009;27(24):4027–34.
 33. Morita S, Yamamoto H, Sugitani Y. Biomarker-based bayesian randomized phase II clinical trial design to identify a sensitive patient subpopulation. *Statistics in medicine*. John Wiley & Sons, Ltd; 2014;
 34. Mackey HM, Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. *Contemporary clinical trials*. Elsevier; 2013;36(2):664–72.
 35. Peterson B, George SL. Sample size requirements and length of study for testing interaction in a $1 \times k$ factorial design when time-to-failure is the outcome. *Controlled clinical trials*. Elsevier; 1993;14(6):511–22.
 36. Lee S. Mistakes in validating the accuracy of a prediction classifier in high-dimensional but small-sample microarray data. *Statistical methods in medical research*. SAGE Publications; 2008;
 37. Sargent DJ, Mandrekar SJ. Statistical issues in the validation of prognostic, predic-

- tive, and surrogate biomarkers. *Clinical Trials*. SAGE Publications; 2013;10(5):647–52.
38. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of dNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute*. Oxford University Press; 2003;95(1):14–8.
39. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: A systematic review. *PLoS medicine*. Public Library of Science; 2012;9(5):e1001221.
40. Janssens AC, Ioannidis J, Bedrosian S, Boffetta P, Dolan SM, Dowling N, et al. Strengthening the reporting of genetic risk prediction studies (gRIPS): Explanation and elaboration. *European journal of clinical investigation*. Wiley Online Library; 2011;41(9):1010–35.
41. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials: Explanation and elaboration. *BMC medicine*. BioMed Central Ltd; 2013;11(1):220.