

Biomarker Discovery and Validation: Statistical Considerations



Fang-Shu Ou, PhD,^{a,*} Stefan Michiels, PhD,^b Yu Shyr, PhD,^c Alex A. Adjei, MD, PhD,^d
Ann L. Oberg, PhD^a

^aDepartment of Health Sciences Research, Mayo Clinic, Rochester, Minnesota

^bDepartment of Biostatistics and Epidemiology, Gustave Roussy, OncoStat U1018, CESP, Inserm, Université Paris-Saclay, labeI Ligue Contre le Cancer, Villejuif, France

^cDepartment of Biostatistics, Vanderbilt University, Nashville, Tennessee

^dDepartment of Oncology, Mayo Clinic, Rochester, Minnesota

Received 1 November 2020; revised 4 January 2021; accepted 7 January 2021
Available online - 2 February 2021

ABSTRACT

Biomarkers have various applications including disease detection, diagnosis, prognosis, prediction of response to intervention, and disease monitoring. In this era of precision medicine, having validated biomarkers to inform clinical decision making is more important than ever. In this article, we discuss best the practices and potential issues in biomarker discovery and validation. We encourage team science partnerships to bring cutting-edge discovery from bench to bedside, leading to improved patient care and outcomes.

© 2021 International Association for the Study of Lung Cancer. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Biomarker; Exploratory analysis; Confirmation analysis; Clinical trial

Introduction

A biological marker (biomarker) is “a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions.”¹ Biomarkers have various applications, such as risk estimation, disease screening and detection, diagnosis, estimation of prognosis, prediction of benefit from therapy, and disease monitoring (Fig. 1). In oncology, biomarker candidates often consist of biological molecules found in cancer cells. The most common biomarkers are cancer-associated proteins, gene mutations, deletions, rearrangements, and extra copy numbers of genes. These molecules are sometimes secreted into the circulation and so may be detected by blood-based assay,

whereas others are present in cancer cells and so require a biopsy to obtain tissue for testing. An ideal biomarker satisfies the following properties: it should be either binary (i.e., present or absent) or quantifiable without subjective assessments; the result should be generated by an assay that is adaptable to routine clinical practice and has a timely turnaround (i.e., in a matter of days rather than weeks); the biomarker assay should be sensitive and specific; and most importantly, the biomarker should be detectable using easily accessible specimens.

Molecular biomarkers are used together with clinical information to achieve precision medicine to customize prevention, screening, and treatment strategies to a group of patients with similar characteristics (Fig. 1). Risk stratification biomarkers may identify patients at higher than usual risk of disease who should be monitored more closely than the general population, for example, smoking increases the risk of lung cancer.² Disease screening and detection biomarkers are used to detect diseases before symptoms manifest, when therapy has a greater likelihood of success, for example, low-dose computed tomography screening is recommended for patients at high risk of lung cancer.² Diagnostic biomarkers detect the

*Corresponding author.

Disclosure: Dr. Shyr reports receiving personal fees from AstraZeneca, Eisai, Janssen, Novartis, Pfizer, and Roche outside of the submitted work. The remaining authors declare no conflict of interest.

Address for correspondence: Fang-Shu Ou, PhD, Department of Health Sciences Research, Mayo Clinic, 200 First St. SW, Rochester, MN 55905. E-mail: ou.fang-shu@mayo.edu

© 2021 International Association for the Study of Lung Cancer. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 1556-0864

<https://doi.org/10.1016/j.jtho.2021.01.1616>

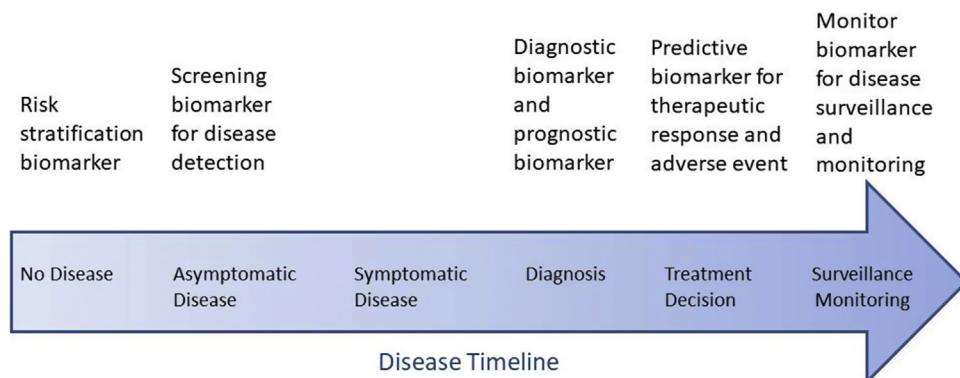


Figure 1. Use of biomarkers in relation to the course of disease.

presence of diseases, for example, biopsies can be used in the diagnosis of lung cancer.² Prognostic biomarkers provide information on overall expected clinical outcomes of a patient, regardless of therapy or treatment selection, for example, sarcomatoid mesothelioma has a poor outcome regardless of therapy.³ Predictive biomarkers inform the overall expected clinical outcome on the basis of treatment decisions in biomarker-defined patients only. The most important predictive biomarkers found for NSCLC, for example, are mutations in the *EGFR* gene, *BRAF*, or *MET* gene and rearrangements involving the *ALK*, *ROS1*, *RET*, and *NTRK* family genes⁴; various targeted therapies are available for patients identified by most of these biomarkers.

A biomarker's journey from discovery to clinical use is long and arduous, but it can be broken into phases or steps.⁵⁻⁸ Biomarker discovery efforts have increased with the emergence of technologies for gathering relevant data; for example, single-cell next-generation sequencing, liquid biopsy (blood sample) for circulating tumor DNA, microbiomics, radiomics, and other types of high-throughput technologies have exploded in popularity in recent years, owing to their ability to produce an enormous volume of data quickly and at relatively low cost. Across the continuum of biomarker data capture and utilization, however, many more challenges lie ahead—from analysis of high-throughput biomarker data to maximum exploitation of the electronic health record, and to the ultimate goal of biomarker-driven clinical practice. Biomarker discovery and validation are essential steps in establishing biomarkers in all applications across the disease course. In this article, we discuss the best practices for biomarker discovery and validation from a statistical perspective (Fig. 2).

Biomarker Discovery

The intended use of a biomarker (e.g., risk stratification and screening)¹ and the target population to be

tested need to be defined early in the development process. The use of a biomarker in relation to the course of a disease and specific clinical contexts should also be pre-specified (Fig. 1). The patients and specimens should both directly reflect the target population and intended use.

Key Considerations for Biomarker Discovery

Key considerations for conducting discovery studies using archived specimens are the patient population represented by the specimen archive, power of the study (through the number of samples and number of events), prevalence of the disease, the analytical validity of the biomarker test, and the preplanned analysis plan.⁹ The most reliable setting in which to perform such (retrospective) studies is by means of specimens and data collected during prospective trials, and the results of one study need to be reproduced in another. Definitions for levels of evidence have been developed to evaluate the clinical use of biomarkers in oncology and medicine.^{9,10}

Bias, a systematic shift from truth, is one of the greatest causes of failure in biomarker validation studies.¹¹ Bias can enter a study during patient selection, specimen collection, specimen analysis, and patient evaluation. Randomization and blinding are two of the most important tools for avoiding bias. Randomization in biomarker discovery should be carried out to control for nonbiological experimental effects owing to changes in reagents, technicians, machine drift, etc., that can result in batch effects.¹² Specimens from controls and cases should be assigned to arrays, testing plates or batches by random assignment, ensuring the distributions of cases, controls, and age of specimen are equally distributed.¹³ Blinding can be carried out by keeping the individuals who generate the biomarker data from knowing the clinical outcomes; it prevents the bias induced by unequal assessment of biomarker result.¹⁴ Randomization and blinding should be used in the process of biomarker data generation and should be incorporated at every stage of a study when possible.

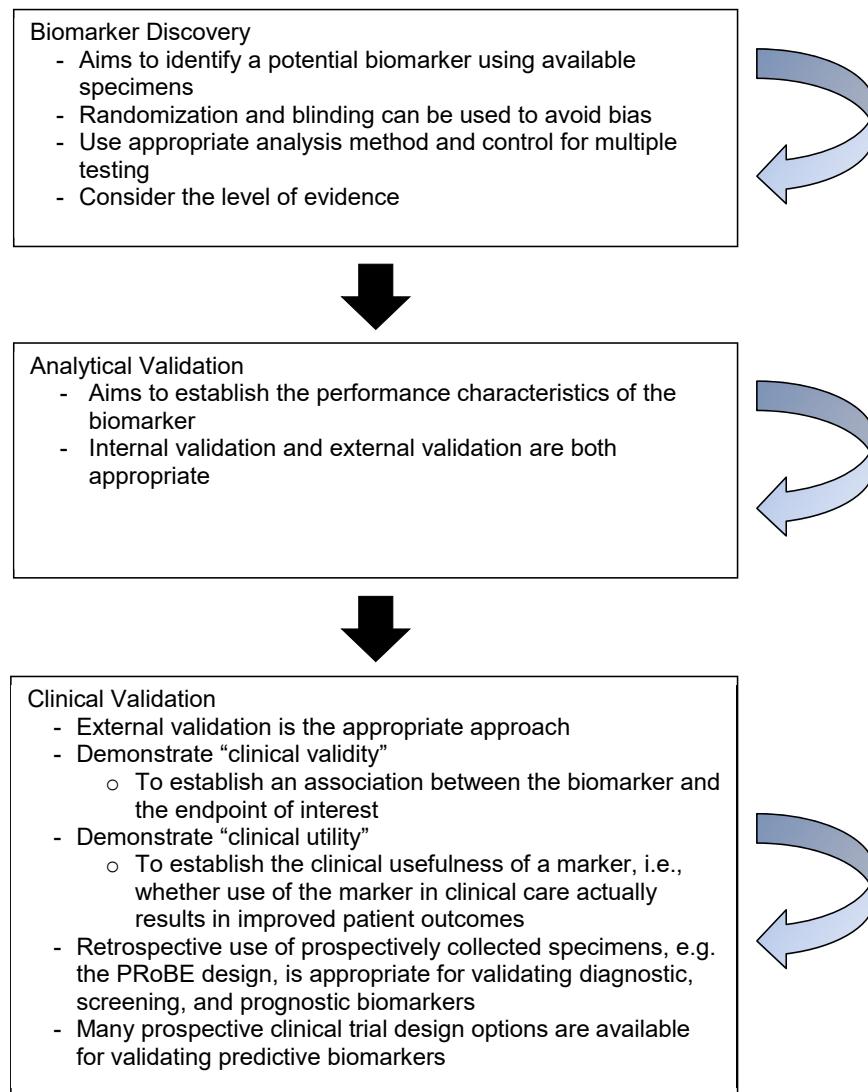


Figure 2. Simplified schematic of biomarker development. PRoBE: prospective-specimen-collection, retrospective-blinded-evaluation.

Prognostic and Predictive Biomarker Identification

A prognostic biomarker can be identified in properly conducted retrospective studies that do not rely solely on convenience samples but use biospecimens prospectively collected from a cohort that represents the target screening population, case-control studies, and single-arm trials. A prognostic biomarker is identified through a main effect test of association between the biomarker and the outcome in a statistical model. An example of a prognostic biomarker is the *STK11* mutation that is associated with poorer outcome in nonsquamous NSCLC.¹⁵ Tissue samples were collected from a consecutive series of patients with nonsquamous NSCLC who underwent curative-intent surgical resection in 2001 to 2006 at two hospitals. An a priori power calculation was performed to ensure a sufficient number of overall

survival events to provide adequate statistical power to assess five candidate biomarkers. Even though convenience samples were used, the prognostic effect was validated in two external datasets which strengthened the validity of the discovery.

A predictive biomarker needs to be identified in secondary analyses using data from a randomized clinical trial, through an interaction test between the treatment and the biomarker in a statistical model. Secondary analyses refer to subsequent correlative studies that may or may not be predefined as a protocol objective. An example of predictive biomarker identification is the IPASS study.¹⁶ The IPASS study enrolled patients with advanced pulmonary adenocarcinoma who were non-smokers or former light smokers and randomly assigned patients to receive gefitinib or carboplatin plus paclitaxel (CP). Patients' *EGFR* mutation status was not known at

the time of enrollment and was determined retrospectively. The interaction between treatment and *EGFR* mutation was statistically significant ($p < 0.001$) and indicated that among patients who have EGFR-mutated tumors, progression-free survival (PFS) was significantly longer (hazard ratio = 0.48; 95% confidence interval [CI]: 0.36–0.64) for those receiving gefitinib compared with those receiving CP. In contrast, among patients who have EGFR wildtype tumors, PFS was significantly shorter (hazard ratio = 2.85; 95% CI: 2.05–3.98) for those receiving gefitinib compared with those receiving CP.¹⁶

Analytical Methods

Analytical methods should be chosen to address study specific goals and hypotheses. Data-driven analyses and the resulting findings are less likely to be reproducible in an independent set of data. Thus, the analytical plan should be written and agreed on by all members of the research team before receiving data to avoid the data influencing an analysis. This includes defining the outcomes of interest, hypotheses that will be tested, and criteria for success. Control of multiple comparisons should be implemented when multiple biomarkers are evaluated; a measure of false discovery rate is especially useful when using large-scale genomic or other high dimensional data for biomarker discovery.¹⁷ During biomarker discovery, evaluation of associations between a biomarker and disease status, demographic or clinical characteristics, such as age, sex, and body mass index, or in diseased patients, stage or other disease characteristics, can inform design of future validation studies. Metrics useful for evaluating biomarkers (Table 1) include differences between groups, sensitivity, specificity, positive and negative predictive values, discrimination (i.e., receiver operating characteristic area under the curve), calibration, and clinical validity and use.^{10,18–21} The appropriate metric depends on the study goals and should be determined by a study

team, including clinicians, scientists, statisticians, and epidemiologists.

It is often the case that information from a panel of multiple biomarkers will be required to achieve better performance than a single biomarker, despite the added potential measurement errors that come from multiple assays. Using each biomarker in its continuous state instead of a dichotomized version retains maximal information for model development, and in turn, greater improvement in panel performance; dichotomization for clinical decision making is best left for later studies. The optimal analytical strategy for combining multiple biomarkers and for choosing which biomarkers to combine depends on both sample size and clinical context. Incorporation of some form of variable selection, such as shrinkage, during model estimation generally minimizes overfitting and maximizes the likelihood of validation; hundreds to thousands of patients are generally required to incorporate nonlinear functions, such as interactions, smoothing splines, or machine learning and artificial intelligence algorithms, without overfitting. It is useful to generate pilot data for use in simulations to inform sample size calculations and plan the appropriate analytical strategy.^{18–20,22,23}

Missing data can lead to biased results. Thus, the analysis plan should include an approach to handle missing data, including assessment of the mechanism responsible for the missingness and an approach to handle the missingness that minimizes potential biases from being introduced into an analysis.²⁴

The EQUATOR network assembles an important collection of guidelines for the design and reporting of diagnostic and prognostic modeling studies (<https://www.equator-network.org/>).

Biomarker Validation

Validation is “a process to establish that the performance of a test, tool, or instrument is acceptable for its intended purpose.”¹ Internal validation establishes a

Table 1. Metrics Useful for Evaluating Biomarker Performance

Metrics	Description
Sensitivity	The proportion of cases that test positive
Specificity	The proportion of controls that test negative
Positive predictive value	Proportion of test-positive patients who actually have the disease; is a function of disease prevalence
Negative predictive value	Proportion of test-negative patients who truly do not have the disease; is a function of disease prevalence
ROC curve	Plot of sensitivity (true positive rate) versus 1 specificity (false-positive rate), with a data point calculated for every value of the marker in the data set
Discrimination	How well the marker distinguishes cases from controls; often measured by the area under the ROC curve; ranges from 0 to 1, with 0.5 indicating performance equivalent to a coin flip and 1 corresponds to perfect ability to distinguish
Calibration	How well a marker estimates the risk of disease or of the event of interest

ROC, receiver operating characteristic.

biomarker's performance in the data in which the biomarker was developed and should be assessed by means of resampling methods, such as bootstrapping or cross-validation, to provide realistic expectations.¹⁸ External validation establishes a biomarker's performance in a completely independent data set not used during development; it must be established using data from different time frames, institutions, or geographic regions which we discuss in subsequent paragraphs. Analytical validation and clinical validation are two distinct aspects of biomarker validation. Use of specimens collected prospectively from the target population before knowing patient outcomes is a critical design feature of all validation studies which minimizes the influence of bias.

Analytical Validation

Analytical validation aims to establish the performance characteristics of a biomarker including sensitivity, specificity, accuracy, precision, interlaboratory reproducibility, and other relevant performance characteristics following a prespecified protocol. The statistical analysis methods used for analytical validation are similar to the methods mentioned in biomarker discovery (Table 1). The goal of analytical validation is to reveal a biomarker's technical performance (i.e., the biomarker will provide consistent measurements to the unknown true values) and not its usefulness.

Clinical Validation

Clinical validation aims to establish an association between the biomarker and the end point of interest (i.e., clinical validity per Teutsch et al.¹⁰) and to reveal the usefulness of the biomarker (i.e., clinical use per Teutsch et al.¹⁰). Clinical validation relies on external validation and can be done by retrospective use of clinical trial data or by prospective clinical trials. Retrospective use of clinical trial data is a form of external clinical validation in which the biomarker evaluation is not part of the original study design.

Establishing clinical utility or usefulness generally requires a prospective clinical trial, a form of external validation, to reveal that use of the biomarker to guide patient care translates into improved health outcomes. An example is the approval of pembrolizumab as the first tissue-agnostic approval granted by the United States Food and Drug Administration (FDA).²⁵ Patients with microsatellite instability-high (MSI-H) tumors treated with pembrolizumab had higher overall response rates compared with those with microsatellite stable tumors regardless of the tumor origin in the KEYNOTE-016 study. The regulatory approval was based on data from five different trials ($N = 149$) in which patients with

MSI-H were retrospectively identified from two prospective studies ($N = 14$) and prospectively identified from three studies ($N = 135$). The objective response rate was 39.6% (7% with complete response) among 149 patients with MSI-H tumor consisting of 15 different tumor types which was considered clinically meaningful (compared with an objective response rate of 0% among patients with colorectal cancer with microsatellite stable tumors in KEYNOTE-016²⁶). At the time of the approval, no companion in vitro diagnostic device was available. Patients were enrolled predominantly on the basis of PCR-based tests for MSI-H and immunohistochemistry-based tests for deficient mismatch repair available in the community as laboratory-developed tests. The FDA determined that the risk to patients with "false positive" tumors is low in this setting and, given the efficacy observed, FDA approved for this use.²⁷ There was commitment from Merck to develop a companion diagnostic test for detection of MSI-H and deficient mismatch repair across all cancers postmarketing.

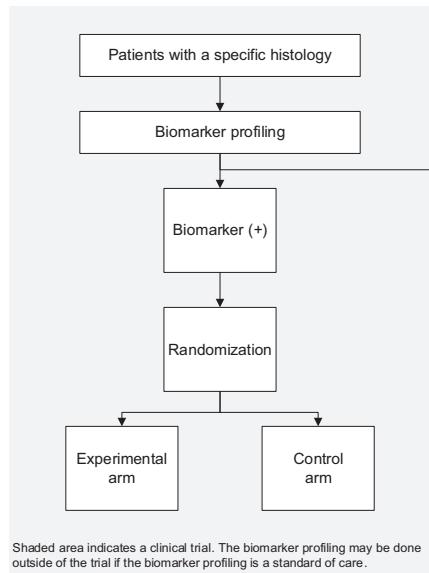
Study Designs for Biomarker Validation

Though costly, biomarker evaluation efforts are enhanced by biobanks of specimens collected prospectively from an observational cohort that represents the target population intended for the biomarker.²⁸ A prospective-specimen-collection, retrospective-blinded-evaluation design²⁹ can be performed in such a setting to validate screening, diagnostic, and prognostic biomarkers. Specimens and clinical data are collected without knowing the patient outcome. Case patients and control patients would be randomly selected on the basis of their outcome status. The biomarker data are then generated for the patients selected, blinded to clinical and outcome information. An example of such design is the MILD study.³⁰ The MILD trial, a randomized prospective clinical trial, enrolled 4099 current or former smokers without history of cancer and randomized them to low-dose computed tomography versus observation. Whole blood was collected at enrollment and subsequent follow-up. Retrospectively, 1000 consecutive plasma samples collected from June 2009 to July 2010 among lung cancer-free individuals enrolled onto the trial were used for validation of a microRNA signature classifier. The classifier was pre-specified with predefined cut points, and risk scores were generated blinded to clinical outcome for individual participants and submitted to an independent research center. Data analysis was completed according to a prespecified statistical analysis plan by the independent research center. This validation study intentionally used the full cohort rather than a random subset of patients to maximize the study power.

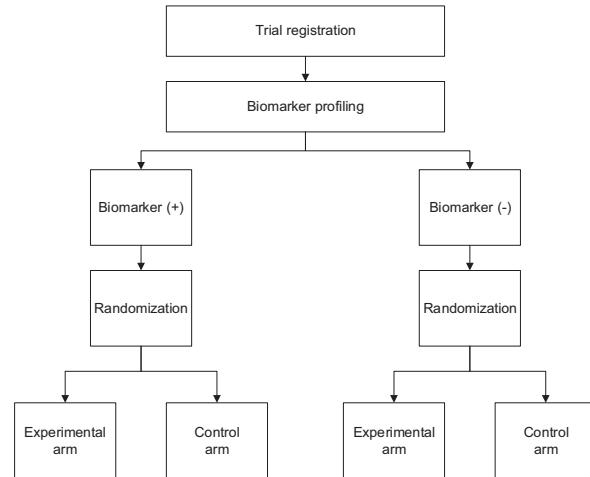
There are several prospective clinical trial designs aimed to validate the clinical use of a predictive biomarker in a clinical setting. Enrichment designs screen all patients for the biomarker but only enroll and randomize those with the desired molecular features. A treatment will be evaluated within the biomarker-defined subgroup only. Enrichment designs are advantageous

when the biomarker prevalence is low (<15%–20%). An example of such a design is the EURTAC trial³¹ which led to the FDA's approval of erlotinib for the first-line treatment of patients with metastatic NSCLC harboring EGFR mutations. The EURTAC trial screened 1227 patients and then randomized 174 patients with EGFR mutations to receive erlotinib or standard chemotherapy (Fig. 3A).

A Enrichment design



B All-comer (stratified by biomarker status) designs



C Subgroup design

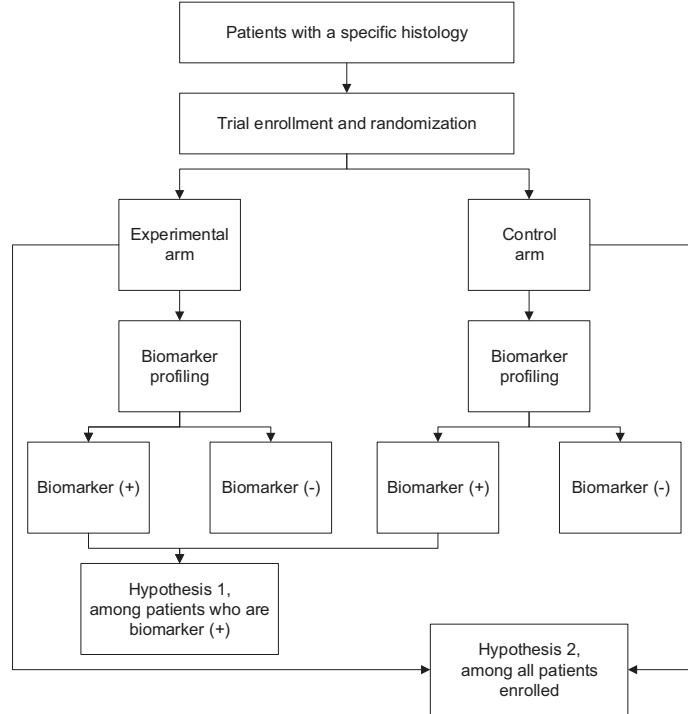


Figure 3. Trial design schema. (A) Enrichment design. (B) All-comer (stratified by biomarker status) design. (C) Subgroup design.

All-comer (stratified by biomarker status) designs screen all patients for the biomarker and then enroll and randomize patients with a valid biomarker result. The randomization can be stratified by the biomarker status (if the turnaround time of biomarker testing is short), and the test of treatment by biomarker interaction is included in the prespecified analysis plans. All-comer designs are appropriate when the treatment benefit needs to be better understood in both patients who test positive and in those who test negative. An example of such a design is the MARVEL trial (N0723, NCT00738881). The MARVEL trial planned to enroll 1196 patients with advanced NSCLC after first-line therapy and patients' EGFR expression by means of fluorescence in situ hybridization (FISH) was evaluated by central pathology review. After the FISH result was available, patients were randomized to receive pemetrexed versus erlotinib, stratified by the FISH status and other factors. The goal was to identify 287 FISH-positive patients and 670 FISH-negative patients (70%) to evaluate whether there are differences in PFS owing to treatment with erlotinib compared with pemetrexed for subsets defined by FISH positivity versus negativity (Fig. 3B).

Subgroup designs validate a predictive biomarker in a specific subgroup of patients and in the overall population using a multiple-hypothesis design.³² In this design, all patients with a particular disease are randomized to experimental therapy versus standard of care, but coprimary objectives are defined to test the superiority of the experimental therapy in the subgroup of patients selected by the biomarker, and for all enrolled patients. This design is advantageous when there is evidence that the experimental therapy will be most effective in patients with the biomarker of interest, but could also have a broad effect in the general disease population. An example is Southwest Oncology Group S0819,³³ which was designed to test the hypothesis that EGFR amplification can identify patients most likely to benefit from EGFR antibodies in combination with chemotherapy in patients with advanced NSCLC. S0819 randomized 1313 eligible patients to chemotherapy with cetuximab versus chemotherapy alone. EGFR-FISH status was not required to be known at trial enrollment and was evaluated at each interim analysis. Coprimary end points were PFS in patients with EGFR-FISH-positive cancer and overall survival in the entire population (Fig. 3C).

Platform-type trial designs, such as umbrella trials (histology specific) and basket trials (biomarker specific and agnostic to histology), can be advantageous in biomarker validation as well.³⁴

There are common features for establishing analytical validity, clinical validity, and clinical use,

that is, there should be a prespecified protocol dealing with the specifics of the validation process, such as specimen collection, specimen handling and storage procedures, biomarker and clinical outcomes of interest, the purposes of the biomarker, and the potential benefits and risks associated with the use of the biomarker.

Conclusions

In this article, we discussed the statistical perspectives on the best practices for biomarker discovery and validation. One aspect that we omitted was the biomarker qualification process with the regulatory agencies.⁷ Readers should note that the FDA requires biomarker candidates to undergo clinical validation and be assessed as a companion diagnostic before receiving regulatory approval. The biomarkers used to direct therapies need to be generated by an assay that is performed in a Clinical Laboratory Improvement Amendments-certified laboratory, which will be the first step toward clinical validation. We encourage investigators to reach out to health authorities early to discuss potential biomarkers of interest.

We would also like to take this opportunity to urge oncologists to resist the temptation of adopting unvalidated biomarker findings into practice. Attempts to discover biomarkers have accelerated through advanced technology in generating relevant data. The potential biomarkers discovered should be considered as hypothesis generating, and the biomarkers need to be validated (both analytically and clinically) before adoption. An example would be the *STK11* and *KEAP1* mutations that seemed to be predictive with emerging data revealing patients with *STK11* and *KEAP1* mutations do not respond to immunotherapy. However, an exploratory analysis using clinical trial data revealed that pembrolizumab monotherapy was associated with improved overall response rates compared with chemotherapy regardless of *STK11* and *KEAP1* mutational status, that is, these mutations were prognostic.³⁵ In addition, an analysis using real-world evidence also revealed that *STK11* and *KEAP1* mutations are prognostic biomarkers and unlikely to be predictive biomarkers for anti-programmed cell death protein-1 and anti-programmed death-ligand 1 therapy.³⁶ *STK11* and *KEAP1* remain unvalidated predictive biomarkers, and clinicians' treatment decisions should not be swayed by the mutation status of these two genes.

The discovery and validation of biomarkers require thorough planning and the collaboration of clinicians, scientists, statisticians, and epidemiologists. The success of these endeavors requires collaborative and cross-

disciplinary approaches. A cohesive and an effective team of collaborative scientists is crucial for biomarker development, and we promote such partnerships to ultimately accelerate the translation of cutting-edge scientific discoveries from bench to bedside thus leading to improved patient care and outcomes.

Acknowledgments

This work was partially supported by the National Institutes of Health (NIH) grant P30CA15083 (Mayo Comprehensive Cancer Center grant; Drs. Ou, Oberg, and Adjei), National Cancer Institute (NCI) grant P50CA136393 (Mayo Clinic Specialized Programs of Research Excellence in Ovarian Cancer grant; Dr. Oberg), NCI grant P50CA102701 (Mayo Clinic Specialized Programs of Research Excellence in Pancreatic Cancer; Dr. Oberg), NCI grant U10CA180882 (Alliance Statistics and Data Management Center; Drs. Ou and Oberg), NIH grant U24CA213274 (Dr. Shyr), NIH grant U54TR002243 (Dr. Shyr), and NIH grant P30CA068485 (Dr. Shyr).

References

1. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and Other Tools). Resource*. Silver Spring, MD: Food and Drug Administration; 2016.
2. National Comprehensive Cancer Network. Lung cancer screening. 2020 version. <https://www.nccn.org/patients/guidelines/content/PDF/lung-early-stage-patient.pdf>. Accessed January 4, 2021.
3. Nicholson AG, Sauter JL, Nowak AK, et al. EURACAN/IASLC proposals for updating the histologic classification of pleural mesothelioma: towards a more multidisciplinary approach. *J Thorac Oncol*. 2020;15:29-49.
4. Remon J, Ahn MJ, Girard N, et al. Advanced-stage non-small cell lung cancer: advances in thoracic oncology 2018. *J Thorac Oncol*. 2019;14:1134-1155.
5. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93:1054-1061.
6. Institute of Medicine (US). Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease. In: Micheel CM, Ball JR, eds. *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. Washington, DC: National Academies Press (US); 2010.
7. Food and Drug Administration. Biomarker qualification: evidentiary framework guidance for industry and FDA staff. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/biomarker-qualification-evidentiary-framework>. Accessed January 4, 2021.
8. Rudin M. Imaging readouts as biomarkers or surrogate parameters for the assessment of therapeutic interventions. *Eur Radiol*. 2007;17:2441-2457.
9. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101:1446-1452.
10. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP working group. *Genet Med*. 2009;11:3-14.
11. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. *J Clin Oncol*. 2010;28:698-704.
12. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11:733-739.
13. Qin LX, Zhou Q, Bogomolny F, et al. Blocking and randomization to improve molecular biomarker discovery. *Clin Cancer Res*. 2014;20:3371-3378.
14. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer*. 2005;5:142-149.
15. Pécuchet N, Laurent-Puig P, Mansuet-Lupo A, et al. Different prognostic impact of STK11 mutations in non-squamous non-small-cell lung cancer. *Oncotarget*. 2017;8:23831-23840.
16. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*. 2009;361:947-957.
17. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100:9440-9445.
18. Harrell FE Jr. *Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. Berlin, Switzerland: Springer; 2015.
19. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. New York, NY: Springer International Publishing; 2019.
20. Pepe M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press; 2003.
21. Mahar AL, Compton C, McShane LM, et al. Refining prognosis in lung cancer: a report on the quality and relevance of clinical prognostic tools. *J Thorac Oncol*. 2015;10:1576-1589.
22. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer Science & Business Media; 2009.
23. James G, Witten D, Hastie T, et al. *An Introduction to Statistical Learning: With Applications in R*. New York, NY: Springer; 2013.
24. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*. Chichester, United Kingdom: John Wiley & Sons; 2019.
25. Boyiadzis MM, Kirkwood JM, Marshall JL, Pritchard CC, Azad NS, Gulley JL. Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *J Immunother Cancer*. 2018;6:35.
26. Le DT, Uram JN, Wang H, et al. Programmed death-1 blockade in mismatch repair deficient colorectal cancer. *J Clin Oncol*. 2016;34(suppl 15):103-103.
27. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA approval summary: pembrolizumab for the treatment of microsatellite instability-high solid tumors. *Clin Cancer Res*. 2019;25:3753-3758.
28. Andre F, McShane LM, Michiels S, et al. Biomarker studies: a call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol*. 2011;8:171-176.

29. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst.* 2008;100:1432-1438.
30. Sozzi G, Boeri M, Rossi M, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study [published correction appears in *J Clin Oncol.* 2014;32:1520]. *J Clin Oncol.* 2014;32:768-773.
31. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* 2012;13:239-246.
32. Hoering A, Leblanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res.* 2008;14:4358-4367.
33. Herbst RS, Redman MW, Kim ES, et al. Cetuximab plus carboplatin and paclitaxel with or without bevacizumab versus carboplatin and paclitaxel with or without bevacizumab in advanced NSCLC (SWOG S0819): a randomised, phase 3 study. *Lancet Oncol.* 2018;19:101-114.
34. Ou FS, An MW, Ruppert AS, Mandrekar SJ. Discussion of trial designs for biomarker identification and validation through the use of case studies. *JCO Precis Oncol.* 2019;3:PO.19.00051.
35. Collingridge D. The Lancet Oncology. Paper presented at: ASCO Virtual Annual Meeting. June 5, 2020.
36. Papillon-Cavanagh S, Doshi P, Dobrin R, Szustakowski J, Walsh AM. STK11 and KEAP1 mutations as prognostic biomarkers in an observational real-world lung adenocarcinoma cohort. *ESMO Open.* 2020;5:e000706.