

Evidence-Based Development and Clinical Use of Precision Oncology Therapeutics

Leah A. Comment¹ , Ashley F. Ward¹ , Alexa B. Schrock¹, David Fabrizio¹, Jeffrey M. Venstrom¹, Priti S. Hegde¹ and Brian M. Alexander^{1,*} 

Precision medicine refers to the “tailoring of medical treatment to the individual characteristics of each patient.”¹ Because cancer is a disease of the genome,² molecular profiling often informs diagnosis, prognosis, and therapeutic decision making. Biomarker development generates actionable, reliable information for therapeutic development and clinical decision making. However, therapeutic development only requires biomarker to identify populations where a drug will be effective. Companion diagnostic development through categorical biomarker definitions may, therefore, not be ideal for biomarker development.

BIOMARKER CLASSIFICATION

Precision oncology relies on the use of biomarkers to guide clinical decision making. Biomarkers can be helpful in diagnosing a disease, outlining prognosis, predicting response to a therapy, or for monitoring response and outcomes. As cancer is a disease of the genome, genomic profiling can play a critical role as a diagnostic biomarker. For example, diagnostic criteria for some tumors now combine molecular information with standard histopathologic criteria. Prognostic biomarkers are not only useful for counselling patients but can also be helpful for clinical decision making. Decisions on adjuvant therapy, for example, are based primarily on risk of recurrence. A therapy with the

same proportional risk reduction confers more clinical benefit by reducing the absolute risk of recurrence from 50% to 20% than from 5% to 2%, for example. Post-treatment biomarkers may also be useful in assessing response and monitoring disease status. Most often, however, biomarkers in support precision medicine are predictive biomarkers, and economic incentives may direct resources towards the development of predictive biomarkers in parallel with therapeutic development.³

PREDICTIVE BIOMARKER DEVELOPMENT: COMPANION DIAGNOSTICS AND BEYOND

Predictive biomarkers are “used to identify individuals who are more likely than

similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent.”⁴ Assays for these biomarkers are frequently established as companion diagnostics as part of the drug development process. Companion diagnostics are devices deemed essential for the safe and effective use of therapeutic products, defining the patient population for whom the benefits of a drug have been demonstrated to outweigh the risks. There are some advantages to developing predictive biomarkers as part of the drug development process. First, development as a companion diagnostic leverages established US Food and Drug Administration (FDA) regulatory pathways, setting a high bar for assay validity and precision, providing a quality benchmark that may be critical for clinical decision making. Second, the process for generating clinical evidence is very expensive; reimbursement for diagnostic tests may not drive the required investment in large clinical trials for each biomarker. Companion diagnostic development allows biomarker evidence generation to be “subsidized” by drug development. However, companion diagnostic development and predictive biomarker development are not the same. In many cases, companion diagnostic development does not provide all of the evidence needed for clinical decision making.

The dotted boxes in **Figure 1** exemplify a population that might be studied to support therapeutic and companion diagnostic development. Despite defining a population that might benefit, conclusive biomarker utility requires knowing the response in the biomarker-negative population. Differential effect is important evidence for decision making, but such evidence is often not generated as part of companion diagnostic development. The

¹Foundation Medicine, Inc., Cambridge, Massachusetts, USA. *Correspondence: Brian M. Alexander (balexander@foundationmedicine.com)

Received May 3, 2020; accepted June 16, 2020. doi:10.1002/cpt.1967

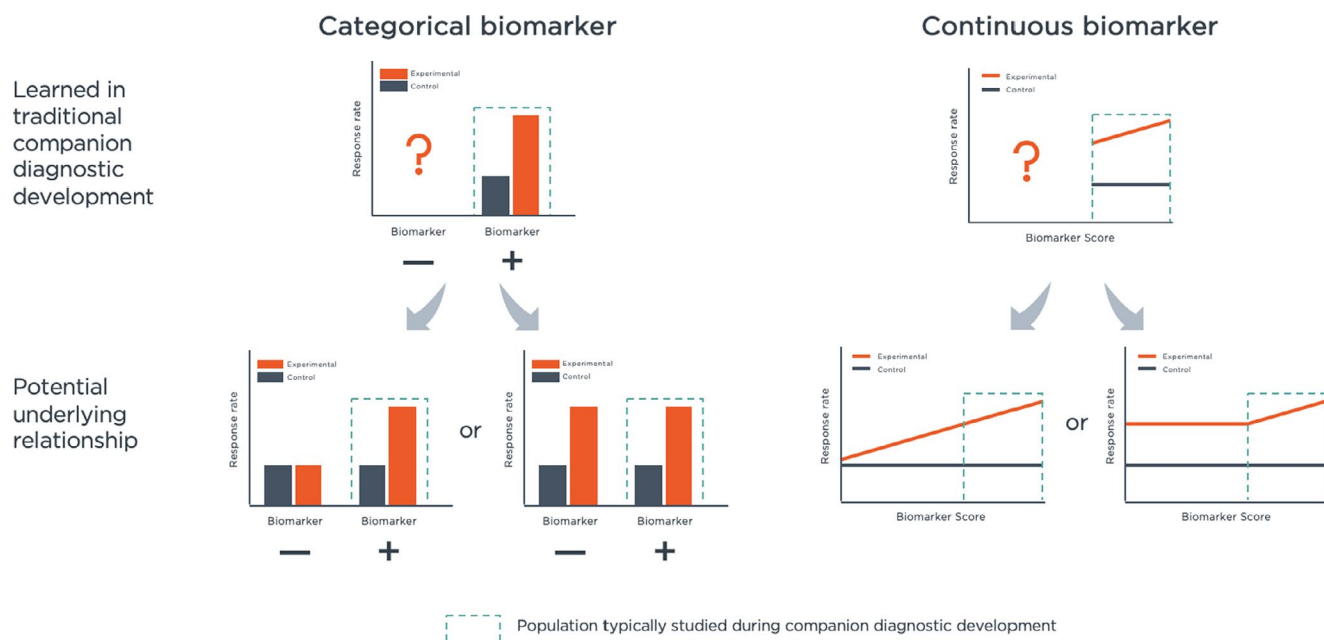


Figure 1 Comparison of evidence generation during companion diagnostic development for novel predictive biomarkers, where a traditional approach (top row) focuses exclusively on characterizing the therapeutic effect among a subpopulation classified as biomarker positive. The assumption that the true therapeutic effect is negligible in the biomarker-negative population can be based on strong biological rationale (left column, bottom left), as with oncology drugs targeting driver mutations. For predictive biomarkers with a truly continuous relationship (right column), the traditional companion diagnostic approach leads to important gaps in evidence for clinical decision making, mainly due to substantial uncertainty about the risks and benefits for patients below the cutoff.

underlying truth could be that the biomarker-negative group still benefits from the drug (Figure 1).

Categorical vs. continuous predictive biomarkers

Categorical biomarkers define discrete populations. Sometimes, categorical biomarkers are defined by the underlying biology. Blood type is an example. Other times, categorical biomarkers are defined from continuous biomarker “cut points.” In these situations, transformation to categories is often based on limited evidence, and information is always lost. For example, age is a continuous biomarker, whereas “young” or “old” based on a cutoff of 65 years is a categorical transformation. Although blood types A vs. B may map to relatively deterministic predictions of transfusion response, age 64 vs. 66 likely provides much less information for clinical decision making. Importantly, this fact does not invalidate the overall utility of age as a biomarker, in that a comparison of age 18 to age 66 could result in deterministic predictions akin to blood types.

In precision oncology, a historical abundance of biologically-defined categorical

biomarkers has obscured key differences between companion diagnostic and predictive biomarker development processes. Examples include EGFR inhibitors for non-small cell lung cancer with exon 19 deletions or exon 21 L858R substitution mutations, alpelisib for PIK3CA mutated breast cancer, PARP inhibitors for homologous recombination pathways mutated prostate or ovarian cancer, and TRK inhibitors for NTRK fusion-positive solid tumors.⁵ When mutations drive tumor growth and progression, targeting these mutations often results in a specific, robust response. Biomarker assays to detect driver mutations are therefore typically developed as companion diagnostics. The predictive aspect of the biomarker is often implicit (i.e., cancers without these driver mutations would not be expected to respond to a drug targeting the driver mutation), and biomarker-negative populations are often not studied clinically.⁶ When the biomarker-negative population is clearly identified and there is a strong biological rationale, development of a companion diagnostic may be sufficiently informative for clinical decision making. Furthermore, if mutations

are mutually exclusive, biomarker panels can directly support clinical decision making heuristics through if-then algorithms.

Despite possible benefits of heuristics and simplified decision tools, some biological processes—particularly complex, multifactorial ones—are better characterized by continuous quantities. However, the interpretation of continuous predictive biomarkers is difficult. In these cases, a marginal increase in the measurement or score is associated with a marginal change in the probability of experiencing therapeutic benefit. For such biomarkers, there is no “right” place to set a cutoff, and no biologic determinism to exploit. Decision making becomes more probabilistic, and the tradeoffs more situation specific. For example, trial eligibility criteria are complicated without biologic determinism to dictate a clear cutoff. Defining an inclusive cut point requires large and lengthy trials to demonstrate treatment effect in the overall population. Defining a restrictive one (i.e., with only high-probability responders) reduces the size and length of the trial, but the population eligible to receive the drug in clinical practice excludes patients who could benefit. Furthermore,

when the trial population is defined by a cutoff applied to a continuous biomarker, as in **Figure 1a**, the evidence necessary to obtain drug approval often does not support evidence-based decision making for patients falling just beyond the cutoff. Patients with biomarker scores just above and just below the cutoff could have indistinguishable clinical courses in response to a drug, but the cutoff sets them on very different clinical paths.

Predicting effects of immune checkpoint inhibitors

Biomarkers predicting response to immune checkpoint inhibitors (ICPIs) are not as straightforward as the biologically defined categorical biomarkers for tyrosine kinase inhibitor response. For example, programmed death ligand 1 (PDL1) immunohistochemistry assays are critical ICPI biomarkers. However, although the true relationship between PDL1 immunohistochemical scores and response probability may be continuous, cutoffs used in clinical trials define specific subgroups. The 22C3 assay cutoffs of 1% tumor proportion score (TPS) in recurrent non-small cell lung cancer and 50% TPS in newly diagnosed non-small cell lung cancer were based on data from the Keynote-10⁷ and Keynote-24 trials, respectively. In Keynote-10, however, the subset of patients with TPS \geq 50% had a better hazard ratio for progression-free survival and overall survival and a higher overall response rate in response to pembrolizumab, than the broader group defined by TPS \geq 1%, suggesting a relationship with more gradation. Similarly, Aguilar *et al.* showed in a multi-institutional retrospective analysis that patients with PDL1 expression in $>$ 90% of tumor cells had better outcomes compared to patients with expression in 50–89% of cells.⁸ Thus, there is likely more information in the continuous analysis of PDL1 that could be brought to bear on clinical decision making. Such decisions would use probabilistic reasoning to assess tradeoffs. For example, a low probability of response may be more acceptable in clinical scenarios with few alternative options, whereas the bar would be higher when selecting among several promising therapies or combinations. Furthermore, erroneous decision making based on threshold effects would be avoided—TPS of 49% and

that of 51% are unlikely to have different outcomes.

Likewise, a continuous biomarker approach may be appropriate for tumor mutational burden (TMB), a surrogate measure for neoantigenicity that has been hypothesized to predict ICPI treatment effect. In theory, greater numbers of foreign-appearing antigens will trigger greater immune system recognition and response. Unlike with driver mutations and target therapies, the biological rationale here does not suggest the existence of natural “cutoffs.” Retrospective analyses support this hypothesis. In one large study of patients treated with ICPI across a variety of tumor types, continuous TMB was positively associated with survival without a specific cutoff identifying “responders” and “nonresponders.”⁹ Still, most clinical development of biomarkers for ICPI focuses on defining the right “cutoff” for therapeutic effect. When the true relationship exhibits gradation, defining a high cutoff increases the power of a trial for a given sample size. But the resulting label indication is unnecessarily restrictive, potentially denying access to patients who would experience substantial therapeutic benefit. Conversely, defining a low cutoff requires larger and longer trials, may dilute the observed treatment effect, and yet there could still be benefit for patients below the cutoff. The FDA complementary diagnostics designation, for biomarkers that aid in benefit-risk decision making, offers a potential regulatory pathway to more fully develop biomarker information to support precision medicine, and has been used for PDL1 immunohistochemistry for ICPI. However, there is no requirement to develop biomarker assays through this pathway, no requirement for using the biomarker in clinical use and comprehensive information about the biomarker/effect relationship is not necessary nor is guidance on use provided. Developing a true predictor of ICPI response is further complicated by the observation TMB and PDL1 are not correlated,¹⁰ contributing independent predictive utility. A future “immunoscore” will likely have many individual biomarker inputs, including an assessment of nontumor factors, such as immune infiltrates and host immune factors, potentially creating tradeoffs between efficacy and interpretability.

CLINICAL TRIALS FOR EVIDENCE GENERATION TO SUPPORT PRECISION MEDICINE

Optimal biomarker evidence for precision oncology requires treatment effect estimation across the entire range of biomarker values. For categorical biomarkers with strong rationale indicating a biomarker-specific effect, current approaches may suffice. In particular, biomarkers for driver mutations and associated inhibitors are generally served well by companion diagnostic development and trials conducted exclusively in biomarker-positive patients. Outside of this biomarker-positive population, the hypothesized mechanism of action for the drug dictates a null or negative treatment effect. On the other hand, more robust evidence is needed to guide decision making with complex biomarkers, such as continuous variables or composite scores. For these, an ideal body of evidence would result from all patients enrolling in randomized studies, with biomarker associations learned through preplanned analyses rather than *a priori* restrictions with artificial categories. This might introduce considerable regulatory risk, however, as “negative” results in the overall population may lead to rejection of promising therapies. Randomize-all designs are also inefficient. However, novel designs that incorporate adaptive components could be used as a compromise between trial efficiency and robust data collection.^{6,11} Trials could start with randomizing all patients but become more biomarker-specific as the evidence supports an association, as in I-SPY2, GBM AGILE, and INSIGHt.¹¹ In addition, trials could start as biomarker-specific based on robust preclinical or prior trial data and then explore a broader range of patients as efficacy is demonstrated.⁶ Ultimately, innovation in regulatory frameworks and clinical trial design could generate better evidence for decision making in precision oncology.

FUNDING

Any funding related to this manuscript was provided by Foundation Medicine, Inc.

CONFLICT OF INTEREST

All authors are employees of Foundation Medicine, Inc., a subsidiary of F. Hoffmann-La Roche AG and receive personal fees from Foundation Medicine and hold equity in Roche. B.M.A. has received research funding from Eli Lilly, Puma, and Celgene.

© 2020 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. (National Academies Press, Washington, DC, 2011).
2. Collins, F.C. A disease of the genome. *Cancer Res.* **67**, PL01 (2007).
3. Stern, A.D., Alexander, B.M. & Chandra, A. How economics can shape precision medicines. *Science* **355**, 1131–1133 (2017).
4. US Food and Drug Administration-National Institutes of Health FDA-NIH Biomarker Working Group. BEST (Biomarkers, Endpoints, and other Tools) Resource. (Food and Drug Administration (US)/National Institutes of Health (US), Bethesda, MD, 2016) <<http://www.ncbi.nlm.nih.gov/books/NBK326791/>>.
5. Foundation One CDx label <<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P170019>>.
6. Trippa, L. & Alexander, B.M. Bayesian baskets: a novel design for biomarker-based clinical trials. *J. Clin. Oncol.* **35**, 681–687 (2017).
7. Herbst, R.S. *et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* **387**, 1540–1550 (2016).
8. Aguilar, E.J. *et al.* Outcomes to first-line pembrolizumab in patients with non-small-cell lung cancer and very high PD-L1 expression. *Ann. Oncol.* **30**, 1653–1659 (2019).
9. Samstein, R.M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206.
10. Yarchoan, M. *et al.* PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI Insight* **4**, e126908 (2019).
11. The Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat. Rev. Drug Discov.* **18**, 797–807 (2019).