REGULATORY SCIENCE

What evidence do we need for biomarker qualification?

Chris Leptak, ¹* Joseph P. Menetski, ²* John A. Wagner, ³* Jiri Aubrecht, ⁴ Linda Brady, ⁵ Martha Brumfield, ⁶ William W. Chin, ⁷ Steve Hoffmann, ² Gary Kelloff, ⁸ Gabriela Lavezzari, ⁹ Rajesh Ranganathan, ¹⁰ John-Michael Sauer, ⁶ Frank D. Sistare, ¹¹ Tanja Zabka, ¹² David Wholley ^{2†}

Biomarkers can facilitate all aspects of the drug development process. However, biomarker qualification—the use of a biomarker that is accepted by the U.S. Food and Drug Administration—needs a clear, predictable process. We describe a multistakeholder effort including government, industry, and academia that proposes a framework for defining the amount of evidence needed for biomarker qualification. This framework is intended for broad applications across multiple biomarker categories and uses.

Copyright © 2017
The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S.

INTRODUCTION

Drug development is a time- and cost-intensive endeavor (1), and any efficiency that can be realized during the development and regulatory processes will speed access of approved therapies to patients. The expected positive impact of biomarkers on drug development is substantial, and coordinated efforts to identify biomarkers are a focus of much research and discussion (2-5). A biomarker is defined as a characteristic that is measured as an indicator of a normal biological process, a pathogenic process, or a response to an exposure or intervention, including a therapeutic intervention (2). A critical process is biomarker qualification, which is a conclusion based on a formal regulatory procedure that the biomarker, within a stated context of use (COU), can be relied upon to have a specific interpretation and application in medical product development and regulatory review (2). It has been challenging to define a predictable process to qualify biomarkers, which requires clear direction regarding considerations of the type and level of evidence needed. Here, we discuss the type and amount of evidence needed to support the qualification of biomarkers for regulatory use within the Center for Drug Evaluation and Research (CDER) of the U.S. Food and Drug Administration (FDA). We will not discuss biomarkers as part of diagnostic tests regulated by the FDA's diagnostic centers or the use of biomarkers that do not require regulatory approval.

The FDA recently highlighted the need to develop and apply biomarkers in the drug de-

velopment and approval process (5). Most recently, CDER published guidance that describes the process it uses in biomarker qualification (6). Once a biomarker is qualified, its use in any drug development program is accepted by the FDA under the COU for which it obtained qualification. To date, the biomarker development stakeholder community has been challenged to identify and develop the evidence criteria needed to support the qualification process. Furthermore, the FDA has acknowledged the importance of coordinated efforts from multiple stakeholders to help "determine what levels of evidence befit different types of biomarkers, based on their context of use" (7) to establish qualification (8).

A coordinated effort by multiple stakeholders, including the FDA, the U.S. National Institutes of Health (NIH), industry, academia, patient groups, and the nonprofit sector, reached alignment on a proposed evidence criteria framework for biomarker development. The goal is to improve the quality of submissions to the FDA, enhance the predictability of the qualification process, and most specifically, clarify the type and extent of evidence needed to support the biomarker's COU. Ultimately, this multistakeholder effort may prove useful to the FDA and other regulators for creating new guidance. Here, we provide a brief history of milestones in biomarker qualification guidelines, summarize the methodology and assumptions used to create this framework, and describe the core features of the general framework.

DEVELOPING AN EVIDENCE FRAMEWORK FOR BIOMARKER QUALIFICATION

The history of biomarker qualification stretches back to classic statistical theory on surrogate end points (9). Since then, the discussion has evolved from initial concepts of validity criteria to more comprehensive views that account for benefitrisk relationships, the level of evidence required, the distinction of the term "qualification," and the separate needs for analytical validation (establishing that the performance characteristics of a test, tool, or instrument are acceptable in terms of its sensitivity, specificity, accuracy, precision, and other relevant performance characteristics using a specified technical protocol) (2, 10–14).

After a 1999 FDA/NIH workshop on biomarkers and the FDA Critical Path publication (2004) that initially called for an emphasis on biomarker development (15), aspects of the biomarker qualification process have been steadily developed. Yet, progress on development of evidence criteria has been intermittent given the challenges of defining a generalized framework, which is applicable across multiple disease areas, types of biomarkers, and COUs. The factors that impact the type and extent of evidence criteria needed to qualify a biomarker are complex, unique to how the biomarker will be used, and often are not easily quantified. To date, no singlestakeholder community has been able to develop evidence criteria in isolation. Therefore, this group effort developed a framework drawing on the expertise and experience of its members while following the lead of prior efforts in using a semiquantitative approach.

One salient earlier effort to develop evidence criteria by Williams *et al.* (16) proposed a semi-quantitative framework that aimed to "qualify biomarkers in terms of cost effectiveness using a set of principles that enable the evaluation of biomarkers even with incomplete knowledge." This approach applies "tolerability of risk" to supportive evidence. It concludes that biomarker qualification essentially rests on the principle that

¹Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA. ²Foundation for the National Institutes of Health (NIH), North Bethesda, MD 20852, USA. ³Takeda Pharmaceuticals International Co., Cambridge, MA 02139, USA. ⁴Pfizer Inc., Groton, CT 06340, USA. ⁵National Institute of Mental Health, NIH, Bethesda, MD 20892, USA. ⁶Critical Path Institute, Tucson, AZ 85718, USA. ⁷Pharmaceutical Research and Manufacturers of America, Washington, DC 20004, USA. ⁸National Cancer Institute, NIH, Rockville, MD 20850, USA. ⁹Biocerna LLC, Maple Lawn, MD 20759, USA. ¹⁰Massachusetts General Hospital, Boston, MA 02114, USA. ¹¹Merck and Co. Inc., West Point, PA 19486, USA. ¹²Genentech Inc., South San Francisco, CA 94080, USA.

^{*}These authors contributed equally to this work.

[†]Corresponding author. Email: dwholley@fnih.org

the value of incremental benefits provided by the true results of the biomarker must exceed the incremental costs (defined as societal harm) of the false results with the biomarker. The practicality of this approach, however, is limited given the difficulty in measuring both benefits as well as harm.

Altar et al. (17) provided the first "evidence map" for qualification with categorical descriptions of different types of scientific evidence required for different levels of risk. The map, however, was complex and did not adequately account for COU. By contrast, a 2010 Institute of Medicine study (18) emphasized the importance of COU as part of a three-part framework, which included analytical validation, biomarker qualification, and utilization components. In that case, however, the model did not address specific decision processes or evaluation criteria. Amur et al. (8, 19) noted the difference between biomarker development within drug-specific development programs and the process outlined by the FDA's Biomarkers Qualification Program (20). These authors emphasized the importance of the COU described in conjunction with definitions of biomarker types, although risk-based evidence criteria were not defined.

The framework proposed here synthesizes key elements of the above efforts to propose an overarching approach to evidence criteria. In particular, it applies some of the concepts of risk and benefit proposed by Williams *et al.* (16) to specific considerations driven by the COU and then uses these to help define an evidence map.

We applied the biomarker nomenclature outlined in the BEST (Biomarkers, Endpoints, and other Tools) glossary first published by the FDA/NIH biomarker working group in 2016 (2). For reference, categories of biomarkers are summarized in the full framework document (21). These categories largely reflect the various COUs that can inform how the biomarker may be applied in drug development and regulatory review.

Under the auspices of the Foundation for NIH (FNIH) Biomarkers Consortium, a multistakeholder group with representatives from FDA, NIH, industry, patient groups, and academia created an early draft of the framework document. A workshop cosponsored by the FNIH Biomarkers Consortium and FDA was held in April 2016 with more than 200 participants. A draft of this framework document was distributed to participants prior to the workshop, along with several case studies. The authors summarized input from the workshop and synthesized

the resulting consensus full framework document (21).

THE EVIDENTIARY CRITERIA FRAMEWORK PROPOSAL

The general evidentiary criteria framework is summarized below and includes five component steps as follows: define a need statement, define a COU statement, assess benefit in light of the COU, assess risk in light of the COU, and populate an evidentiary criteria map (Fig. 1). The current framework is intended to support constructive discussions between biomarker developers needing to submit to the biomarker qualification program and regulators at the FDA, with the goal of enabling refinements of the COU as the data mature (Fig. 1).

Need statement

The need statement is a concise and coherent description of the knowledge gap or drug development need (for example, improved safety monitoring) that the biomarker developer plans to address. It accounts for current scientific understanding of the biomarker and describes the scope of how a biomarker program, if successful, could positively impact drug development.

COU statement

The COU is central to a biomarker qualification submission. When a biomarker developer begins the process of defining a COU, there are a number of questions that should be considered to aid in the clear articulation of the COU and to promote a common understanding of the intended use of the biomarker in drug development (21). Although these questions may not apply to all biomarkers or COUs, they are intended to be elaborated before the COU is defined and can be readdressed as new information becomes available. For the purposes of this framework, the COU statement is simplified to a concise description of how a biomarker is intended to be used in drug development. It addresses two main questions: (i) What BEST category of biomarker is proposed and what information content would it provide? (ii) What is the biomarker's specific fit-for-purpose use?

Assessment of benefit and risk

Once the COU statement is articulated, benefit and risk need to be assessed separately within the broader context of the impact on patients (for example, the role of the proposed biomarker in drug development, the severity of the disease or condition, and the availability of other tools to advance drug development in that disease or condition). Thus, the determi-

nation of benefit and risk takes into account the previously described need statement and results from the COU assessment.

Although precise quantification of benefit and risk is not feasible or practical, a thorough evaluation of the reasonable benefit and risk, and a data-driven, semiquantitative assessment that encompasses all of the relevant components of the relationship between benefit and risk, are generally sufficient for decision-making. The evidentiary criteria framework contains a spectrum of benefit and risk outcomes (that is, favorable to unfavorable). Recognizing that not all information will fit perfectly into any standard category, it will be important to assess all elements of potential risk and benefit associated with the biomarker's COU relative to the need statement to make an informed assessment as to where on the continuum the biomarker fits best.

LINKING BENEFIT AND RISK TO THE EVIDENCE REQUIRED

The evidence maps in this evidentiary criteria framework are inspired by, but are considerably less complex than, the map used by Altar et al. (17). The identified need and choices made in the COU section impact the overall relative level of benefit and risk, which in turn determines the level of evidence needed to evaluate the biomarker for qualification. Note that the type, quality, and amount of data required will be linked to the benefit-risk relationship through the evidentiary criteria grid relative to the need statement (Fig. 1). Thus, a candidate biomarker with a high negative consequence if incorrect and low improvement over the current standard would require a more extensive data package than a candidate biomarker with a low negative consequence of failure and a high benefit. A list of questions to help guide the amount of evidence needed is included in the full framework document (21).

Consider a favorable benefit-risk scenario (requiring a minimal level of evidence) in which a biomarker is used for stratification of patients to ensure equal distribution of biomarker-positive and biomarker-negative individuals in the different arms of a clinical trial. If the biomarker does not perform as expected, the loss consists of the resources spent on the biomarker assay and would not influence the trial outcome to the patient or patient safety. An example of this would be a predictive biomarker of drug efficacy that attempts to limit the trial to only those who are most sensitive to the drug. If the biomarker fails, the subjects in the trial would have the same risk and benefit from treatment that would be found in the absence of the biomarker.

Proposed five-component process Drug development Need Context statement of use · Class of biomarker? · Knowledge gap? Informs required · What question does the • Drug development stringency of biomarker address? evidentiary criteria need? **F**valuate To patient compared to status quo **Benefit** Risk · Consequence of · Improved sensitivity false positive · Improved selectivity Consequence of · Mechanistic context false negative **Workflow and decision process**

Characterizing relationship of biomarker to clinical outcome.

- of biomarker to clinical outcome

 Biological rationale for
- use of biomarker

 Type of data and study
- design (i.e., prospective, retrospective, etc.)
- Independent data sets for qualification
- Comparison to current standard
- · Assay performance
- · Statistical methods to use

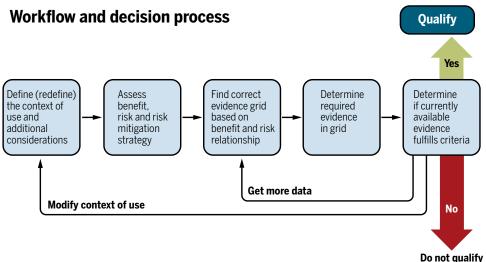


Fig. 1. Proposed evidentiary criteria framework. An illustration of the steps in the process for defining the appropriate amount of evidence needed to use a biomarker from the point of view of regulatory decision-making. The process requires defining the detailed limits of the decision and collecting the appropriate data based on how the decision will affect patients. Given the complexity of data collection, the process involves multiple conversations with the regulatory agency and can circle back if the decision or COU needs to be changed during the process.

An example of a challenging benefit-risk relationship (potential for high benefit while accompanied by high risk) is a biomarker used as a surrogate end point, which would require a high level of evidence. If the biomarker is not truly a surrogate end point in terms of predicting clinical benefit, the results of the clinical trial would be invalid, and inappropriate approval decisions would be made. This would lead to potentially ineffective drugs being marketed or patients being denied access to effective therapy. Yet, another example of a challenging COU is

a safety biomarker intended to replace the current standard and be used routinely to identify organ toxicity. In both of these examples, a high level of evidence would be needed to support the biomarker's COU. On the other hand, a safety biomarker intended to supplement existing standards to be used only in certain well-defined circumstances would be anticipated to require more moderate evidence.

As part of the evidence needed to support a biomarker's development and COU, two critical factors include the characterization and validation of the biomarker assay and the statistical plan to ensure that the evidence generated supports the eventual use of the biomarker. For the assay, biases introduced will affect the clinical interpretation of the biomarker's accuracy and its value as a drug development tool. Similarly, a careful evaluation of statistical approaches is critical because some assumptions/claims may be unfounded if based on inappropriate statistical analyses.

THE EVIDENTIARY CRITERIA ASSESSMENT MAP

Our current inability to quantitate the benefit, risk, or value of individual data sources prevents a strictly quantitative link from the amount of benefit and risk to the amount of evidence needed to qualify a biomarker. It is generally agreed, however, that we can categorize benefits, risks, and evidence into broad semiquantitative groups. Because regulatory science is not sufficiently advanced to provide this direct link, we have proposed the use of categorical descriptions for what constitutes a high and minimal level of evidence. An assessment of the strength of the evidence needed for a biomarker to be linked to drug development decision-making has been separated into several categories. These categories are captured in the evidentiary criteria assessment map and include elements of scientific understanding and biological performance and the types of clinical data and samples proposed to establish qualification. Although the process described herein refers to clinical qualification, clinical data can be augmented by nonclinical data to provide additional weight of evidence or to provide bridging information if clinical data are not feasible. Scientific understanding includes the biological rationale, understanding of the molecular mechanisms, and the link of the proposed biomarker to regulatory understanding of the scientific impact. Biological performance includes consisten-

cy of correlation of the biomarker changes, presence of a dose response, temporal relationship to the magnitude of the biomarker response and outcome, and sensitivity and specificity of the biomarker response. The types and amount of data/samples required include the quality of the data source and whether prospective or retrospective. To provide initial context, we have suggested example descriptions of expectations for several areas that have been discussed in this document [see Table 1 in (21)]. It is important to note that the evidentiary criteria for qualification are dependent

on the COU. For a broad COU, the criteria for each area may require a high level of evidence. Alternatively, a biomarker that is used in a narrow context could be fulfilled with less evidence. The criteria will be modified as more data become available and as the COU is refined during the biomarker qualification process and the field advances as a whole.

An evidentiary criteria assessment map provides a visual representation of an evidence map for biomarker development (21). For example, with assay validation, a high level of evidence should be required for regulatory clearance or approval for marketing as a diagnostic, whereas a minimal level of evidence should be deemed "fit-for-purpose" validation with acceptable performance characteristics. The evidence map can be used as a communication tool for gaining alignment between submitters and FDA reviewers at several key milestones for a biomarker development plan: (i) initial discussions to align expectations; (ii) purposeful interim progress updates to ensure that evidence expectations have been met before proceeding further; and (iii) review evaluation to support the qualification outcome. An evidentiary criteria assessment map also can be used internally by the submitter to track the current level of evidence of the biomarker relative to the intended level of evidence to meet the qualification claim.

CONCLUSION

This evidentiary crit eria framework represents alignment of multiple, diverse stakeholders and sets forth consistent, comprehensive, and semiquantitative parameters for biomarker qualification, bringing a greater degree of clarity, predictability, and harmonization. It is intended to be broadly applicable across multiple categories of biomarkers and COUs to support qualification of biomarkers for regulatory use in drug development. Given that each category of biomarker and COU has unique factors to consider as part of the development process (for example, analytical considerations for confident measurement or appropriate statistical approaches to data analysis), we propose that detailed illustrative examples or modules should be created to address these more specific issues.

REFERENCES AND NOTES

- A. Sertkaya, A. Birkenbach, B. A., J. Eyraud, "Examination of Clinical Trial Costs and Barriers for Drug Development' (2014); http://aspe.hhs.gov/report/examinationclinical-trial-costs-and-barriers-drug-development.
- FDA-NIH Biomarker Working Group, BEST (Biomarkers, EndpointS, and other Tools) Resource (Food and Drug Administration and National Institutes of Health, 2016).
- B. Celli, K. Tetzlaff, G. Criner, M. I. Polkey, F. Sciurba, R. Casaburi, R. Tal-Singer, A. Kawata, D. Merrill, S. Rennard; COPD Biomarker Qualification Consortium, The 6-minute walk test as a COPD stratification tool: Insights from the COPD biomarker qualification consortium. Am. J. Respir. Crit. Care Med. 194, 1483–1493 (2016).
- L. Landeck, C. Kneip, J. Reischl, K. Asadullah, Biomarkers and personalized medicine: Current status and further perspectives with special focus on dermatology. *Exp. Dermatol.* 25, 333–339 (2016).
- G. Lavezzari, A. W. Womack, Industry perspectives on biomarker qualification. *Clin. Pharmacol. Ther.* 99, 208–213 (2016).
- Center for Drug Evaluation and Research, Guidance for Industry and FDA Staff Qualification Process for Drug Development Tools (2014); www.fda.gov/downloads/ drugs/guidancecomplianceregulatoryinformation/ guidances/ucm230597.pdf.
- Biomarker Qualification Program, www.fda.gov/ Drugs/DevelopmentApprovalProcess/ DrugDevelopmentToolsQualificationProgram/ ucm284076.htm [accessed 4 October 2016].
- S. Amur, L. LaVange, I. Zineh, S. Buckman-Garner,
 J. Woodcock, Biomarker qualification: Toward a multiple
 stakeholder framework for biomarker development,
 regulatory acceptance, and utilization. Clin. Pharmacol.
 Ther. 98, 34–46 (2015).
- R. L. Prentice, Surrogate endpoints in clinical trials: Definition and operational criteria. Stat. Med. 8, 431–440 (1989).
- 10. The ICH E9 Guidance. Fed. Regist. 63, 49853 (1998).
- J.-P. Boissel, J.-P. Collet, P. Moleur, M. Haugh, Surrogate endpoints: A basis for a rational approach. Eur. J. Clin. Pharmacol. 43, 235–244 (1992).
- J. W. Lee, V. Devanarayan, Y. C. Barrett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J. A. Rogers, R. Millham, P. J. O'Brien, J. Sailstad, M. Khan, C. Ray, J. A. Wagner, Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* 23, 312–328 (2006).
- L. J. Lesko, A. J. Atkinson Jr., Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: Criteria, validation, strategies.
 Annu. Rev. Pharmacol. Toxicol. 41, 347–366 (2001).
- J. A. Wagner, Overview of biomarkers and surrogate endpoints in drug development. *Dis. Markers* 18, 41–46 (2002)
- FDA critical path opportunities report (2004); www.fda. gov/downloads/scienceresearch/specialtopics/

- critical path initiative/critical path opportunities reports/ucm077254.pdf.
- S. A. Williams, D. E. Slavin, J. A. Wagner, C. J. Webster, A cost-effectiveness approach to the qualification and acceptance of biomarkers. *Nat. Rev. Drug Discov.* 5, 897–902 (2006).
- C. A. Altar, D. Amakye, D. Bounos, J. Bloom, G. Clack, R. Dean, V. Devanarayan, D. Fu, S. Furlong, L. Hinman, C. Girman, C. Lathia, L. Lesko, S. Madani, J. Mayne, J. Meyer, D. Raunig, P. Sager, S. A. Williams, P. Wong, K. Zerba, A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clin. Pharmacol. Ther.* 83, 368–371 (2008).
- Evaluation of biomarkers and surrogate endpoints in chronic disease (2010); www.nationalacademies.org/ hmd/~/media/Files/Report%20Files/2010/Evaluationof-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease/Evaluation%200f%20Biomarkers%20-%20 2010%20Report%20Brief.ashx.
- S. G. Amur, S. Sanyal, A. G. Chakravarty, M. H. Noone, J. Kaiser, S. McCune, S. Y. Buckman-Garner, Building a roadmap to biomarker qualification: Challenges and opportunities. *Biomark. Med.* 9, 1095–1105 (2015).
- Innovation or stagnation: Challenges and opportunity on the critical path to new medical products (2014); www.fda.gov/ScienceResearch/SpecialTopics/ CriticalPathInitiative/CriticalPathOpportunitiesReports/ ucm077262.htm.
- Evidentiary Criteria Writing Group, Framework for Defining Evidentiary Criteria for Biomarker Qualification: Final Version (2016); http://bit.ly/2xlfs3g

Acknowledgments: The Biomarker Evidentiary Criteria Writing Group included 15 members. Content included here represents the views of individual authors. The Writing Group would like to thank all of the participants in the April 2016 workshop, "Biomarker Qualification Workshop: Framework for Defining Evidentiary Criteria," for excellent discussion and input, We thank J. Hepker of Prescott Medical Communications Group for writing and editing assistance. Competing interests: W.W.C. is employed by PhRMA, S.H. is the Scientific Program Manager for the FNIH, G.L. is employed by Biocerna LLC and is a former PhRMA employee, R.R. is the Director of Translational Medicine Initiatives at Massachusetts General Hospital and is a former PhRMA employee, J.A.W. is employed by Takeda Pharmaceuticals International Co., J.A. is employed by Pfizer Inc., F.D.S. is employed by Merck and Co., and T.Z. is employed by Genentech Inc. All other authors declare that they have no competing interests.

10.1126/scitranslmed.aal4599

Citation: C. Leptak, J. P. Menetski, J. A. Wagner, J. Aubrecht, L. Brady, M. Brumfield, W. W. Chin, S. Hoffmann, G. Kelloff, G. Lavezzari, R. Ranganathan, J.-M. Sauer, F. D. Sistare, T. Zabka, D. Wholley, What evidence do we need for biomarker qualification? *Sci. Transl. Med.* **9**, eaal4599 (2017).

Science Translational Medicine

What evidence do we need for biomarker qualification?

Chris Leptak, Joseph P. Menetski, John A. Wagner, Jiri Aubrecht, Linda Brady, Martha Brumfield, William W. Chin, Steve Hoffmann, Gary Kelloff, Gabriela Lavezzari, Rajesh Ranganathan, John-Michael Sauer, Frank D. Sistare, Tanja Zabka and David Wholley

Sci Transl Med 9, eaal4599. DOI: 10.1126/scitranslmed.aal4599

ARTICLE TOOLS http://stm.sciencemag.org/content/9/417/eaal4599

REFERENCES This article cites 13 articles, 0 of which you can access for free

http://stm.sciencemag.org/content/9/417/eaal4599#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service