Regulatory Advancements for Patients

FRIENDS
of CANCER
RESEARCH

FRIENDS of CANCER RESEARCH

INTRODUCTION

Friends of Cancer Research (*Friends*) has led the creation and implementation of policies to catalyze development and support patient access to safe and beneficial therapies for the past 25 years. Through facilitating meaningful, collaborative efforts, *Friends* has successfully engineered solutions that address shared challenges and accelerate the field forward. Each year, *Friends* convenes researchers, industry experts, and regulators to develop revolutionary, yet actionable solutions.

By convening working groups, hosting scientific conferences, and conducting research on a range of topics *Friends* helps inform regulatory policy, oncology development, and clinical practice. This emphasis on collaboration in our work creates venues and partnerships that encourage dialogue between stakeholders and, ultimately help ignite strategic advances in science and regulatory policy. In 2020, *Friends*' projects informed the many white papers, scientific abstracts, and peer-review manuscripts included in this report.

These publications represent *Friends*' mission to power advances in science, policy, and regulation that speed life-saving treatments to patients. This report provides resources intended to inform those interested in science and regulatory issues in oncology. *Friends*' 2020 publications included herein can be characterized by several key themes:

❶ **Real-world evidence:** Exploring Real-World Data to Characterize Real World Outcomes

❷ **Patient-focused drug development:** Aligning Patients' Needs with Oncology Drug Development

❸ **Complex biomarkers:** Aligning Patients' Needs with Oncology Drug Development

❹ **Optimal drug development:** Identifying Opportunities for Modernization and Innovation

REAL-WORLD EVIDENCE: EXPLORING REAL-WORLD DATA TO CHARACTERIZE REAL WORLD OUTCOMES

- RWE White Paper: Considerations for Use of Real-world Evidence in Oncology: Lessons Learned from Friends of Cancer Research Collaborations
- Overall survival (OS) in advanced non-small cell lung cancer (aNSCLC) patients treated with frontline chemotherapy or immunotherapy by comorbidity: A real-world data (RWD) collaboration.
- Trends in immunotherapy use in patients with advanced non-small cell lung cancer (aNSCLC) patients: Analysis of real-world data.

PATIENT-FOCUSED DRUG DEVELOPMENT: ALIGNING PATIENTS' NEEDS WITH ONCOLOGY DRUG DEVELOPMENT

- Use of Patient-Reported Outcomes to Understand & Measure the Patient Experience of Novel Cell and Gene Therapies
- Advocacy and Patient Involvement in Clinical Trials BT-Principles and Practice of Clinical Trials
- A National Assessment of Diagnostic Test Use for Patients with Advanced NSCLC and Factors Influencing Physician Decision-Making

COMPLEX BIOMARKERS: INFORMING STANDARDIZATION AND HARMONIZATION OF DIAGNOSTIC TESTING

- Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project
- Harmonization and Standardization of Panel-Based Tumor Mutational Burden Measurement: Real-World Results and Recommendations of the Quality in Pathology Study
- International liquid biopsy standardization alliance white paper

OPTIMAL DRUG DEVELOPMENT: IDENTIFYING OPPORTUNITIES FOR MODERNIZATION AND INNOVATION

- Accelerating the development of innovative cellular therapy products for the treatment of cancer
- Biomarker-driven therapies for previously treated squamous non-small-cell lung cancer (Lung-MAP SWOG S1400): a biomarker-driven master protocol
- Annual Meeting White Paper: Modernizing Expedited Development Programs. Nov. 2020
- Annual Meeting White Paper: Optimizing the Use of Accelerated Approval. Nov. 2020
- New drug approvals in oncology
- A regulatory science initiative to harmonize and standardize digital pathology and machine learning processes to speed up clinical innovation to patients

FRIENDS
of CANCER
RESEARCH

A FRIENDS OF CANCER RESEARCH WHITE PAPER

Considerations for Use of Real-World Evidence in Oncology

LESSONS LEARNED FROM FRIENDS OF CANCER RESEARCH COLLABORATIONS

Clinical trials, from Phase I dose-finding and safety trials to Phase III randomized trials examining efficacy, form the backbone of the drug development pipeline and inform regulatory approvals. While the centrality of clinical trials remains, there has been increasing interest in the potential contributions of real-world evidence (RWE) that results from analyses of real-world data (RWD). RWD refers to information that is collected during standard clinical care or health care billing, such as in electronic health records (EHR) or health insurance claims data, and can be leveraged for research and analytic purposes. The resulting evidence generated, called RWE, can reflect broader, more diverse patient populations than are typically included in traditional clinical trials and can be applied across multiple use cases, including to answer timely clinical questions, assess endpoints measures, perform comparative effectiveness research, and study long-term drug safety. Still, challenges remain on how to realize the full potential of RWE to support clinical research, drug development, and regulatory decision-making. Standardized variable definitions within datasets, harmonization across datasets, and application of appropriate analytical methods remain important considerations and challenges.

Recent implementation of legislative and regulatory policies focused on RWE, such as the 21st Century Cures Act, Prescription Drug User Fee Act (PDUFA) Reauthorization of 2017, and FDA Framework on Real-World Evidence, highlight the interest in using RWE applications across the drug development life cycle. Building trust in routine use of RWE for regulatory decisions will require a firm understanding of the question being asked, underlying data across real-world datasets, including the various sources of available data, their strengths and limitations, and the implications for observed endpoints. Well-validated endpoints must also be assessed as real-world endpoints to support the acceptance of RWE. Multi-

Objectives

▾ **Discuss methods and considerations for extracting data on patient characteristics**

▾ **Standardizing definitions/methodology across multiple RW datasets with the intent of aligning to similar patient populations**

▾ **Describe opportunities and potential problems in allowing flexibility in definitions**

▾ **Processes for assembling "fit-for-purpose" real-world datasets**

Friends of Cancer Research extends our thanks to the project partner organizations and working group members

PARTNER ORGANIZATIONS

ASCO

CHU Nantes

ConcertAI

COTA

Flatiron Health

Health Data Insight CIC

IQVIA

Kaiser Permanente

McKesson

National Institutes of Health

Optum Labs & Mayo Clinic

Owkin

Syapse

Tempus

U.S. Food and Drug Administration

WORKING GROUP MEMBERS

ASCO
Suanna S. Bruinooge
Elizabeth Garrett-Mayer

CHU Nantes
Pr Brigitte Dréno
Cécile Frénard
Romain Goussault
Emilie Varey

ConcertAI
Whitney Rhodes
Mark S. Walker

COTA
Andrew Belli
Eric Hansen
Keshava Dilwali
C.K. Wang

Flatiron Health
Shrujal Baxi
Aaron B. Cohen
Nicole Mahoney
Erik Rasmussen
Olga Tymejczyk
Aracelis Z. Torres

IQVIA
Jennifer Christian
Nancy Dreyer
Christen Gray
Pia Horvat
Adam Reich
Joe Wagner
Xinyan Yu

Kaiser Permanente
Lawrence H. Kushi
Lori C. Sakoda
Emily Valice

McKesson
Marley Boyd
Janet Espirito
Nicholas Robert

NIH
Eric Boyd (contractor with Information Management Services Inc., Calverton.)
Lindsey Enewold
Donna Rivera
Elad Sharon

OptumLabs
Henry Henk

Owkin
Etienne Bendjebbar
Maxime He
Anna Huyghues-Despointes
Anne-Laure Moisson
Radha Patel

Syapse
Monika Izano
Yanina Natanzon
Connor Sweetnam

Tempus
Ruth Pe Benito
Rebecca Honnold

stakeholder collaboration is necessary to develop robust recommendations to maximize the quality and utility of RWD analyses. This includes selecting datasets and sources that are appropriate and fit-for-purpose to address the question being addressed, as well as the subsequent evidence generation that is needed in support of oncology research, including drug development.

Informed by several pilot projects leveraging a common protocol (established in the RWE 1.0 Pilot Project and expanded upon in subsequent pilots, described below) among multiple real-world data partners, US and international populations, and oncology-specific disease settings, Friends of Cancer Research (*Friends*) and collaborators identified implications of dataset specifics and patient characteristics on real-world endpoints and recommendations for developing a RWE framework to encourage and guide future RWE studies that leverage multiple data sources to answer a single question through a harmonized protocol.

Friends of Cancer Research Real-World Evidence Pilots

RWE 1.0 Pilot Project

The initial *Friends* RWE Pilot, 1.0, brought together six data partners to evaluate the performance of real-world endpoints across multiple data sources by focusing on a common clinical question: What endpoints for advanced non-small cell lung cancer (aNSCLC) patients treated with immune checkpoint inhibitors can be evaluated and compared across all of these data sources? To answer this question, the RWE Pilot 1.0 members aligned on a framework of necessary data elements, characteristics, definitions for real-world (rw) endpoints, based on data availability in electronic health record (EHR) and claims systems. The preliminary goal was to evaluate whether the various datasets included in this study could achieve similar results when measuring treatment effect using a common framework. The protocol developed through the RWE Pilot 1.0 served as the basis for several additional pilot projects aimed at (1) identifying minimum data quality and reporting standards to aid the interpretation of individual RWD studies, and comparisons across studies performed using different RWD sources, (2) evaluating the ability to estimate and compare effectiveness endpoints for different therapies across the data sources, and (3) adapting this framework for evaluation of rw endpoints in the context of a specific research question. Results from RWE Pilot 1.0 showed that similar patient populations could be extracted across datasets with differing underlying data sources using aligned baseline characteristic definitions and a harmonized protocol, and that certain rw endpoints, time-to-treatment discontinuation (rwTTD), were correlated with rw overall survival (rwOS).

RWE Pilot 1.0 was then extended to other data sources and disease settings in an effort to further examine the generalizability of the findings and the framework.

RWE Pilot 1.0: RWE Framework in the United Kingdom Cancer Analysis System

Through a collaboration with IQVIA and Health Data Insight CIC, using data from the Cancer Analysis System (CAS) database in the United Kingdom, we sought to apply the framework established in Pilot 1.0 to confirm previously observed associations between rwOS and potential proxy endpoints (rwTTD/TTNT) in a nationally sourced, population-level, dataset. The CAS study followed the *Friends* RWE Pilot 1.0 protocol and compared the original findings from six US data sources to a UK Cancer Registry (CAS database) to compare RWD in aNSCLC. CAS is a cancer registry that includes more than 99% of all cancer patients in England and contains data on patient and tumor characteristics, treatments, hospitalizations, and mortality.

This study supported the findings from the original RWE Pilot 1.0 project and demonstrated a high level of correlation between rwOS and other rw endpoints, indicating the potential use of rwTTD/TTNT as a proxy endpoint for OS in real-world studies.

RWE Pilot 1.0: RWE Framework in Melanoma Patients from the RIC-Mel Database

In collaboration with Owkin/Centre Hospitalier Universitaire (CHU) de Nantes Pilot, we investigated the broader applicability of the RWE Pilot 1.0 framework in patients treated with immune checkpoint inhibitors for melanoma (anti-PD-1 monotherapy or anti-PD-1 combined with anti-CTLA-4 therapy as a first line of treatment or as a later line of treatment in advanced and metastatic melanomas). The project utilized the RIC-Mel database, which federates key patient information across 49 research institutions in France, with near comprehensive coverage and data that is highly curated and harmonized as all data collection is unified under a common CRF and digital platform interface for melanoma patients. Extending the RWE framework to patients with melanoma also provided an opportunity to align on data quality, standards, and investigate rw endpoints and their correlation to OS in other disease settings outside NSCLC. The applicability of the RWE framework in melanoma, supports further development of the framework as the structure for future studies.

RWE Pilot 2.0: Treatment Comparisons Analysis

Building on the work of the RWE Pilot 1.0, in 2019, *Friends* convened ten data partners, including organizations with data from EHRs or insurance claims, to conduct a parallel study where the different data sources were used to assess endpoints among aNSCLC patients receiving different first-line treatment regimens. Given the accumulating clinical experience with immune-oncology (IO) therapies, the RWE Pilot 2.0 was performed to assess treatment effect between platinum doublet chemotherapy, PD-L1 monotherapy, and PD-L1 in combination with platinum doublet chemotherapy using a common protocol. Patients meeting broad inclusion and exclusion criteria (treated with a qualifying therapy in first-line for aNSCLC, see SAP) were included to reflect the real-world population represented by the different data sources. Results of the study trended towards better outcomes in patients receiving IO over chemotherapy, directionally consistent with the findings of recent clinical trials.

RWE Pilot 2.0: Internal Consistency Analysis

Five RWE Pilot 2.0 data partners with data sourced from EHRs formed a subgroup to assess the consistency of findings in relation to trial results that examined the same treatment effects. Using the initial RWE Pilot 2.0 protocol as the basis for this study, additional patient inclusion/exclusion criteria were applied, leveraging EHR and lab data. The criteria, based on the KEYNOTE-189 clinical trial (platinum doublet chemotherapy versus PD-L1 in combination with platinum doublet chemotherapy in first-line aNSCLC), guided cohort selection to compare treatment effects in a more homogenous 'trial-like' real-world population, using rw endpoints of OS, TTNT, and TTD. The inclusion/exclusion criteria in this analysis were selected to facilitate greater alignment of baseline characteristics across datasets and greater similarity to clinical trial populations, although significant differences remain, on which approval of immune checkpoint inhibitors had been based (no balancing or weighting was applied). Treatment effects were compared at multiple restriction steps by select trial-based criteria, to assess how inclusion/exclusion criteria may have contributed to differences in observed treatment effect estimates. Additionally, the application of nuanced trial-based inclusion/

Box 1
Pilot 2.0 Research Questions

▼ **Objective 1:**
Describe demographic and clinical characteristics of patients with aNSCLC receiving frontline chemotherapy doublet, PD-(L)1 monotherapy, or PD-(L)1 + doublet chemotherapy.

▼ **Purpose:**
Provide baseline understanding of the similarities/differences among the datasets to describe what confounding factors may need to be considered when interpreting the data.

▼ **Objective 2:**
Evaluate treatment effect size in frontline therapy regimens using real-world endpoints.

▼ **Purpose:**
Agree on data-source specific definitions and measurement of endpoints assessed through real-world data, in order to ensure reliability, consistency, and preservation of clinical meaning.

exclusion criteria to EHR data yielded important insights regarding data capture and the ability of RWD to identify precisely defined patient populations and characteristics.

The results from the above analyses were shared amongst all participating groups, to facilitate discussion of the combined learnings, and to subsequently develop a list of considerations for the design, conduct and interpretation of RWD studies from different data sources. Manuscripts are pending for each of the four expansion pilots.

Considerations for a Real-world Evidence Framework

These five RWE pilot programs have yielded important lessons learned regarding establishing a RWE framework across multiple data partners in order to answer a common clinical question and we summarize these below.

Establishing a Research Question

To begin, defining and aligning on the clinical research question or objective is the key for any RWD study. All subsequent study considerations, including whether a data source is fit for purpose (meets certain data quality and completeness to address a specific question) and acknowledging potential limitations of data collection in real-world practice settings), will be guided by the clinical question. The considerations addressed in this whitepaper reflect lessons learned in the context of the RWE Pilot 2.0 research questions (Box 1).

Standardizing a common set of data elements

Establishing a core set of data elements to collect and standardize definitions could enable greater comparability across RWE studies, independent of data source(s). Demographics such as age and sex are minimal, structured, data elements that are typically readily available across independent data sources. However, eligibility criteria and definitions for other data elements demand thoughtful consideration and transparency such as: a) variables available in different formats (for example, PD-L1 biomarker positive/negative indicator vs. percent staining), b) variables requiring a curated definition (for example, ICD codes vs. lab values in the definition of organ function), or c) variables requiring extraction from unstructured data (for example, status of advanced cancer at initial diagnosis vs. progression after initial, earlier-stage diagnosis).

Using the RWE Pilot 2.0 as a case study, we propose a core set of data elements for consideration in real-world oncology studies **(Appendix)**. We include further considerations for harmonizing definitions across datasets with the prerequisite core data elements to address the pre-specified research question.

Considerations

1.   Identify a core set of data elements that can be systematically defined across real-world data sets for the proposed study.

Important considerations when creating a core set of data elements include commonality and availability of data across datasets, the clinical setting, and study objectives. Data elements that are consistently available across all or the majority of datasets will reduce data variability and increase understanding of data missingness within the datasets. The completeness of each data element within each dataset, as well as across datasets, should be evaluated and reported. Selection of core data elements should also consider the clinical context. As a result, a core set of data elements **(Appendix)**, will require modifications when applying across disease or therapeutic class. For example, smoking status may be relevant across multiple diseases but provide particularly important prognostic information in lung cancer, as opposed to other cancer types, such as melanoma. The phenotype of melanoma requires different information such LDH, BRAF and histologic details of primary lesion. However, age, sex, and stage of disease are important data elements across all of the RWE Pilots.

Additionally, consider that the prognostic value of a characteristic such as smoking status will depend upon the level of variable completeness and definition used for this data element (for example, patient was never a smoker vs. there is no evidence of smoking history). Last, consider the study objectives and endpoints to be measured when selecting core data elements as this will help with selecting the most appropriate data elements. For example, patient age at advanced diagnosis would be a more appropriate characteristic than age at initial diagnosis (where a patient presented with early stage disease and now has aNSCLC) for a study objective to observe treatment effect in patients with aNSCLC.

2.   Identify the analytic variables that require a high level of harmonization vs. those that can accommodate variability across data sources.

Harmonized definitions should be employed wherever possible and particularly for data elements with high likelihood of impact on endpoint calculations. However, standardized definitions are not always feasible and variability across datasets may be acceptable as assessed on a per study basis. For example, even when using a common case report form across institutions within a data source, heterogeneity of information reporting can persist between institutions (e.g. lymph node removals can be coded differently depending upon the clinical site). To identify data elements where variability may be acceptable, consider 1) whether harmonization is possible given each source of the data element (e.g., EHR vs. claims data; diagnosis vs. laboratory value for defining a comorbidity) and the underlying population and 2) whether harmonization is necessary. For example, treatment initiation date could be sourced from administrative claims, an electronic

prescription order, or date of administration within an EHR, and a flexible definition could ensure more comprehensive identification of patients receiving a particular treatment. Similarly, practice patterns can vary across geographic regions and clinical practices and can impact how a frontline therapy is defined within different datasets. Flexible definitions may be needed that place greater emphasis on accurate identification of appropriate patient populations within the context of each dataset compared to harmonization of variable definitions among datasets. Last, a harmonized definition may not be necessary for some data elements, particularly where little to no impact on the included patient population or calculation of endpoints is expected or where stringent definitions could limit potential observations. For example, the RWE Framework broadly identified inclusion based upon treatment with a platinum doublet chemotherapy or IO monotherapy or IO in combination with any platinum doublet chemotherapy but did not restrict to specific drugs or pre-defined procedure codes (Healthcare Common Procedure Coding System (HCPCS)/Drug Codes) for specific regimens. As a result of allowing for inclusion based upon a class of drugs, the study evaluated on and off-label use that might have been excluded from the study if more stringent inclusion criteria had been applied.

Similarly, if the variable in question is included in an analysis as a potential confounder (i.e., to adjust for confounding), the specific form the confounder takes in the model may be less relevant than other variables for which specific inferences are intended. The strength of confounding exhibited by each variable is also important and consistent modeling of each variable across datasets will be important to control for confounding.

3.   Align on harmonized definitions where appropriate.

Harmonizing the definition for key data elements can help account for variability likely to exist among datasets in terms of the source and patient population represented. Factors to address when aligning definitions include accuracy, extent of missingness, and granularity. Similarly, categories of reference values for classification of covariates (such as lab values) should be agreed upon and used consistently.

First, data accuracy is important to consider for harmonization of definitions. For example, the definition of covariates such as organ function, which can be extracted from ICD codes or laboratory test results, should be considered for potential implications on results. Analyses utilizing extracted laboratory values are likely to have greater granularity when comparing magnitudes of organ failure (normal, mild, moderate, or severe) as compared to definitions based on structured ICD codes, which may communicate less information (for example, evidence of organ disease) but be recorded more frequently than lab test results. Different considerations for missingness should be accounted for when comparing diagnostic data [ICD codes] vs. lab value data, where absence of ICD codes or test results does not equate to absence of a condition. Differentiation between patients with no evidence of organ disease and patients with unknown organ function will be difficult or impossible, particularly when utilizing only structured ICD codes.

Second, consider the source/level of detail of data elements when harmonizing definitions, particularly where there is variability in how the data element is documented. For example, PD-L1 expression can be reported in a variety of ways (pathology report vs physician reported) and additionally have different thresh-

olds for what constitutes a "positive" or "negative" result. When both sources of biomarker status are used, definitions should reflect the existing variation and attempt to align populations where possible.

Third, when addressing data missingness and the reason for missingness (whether or not missingness is at random) it is important to understand the indication(s) for measuring covariates such as organ function. While certain tests of organ function may be done routinely in line with clinical guidelines, some may be ordered specifically if patients have preexisting conditions or present with certain symptoms. In that case, the ascertainment is biased and will impact endpoint estimates. Similarly, HIV testing is not routinely done outside of clinical trial selection. It is also important to consider that for some comorbidities, such as hypertension, using diagnostic vs. lab data for identification could lead to different endpoint estimates. Patients who have hypertension that is controlled through medication may be identified by ICD codes for hypertension or diabetes in some patient records. However, those same patient records would indicate normal blood pressure values due to control with anti-hypertensive medications. As a result, the use of diagnostic codes vs. lab values may not yield the same value for some covariates. The same would apply to use of diagnostic codes vs lab values of blood glucose to identify diabetes in a patient on medications to normalize blood glucose level. The ascertainment window for laboratory values will also impact this measure, taking into account proximity of data ascertainment to timeframe of interest and how to address reporting of multiple laboratory values during the study period.

Last, consider the granularity of definitions. For example, identification of adverse events is of particular importance in RWE but is especially difficult to measure if relying primarily on structured data to attribute to a particular therapy or treatment. Assignment of attribution requires chart review, which is variable and time consuming. A related example is use of the term "advanced" (this includes both Stage IIIB/C and IV disease) to identify aNSCLC patients, which can have an impact on observed endpoints for the specified population. The term "advanced" may be defined as a patient with a certain stage of disease at diagnosis or having developed metastatic disease independent of initial stage, but if the focus of the study is the treatment of metastatic NSCLC then only Stage IIIB/C patients should be eligible if they progressed with metastatic disease. The definition of advanced in this case will depend upon the ability of each dataset to capture and identify progression as a disease indicator or as a clinical endpoint within each dataset. Similarly, large amounts of missingness in progression data may also impact the patient population selected for a study and should be considered when aligning definitions and interpreting results. Other considerations include clinical guidelines and workflows for disease surveillance (following treatment), that may or may not differ across practice settings, and duration of follow-up.

4.   Review of the distribution of identified variables by collaborators.

Lastly, review the distribution of the pre-specified variables for each population across datasets as an internal check on the study definitions and alignment on methods implementation. Specifically, use this informal assessment to identify unexplained outliers associated with a collaborator/data source (e.g., high number of early deaths or very long survival, or high percentage of advanced diagnosis) and missingness within datasets. This can help to not only identify where additional checks are needed to determine if an error has been made (e.g., in the harmonization process) but data sources to exclude for certain analyses because data is

not fit-for-purpose (at least for the study at hand).

Methodological considerations for interpreting endpoints

The information that can be gleaned from a real-world study depends on the methodology and definitions used to select the patient population. In addition to a core set of data elements and harmonized definitions for patient selection, it is essential to align on common statistical methodology for analyzing and interpreting endpoints. The level of specification in the real-world methodology is paramount as this will help to reduce confounding and variability in implementation due to differences in interpretation of the protocol. Ultimately, a thorough methodology will be important to facilitate earlier engagement with regulatory agencies for RWD to support drug development and regulatory decision making as well.

- **Identify and summarize the source of the endpoint information and how the endpoint was derived.**

A thorough understanding of the source of the data used to derive the endpoint, including limitations and completeness of the data, is necessary to draw accurate conclusions. The source, completeness, and accuracy of mortality data in observational studies can impact comparative effectiveness inference. For example, death information is not systematically captured in routine clinical care in the U.S., potentially requiring multiple sources of information to be used to capture mortality. Further, calculating sensitivity and specificity for mortality in EHR data requires linking to a gold standard and may present challenges for de-identified data sources, particularly in the U.S. data sources. Even though a centralized mortality database, the National Death Index, exists for research use, it is often not linked to in the context of RWD, and regardless is not complete in a timely manner, precluding its use for evaluation of new therapies. Reporting metrics including data completeness, sensitivity, and specificity for mortality should be established. In cases where EHR data is used, more information can be obtained by performing chart review as opposed to strictly depending on structured data.

- **Determine appropriate endpoints.**

The specific research question and clinical context will drive selection of an appropriate endpoint **(Tables 1 and 2)**. For example, objective endpoints, such as OS are susceptible to factors such as post baseline events, such as treatment crossover, and may make treatment effects harder to interpret. OS may also suffer from substantial missingness. Some endpoints such as progression free survival or overall response rate may not be appropriate for RWE studies due to the difficulty of identifying progression or response in structured RWD. Specifically, progression/response is not consistently reported in real-world care and, where it is available, requires time consuming chart review to extract. Further, while clinical trials rely upon objective and well-defined Response Evaluation Criteria in Solid Tumors (RECIST) criteria to measure progression, in RWD, progression/response assessments may not be as rigorous as RECIST, with more subjective clinician interpretation, variability in the scheduling of imaging tests than in trials, and less rigorous reporting in the EHR. Consensus in how to define and document progression/response in structured data would make this endpoint more readily available and appropriate as a rw endpoint.

Conversely, treatment-based endpoints, such as rw time-to-treatment discontinuation (rwTTD) or time-to-next treatment (rwTTNT), are more objective, may be more readily interpreted, and may present advantages regarding completeness. However, these endpoints do not explicitly capture differences in drug effectiveness. Interpretations are complicated by the diversity of reasons for treatment discontinuation or switch (such as toxicity or patient preference), as well as differences in expected treatment duration (e.g. pre-defined number of cycles vs indefinitely) across therapies or indications. For example, treatment is arbitrarily stopped after 4-6 cycles in some cancers regardless of the status of disease. Similarly, it is difficult to assess the end date for oral oncolytics using only structured data from EHRs. Despite different reasons for discontinuation, the clinical endpoints may be more relevant to the patient. The research question, clinical context, quality of mortality variable, and availability of additional data (e.g., on post-baseline therapies or reasons for treatment discontinuation) should help guide endpoint selection.

- **Provide transparency on endpoint derivation, definition and transformation.**

Transparency regarding how endpoints are derived (e.g. detailed documentation of deviation in methodology regarding the source of the data and what transformation is conducted to derive the endpoint) is important for 1) standardizing methodology and confirming comparability of results, 2) performing validation studies of the endpoints, and 3) building trust in the results of RWD studies. Variability in data sources, completeness and quality of data, as well as limitation of analysis plans, including defining exposure, endpoints, and key covariates, and potential resulting biases all need to be considered.

It may be preferable, when comparing data from disparate RWD, to pre-specify more than one estimate or measure of association for comparison (for example, proportion of patients that are event-free at pre-specified timepoints [the survival function] and a hazard ratio) **(Table 2)**. Various measures may be affected differently by dataset characteristics and study design elements, including distribution of exposure to treatments over time, duration of follow-up per treatment arm, and crossover from one treatment arm to another. Characterization of adjusted survival curves can be considered.

- **Ensure comparability of index dates**

When conducting real-world studies to assess treatment effect, the comparability of index date (e.g., the start of the time when patients experience the qualifying event and become at-risk of having the endpoint of interest) is of critical importance (Box 2). For example, if entry into a dataset is linked to post-baseline data, this may generate bias and render estimates not comparable across data sources.

- **Ensure comparability of censoring rules and event dates**

Harmonization, and transparency where harmonization is not possible, of censoring rules and event dates will increase interpretability of results. Important considerations include the length of follow-up available (and continuity of follow-up), the therapy being investigated, and the endpoints to be measured **(Table 1)**. Various types of patient activity recorded in different datasets (e.g., structured visits, labs, abstracted dates)

Box 2
RWE Pilot Definitions for Identification of Index Date

▼ **Frontline will be defined as the first systemic drug regimen given to NSCLC patients subsequent to the patient's date of advanced diagnosis and after the start of time period of the study.**

▼ **Frontline will begin on the date that the systemic cancer treatment was initiated and include all agents received within 30 days following the day of first administration.**

▼ **The index date will be based on the earliest drug episode (e.g., first administration or non-cancelled order) of the frontline therapy for advanced disease.**

and their applicability for use as censoring dates for the chosen endpoint, should be considered. For example, structured visit or claim dates might be most readily available and could be standardized for a mortality endpoint, but endpoints such as progression-free survival require a finer distinction between types of clinical encounters.

A data cutoff should be pre-specified, to ensure ascertainment of event and censor dates over a standard timeframe. The selection of data cut-off should be informed by the research question (how much follow-up is expected to be necessary to observe a treatment effect for the disease?) and the endpoint selected (how much follow-up time is necessary to accrue a meaningful number of events, and, in case of mortality, for datasets to optimize sensitivity of event capture from external data sources?).

When defining an event where a combination therapy is used, censoring rules must be harmonized across datasets to ensure consistent assignment of discontinuation (e.g., do both therapies within a combination need to be discontinued or does discontinuation of a single therapy within a combination constitute an event?).

• **Assess "fitness-for-purpose" to increase confidence in endpoint.**

Studies to assess the reliability of the endpoint used are necessary until a larger body of RWE comparative effectiveness in oncology literature is available. These studies may confirm accuracy of OS (i.e., identify if differences in OS estimates are true or artifactual due to incomplete mortality data) or support correlation of a proxy endpoint to a gold-standard endpoint.

• **Contextualize methods and results against other data sources, as applicable.**

It is important to consider factors that increase confidence in the real-world measure/endpoint in RWD. This could also be addressed by examining associations of each covariate (e.g., age, sex, etc.) with survival endpoints to assess whether observed effects fall in line with expectation and whether their directions and magnitudes of association are comparable across study populations. Further, where possible, comparability of survival curves for the selected endpoint to similar epidemiological studies performed within similar

populations, such as with SEER data, should be conducted. However, this requires confidence that the study population is similar to the patient population in the comparison population. This may be difficult to achieve given the possible number of patient characteristics that are un-identifiable or confounding within an observational dataset. Potential for confounding also contributes to the lack of agreement or inability to conduct appropriate RWD-based analyses of clinical trial results, and in addition to a lack of endpoint comparisons as described above, as PFS assessment is often not readily available in RWE. Similarly, clinical trials involve detailed protocols for care delivery (e.g., how standard of care or the investigational agent is delivered) whereas differences in real-world protocols can confound comparability among studies. Where possible, additional inclusion/exclusion criteria, as well as weighting, can be applied to better align real-world data populations in comparison to other data sources, (such as clinical trial or other observational data) to enhance comparability of findings. Ongoing evaluation of the study objectives and endpoints, and revisions, where necessary, should occur throughout the study to promote comparability.

- **Ensure replicability of endpoint measures.**

Similar to comparability to other studies, efforts should be taken to increase replicability of endpoint measures. For example, endpoint measures such as treatment administration date or date of service to derive a time to next treatment may be readily identified in RWD and, thus, can be replicated across datasets as compared to identification of progression, which may require date of and results of radiographs, laboratory tests, and/or clinician assessments. It is essential to conduct sensitivity analyses when comparing across datasets because of potential variability to ensure robustness of the results and stability of the estimates, especially with variance in statistical methodologies which may account for differences.

A process for assembling "fit-for-purpose" real-world datasets

Although there exists certain challenges and limitations with RWE, with a thorough understanding of the data provenance and well-designed study protocols, real-world datasets can be assembled that produce robust analyses that complement those of clinical trials and other datasets. By applying the *Friends* RWE Framework in several clinical settings and diverse data sources, we developed a process for assembling a "fit-for-purpose" real-world dataset to guide future real-world studies **(Figure 1)**. It is important to emphasize that the recommendations enumerated in this whitepaper are intended to inform a process for assembling fit-for-purpose datasets and methodologies based upon lessons learned from the *Friends* RWE collaborations. The exact core data elements, definitions, and protocols may not necessarily apply to all clinical settings or datasets as the RWE Pilots were developed to inform treatment effect of IO therapies in a specific disease setting. Certainly, modification of this framework will be appropriate to expand to other drug classes, other cancer types, different health systems (international studies), and beyond oncology. Specifically, consideration of the added complexity is necessary to adapt a standard protocol across multiple settings, e.g., different study periods due to different scope/timeline of regulatory approvals and existence of different regulatory and clinical guideline bodies, as well as considerations around accessing sensitive patient data under different patient privacy restrictions or using de-identified patient data. With more widespread application of this framework, we can begin to accumulate a body of evidence in support of various real-world endpoints and inform regulatory policy.

Table 1: Real-world Endpoint Definitions

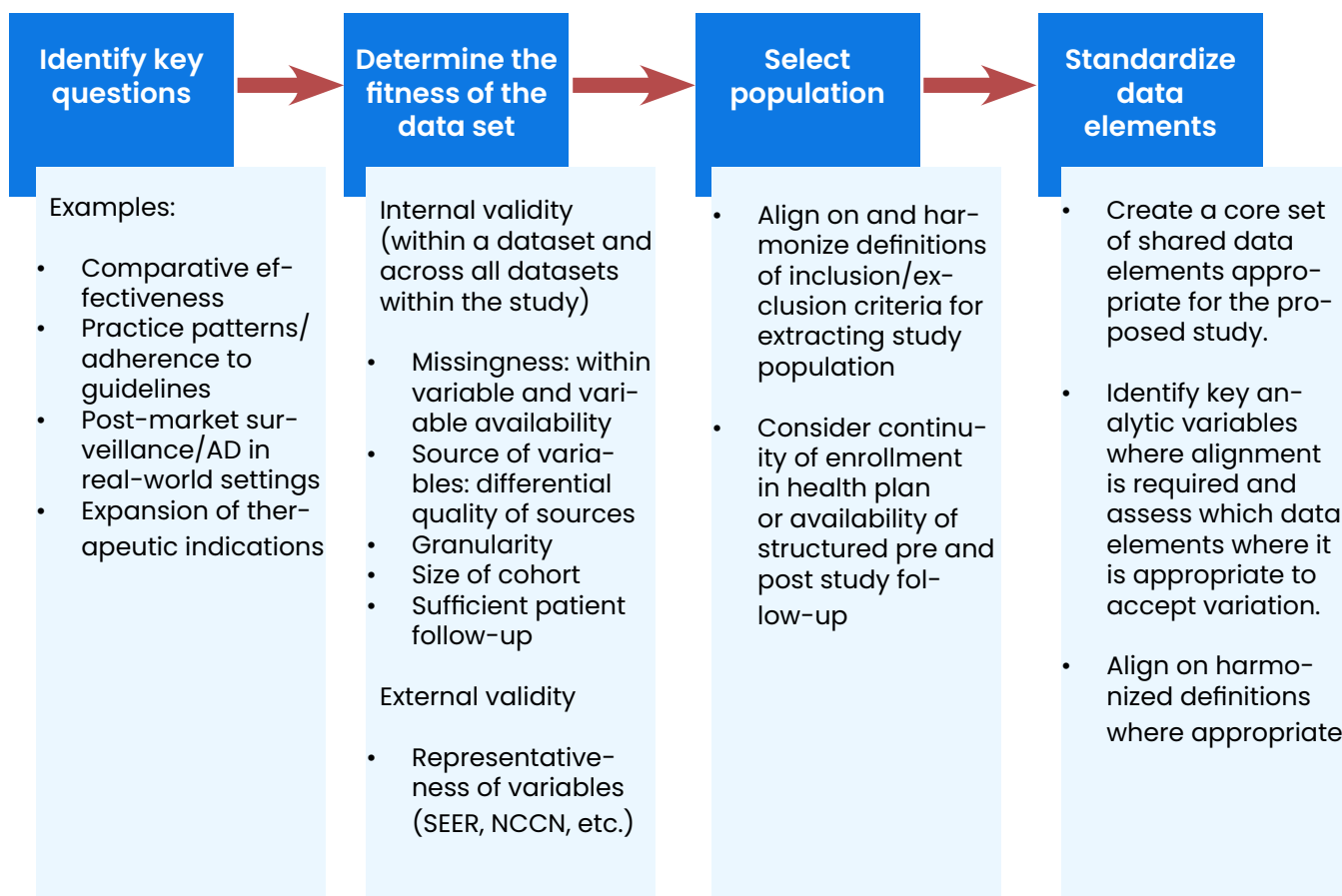| rwEnd-point	Definition	Censor Date	Considerations for Definitions and Alignment
rwOS	Length of time from the index date to the date of death, or disenrollment (need to define gap in enrollment). For claims data, health plan disenrollment date is incorporated if deaths are not captured among those who leave health plan coverage.	Last structured recorded clinical activity within the real-world database including prescription, office or institutional billing claims data, or end of follow-up period, whichever occurs earliest.	• Definition variability appropriate. • Separate definitions required for EHR-based vs claims-based data sources for all endpoints. • Consider the completeness of vital status data.
rwTTNT	Length of time from the index date to the date the patient received an administration of their next systemic treatment regimen or to their date of death if there is a death prior to having another systemic treatment regimen.	Last known activity or end of follow-up.	• Length of patient follow-up to capture subsequent treatment regimens.
rwTTD	Length of time from the index date to the date the patient discontinues frontline treatment (i.e., the last administration or non-cancelled order of a drug contained within the same frontline regimen). Discontinuation is defined as having a: • having a subsequent systemic therapy regimen after the frontline treatment; • having a gap of more than 120 days with no systemic therapy following the last administration; or, • or having a date of death while on the frontline regimen.	Last known usage (i.e., administration or non-cancelled order) of frontline treatment.	• Consider standard duration of frontline treatment regimen.

Table 2: Comparison of rwEndpoint Event Estimates to Assess Treatment Effect

Group	Treatment Regimen	rwOS				rwTTD				rwTTNT			
		N of Patients	12 month Survival Estimate	95% Confidence Interval (lower)	95% Confidence Interval (upper)	N of Patients	12 month Treatment Continuation Estimate	95% Confidence Interval (lower)	95% Confidence Interval (upper)	N of Patients	12 month Next Treatment Estimate	95% Confidence Interval (lower)	95% Confidence Interval (upper)
A	Platinum Doublet Chemotherapy	1542	0.530	0.504	0.555	1542	0.035	0.026	0.045	1542	0.230	0.208	0.252
	PD-(L)1 Monotherapy	257	0.501	0.428	0.569	257	0.322	0.253	0.393	257	0.409	0.340	0.476
	PD-(L)1 + doublet chemotherapy	132	0.506	0.401	0.602	132	0.330	0.229	0.439	132	0.435	0.333	0.532
B	Platinum Doublet Chemotherapy	697	0.556	0.518	0.592	1572	0.047	0.033	0.065	1572	0.246	0.215	0.279
	PD-(L)1 Monotherapy	429	0.615	0.567	0.659	492	0.263	0.223	0.306	492	0.440	0.392	0.486
	PD-(L)1 + doublet chemotherapy	233	0.570	0.503	0.630	233	0.227	0.176	0.283	233	0.437	0.373	0.500
C	Platinum Doublet Chemotherapy	997	0.629	0.596	0.661	998	0.042	0.030	0.056	995	0.216	0.189	0.244
	PD-(L)1 Monotherapy	160	0.575	0.470	0.667	160	0.008	0.001	0.039	159	0.286	0.206	0.372
	PD-(L)1 + doublet chemotherapy	56	0.734	0.565	0.846	56	--	--	--	56	--	--	--
D	Platinum Doublet Chemotherapy	1351	0.494	0.467	0.521	1351	0.033	0.025	0.044	1351	0.230	0.207	0.253
	PD-(L)1 Monotherapy	235	0.515	0.448	0.577	235	0.191	0.144	0.244	235	0.315	0.256	0.375
	PD-(L)1 + doublet chemotherapy	75	0.456	0.339	0.565	75	0.187	0.108	0.282	75	0.290	0.190	0.396

Group	Therapy												
E	Platinum Doublet Chemotherapy	1146	0.461	0.432	0.489	1146	0.044	0.033	0.057	1146	0.180	0.159	0.203
	PD-(L)1 Monotherapy	90	0.542	0.433	0.638	90	0.216	0.136	0.307	90	0.340	0.242	0.439
	PD-(L)1 + doublet chemotherapy	33	0.636	0.450	0.775	33	0.394	0.231	0.554	33	0.455	0.282	0.612
F	Platinum Doublet Chemotherapy	9707	.430	.420	.440	9707	.213	.204	.222	9707	.185	.177	.193
	PD-(L)1 Monotherapy	--	--	--	--	--	--	--	--	--	--	--	--
	PD-(L)1 + doublet chemotherapy	--	--	--	--	--	--	--	--	--	--	--	--
G	Platinum Doublet Chemotherapy	1453	0.582	0.553	0.612	1453	0.039	0.029	0.053	1453	0.231	0.208	0.257
	PD-(L)1 Monotherapy	132	0.657	0.565	0.763	132	0.253	0.164	0.393	132	0.461	0.367	0.580
	PD-(L)1 + doublet chemotherapy	57	0.700	0.577	0.848	57	--	--	--	57	0.507	0.381	0.675
H	Platinum Doublet Chemotherapy	4752	0.463	0.447	0.478	4752	0.039	0.033	0.046	4752	0.170	0.159	0.182
	PD-(L)1 Monotherapy	601	0.594	0.550	0.636	601	0.319	0.278	0.361	601	0.442	0.398	0.485
	PD-(L)1 + doublet chemotherapy	243	0.564	0.495	0.628	243	0.319	0.256	0.382	243	0.410	0.343	0.475

*-- data unavailable due to small sample size, insufficient follow-up, or lag in availability of data within data source

Figure 1. Step-wise approach to assembling "fit-for-purpose" real-world dataset.

Identify key questions	Determine the fitness of the data set	Select population	Standardize data elements

Identify key questions

Examples:

- Comparative effectiveness
- Practice patterns/adherence to guidelines
- Post-market surveillance/AD in real-world settings
- Expansion of therapeutic indications

Determine the fitness of the data set

Internal validity (within a dataset and across all datasets within the study)

- Missingness: within variable and variable availability
- Source of variables: differential quality of sources
- Granularity
- Size of cohort
- Sufficient patient follow-up

External validity

- Representativeness of variables (SEER, NCCN, etc.)

Select population

- Align on and harmonize definitions of inclusion/exclusion criteria for extracting study population
- Consider continuity of enrollment in health plan or availability of structured pre and post study follow-up

Standardize data elements

- Create a core set of shared data elements appropriate for the proposed study.
- Identify key analytic variables where alignment is required and assess which data elements where it is appropriate to accept variation.
- Align on harmonized definitions where appropriate

Appendix: Core Data Elements

Variable Name	RWE Protocol Definition	Considerations for Definitions and Alignment
Minimal Structured Data Elements		
Characterization of the study population		
Advanced diagnosis date	• Date of diagnosis of advanced disease: 　○ Initial diagnosis with stage IIIB, IIIC, or IV or 　○ First recurrence/progression after earlier stage diagnosis	• Definition should be aligned or transparency of variability must be provided • Dependent upon ability to identify recurrence or progression within dataset • Protocol for identification of progression may differ across datasets
Age at index	• Age at the start of frontline therapy, as previously noted in the index date definition • Reported as continuous, categorical, and binary. • Categorical: 　○ <49 years 　○ 50-64 years 　○ 65-74 years 　○ 75+ years • Binary: 　○ <75 years 　○ 75+ years	• Binary categorization should reflect study objectives. If comparison to clinical trial data is of interest, binary age should be reflective.
Sex	• Male • Female • Other / Unknown	
Region	• Based on patient's state of residence: 　○ Midwest 　○ Northeast 　○ South 　○ West 　○ Other/Missing	

Race	• White • Black or African American • Asian or Pacific Islander • Other (Native American, Alaskan Native) • Unknown/Missing	• The level of harmonization for this covariate should consider key objectives for the study as granularity could impact endpoints.
Histology	• Non-squamous cell carcinoma • Squamous cell carcinoma • NSCLC histology not otherwise specified (NOS)	• The level of harmonization for this covariate should consider key objectives for the study as granularity could impact endpoints.

First-line regimen	• First regimen patient received in the advanced setting. Eligible frontline therapies include:
	  o Platinum Doublet chemotherapy (cisplatin, carboplatin, oxaliplatin, or nedaplatin with pemetrexed, paclitaxel, nab-paclitaxel, gemcitabine)
	  o PD-(L)1 monotherapy (pembrolizumab, nivolumab, atezolizumab)
	  o Any PD-(L)1 + doublet chemotherapy combination (pembrolizumab, pemetrexed and platinum or pembrolizumab, platinum and paclitaxel or nab-paclitaxel)
		• Variability in defining a frontline regimen may be acceptable depending upon clinical practice variation that may exist, particularly if using internationally sourced data, while minimizing misclassification of exposure.
		• Consider whether granularity regarding the exact drug or class of drug is important.
		• Consider how distinguish monotherapy vs combination therapy

Clinical characterization of the study population

Smoking status	• Patient's smoking status as documented at any point prior to the data cutoff for this study:
	  o History of smoking
	  o No history of smoking
	  o Unknown/not documented
		• A more granular definition such as smoking status at diagnosis or at frontline treatment initiation (index date) may be a better barometer.
		• "History of smoking" does not distinguish between current vs. former smoking status.

Group stage	• Stage of disease at the time of initial diagnosis with NSCLC:
	  o 0
	  o I
	  o II
	  o III
	    - IIIa
	    - IIIb
	    - IIIc
	  o IV
	  o Group stage is not reported	

PD-L1 tested (before or 30 days after the index date)	• Tested for PD-L1
	  o Tested
	  o Untested
	
	• **NOTE:** Testing may occur at any point before or up to the index date, defined previously.
	  o Where available: the test date will be identified as the most recent date available across the "specimen collected" date, "specimen received" date, and "result date" variables.	

PD-L1 status (before or 30 days after the index date)	• Result for PD-L1 test among those with documented testing
	  o PD-L1 positive
	  o PD-L1 negative/not detected
	  o PD-L1 equivocal
	  o No interpretation given in report
	  o Results pending/Unknown
	
	• Testing may occur at any point before or up to the index date.
		• Alignment necessary for testing time-frame
		• Consider study objectives when defining look-back and cut-off dates.
		• This result is based on test interpretation as reported to the physician, which may consist of differing cut-offs.

PD-L1 staining (before or 30 days after the index date)	• Staining level result for PD-L1 test among those with documented testing
	  o <1%
	  o 1% - <50%
	  o ≥50%
	  o Unknown
	
	• **NOTE:** Testing may occur at any point before or up to the index date, defined previously.
		• Alignment necessary for testing time-frame
		• Consider study objectives when defining look-back and cut-off dates.
		• Unstructured data may not be widely available.
		• Greater precision is available but extent missingness should be addressed.

Organ function at index	• Patient's renal/hepatic function at the index date, as previously described, based on structured lab data	• Alignment necessary for testing time-frame and definition of severity categories.
	o Severe renal/hepatic failure	• Consider study objectives when defining look-back and cut-off dates.
	o Moderate renal/hepatic failure	• If data from more than one lab test are available for a given individual, use the most recent value.
	o Normal renal/hepatic function	
	NOTE: Restricted to patients with creatinine serum lab values for renal function or total bilirubin, aspartate aminotransferase (AST), or alanine transaminase (ALT) for hepatic function up to 30 days before the index date	
	o Categorization of renal function defined as:	
	o Severe: >3x upper limit of normal	
	o Moderate: 1.5-3x upper limit of normal	
	o Mild: >ULN-1.5x ULN	
	o Normal: ≤ULN	
	o Categorization of hepatic function defined as:	
	- Severe defined by one of the following:	
	• Total bilirubin >3x upper limit of normal	
	• AST >5x upper limit of normal	
	• ALT >5x upper limit of normal	
	- Moderate defined by one of the following:	
	• Total bilirubin 1.5-3x upper limit of normal	
	• AST 3-5x upper limit of normal	
	• ALT 3-5x upper limit of normal	
	- Mild defined by one of the following:	
	• Total bilirubin >ULN-1.5x ULN	
	• AST >ULN-3x ULN	
	• ALT >ULN-3x ULN	
	- Normal defined by meeting none of the above criteria	

Field	Definition	Considerations

Presence/ab-sence of chronic organ disease (ICD9 code)	• Defined as at least one diagnostic code any time prior to and including the index date, defined as having one of the following ICD9/10 codes: [N18.x, N19.x.] or [585.X, 586.X] for kidney disease or [K70-K77] or [570.X, 571.X, 572.X, 573.X] for liver disease.
	
	   o Yes
	
	   o Unknown
		• Alignment of ICD codes required
		• Limited in ability to distinguish no organ disease vs no evidence of organ disease.
		• See SAP for complete list of ICD codes used.

Performance status (ECOG) at index	• Patient's ECOG status at the time of the index date, defined previously
	
	   o 0
	
	   o 1
	
	   o 2+
	
	   o Unknown
	
	• **NOTE:** ECOG may have been recorded up to 30 days prior to the index date, OR up to 7 days after the index date, whichever is closest to the index date. If there are multiple ECOG values at the same absolute distance from the index date, priority is given to the ECOG value that precedes the index date. For patients with multiple ECOG values recorded on the same day, the highest value will be selected.
		• Alignment necessary for time-frame
		• Consider study objectives when defining look-back and cut-off dates.

CNS Metastases	• Defined as at least one diagnostic code up to 30 days after the index date.
	• Complete list of ICD codes was created.
	• Secondary CNS neoplasm codes:
	
	   o ICD9 codes: 198.3, 198.4/
	
	   o ICD10 codes: C79.31, C79.32, C79.49.
	
	• Primary malignant neoplasm codes should be included if they occur after index because the likelihood of a CNS primary after a metastatic NSCLC is very unlikely and miscoding is common. Primary CNS neoplasm codes: ICD9: 191.-191.9/ ICD10: C71.0-C71.9
		• Alignment necessary
		• Consider structured ICD codes.
		• See SAP for complete list of ICD codes used.

Endpoints

Field	Definition	Considerations
Date of death	• rwOS event date	• Alignment necessary • Consider minimum standards for data completeness
Start date of regimen after frontline	• Start date of regimen immediately after frontline (i.e., second-line) • **NOTE:** Patients with a death prior to having another line will be considered as having an event • rwTTNT event date for endpoints of regimen after frontline [i.e., second-line] or death	

Data element	Definition	Considerations
Last confirmed activity date	• Patient's last known structured recorded clinical activity • For calculation of structured follow-up time • rwTNT or rwOS censor date	• Definition variability appropriate. Separate definitions required for EHR-based vs claims-based data sources. • Various types of patient activity recorded in different datasets (e.g. structured visits, labs, abstracted dates) and their applicability for use as censoring dates for the chosen endpoint, should be considered
Frontline discontinuation date	• Date of frontline regimen discontinuation • rwTTD event date	
Frontline last continuing date	• Date of last known frontline regimen when there is no frontline discontinuation (i.e., still on frontline therapy) at the data cutoff • Or Last observed administration date of frontline. • rwTTD censor date	
Additional Data Elements for Consideration		
Age at advanced diagnosis	• Age at advanced diagnosis (continuous)	
Ethnicity	• Hispanic • Non-Hispanic • Unknown/Missing	• Same considerations as for race

Median income (quartile)	• Median household income (zip-level quartiles)
	  o 1 (lowest median household income)
	  o 2
	  o 3
	  o 4 (highest median household income)
	  o Unknown	

Other Biomarkers	• ALK/EGFR	• Consider clinical implications of additional biomarkers as exclusion criteria

Overall survival (OS) in advanced non-small cell lung cancer (aNSCLC) patients treated with frontline chemotherapy or immunotherapy by comorbidity: A real-world data (RWD) collaboration.

Donna R. R Rivera, Laura Lasiter, Jennifer Christian, Lindsey Enewold, Janet L. Espirito, Eric Hansen, Henry J. Henk, Lawrence H. Kushi, Daniel Lane, Yanina Natanzon, Ruth Pe Benito, Erik Rasmussen, Nicholas J. Robert, Mark Stewart, Connor Sweetnam, Olga Tymejczyk, Emily Valice, Joseph Wagner, Alia Zander, Jeff Allen

National Cancer Institute, Rockville, MD; Friends of Cancer Research, Washington, DC; IQVIA, Research Triangle Park, NC; National Cancer Institute, Bethesda, MD; McKesson Life Sciences, The Woodlands, TX; COTA, Inc., New York, NY; OptumLabs, Shakopee, MN; Kaiser Permanente, Oakland, CA; COTA Healthcare, Boston, MA; Syapse, San Francisco, CA; Tempus, Chicago, IL; Flatiron Health, New York, NY; 10101 Woodloch Forest, The Woodlands, TX; Syapse Inc., San Francisco, CA; Kaiser Permanente Division of Research, Oakland, CA; IQVIA, Plymouth Meeting, PA

Abstract

Background: Friends of Cancer Research convened 9 data partners to identify data elements and common definitions for real world (rw) endpoints to evaluate populations typically excluded from clinical trials. Here we report on rwOS by frontline treatment and comorbidities. **Methods:** A retrospective observational analysis of patients with aNSCLC initiating frontline platinum doublet chemotherapy (chemo) or PD-(L)1-based immuno-oncologic (IO) therapy (monotherapy or chemo combination) between 1 Jan 2011 to 31 Mar 2018 was conducted using administrative claims, EHR, and cancer registry RWD. We evaluated rwOS from frontline therapy initiation using Kaplan-Meier methods, stratified by ECOG status, brain metastases (ICD), history of chronic kidney or liver disease (CKD/ CLD, ICD), and evidence of kidney or liver dysfunction (KD/ LD, lab-based). **Results:** A total of 33,649 patients were included (N 972-17,454) with 10 to 26% of patients receiving IO as frontline therapy. There was a broad range of comorbidity prevalence across datasets and patients with evidence of comorbidity had comparatively shorter 12-month OS (Table). **Conclusions:** RWD analyses can generate expanded evidence on patient outcomes for populations routinely excluded from clinical trials and may help inform decision making where sparse data exist on appropriate treatment approaches. Additional understanding of data missingness, sensitivity of definitions, and covariate adjustment are needed to make direct comparisons across regimens and data sources.

Trends in immunotherapy use in patients with advanced non-small cell lung cancer (aNSCLC) patients: Analysis of real-world data.

Lawrence H. Kushi, Laura Lasiter, Andrew J. Belli, Marley Boyd, Suanna S. Bruinooge, Jennifer Christian, Elizabeth Garrett-Mayer, Eric Hansen, Rebecca Honnold, Ruth Pe Benito, Yanina Natanzon, Lori Sakoda, Donna R. R Rivera, Whitney Rhodes, Nicholas J. Robert, Elad Sharon, Connor Sweetnam, Joseph Wagner, Mark S. Walker, Jeff Allen

Kaiser Permanente, Oakland, CA; Friends of Cancer Research, Washington, DC; COTA, Inc., New York, NY; McKesson Life Sciences, The Woodlands, TX; American Society of Clinical Oncology, Alexandria, VA; IQVIA, Research Triangle Park, NC; Tempus Labs, Chicago; Tempus, Chicago, IL; Syapse, San Francisco, CA; Kaiser Permanente Northern California, Division of Research, Oakland, CA; National Cancer Institute, Rockville, MD; Concerto HealthAI, Memphis, TN; 10101 Woodloch Forest, The Woodlands, TX; National Cancer Institute, Bethesda, MD; Syapse Inc., San Francisco, CA; IQVIA, Plymouth Meeting, PA; ACORN Research LLC, Memphis, TN

Abstract

Background: Leveraging data from a collaboration with 9 data partners, Friends of Cancer Research convened the Real-world Evidence Pilot 2.0, to examine trends and real world (rw) data endpoints in immunotherapy (IO) use for the front line treatment of aNSCLC. **Methods:** This study leveraged parallel analyses of rw data elements across heterogenous data sources (EHR, administrative claims, and registry) to: a) describe trends in uptake and use of novel IO frontline therapy after advanced diagnosis in NSCLC patients treated in usual care settings and b) examine associations between treatment and rw outcomes at one-year follow-up. The proportion of patients treated on each regimen (IO single agent, chemo, or IO + chemo) from 2011 through 2017 were calculated. Analysis included proportion of patients across treatment regimens stratified by year to describe post approval uptake of IO. Kaplan-Meier survival estimates were reported to adjust for follow-up time and stratified by PD-L1 status and stage. **Results:** Seven datasets identified a range of 999 to 4617 patients per dataset for this analysis. Across datasets, 2508, 3446, and 4176 patients initiated treatment in 2015, 2016, and 2017, respectively. No patients received IO or IO + chemo regimens prior to 2015. Initial approvals for IO use in aNSCLC occurred in October 2015 and for first line in metastatic NSCLC in October 2016. When examining survival at 1 year, overall, OS in PD-(L)1 + patients appeared longer than those with a PD-(L)1 - status. **Conclusions:** RWE analyses may reveal important trends in clinical cancer patient care including patterns of off-label use. The heterogeneity in the timing of IO uptake across datasets ranged from immediately after approval to ~12 months post-approval.

COMMENTARY

Use of Patient-Reported Outcomes to Understand & Measure the Patient Experience of Novel Cell and Gene Therapies

Laura Lasiter, PhD[1] · Alicyn Campbell, MPH[2] · Ethan Basch, MD[3] · Stacie Hudgens, MSc[4] · Mark Stewart, PhD[1] · James J. Wu, MSc, MPH[5] · Allison Barz Leahy, MD[6,7] · Jeff Allen, PhD[1]

Abstract

Patient reported outcomes (PROs) are the gold standard for assessing patients' experience of treatment in oncology, defined in the 21st Century Cures Act as information about patients' experiences with a disease or condition, including the impact of a disease or condition, or a related therapy or clinical investigation on patients' lives; and patient preferences with respect to treatment of their disease or condition [1]. PROs provide a comprehensive assessment of the benefits and risks of new medical products, as well as essential data to inform real-world use. Although RCTs are the ultimate source for information for evaluating products in development, they are not always feasible for rare diseases with few or no effective treatment options available. Thus, it is important to consider other measures that can help to improve the strength of evidence for cell and gene therapies targeting rare indications. While collection of PROs and other patient experience endpoints does not resolve the difficulty of conducting trials in small populations, doing so contributes empirical evidence that informs both product development and patient access. Additionally, including routine collection of PROs in registries may provide supplemental data to further characterize the benefit:risk profile of cell and gene therapies at follow-up times that would be infeasible to operationalize in a clinical trial setting.

Introduction: Cell & Gene Therapies

Therapies derived from human cells and genes are providing novel treatment options for patients with life-threatening conditions. Gene therapies seek to modify a patient's genes to treat or cure disease. The transferred genetic material changes how a single protein or group of proteins is produced by the cell. Gene therapy can be used to reduce levels of a disease-causing version of a protein, increase production of disease-fighting proteins, or to produce new/modified proteins. Cell therapy is the transfer of intact, live cells into a patient to help lessen or cure a disease. Cell therapies alter the biological properties of living cells, either a patient's own cells as in autologous cell therapies, or from a donor as in allogeneic cell therapies, for therapeutic use. The cells used in cell therapy can be classified by their potential to transform into different cell types. Though cell and gene therapies have different mechanisms of action, the US FDA regulates both treatment modalities as gene therapies.

Increasing the Empirical Evidence Base for Novel Cell and Gene Therapies

Both private and government payers have begun to recognize the value that PROs add to the evidence base for new therapies. Most recently, and perhaps most significantly, the Center for Medicare & Medicaid Services (CMS) sought input from an independent advisory committee, the Medicare Evidence Development & Coverage Advisory

✉ Laura Lasiter
llasiter@focr.org

1 Friends of Cancer Research, 1800 M St. NW Suite 1050S, Washington, DC 20036, USA

2 Patient Relevant Evidence, San Francisco, USA

3 University of North Carolina at Chapel Hill, Chapel Hill, USA

4 Clinical Outcomes Solutions, Tucson, USA

5 Amgen, Inc, Thousand Oaks, USA

6 University of Pennsylvania, Philadelphia, USA

7 Children's Hospital of Philadelphia, Philadelphia, USA

✿ Springer

Committee (MEDCAC), on how to incorporate existing PRO assessment tools into future clinical studies, specifically for new classes of therapies such as Chimeric Antigen Receptor T cell (CAR-T) Therapies. The 2017 FDA approvals of tisagenlecleucel and axicabtagene ciloleucel, the first two CAR-T therapies approved for cancer indications in the USA, had created a new class of commercially available cell and gene therapies and, importantly, a potential unmet need for payer review and guidance due to the expected curative benefit and likely high costs associated with both approved products. The MEDCAC meeting was convened as part of the May 2018 announcement that CMS would conduct a National Coverage Determination (NCD) for CAR-T used to treat advanced cancer in Medicare patients [2, 3]. As autologous cell therapies (including CAR-Ts and also emerging technologies such as TCR-based therapies) are individualized per patient, robust clinical trial data are difficult to obtain. These challenges are amplified among Medicare patients, who by simple life expectancy may not experience the same duration of survival as younger patients. CMS was interested in how PRO assessment tools could support health outcomes research and, consequently, coverage determinations following the approval of the first CAR-T therapies. MEDCAC panel unanimously recommended inclusion of PROs in the NCD.

The NCD for CAR-T was ultimately published on August 7, 2019, without the requirement for PRO collection as part of a larger administrative effort to recognize the significance of the curative potential of these new treatments and to encourage broad access to them (coverage with evidence development and collection of PROs were removed to avoid any potential burden placed on providers created by reporting requirements) [4]. However, the inclusion of PRO collection in the proposed NCD by CMS did signal a recognition of the importance and usefulness of PROs to enhance the empirical evidence base and our understanding of the long-term value for cell and gene therapies.

Recommendations

A multi-stakeholder group convened by Friends of Cancer Research (*Friends*), in response to the proposed NCD, met regularly in late 2018/early 2019 to consider inclusion of PROs as a factor in coverage decisions for CAR-T therapies, particularly where they pertain to breakthrough designated therapies and where investigational CAR-Ts have the potential to significantly improve health-related quality of life. This group of recognized subject-matter experts in their field included clinicians, academics, and industry representatives with extensive expertise in PROs and/or CAR-T clinical trials, and deep understanding of the requirements for US and EU drug approval applications. The collective expertise of

the working group members, supported by available literature, were leveraged for the development of a PRO collection framework, focused on the core concepts of interest most relevant to patients under treatment, that could inform future payer decisions. While developed with respect to the recent CAR-T approvals and announcement of a NCD by CMS, this framework is expected to be applicable for evidence development across cell and gene therapies, and in rare diseases with moderate survival expectations.

PRO Assessment of CAR-T Therapies

Key data elements for PRO assessment are described in Table 1. PRO measures should be selected that are most appropriate to address relevant questions at the applicable timepoint related to first dose. We suggest this be divided into 3 phases: acute, sub-acute, and long term (Fig. 1). Inclusion of the acute phase collection is vital, since a lag in PRO collection after treatment initiation will miss immediate toxicity associated with the acute phase of treatment (neurological toxicity and cytokine release syndrome). We encourage consistent collection before, during and immediately following active treatment to most accurately assess the patient experience during the acute and sub-acute phases. Further, because of the curative expectations of cell therapies, multi-year follow-up should be considered to capture long-term events associated with CAR-Ts and other cell and gene therapies (FDA approval for both CAR-Ts included 15 year follow-up post-market requirements) and to assess long-term health-related quality of life; a metric which will be of increasing importance as these therapies become a new treatment paradigm [5–8]. As such, we encourage long-term follow-up timepoints, a minimum of 5 years, such that it aligns with the long-term patient experience and timeframe for projected efficacy benefits and sponsor regulatory commitments for surveillance. For monitoring of late toxicities, an approach in which immediate post-treatment questionnaires are administered monthly during the initial 6 months, then spaced out every 6 months for three years, and then annually, is consistent with commonly used approaches [9, 10]. In contrast, during active therapy, weekly PRO collection is more appropriate. For all phases of PRO assessment quantitative, rather than qualitative assessment is standard and less subject to heterogeneity or bias where multiple interviewers may be involved in administration over time.

To enable researchers to systematically include the concepts of interest in a standardized manner, Table 2 was constructed which lists potential cell and gene therapy side effects, their timing in the course of treatment, and their corresponding well-defined measurement system and scoring to facilitate the integration of these tools into research in a consistent manner [11–16]. Tools proposed by MEDCAC and most frequently used within sponsor trials were included.

Table 1 Key Data Elements to Assess Chimeric Antigen Receptor (CAR) T Cell Therapy and Patient-Reported Outcomes.

Source	Clinical Outcome	Utility of Elements	Timing of Collection
Efficacy data	Response rate (RR) Progression free survival (PFS) Overall survival (OS)	Need to demonstrate efficacy using well recognized end-points	Pre-approval Pre-approval Post-market
Clinician-derived Safety/tolerability	Common Terminology Criteria for Adverse Events (CTCAE)	Traditionally used signals of adverse reactions/tolerability; reported by the clinician/healthcare professional; remain important for determining tolerability and should continue to be routinely captured	Pre-approval and post-market
Patient-derived Adverse event data	Symptomatic adverse events Global side effect impact/bother/burden Global health status	Suitable PRO tools should be selected that capture patient-derived data concerning the impact of the adverse events of the therapy and the overall treatment burden for the patient	Pre-approval and post-market
Additional supportive patient-derived data	Physical function Other functional domains (e.g., emotional, social, role, cognitive) Specific key symptoms of disease (e.g., pain, fatigue, nausea, vomiting, anorexia) as single items or composite scale	Depending on the objectives of the study and the type and intensity of therapy (including known adverse events of special interest), other elements may contribute to defining the tolerability of a treatment regimen	Pre-approval and post-market
Healthcare utilization	Hospitalization rates/duration Emergency department visits Supportive care medication use	These items may provide a more holistic healthcare view of the tolerability of a treatment for a patient and may help determine the requirements for managing medical needs	Post-market

Table adapted from a Friends of Cancer Research whitepaper: https://www.focr.org/sites/default/files/Comparative%20Tolerability%20Whitepaper_FINAL.pdf.
CAR, chimeric antigen receptor; CTCAE, Common Terminology Criteria for Adverse Events; PFS, progression free survival; RR, response rate; OS, overall survival.
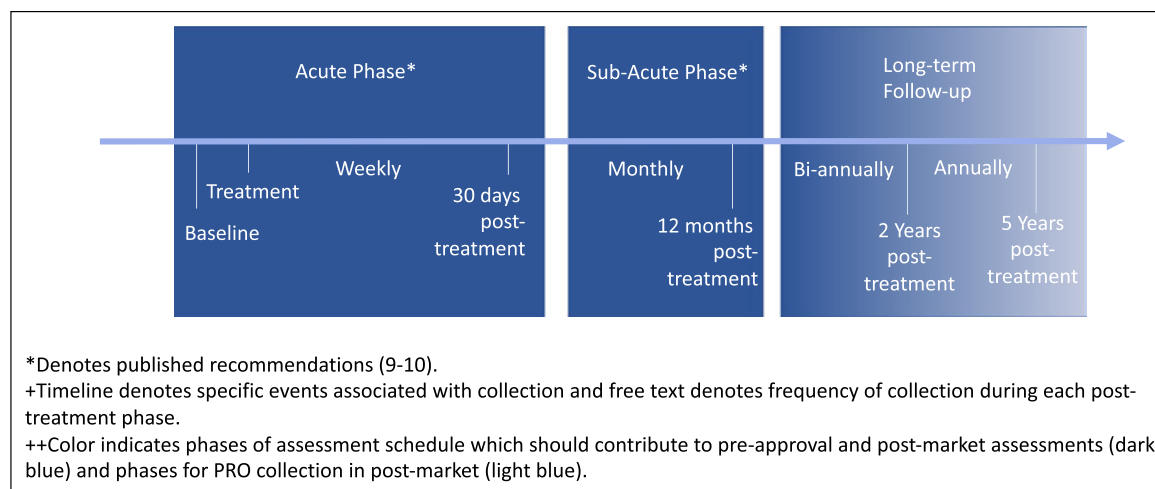
Figure 1 Assessment Schedule for PRO Collection.+, ++.

The Patient-Reported Outcomes Measurement Information System (PROMIS), another of the MEDCAC recommended tools, is used for monitoring patient physical, mental, and social well-being. Given that it can be administered via computer-adapted technologies and its extensive library of items, it was not included here.

Data Collection Infrastructure

Given that CAR-T administration is limited to a select number of specialized clinical locations by the Risk Evaluation and Mitigation System required by FDA, PRO data reporting is expected to be relatively straightforward during the acute phase of treatment. However, the extended assessment periods recommended in this commentary will expand data reporting requirements into different care settings. Researchers, CMS, and other payers will need to be mindful when developing methodologies and policies to account for potential disruptions in data collection as patients transition from hospital inpatient to out-patient settings or from academic medical centers back into routine care as a standard infrastructure to seamlessly collect this data from clinic to routine care is currently lacking. Oncologists in routine practice, including standard practice and community oncology practices, are less likely to have experience with PRO collection and fewer resources to devote to administration of PRO instruments. There are a variety of third-party vendors and real-world evidence suppliers to support the extended assessment requirements and reduce financial and resource burdens placed on practices in those settings and increase collection compliance. When assessing an appropriate vendor, availability of patients, sites that are accessible in the third-party system, the comprehensiveness of the clinical

record (e.g., clinical outcomes), integration of patient facing symptom collection capabilities, quality of the design of the use experience, and impact of symptom collection processes on the health system practice, and analytic capabilities are key factors to consider. Mobile health monitoring and electronic data collection (ePRO) should also be encouraged as it may facilitate real-time monitoring of compliance for backup data collection and easy data transfer. The ability to utilize ePRO will be particularly important for collecting data on late toxicities or as patients are transferred to out-patient settings. Further, ePRO systems may provide more flexible platforms for customizable data collection, though not necessarily appropriate for all studies or patient populations and additional cost may be a complication to their use [17]. A number of vendors offer stand-alone ePRO software for data collection, and increasingly electronic health record (EHR) systems and real-world evidence companies offer tools for collection of PRO data within the EHR or accompanying patient portals. Patient registries are identified as another source of real-world data that can be used to generate RWE in the Framework for FDA's Real-World Evidence Program and are particularly useful in data collection for rare diseases [18, 19]. Looking forward, more integration of standard patient outcomes into the medical record and routine care not only for cell and gene therapies but across also cancer treatments should be a priority for both regulator, payers, providers and health systems.

Use of PROs to Inform Coverage Determinations

Considering that cell therapies, such as CAR-Ts, are expected to extend median overall survival (OS) by a number of years, well beyond the usual scope of a clinical trials,

Table 2 CAR-T Applicable PROs and Their Representation in MEDCAC-Approved Tools.

Concepts	Symptom	Acute	Sub-acute	Long Term (1 + years)	Item	Response Format, Basic Scoring
Adverse events (AE)/toxicity						
AE/GI	Nausea	X	X		EORTC: Have you felt nauseated?	4 point (pt) Likert scale, scored as single item
					MDASI: Your nausea at its WORST?	11 point NRS
					PRO-CTCAE: Nausea	F(Frequency), S (severity)
	Vomiting	X	X		EORTC: Have you vomited?	4 pt Likert scale, scored as single item
					MDASI: Your vomiting at its WORST?	11 point NRS
					PRO-CTCAE: Vomiting	F, S
	Diarrhea	X	X		EORTC: Have you had diarrhea?	4 pt Likert scale, scored as single item
					PRO-CTCAE: Diarrhea	F
					MDASI: Your diarrhea at its WORST?	11 point NRS
	Constipation	X	X		EORTC: Have you been constipated?	4 pt Likert scale, scored as single item
					PRO-CTCAE: Constipation	S
					MDASI: Your constipation at its WORST?	11 point NRS
	Anorexia	X	X		EORTC: Have you lacked appetite?	4 pt Likert scale, scored as single item
					MDASI: Your problem with lack of appetite at its WORST?	11 point NRS
					PRO-CTCAE: Decreased appetite	S, I (Interference)
AE/CRS	Fever, Chills	X	X		PRO-CTCAE:	F, S
					Chills	F, S
					Increased sweating	F, S
					Hot flashes	F, S
					Heart palpitations	11 point NRS
					MDASI: Your fever or chills at its WORST?	11 point NRS
					Your feeling of malaise (not feeling well) at its WORST?	
	Edema	X	X		EORTC: Have you experienced any swelling in certain parts of your body (e.g., ankles, legs or around your eyes)?	4 pt Likert scale, scored as single item
					PRO-CTCAE: Swelling	F, S, I
					MDASI:	1 point NRS
					Your swelling of your hands, legs, feet, abdomen, or around your eyes at its WORST?	11 point NRS
					Your problem with ankle swelling at its WORST?	
AE/Constitutional	Fatigue	X	X	X	EORTC: Were you tired?	4 pt Likert scale, scored as single item
					MDASI:	
					Your fatigue (tiredness) at its WORST?	11 point NRS
					Your problem with lack of energy at its WORST?	11 point NRS
					PRO-CTCAE: Fatigue	S, I

Table 2 (continued)

Concepts	Symptom	Acute	Sub-acute	Long Term (1 + years)	Item	Response Format, Basic Scoring
	Myalgia	X	X		EORTC: Have you felt weak? Have you had pain? Did pain interfere with your daily activities? PRO-CTCAE: Muscle pain MDASI: Your muscle weakness at its WORST? Your muscle soreness or cramping at its WORST?	4 pt Likert scale, scored as single item F, S, I 11 point NRS 11 point NRS
	Arthralgia	X	X		EORTC: Have you had pain? Did pain interfere with your daily activities? PRO-CTCAE: Joint Pain MDASI: Your joint stiffness or soreness at its WORST?	4 pt Likert scale, scored as single item F,S,I 11 point NRS
AE/CNS	Headache	X	X		PRO-CTCAE: Headache MDASI: Your headache at its WORST?	F,S,I 11 point NRS
	Tremor	X	X			
	Dizziness	X	X		PRO-CTCAE: Dizziness	S, I
	Confusion	X	X		EORTC: Have you had difficulty remembering things? PRO-CTCAE: Concentration Memory MDASI: Your problem with remembering things at its WORST?	4 pt Likert scale, scored as single item S,I S,I 11 point NRS
	Aphasia	X	X		EORTC: Have you had trouble finding the right words to express yourself? Did you have difficulty speaking? Did you have trouble communicating your thoughts? MDASI: Your difficulty speaking (finding the words) at its WORST?	4 pt Likert scale, scored as part of a 3 item communication deficit sub-scale 11 point NRS
	Insomnia	X	X		EORTC: Have you had trouble sleeping? PRO-CTCAE: Insomnia MDASI: Your disturbed sleep at its WORST?	4 pt Likert scale, scored as single item S, I 11 point NRS

Table 2 (continued)

Concepts	Symptom	Acute	Sub-acute	Long Term (1 + years)	Item	Response Format, Basic Scoring
	Anxiety	X	X		EORTC: Did you feel irritable? Did you feel depressed? Did you feel tense? Did you worry? MDASI: Your feeling of being distressed (upset) at its WORST? Your feeling sad at its WORST? PRO-CTCAE: Anxious Discouraged Sad	4 pt Likert scale, scored as single item 11 point NRS 11 point NRS F, S, I F, S, I F, S, I
AE/respiratory	Dyspnea	X	X		EORTC: Were you short of breath? three item scale: [1] Were you short of breath when you rested? [2] Were you short of breath when you walked? [3] Were you short of breath when you climbed stairs? PRO-CTCAE: Shortness of breath MDASI: Your shortness of breath, at its worst?	4 pt Likert scale, scored as single item 4 pt Likert scale, scored as multi-item sub-scale S, I 11 point NRS
	Cough	X	X		EORTC: How much did you cough? PRO-CTCAE: Cough MDASI: Your coughing at its WORST?	4 pt Likert scale, scored as single item. S, I 11 point NRS
General symptom (disease or treatment)	Pain	X	X	X	EORTC: Have you had pain? Did pain interfere with your daily activities? PRO-CTCAE: General pain Muscle pain Joint pain MDASI: Your pain at its WORST?	4 pt Likert scale, scored as single item. F, S, I F, S, I F, S, I 11 point NRS

Table 2 (continued)

Concepts	Symptom	Acute	Sub-acute	Long Term (1 + years)	Item	Response Format, Basic Scoring
Physical function		X	X	X	(EORTC) Do you have trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase? Do you have any trouble taking a *long* walk? Do you have any trouble taking a *short* walk outside of the house? Do you need to stay in bed or a chair during the day? Do you need help with eating, dressing, washing yourself or using the toilet? MDASI Interference scale[a] items: Walking Activity Working (including housework) Relations with other people Enjoyment of life Mood	4 pt Likert scale, scored as single sub-scale 11 point NRS
Instrumental activities of daily living (IADLs)		X	X	X	Role Function (EORTC) Were you limited in doing either your work or other daily activities? Were you limited in pursuing your hobbies or other leisure time activities? MDASI Interference scale[a]	4 pt Likert scale, scored as single sub-scale 11 point NRS
Social functioning		X	X	X	EORTC: Has your physical condition or medical treatment interfered with your *family* life? Has your physical condition or medical treatment interfered with your *social* activities? MDASI Interference scale[a]	4 pt Likert scale 11 point NRS
Financial		X	X	X	Has your physical condition or medical treatment caused you *financial* difficulties?	4 pt Likert scale

As part of this work, the PRO subject-matter experts on the multi-stakeholder group drafted this comprehensive table mapping concepts relevant to CARTs, cell and gene therapies to their respective and most commonly used PRO tools also endorsed by CMS for evidence generation (e.g., PRO-CTCAE, EORTC, and MDASI).

[a]The items from the interference scale capture interference with daily living caused by these symptoms.

AE, adverse events; CNS, central nervous system; CRS, cytokine release syndrome; EORTC, European Organization for Research and Treatment of Cancer; F, frequency; GI, gastrointestinal; I, interference; IADLs, instrumental activities of daily living; MDASI, MD Anderson Symptom Inventory; NRS, numeric rating scale; PRO-CTCAE, Patient-Reported Outcomes-Common Terminology Criteria for Adverse Events; pt, point; S, severity.

it will be essential to support ongoing evaluation of these products in the post-market setting. Ongoing evaluation will be particularly relevant as the current approved CAR-T products rapidly expand their labels into new indications and patient populations. Tisagenlecleucel-t, first approved in August 2017 for relapsed and refractory pediatric and young adult patients with B cell acute lymphoblastic leukemia (ALL), received a May 2018 label expansion for use in diffuse large-B cell lymphoma, a type of non-Hodgkin's lymphoma occurring most frequently in individuals over 60 years of age. PROs will be a metric to consider in this ongoing surveillance to ensure a holistic approach to performance evaluations, including:

1. Determining appropriate patient populations;
2. Expanding indications;
3. Considering new care settings, and;
4. Informing long-term value.

PROs will be metrics for informing coverage as new patient populations and new uses of CAR-T cell therapies are identified. However, since the long-term effects of CAR-T cell therapies are expected to last years, well beyond the acute phase as the patient transitions back into routine care, the impact of treatment setting will also be relevant to studying long-term effects and should be included as a consideration in PRO collection requirements. Therefore, we support the collection of PRO information in both the out-patient and inpatient settings and in routine practice. In addition, given the novelty of these therapies, we believe that the REMS restrictions provide a unique opportunity to collect clinician reported and patient-reported outcomes, as the conclusions from each may not necessarily be the same. Further, special considerations will be needed due to the individualized nature of CAR-T-therapy manufacturing. Specifically, little is currently being collected to assess the potential impact that product or process changes have on health outcomes. PRO data could be used to monitor technology progression for changes in patient experience and health outcomes that may not otherwise be identifiable due to the extended timeline of surveillance and coordinated data-sharing infrastructure needed. As the manufacturing process is improved and yields safer, more efficacious therapies and evidence evolves to demonstrate the value of cellular therapy over other treatment options, this may change patient/provider calculations and shared medical decision-making.

Conclusions

Integration of this framework into coverage decisions will require identification of key PRO measures for value

assessments and development of standards to ensure they are uniformly collected and obtained from well-defined PRO instruments. Once quality standards and methods for integrating PRO measures with traditional clinical trial measures have been developed, PROs can be systematically included to increase the evidence base for these new therapies [20]. Additionally, PRO measures provide data that are integral to comprehensive disease modeling using health-related quality of life (HRQoL) metrics and values. Efforts to improve disease modeling for CAR-T therapies are ongoing; these models risk inaccuracy without the ability to include long-term toxicity and HRQoL data. When incorporating PROs, it will be important to delineate the research objective and endpoints in the study, as this will affect the PRO tools selected and the methodology employed.

Notes

No data, models, or methodology used in this manuscript are proprietary.

The publication of study results was not contingent on the sponsor's approval or censorship of the manuscript.

Author Contributions
All authors were involved in the conception, drafting & revising, and approval processes for this commentary. Each author agrees to be accountable for all aspects of the work.

Compliance with Ethical Standards

Conflict of interest
Dr. Basch reports grants from the National Cancer Institute and the Patient Centered Outcomes Research Institute, outside the submitted work; and Dr. Basch is on the Editorial Board of the Journal of the American Medical Association and consults on research projects at Memorial Sloan Kettering Cancer Center, Dana Farber Cancer Institute, Centers for Medicare & Medicaid Services, and Research Triangle Institute. Dr. Basch is a Scientific Advisor for Sivan Healthcare, Care-Vive Systems, and Navigating Cancer. Authors Laura Lasiter, Alicyn Campbell, Stacie Hudgens, Mark Stewart, James J. Wu, Allison Barz Leahy, and Jeff Allen have nothing to disclose.

Springer

References

1. Pubic Law 114-255: 21st Century Cures Act. (130 STAT. 1034; 12/13/2016; H.R. 34). Retrieved from https://www.Congress.gov.
2. Centers for Medicare and Medicaid. 2018. August 22). MEDCAC Meeting – Chimeric Antigen Receptor (CAR) T-cell Therapy and Patient Reported Outcomes. Retrieved from https://www.cms.gov/medicare-coverage-database/details/nca-proposed-decision-memo.aspx?NCAId=291.
3. Centers for Medicare and Medicaid. 2018. National Coverage Analysis Tracking Sheet for Chimeric Antigen Receptor (CAR) T-cell Therapy for Cancers (CAG-00451 N). Retrieved from https://www.cms.gov/medicare-coverage-database/details/nca-tracking-sheet.aspx?NCAId=291.
4. Centers for Medicare and Medicaid. 2019. Decision Memo for Chimeric Antigen Receptor (CAR) T-cell Therapy for Cancers (CAG-00451 N). Retrieved from https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=291.
5. Food and Drug Administration. 2017, October 18. FDA approves CAR-T cell therapy to treat adults with certain types of large B-cell lymphoma [Press release]. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-approves-car-t-cell-therapy-treat-adults-certain-types-large-b-cell-lymphoma.
6. Food and Drug Administration. 2017. Summary Basis for Regulatory Action Template- Yescarta. Retrieved from https://www.fda.gov/media/108788/download.
7. Food and Drug Administration. 2017, August 30. FDA approval brings first gene therapy to the United States [Press release]. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-approval-brings-first-gene-therapy-united-states.
8. Food and Drug Administration. 2018. Summary Basis for Regulatory Action Template- Kymriah. Retrieved from https://www.fda.gov/files/vaccines%2C%20blood%20%26%20biologics/published/April-13–2018-Summary-Basis-for-Regulatory-Action—KYMRIAH.pdf.
9. Chakraborty R, Sidana S, Shah GL, Scordo M, Hamilton BK, Majhail NS. Biol blood marrow transplant. Patient-reported outcomes with chimeric antigen receptor T cell therapy: challenges and opportunities. Biol Blood Marrow Transplant. 2019;25(5):155–62. https://doi.org/10.1016/j.bbmt.2018.11.025.
10. Ruark J, Mullane E, Cleary N, Cordeiro A, Bezerra ED, Wu V, et al. Patient-reported neuropsychiatric outcomes of long-term survivors after chimeric antigen receptor T cell therapy. Biol Blood Marrow Transplant. 2019;26(1):34–43. https://doi.org/10.1016/j.bbmt.2019.09.037.
11. Aaronson N, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez N, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365–76.
12. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, et al. EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: EORTC; 2001.
13. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014. https://doi.org/10.1093/jnci/dju244.
14. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. 2015 Nov. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). National Cancer Institute PRO-CTCAE Study Group. JAMA Oncol. 2015 Nov;1(8):1051-9. https://doi.org/10.1001/jamaoncol.2015.2639. Erratum in: JAMA Oncol. 2016 Jan;2(1):146.
15. Kluetz PG, Chingos DT, Basch EM, Mitchell SA. Patient-reported outcomes in cancer clinical trials: measuring symptomatic adverse events with the national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Am Soc Clin Oncol Educ Book. 2016;35:67–73. https://doi.org/10.14694/EDBK_159514.
16. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients: the M. D. Anderson Symptom Inventory. Cancer. 2000;89:1634–46.
17. Schwartzberg L. Electronic Patient-Reported Outcomes: The Time Is Ripe for Integration Into Patient Care and Clinical Research. Am Soc Clin Oncol Educ Book. 2016;35:e89–96. https://doi.org/10.14694/EDBK_158749.
18. Food and Drug Administration. (2018) Framework for FDA's Real-World Evidence Program. Retrieved from https://www.fda.gov/media/120060/download.
19. Gliklich RE, Dreyer NA, Leavy MB, editors. Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Apr. 20, Rare Disease Registries. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK208609/.
20. Calvert MJ, O'Conner DJ, and Basch EM. 2019. Comment. Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes. Nature Reviews: Drug Discovery.

Springer

Advocacy and Patient Involvement in Clinical Trials

Ellen Sigal, Mark Stewart, and Diana Merino

Contents

Abstract

Patient engagement in research and clinical trials has evolved over time. Patients are no longer simply passive research subjects but are increasingly being integrated into research teams and protocol review teams to help design, implement, and disseminate clinical trial findings. While potential barriers exist for meaningful patient engagement, mechanisms and methods to effectively engage patients and advocacy groups are evolving, and resources and best practices are continually being developed to assist researchers and patients. Additionally, legislation and regulatory guidance are being instituted to promote patient engagement and ensure it is a routine process for clinical trial development. Developing patient-centered clinical trial designs has led to development of innovative clinical trial infrastructures and statistical methods. Patient advocates and organizations are also increasingly developing their own data sources and

E. Sigal (✉) · M. Stewart · D. Merino
Friends of Cancer Research, Washington, DC, USA
e-mail: esigal@focr.org; mstewart@focr.org; dmerino@focr.org

clinical trials, which represent unique opportunities for researchers to partner with patient groups to rapidly advance drug development.

Introduction

The role of patients and advocates in clinical research and their involvement in the regulation and oversight of clinical trials have substantially grown over time. In just a few decades, patients have gone from being considered passive human subjects whose clinical measures would contribute to answering research questions to active participants and engaged stakeholders. This growing movement toward a more patient-centered approach aims to provide the best healthcare for each patient, which takes into consideration the patient's own goals, values, and preferences (Manganiello and Anderson 2011). This movement is rooted in early advocacy efforts led by the HIV/AIDS community dating back to 1988 and resulted in fundamental changes to the medical research paradigm.

The path from initial development of a new drug to entry of the new therapy into the patient community relies on clinical trials, which represent the final step in evaluating the safety and efficacy of new therapeutic approaches. Along this developmental path, patients can provide critical input from collecting natural history information; involvement in endpoint selection; protocol design; consent and eligibility; clinical trial recruitment and retention strategies; design of post-market safety studies; and dissemination of trial findings (Fig. 1).

A detailed analysis of several clinical trials indicates that 48% of all sites in a given trial fail to meet their enrollment targets and more than 11% never enroll a single patient (Kaitin 2013). It is estimated that less than 5% of adult cancer patients enroll in a clinical trial despite many indicating their desire to participate in clinical
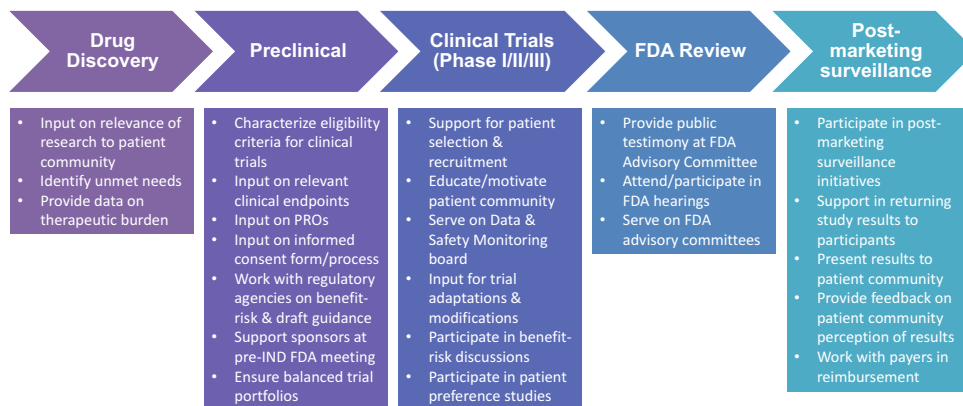
Fig. 1 Opportunities for patient involvement in the drug development process

trials (Comis et al. 2003; Unger et al. 2016). Thus, significant barriers such as clinical trial access, demographic and socioeconomic challenges, inappropriate or excessive procedures, broad exclusion criteria, lack of patient-centric trial designs, and patient and physician attitudes remain that hinder trial participation. While not every barrier may be readily overcome, engaging patients early and often throughout the entire research and drug development process can help ensure appropriately designed trials that are viewed favorably by patients, answer questions important to the patient community, and ultimately encourage participation.

A growing body of evidence describing the benefits of patient involvement in research and clinical trials is slowly changing scientific, medical, and regulatory practices. In their systematic review, Domecq and colleagues found that patient engagement positively influenced research by increasing study enrollment rates and helping researchers in securing funding, designing study protocols, and choosing relevant outcomes (Domecq et al. 2014). Greater patient engagement in research and clinical trials would help drug developers sponsor trials that are more informed about the needs of the patients, which would translate to more feasible and streamlined trial design generating better outcomes (Hanley et al. 2001; Tinetti and Basch 2013). Increased engagement could also reduce patient accrual time due to improved enrollment, reduce patient attrition, and make findings more applicable and relevant to the target population (Bombak and Hanson 2017), which would significantly decrease trial costs. Implementation of mechanisms for patient engagement can vary.

Patient Engagement in Research and Drug Development

Acknowledging that patients are central to research and drug development, several national and international organizations have invested in clearly defining the role of patient involvement in research practices and the need for the development of innovative infrastructures that will help facilitate the incorporation of the patient voice in all stages of the research process, including design, execution, and translation of research (Domecq et al. 2014). The Patient-Centered Outcomes Research Institute (PCORI) was established in 2010 to improve the quality and relevance of evidence available to help stakeholders make better-informed health decisions and requires that all its funded research projects include patient input throughout the entire research study (www.pcori.org). Patient engagement has been defined by PCORI as "involvement of patients and other stakeholders throughout the planning, conduct, and dissemination of the proposed project" and is becoming institutionalized and incorporated into several funding schemes (PCORI 2018). Patient-driven research activities have ranged from pre-discovery funding for development and acquisition of animal models and cell lines all the way to post-market study design and value discussions.

The US Food and Drug Administration (FDA) recognizes that patients are experts on living with their conditions, and as such, their voice is uniquely positioned to inform stakeholders and provide the right therapeutic context for drug

development as well as perspective on the outcome measures that are most relevant to patients and evaluation by regulatory agencies (Anderson and McCleary 2016). Patients may voice their concern or support for the development of certain drugs and provide a firsthand perspective on the proper balance of risk to benefit for a particular disease or patient population. For instance, the patient voice was crucial when reintroducing Tysabri, a monoclonal antibody used to treat multiple sclerosis, which had been previously removed from the market following reports of lethal side effects. After the thorough review of safety information, the FDA convened an advisory committee where patients and caregivers were invited to testify. Weighing all evidence, including the advocates' testimonies, the FDA found enough support to remarket the drug under a special prescription program (Schwartz and Woloshin 2015). Additionally, the FDA has formalized several initiatives to encourage the inclusion of the patient voice in medical product development. Under the fifth authorization of the Prescription Drug User Fee Act (PDUFA V) signed into law in 2012, the FDA began the Patient-Focused Drug Development (PFDD) program with the intent to more systematically incorporate the patient perspective into drug development (FDA 2018). From 2012 to 2017, the FDA organized 24 disease-specific PFDD meetings that have helped capture patients' experiences, perspectives, and priorities and enabled the incorporation of this meaningful information into the drug development process and its evaluation. Duchenne muscular dystrophy advocacy organizations helped to exemplify how patient and advocates can successfully inform regulators, provide meaningful input into benefit and risk assessments, and identify treatment priorities. To build on this success and enable more patient advocacy organizations to shape and influence drug development, the twenty-first Century Cures Act and PDUFA VI have tasked FDA with developing additional guidance to describe approaches to gather patient experience data, quantifying benefit and risks, and using patient-reported outcomes in treatment development. Moreover, the newly formed FDA Oncology Center of Excellence (OCE) has made PFDD a priority and is exploring innovative regulatory strategies that incorporate patient input. Additionally, the National Cancer Institute (NCI) also encourages patient advocates to be involved in the clinical trial process. The SWOG Cancer Research Network, one of five NCI cooperative cancer research groups, has an advocate assigned to every research committee and who is involved in every stage of the process.

Primary Areas of Engagement

A systematic review that searched for reporting of patient engagement on controlled trials and nonrandomized comparative trials conducted from May 2011 to June 2016 reviewed 2777 citations, of which only 23 clinical trials (17 randomized controlled trials and 6 nonrandomized comparative studies) reported patient engagement practices (Fergusson et al. 2018). The methods of engagement most commonly reported involved the development of the research question, selection of outcome, dissemination and implementation of results, and other activities, such as the

refinement of the study intervention and protocol review (Fergusson et al. 2018). Thus, there is evidence showing that researchers have engaged patients, especially in trials that reported following the community-based participatory research (CBPR) methods as part of the study design; however, there is still more work needed to get patients meaningfully involved in clinical research. Innovative methodologies, such as CBPR, which aim to have more meaningful relationships with the target population and more effective dissemination and implementation of results are key in improving patient involvement in research (Chhatre et al. 2018).

Another systematic review assessed patient engagement in research including randomized control trials, qualitative studies, single cohort studies, cross-sectional studies, case reports, and systematic reviews (Domecq et al. 2014). This study found that engagement was feasible and most commonly done in the beginning of the research process (agenda setting and protocol development) and less commonly during the execution and translation of research. The study also found no comparative effectiveness research on patient engagement methods. The authors concluded that the lack of this evidence is what may have led to inconsistent and vague reporting of patient engagement research, preventing the incorporation of effective reporting methods.

Using the 2014 Health Information National Trends Survey, one study investigated three aspects of patient engagement: interest, awareness, and participation as research partners in the medical research process to identify different levels of engagement and barriers that prevent engagement (Hearld et al. 2017). The study consisted of a cross-sectional analysis that suggested modest levels of interest in engaging in the research process among respondents. The study also found low levels of awareness of ways in which patients could become involved in research and very low levels of actual participation. Several factors, such as patient health status, attitudes about their health and healthcare, and sociodemographic characteristics, were also examined to provide insights into the types of patients most likely to be engaged in the research process. The study suggested that higher socioeconomic status and positive patient attitudes were associated with increased interest in becoming involved in research but there was no association between respondents with different demographic, socioeconomic, and environmental characteristics to actual participation. The authors concluded that raising awareness of engagement opportunities would improve people's interest in being engaged in research. Moreover, they suggested further research to identify why patients who may be aware of research opportunities are still reluctant to become active participants of the research process.

Challenges Associated with Incorporating Patients into Research and Drug Development

Attitudes toward a more patient-centered or patient-focused approach to care and research are continuing to shift, in part, because of the increasing awareness that active patient participation in research can lead to improvements in the credibility

of the study findings and their direct applicability to patients. In addition to the benefits observed for study sponsors and participants, greater patient involvement is also driven by a compelling ethical rationale that lies behind the participation of patients in the democratization of the research process (Domecq et al. 2014). Data shows a compelling relationship between the incidence of clinical trial enrollment and improvement in cancer population survival, and a recent survey indicates the value patient engagement can have on improving patient retention and accelerating trial accrual (Smith et al. 2015; Unger et al. 2016). However, several challenges and concerns remain about the way patient engagement is being conducted (Bombak and Hanson 2017).

Barriers to Patient Engagement

The most commonly described patient engagement barriers were related to logistics and a concern of tokenistic engagement (Domecq et al. 2014). Tokenism refers to involving patients superficially. This can often occur when a small number of participants, who may be involved in the research process minimally, are considered to represent a far larger and diverse patient group. This insincere act of patient inclusion hinders patients from seeking greater involvement in the research process, and it lessens the credibility of the patient voice. Indeed, various research studies have identified that people frequently find that participating in clinical trials is meaningless or disempowering (Mullins et al. 2014), yet people often want to be informed, empowered, and engaged in their medical management (Davis et al. 2005). Some programs may require patients to undergo intense forms of training and involve abundant time, interest, and potentially resources (Bombak and Hanson 2017). These requirements may create preference for observable or quantitative skills over instinct and intuition and may bias the perspectives shared as part of the study. The lack of incentives or payment for a patient's time may also be a barrier for some patients to become engaged in research. Moreover, various erroneous perceptions have been identified as barriers for engagement. Some studies have identified the detrimental perception that patients will not be objective in their decisions and will become a hurdle in the design and development process or that patients and advocates are naïve about the research process and funding problems (Hanley et al. 2001; Bombak and Hanson 2017). These barriers should be assessed in more detail, and greater efforts should be placed on overcoming any perceived drawback that would prevent patients from engaging and getting involved in scientific research.

Historically, few mechanisms existed for systematic engagement of patients in the drug development continuum, and in the very seldom cases in which structures for patient participation exist, they may be disorganized or confusing (Hohman et al. 2015). Efforts to overcome these should be undertaken, and learning modules and information are available to provide best practices. In recognition of these potential barriers, many patient advocacy organizations have research training programs designed specifically for patients to help inform and prepare them to support research

studies. They can also provide mechanisms to connect patients with opportunities to participate on advisory boards and research teams to support the development of clinical trials. Most notably, the National Breast Cancer Coalition developed Project LEAD Institute, which provides a series of courses that establish a foundation of scientific knowledge to empower patients to participate actively and collaborate with physicians, industry, and regulatory agencies. In addition, Fight Colorectal Cancer has a Research Advocacy Training and Support (RATS) program, and Susan G. Komen and the American Association for Cancer Research also have programs to train advocates to support research studies. The Clinical Trials Transformation Initiative (CTTI), a public-private partnership, helps develop and drive adoption of practices within physician and patient communities to support patient engagement that will increase the quality and efficiency of clinical trials.

The inclusion of patients as reviewers and on research teams has led to more appropriately designed trials and the development of innovative clinical trial designs and statistical methods. Additionally, studies have demonstrated that patient involvement in the design and development of clinical trials is necessary to improve the efficiency and relevance of drug development and evaluation.

The Contribution of Patient Advocacy to Research and Drug Development

The incorporation of the patient voice has directly impacted the way trials are designed and conducted (Mullins et al. 2014). The way in which clinical trials are designed can transform the evidence generation process to be more patient centered, providing people with an incentive to participate or continue participating in clinical trials. Providing better information to participants and incorporating alternative trial designs will minimize concerns that clinical trials aren't patient centered and will dispel any doubts or concerns that prevent patients from becoming meaningful participants in the planning and design of clinical trials. Addressing the concerns and desires of patients has led to innovative strategies and designs to make trials more patient centric.

Trial Designs and Endpoint Selection

Many new therapies in oncology are molecularly targeted against specific oncogenic driver mutations that may be present in only a fraction of the patient population. Although the advent of targeted therapies holds great promise for patients, it also means that many patients may need to be screened before enough patients harboring the necessary mutation are found. Additionally, patients may not have the mutation of interest and will potentially have to seek out a variety of trials before finding a match. Master protocols are one mechanism to assist with the development and investigation of targeted therapies (Woodcock and LaVange 2017). Perhaps one of the greatest efficiencies of the collaborative clinical trial system is its increased

benefit to patients seeking access to genomic screening technologies and experimental therapies. Rather than being forced to undergo multiple screening attempts and to move from trial to trial before ever being matched with a trial and treatment arm, patients who are screened for inclusion in a master protocol study need only be tested once to have a high likelihood of eventually participating in the study. The variety of patient subgroups that are evaluated over the course of a master protocol, as well as the use of non-match substudies, greatly increases patients' chances of receiving a study treatment. Moreover, patients who participate in master protocols are given access to a broad-based screening technology such as next-generation sequencing (NGS), which efficiently screens patients for a multitude of genomic markers and matches them to treatment arms based upon this information. Some select master protocols include the BATTLE program, LUNG-MAP for patients with lung cancer, and NCI-MATCH for patients with solid tumors, lymphomas, and myeloma.

Other patient-centric trial designs include pragmatic trials, adaptive trials, and trials that incorporate Bayesian statistics and allow patient crossover to the experimental treatment (Mullins et al. 2014). Pragmatic clinical trials can produce results that more accurately reflect the outcomes a typical person could expect to experience. Adaptive clinical trial designs allow for modifications to occur partway through the study based on information collected through the trial's progress. The incorporation of Bayesian statistics allows trialists to use prior information learned during the course of the trial and is often employed within adaptive trials. The subsequent Bayesian statistical analysis would describe the probability of a treatment's effect. While these trials provide many advantages for patients, they do have limitations. They can create logistical complications attributable to data management and study design as well as pose risks in the interpretability of the trial results. Trials that allow patients to crossover to the treatment arm, if shown to be superior to the control arm, can attenuate the treatment effect size. Additionally, the specific therapy under study may dictate which trial design is most optimal, particularly if interim results are unattainable to inform an adaptive methodology. The needs of patients and the need to generate solid evidence of efficacy will always need to be balanced.

It is important to engage patients early to understand the endpoints that matter most to them in all settings and stages of a disease. Mortality, for example, is an important outcome measure but is often not the only important outcome to patients. Especially in circumstances when chances of survival can be relatively low, other outcomes such as unnecessary diagnostic procedures or progression-free survival (PFS) are also important to patients. Clinical trials, therefore, must be designed with the patient's needs and preferences in mind within a given disease context. While certain endpoints may be more meaningful to researchers, these endpoints may ultimately not be meaningful to the patient group affected by the clinical trial. With the exception of validated surrogate endpoints, a primary endpoint should generally be a measure of something that is important to the patient (Vroom 2012). These endpoints should measure not only how a patient survives but also how a patient feels and functions.

The ascertainment of certain meaningful clinical endpoints, however, may be burdensome and time-consuming for researchers, hindering potentially lifesaving access for patients to the innovation under investigation. Recognizing this problem, Friends of Cancer Research and the Brookings Institute convened a panel of experts at a 2011 conference to discuss potential methods for streamlining the FDA approval process for drugs that show large treatment effects early in development while still ensuring drug safety and efficacy. The discussion at this conference informed the creation of the "Advancing Breakthrough Therapies for Patients Act" which established the FDA's Breakthrough Therapy Designation (BTD). This designation defines a breakthrough therapy as a drug intended to treat a serious or life-threatening disease or condition and for which preliminary evidence indicates that the drug may demonstrate substantial improvement over existing therapies (FDA Fact Sheet: Breakthrough Therapies). Once BTD is requested by the drug sponsor, the FDA and sponsor work together to determine the most efficient path forward, and if the designation is granted, the FDA will work closely with the sponsor to help expedite the development and review of the drug. Because innovative designation and approval pathways such as BTD take into consideration novel approval endpoints for clinical trials demonstrating higher rates of benefit in carefully selected patients, it is especially critical that patients are involved in identifying and defining the endpoints most important to them.

Given the broad benefits associated with patient involvement in scientific research and clinical trials, it is crucial to focus on greater dissemination and awareness. Strategies for the uptake and implementation of mechanisms for patient involvement should involve patients and patient advocates, health professionals, and drug developers. The creation of more educational resources to support researchers and patients when coordinating the incorporation of the patient voice in clinical trials would also improve the uptake of these mechanisms.

Capturing and Measuring Patient Experience

The patient voice is more commonly being incorporated in regulatory decision-making and has enabled the creation of more modern regulatory pathways. A patient's and their caregiver's experience with the disease and treatment-related symptoms, which may alter their function and health-related quality of life, is important. Capturing this rich experience from both patients and their caregivers helps provide key outcome information to consider in the evaluation of new agents. A recent policy review article written by international regulatory professionals from the USA, Europe, and Canada highlights the need for capturing the patient experience from different sources and focuses on the use of rigorous PRO measures to facilitate the regulatory decision-making process (Kluetz et al. 2018). Among the many advantages that PRO measures provide, these data are critical for supporting the benefit-risk assessment of experimental agents and useful when incorporated into prescribing and product information as descriptive data to inform safety

and tolerability (Kim et al. 2018) or as a claim of treatment benefit. This information is particularly important for concerns with quality of life issues that patients and caregivers may have.

All international regulatory agencies acknowledge that robust and accurate data collected from the patient experience can be useful, as it complements existing measurements of safety and efficacy, but warn that poorly defined PRO methodology using heterogeneous analytical methods greatly hinders the incorporation of PRO data in regulatory decision-making (Kluetz et al. 2018; Kuehn 2018; Bottomley et al. 2018). It recommends that sustained international collaboration among regulatory agencies is required to improve patient experience collection and standardize the assessment, analysis, and interpretation of patient data from clinical trials.

The FDA has recognized that a central aspect of PFDD is the use of patient-reported outcomes (PROs) as a way to incorporate the patient voice in drug development and regulatory decisions. PROs are directly reported by the patient and provide a status of the patient's health, quality of life, or functional status (FDA-NIH Biomarker Working Group 2016). PRO measures can provide a better understanding of treatment outcomes and tolerability from a patient perspective and complement current measures of safety and efficacy (Kim et al. 2018). In 2009, the FDA released guidance for industry on the use of PROs in medical product development to support labeling claims and has worked with other advocacy organizations, such as the Critical Path Institute, and industry to form working groups that seek to engage patients and caregivers in the development of robust symptom-measuring tools, such as the PRO Consortium. Although challenges exist when seeking to collect patient and caregiver experience data, such as the need for more personalized and dynamic measuring tools that keep up with the diversity of novel drug classes with wide variety of toxicities, greater efforts to ensure consistency, reliability, and applicability of these data are warranted to support robust use in the drug development space.

Contributors to Data Generation

Patients and advocacy organizations are also actively establishing their own data sources to support clinical drug development and, in some instances, establishing their own clinical trials. These include patient registries, online data-sharing communities, wearable devices, and social media tools for capturing longitudinal data points. Organizations such as the Genetic Alliance, the National Organization for Rare Disorders, and Parent Project Muscular Dystrophy have launched registries to study the natural history of disease, burden of disease, expectations for treatment benefits, and perspectives on tolerable harms and risks. These tools can help inform academia and industry and incentive further study into a particular disease state. Through public-private partnerships, advocacy organizations are also initiating clinical trials within their patient communities. For example, the Leukemia and Lymphoma Society is leading the Beat AML Master Trial, which is a collaborative

trial to test targeted therapies in patients with acute myeloid leukemia (AML) (Helwick 2018). Principle investigators should look for opportunities to utilize and integrate these data collection efforts into their research questions and studies in order to develop innovative partnerships that improve research logistics, outreach and communication, funding, and the prioritization of clinical trials.

Future Areas of Innovation and the Evolving Clinical Trial Landscape

There has been great progress in the area of patient engagement in clinical trials and the advancements being made by patient advocacy groups, and additional areas of opportunity continue to be identified. The development of more refined frameworks, models, best practices, and guidelines will help ensure early investigators have foundational knowledge to meaningfully engage patients and advocacy organizations in their research questions and drug development programs. Biopharma is investing heavily to accelerate development timelines. TransCelerate BioPharma Inc., a nonprofit organization that creates collaborations across biopharmaceutical research and development community, has recently launched a new initiative around patient awareness and access (TransCelerate 2018). Toolkits are available to assist research teams in engaging patient advocacy organizations and participants to optimize clinical trial designs. Additionally, some healthcare systems are partnering with cognitive computing platforms to help physicians match, enroll, and support patients (Bakkar et al. 2018).

The incorporation of external data sources to streamline, augment, and support clinical trial development is growing rapidly, due in large part to the advent of technological solutions that include patient collaboration programs, crowdsourcing, and the collection of big data and analytics. The US FDA is currently developing guidance and a framework to describe how real-world evidence can support drug development and regulatory decision-making. These external data sources represent an opportunity to augment clinical trial data and can potentially result in more streamlined drug development with fewer patients. These novel mechanisms of data collection, as well as their use and implementation, will continue to require the involvement of active advocates and consumers, who, through their experience, will contribute greatly to the oversight and eventual success of future clinical trials.

Cross-References

- ► Basket Protocols
- ► Bayesian Adaptive Designs
- ► Compassionate Use
- ► Creating the Initial Design of a Trial
- ► Crossover Trials
- ► Implementing the Trial Protocol

► Orphan Drugs and Rare Diseases
► Participant Recruitment, Screening, and Enrollment
► Patient Reported Outcomes
► Protocol: Development and amendments; SPIRIT
► Randomized Trials Using Claims or Electronic Health Data
► Trials in Minority Populations
► Trials in Rare Diseases
► Trials in Small Populations
► Trials in the Elderly
► Umbrella and Platform Trials
► Use of Real World Data in Trials

References

Anderson M, McCleary KK (2016) On the path to a science of patient input. Sci Transl Med 8:1–6. https://doi.org/10.1126/scitranslmed.aaf6730

Bakkar N, Kovalik T, Lorenzini I et al (2018) Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. Acta Neuropathol 135:227–247. https://doi.org/10.1007/s00401-017-1785-8

Bombak AE, Hanson HM (2017) A critical discussion of patient engagement in research. J Patient Cent Res Rev 4:39–41. https://doi.org/10.17294/2330-0698.1273

Bottomley A, Pe M, Sloan J et al (2018) Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. Clin Trials 15:624–630. https://doi.org/10.1177/1740774518795637

Chhatre S, Jefferson A, Cook R et al (2018) Patient-centered recruitment and retention for a randomized controlled study. Trials 19:205. https://doi.org/10.1186/s13063-018-2578-7

Comis RL, Miller JD, Aldigé CR et al (2003) Public attitudes toward participation in cancer clinical trials. J Clin Oncol 21:830–835. https://doi.org/10.1200/JCO.2003.02.105

Davis K, Schoenbaum SC, Audet AM (2005) A 2020 vision of patient-centered primary care. J Gen Intern Med 20:953–957. https://doi.org/10.1111/j.1525-1497.2005.0178.x

Domecq JP, Prutsky G, Elraiyah T et al (2014) Patient engagement in research: a systematic review. BMC Health Serv Res 14:1–9. https://doi.org/10.1016/j.transproceed.2016.08.016

FDA (2018) FDA voices: perspectives from FDA experts. https://www.fda.gov/newsevents/newsroom/fdavoices/default.htm. Accessed 12 Nov 2018

FDA Fact Sheet: Breakthrough Therapies. https://www.fda.gov/regulatoryinformation/lawsenforcedbyfda/significantamendmentstothefdcact/fdasia/ucm329491.htm. Accessed 12 Nov 2018

FDA-NIH Biomarker Working Group (2016) BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Food and Drug Administration (US), Silver Spring; Co-published by National Institutes of Health (US), Bethesda

Fergusson D, Monfaredi Z, Pussegoda K et al (2018) The prevalence of patient engagement in published trials: a systematic review. Res Involv Engagem 4:17. https://doi.org/10.1186/s40900-018-0099-x

Hanley B, Truesdale A, King A et al (2001) Involving consumers in designing, conducting, and interpreting randomised controlled trials: questionnaire survey. BMJ 322:519–523

Hearld KR, Hearld LR, Hall AG (2017) Engaging patients as partners in research: factors associated with awareness, interest, and engagement as research partners. SAGE Open Med 5:205031211668670. https://doi.org/10.1534/genetics.107.072090

Helwick C (2018) Beat AML trial seeking to change treatment paradigm. [Internet] The ASCO Post

Hohman R, Shea M, Kozak M et al (2015) Regulatory decision-making meets the real world. Sci Transl Med 7:313fs46. https://doi.org/10.1126/scitranslmed.aad5233

Kaitin K (2013) 89% of trials meet enrollment, but timelines slip, half of sites under-enroll. Tufts Cent Study Drug Dev Impact Rep 15:1–4

Kim J, Singh H, Ayalew K et al (2018) Use of pro measures to inform tolerability in oncology trials: implications for clinical review, IND safety reporting, and clinical site inspections. Clin Cancer Res 24:1780–1784. https://doi.org/10.1158/1078-0432.CCR-17-2555

Kluetz PG, O'Connor DJ, Soltys K (2018) Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol 19:e267–e274. https://doi.org/10.1016/S1470-2045(18)30097-4

Kuehn CM (2018) Patient experience data in US Food and Drug Administration (FDA) regulatory decision making: a policy process perspective. Ther Innov Regul Sci 52:661–668. https://doi.org/10.1177/2168479017753390

Manganiello M, Anderson M (2011) Back to basics: HIV/AIDS advocacy as a model for catalyzing change. AIDS 1–29. https://www.fastercures.org/assets/Uploads/PDF/Back2BasicsFinal.pdf

Mullins CD, Vandigo J, Zheng Z, Wicks P (2014) Patient-centeredness in the design of clinical trials. Value Health 17:471–475. https://doi.org/10.1016/j.jval.2014.02.012

PCORI (2018) The value of engagement. https://www.pcori.org/about-us/our-programs/engagement/value-engagement. Accessed 12 Nov 2018

Schwartz L, Woloshin S (2015) FDA and the media: lessons from Tysabri about communicating uncertainty. NAM Perspect 5. https://doi.org/10.31478/201509a

Smith SK, Selig W, Harker M et al (2015) Patient engagement practices in clinical research among patient groups, industry, and academia in the United States: a survey. PLoS One 10:e0140232

Tinetti ME, Basch E (2013) Patients' responsibility to participate in decision making and research. JAMA 309:2331–2332. https://doi.org/10.1001/jama.2013.5592

TransCelerate (2018) Patient experience. http://www.transceleratebiopharmainc.com/initiatives/patient-experience/. Accessed 12 Nov 2018

Unger JM, Cook E, Tai E, Bleyer A (2016) The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 35:185–198. https://doi.org/10.14694/EDBK_156686

Vroom E (2012) Is more involvement needed in the clinical trial design & endpoints? Orphanet J Rare Dis 7:A38. https://doi.org/10.1186/1750-1172-7-S2-A38

Woodcock J, LaVange LM (2017) Master protocols to study multiple therapies, multiple diseases, or both. N Engl J Med 377:62–70. https://doi.org/10.1056/NEJMra1510062

ORIGINAL RESEARCH

A National Assessment of Diagnostic Test Use for Patients with Advanced NSCLC and Factors Influencing Physician Decision-Making

Madison M. Wempe, BS; Mark D. Stewart, PhD; Daniel Glass, PhD; Laura Lasiter, PhD;
Diana Merino Vega, PhD; Nisha Ramamurthy; Jeff Allen, PhD; Ellen V. Sigal, PhD

**Stakeholder Perspective,
page 118**

Disclosures are at end of text
Supplemental material online

BACKGROUND: Diagnostic tests, including US Food and Drug Administration (FDA)-approved tests and laboratory-developed tests, are frequently used to guide care for patients with cancer, and, recently, have been the subject of several policy discussions and insurance coverage determinations. As the use of diagnostic testing has evolved, stakeholders have raised questions about the lack of standardized test performance metrics and the risk this poses to patients.

OBJECTIVES: To describe the use of diagnostic testing for patients with advanced non–small-cell lung cancer (NSCLC), to analyze the utilization of FDA-approved versus laboratory-developed diagnostic tests, and to evaluate the impact of existing regulatory and coverage frameworks on diagnostic test ordering and physician treatment decision-making for patients with advanced NSCLC.

METHODS: We conducted a 2-part study consisting of an online survey and patient chart review from March 1, 2019, to March 25, 2019, of physicians managing patients with advanced NSCLC. Respondents qualified for this study if they managed at least 5 patients with advanced NSCLC per month and had diagnosed at least 1 patient with advanced NSCLC in the 12 months before the survey. A total of 150 physicians completed the survey; before completing the survey, they were instructed to review between 4 and 8 charts of patients with stage IV NSCLC from their list of active patients.

RESULTS: A total of 150 practicing oncologists who manage patients with advanced NSCLC responded to the survey and reviewed a total of 815 patient charts. Of these 815 patients, 812 (99.6%) were tested for at least 1 biomarker, including 73% of patients who were tested for *EGFR*, 70% tested for *ALK*, 58% tested for *BRAF* V600E, and 38% of patients tested for *ROS1*, by FDA-approved diagnostic tests. In all, 185 (83%) patients who tested positive for *EGFR* and 60 (83%) patients who tested positive for *ALK* received an FDA-approved targeted therapy for their biomarker. A total of 98 (65%) physicians responded that the patient's insurance coverage factored into their decision to order diagnostic tests and 69 (45%) physicians responded that cost or the patient's insurance coverage could influence them not to prescribe an indicated targeted therapy.

CONCLUSION: The survey results indicate that diagnostic testing has become routine in the treatment of patients with advanced NSCLC, the use of FDA-approved diagnostic tests has increased, and insurance coverage and cost influence patient access to diagnostic testing as well as to targeted treatment options.

KEY WORDS: advanced NSCLC, diagnostic tests, FDA-approved tests, insurance coverage, laboratory-developed tests, lung cancer biomarkers, physician decision-making, targeted therapies

Ms Wempe is Research Associate, Dr Stewart is Vice President, Science Policy, Dr Lasiter is Director, Health Policy, Dr Vega is Director, Research Partnerships, Ms Ramamurthy was Summer Intern, Dr Allen is President and CEO, and Dr Sigal is Chairperson and Founder, Friends of Cancer Research, Washington, DC; Dr Glass is Managing Director, Global Market Research, Deerfield Management, New York, NY.

Molecular diagnostic tests are increasingly being used to guide care for patients with cancer, as the understanding of how different biomarkers contribute to tumor growth improves, technologies allowing for the simultaneous evaluation of multiple tumor alterations become available, and a growing number of transformative biomarker-targeted therapies are developed.[1]

These targeted therapies have proved beneficial to many patients with cancer, and in certain cancer indications, including *EGFR* mutation–positive advanced non–small-cell lung cancer (NSCLC), these therapies have led to notable improvements in patient outcomes, such as 5-year survival rates and progression-free survival.[2]

Recognizing the growing evidence that supports the use of diagnostic tests to identify appropriate therapies for patients with advanced NSCLC, leading organizations, such as the National Comprehensive Cancer Network (NCCN), the American Society of Clinical Oncology, the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology, have established clinical guidelines that universally recommend diagnostic testing for most patients with advanced or metastatic NSCLC.[3-5] These guidelines are intended for use by oncologists in the diagnosis and treatment decision-making processes.[3-5]

In addition, payers such as the Centers for Medicare & Medicaid Services (CMS), the largest payer in the US healthcare system, have recognized the importance of molecular diagnostic tests to their patient populations. In March 2018, CMS released a coverage determination in which US Food and Drug Administration (FDA)-approved or -cleared diagnostic tests that use next-generation sequencing technologies were deemed reasonable and necessary for patients with recurrent, relapsed, refractory, metastatic, or advanced stage III or IV cancer.[6] This decision ensured consistent coverage policy of FDA-approved and FDA-cleared diagnostic tests for Medicare patients at the national level. Recent data also suggest that these multigene panels may be more cost-effective than single-marker genetic tests.[7]

Diagnostic test results must be accurate, reliable, and clinically meaningful for patients to reap the benefits of precision medicine and biomarker-targeted therapies. Currently, the performance metrics of diagnostics are evaluated under a bifurcated regulatory system. The FDA was given authority to regulate all in vitro diagnostics as medical devices under the 1976 Medical Device Amendments to the Federal Food, Drug, and Cosmetic Act.[8] However, the agency has generally exercised discretion and has not enforced the device provisions of this legislation consistently on the subset of in vitro diagnostics (referred to as laboratory-developed tests); this is because, historically, laboratory-developed tests used relatively simple technology and have had limited availability.[8,9]

Instead, laboratory-developed tests have been regulated under the Clinical Laboratory Improvement Amendments (CLIA) program, which is overseen by CMS.[10] In contrast to rigorous FDA review, which requires demonstrations of analytical and clinical validity, the CLIA certification primarily assesses a laboratory's ability to properly conduct tests through protocol adherence and personnel qualification, but typically not the performance metrics of the test itself.[10]

Consequently, as the practice of medicine has evolved, and diagnostic tests have become increasingly complex and vital to clinical care, stakeholders such as patient advocacy organizations, congressional legislators, and the FDA have raised questions about the lack of standardized performance metrics for FDA-approved tests and for laboratory-developed tests, emphasizing the risk that this lack of standardization poses to patients.[11]

In fact, after identifying multiple problems with the performance of several high-risk laboratory-developed tests in 2010, the FDA announced its intent to reconsider its policy of enforcement discretion and took several subsequent actions. These actions included the development of draft guidance that outlined an approach to laboratory-developed test oversight in 2014, the publication of a report with 20 case studies that documented the public health impact of problematic laboratory-developed tests in 2015, and the release of a discussion paper on such tests in 2017.[9,11,12]

In December 2018, congressional leaders working

with the FDA and other stakeholders, including clinical laboratories, diagnostic test manufacturers, trade associations, and patient advocacy groups, released draft legislation aimed at establishing a uniform regulatory framework for all diagnostic tests under the authority of the FDA, with the intent of reducing the lack of standardization and inconsistent regulatory requirements for FDA-approved tests and laboratory-developed tests.[13]

This draft legislation was updated to incorporate extensive comments from the FDA and the broader community and was introduced to the US House of Representatives in early 2020, but its legislative path forward remains unclear.[14]

As part of our 2015 national survey of oncologists who managed patients with advanced NSCLC, we explored the use of FDA-approved diagnostic tests and laboratory-developed tests.[15] The findings from this original survey indicated that most patients with advanced NSCLC received testing for *EGFR* mutations and *ALK* rearrangements, with testing for *EGFR* more frequently performed in privately owned, academic, and community-based treatment settings using laboratory-developed tests, and testing for *ALK* evenly split between laboratory-developed tests and FDA-approved diagnostics across these settings. This utilization of different tests within the same treatment setting raised concerns that an unknown degree of variability could exist between tests with the same intended use.[15]

In the period since this original survey was conducted, tests have increased in complexity and have become more widely available, policy discussions have evolved, and major coverage determinations have been made; therefore, a new landscape analysis is warranted.

Our current study updates the original survey data[15] by investigating the use of FDA-approved tests and laboratory-developed tests based on a 2019 national survey of oncologists who managed patients with advanced NSCLC. In addition, this study includes survey questions regarding the factors that influence the diagnostic test ordering and treatment decision-making practices of physicians who manage patients with advanced NSCLC under current regulatory and coverage frameworks.

Methods

The study sample was based on a national panel of oncologists maintained by M3 Global Research. This panel includes more than 1000 physicians and is broadly representative of all oncologists in the United States across the demographic dimensions of region and years in medical practice. Study invitations were sent to all oncologists within the M3 panel. Respondents qualified for this study if they managed at least 5 patients with advanced NSCLC per month and had personally diagnosed at least 1 patient with advanced NSCLC in the 12-month period before the survey fielding (March 2018-February 2019).

The study consisted of 2 parts: a short survey and a patient chart review. A total of 150 physicians completed the study, translating to a response rate of 14%. The participants were offered an industry-standard honorarium as compensation for their time to complete the study. The study was administered online and was fielded from March 1, 2019, to March 25, 2019.

A data collection instrument was developed to capture deidentified information on patients with stage IV NSCLC in the United States for use in the chart review portion of the study. Based on a similar instrument fielded in 2015, we made updates to reflect new treatment and diagnostic test options that were available at the time of the study.

Before completing the survey, responding physicians were instructed to choose between 4 and 8 patients with stage IV NSCLC from their list of active patient charts. To facilitate the selection of random charts, oncologists were instructed to identify patients based on the assignment of random letters to correspond with the first letter of the patients' last names. The patient charts were required to have been active within the practice during the past 12 months.

Patient information, including age, weight, sex, ethnic origin, concomitant conditions, insurance type, smoking status, diagnosis year, genetic testing information, and all treatment lines, was recorded by physicians using the data collection instrument for each randomly selected patient chart. A total of 815 patient charts were included in the study, which reflected the practices of 150 responding oncologists who managed patients with advanced NSCLC.

Data Analysis

All survey data, including the identities of the responding physicians, were analyzed in aggregate and were completely anonymized. When the specific genetic test types were unknown to the responding physician, consent was obtained to contact the affiliated hospital pathology laboratory. Follow-up phone calls were made in these instances to determine which genetic testing platform or external testing services a specific hospital uses. The data were analyzed across all patients, as well as across the histologic subtypes of advanced NSCLC.

Key patient demographics, such as patient insurance type and practice setting, were also analyzed for patterns in genetic testing. Patients were included in the analysis of the use of FDA-approved and laboratory-developed tests if they were diagnosed with advanced NSCLC after the first FDA approval of a diagnostic test for a given NSCLC-related mutation, and if their test type could be

determined. Patients who were diagnosed with advanced NSCLC before the first FDA approval of a diagnostic test for a given NSCLC-related mutation and whose test type could not be determined were excluded from this analysis.

The analysis of insurance type was limited to patients with private insurance or with Medicare coverage. Patients with Medicaid, military insurance, self-insurance, no insurance, or unknown insurance types were excluded because of low sample sizes or because of the possibility of inherent confounding variables in these populations.

Statistical Analysis

All subgroup differences in proportions were tested using a chi-square analysis. Post-hoc pairwise comparisons used a Bonferroni correction to reduce the risk for type 1 errors. The pairwise comparisons were tested for significance at the .05 level. All statistical analysis was performed using IBM SPSS Statistics Version 20.0 (IBM; Armonk, NY).

Results

The study physician and patient characteristics are presented in **Table 1**. Our sample of physicians was broadly representative of oncology physicians across factors of interest such as years in practice and geographic region. In addition, the patient population included in our chart review was representative of the wider population of patients with advanced NSCLC across factors of interest such as age and histologic subtype.[16]

Biomarker testing rates among patients selected in the chart review portion of our study were examined. Of the 815 patients with advanced NSCLC in this study, 812 (99.6%) patients were tested for at least 1 mutation. A total of 669 (82%) patients were tested for *EGFR* mutations, 586 (72%) for *ALK* rearrangements, 298 (37%) for *BRAF* V600E mutations, and 380 (47%) for *ROS1* rearrangements. Other biomarkers were tested at lower frequencies. Of the 601 patients who were not classified as having squamous-cell carcinoma, 203 (34%) were tested for all 4 of these mutations.

The use of FDA-approved diagnostic tests for *EGFR* mutations, *ALK* rearrangements, *BRAF* V600E mutations, and *ROS1* rearrangements across treatment setting and insurance type is shown in **Table 2**. The differences in the use of FDA-approved diagnostic tests across treatment settings were significant (*P* <.05), indicating that the setting in which a patient receives treatment may influence whether he or she receives an FDA-approved test or a laboratory-developed test. Specifically, the difference in the use of FDA-approved tests for *EGFR* mutations between the academic (65%) and private (82%) practice settings was determined to be significant (*P* <.05).

Table 1	Characteristics of Physicians Who Completed the Survey and Patient Chart Review and Distribution of Patient Population Across Factors of Interest[a]		
Physician characteristics	**Responding physicians, N (%)**	**Patient characteristics**	**Patient charts, N (%)**
Sex		Sex	
Female	27 (18)	Female	320 (39)
Male	122 (81)	Male	495 (61)
Unknown	1 (1)	Age-group	
Years in practice		18-39 yrs	25 (3)
<10 yrs	34 (23)	40-64 yrs	294 (36)
10-19 yrs	43 (29)	≥65 yrs	496 (61)
20-29 yrs	23 (15)	Geographic region	
≥30 yrs	49 (33)	Midwest	190 (23)
Unknown	1 (1)	Northeast	184 (23)
Practice type		South	209 (26)
Academic center	57 (38)	West	167 (20)
Community-based center	29 (19)	Unknown	65 (8)
Private clinic	64 (43)	Cancer center setting	
Region		Academic center	341 (42)
East North Central	28 (19)	Community-based center	151 (18)
East South Central	5 (3)	Private clinic	323 (40)
Middle Atlantic	28 (19)	Insurance coverage	
Mountain	9 (6)	Medicaid	95 (12)
New England	5 (3)	Medicare	316 (39)
Pacific	21 (14)	Medicaid + Medicare	17 (2)
South Atlantic	25 (17)	Private	325 (40)
West North Central	7 (5)	Other	45 (5)
West South Central	11 (7)	Unknown	17 (2)
Unknown	11 (7)	Ethnicity	
Practice ownership		Caucasian	461 (56)
Physician owned	78 (52)	African American	181 (22)
Hospital owned	72 (48)	Asian	104 (13)
Other	0 (0)	Hispanic	63 (8)
Payer distribution (by patient's coverage)		Other	6 (1)
Majority Medicaid	13 (9)	Histologic subtypes	
Majority Medicare	62 (41)	Squamous-cell carcinoma	214 (26)
Majority private	47 (31)	Adenocarcinoma	536 (66)
Mix	28 (19)	Other type	65 (8)
		Smoking status	
		Current smoker	183 (22)
		Past smoker	405 (50)
		Passive smoker	48 (6)
		Never smoked	172 (21)
		Unknown	7 (1)

[a]Information on 815 patients provided by 150 responding physicians.

Table 2 Use of FDA-Approved versus Laboratory-Developed Diagnostic Tests Across Patient Treatment Settings and Insurance Types

Test used, by molecular abnormalities	Total sample, N (%)	Patient treatment setting			Patient insurance type				
		Academic center, N (%)	Community-based center, N (%)	Private clinic, N (%)	Medicare only, N (%)	Private, N (%)	Medicaid only, N (%)	Medicare + Medicaid, N (%)	Other, N (%)
EGFR									
FDA-approved	273 (73)	96 (65)	47 (69)	130 (82)	103 (75)	103 (73)	36 (72)	5 (56)	26 (72)
Laboratory-developed	101 (27)	51 (35)	21 (31)	29 (18)	35 (25)	38 (27)	14 (28)	4 (44)	10 (28)
ALK									
FDA-approved	219 (70)	78 (56)	40 (71)	101 (86)	85 (71)	85 (71)	28 (70)	4 (50)	17 (61)
Laboratory-developed	95 (30)	65 (44)	16 (29)	17 (14)	34 (29)	34 (29)	12 (30)	4 (50)	11 (39)
BRAF V600E									
FDA-approved	75 (58)	22 (44)	15 (48)	38 (79)	24 (57)	32 (59)	16 (73)	1 (33)	2 (25)
Laboratory-developed	54 (42)	28 (56)	16 (52)	10 (21)	18 (43)	22 (41)	6 (27)	2 (67)	6 (75)
ROS1									
FDA-approved	62 (38)	24 (41)	6 (18)	32 (44)	20 (35)	27 (42)	11 (42)	1 (20)	3 (30)
Laboratory-developed	101 (62)	34 (59)	27 (82)	40 (56)	37 (65)	38 (58)	15 (58)	4 (80)	7 (70)

NOTE: This table only includes information about patients for whom we were able to identify whether their diagnostic test was FDA-approved or laboratory-developed. Differences in the use of FDA-approved tests for *EGFR* mutations between the academic (65.3%) and private (81.8%) practice settings were significant (P <.05), based on chi-square tests. Differences in the use of FDA-approved tests for *ALK* rearrangements between the academic (55.7%) and private (85.6%) practice settings were significant (P <.05), based on chi-square tests. FDA indicates US Food and Drug Administration.

The differences in the use of FDA-approved tests for *BRAF* V600E mutations between the academic (44%) and private (79%) practice settings were also significant (P <.05). The differences in the use of FDA-approved tests for *ROS1* rearrangements between the community-based (18%) and private (44%) practice settings were significant as well (P <.05). No significant differences were observed between patients with private insurance and patients with Medicare coverage.

The receipt of an appropriate targeted therapy among patients who tested positive for an actionable biomarker was evaluated for *EGFR* mutations and *ALK* rearrangements across the treatment settings and insurance types. Therapies were deemed "appropriate" in this analysis if they were FDA-approved for advanced NSCLC, and if they targeted the biomarker(s) for which a patient tested positive.

Of the 669 patients who were tested for *EGFR* mutations, 223 (33%) tested positive for that mutation. In all, 185 (83%) of these patients received a targeted therapy indicated for patients with *EGFR*-positive advanced NSCLC. Similarly, of the 586 patients who were tested for *ALK* rearrangements, 72 (12%) tested positive for that mutation. A total of 60 (83%) of these patients received a targeted therapy indicated for patients with *ALK*-positive advanced NSCLC.

The differences in the use of an appropriate targeted therapy for patients testing positive for *EGFR* mutations between the academic (92%) and community-based (75%) practice settings were determined to be significant (P <.05). The difference in the use of an appropriate targeted therapy for patients testing positive for *EGFR* mutations between the academic (92%) and private (77%) practice settings was also significant (P <.05).

These significant differences indicate that a patient's treatment setting may influence whether he or she receives a targeted treatment. The use of an appropriate targeted therapy for patients testing positive for *ALK* rearrangement was significant overall (P <.05), but there were no significant differences in the pair-wise comparisons. No significant differences were observed between patients with private insurance and those with Medicare coverage.

A total of 66 (44%) physicians who responded to the survey indicated that the availability of targeted therapies influences their decision to order diagnostic tests for patients with advanced NSCLC. In all, 98 (65%) physicians indicated that the patient's type of insurance coverage factors into their decision to order diagnostic tests for patients with advanced NSCLC (**Figure Part A**).
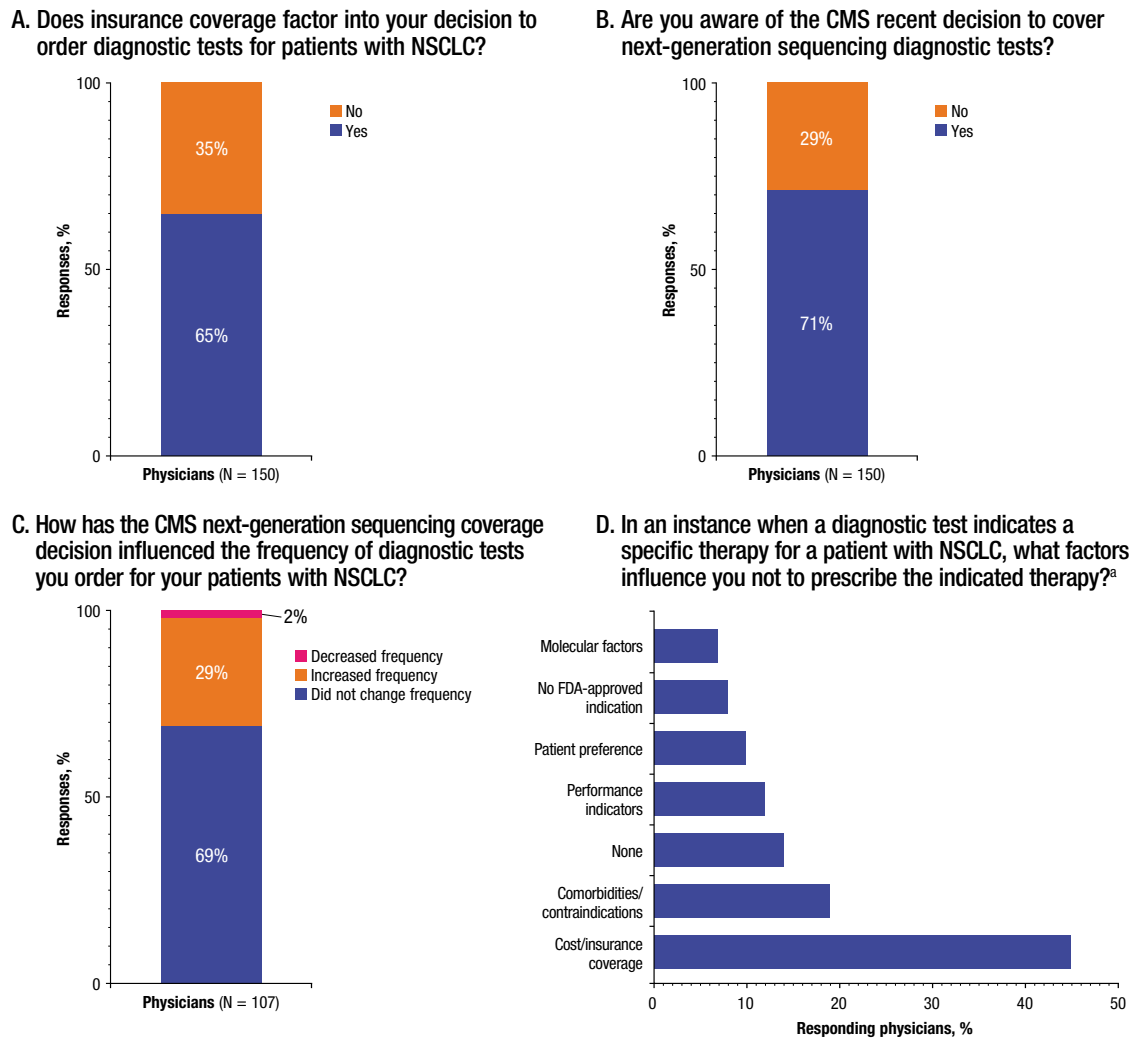
A total of 107 (71%) respondents indicated that they were aware of the CMS next-generation sequencing coverage determination (**Figure Part B**), and 74 (69%) of those 107 physicians indicated that this determination increased the frequency of diagnostic tests that they

| Figure | Factors Influencing Use of Diagnostic Tests and Prescription of Targeted Therapies: Selected Physician Questionnaire Responses |

A. Does insurance coverage factor into your decision to order diagnostic tests for patients with NSCLC?

■ No
■ Yes

35%
65%

Physicians (N = 150)

B. Are you aware of the CMS recent decision to cover next-generation sequencing diagnostic tests?

■ No
■ Yes

29%
71%

Physicians (N = 150)

C. How has the CMS next-generation sequencing coverage decision influenced the frequency of diagnostic tests you order for your patients with NSCLC?

■ Decreased frequency
■ Increased frequency
■ Did not change frequency

2%
29%
69%

Physicians (N = 107)

D. In an instance when a diagnostic test indicates a specific therapy for a patient with NSCLC, what factors influence you not to prescribe the indicated therapy?[a]

Molecular factors
No FDA-approved indication
Patient preference
Performance indicators
None
Comorbidities/contraindications
Cost/insurance coverage

Responding physicians, %

[a]This survey question was open-ended and responses were coded into the categories listed.
CMS indicates Centers for Medicare & Medicaid Services; FDA, US Food and Drug Administration; NSCLC, non–small-cell lung cancer.

order for patients with advanced NSCLC (**Figure Part C**). In all, 149 (99%) physicians indicated that they are confident that diagnostic tests yield high-quality data about their patients with advanced NSCLC.

A total of 69 physicians (45%) responded that in a case when a diagnostic test indicated a specific therapy for a patient with advanced NSCLC, cost issues or the patient's insurance coverage could influence them not to prescribe the indicated therapy (**Figure Part D**). The full physician questionnaire can be found in the **Appendix**

(available at **www.AHDBonline.com**). Selected questionnaire responses are shown in the Figure.

Discussion

Given the increasing importance of molecular diagnostic tests and targeted therapies in the diagnosis and treatment of patients with cancer, we examined the diagnostic test ordering and treatment decision-making practices of physicians who are managing patients with advanced NSCLC under the current federal regulatory

and healthcare coverage frameworks. Our analysis of 150 physician survey responses and 815 patient records produced 3 key findings.

First, our findings indicate that diagnostic testing has become routine in the diagnosis of and treatment decision-making for patients with advanced NSCLC. The survey respondents reported that almost all patients included in the chart review received at least 1 diagnostic test, which is consistent with previous findings in the literature that indicate high rates of diagnostic testing in patients with lung cancer.[17-20]

In line with evidence-based clinical practice guidelines issued by the NCCN,[5] we observed that most patients were tested for *EGFR* mutations and *ALK* rearrangements, and that the testing rates differed across the histologic subtypes. Although most patients were tested for these individual genetic markers, few patients received the NCCN-recommended comprehensive biomarker testing for their subtypes,[5] indicating that barriers to comprehensive biomarker testing may still exist.

In addition, our survey highlights that frameworks for the appropriate insurance coverage of diagnostic tests and targeted therapies are important to the access of care for patients with advanced NSCLC. The survey respondents indicated that insurance coverage influences their decision to order diagnostic tests for patients with advanced NSCLC. Furthermore, insurance coverage and cost were the most frequently reported factors that influenced a physician not to prescribe an indicated therapy.

Although we did not observe significant differences in the ordering of diagnostic tests and in the prescription of targeted therapies across the insurance types it is likely because, as a result of sample size concerns, our analyses were limited to patients with private insurance and Medicare, and the value of diagnostic tests and targeted therapies is widely acknowledged by these payers. Further examination of the impact of insurance type on a patient's access to diagnostic tests and targeted therapies is warranted, and if inequities in access to care are observed, intervention may be required at the societal and governmental levels.

Finally, we observed a shift toward the use of FDA-approved diagnostic tests since the publication of our 2015 survey,[15] but the magnitude of this shift differed across the treatment settings. In our 2015 survey, only 13% of patients tested for *EGFR* and 51% of patients tested for *ALK* received FDA-approved diagnostic tests,[15] compared with 73% and 70%, respectively, in our current survey (Table 2). This shift may indicate that physicians and, ultimately, pathologists value FDA regulation as well as newly approved technologies, such as next-generation sequencing panels, which were approved for use in this population after 2015.

The differences in the magnitude of this shift across the treatment settings may reflect, in part, that physicians in the academic setting preferentially use tests developed by their individual institutions. Our results also suggest that the overall increase in the use of FDA-approved tests may be explained by the increasing number of diagnostic assays approved by the FDA (eg, 2 tests approved in 2015 for *EGFR* vs 4 in 2019; 1 test approved in 2015 for *ALK* vs 3 in 2019) and the length of time since the initial FDA approval of a diagnostic test for a given mutation (eg, the first *EGFR* mutation test was approved in 2013; in 2015, that test had been on the market for 2 years vs 6 years in 2019).

Policy Implications and Recommendations

Based on the results of this 2-part study consisting of a survey and chart review, we recommend 2 policy changes to ensure patient access to high-quality, well-validated diagnostic tests and to their indicated biomarker-targeted therapies.

First, we recommend that as legislators and other stakeholders continue to work toward improving the current regulatory system for the benefit of patients, they prioritize the development of a predictable regulatory framework that fosters and encourages innovation while maintaining uniform oversight. Although we observed a shift toward the use of FDA-approved tests in our survey, a significant number of patients with advanced NSCLC still received molecular assessments that are subject to regulatory requirements different from those pursuing FDA premarket review.

However, this study did not seek to address the relative quality of laboratory-developed tests and FDA-approved diagnostic tests, and further research on potential variability in performance metrics and comparative outcomes is warranted. An improved framework should maintain the FDA's standards for analytical and clinical validity but not impose an excessive burden on stakeholders involved in diagnostic test innovation, such as academic laboratories that have voiced concerns that seeking FDA approval would be onerous and expensive.

Second, we recommend that as payers consider future coverage decisions, in the absence of a uniform regulatory framework, they develop a minimum set of performance characteristics necessary to support determinations of coverage for diagnostic tests. Because diagnostic test results are frequently used in treatment decision-making processes, it is critical that accurate results are produced to prevent patients from being exposed to nonefficacious treatments and the unnecessary toxicities that would result from inappropriate identification of candidates for treatment.

As our survey demonstrated, physician decision-making is affected by insurance coverage, and the latest shifts toward the use of FDA-approved diagnostic tests coin-

cided with the recent CMS coverage determination, which indicates that coverage frameworks may have the potential to alter physicians' prescribing patterns. Furthermore, the continued coverage of biomarker-targeted therapies and mechanisms for rapidly incorporating new diagnostic test and drug approvals into coverage frameworks are essential to preserve patient access.[21]

Limitations

This study has several limitations. First, this survey focused on oncologists, not pathologists. Although pathologists may have access to more diagnostic test–related information, oncologists were more appropriate for this study because our goals were to evaluate the use of diagnostic tests for patients with advanced NSCLC and to identify factors influencing physician decision-making under existing regulatory and coverage frameworks.

In addition, our study was not designed to address the comparative outcomes of patients who were tested with laboratory-developed tests versus FDA-approved tests.

Furthermore, as with most surveys, the potential for response bias and for nonresponding physicians bias exists.

The potential impact of confounders, such as a patient's inability to receive a diagnostic test because of insufficient tissue, is unknown.

Finally, a portion of the patient records (and associated pathology reports) did not include information on the type of test used to detect lung cancer mutations, even after follow-up phone calls, and had to be excluded from further analysis, including 295 (44%) patients who were tested for *EGFR* mutations, 272 (46%) patients who were tested for *ALK* rearrangements, 169 (57%) patients tested for *BRAF* V600E mutation, and 217 (57%) patients tested for *ROS1* rearrangements.

Conclusion

Molecular diagnostic tests and biomarker targeted therapies are routinely used in oncology care and will continue to drive the concept of precision medicine forward. Our study presents novel survey and chart review data that illustrate the routine use of diagnostic tests in the treatment of patients with advanced NSCLC and demonstrate an increase in the use of FDA-approved diagnostic tests. We also identify factors, such as insurance coverage and cost, that influence physicians' diagnosis and treatment decision-making processes. Together, these findings illustrate that optimized regulatory and coverage frameworks are critical to an oncology patient's access to care.

Author Disclosure Statement
Ms Wempe, Dr Stewart, Dr Glass, Dr Lasiter, Dr Vega, Dr Allen, Dr Sigal, and Ms Ramamurthy have no conflicts of interest to report.

References

1. Berberabe T. With biomarker expansion, combinations and personalized medicine continue to rise. *Target Ther Oncol.* July 5, 2019. www.targetedonc.com/publications/targeted-therapy-news/2019/July-2019/with-biomarker-expansion-combinations-and-personalized-medicine-continue-to-rise. Accessed September 6, 2019.
2. Lin JJ, Cardarella S, Lydon CA, et al. Five-year survival in *EGFR*-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. *J Thorac Oncol.* 2016;11:556-565.
3. Kalemkerian GP, Narula N, Kennedy EB, et al. Molecular testing guideline for the selection of patients with lung cancer for treatment with targeted tyrosine kinase inhibitors: American Society of Clinical Oncology endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology clinical practice guideline update. *J Clin Oncol.* 2018;36:911-919.
4. Lindeman NI, Cagle PT, Aisner DL, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. *Arch Pathol Lab Med.* 2018;142:321-346.
5. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Non-Small Cell Lung Cancer: NCCN Evidence Blocks. Version 7.2019. August 30, 2019. www.nccn.org/professionals/physician_gls/pdf/nscl_blocks.pdf. Accessed September 6, 2019.
6. Centers for Medicare & Medicaid Services. Decision memo for next generation sequencing (NGS) for Medicare beneficiaries with advanced cancer (CAG-00450N). March 16, 2018. www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=290. Accessed September 6, 2019.
7. Steuten L, Goulart B, Meropol NJ, et al. Cost effectiveness of multigene panel sequencing for patients with advanced non–small-cell lung cancer. *JCO Clin Cancer Inform.* 2019;3:1-10. doi.org/10.1200/CCI.19.00002.
8. Federal Food, Drug, and Cosmetic Act, 21 USC §321 (2013). www.govinfo.gov/content/pkg/USCODE-2013-title21/html/USCODE-2013-title21-chap9-subchapII-sec321.htm. Accessed September 6, 2019.
9. US Food and Drug Administration. The public health evidence for FDA oversight of laboratory developed tests: 20 case studies. November 16, 2015. http://wayback.archive-it.org/7993/20171115144712/https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM472777.pdf. Accessed September 6, 2019.
10. The Clinical Laboratory Improvement Amendments. *J Natl Cancer Inst.* 2004;96:501. doi.org/10.1093/jnci/96.7.501.
11. US Food and Drug Administration. Discussion paper on laboratory developed tests (LDTs). January 13, 2017. www.fda.gov/media/102367/download. Accessed September 6, 2019.
12. US Food and Drug Administration. Draft guidance for industry, Food and Drug Administration staff, and clinical laboratories: framework for regulatory oversight of laboratory developed tests (LDTs). October 3, 2014. www.fda.gov/media/89841/download. Accessed September 6, 2019.
13. Verifying Accurate Leading-edge IVCT Development Act of 2018. Preliminary discussion draft. December 6, 2018. https://degette.house.gov/sites/degette.house.gov/files/valid_act_discussion_draft_12.6.18.pdf. Accessed September 6, 2019.
14. Verifying Accurate Leading-edge IVCT Development (VALID) Act of 2020. H.R. 6102. 116th Congress. March 5, 2020. www.congress.gov/116/bills/hr6102/BILLS-116hr6102ih.pdf. Accessed May 21, 2020.
15. Audibert CM, Shea MB, Glass DJ, et al. Use of FDA-approved and laboratory-developed tests in advanced non–small cell lung cancer: results of a retrospective market analysis. *Personalized Med Onc.* 2016;5(7):278-284.
16. Stewart M, Norden AD, Dreyer N, et al. An exploratory analysis of real-world end points for assessing outcomes among immunotherapy-treated patients with advanced non–small-cell lung cancer. *JCO Clin Cancer Inform.* 2019;3:1-15. doi.org/10.1200/CCI.18.00155.
17. Mason C, Ellis PG, Lokay K, et al. Patterns of biomarker testing rates and appropriate use of targeted therapy in the first-line, metastatic non-small cell lung cancer treatment setting. *J Clin Pathw.* 2018;4:49-54.
18. Lim C, Tsao MS, Le LW, et al. Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. *Ann Oncol.* 2015;26:1415-1421.
19. Schink JC, Trosman JR, Weldon CB, et al. Biomarker testing for breast, lung, and gastroesophageal cancers at NCI designated cancer centers. *J Natl Cancer Inst.* 2014;106:dju256. doi.org/10.1093/jnci/dju256.
20. Lee DH, Tsao MS, Kambartel KO, et al. Molecular testing and treatment patterns for patients with advanced non small cell lung cancer: PIvOTAL observational study. *PLoS One.* 2018;13:e0202865. doi.org/10.1371/journal.pone.0202865.
21. Alexandria Real Estate Equities; Friends of Cancer Research. Blueprint for breakthrough forum: research and reimbursement in the age of precision medicine. September 2018. www.focr.org/sites/default/files/pdf/Friends Alexandria Blueprint White Paper_October.pdf. Accessed September 6, 2019.

Stakeholder Perspective next page

Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project

Diana M Merino [ORCID],[1] Lisa M McShane,[2] David Fabrizio,[3] Vincent Funari,[4] Shu-Jen Chen,[5] James R White,[6] Paul Wenz,[7] Jonathan Baden,[8] J Carl Barrett,[9] Ruchi Chaudhary,[10] Li Chen,[11] Wangjuh (Sting) Chen,[12] Jen-Hao Cheng,[5] Dinesh Cyanam,[10] Jennifer S Dickey,[13] Vikas Gupta,[14] Matthew Hellmann,[15] Elena Helman,[16] Yali Li,[3] Joerg Maas,[17] Arnaud Papin,[18] Rajesh Patidar,[11] Katie J Quinn,[16] Naiyer Rizvi,[19] Hongseok Tae,[12] Christine Ward,[8] Mingchao Xie,[20] Ahmet Zehir,[15] Chen Zhao,[7] Manfred Dietel,[17] Albrecht Stenzinger,[21] Mark Stewart,[1] Jeff Allen,[1] On behalf of the TMB Harmonization Consortium

ABSTRACT

Background Tumor mutational burden (TMB), defined as the number of somatic mutations per megabase of interrogated genomic sequence, demonstrates predictive biomarker potential for the identification of patients with cancer most likely to respond to immune checkpoint inhibitors. TMB is optimally calculated by whole exome sequencing (WES), but next-generation sequencing targeted panels provide TMB estimates in a time-effective and cost-effective manner. However, differences in panel size and gene coverage, in addition to the underlying bioinformatics pipelines, are known drivers of variability in TMB estimates across laboratories. By directly comparing panel-based TMB estimates from participating laboratories, this study aims to characterize the theoretical variability of panel-based TMB estimates, and provides guidelines on TMB reporting, analytic validation requirements and reference standard alignment in order to maintain consistency of TMB estimation across platforms.

Methods Eleven laboratories used WES data from The Cancer Genome Atlas Multi-Center Mutation calling in Multiple Cancers (MC3) samples and calculated TMB from the subset of the exome restricted to the genes covered by their targeted panel using their own bioinformatics pipeline (panel TMB). A reference TMB value was calculated from the entire exome using a uniform bioinformatics pipeline all members agreed on (WES TMB). Linear regression analyses were performed to investigate the relationship between WES and panel TMB for all 32 cancer types combined and separately. Variability in panel TMB values at various WES TMB values was also quantified using 95% prediction limits.

Results Study results demonstrated that variability within and between panel TMB values increases as the WES TMB values increase. For each panel, prediction limits based on linear regression analyses that modeled panel TMB as a function of WES TMB were calculated and found to approximately capture the intended 95% of observed panel TMB values. Certain cancer types, such as uterine, bladder and colon cancers exhibited greater variability in panel TMB values, compared with lung and head and neck cancers.

Conclusions Increasing uptake of TMB as a predictive biomarker in the clinic creates an urgent need to bring stakeholders together to agree on the harmonization of key aspects of panel-based TMB estimation, such as the standardization of TMB reporting, standardization of analytical validation studies and the alignment of panel-based TMB values with a reference standard. These harmonization efforts should improve consistency and reliability of panel TMB estimates and aid in clinical decision-making.

BACKGROUND

Immune checkpoint inhibitors (ICIs) have recently emerged as a pillar of cancer care, providing the potential for durable responses and improved survival for patients across multiple cancer types.[1–3] An intensive clinical development pipeline investigating ICIs is ongoing as a result. However, not all patients with cancer respond to ICIs, with modest

response rates for several approved indications (approximately 20% or less in lung cancer, bladder cancer and cancers of the head and neck, among others) and high treatment costs. There is a crucial interest in the development of biomarker assays to predict which patients are most likely to respond and benefit from ICIs, and to improve clinical decision-making and disease management.[4 5]

Expression of the programmed cell death ligand protein-1 (PD-L1) by immunohistochemistry (IHC) has been studied extensively as a biomarker of response to anti-PD-L1 and programmed cell death protein 1 (PD-1) therapy. Several assays have been developed to quantify tumor PD-L1 immuno-positivity; however, quantitation is imperfect, and lack of standardization across platforms and scoring systems precludes assay interchangeability.[6] Tumor mutational burden (TMB), which measures the number of somatic mutations per megabase (Mb) of the interrogated genomic sequence of a tumor, has been most recently identified as a biomarker of response to ICIs in several cancer types. High TMB is associated with improved outcomes in patients with melanoma treated with cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) blockade[7–9] and PD-1/PD-L1 blockade across several cancer types, including melanoma,[10 11] non-small-cell lung carcinoma,[12–15] bladder cancer,[16] microsatellite instability cancers[3 17] and pan-tumor cohorts.[18–20] High TMB has also been associated with improved outcomes in patients treated with a combination of PD-1/PD-L1 and CTLA-4 inhibitors.[21–24]

Initial assessments of TMB involved whole exome sequencing (WES) of matched tumor tissue and normal specimens using next-generation sequencing (NGS).[3 8–10] However, WES is not currently routine in clinical practice due to substantial cost and turnaround time, which has led assay manufacturers and commercial and academic labs to develop targeted NGS panels. These targeted panels, which cover several hundred genes, are already routinely used in clinical practice, and are currently being adapted to estimate TMB. TMB estimated from targeted NGS panels has generally correlated well with TMB determined by WES, however the reliability of this technology is still being assessed.[13–16 20 22 25–30]

There are several targeted NGS panels at different stages of development that estimate TMB. To date, the Foundation Medicine FoundationOne CDx test[31] is currently the only Food and Drug Administration (FDA)-approved panel, which includes TMB as part of its tumor profiling claim, while the Memorial Sloan Kettering Cancer Center MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets)[32] has received FDA authorization. Additionally, there are many more commercial and laboratory-developed test panels currently under development. Each panel has unique features integrated into their design that may impact TMB estimation. For example, each panel may include different numbers and types of genes, use different sequencing platforms, have different methods of filtering germline mutations, incorporate different mutation types in the quantification of TMB and use proprietary bioinformatics protocols to calculate TMB.[33 34] Thus, TMB estimates will vary according to the targeted panel used.[35] This is a crucial time to understand the differences in TMB estimation across panels, standardize the way TMB is reported, begin to harmonize methods for TMB quantification and identify optimal approaches to promote TMB alignment across different targeted NGS panels.

Friends of Cancer Research (*Friends*) convened a consortium of key stakeholders, including diagnostic manufacturers, academics, pharmaceutical companies, the National Cancer Institute and the FDA, to recommend best practices and approaches for TMB measurement, validation, alignment and reporting well ahead of the adoption of this powerful biomarker in clinical decision-making. Leveraging the expertise and insights of this comprehensive group of stakeholders, the *Friends* TMB harmonization project seeks to establish a uniform approach to measure and report TMB across different sequencing panels by harmonizing the definition of TMB, proposing best practices for analytic validation studies and ensuring consistency of TMB calculation through alignment with a universal reference standard. The project consists of a stepwise approach broken down into three phases: phase I, reported here, comprises the in silico analysis, which by using publicly available data from The Cancer Genome Atlas (TCGA) representing 32 cancer types, aims to identify the theoretical variability of panel-derived TMB estimates (panel TMB) relative to a common, standardized WES-derived TMB (WES TMB) across various panels. Building on the results of the in silico analysis, phase II will analyze human tumor clinical sample material to objectively measure variation across panels using patient formalin-fixed paraffin-embedded (FFPE) tissue samples. This empirical analysis will also compare panel TMB results to an agreed on universal reference standard, consisting of a collection of human tumor-derived reference cell lines that span a clinically meaningful TMB dynamic range. FFPE tissue samples will also be used to validate the use of the cell line standard. Finally, phase III will involve a clinical study that seeks to retrospectively analyze samples from patients treated with ICIs to evaluate optimal cut-off values that will help guide the clinical application of TMB (see online supplementary figure 1).

The need for harmonization of TMB is a global effort, which is portrayed by the representation of national and international diagnostic companies in the consortium. Moreover, in seeking to complement the consortium's work, the *Friends* TMB harmonization project has partnered with the technical comparability study conducted by Quality in Pathology in Germany,[36] leading to the identification of common and panel-specific factors that influence TMB estimation and the development of global recommendations, which have been published previously.[33]

Due to the large scale and collaborative nature of this effort, study results will greatly contribute to understanding and refining how to best quantify and interpret TMB as a biomarker, help establish standards that will facilitate harmonization across different testing platforms and inform future harmonization efforts that seek to ensure consistency across diagnostic platforms.

METHODS
In silico dataset
Mutation calls generated using Multi-Center Mutation calling in Multiple Cancers (MC3) WES data from TCGA project were used for this analysis.[37] Variants that overlapped with the CCDS, using bedtools (-wa option)[38] were extracted from the publicly available mc3.v0.2.8.PUBLIC.maf file (https://gdc.cancer.gov/about-data/publications/pancanatlas). Finally, the data were filtered for any overlap or redundancy using the 'merge' function. The consortium created a final bed file that covered 32.102 Mb of the genome after intersecting the data found in the MAF files and filtering for any overlap or redundancy (see online supplementary methods). The final bed file size was used as the denominator for calculating WES TMB in this study. Three different consortium laboratories independently calculated WES TMB using the same dataset and analytical methodology with 100% concordance.

Ten thousand two hundred ninety-five tumor samples with matched normal initially composed part of the cohort. Only samples with at least one variant which PASSED variant review filter were used (see online supplementary methods for variant quality filters). Low quality samples based on variant filters and those with low purity were also removed from further analysis. The remaining cases (n=8291) were randomly assigned to training (n=4157) and validation (n=4134) datasets with similar median candidate mutations and cancer types (online supplementary figure 2). Participants, though not required, could use the 'training' set for their own algorithm or parameter testing. However, all analyses described herein were conducted using the validation dataset.

The evaluations reported in the present study are those comparing panel TMB to WES TMB on the validation set, with no adjustments made to the panel TMB algorithms once the validation set analyses began. All analyses focused on tumors for which WES TMB was ≤40 because >98% of the TCGA dataset tumors investigated had TMB ≤40 in the TCGA dataset and all members of the consortium agreed that this range would have the greatest relevance for clinical decision-making. Of the 4134 tumors initially represented in the validation set, 4065 remained after excluding those with WES TMB >40. All results were blinded to the entire consortium, with the exception of the project statistician and data manager (LMMS and DMM) who were regarded as neutral parties not affiliated with any of the participating laboratories.

Statistical analysis
Statistical analyses interrogated the relationship between WES TMB and panel TMB values. The first analysis focused on the combined data from all 32 tumor types. Spearman's R correlation values were calculated, and scatterplots and difference plots were created to assess linearity of the relationship between panel TMB and WES TMB and to evaluate whether variance of panel TMB was constant across the range of WES TMB values.

Next, the 32 tumors were divided into three strata according to the number of samples within each tumor type that had TMB values spanning the range 0–40 mut/Mb (see online supplementary methods and figure 3). Stratum 1 contained eight tumor types (see online supplementary table 1A—stratum 1) displaying a good distribution of TMB values spanning the range of interest (0–40 mut/Mb). Seventy-seven per cent of samples (1257/1627) had TMB ≤10 mut/Mb, 19% (306/1627) had TMB 10–40 mut/Mb and 4% (64/1627) had TMB ≥40 mut/Mb and were thus eliminated from further analyses. Stratum 2 was represented by 11 tumor types (see online supplementary table 1B—stratum 2) whose samples had generally low TMB values (≤10 mut/Mb, 98%, 1723/1754), and only 1.5% (26/1754) of samples had TMB 10–40 mut/Mb. Only five samples (0.29%) had TMB ≥40 mut/Mb and were thus eliminated from further analyses. Stratum 3 was represented by 13 tumor types (see online supplementary table 1C—stratum 3) whose samples had very low TMB values (≤5 mut/Mb, 99.5%, 749/753) and only 4 samples (0.5%) had samples with TMB between 5 and 10 mut/Mb. Regression modeling using weighted least squares was implemented to account for the heteroscedasticity in errors, referring to the variability in panel TMB values about the fitted regression line, which was observed to increase with the mean and with WES TMB. This modeling was conducted for all strata, although we focused on stratum 1 considering strata 2 and 3 provided less stable and unreliable estimates due to the large number of samples that concentrated in the lower end of the TMB range.

For each regression, the mean panel TMB was modeled as a simple linear function of the WES TMB, and five different models for the error variance were considered (see online supplementary methods). Restricted maximum likelihood analysis using the gls function available in the R package nlme was performed to estimate the model parameters and select a best fitting variance structure based on minimum Akaike information and Bayesian information criteria.

Whole exome analysis
The whole exome analysis of the TCGA MC3 validation dataset used an agreed on methodology to calculate the WES TMB values, termed the Uniform TMB Calculation Method (see online supplementary table 2). The goal of phase I of this harmonization study is to assess the theoretical variability across panels. Given that the participating panels were at different stages of development and

COMPLEX BIOMARKERS: INFORMING STANDARDIZATION AND HARMONIZATION OF DIAGNOSTIC TESTING

Table 1 Description of the 11 participating diagnostic NGS panels

Laboratory	Panel name	# genes	Total region covered (Mb)	TMB region covered* (Mb)	Type of exonic mutations included in TMB estimation	Published performance characteristics (ref.)
ACT Genomics	ACTOnco+	440	1.8	1.12	Non-synonymous†, synonymous	NA
AstraZeneca	AZ600	607	1.72	1.72	Non-synonymous, synonymous	NA
Caris	SureSelect XT	592	1.60	1.40	Non-synonymous	Vanderwalde et al[40]
Foundation Medicine	FoundationOne CDx‡	324	2.20	0.80	Non-synonymous, synonymous	Frampton et al[41] Chalmers et al[25] Fabrizio et al[42] US FDA SSED[31]
Guardant Health	GuardantOMNI§	500	2.15	1.00	Non-synonymous, synonymous	Quinn et al[43]
Illumina	TSO500 (TruSight Oncology 500)	523	1.97	1.33	Non-synonymous, synonymous	NA
Memorial Sloan Kettering Cancer Center	MSK-IMPACT¶	468	1.53	1.14	Non-synonymous	Cheng et al,[44] Zehir et al,[30] US FDA[32]
NeoGenomics	NeoTYPE Discovery Profile for Solid Tumors	372	1.10	1.03	Non-synonymous, synonymous	NA
Personal Genome Diagnostics	PGDx elio tissue complete	507	2.20	1.33	Non-synonymous, synonymous	Wood et al[45]
QIAGEN	QIAseq TMB panel	486	1.33	1.33	Non-synonymous, synonymous	NA
Thermo Fisher Scientific	Oncomine Tumor Mutation Load Assay	409	1.70	1.20	Non-synonymous	Chaudhary et al[46] Endris et al[35]

*Coding region used to estimate TMB regardless of the size of the region assessed by the panel.
†Non-synonymous mutations include single nucleotide variants, splice-site variants and short insertions and deletions (indels).
‡FoundationOne CDx assay has been approved by the US FDA as an IVD.[31]
§GuardantOMNI is a plasma-based circulating tumor DNA assay.
¶MSK-IMPACT assay has been authorized by the US FDA[32]
NA, not available.

had different sensitivity levels, the consortium decided to use the Uniform TMB Calculation Method, which would enable the selection of high-quality variants that all laboratories were able to assess as part of their panels. The consortium created a custom bed file covering 32.102 Mb of the genome which was used to calculate the reference WES TMB values. The calculated WES TMB values comprised the reference dataset for this study. The uniform method for analysis of WES TMB included minimum thresholds for median target coverage (median 300X as this was identified as the point where sensitivity for the lower allele frequency variants drops drastically) (see online supplementary figure 4), variant allele frequency (≥0.05), read depth (≥25) and variant count (≥3), and synonymous variants were excluded.

Panel analysis
Each participating laboratory calculated TMB from the subset of the exome restricted to the genes covered by their targeted panel and using their own unique bioinformatics pipeline (panel TMB). If available, the laboratory's bioinformatics analysis has been reported in table 1.

The panel-derived TMB datasets were sent to a neutral third party (DMM) who assigned coded identifiers to the laboratories to mask which laboratory contributed each dataset. All subsequent data analyses were conducted by LMMS and DMM. Participating laboratories were not involved in the analyses and were not provided the key to the coded lab identifiers.

RESULTS
In silico assessment of theoretical TMB variation across panels
Eleven academic and commercial laboratories with targeted gene panels in different stages of development participated in this study (table 1). The size of the coding region used to estimate TMB from these gene panels ranged between 0.80 and 1.72 Mb. And the number of genes in each of the gene panels ranged between 324 and 607 genes. All participating laboratories included exonic somatic non-synonymous, frameshift and splice site variants and short indels when estimating TMB. Eight panels (8/11, 73%) also included synonymous variants in their
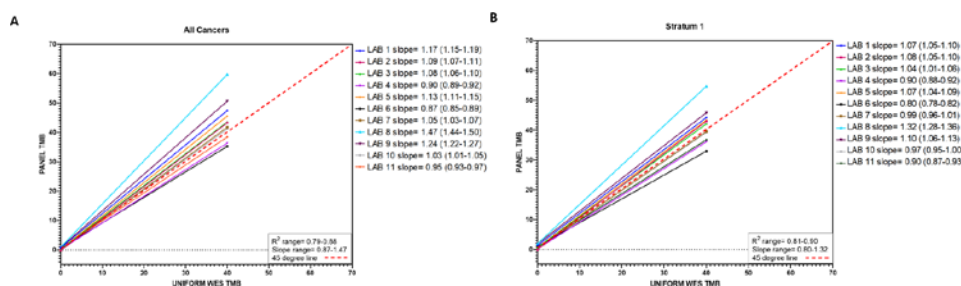
Figure 1 Estimated regression lines for panel tumor mutational burden (TMB) as a function of whole exome sequencing (WES) TMB for each of the 11 participating laboratories analyzing (A) all cancer types combined and (B) stratum 1 cancer types combined. Solid lines represent the fitted regression lines. Red dashed line represents 45° line.

estimation. Each laboratory used their own bioinformatics algorithms and workflows, which were optimized using the sequencing methods, mutation types and filters that best suited their own panel specifications. Since the participating panels were in different stages of development, only a few had published panel performance characteristics (table 1).

The WES TMB values were calculated using the TCGA MC3 Mutation Annotation Format (MAF) validation dataset and an agreed on methodology (see 'Whole exome analysis' section and online supplementary table 2). The panel TMB values on the same validation dataset were estimated by down-sampling to the regions covered by each of the laboratories' panels and applying their own bioinformatics algorithms. To prevent the misinterpretation of this study's results as an interlab performance study, all laboratories agreed for the results to be blinded with respect to the lab generating each dataset.

First, all 32 cancer types in the TCGA MC3 dataset were investigated together using weighted linear regression analysis (generalized least squares, see 'Methods' section). Some variation was observed across panels, with Spearman's rank correlation values (R) ranging from 0.79 to 0.88, and slope values ranging from 0.87 to 1.47 (figure 1A, online supplementary figure 5). Eight laboratories (73%) had slope values >1, demonstrating an overestimation of TMB. Panel factors that may influence TMB overestimation were not assessed due to the blinded study design but may have included the type of mutations counted for the panel TMB value (eg, synonymous alterations included in panel TMB that were excluded from the WES estimation), among others.

In silico assessment of theoretical TMB variation across panels by cancer type

A limitation of analyzing all cancer types together is the variable distribution of TMB across different cancer types, with some cancer types displaying large dynamic ranges of TMB values up to several hundred mutations per Mb and others with very limited distributions with very few samples reaching 20 mutations per Mb (see online supplementary figure 3). To account for this limitation, cancer types were categorized into strata by their distribution of WES TMB values. Stratum 1 (n=1563 samples with <40 mut/Mb)

had samples with a good distribution of WES TMB values covering 0–40 mut/Mb, which enabled a more robust regression analysis across a clinically relevant TMB range. The eight cancer types in stratum 1 were: bladder urothelial carcinoma (BLCA, n=195), colon adenocarcinoma (COAD, n=128), head and neck squamous cell carcinoma (HNSC, n=232), lung adenocarcinoma (LUAD, n=228), lung squamous cell carcinoma (LUSC, n=228), skin cutaneous melanoma (SKCM, n=166), stomach adenocarcinoma (STAD, n=189) and uterine corpus endometrial carcinoma (UCEC, n=197).

Regression analyses restricted to stratum 1 tumors revealed an association between WES TMB and panel TMB similar to that for all cancer types analyzed together (Spearman's R: 0.81–0.90 and slope 0.80–1.32, figure 1B, per laboratory online supplementary figure 6 and table 3). The slopes calculated when stratum 1 tumors were analyzed were consistently lower than when all cancers were analyzed. The greatest differences in slope values when comparing slopes estimated for all cancers and for stratum 1 tumors only, were observed for labs 8 (all cancers 1.47 vs stratum 1 1.32) and 9 (all cancers 1.24 vs stratum 1 1.1) (both Δ 0.15), while labs 4 (all cancers 0.904 vs stratum 1 0.897) and 2 (all cancers 1.087 vs stratum 1 1.076) had the least differences (Δ 0.007 and 0.01, respectively). When stratum 1 tumors were analyzed, only six laboratories (55%) reported overestimation of TMB with slope values >1.

Regression analyses with stratum 2 and 3 were not robust, as the WES TMB values did not adequately cover the entire clinically meaningful range (see online supplementary figures 7 and 8, and table 3).

Lastly, the eight cancers in stratum 1 were analyzed separately. UCEC, BLCA and COAD had the broadest range of slope values (UCEC: range 0.755–1.602, Δ 0.847; BLCA: range 1.042–1.79, Δ 0.748; COAD: range 0.75–1.486, Δ 0.736) (figure 2, online supplementary table 4), and most laboratories consistently overestimated these cancer types, with BLCA as the only cancer type for which all 11 laboratories (100%) consistently overestimated their panel TMB values relative to WES TMB. Conversely, LUAD, LUSC and HNSC had the tightest range of slope values with no consistent bias to overestimating or underestimating TMB
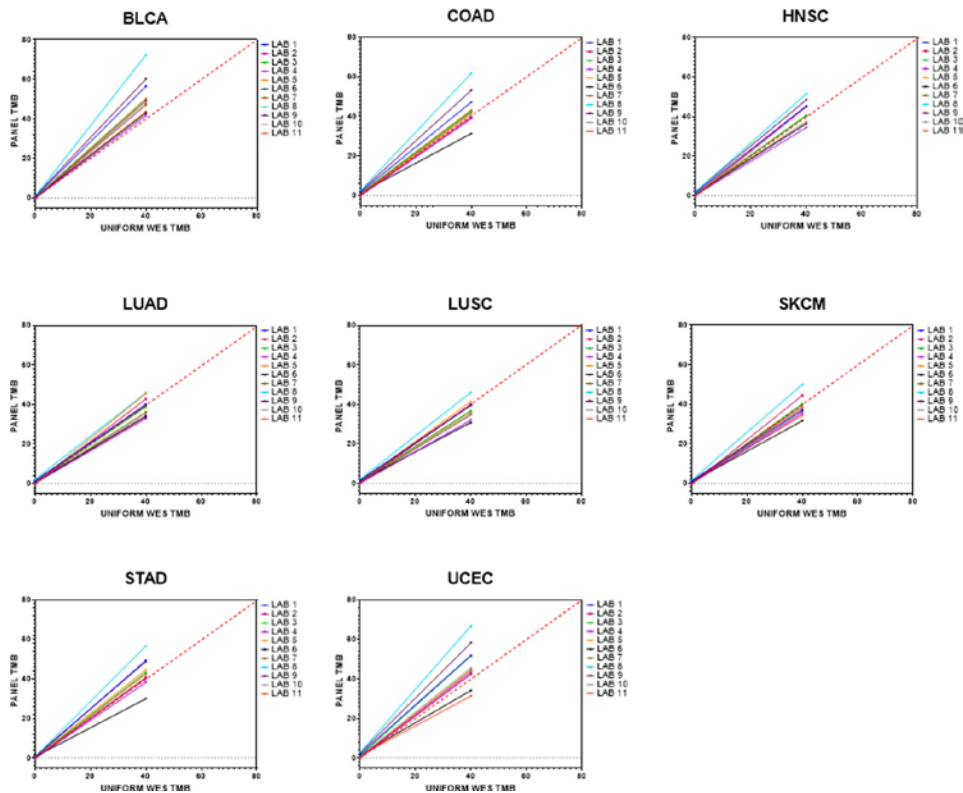
Figure 2 Estimated regression lines for panel tumor mutational burden (TMB) as a function of whole exome sequencing (WES) TMB for the eight cancer types within stratum 1. All cancer types had a good distribution of WES TMB values from 0 to 40 mut/Mb. Solid lines represent the fitted regression lines. Red dashed line represents 45° line. BLCA, bladder urothelial carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; LUAD, lung adenocarcinoma ; LUSC, lung squamous cell carcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; UCEC, uterine corpus endometrial carcinoma.

(LUAD: range 0.817–1.135, Δ 0.318; LUSC: range 0.741–1.099, Δ 0.358; HNSC: range 0.854–1.244, Δ 0.39).

Defining the theoretical variation in TMB across panels and by cancer type

Prediction limits for the observed panel TMB at fixed WES TMB (5, 10, 15 and 20 mut/Mb) were calculated to quantify the variability around the regression line at those selected WES TMB values. The limits were designed to capture approximately 95% of the panel TMB values expected to be observed at a given WES TMB. Some laboratories had consistently tighter (narrower) prediction intervals, while others demonstrated more variability (wider intervals), but for all laboratories, the prediction intervals became tighter with decreasing WES TMB value, indicating greater variability in panel TMB at larger WES TMB values (figure 3). Generally, the prediction intervals observed for each participating laboratory were similar, with laboratories demonstrating intervals that spanned as small as ±4.7 mut/Mb or as large as ±12.3 mut/Mb when the WES TMB was 10 mut/Mb, which is a TMB threshold that has been previously used to define a TMB-high cohort using NGS panels.[33] When prediction limits were assessed by strata, the variability of the intervals was very large for

cancer types in strata 2 and 3 compared with stratum 1 because most TMB values for the cancers in these strata accumulate in the lower end of the TMB spectrum, thus resulting in more uncertainty in the fitted regression lines and wide scatter in panel TMB values around those lines (see online supplementary figure 9). When the eight stratum 1 cancers were analyzed separately, prediction intervals at the discreet value of WES TMB=10 mut/Mb were observed to be wider for BLCA and UCEC, while LUAD, LUSC, HNSC and SKCM had the tightest intervals (figure 4). This is similar to the observed variation in fitted regression lines for BLCA and UCEC across laboratories (figure 2). The theoretical variability around the regression was also seen to increase (wider intervals) with increasing TMB value in individual cancer types (see online supplementary figure 10).

DISCUSSION

Eleven laboratories with distinct NGS targeted gene panels and bioinformatic approaches participated in phase I of the Friends of Cancer Research TMB Harmonization Project and provided early insights into the theoretical variability across different targeted gene panels
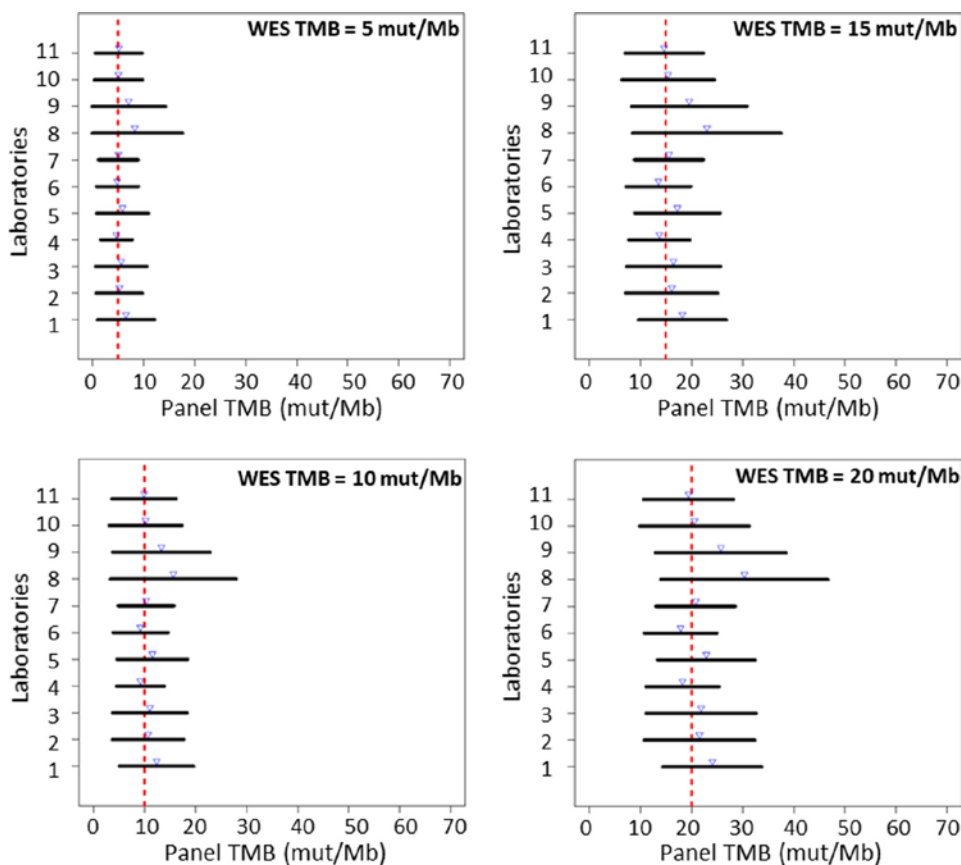
Figure 3 Ninety-five per cent prediction intervals for panel tumor mutational burden (TMB) estimated at discreet whole exome sequencing (WES) TMB values (5, 10, 15 and 20 mut/Mb), by laboratory across all laboratories. Blue arrows represent the estimated mean panel TMB for each laboratory. Red dashed line represents the discreet WES TMB value at which prediction interval is calculated.

that estimate TMB. The goal of the first phase of the project was to describe the variability in TMB estimates across several uniquely designed panel-based diagnostic assays and to further elucidate the theoretical variation in TMB quantification using an in silico approach with a large publicly available dataset with high-quality reads and a common reference TMB standard calculated from the entire exome. Moreover, dependence of the association between panel TMB and WES TMB on cancer type was investigated.

Variability in panel TMB across different panels was observed, with some panels consistently overestimating or underestimating TMB, suggesting that panel size and composition, as well as laboratories' bioinformatics algorithms, including types of mutations counted and variant filters used in the TMB calculation, were likely contributors to the differences. Because of the blinded design of this study, the influence of these factors on panel TMB variability was not evaluated in this early phase of the project but will be assessed in the following empirical phase to be reported subsequently. Additionally, other studies have recently reported on the impact of panel

size, DNA input and variant filtering on panel-based TMB estimates.[35][39]

The study evaluated a robust dataset containing 32 cancer types, with very few cases having TMB >40 mut/Mb (n=69/4134, 1.7%), so it was not possible to robustly estimate the association between panel TMB and WES TMB for cases with values >40 mut/Mb. TMB data were thus capped at 40 mut/Mb and a linear relationship was used to model the relationship in that range. Factoring the limited dynamic range of TMB values observed in some cancer types, a subset of eight cancers was identified and named stratum one for the primary analysis. Stratum 1 cancer types included lung, bladder, head and neck, skin, colon, uterine and gastric cancers, all of which have been shown to respond to immune checkpoint inhibitors. Evaluating these cancers separately revealed distinct levels of variability in the association between panel TMB and WES TMB across panels, with some cancer types having less variability (eg, lung and head and neck cancers), and some having greater variability (eg, uterine, bladder and colon cancers). As our initial findings suggest that panels may perform differently on certain cancer types, further
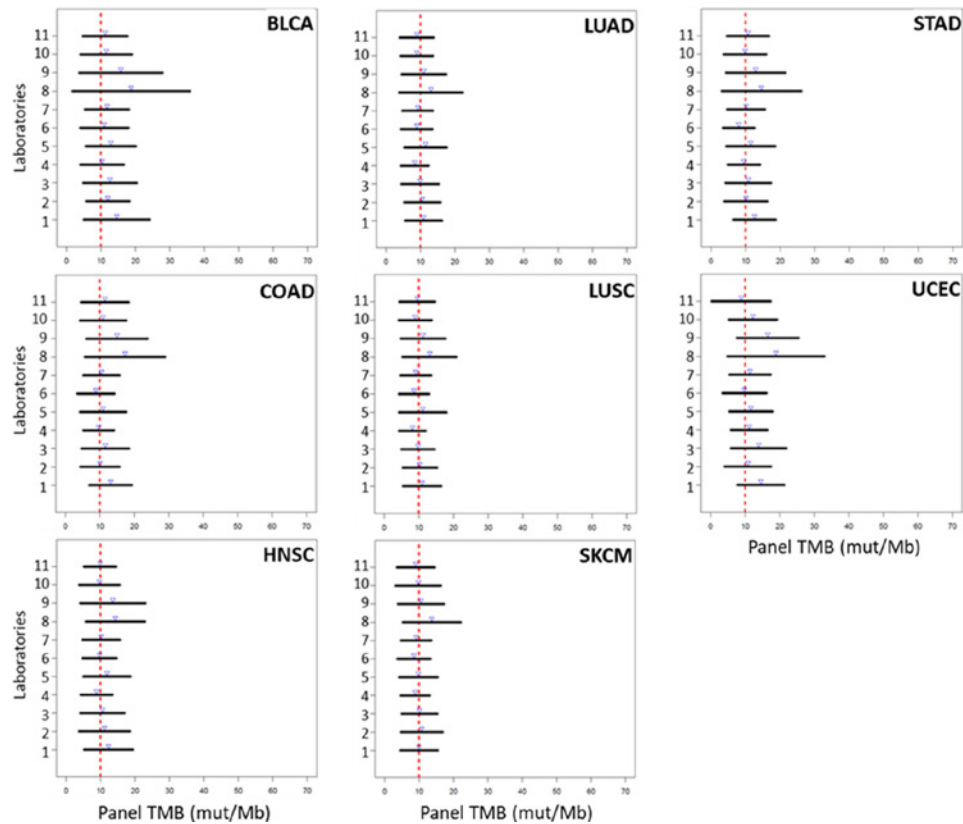
Figure 4 Ninety-five per cent prediction intervals for panel TMB (x-axis) of stratum 1 tumor types at whole exome sequencing tumor mutational burden (WES TMB) 10 mut/Mb by laboratory (y-axis). Blue arrows represent the estimated mean panel TMB for each laboratory. Red dashed line indicates WES TMB value=10 mut/Mb. BLCA, bladder urothelial carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; LUAD, lung adenocarcinoma ; LUSC, lung squamous cell carcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; UCEC, uterine corpus endometrial carcinoma.

work is required to understand the factors contributing to any disease-specific TMB variability, and the relationships beyond the analyzed TMB range. However, the composition of the panels' genes, types of mutations counted or methods used to train their respective TMB algorithms could be future areas of focus.

Despite these cancer-dependent findings, our study found that panel TMB values were strongly correlated with WES TMB across laboratories. Additionally, the calculated 95% prediction intervals permitted estimation of the linear relationship between panel TMB and WES TMB as well as quantification of the range in which 95% of the observed TMB panel values would be expected to fall for tumors with various fixed WES TMB values. This provides a framework for understanding the theoretical variability likely to be incurred in the clinical application of TMB estimation across panels, but also suggests that harmonization of TMB estimates could be achieved through alignment using external reference materials. There is still, however, much that can be done to improve the reliability of using NGS panels for TMB estimation.

The selection of high-quality variants from the TCGA MC3 dataset was used to assess the theoretical variability

of TMB across panels in this study ensuring the interpretability of the findings where the assessment of variability was limited to factors such as panel size and composition or bioinformatics pipeline, instead of perceived differences regarding sensitivity and specificity of individual variant calling. However, we acknowledge that in a clinical setting the estimation of TMB from FFPE tissue may introduce variants of lesser quality and panels should aim to validate the sensitivity and specificity of individual variant calling separately from TMB validation.

As TMB measurements are most likely to be impactful in treatment decisions for stratum one cancer types, including these tumors as part of a laboratory's analytical validation studies to achieve optimal accuracy and consistency is critical. On the other hand, it is also important to recognize that there are cancer types with generally low TMB values that may have a few cases with high TMB values that may benefit from reliable panel TMB results. Moreover, because of the cancer type-dependent distribution of TMB values, studies aiming to evaluate the clinical utility of TMB and determine optimal TMB cut-offs for treatment decisions may need to account for specific cancer types. This would be consistent with a recent

Table 2 Consensus recommendations for the standardization of analytical validation studies of targeted NGS panels that estimate TMB

Parameters	Recommendations
Accuracy*	▶ Accuracy or agreement should be measured by comparing the TMB values generated by the assay requiring validation against reference TMB values generated either from: – A comparable companion diagnostic approved by a regulatory agency, such as the FDA, if available, OR – A WES assay with validated performance characteristics and using an accepted WES TMB calculation method, such as the common method reported in this study (see online supplementary table 1). ▶ The minimum number of samples used for evaluation of accuracy should be at least 30. Samples should have TMB values that span the entire analytical range being investigated (0–40 mut/Mb is recommended currently). ▶ *TMB as a continuous score:* This analysis will characterize the analytical performance of the assay over the analytical range of interest per the intended use. – Quantification of performance based on TMB as a continuous variable should include an appropriate regression analysis and a scatter plot showing the association between the panel and reference TMB values. – Additionally, quantification of performance should examine the pointwise prediction intervals for panel TMB values as obtained from the regression analysis, at predefined reference TMB values. A description of the absolute deviation from the mean should also be reported. ▶ *For TMB as a categorical call:* accuracy should be analytically validated using a single or multiple discreet TMB cut-off values within the analytical range of interest. – Quantification of performance of TMB as a categorical call should be based on 2×2 agreement tables to inform the positive per cent agreement, negative per cent agreement and overall per cent agreement informed by 2×2 agreement tables. – For assays pursuing a companion diagnostic claim, accuracy should be examined using a predetermined discrete cut-off value investigated in a study using clinical samples covering the spectrum of conditions from a defined, or intent-to-treat (ITT) population. If the ITT population includes multiple cancer types, stratified analyses should be conducted. ▶ *Alteration level agreement:* as a supplemental analysis, the agreement between alteration level calls (including single nucleotide variants (SNVs) and short indels) between the platform and the reference should be provided for each alteration included in the TMB panel and restricted to the overlapping genomic regions between the two assays. – Report the concordance of variant calls between variants identified by WES and panel as a function of the panel variant allele frequency (VAF). ▶ Characterize the percentage of tests passing QC by reporting first pass acceptability rate and overall acceptability rate (after samples have been retested, if necessary).

Continued

Table 2 Continued

Parameters	Recommendations

Precision*	▲ Precision should be evaluated using several samples. For each sample, separate analyses should be performed as described in the *TMB as a continuous score* and *TMB as a categorical call* sections below.
	
	▲ Analytical validation of precision of TMB as both a continuous score and a categorical call will improve reliability of TMB as a biomarker.
	▲ Because TMB is a composite estimate composed of different variants, its precision should be evaluated as a composite score (mut/Mb).
	▲ *TMB as a continuous score:*
	– Precision studies of quantitative TMB estimates should evaluate the mean, SD and coefficient of variation of TMB values obtained from testing aliquots of the same sample under stipulated precision conditions (eg, replicates, runs, instruments, lots, operators) for a range of samples (5–6 samples with 20 TMB results distributed across the precision conditions each) with TMB values within the analytical range (0–40 mut/Mb is recommended currently), and include different levels of tumor content and VAF values.
	   – Note: identification of the TMB range to be evaluated should be guided by the most recently published clinically relevant studies.
	– Precision of the TMB score should be estimated using a variance component analysis to estimate between-run, within-run, between instruments, between lots and between operator SD for each sample.
	– Quantification of performance should include calculation of repeatability and within-lab SD for each sample corresponding to several discreet TMB values, given that a single average value of variation (eg, coefficient of variation pooled across several samples having different TMB levels) may not best reflect the changing variability across the TMB range.
	
	▲ *TMB as a categorical call:* consistency of the categorical calls (eg, TMB high vs TMB low according to a discreet cut-off) should be evaluated based on both repeatability and within-laboratory precision for the TMB results according to a single or multiple discreet cut-off values.
	– For repeatability, calculate the per cent of TMB high calls (if majority call is high) or per cent of TMB low calls (if majority call is low) between replicate samples tested under the same lab conditions.
	– For within laboratory precision, calculate the per cent of TMB high calls (if majority call is high) or per cent of TMB low calls (if majority call is low) and the mean TMB score from replicate samples tested under varying within-lab conditions.
	   – Note: the number of aliquots tested per sample should be sufficient to account for the various sources of assay variability, such as the ones described above (TMB as a continuous score). Moreover, the number of samples tested should be similar.
	– Emphasis should be placed on evaluating samples with TMB values:
	   – Significantly below cut-off (approximates limit of blank, expect TMB low, expect TMB low almost 100% of time).
	   – Near and below cut-off (expect TMB low 95% of time).
	   – Near and above cut-off (expect TMB high 95% of time).
	   – Significantly above cut-off (expect very high TMB almost 100% of time).
	
	▲ *Alteration level precision:* the evaluation of precision for each individual alteration call used to estimate TMB is not necessary but may be performed as an exploratory analysis to provide insight into the mechanisms that contribute to the TMB score variability.
	▲ Per cent tumor content should be collected when evaluating precision and reported, if applicable.
	▲ Characterize the percentage of tests passing QC by reporting first pass acceptability rate and overall acceptability rate (after samples have been retested, if necessary).

Table 2 Continued

Parameters	Recommendations
Sensitivity*	▶ The impact of tumor content of a sample on the TMB categorical call (high, low) should be evaluated using multiple samples, taking into consideration the precision of the TMB score as a function of decreasing tumor content. — Undiluted samples should have a range of expected TMB scores, a range of VAF values for somatic mutations and a ratio of SNVs and indels that are representative of clinical samples. ▶ The evaluation of panel sensitivity to tumor content should be done using: — Samples: 6–10 undiluted samples where each sample is diluted to at least 5 levels of tumor content. — Each sample should have a dilution series ranging from well above and below the expected sensitivity limits for tumor content. — Note: It is likely that matched normal will be required to generate each respective dilution value. Consideration should be given to technical and biological factors that may impact the choice of the normal sample and design of the dilution series. — At each dilution level, at least 10 replicate samples should be tested. ▶ For each sample, the evaluation should include a calculation of the per cent of TMB high (above a predefined TMB threshold) across replicates at each dilution and a probit regression of per cent TMB high versus tumor content. From the regression, report the estimated tumor content where the probability of detecting TMB high is 95%.
Limit of blank*	▶ Non-tumor samples should be used to establish the limit of blank for TMB, yielding results close to, but not always equal to 0 mut/Mb. — Considerations should be given to technical and biological factors, such as age of patient and distance from tumor lesion, among others.
Percentage of tests passing QC	▶ Report percentage of tests passing TMB QC metrics in routine testing. ▶ Example QC metrics for TMB might include: median exon coverage, coverage uniformity, contamination rate.

*The definitions of the terms used in this table are based on the Clinical and Laboratory Standards Institute Harmonized Terminology Database at http://htd.clsi.org/.
FDA, Food and Drug Administration; NGS, next-generation sequencing; QC, quality control; TMB, tumor mutational burden ; WES, whole exome sequencing .

report that found that in patients who received ICI, those who had high TMB had longer survival than those who had low TMB, but TMB-high cut-offs were cancer-type dependent.[20]

The Friends of Cancer Research TMB Harmonization Consortium includes the participation of several leading commercial and academic laboratories as well as a diverse group of stakeholders, who together identified opportunities for standardization to promote the harmonization of TMB estimation. These have led the consortium to recommend best practices for panel developers that seek to promote consistency in alignment and facilitate commutability across panels table 2. These recommendations revolve around the following three items.

1. Ensure reporting consistency: TMB should be reported in mutations/megabase (mut/Mb)

The current practice of reporting WES-derived TMB values as number of somatic mutations, while panel-derived TMB values are reported as a density of somatic mutations per Mb of genomic region covered by the panel (mut/Mb), precludes the aggregate analysis of TMB being derived from WES or targeted panels, especially since the size of the exome interrogated using different platforms may not be consistent. Reporting TMB as mutations/megabase (mut/Mb) in order to keep these values consistent and comparable is recommended by the consortium.

2. Analytical validation studies for TMB estimation should be standardized to include assessment of analytical accuracy, precision and sensitivity

Size of targeted gene panel, technical sensitivity of the assay and pre-analytical and analytical variables are known to contribute to variability in panel-based TMB estimates.[33] The same in silico data were used by every participating laboratory, which created a theoretical setting that focused the investigation on potential sources of variability that are unique to the technical specifications of the panel (eg, size and composition), and the bioinformatics approaches of each laboratory (eg, mutation types counted and germline and hotspot mutation filtering). Some of these factors cannot be easily modified and standardized across laboratories as panel assays are, for the most part, proprietary and have been designed to optimize their respective technical specifications and conditions. However, harmonization of TMB estimates may be achieved across laboratories by ensuring that the analytical validation studies for each panel follow a standard approach including alignment of panel TMB values to an external reference standard. Recommendations for analyzing accuracy, precision and sensitivity of TMB values to tumor content when used both as a continuous score and a categorical call have been proposed by the consortium (table 2). These recommendations will ensure that regardless of the type of panel or bioinformatics pipeline a laboratory decides to use, TMB estimates are held to a standard of acceptable reliability.

3. Consistency across panels could be ensured through alignment of panel TMB values to WES-derived universal reference standard

Comparison to WES TMB is currently the most recognized way to determine accuracy of panel TMB. However, it should be noted that differences in performance between panel TMB and WES TMB are to be expected based on differences in coverage depth between the two methods, with typically greater depth and higher variability observed in panels.

Universal reference standards, with TMB values spanning a clinically relevant range (eg, 0–40 mut/Mb), represent a promising tool to achieve alignment or calibration in order to ensure consistency of the TMB estimation across platforms, regardless of known sources of variability. An ideal reference standard for TMB estimation should be generated from a renewable source and its TMB values should be calculated using WES. To mitigate differences resulting from comparisons using multiple different WES assays, a universally accepted, predefined bioinformatics pipeline and statistical methods should be implemented. A calibration curve generated using the reference standard should be used to normalize and compare across panels, which should promote alignment and aid in the analytical validation of panel TMB values.

CONCLUSIONS

Harmonization of methodologies for the accurate measurement of complex continuous biomarkers is an ongoing effort. The Friends of Cancer Research TMB Harmonization Project has convened key stakeholders early in the development of NGS assays that estimate TMB to more effectively identify avenues for the harmonization of estimation approaches and to emphasize the need for the uptake and implementation of these harmonization recommendations. The results included in this report are the initial results from this stepwise approach, but future studies will focus on assessing the feasibility of using tumor-derived cell lines as external reference standards to help facilitate alignment of panel TMB values. Additional empirical analyses will also be conducted to investigate the influence of biologic factors (eg, specimen type, cancer type) and technical factors (eg, sequencing technology) on panel TMB, continue refining best practices for panel assessment of TMB, and developing alignment approaches to improve interchangeability between TMB estimates generated from different targeted gene panels.

Lastly, the collaborative efforts of the TMB Harmonization Consortium will serve as a framework for future harmonization initiatives that seek to standardize complex quantitative biomarker assays and promote the reliability of biomarker testing.

Author affiliations
[1]Friends of Cancer Research, Washington, DC, USA
[2]National Cancer Institute, Bethesda, Maryland, USA

[3]Foundation Medicine Inc, Cambridge, Massachusetts, USA

[4]NeoGenomics Laboratories, Aliso Viejo, California, USA

[5]ACT Genomics, Taipei, Taiwan

[6]Resphera Biosciences, Baltimore, Maryland, USA

[7]Clinical Genomics, Illumina Inc, San Diego, California, USA

[8]Bristol-Myers Squibb Co, Princeton, New Jersey, USA

[9]Translational Medicine, Oncology Research and Early Development, AstraZeneca Pharmaceuticals LP, Boston, Massachusetts, USA

[10]Clinical Sequencing Division, Thermo Fisher Scientific, Ann Arbor, Michigan, USA

[11]Molecular Characterization Laboratory, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

[12]Caris Life Sciences Inc, Phoenix, Arizona, USA

[13]Personal Genome Diagnostics, Baltimore, Maryland, USA

[14]QIAGEN Inc, Aarhus, Denmark

[15]Memorial Sloan Kettering Cancer Center, New York, New York, USA

[16]Bioinformatics, Guardant Health Inc, Redwood City, California, USA

[17]Quality in Pathology (QuIP), Berlin, Germany

[18]QIAGEN Inc, Waltham, Massachusetts, USA

[19]Division of Hematology/Oncology, Department of Medicine, Columbia University, New York, New York, USA

[20]AstraZeneca Pharmaceuticals LP, Waltham, Massachusetts, USA

[21]Institute of Pathology, University Hospital Heidelberg, Heidelberg, Baden-Württemberg, Germany

Twitter Diana M Merino @d2merino

ORCID iD
Diana M Merino http://orcid.org/0000-0002-6173-6957

REFERENCES

1. Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 2015;27:450–61.
2. Sharma P, Allison JP. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* 2015;161:205–14.
3. Le DT, Uram JN, Wang H, *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015;372:2509–20.
4. Delamarre L, Mellman I, Yadav M. Neo approaches to cancer vaccines. *Science* 2015;348:760–1.
5. Brahmer JR. Immune checkpoint blockade: the hope for immunotherapy as a treatment of lung cancer? *Semin Oncol* 2014;41:126–32.

6 Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, Pathologist-Based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. JAMA Oncol 2017;3:1051–8.

7 Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med 2014;371:2189–99.

8 Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 2015;350:207–11.

9 Hugo W, Zaretsky JM, Sun L, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell 2016;165:35–44.

10 Johnson DB, Pollack MH, Sosman JA. Emerging targeted therapies for melanoma. Expert Opin Emerg Drugs 2016;21:195–207.

11 Riaz N, Havel JJ, Makarov V, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. Cell 2017;171:934–49.

12 Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124–8.

13 Kowanetz M, Zou W, Shames DS, et al. Tumor mutation load assessed by FoundationOne (FM1) is associated with improved efficacy of atezolizumab (atezo) in patients with advanced NSCLC. Ann Oncol 2016;27:vi23.

14 Carbone DP, Reck M, Paz-Ares L, et al. First-Line nivolumab in stage IV or recurrent non-small-cell lung cancer. N Engl J Med 2017;376:2415–26.

15 Rizvi H, Sanchez-Vega F, La K, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. J Clin Oncol 2018;36:633–41.

16 Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet 2016;387:1909–20.

17 Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357:409–13.

18 Goodman AM, Kato S, Bazhenova L, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther 2017;16:2598–608.

19 Cristescu R, Mogg R, Ayers M, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science 2018;362:eaar3593.

20 Samstein RM, Lee C-H, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet 2019;51:202–6.

21 Hellmann MD, Nathanson T, Rizvi H, et al. Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. Cancer Cell 2018;33:843–52.

22 Hellmann MD, Ciuleanu T-E, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. N Engl J Med 2018;378:2093–104.

23 Hellmann MD, Callahan MK, Awad MM, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. Cancer Cell 2018;33:853–61.

24 Peters S, Cho BC, Reinmuth N, et al. Tumor mutational burden (TMB) as a biomarker of survival in metastatic non-small cell lung cancer (mNSCLC): Blood and tissue TMB analysis from MYSTIC, a Phase III study of first-line durvalumab ± tremelimumab vs chemotherapy. In: Proceedings of the American association for cancer research AACR Anual meeting 2019. Atlanta, GA, 2019.

25 Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 2017;9:1–14.

26 Johnson DB, Frampton GM, Rioth MJ, et al. Targeted next generation sequencing identifies markers of response to PD-1 blockade. Cancer Immunol Res 2016;4:959–67.

27 Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. Cell 2017;171:1042–56.

28 Fabrizio DA, George TJ, Dunne RF, et al. Beyond microsatellite testing: assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition. J Gastrointest Oncol 2018;9:610–7.

29 Sorscher S, Desnoyers R. A patient with a microsatellite stable (MSS) and high mutational burden metastatic colorectal cancer responding to checkpoint inhibitor therapy. MOJ Clin Med Case Reports 2016;5:230–1.

30 Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703–13.

31 US Food and Drug Administration. Summary of safety and effectiveness data (SSED): FoundationOne CDxTM, 2017. Available: https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019B.pdf

32 U.S. Food and Drug Administration. Evaluation Of Automatic Class III Designation For MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets) - Decision Summary, 2017. Available: https://www.accessdata.fda.gov/cdrh_docs/reviews/den170058.pdf

33 Stenzinger A, Allen JD, Maas J, et al. Tumor mutational burden standardization initiatives: recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions. Genes Chromosom Cancer 2019;58:578–88.

34 Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol 2019;30:44–56.

35 Endris V, Buchhalter I, Allgäuer M, et al. Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels. Int J Cancer 2019;144:2303–12.

36 Stenzinger A, Endris V, Budczies J, et al. Harmonization and standardization of panel-based tumor mutational burden (TMB) measurement: real-world results and recommendations of the QuIP study. J Thorac Oncol 2020. doi:10.1016/j.jtho.2020.01.023. [Epub ahead of print: 28 Feb 2020].

37 Ellrott K, Bailey MH, Saksena G, et al. Scalable open science approach for mutation calling of tumor Exomes using multiple genomic pipelines. Cell Syst 2018;6:271–81.

38 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.

39 Buchhalter I, Rempel E, Endris V, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. Int J Cancer 2019;144:1–11.

40 Vanderwalde A, Spetzler D, Xiao N, et al. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. Cancer Med 2018;7:746–56.

41 Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013;31:1023–31.

42 Fabrizio DA, Milbury C, Yip W-K, et al. Analytic validation of tumor mutational burden as a companion diagnostic for combination immunotherapy in non-small cell lung cancer. Annals of Oncology 2018;29:viii16–57.

43 Quinn K, Helman E, Nance T, et al. Development and analytical validation of a plasma-based tumor mutational burden (TMB) score from next-generation sequencing panels. Ann Oncol 2018;29:viii41.

44 Cheng DT, Mitchell TN, Zehir A, et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn 2015;17:251–64.

45 Wood DE, White JR, Georgiadis A, et al. A machine learning approach for somatic mutation discovery. Sci Transl Med 2018;10:eaar7939.

46 Chaudhary R, Quagliata L, Martin JP, et al. A scalable solution for tumor mutational burden from formalin-fixed, paraffin-embedded samples using the Oncomine tumor mutation load assay. Transl. Lung Cancer Res 2018;7:616–30.

IASLC

Check for updates

Harmonization and Standardization of Panel-Based Tumor Mutational Burden Measurement: Real-World Results and Recommendations of the Quality in Pathology Study

Albrecht Stenzinger, MD,[a,b,*] Volker Endris, PhD,[a] Jan Budczies, PhD, PD,[a]
Sabine Merkelbach-Bruse, PhD,[c] Daniel Kazdal, PhD,[a,b] Wolfgang Dietmaier, PhD,[d]
Nicole Pfarr, PhD,[e] Udo Siebolts, MD, PhD, PD,[f] Michael Hummel, PhD,[g]
Sylvia Herold, PhD,[h] Johanna Andreas, BSc,[i] Martin Zoche, PhD,[j] Lars Tögel, PhD,[k]
Eugen Rempel, PhD,[a] Jörg Maas, Dipl.-Pol.,[l] Diana Merino, PhD,[m]
Mark Stewart, PhD,[m] Karim Zaoui, MD,[n] Matthias Schlesner, PhD,[o]
Hanno Glimm, MD,[p] Stefan Fröhling, MD,[q] Jeff Allen, PhD,[m] David Horst, MD,[g]
Gustavo Baretton, MD,[h] Claudia Wickenhauser, MD,[f] Markus Tiemann, MD,[i]
Matthias Evert, MD,[d] Holger Moch, MD,[j] Thomas Kirchner, MD,[r]
Reinhard Büttner, MD,[c] Peter Schirmacher, MD,[a] Andreas Jung, MD,[r]
Florian Haller, MD,[k] Wilko Weichert, MD,[e] Manfred Dietel, MD[l]

[a]Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[b]Translational Lung Research Center (TLRC) Heidelberg, German Center for Lung Research (DZL), Heidelberg, Germany
[c]Institute of Pathology, University Hospital Cologne, Cologne, Germany
[d]Institute of Pathology, University Regensburg, Regensburg, Germany
[e]Institute of Pathology, Technical University Munich (TUM), Munich, Germany
[f]Institute of Pathology, University Hospital Halle, Halle, Germany
[g]Institute of Pathology, Charité University Hospital, Berlin, Germany
[h]Institute of Pathology, University Hospital Dresden, Dresden, Germany
[i]Institute of Hematopathology, Hamburg, Germany

[j]*Institute of Pathology, University Hospital Zurich, Zurich, Switzerland*
[k]*Institute of Pathology, University Hospital Erlangen, Erlangen, Germany*
[l]*Quality in Pathology (QuIP), Berlin, Germany*
[m]*Friends of Cancer Research (FoCR), Washington, District of Columbia*
[n]*Department of Otorhinolaryngology, University Hospital Heidelberg, Heidelberg, Germany*
[o]*Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany*
[p]*Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT Dresden) and University Hospital Carl Gustav Carus, Dresden, and Translational Functional Cancer Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany and German Cancer Consortium (DKTK), Dresden, Germany*
[q]*Division of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg and German Cancer Research Center (DKFZ), Heidelberg, Germany*
[r]*Institute of Pathology, Ludwig-Maximilians University (LMU), Munich, Germany*

ABSTRACT

Introduction: Tumor mutational burden (TMB) is a quantitative assessment of the number of somatic mutations within a tumor genome. Immunotherapy benefit has been associated with TMB assessed by whole-exome sequencing (wesTMB) and gene panel sequencing (psTMB). The initiatives of Quality in Pathology (QuIP) and Friends of Cancer Research have jointly addressed the need for harmonization among TMB testing options in tissues. This QuIP study identifies critical sources of variation in psTMB assessment.

Methods: A total of 20 samples from three tumor types (lung adenocarcinoma, head and neck squamous cell carcinoma, and colon adenocarcinoma) with available WES data were analyzed for psTMB using six panels across 15 testing centers. Interlaboratory and interplatform variation, including agreement on variant calling and TMB classification, were investigated. Bridging factors to transform psTMB to wesTMB values were empirically derived. The impact of germline filtering was evaluated.

Results: Sixteen samples had low interlaboratory and interpanel psTMB variation, with 87.7% of pairwise comparisons revealing a Spearman's ρ greater than 0.6. A wesTMB cut point of 199 missense mutations projected to psTMB cut points between 7.8 and 12.6 mutations per megabase pair; the corresponding psTMB and wesTMB classifications agreed in 74.9% of cases. For three-tier classification with cut points of 100 and 300 mutations, agreement was observed in 76.7%, weak misclassification in 21.8%, and strong misclassification in 1.5% of cases. Confounders of psTMB estimation included fixation artifacts, DNA input, sequencing depth, genome coverage, and variant allele frequency cut points.

Conclusions: This study provides real-world evidence that all evaluated panels can be used to estimate TMB in a routine diagnostic setting and identifies important parameters for reliable tissue TMB assessment that require careful control. As complex or composite biomarkers beyond TMB are likely playing an increasing role in therapy prediction, the efforts by QuIP and Friends of Cancer Research also delineate a general framework and blueprint for the evaluation of such assays.

Introduction

Immune checkpoint inhibitors (ICIs) have greatly expanded therapeutic options in oncology.[1] Although many clinical trials have reported strong clinical responses across various tumor types, evidence is increasing that even in generally responsive tumor entities, many tumors are resistant at baseline or develop resistance to ICIs, for example, by immunoediting.[2] Moreover, adverse events associated with ICIs have been noted, particularly with combinatorial regimens that target cytotoxic T lymphocyte–associated protein 4 in addition to programmed cell death protein 1 or programmed death-ligand 1 (PD-L1).[3] Collectively, these observations argue for a sophisticated biomarker approach that reflects the interplay between the host's immune system and the cancer cells and is able to reliably separate likely responders from nonresponders.

To date, two predictive ICI-specific biomarkers have been approved in certain cancer types, which are as follows: (1) PD-L1, assessed by immunohistochemistry (IHC) with a wide range of different scoring systems and cut points depending on cancer type–specific trial results, and (2) high-level microsatellite instability or mismatch repair deficiency, assessed by either polymerase chain reaction (PCR) or IHC.[4,5] Whereas the former approach measures a continuous variable that serves as an approximation for T-cell anergy or tumor cells escaping immune response, the latter identifies a subgroup of cancers with a high

mutational burden and thus increased neoantigen load, which likely results in a higher propensity of immune cell–mediated tumor cell killing.

However, many cancer types, including NSCLC, do not harbor deleterious mutations in one of the DNA mismatch repair genes but have increased tumor mutational burden (TMB) associated with higher loads of neoantigens, which is caused by DNA damage through external noxae (e.g., ultraviolet light and smoking) or deleterious mutations affecting other DNA repair genes.[6]

Although clinical trials assessing the utility of TMB prospectively are ongoing, many retrospective analyses of individual patient cohorts and clinical trials have reported that TMB can be successfully used for patient stratification. Initial seminal studies employed whole-exome sequencing (WES) to measure TMB.[7-10] Because this approach has several limitations, including sample requirements, necessity for concurrent germline sequencing, extensive laboratory capacity for diagnostic application, and economic constraints in consideration of a diagnostic outreach setting, gene panels were designed and used to estimate TMB values, primarily in formalin-fixed and paraffin-embedded (FFPE) tissue and, more recently, in cell-free circulating tumor DNA.[11-13] Such assays have been successfully used under controlled trial conditions or at specific academic cancer centers. However, a detailed evaluation of the overall performance of commercially available sequencing panels that can be used as laboratory-developed tests and of the parameters affecting its diagnostic applicability is missing.

To address this important issue, we present the results of the multi-institutional Quality in Pathology (QuIP) study on a comparative assessment of TMB estimated by gene panel sequencing (psTMB) from 11 different institutes of pathology and four industrial laboratories. Analyzing 20 different FFPE cancer samples from routine diagnostics that reflect the full continuum of TMB, as measured by WES (wesTMB), provides real-world data on the following six different targeted gene panels designed for TMB estimation: Oncomine Tumor Mutational Load Assay (OTML; Thermo Fisher Scientific, Waltham, MA), QIAseq TMB panel (QIAseq; QIAGEN GmbH, Hilden, Germany), NEOplus RUO assay (NEOplus; NEO New Oncology, Cologne, Germany), TruSight Oncology 500 panel (TSO500; Illumina, San Diego, CA), a custom-designed academic panel (ACADEMIC; Agilent, Santa Clara, CA), and the FoundationOne assay (F1; Foundation Medicine, Cambridge, MA). Together with the efforts led by the Friends of Cancer Research,[14,15] this study sets the basis for harmonization of panel-based TMB measurement and supports implementation of TMB in routine diagnostic laboratories.

Materials and Methods

Samples

All patients provided written informed consent under an institutional review board–approved protocol, and the study was conducted in accordance with the Declaration of Helsinki. FFPE tissue specimens of 10 lung adenocarcinoma (LUAD), seven head and neck squamous cell carcinoma, and three colon adenocarcinoma were prepared and diagnosed at the Institute of Pathology Heidelberg, Germany. See Supplementary Table 1 for further detailed sample information. Only one block per tumor was selected, and consecutive sections were used for DNA extraction by the different laboratories. Tumor content was controlled using hematoxylin and eosin–stained slides on the first and last sections to ensure homogeneity throughout the slices.

Library Preparation and Sequencing

Protocols for the six applied panel-sequencing approaches (OTML, QIAseq, NEOplus, TSO500, ACADEMIC, and F1) and for WES are detailed in the Supplementary Materials and Methods (Supplementary Table 2). All assays were performed according to the manufacturers' protocols if not specified otherwise.

Data Analysis and Visualization

Data analysis and visualization were performed using the statistical programming language R (version 3.51).[16] Levels of psTMB were visualized as boxplots and as heatmaps, including hierarchical clustering of experiments (Manhattan distance, average linkage clustering). Spearman's correlations (ρ) and Pearson's correlations (R) of psTMB were calculated between pairs of experiments, clustered (Euclidean distance, average linkage clustering), and visualized as heatmaps. Error bars were plotted using the function plotCI from the R package gplots. Violin plots were generated using the R package vioplot.

Linear models without intercept were fitted to psTMB levels with wesTMB levels. Measurement of psTMB is influenced by different factors. Although misclassification of germline mutations as somatic mutations is independent of the TMB level, other factors, including the subsampling error caused by interrogation of only a limited part of the coding sequence, increase with a higher TMB.[17] Because the exact shape of the mathematical dependence of the TMB error on the level of the TMB is not known, linear models were fitted in the following two different ways: (1) standard linear regression (least square regression, LS) corresponding to constant error contributions, and (2) weighted linear regression (weighted least squares, WLS) with weights equal to the

reciprocal of TMB taking into account heteroscedasticity. The shape of the weights used in the WLS model reflects the mathematical law for the variation of psTMB that we recently uncovered and described—a linear increase of the variation of psTMB proportional to the level of TMB.[17]

Results

Study Outline

In this study (Fig. 1), FFPE tissue samples of 20 tumors (Supplementary Table 1) with existing matched WES data were analyzed using four commercial panel-sequencing TMB assays (Supplementary Table 2). Each assay was run by four different pathology laboratories and a reference laboratory of the panel provider on all samples. In addition, three pathology laboratories tested the ACADEMIC assay, and all samples were analyzed using the F1 assay. The analyzed study cohort was selected to represent the full spectrum of TMB values as characterized by The Cancer Genome Atlas for LUAD,

head and neck squamous cell carcinoma, and colon adenocarcinoma, but it has a higher proportion of tumors with an intermediate TMB (100–300 mutations) (Supplementary Figure 1). In total, panel sequencing and psTMB measurement were successful in 467 of the 480 performed analyses (97.3%).

TMB Levels and Correlations

Measurements of psTMB in the 20 tumor tissue samples ranged between 0 and 244 mutations per megabase pair (muts/Mbp) with a median of 9.2 muts/Mbp (Fig. 2A). With respect to interlaboratory and interpanel variance, four of the tumor samples (T4, T7, T13, and T15) stood out by having a larger interquartile range of psTMB compared with the remaining samples. This was mainly owing to unfavorable preanalytic quality parameters (degraded DNA or low tumor cellularity) (Fig. 2A). Two samples (T4 and T15) had a large interlaboratory variance of psTMB when each of the panels was analyzed separately, whereas this was not
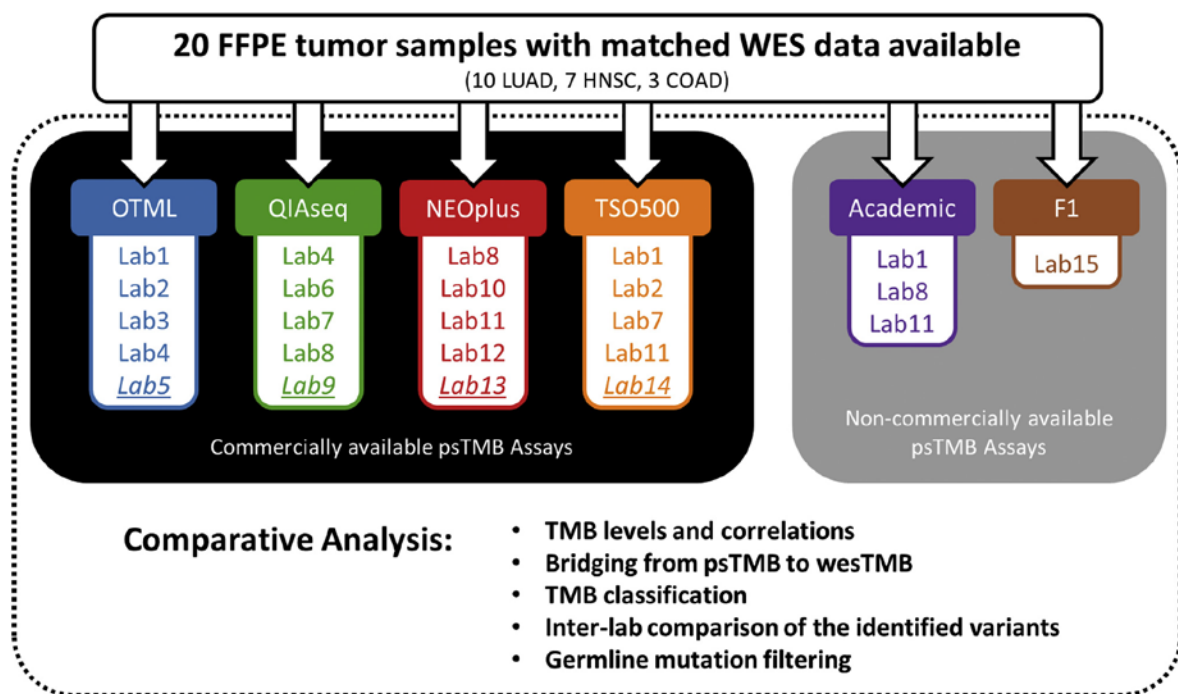
Figure 1. Outline of the QuIP TMB harmonization study. In this comparative study, FFPE tissue samples from 20 tumors were analyzed using four commercial panel-sequencing TMB assays. Each assay was tested by four independent pathology laboratories using all 20 samples and by a reference laboratory of the panel provider. In addition, all samples were analyzed using an "ACADEMIC" assay in three hospital laboratories and by applying the F1 assay. The study cohort consisted of 20 samples from patients with LUAD (n = 10), HNSCC (n = 7), and COAD (n = 3). For all tumors, wesTMB using fresh-frozen tumor tissue samples and paired blood samples was available. ACADEMIC, custom-designed academic panel; COAD, colon adenocarcinoma; F1, FoundationOne assay; FFPE, formalin-fixed and paraffin-embedded; HNSCC, head and neck squamous cell carcinoma; LUAD, lung adenocarcinoma; NEOplus, NEOplus RUO assay; OTML, Oncomine Tumor Mutational Load Assay; psTMB, TMB assessed by gene panel sequencing; QIAseq, QIAseq TMB panel; QuIP, Quality in Pathology; TMB, tumor mutational burden; TSO500, TruSight Oncology 500 panel; WES, whole-exome sequencing; wesTMB, TMB assessed by WES.
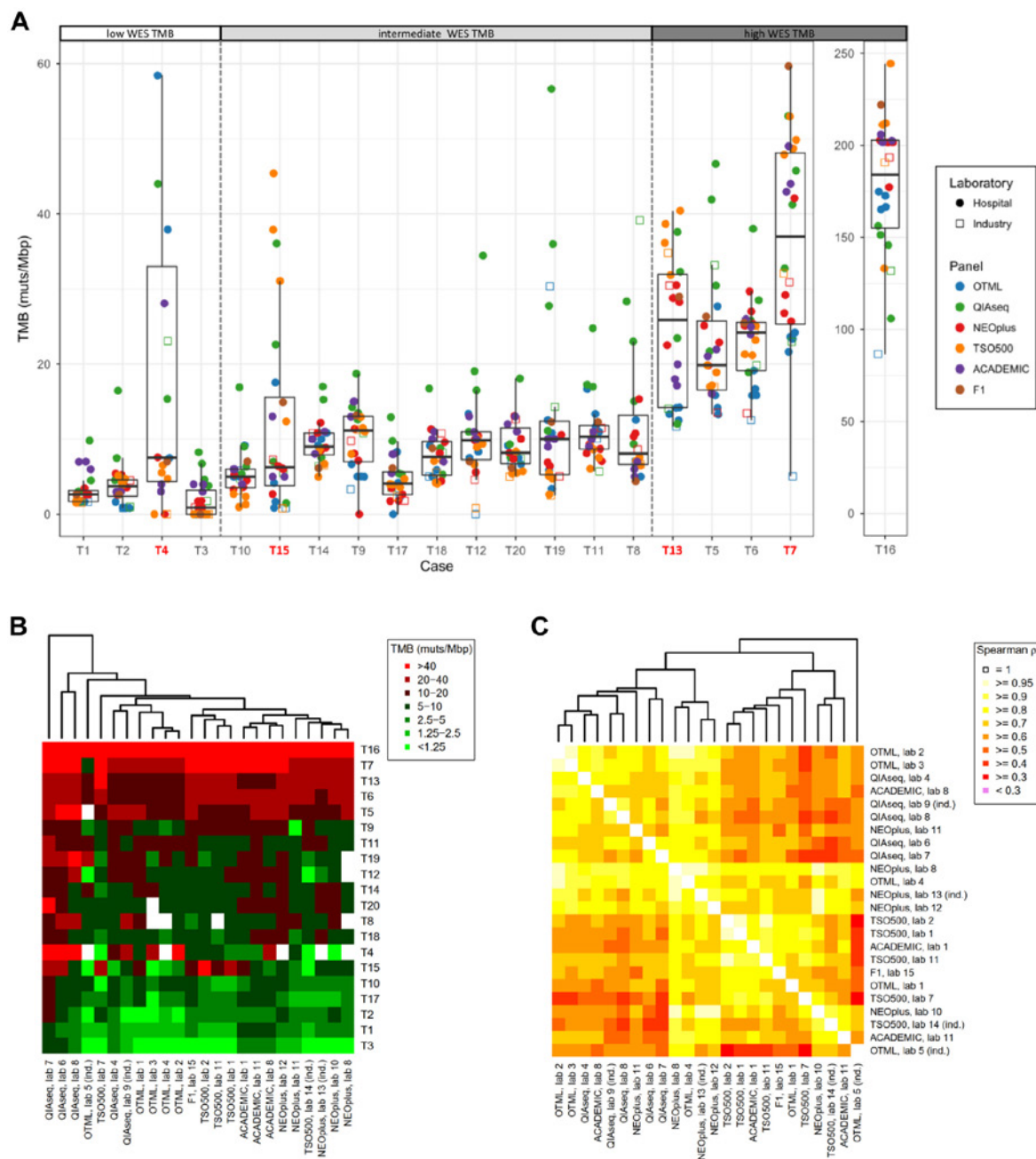
Figure 2. (*A*) Overview of the generated psTMB estimates with tumors ordered by increasing wesTMB levels. Applying a three-tier classification system, four tumors (T1–T4) were classified as TMB low (<100 missense mutations), 11 tumors were classified as TMB intermediate (100–300 missense mutations), and five tumors (T5–T7, T13, and T16) were classified as TMB high (≥300 missense mutations). Four samples stood out by high interquartile ranges and are marked by red IDs. Preanalytic quality parameters were unfavorable for three of these samples (T15: low tumor cellularity; T4 and T13: degraded DNA). (*B*) Heatmap of psTMB levels. Red color indicates psTMB level greater than 10 muts/Mbp. Green color indicates psTMB level less than 10 muts/Mbp. White color indicates insufficient DNA quality. (*C*) Spearman's correlations between psTMB and wesTMB levels in the study cohort. ACADEMIC, custom-designed academic panel; F1, FoundationOne assay; muts/Mbp, mutations per megabase pair; NEOplus, NEOplus RUO assay; OTML, Oncomine Tumor Mutational Load Assay; psTMB, TMB assessed by gene panel sequencing; QIAseq, QIAseq TMB panel; QuIP, Quality in Pathology; TMB, tumor mutational burden; TSO500, TruSight Oncology 500 panel; WES, whole-exome sequencing; wesTMB, TMB assessed by WES.

the case for the two other tumor samples (T7 and T13), in which interpanel variance was an important confounder (Supplementary Figure 2).

In a heatmap, including hierarchical clustering of the psTMB levels, data readouts based on the same sequencing panel often clustered together, indicating independence from the operating laboratory (Fig. 2B). Among most of the sequencing results, moderate to strong pairwise correlations of psTMB measurements were observed: of all pairwise Spearman correlations, 65.9% were greater than or equal to 0.7, 87.7% were greater than or equal to 0.6, and 95.7% were greater than or equal to 0.5 (Fig. 2C). In the study cohort, the strength of Pearson's correlations was dependent on the inclusion or exclusion of a single sample (T16, *POLE*-mutated colorectal carcinoma) that had a very high TMB (>100 muts/Mbp) (Supplementary Figure 3). Hence, the Spearman's correlation was a more suitable approach for the measurement of the psTMB correlations than the Pearson's method.

Bridging From psTMB to wesTMB

Linear regression models were fitted for bridging from psTMB to wesTMB (Fig. 3). To this end, we performed LS but also WLS (see the Materials and Methods section for details) for each of the panels tested in the study. Bridging factors (BFs) for transformation of psTMB to wesTMB were calculated as reciprocals of the regression slopes (Supplementary Table 3). For most of the assays, the BF determined by WLS was very close to the BF determined by LS. However, for the ACADEMIC assay, the WLS BF was slightly lower than the LS BF (17.7 versus 19.8), whereas it was considerably lower for the QIAseq assay (15.8 versus 25.6).

A clinically relevant psTMB cut point of 10 muts/Mbp in NSCLC was established in the CheckMate 568 study using the F1 panel, evaluated in the CheckMate 227 study, and bridged to a wesTMB cut point of 199 mutations using data from the CheckMate 026 study.[18-20] Based on these findings, psTMB cut points corresponding to 199 mutations were calculated for each of the investigated assays (Supplementary Table 3). For most of the psTMB assays, the calculated cut points were consistently in the range of 9.4 to 11.5 muts/Mbp. There were two exceptions, as follows: considerably different cut points were obtained for the OTML assay (LS: 7.8 muts/Mbp, WLS: 7.9 muts/Mbp) and the QIAseq assay (LS: 7.8 muts/Mbp, WLS: 12.6 muts/Mbp).

TMB Classification

Next, we evaluated and compared a two-tier system with a three-tier system for TMB classification (Fig. 4) after a recent indication to improve the misclassification ratio. For the two-tier approach, a dichotomization into "low TMB" and "high TMB" was conducted using a wesTMB cut point of 199 mutations. The three-tier approach classified TMB as "low" (<100 mutations), "intermediate" (100–300 mutations), and "high" (≥300 mutations). Classification with alternative cut points (150 and 250 mutations) is found in Supplementary Figure 4. For each of the panel-sequencing platforms, psTMB values were converted to wesTMB values using the BFs obtained by WLS regression. Altogether (20 samples × 24 experiments), we observed an agreement between psTMB and wesTMB classifications in 74.9% of the cases using the two-tier approach. For the three-tier approach, a "strong misclassification" was defined by a high TMB tumor classified as low TMB or vice versa (difference spanning two tiers), whereas a misclassification by a single tier (e.g., intermediate TMB to low TMB) was termed "weak misclassification." Here, we observed an agreement in 76.7% of cases, compared with a weak and strong misclassification in 21.8% and 1.5% of the cases, respectively. Of note, strong misclassification occurred only for a single tumor sample (T4) that was classified as low TMB by WES but as high TMB in seven psTMB assays and was not analyzable in five psTMB approaches. Assessment of this tumor (LUAD) was priori expected to be challenging owing to highly degraded DNA.

TMB classifications using BFs determined either by WLS or LS regression were similar, as LS regression resulted in 74.3% agreement for two-tier classification and 75.0% agreement, 23.1% weak misclassifications, and 1.9% strong misclassifications for the three-tier classification.

Interlaboratory Comparison of the Identified Variants

In-depth analysis of called variants included in the calculation of TMB identified key factors that influence precise psTMB estimation from the FFPE tissue (Fig. 5). A sequencing approach without an application for PCR duplicate removal, known as deduplication, has a higher probability of erroneous calling of C>T or G>A fixation artifacts and subsequent overestimation of psTMB, especially in highly fragmented, low-quality DNA samples. Methods for deduplication include specialized software solutions and the use of unique molecular identifiers (or molecular barcodes).

False-positive variants in the generated data set were identified by a side-to-side comparison of all variants identified by the different laboratories using the same panel. Variants were classified into nonreproducible variants (detected by a single laboratory), partially reproducible variants (detected by more than one laboratory,
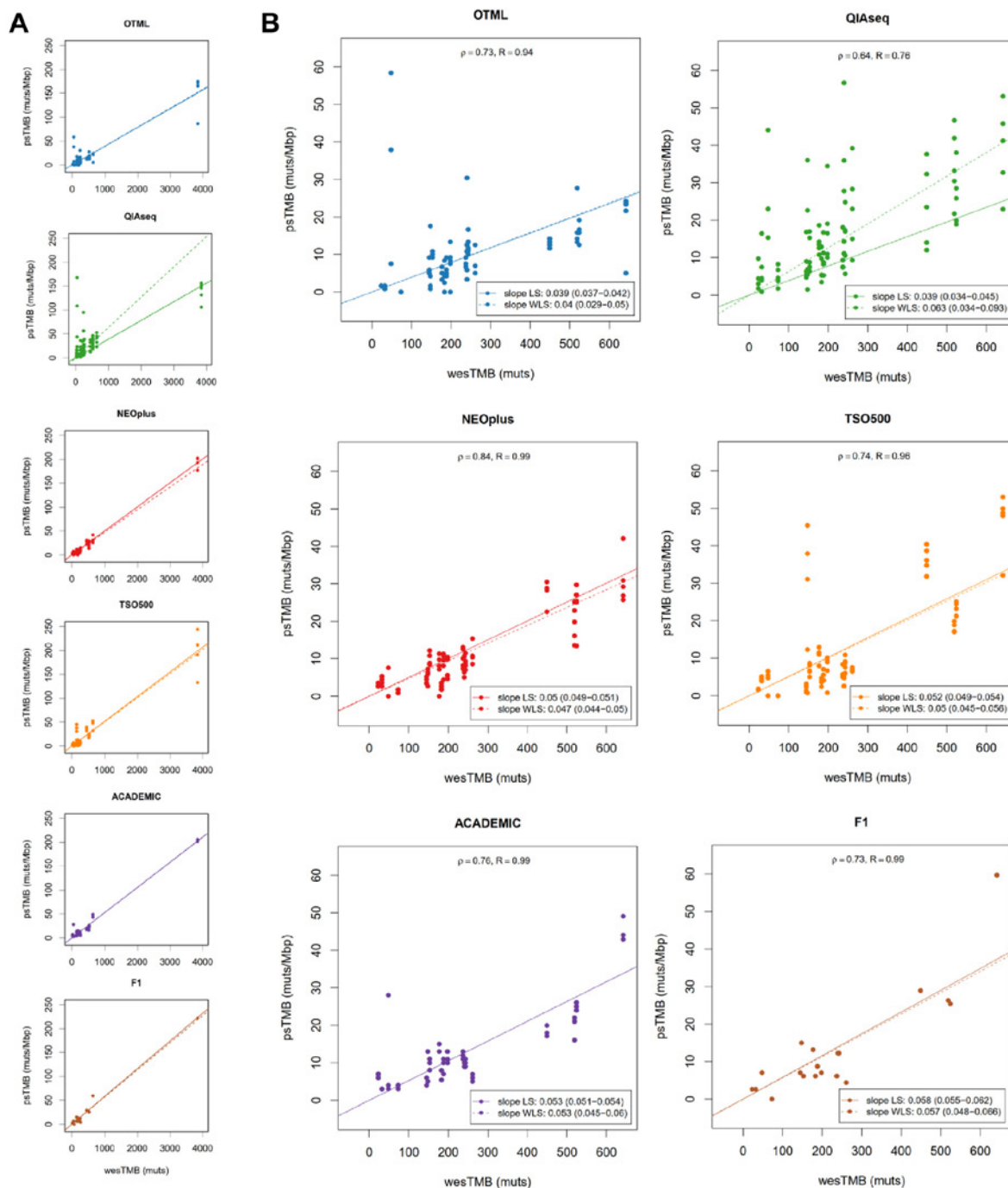
Figure 3. Calibration of TMB measured by psTMB against wesTMB. Linear fits using LS and WLS regression. (A) Overview plots revealing all psTMB and wesTMB measurements. (B) Zoom-ins to the intervals (0, 650) of wesTMB and (0, 65) of psTMB. The intercepts in the linear regression models were set to zero. ACADEMIC, custom-designed academic panel; F1, FoundationOne assay; LS, least squares; muts/Mbp, mutations per megabase pair; NEOplus, NEOplus RUO assay; OTML, Oncomine Tumor Mutational Load Assay; psTMB, TMB assessed by gene panel sequencing; QIAseq, QIAseq TMB panel; TMB, tumor mutational burden; TSO500, TruSight Oncology 500 panel; WES, whole-exome sequencing; wesTMB, TMB assessed by WES; WLS, weighted least squares.
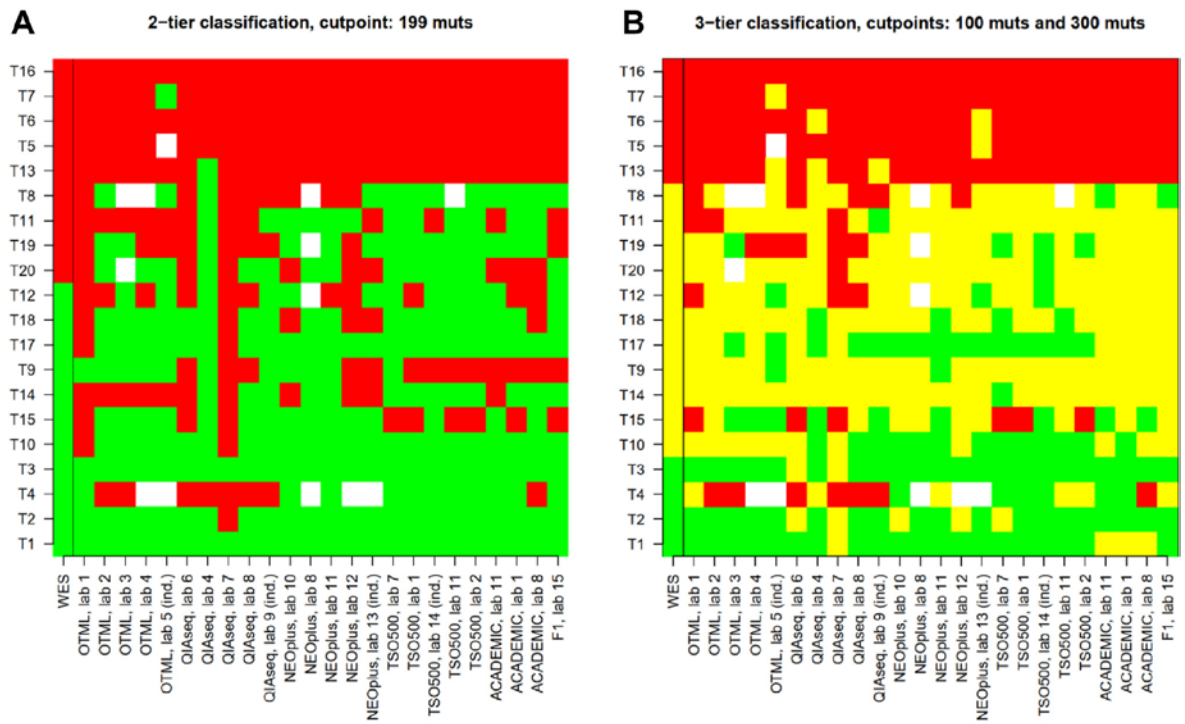
Figure 4. TMB classification by panel sequencing compared with TMB classification by WES. Measurements of psTMB were converted to wesTMB using the bridging factors in Supplementary Table 3. (A) Two-tier classification using the cut point of 199 mutations. Misclassifications: 25.1%. (B) Three-tier classification using the cut points of 100 and 300 mutations. Red indicates high TMB, yellow indicates intermediate TMB, and green indicates low TMB. Strong misclassifications (=misclassifications mixing TMB high and TMB low cases): 1.5%. Weak misclassifications (=misclassifications mixing intermediate TMB cases with TMB high or TMB low cases): 21.8%. ACADEMIC, custom-designed academic panel; F1, FoundationOne assay; NEOplus, NEOplus RUO assay; OTML, Oncomine Tumor Mutational Load Assay; psTMB, TMB assessed by gene panel sequencing; QIAseq, QIAseq TMB panel; TMB, tumor mutational burden; TSO500, TruSight Oncology 500 panel; WES, whole-exome sequencing; wesTMB, TMB assessed by WES.

but not by all laboratories), and fully reproducible variants (detected by all laboratories). Variant allele frequencies (VAFs) were considerably lower for variants with low degrees of interlaboratory reproducibility, and many of the nonreproducible variants had VAFs close to the VAF cut point (Fig. 5A). Thus, low-frequency variants close to the VAF cut point contributed considerably to psTMB variation. To minimize the rate of false-positive calls, specific thresholds for VAF were used for each panel according to the assay provider: VAFs greater than or equal to 10% was applied for the OTML and NEOplus panels and VAFs greater than or equal to 5% for the remaining panels. The number of nonreproducible variants was considerably higher for the OTML assay (3497 variants), which did not include deduplication, compared with the other assays (QIAseq: 1055; NEOplus: 94; TSO500: 70; ACADEMIC: 691). In addition, as illustrated in Figure 5B, the ratio of C>T or G>A transitions was considerably higher for nonreproducible variants (red pie charts) detected by the OTML panel (86%) compared with the other panels (22%–42%), and compared with the ratio

of C>T or G>A of variants that were detected by all laboratories (gray pie charts). These data identify paraffin fixation artifacts and resulting C>T or G>A transitions as important parameters contributing to false-positive variant detection for assays that do not employ deduplication.

False-negative calls (defined here as mutations called by all but one laboratory) can be connected to insufficient depth of coverage at the respective positions. Because the pipelines for capture-based fragment libraries typically include deduplication and unique molecular identifier filtering, the depth of coverage directly correlates with the amount of DNA input, as found representatively for the TSO500 panel in Figure 5C. Here, the median exon coverage that could be analyzed was significantly higher ($p < 0.01$) in laboratory 1 using 80 ng as DNA input compared with 40 ng that was used for the other TSO500 approaches (laboratories 2, 7, 11, and Illumina). Furthermore, the amount of DNA input had a strong impact on the average size of the covered sequencing region (Fig. 5C, middle). Although the
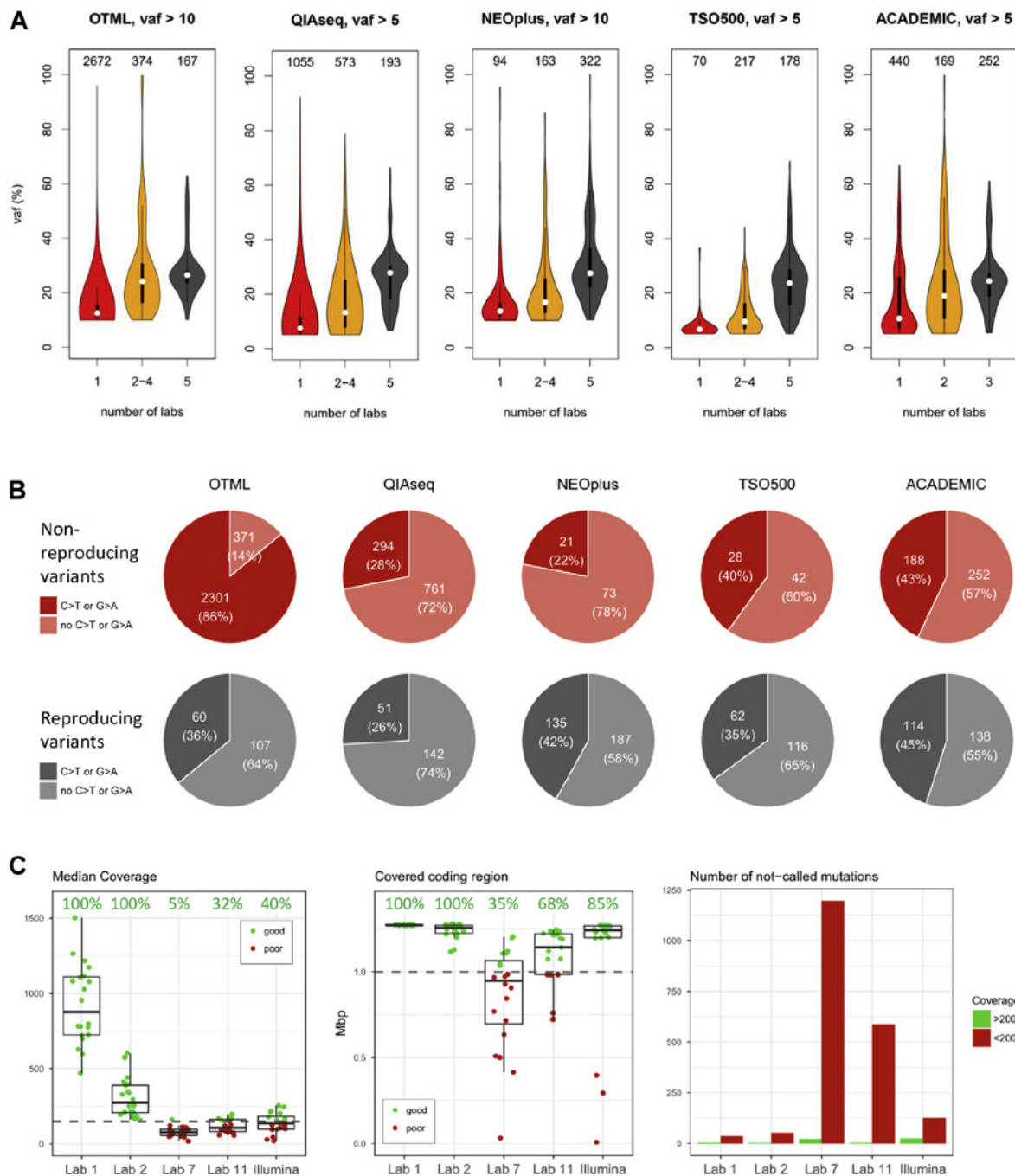
Figure 5. Interlaboratory reproducibility of the detected mutations (pooled analysis of 20 samples). (*A*) Distribution of VAFs in dependence of the number of laboratories that detected the mutation. (*B*) Mutation type (C>T, G>A, or other) of the mutations detected by a single laboratory. (*C*) Impact on DNA input is representatively revealed for the TSO500 panel. DNA input: 80 ng (laboratory 1), 40 ng (labs 2, 7, 11, and Illumina). Left: Median exon coverage for each sample; the number on top gives the percentage of cases with a median exon coverage of more than 150 times. Middle: covered coding region size for each sample. The number on top gives the percentage of cases with a covered coding region of more than 1.0 Mbp. Right: coverage of mutations not-called by a single laboratory (false negatives). ACADEMIC, custom-designed academic panel; NEOplus, NEOplus RUO assay; OTML, Oncomine Tumor Mutational Load Assay; QIAseq, QIAseq TMB panel; TSO500, TruSight Oncology 500 panel; VAF, variant allele frequency.

maximum covered coding region size of 1.28 Mbp was reached for all samples using 80 ng (laboratory 1), lower DNA input resulted in significantly ($p < 0.01$) lower covered coding region sizes, which were larger than 1.0 Mbp in 35% to 100% of the analyzed samples. To enhance specificity, only mutations with minimum coverage of 100 times were included in the psTMB calculation. Therefore, and connected to the lower coverage, we observed a higher rate of false-negative variants in analyses using 40 ng DNA (laboratories 2, 7, 11, and Illumina) compared with 80 ng (Fig. 5C, right). Similar findings were seen for 100-ng versus 200-ng DNA input using the NEOplus assay (data not shown).

Germline Mutation Filtering

Germline mutation filtering is an important step in the calculation of psTMB because only the tumor's somatic mutations are relevant for recognition by the immune system. In the absence of sequencing of paired normal tissue or blood samples in most diagnostic scenarios, germline mutation filtering needs to be performed in silico. For all assays evaluated in the current TMB harmonization study, the bioinformatic pipelines include a step of negative filtering for entries in single-nucleotide polymorphism (SNP) databases, such as gnomAD, ExAC, and dbSNP. In addition, some of the pipelines include further steps, for example, filtering by algorithms specifically designed to distinguish somatic versus germline mutations such as somatic-germline zygosity or filtering with respect to the mutations detected by panel sequencing of reference cohorts of normal samples (e.g., NEOplus and ACADEMIC panel).[21] We evaluated the performance of filtering using SNP databases for the LUAD samples (n = 10) (Supplementary Figure 5). Variants detected by WES in matched blood samples were used as a reference. The sensitivity for classifying mutations as somatic was 87%, 90%, and 79%, with corresponding positive predictive values of 90%, 90%, and 91% when using gnomAD, ExAC, and dbSNP for filtering (pooled analysis of the 10 tumor samples). Filtering out only common SNPs (minor allele frequency > 0.001 in gnomAD) increased sensitivity to 98% but decreased positive predictive value to 81%.

Although germline mutation filtering using gnomAD and ExAC performed well, rs-filtering (dbSNP) seemed to be too stringent. Restriction of filtering to common SNPs considerably decreased the number of false negatives but increased the number of false positives. Additional filters that are implemented in the panel-specific bioinformatic pipelines, such as somatic-germline zygosity algorithm or the TSO500 "proxy filter" (Supplementary Figure 6), can further improve germline mutation filtering.

Discussion

Tumor versus matched blood WES was used in many initial clinical immuno-oncology studies and may be considered a reference standard for TMB assessment. However, clinical implementation of WES-based TMB assessment may be impractical, considering the financial costs and the limited availability of appropriately preserved samples or quality DNA, and matched normal samples for germline sequencing. Gene panel assays offer a number of economical and practical advantages for clinical assessment of patient samples, including increased sequencing depth, in silico germline subtraction (negating the requirement for matched samples), and concurrent evaluation of actionable mutations.

The QuIP study provides a thorough analysis of real-world performances of six select TMB panels. Using real-world diagnostic FFPE samples, which included different types of challenging cases with poor DNA quality, heavy fixation artifacts, or low tumor cellularity, our results reveal that, in principle, all assays tested in this study were able to estimate TMB values and could be applied in a diagnostic setting.

The effect of panel size and coverage on the accuracy of psTMB assessment has been previously studied using in silico simulations of gene panels derived from WES data.[17,22] The gene panels used in the laboratory-developed tests covered at least 1 Mbp of the coding sequence, which was found to be essential for valid panel-based TMB assessment.[22] However, even with these large panels, variability of the TMB score can be expected because psTMB measurement has a probabilistic nature: the overall TMB is extrapolated by investigating only a fraction (about 1:30) of the exome. Simulating five commercial panels in WES data, only 17% to 28% of additional error occurred on top of the unavoidable probabilistic error, demonstrating that sufficient panel size is more critical than the particular localization of the panel in the exome.[17]

There is a multitude of other wet-laboratory parameters, ranging from biological factors (e.g., tumor heterogeneity) and preanalytics (e.g., DNA quality) to sequencing (e.g., coverage) and bioinformatics parameters (e.g., germline subtraction) that can influence TMB scores.[23-26] Hence, as expected, absolute TMB values slightly varied. This scenario is not unknown to pathology in general and immune oncology response prediction in particular: just as for TMB, the established PD-L1 IHC assay for NSCLC quantifies a continuous variable in tumor cells ranging from 0% to 100% PD-L1 expressing cells, and several parameters, such as tumor heterogeneity and fixation, are known to influence PD-L1 scores.[4,27] Just as with PD-L1, for clinical purposes, TMB as a continuous measure must be categorized. In our

approach, we stratified samples into one of three groups, which categorized the continuum of TMB ranging from 0 to greater than 200 muts/Mbp: low, intermediate, and high TMB, according to a concept proposed by us recently.[17] In contrast to a two-tier system with one defined cutoff, this concept allows for a definition of a certain "intermediate" gray zone of TMB measurements in an area around the currently proposed clinical cut point. Using cut points of 100 mutations (corresponding to approximately 5 muts/Mbp) and 300 mutations (corresponding to approximately 15 muts/Mbp), strong misclassifications occurred only for a single tumor sample (T4), a case that was particularly challenging because of poor DNA quality, which would justify to decline analysis in a clinical setting. Misclassification of other highly degraded samples (T12 and T19) or critical cases with a low tumor-cell content (T15), high-level microsatellite instability status (T13 and T15), or a loss-of-function mutation in *POLE* (T16) was prevented using the three-tier system instead of the two-tier system.

We observed an overall low influence of the specific laboratory performing the analysis; data generated by the industrial partners for their specific panel were in the range of the respective TMB scores determined by the hospital laboratories. Moreover, we found that most panels had moderate to strong correlations with TMB measured by WES ($\rho = 0.64–0.84$).

Our study also found that germline subtraction using bioinformatic pipelines can be used to identify likely somatic variants in the probabilistic setting of psTMB measurement. Nevertheless, as revealed by us recently, incorrect filtering can influence TMB scores in individual cases, and future studies are warranted to further investigate the influence of in silico versus blood-based subtraction of germline events.[25,28] As current germline variant databases are biased toward, for example, white populations, ethnicity-related aspects require careful analysis in this context.

We identified assay-independent and assay-specific parameters (Fig. 6) that will require careful control when psTMB is implemented in routine diagnostics. Of these, the effects of tumor-cell content, DNA input, and coverage are most critical to prevent the miss of mutations which would result in too low psTMB scores. Another important aspect are deamination artifacts (C>T transitions) created by formalin fixation, which can be diagnostically challenging when left uncontrolled. In this regard, DNA amplification during panel sequencing
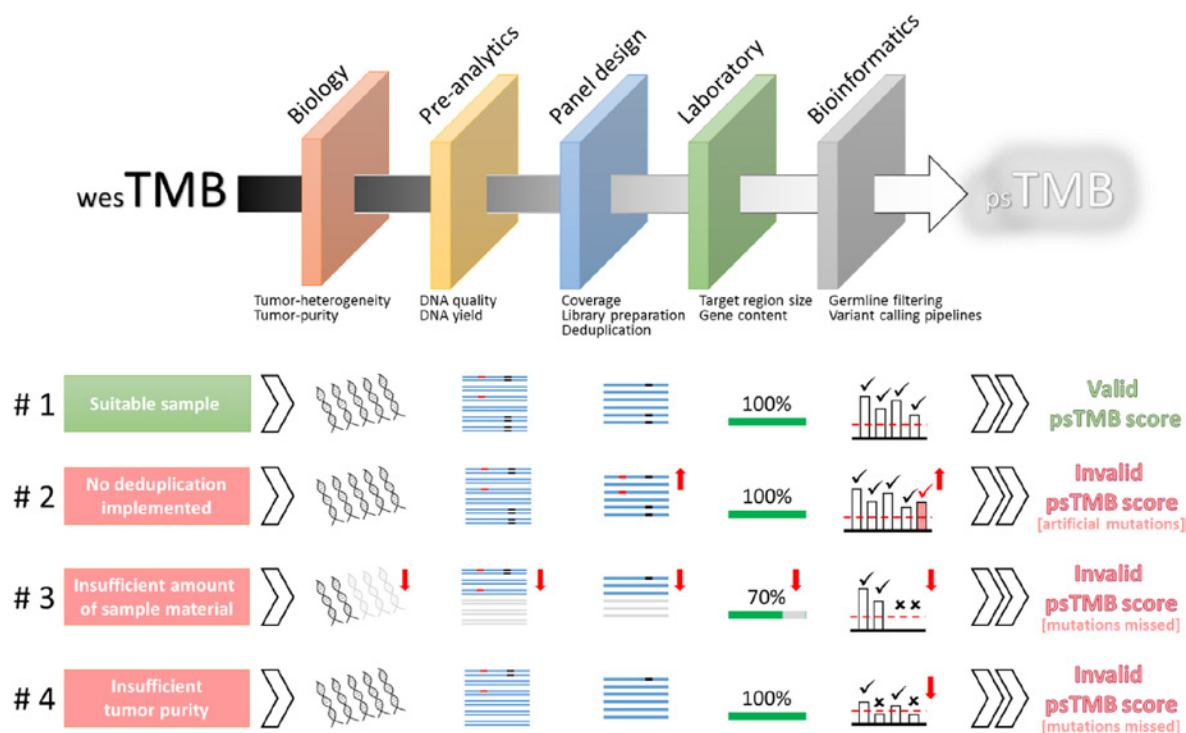
Figure 6. Schematic representation of assay-independent and assay-specific parameters influencing the accuracy of psTMB scores. Lower lane: four representative samples revealing the effect of deduplication strategies (#2), insufficient sample material (#3) or low tumor purity (#4) on DNA input, coverage, covered coding sequence, variant calling, and the resulting psTMB scores. Red arrow pointing down indicates false-negative effect. Red arrow pointing up indicates false-positive effect. psTMB, TMB assessed by gene panel sequencing; wesTMB, TMB assessed by WES.

can be critical as overamplification of artifacts or additional errors during replication can occur, leading to false-positive mutation calls and subsequently to overestimated psTMB scores. This issue can be compensated by setting an appropriate limit of detection (LOD) for the allelic frequency and especially by applying in silico or technical (molecular barcodes) approaches, or both, for deduplication (removal of PCR duplicates). In the present data set, a LOD of 5% in combination with deduplication yielded reliable mutation calling, and eventually TMB values. Hybrid-capture–based target enrichment was favorable in this context. In panels without deduplication, deamination artifacts may be controlled by increasing the LOD to, for example, 10%, rendering cases with low tumor cellularity challenging owing to the impaired sensitivity. Recent reports indicate that the application of uracil-DNA glycosylase, an enzyme selectively digesting uracil-containing nucleic acid, can reduce deamination artifacts, when assessing TMB in FFPE samples using assays without a deduplication approach.[29-31] However, the effect of this approach was not tested in this study.

We also calculated BFs to convert psTMB to wesTMB for the assessed panels. Although future studies exploring larger sample sets will likely improve this analysis, we believe that the data revealed here provide a strong and sound basis that will facilitate the comparison of TMB values obtained by different panels.

A limitation of our study is the limited number of cases and the use of three different cancer types. The latter selection was influenced by (1) a case mix that reflects the continuum of TMB, (2) avoiding result bias owing to a single cancer type, (3) tissue availability for the entire study and all partners, and (4) availability of corresponding WES data. Because the predictive power of TMB is currently being tested in many immuno-oncology trials across various cancer types, and as our study is primarily aimed at investigating the ability of panels to measure TMB, we believe that these points do not interfere with our results and conclusions.

In summary, the QuIP study provides real-world evidence that all panels tested in this study can be used to estimate TMB by panel sequencing from FFPE samples in a routine diagnostic setting. However, this study has identified several critical parameters, including sample fixation, DNA input, sequencing depth, genome coverage, and VAF cut points, that may confound psTMB estimation and require careful control to achieve successful and reliable psTMB analysis. Beyond TMB, in conjunction with efforts by the Friends of Cancer Research, this study provides a blueprint and framework for the systematic analysis of complex or composite predictive biomarkers, which will likely play an increasing role in guiding oncological therapy.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at https://doi.org/10.1016/j.jtho.2020.01.023.

References

1. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19:133–150.
2. Jenkins RW, Barbie DA, Flaherty KT. Mechanisms of resistance to immune checkpoint inhibitors. *Br J Cancer*. 2018;118:9-16.
3. Brahmer JR, Lacchetti C, Schneider BJ, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2018;36:1714-1768.
4. Buttner R, Gosney JR, Skov BG, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. *J Clin Oncol*. 2017;35:3867–3876.
5. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;372:2509-2520.

6. Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171:1042-1056.e10.

7. Carbone DP, Reck M, Paz-Ares L, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med*. 2017;376:2415-2426.

8. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124-128.

9. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014;371:2189-2199.

10. Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350:207-211.

11. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9:34.

12. Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. 2018;24:1441-1448.

13. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23:703-713.

14. Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB harmonization project. *J Immunother Cancer*. 2020. In press.

15. Merino DM, McShane L, Butler M, et al. TMB standardization by alignment to reference standards: phase II of the Friends of Cancer Research TMB Harmonization Project [abstract]. *J Clin Oncol*. 2019;37(suppl 15):2624.

16. R Core Team. R: a language and environment for statistical computing. http://www.r-project.org. 2018. Accessed March 11, 2020.

17. Budczies J, Allgauer M, Litchfield K, et al. Optimizing panel-based tumor mutational burden (TMB) measurement. *Ann Oncol*. 2019;30:1496-1506.

18. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378:2093-2104.

19. Hellmann MD, Paz-Ares L, Bernabe Caro R, et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med*. 2019;381:2020-2031.

20. Ready N, Hellmann MD, Awad MM, et al. First-line nivolumab plus ipilimumab in advanced non-small-cell lung cancer (CheckMate 568): outcomes by programmed death ligand 1 and tumor mutational burden as biomarkers. *J Clin Oncol*. 2019;37:922-1000.

21. Sun JX, He Y, Sanford E, et al. A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal. *PLoS Comput Biol*. 2018;14, e1005965.

22. Buchhalter I, Rempel E, Endris V, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int J Cancer*. 2019;144:848-858.

23. Ascierto PA, Bifulco C, Palmieri G, Peters S, Sidiropoulos N. Preanalytic variables and tissue stewardship for reliable next-generation sequencing (NGS) clinical analysis. *J Mol Diagn*. 2019;21:756-767.

24. Buttner R, Longshore JW, López-Rios F, et al. Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open*. 2019;4: e000442.

25. Kazdal D, Endris V, Allgäuer M, et al. Spatial and temporal heterogeneity of panel-based tumor mutational burden in pulmonary adenocarcinoma: separating biology from technical artifacts. *J Thorac Oncol*. 2019;14:1935-1947.

26. Stenzinger A, Allen JD, Maas J, et al. Tumor mutational burden standardization initiatives: recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions. *Genes Chromosomes Cancer*. 2019;58:578-588.

27. Kerr KM, Tsao MS, Nicholson AG, et al. Programmed death-ligand 1 immunohistochemistry in lung cancer: in what state is this art? *J Thorac Oncol*. 2015;10:985-989.

28. Chang H, Sasson A, Srinivasan S, et al. Bioinformatic methods and bridging of assay results for reliable tumor mutational burden assessment in non-small cell lung cancer. *Mol Diagn Ther*. 2019;23:507-520.

29. Alborelli I, Leonards K, Rothschild SI, et al. Tumor mutational burden assessed by targeted NGS predicts clinical benefit from immune checkpoint inhibitors in non-small cell lung cancer. *J Pathol*. 2020;250:19-29.

30. Do H, Molania R, Mitchell PL, Vaiskunaite R, Murdoch JD, Dobrovic A. Reducing artifactual *EGFR* T790M mutations in DNA from formalin-fixed paraffin-embedded tissue by use of thymine-DNA glycosylase. *Clin Chem*. 2017;63: 1506-1514.

31. Heeke S, Benzaquen J, Long-Mira E, et al. In-house implementation of tumor mutational burden testing to predict durable clinical benefit in non-small cell lung cancer and melanoma patients. *Cancers (Basel)*. 2019;11:1271.

European School of Oncology – Review

International liquid biopsy standardization alliance white paper

Dana Connors [a], Jeff Allen [b], J.D. Alvarez [c], Jennifer Boyle [d], Massimo Cristofanilli [e], Carolyn Hiller [c], Susan Keating [f], Gary Kelloff [g], Lauren Leiman [h], Robert McCormack [i], Diana Merino [b], Emily Morgan [a], Klaus Pantel [j], Christian Rolfo [k,*], Maria Jose Serrano [l], A. Pia Sanzone [d], Thomas Schlange [m], Caroline Sigman [f], Mark Stewart [b]

[a] Foundation for the National Institutes of Health, United States
[b] Friends of Cancer Research, United States
[c] Medical Device Innovation Consortium, United States
[d] National Institute for Biological Standards and Control (NIBSC), UK
[e] Northwestern Medicine, International Society of Liquid Biopsy (ISLB), United States
[f] CCS Associates, Inc., United States
[g] Biomarkers Consortium Cancer Steering Committee, United States
[h] BloodPAC, United States
[i] FNIH ctDNA QCM Project, United States
[j] European Liquid Biopsy Society, Germany
[k] University of Maryland Medical Center, International Society of Liquid Biopsy (ISLB), United States
[l] Centro GENyO, International Society of Liquid Biopsy (ISLB), United States
[m] Cancer-ID, Czech Republic

ARTICLE INFO

ABSTRACT

The promise of precision medicine as a model to customize health care to the individual patient is heavily dependent upon new genetic tools to classify and characterize diseases and their hosts. Liquid biopsies serve as a safe alternative to solid biopsies and are thus a useful and critical component to fully realizing personalized medicine. The International Liquid Biopsy Standardization Alliance (ILSA) comprises organizations and foundations that recognize the importance of working towards the global use of liquid biopsy in oncology practice to support clinical decision making and regulatory considerations and seek to promote it in their communities. This manuscript provides an overview of the independent liquid biopsy- and standardization-based programs engaged with ILSA, their objectives and progress to date, and the tools and resources each is developing to contribute to the field. It also describes the unique areas of effort as well as synergy found within the group.

1. Introduction

The promise of precision medicine as a model to customize health care to the individual patient is heavily dependent upon new genetic tools to classify and characterize diseases and their hosts. To accomplish these goals with traditional tumor sampling, invasive procedures to obtain genetic material would necessarily increase to provide enough material to accurately capture and describe genomic variations and their phenotypes. Liquid biopsies serve as a safe alternative to solid biopsies (Ma et al., 2020) and are thus a useful and critical component to fully realizing personalized medicine (Rolfo et al., 2020; Heitzer et al., 2019; Keller and Pantel, 2019; Pantel et al., 2019). Liquid biopsy may include tumor-derived nucleic acid such as circulating tumor DNA (ctDNA), circulating tumor RNA (ctRNA), circulating tumor microRNA (ctmiRNA), RNA or DNA from exosomes, and circulating tumor cells (CTC) collected from peripheral blood, which can be obtained non-invasively through a simple venipuncture (Anfossi et al., 2018; Kalluri et al., 2020). This material is now being used to identify actionable mutations for targeted therapy and through their less invasive nature can be utilized to monitor therapy response serially and screen for early detection of disease (Keller and Pantel, 2019; Russo et al., 2019).

Clinical oncology is being transformed with the adoption of next-generation sequencing (NGS-) based diagnostics. This new technology can enable the rapid identification of potentially significant genetic variations across the human genome, and the results are increasingly being used to determine the best course of treatment for oncology patients (MacConaill, 2013). Ensuring that these patients receive accurate results is imperative given that a false negative or false positive result could cause a patient to be diverted from a more beneficial therapeutic option, assigned to the wrong arm of a clinical trial, or unnecessarily subjected to adverse drug effects. However, lack of agreed-upon, well-characterized and community-validated reference samples and data benchmarks creates a potential challenge for the efficient development of these critical tests and for understanding their results (Normanno et al., 2013; Lampignano et al., 2020).

The **International Liquid Biopsy Standardization Alliance (ILSA)** comprises organizations and foundations that recognize the importance of working towards the global use of liquid biopsy in oncology practice to support clinical decision making and regulatory considerations and seek to promote it in their communities. Multiple efforts have joined together under the auspices of the ILSA group to share the scope of their work, discuss lessons learned, and disseminate the tools and data they have developed as a coordinated effort. The group recognizes that a preponderance of technologies and pursuits in the field of liquid biopsy has the potential to confound the field and obscure important progress. As an alliance, the ILSA partners have found strong value in the exercise of information exchange through in-person meetings and teleconferences, as well as the dissemination of their collective efforts. This manuscript is a further step to inform the scientific community about the formation of the collaborative group and the availability of resources (Fig. 1).

This manuscript provides an overview of the independent liquid biopsy- and standardization-based programs engaged with ILSA, their objectives and progress to date, and the tools and resources each is developing to contribute to the field. It also describes the unique areas of effort as well as synergy found within the group. Collaborative opportunities, tools, and resources the group desires to make available are also shared for interested researchers and clinicians.

2. Overview of liquid biopsy efforts

The **BloodPAC** (Anon, 2020a) mandate is to accelerate the development and validation of liquid biopsy assays to improve the outcomes of patients with cancer. To do this the group developed a collaborative infrastructure that enables sharing of information between stakeholders in the public, industry, academia, and regulatory sectors. There are currently six active BloodPAC Working Groups. The Data Experience Working Group will determine how data are aggregated, reviewed, and processed. The Pre-analytical Variables Working Group identified and created 11 Minimal Technical Data Elements (MTDEs) from the critical variables required for new studies submitted into the Data Commons. The Analytical Variables Working Group focuses on developing analytical protocols, while the Patient Context Variables Working Group identified 5 critical required MTDEs and an additional 11 suggested MTDEs. The Reimbursement Working Group will discuss their developing project plan to address reimbursement for liquid biopsy assays at the annual meeting of American Society of Clinical Oncology (ASCO). The JFDI Working Group, which leads the effort, has increased to 10 members aligned to develop a framework for evidence generation to bring liquid biopsy into routine clinical practice.

The BloodPAC consortium recognizes data sharing and evidence generation as the two fundamental requirements for success and is pursuing them through two dedicated workstreams:

- To create a BloodPAC Data Commons to serve the community at large for all liquid biopsy stakeholders; and to
- Align around a framework for evidence generation to bring liquid biopsy into routine clinical practice.

The University Medical Center Hamburg-Eppendorf established a novel EU consortium called the **European Liquid Biopsy Society (ELBS)** (Anon, 2020b) which aims to become the leading hub for liquid biopsy research in Europe with the key goal to translate liquid biopsy assays into clinical practice for the benefit of patients. 40 European Institutions from academia and private industry attended the kickoff meeting in 2019 in Hamburg, and the number of candidate institutions has increased to 97 within one year, demonstrating the enormous interest in the ELBS consortium, which also welcomes non-EU members. ELBS will replace the public-private partnership CANCER-ID (Anon, 2020c), a five-year Innovative Medicines Initiative (IMI) (Anon, 2020d) consortium, which came to an end in 2019. In the course of its activities, CANCER-ID published best-practice protocols and the results of ring studies based on the implementation of harmonized protocols and standard materials (Kalluri et al., 2020; Kloten et al., 2019). The results are the basis for implementation of liquid biopsy protocols in ongoing clinical studies which are evaluating the predictive value and clinical utility of CTC and ctDNA assays in patients treated with immune

Fig. 1. Member organizations of ILSA perform synergistic functions from pre-analytical variables analysis to clinical utility activities to bring the promise of globally standardized and evaluable precision medicine to clinical practice.

checkpoint inhibitors (Hofman et al., 2019).

ELBS is an institutional network (not individual membership organizations) with the following goals, which build upon and extend the CANCER-ID objectives:

- Foster the introduction of liquid biopsy into clinical trials and practice;
- Support liquid biopsy research (CTCs, ctDNA, cell-free microRNA (cfMiRNA), extracellular vesicles (EVs), platelets, proteins, etc.);
- Provide a partner for regulatory agencies, health care providers and patient advocacy groups;
- Encourage interactions between academia and industry;
- Develop guidelines and provide training in liquid biopsy for medical scientists;
- Disseminate knowledge about liquid biopsy to the medical community through regular symposia, publications, and press releases;
- Increase European visibility as a leading hub for liquid biopsy research;
- Outreach to non-EU networks of liquid biopsy research (in the US, Asia, and Australia).

The main aim of the **International Society of Liquid Biopsy (ISLB)** (Anon, 2020e) is to introduce recommendations to develop reliable and sustainable diagnostics and prognostics tools using liquid biopsies that will benefit patient health management and wellness. Founded in 2017, the ISLB was created considering that: technologies evolve day by day, and the concept of liquid biopsy requires continuous attention and research to provide patient and community benefit; there is a growing worldwide effort under way that combines the knowledge and motivation of clinicians, biotech companies, and the pharmaceutical industry, all considered essential in the development of this field; there is an urgent need to establish criteria and provide guidelines for the design, development, and validation studies necessary before clinical application can be made with confidence; an opportunity exists today to coordinate efforts and strategies between key players through communication and collaboration to establish research priorities and avoid overlap of efforts that simply delay the implementation of these technologies; and there is a fundamental need to support collaboration with all professionals and colleagues around the world to ensure the ability to translate the benefits of liquid biopsy to all communities irrespective of social and economic status, providing them access to both research and clinical application of these new technologies.

Based on these considerations the ISLB was created to provide a forum for the exchange of ideas and to represent the efforts of stakeholders and professionals interested and active in this field. The group maintains and supports regular multisectoral meetings to better understand global advances and existing problems of implementation and to respond to the needs of patients, health systems, and laboratory services. The group also works to assess and advise on existing technologies and their suitability for clinical practice and implementation, improvements that can be made, and the practical benefits that will accrue, and promote and facilitate initial and ongoing training of experts and professionals interested in the development of liquid biopsies. ISLB collaborates with other professional societies, associations, and groups with similar objectives to support researchers, clinicians, and laboratories across the world, including developing countries where economic difficulties may limit early access to these technologies.

To enable the ISLB to act as an ethical conversation point for stakeholders in the emerging liquid biopsy field, the group ensures communications regarding progress are made on a global basis and in a timely manner, and encourage where possible the standardization of procedures when translating research into clinical diagnostic protocols. The ISLB places no limits on those who wish to join and has created membership categories for individuals, institutions, and corporations. Undergraduate and post-graduate medical and biology and genetics students are encouraged to join or participate. The group has created a

series of supervisory committees and plans to develop key regional supporting committees over time.

The ISLB encourages conversations about liquid biopsies to help advance patient care by creating and delivering world class workshops, symposia and conferences in Europe, North America, Asia and South America. In the future, it is expected ISLB will create members' only sections of the ISLB website, as well as access and contact points for non-members, such as the public, media, and health care providers, as well as those with a specialist interest in liquid biopsies. The ISLB hopes to stimulate an active and continuous flow of interest, ideas, and information between members, as well as interested non-members, and to act as an important reference site for health care professionals and governments.

The **Foundation for the National Institutes of Health (FNIH) Biomarkers Consortium (BC)** is a public-private biomedical research partnership that endeavors to discover, develop and seek regulatory approval for biomarkers to support new drug development, preventative medicine and medical diagnostics. The organization works to combine the forces of the public and private sectors in working groups and project teams to accelerate the development of biomarker-based technologies, medicines and therapies for prevention, early detection, diagnosis and treatment of disease. The **ctDNA Quality Control Materials Project** (Anon, 2020f) is an approved collaboration of the BC which gathered multiple private sector, government agency, academic, and not-for-profit partners together to characterize Quality Control Materials (QCMs) and demonstrate their comparable performance to ctDNA and their suitability to establish performance characteristics for NGS assays already in use.

An initial liquid biopsy working group of the BC Cancer Steering Committee, which included members from ASCO, the Association for Molecular Pathology, the College of American Pathologists, the National Institute of Standards and Technology, and regulatory representatives among multiple other stakeholders, identified 14 common and clinically relevant, and actionable variants across the four variant classes to establish a workflow. The project team has conducted a phase 1 performance evaluation of reference materials from three commercial manufacturers, which include these variants. The team will soon launch a phase 2 contrived sample functional characterization (commutability) study in coordination with input from the U.S. Food and Drug Administration (FDA) and the other project stakeholders to determine if the QCMs perform similarly to clinical specimens. During phase 3 a clinical pilot will engage multiple external laboratories across four continents to further evaluate the QCMs in their daily runs and alongside clinical samples for additive performance information.

The manufacturers of the reference materials will be provided final data from these assays that may be used in submission for regulatory approval of standardized QCMs that can be used in laboratories across multiple assay types to accurately identify variant allele fractions of the original 14 clinically relevant variants. Together, the project team hopes to define a cost-effective, practical approach for the validation of ctDNA QCMs that includes a demonstration of commutability and which could serve as a "roadmap" for other groups in the future.

Friends of Cancer Research (Friends) (Anon, 2020g) is an advocacy organization that drives collaboration among partners from every healthcare sector to power advances in science, policy, and regulation that speed life-saving treatments to patients. Friends has been instrumental in the creation and implementation of policies ensuring patients receive the best treatments in the fastest and safest way possible. **ctDNA to MONItor Treatment Response (ctMoniTR) Project** (Anon, 2020h) is a pilot project involving several key pharmaceutical stakeholders with the objective to harmonize the use of ctDNA to evaluate or monitor patient response, and to better answer the pressing question: Do changes in ctDNA levels accurately reflect the therapeutic effect of cancer therapies? The hypothesis under which this project operates is that broad changes in ctDNA levels can detect tumor response to cancer therapies, including immune checkpoint inhibitors, targeted therapies and

chemotherapy.

The project has a multi-step approach including Step 1: the study of ctDNA as a monitoring tool in previously collected trial data from a subset of trials (e.g. NSCLC treated with ICI), and Step 2: the study of ctDNA as a monitoring tool in prospectively collected trial data from various types of advanced cancer and therapies (e.g. immune checkpoint inhibitors (ICIs), tyrosine kinase inhibitors, chemotherapy). Step 1 will assess the feasibility of investigating the directionality of change in ctDNA levels from baseline and investigate its association with patient response by looking at a smaller cohort of studies with a specific indication. Results from Step 1 will inform Step 2, which will investigate the ability of ctDNA to detect early therapeutic response in clinical trials for different indications and different treatments that share a uniform plasma and data collection methodology.

The ctMoniTR project will seek to promote a harmonized advancement of the field of liquid biopsies and to facilitate evidence development that will generate the necessary foundation for regulatory evaluation of ctDNA as a monitoring tool and early predictor of treatment response. As such, it will propose several key methodological and knowledge questions and define what type of data will be needed to respond to these relevant questions. Moreover, the project will seek to align plasma collection methodologies, and suggests uniform approaches for reporting relative directional changes in ctDNA using NGS panels.

The **Japanese bio-Measurement and Analysis Consortium (JMAC)** (Anon, 2020i) is an industrial group established in 2007 to support biotechnology with international standardization activities. The International Organization for Standardization (ISO) activity of JMAC is continuously expanding to cover clinical laboratory (TC 212), food (TC 34), biotechnology (TC 276), and nanotechnology (TC 229). In parallel to the ISO activities, JMAC participates in several research project including miRNA cancer biomarker discovery project and a central nervous system biomarker discovery project pursuing liquid biopsy technologies. JMAC drafted the standards document "Molecular biomarker analysis - general definitions and requirements for microarray detection of specific nucleic acid sequences." Member companies of JMAC benefit from the latest updated information in the biotech industry through these activities and the group is open to new members.

The **Medical Device Innovation Consortium (MDIC)** (Anon, 2020j) is a public-private partnership created in 2012 to advance medical device regulatory science for patient benefit. MDIC brings together representatives of the government, industry, not-for-profit and patient organizations to improve processes for the development, assessment and review of new medical technologies.

MDIC aims to identify and pursue projects that will improve diagnostic testing and product development using novel regulatory science approaches developed through collaboration among MDIC stakeholders. Providing a predictable path for innovation will help patients benefit through quicker access to more cost-effective advanced diagnostic technologies in less time. One focus of this work is the establishment of a public-private partnership to guide the development of reference samples that can be used to develop and validate NGS-based oncologic tests.

Currently, many test developers, including commercial manufacturers and clinical laboratories, are developing their own contrived samples and sample mixes for validation of oncology tests since well-characterized and agreed-upon oncology samples/reference materials do not exist. This makes it difficult to efficiently develop or compare tests and methodologies. Reference samples that can be used to more efficiently develop and assess the various components of an NGS test are needed to ensure confidence in the results being provided by different NGS clinical tests.

The objective of the MDIC's Somatic Reference Samples (SRS) initiative is to guide the development of reference samples that can be used to develop and validate NGS-based oncologic tests, with the focus on solid tumors. These samples are to be properly consented, widely shareable reference samples to be made available to the public and scalably produced in order to enable efficient development and improve the accuracy, reliability and transparency of tissue-based oncology tests. These samples will be quality checked and validated, made available in varying forms (e.g., cells, DNA/RNA, formalin-fixed paraffin-embedded tissue [FFPE]), represent the majority of potential variations and allele fractions of interest (e.g., ploidy, fusions, large/small indels, copy number variations (CNVs), homopolymeric regions), and represent tumor/normal matched pairs. The output of the effort will be shared broadly with the community, with the intent to have any reference samples developed through this effort scaled for commercial distribution.

A cross-functional working group is developing processes for identifying and acquiring appropriately consented tumor/normal matched samples containing variants of interest, consistent production of materials at scale to facilitate public availability in various formats, sequencing data integration and consensus call determination to develop high-confidence truth sets, and long-term maintenance of reference samples and accessibility of data sets. The ultimate goal is to pilot the production of several high priority NGS reference samples for the oncology community. When the project outputs are validated and characterized, they will be made widely available to allow for scale up, production, and broad public accessibility by interested commercial entities.

The **National Institute for Biological Standards and Control (NIBSC, UK)** (Anon, 2020k) is a World Health Organization (WHO) International Laboratory for Biological Standards and is the world's primary producer of WHO International Standards. These highest order reference materials, which are typically prepared as a single homogeneous batch of several thousand ampoules intended to last many years, serve to harmonize the measurement of biological activity in internationally agreed units through traceability to a single common standard; they are not intended for routine use, but rather as calibrators of assays and secondary standards, which in turn may be used as assay run controls. Their use enables comparability between laboratories and methods, towards improved public health via accurate and sensitive measurement in ensuring the quality of biological medicines, diagnostic testing, and therapeutic response in patients worldwide. All NIBSC-produced WHO standards are provided on a non-for-profit basis to facilitate global availability.

NIBSC has endorsement from the WHO Expert Committee on Biological Standardization to generate the WHO 1st International Standards for ctDNA. The intention is to first produce ctDNA standards for the most frequent clinically-associated variants in EGFR, including p.L858R, p. T790M, and exon 19 deletions, with the aim to subsequently address other solid tumor-associated gene variants as their utility in liquid biopsy analyses increases (Rolfo et al., 2018). These standards should ideally capture and allow harmonization of the multiple variables associated with ctDNA measurement, including variant percentage, DNA fragment size(s), ctDNA yield, and gene copy numbers. They should also demonstrate excellent stability since International Standards are typically prepared as a single homogeneous batch of several thousand ampoules intended to last many years.

Commutability, i.e. comparability to real clinical samples, is also a critical component of these standards, since they should perform equally well in laboratory assays and are appropriate for in vitro diagnostic use worldwide. NIBSC is currently assessing the performance of several matrices with various cell-line derived fragmented DNAs to determine the optimal format for the standards, while cross-referencing to patient ctDNA materials. Co-incident with this program, NIBSC is also developing genomic DNA standards for EFGR variants from the same source cell lines; this is intended to facilitate the alignment of both liquid and solid tumor biopsy and aid the transition to a non-invasive diagnostic approach. The establishment of these standards must coincide with standardized protocols for sample collection and preparation, testing, analysis, and reporting, towards a fully harmonized liquid biopsy approach.

COMPLEX BIOMARKERS: INFORMING STANDARDIZATION AND HARMONIZATION OF DIAGNOSTIC TESTING

3. Synergy and unique areas of development

The ILSA partners described above came together due to a similar objective: to bring about the ubiquitous and meritorious use of liquid biopsy in oncology practice to support clinical decision making and regulatory considerations and to obviate the need for invasive solid tumor biopsies. Given the overlapping interests of the members, at its first meeting participants were surprised to find that the efforts are largely non-duplicative. In fact, the efforts noted here are synergistic and cover the spectrum of need to bring liquid biopsy into routine clinical application.

As outlined above, the BloodPAC effort and Cancer-ID/ELBS Consortium are working from U.S. and European perspectives to develop pre-analytical variables considering laboratory handling and patient contact. Both consortia have published reviews outlining developments in liquid biopsy (Pantel et al., 2019c; Alix-Panabieres, 2020). They have also considered downstream use of the materials to establish naming convention and appropriate standards for collection and data coordination.

The FNIH Biomarkers Consortium ctDNA QCM and MDIC SRS projects dovetail with these pre-analytical development steps to provide examples, roadmaps, and initial efforts at reference material development. These groups have noted that they seek to provide U.S. laboratories with reference materials for an initial set of variants across the variant classes and multiple diseases and clinically relevant applications, while leaving room for further material development and improvement upon the processes they establish.

The ELBS and Friends of Cancer Research ctMoniTR Project are taking the next steps to develop ring studies in Europe and clinical trials in the United States, respectively, that will utilize established standardized methods to devise clinical collection standards for downstream implementation and to make the case for clinical utility in practice. ELBS and ILBS participants have collaborated to publish an international expert consensus paper on the clinical use of CTCs in breast cancer (Cristofanilli et al., 2019). The ILBS and MDIC's SRS initiative will build upon these early steps to support the use of liquid biopsy in research and clinical trials in countries around the world, while JMAC and NIBSC provide the infrastructure needed for laboratories to collaborate and build complementary data sets.

Among the multiple stakeholders participating in each of the listed projects are academic partners, private sector organizations and industry, government agencies and patient advocacy. FDA and the European Medicines Agency (EMA) regulatory representatives have also been supportive of the efforts and are interested to provide their support to discuss and develop liquid biopsies for use in clinical research and drug development.

4. Collaborative opportunities

A consistent theme from the collaborative discussions of the ILSA group is the continued communication of ongoing efforts. To that end, ILSA has committed to building a central repository for information sharing hosted by the BloodPAC Data Commons. The platform will share links to each organization to help researchers and clinicians recognize existing collaboration and avoid duplicative work.

Future development of the repository is planned to host protocols, white papers that highlight best practices, relevant policies for different countries and regions, and eventually de-identified data sets from studies that can be utilized in continued research.

The groups are interested in sharing the outputs of their projects as well. The FNIH Biomarkers Consortium ctDNA QCM project team will work with reference material manufacturers to qualify their materials with the FDA and eventually disseminate them for real-world-use ctDNA testing in the field. At the same time, the SRS initiative seeks to guide the development of reference samples used to validate NGS-based oncologic tests with an initial focus on solid tumors. Both groups are seeking to

produce properly consented, widely shareable reference samples to be made available to the public and scalably produced in order to enable efficient development and improve the accuracy, reliability and transparency of blood- and tissue-based oncology tests. Developed materials from these projects will represent the majority of potential variations and allele fractions of interest (e.g., ploidy, fusions, indels, CNVs), and will be shared broadly with the community. The teams within ILSA have the opportunity to share beset practices to ensure that reference samples developed through their efforts can be scaled for commercial distribution.

In developing the WHO 1st International Standards for ctDNA, NIBSC must likewise ensure materials are well suited for their intended purpose in harmonizing ctDNA measurement across laboratories, assays, and secondary standards. It is thus imperative that the groups seek input from liquid biopsy community members worldwide. ILSA serves as a connection point and open market for commercial vendors, clinical centers, diagnostic laboratories, and other consortia, and welcomes further engagement. Given that participants benefit from and seek global agreement for the assignment of values and standards, they recognize the need for continued education and collaboration towards global alignment of measurement.

5. Resources for the field

As a first step to support the larger field, the ILSA partners have proactively linked partner efforts on their websites to promote and disseminate their work (Anon, 2020l). Each effort has also provided descriptive information to include on a central repository including contact information, organizational descriptions, website listings, and links to additional references and collaborative groups (Anon, 2020m). The ILSA partners are identified in a matrix form to provide interested parties with easily accessible and relevant information.

As a second step, the ILSA group is working to translate the current alliance into a recognized Collaborative Community in which the FDA can participate and provide supportive regulatory perspective (Anon, 2020n). The EMA is also active with the group and sees it as a forum in which their unique regulatory perspective can be shared as well, facilitating future coordinated efforts in this field to more precisely refine and describe pathways for qualification of standards and assays. The group provides a forum to ensure standards organizations are frequently engaged with to discuss regulatory pathways and are able to share insights and recommendations with their own networks.

In the future, the group aspires to build a database with various tiers of information that can be accessed by other efforts wishing to partner with one or multiple of the members of ILSA. First tier of information will provide simple descriptions of each effort and contact information. Second tier of information will include publications, white papers, and standard operating procedures from the various groups, diving into detail on their specific efforts. This information will serve both as an educational resource for the liquid biopsy community, as well as a way for other efforts to replicate different aspects of the liquid biopsy development pipeline for which the members of ILSA have developed lessons learned and best practices. Finally, a third tier of information the ILSA group envisions is a data resource with the original data from the experiments being conducted by each endeavor. This resource could then be used to further refine clinical use of liquid biopsy, standards development, and best practices.

6. Call to action

The ILSA group would like to invite additional members of the liquid biopsy field to join their efforts to synergize the science being conducted in this space. ILSA members wish to create a comprehensive repository of resources, lessons learned, and best practices in the liquid biopsy space in order to harmonize the disparate efforts that continue to hinder progress toward liquid biopsy achieving the status of full "clinical

utility" in the many contexts of use in which it is being used (Merker et al., 2018).

Eventually, ILSA would like to develop a uniform end-to-end process identified from pre-analytic capture through technology/assay validation to clinical collection and analysis. The group would then hope to promote the use of this process to allow for standardized data capture for liquid biopsy to build the necessary level of evidence required to have multiple contexts of use qualified by the FDA for clinical implementation.

Declaration of Competing Interest

Dr. Rolfo reports speaker bureau for MSD and AstraZeneca; advisory board role for ARCHER, Inivata, Merck Serono; consultant role for Mylan and Oncopass; supported research grant from Lung Cancer Research Foundation-Pfizer; research support from Guardant Health and Biomark inc. Dr. Cristofanilli reports consultant role for Foundation Medicine, Lilly, Pfizer, Cytodyn. Dr Serrano reports consultant role for Astellas. The other authors do not report any conflict of interest.

References

Alix-Panabieres, 2020. The future of liquid biopsy. Nature 579, S9. https://doi.org/10.1038/d41586-020-00844-5.

Anfossi, S., et al., 2018. Clinical utility of circulating non-coding RNAs – an update. Nat. Rev. Clin. Oncol. 15 (September (9)), 541–563. https://doi.org/10.1038/s41571-018-0035-x.

Anon, https://www.bloodpac.org/.

Anon, www.elbs.eu.

Anon, www.cancer-id.eu.

Anon, https://www.imi.europa.eu.

Anon, https://www.isliquidbiopsy.org/.

Anon, https://fnih.org/what-we-do/biomarkers-consortium/programs/ctdna-reference-materials.

Anon, https://www.focr.org.

Anon, www.focr.org/ctDNA.

Anon, https://www.jmac.or.jp/en/.

Anon, https://mdic.org/.

Anon, https://www.nibsc.org/science_and_research/advanced_therapies/genomic_reference_materials.aspx.

Anon, https://fnih.org/what-we-do/biomarkers-consortium.

Anon, https://fnih.org/what-we-do/programs.

Anon, https://www.fda.gov/about-fda/cdrh-strategic-priorities-and-updates/collaborative-communities-addressing-health-care-challenges-together.

Cristofanilli, Massimo, et al., 2019. The clinical use of circulating tumor cells (CTCs) enumeration for staging of metastatic breast cancer (MBC): international expert consensus paper. Crit. Rev. Oncol. Hematol. 134 (February), 39–45. https://doi.org/10.1016/j.critrevonc.2018.12.004. Epub 2018 Dec 19.

Heitzer, E., et al., 2019. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat. Rev. Genet. 20 (2), 71–88. https://doi.org/10.1038/s41576-018-0071-5.

Hofman, P., et al., 2019. Liquid Biopsy in the Era of Immuno-Oncology: Is It Ready for Prime-Time Use for Cancer Patients? Ann. Oncol. 30 (September (9)), 1448–1459. https://doi.org/10.1093/annonc/mdz196.

Kalluri, R., et al., 2020. The biology, function, and biomedical applications of exosomes. Science 367 (6478), eaau6977. https://doi.org/10.1126/science.aau6977, 07 Feb.

Keller, L., Pantel, K., 2019. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. Oct. Nat Rev Cancer. 19 (10), 553–567. https://doi.org/10.1038/s41568-019-0180-2.

Kloten, Vera, et al., 2019. Multicenter evaluation of circulating plasma MicroRNA extraction technologies for the development of clinically feasible reverse transcription quantitative PCR and next-generation sequencing analytical work flows. Clin. Chem. 65 (9), 1132–1140. https://doi.org/10.1373/clinchem.2019.303271, 1 September.

Lampignano, R., et al., 2020. Multicenter evaluation of circulating cell-free DNA: extraction and downstream analyses for the development of standardized (Pre) analytical work. Clin. Chem. 66, 1149–1160.

Ma, Ning, et al., 2020. Deciphering Cancer clues from blood. Science 367 (6485), 1424–1425. https://doi.org/10.1126/science.abb0736, 27 March.

MacConaill, L.E., 2013. Existing and emerging technologies for tumor genomic profiling. J. Clin. Oncol. 31 (15), 1815–1824. https://doi.org/10.1200/JCO.2012.46.5948.

Merker, Jason D., et al., 2018. Circulating tumor DNA analysis in patients with cancer. Am. Soc. Clin. Oncol. Coll. Am. Pathol. Joint Rev. 36 (June (16)), 1631–1641. https://doi.org/10.1200/JCO.2017.76.8671. Epub 2018 Mar 5.

Normanno, N., et al., 2013. Molecular diagnostics and personalized medicine in oncology: challenges and opportunities. J. Cell. Biochem. 114 (3), 514–524. https://doi.org/10.1002/jcb.24401.

Pantel, K., et al., 2019. Liquid biopsy and minimal residual disease - latest advances and implications for cure. Nat. Rev. Clin. Oncol. 16 (July (7)), 409–424. https://doi.org/10.1038/s41571-019-0187-3.

Pantel, Klaus, et al., 2019c. Circulating tumor cells in prostate cancer: from discovery to clinical utility. Clin. Chem. 65 (January (1)), 87–99. https://doi.org/10.1373/clinchem.2018.287102.

Rolfo, C., et al., 2018. Liquid biopsy for advanced non-small cell lung cancer (NSCLC): a statement paper from the IASLC. J. Thorac. Oncol. 13 (9), 1248–1268. https://doi.org/10.1016/j.jtho.2018.05.030.

Rolfo, C., et al., 2020. Challenges and opportunities of cfDNA analysis implementation in clinical practice: perspective of the International Society of Liquid Biopsy (ISLB). Crit. Rev. Oncol. Hematol., 102978 https://doi.org/10.1016/j.critrevonc.2020.102978. Published online May 5.

Russo, A., et al., 2019. Liquid biopsy tracking of lung tumor evolutions over time. Expert Rev. Mol. Diagn. 1–10. https://doi.org/10.1080/14737159.2020.1680287. Published online October 16.

MEETING REPORT

Accelerating the development of innovative cellular therapy products for the treatment of cancer

Mark D. Stewart[1,*], Anne Keane[2], Lisa H. Butterfield[3], Bruce L. Levine[4], Bruce Thompson[2], Yuan Xu[5], Chris Ramsborg[6], Ann Lee[6], Michael Kalos[7], Chin Koerner[8], Timothy Moore[9], Ingrid Markovic[10], Laura Lasiter[1], Ramy Ibrahim[3], Jeffrey Bluestone[3,11], Ellen Sigal[1], Jeff Allen[1]

[1] Friends of Cancer Research, Washington, DC, USA
[2] Lyell Immunopharma, San Francisco, California, USA
[3] Parker Institute for Cancer Immunotherapy, San Francisco, California, USA
[4] University of Pennsylvania, Philadelphia, Pennsylvania, USA
[5] Legend Biotech, Piscataway, New Jersey, USA
[6] Juno Therapeutics, A Bristol-Myers Squibb Company, Seattle, Washington
[7] Arsenal Biosciences, South San Francisco, California, USA
[8] Novartis Pharmaceuticals, Rockville, Maryland, USA
[9] PACT Pharma, South San Francisco, California, USA
[10] Genentech, A Member of the Roche Group, Washington, DC, USA
[11] University of California San Francisco, San Francisco, California, USA

ABSTRACT

The field of cell therapy is rapidly emerging as a priority area for oncology research and drug development. Currently, two chimeric antigen receptor T-cell therapies are approved by the US Food and Drug Administration and other agencies worldwide for two types of hematologic cancers. To facilitate the development of these therapies for patients with life-threatening cancers with limited or no therapeutic options, science- and risk-based approaches will be critical to mitigating and balancing any potential risk associated with either early clinical research or more flexible manufacturing paradigms. Friends of Cancer Research and the Parker Institute for Cancer Immunotherapy convened an expert group of stakeholders to develop specific strategies and proposals for regulatory opportunities to accelerate the development of cell therapies as promising new therapeutics. This meeting took place in Washington, DC on May 17, 2019. As academia and industry expand research efforts and cellular product development pipelines, this report summarizes opportunities to accelerate entry into the clinic for exploratory studies and optimization of cell products through manufacturing improvements for these promising new therapies.

© 2020 International Society for Cell and Gene Therapy. 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/)

Introduction

Advancements in cancer immunology and recent clinical experience with emerging cellular therapeutics, such as tumor-infiltrating lymphocytes, engineered T-cell receptor T cell and chimeric antigen receptor (CAR) T-cell therapies, are generating huge interest and activity both academically and industrially. Additional platforms and technologies, including cellular therapies based on natural killer and other immune cells, as well as novel gene-editing approaches have entered into clinical trials. These emerging therapeutics have the potential to rapidly change cancer treatment and represent a new paradigm in medical therapy for cancer.

To date, CAR T-cell therapies have been approved by the US Food and Drug Administration (FDA) only for certain types of leukemia and lymphoma; other T-cell based therapies have shown remarkable activity in a limited number of solid tumors but have not yet progressed to approval [1−5]. There is great interest in exploring these new treatment modalities to encompass the treatment of solid tumors, which comprise 90% of all cancers and the majority of cancer deaths [6]. Currently, multiple scientific challenges exist for the successful use of T-cell−based therapies in solid tumors, including issues related to target antigen selectivity and expression, the immunosuppressive nature of the tumor microenvironment, tumor T-cell infiltration and T-cell exhaustion. Academia and industry are attempting to address these barriers, and numerous T-cell−based product

* Correspondence: Mark Stewart, PhD, Friends of Cancer Research, 1800 M Street NW, Washington, DC 20036, USA.
E-mail address: mstewart@focr.org (M.D. Stewart).

candidates are being developed, involving various cell subtypes, autologous and allogeneic approaches, molecular engineering strategies and many different antigen targets. However, pre-clinical *in vivo* animal models are limited in their ability to predict product safety and efficacy for T-cell−based therapeutics, which limits progress.

To advance these therapies ultimately for a larger number of patients, and in particular for those with solid tumors, it would be desirable to design small, data-intensive exploratory clinical studies to differentiate which approaches warrant continuation and further focus. These studies would provide an opportunity to optimize the candidates to subsequently advance into full product development by generating knowledge that cannot be gained using currently available pre-clinical models. Small, early clinical studies also have the potential to facilitate a better understanding of the biology of T-cell−based therapeutics and the product attributes driving efficacy and safety. However, clinical data can be obtained only after the compilation and submission of an investigational new drug application (IND) for each candidate to be evaluated. These IND procedural requirements can make it prohibitively slow and expensive to pursue this critical opportunity for more than a select few product candidates. This can be particularly problematic for academic researchers and small biotechnology companies who generate much of the most innovative science in the field.

Furthermore, there can be varying interpretations of FDA guidance regarding phase appropriate current Good Manufacturing Practice (cGMP) requirements for reagents, plasmids, peptides, vectors and T-cell infusion products for use in the early investigational setting. Consequently, some institutions have imposed very strict cGMP requirements on all investigators that are more applicable for later-stage clinical development. These strict requirements significantly increase the cost and time for academic centers and industry to manufacture early investigational cell products and extend the time to evaluate which approaches should be taken to late-stage/pivotal clinical trials.

Ensuring that the most effective T-cell−based therapeutics are developed for the largest number of patients requires the adoption of a new adaptive manufacturing paradigm as more patients are treated, and more clinical, translational and product quality data are collected during a product lifecycle. In the late-stage development and post-licensure settings, as product and process knowledge increases, a strategy that enables adjustment of the manufacturing process conditions based on patient or patient-specific raw material information to maximize product quality for all patients would be beneficial without the need to conduct costly and lengthy studies. Furthermore, there is an opportunity to develop a regulatory framework for expedited clinical development to facilitate this adaptive learning-based manufacturing paradigm to allow for patient-level modifications for a subset cohort of patients, especially as understanding of the linkage between product quality attributes, manufacturing processes, clinical efficacy and safety evolves through late-stage development and post licensure.

Opportunities to Accelerate Early Discovery Through IND Application Flexibility

The FDA's 2006 Exploratory IND Studies Guidance acknowledged the need to reduce the time and resources expended on candidate products that are unlikely to proceed to licensure and described early phase 1 exploratory approaches that are consistent with regulatory requirements and maintain needed human subject protection, but which involved lesser requirements and lower costs, enabling sponsors to progress more efficiently [7]. This guidance also acknowledged that there is a great deal of flexibility in the amount of data that needs to be submitted with an IND application.

Application of the exploratory IND concept to early, small clinical studies for the purpose of candidate selection for T-cell−based

therapeutics would facilitate the critical opportunities described above. However, modifications would be needed. The current guidance explicitly states that an exploratory IND study is intended to involve very limited human exposure and to have no therapeutic or diagnostic intent. Post-infusion expansion of cellular therapies, the durable nature of cellular products and the ethical requirement to ensure clinical equipoise for patients with life-threatening cancers necessitate that they be dosed at therapeutic levels and with therapeutic intent. Nonetheless, a science- and risk-based approach to an expansion of the exploratory IND concept as it is applied to T-cell−based therapies is possible and appropriate. This would facilitate the evaluation of the safety and activity of next-generation T-cell−based therapeutics that could fundamentally improve their efficacy via small, data-intensive clinical studies.

An expanded exploratory IND pathway would facilitate the efficient generation of clinical data on multiple T-cell−based product candidates or hypotheses in small (N generally less than 30 patients per cohort) studies, reducing the regulatory burden for the sponsor. To ensure patient protection, enrollment in exploratory studies should be limited to patients with advanced cancers of unmet need and limited or no treatment alternatives. The total numbers of patients to be treated under an exploratory IND should be limited to the number required to evaluate the hypotheses to be tested.

Exploratory-phase protocols should be designed with a focus on patient safety and should incorporate opportunities to minimize risks. Therefore, appropriate consideration should be given to protocol design features (Table 1) [8,9].

The following sections outline how phase-appropriate cGMP compliance focused on product quality and patient safety and a streamlined parent-child IND alternative to the current single IND per drug product process would further facilitate the conduct of these studies under an expanded exploratory IND paradigm. T-cell−based therapies are used as specific examples, although the proposal could apply to other cellular products. Table 2 provides a summary of proposed strategies to facilitate the development of cellular therapies.

Phase Appropriate cGMP Compliance Focused on Product Quality and Patient Safety

The US FDA's 2008 Guidance for Industry: cGMP for Phase 1 Investigational Drugs provides a framework whereby more phase-appropriate manufacturing can occur for early studies [10]. The recognition that smaller-scale manufacturing processes may be excluded from some of the controls required for later stages of development where larger numbers of patients are exposed to treatment or for commercialization is critical to support innovative research and establish a better understanding of the biological impact of new therapeutics in small investigational human studies. However, consistent understanding and interpretation of this guidance, especially as it would apply to exploratory cellular therapy INDs, is needed. We provide several key examples below where explicit alignment between FDA, academic and government institutions and industry would facilitate the early exploratory clinical studies described above.

Table 1
Protocol design features to minimize risk.

Judicious dose escalation, cohorts and DLT windows
Adequate dosing intervals and safety assessments
Ongoing assessment by a safety monitoring committee
Incorporation of pre-specified safety, efficacy and futility decision points (Simon two-stage design)
Pre-planned early reporting of safety results
Explicit characterization of studies as "exploratory" in protocol and informed consent form

DLT, dose-limiting toxicity.

Table 2
Summary of strategies to facilitate development of cellular therapies.

Strategy	Description/process
Early phase	
Parent-child IND	Parent IND would contain common sections providing all relevant information for the candidates or manufacturing alterations. Each child IND would cross-reference common sections while providing only the candidate- or process-specific information.
Phase-appropriate cGMP compliance	Phase0appropriate manufacturing requirements would focus on product quality and patient safety using a risk-based approach to enable more efficient manufacturing processes in early-phase development.
Exploratory IND paradigm	Enrollment in trials with exploratory cellular therapy INDs would be limited to patients with advanced cancers and limited or no treatment alternatives. Early planned safety reporting would support an open regulatory dialogue. This pathway would facilitate the efficient generation of clinical data and inform whether more formal trials should be pursued.
Late phase	
Adaptive manufacturing process	An adaptive manufacturing process with the goal of generating a highly similar drug product from the patient-specific starting material is needed. A regulatory strategy that adjusts a process as product and process knowledge increases and based on patient or patient-specific raw material information to maximize product quality for all patients will permit the avoidance of extensive costly and lengthy clinical studies.
Post-marketing product optimization and modifications	As we gain stronger product knowledge and process understanding, modifications to manufacturing processes could be managed via a pre-negotiated plan with health authorities (e.g., Post-Approval Lifecycle Management or Comparability Protocol). Filing requirements for the change may include a combination of an analytical comparability assessment and/or a small clinical study, analogous to a bioequivalence study for a new process. A post-market commitment could be considered to demonstrate/confirm the efficacy of the new process.

Implicit in any approach for manufacturing phase 1−appropriate materials is a focus on patient safety, and the concepts below are proposed with an emphasis on risk assessments and analytical testing to determine and manage potential impact to patient safety. As such, T-cell−based cellular products would undergo release testing following manufacture for standard safety attributes, such as sterility, absence of mycoplasma and endotoxin, viral integration elements (vector copy number), identity, purity and potency.

We note that if remarkable efficacy were seen for a product development candidate tested in an exploratory IND, the requirement for a full IND with more burden of proof for manufacturing process and product knowledge and control would still apply with the potential for increased associated effort and cost. Sponsors may decide to mitigate this risk by pursuing limited process development activities in parallel with clinical studies under an exploratory IND.

A risk-based approach to requirements for the production of raw materials and drug substance (e.g., viral vectors) for T-cell−based therapeutics could more rapidly lead development teams to better test improved combinations of therapeutics, single-chain variable fragments alterations, novel manufacturing interventions, etc., which would lead to more robust and effective products that do not fail in later-stage development studies. Flexibility to permit the use of representative viral vectors not necessarily produced in a GMP facility, but rather in well-controlled laboratories, would result in significant monetary and time savings with little risk to patients. These opportunities could reduce the total time to manufacture investigational T-cell−based therapeutic candidates for use in an early clinical study under an exploratory IND by approximately 50%, as depicted in Table 3 and described in greater detail in subsequent sections.

Reduction in the infrastructure requirements for the manufacture of plasmids

Currently, production of plasmid DNA for downstream production of viral vectors and/or for gene-editing tools is often outsourced to a limited number of companies, resulting in high costs and long manufacturing queues. Generally, sponsors and academic researchers have the technical capabilities to produce these plasmid DNA's, but interpretations of FDA guidance have led to local institutional policies requiring cGMP-grade plasmids for clinical studies. Due to the high infrastructure requirements (International Standards Organization-7 clean rooms, fully developed quality systems and cGMP trained personnel and associated resources) needed to produce cGMP-grade plasmid DNA, many institutions have not invested in the development of the manufacturing and quality infrastructure to produce these raw materials internally. In the industry setting, the impression that cGMP-grade plasmids may be required increases the cost and time associated with manufacturing investigational cellular products. Manufacture of cGMP-grade plasmids for small, exploratory clinical trials of multiple early cellular product candidates unnecessarily increases the cost and time to conduct these studies because it is expected that many of the candidates would not progress into full product development and licensure.

As an alternative to a requirement for cGMP-grade plasmids, high-quality (HQ) fit-for-purpose plasmids may be acceptable. Plasmid DNA can be tested and characterized to confirm its suitability for downstream use in early, exploratory clinical trials with little risk to patient safety.

For example, the regulatory burden associated with the manufacture of HQ DNA plasmids for exploratory clinical studies could be reduced by eliminating the need for an *Escherichia coli* master cell bank. Note that a sponsor could also make a business decision to create the master cell bank and then freeze it, deferring the need for time-consuming and expensive testing until a decision was made to go forward with full development with that product candidate. Manufacturing could occur with review of production protocols, analytical results, manufacturing batch records and release tests performed by a second independent technical expert rather than quality assurance personnel. A certificate of testing (CoT) could be produced summarizing the test results. The authors proposed a representative fit-for-purpose CoT for plasmid DNA for early-phase clinical trials in a previous publication [11]. In essence, a CoT is similar to a certificate of analysis but differs in a few key elements: (i) tests are mostly compendial and may not be fully qualified/validated; (ii) tests may be peer reviewed by a technical expert (in lieu of a quality assurance resource); and (iii) test results have a Target Value in lieu of Acceptance Criterion. In addition, because the plasmid DNA materials are stable when frozen and anticipated to be used quickly in downstream manufacturing of viral vectors, at this stage the need to generate stability data could be weighed against the timing of use and available research data and, in some cases, waived.

Table 3
Summary of efficiencies gained through early stage manufacturing.

CMC activity	Typical time[a] investment	Areas of proposed flexibility	Potential time[a] savings	Potential cost savings[c]
Use of R&D reagents	3–6 mo	Increasing options for use of R&D reagents and reducing cost and time to either enable or negotiate GMP manufacture of reagents	1–3 mo	$–$$$
Plasmid manufacturing	4 mo (+3–6 mo in queue)	Reduced plasmid characterization and infrastructure requirements	5–7 mo	$$
Viral manufacturing	6 mo (+9–12 mo in queue)	Waive RCL testing in lieu of surrogate testing; reduced cGMP requirements for ancillary reagents	4 mo	$
Cell product engineering runs (3 runs)	3 mo	Use representative pilot virus for parallel cell product engineering runs	2 mo[b]	N/A

LVV, lentiviral vector; N/A, not applicable; R&D, research and development; RCL, replication competent lentivirus.
[a] All time estimates are approximate.
[b] There is some overlap in the time savings between the shortened LVV manufacturing timelines, and the engineering runs using pilot materials. Overall, the ability to demonstrate process control using representative materials means that activities are not reliant on manufacturing and release of LVV.
[c] $, hundreds; $$, thousands; $$$, tens of thousands.

Use of phase-appropriate vector testing strategies, including reductions in the replication competency testing requirements

The current replication competency virus assay is based on testing vector supernatant or end of production cells on susceptible human cells over an 8- to 10-week period; this requirement adds significant expense and time to vector release timelines, and, hence, to overall product manufacturing activities. Despite theoretical concerns, the risk of replication competency-related recombination events using third-generation viral vectors is extremely low because the elements required for virus replication are separated across three or four different plasmid DNAs and the 3' untranslated region portion of the transfer plasmids have been modified, resulting in transcriptional inactivation of the long terminal repeat in the proviruses after integration. With respect to viral vectors currently used in cell therapy products, researchers have documented that, to date, no viral vector recombination events have been observed in T cells across hundreds of patient product tests [12,13].

Alternative vector release testing based on a surrogate qualified/validated qualitative polymerase chain reaction (qPCR) test for the glycosaminoglycans and vesicular stomatitis virus G glycoprotein or similar envelope gene sequences depending on the viral vector pseudotype, as has been recently suggested by Skrdlant *et al.* may be acceptable, particularly in the context of early, exploratory studies in patients with limited or no remaining treatment options and poor long-term survival [14].

Vector and cellular drug product release decisions for such exploratory studies could be made on the basis of surrogate testing; if required, full, culture-based replication competency-based testing could be conducted in parallel in the background. The results of the full-culture testing would be available within the period of post-infusion patient follow-up during which time patients would be followed up for the development of treatment-related malignancy.

Use of a risk-based approach for determining safety of reagents used in early clinical trials

Extensive manufacturing requirements for reagents (e.g., activation beads, selection reagents, cytokines and recombinant growth factors) create a time and cost burden in early development. For early clinical trials, it may be reasonable to reduce these requirements, based on an appropriate risk assessment. Typically, these reagents are produced and stored frozen at higher concentrations to ensure greater stability. During manufacturing, a reagent is thawed and diluted to the working concentration and then added to a much larger culture volume. Unless the reagent is used constantly throughout the entire manufacturing process, several rounds of washing, media changes and formulation of the final cell product will significantly dilute the reagent. Similar to the manufacturing requirements for plasmid DNA, fit-for-purpose

requirements (relying on science- and risk-based approaches to ensure patient safety and quality of the reagent) for HQ reagents used within the manufacturing process for early-phase clinical studies would significantly reduce the cost and time burden associated with using innovative reagents, while ensuring patient safety. An emphasis on risk assessments to identify potential impact to patients (e.g., sterility/bioburden and products of animal origin) could provide guidance to academic researchers and industry partners. For non-pharmacopoeial reagents of non-biological origin, a review of a CoT may provide assurance that a reagent is fit-for-purpose for use in the manufacturing of cellular products for small, early clinical studies. For reagents of biological origin (e.g., human serum), purchasing from an accredited supplier, along with a certificate of analysis (source, sterility, endotoxin, infectious agents, mycoplasma) can confer suitability of use. Academic or small industry sponsors may benefit particularly from opportunities to share costs and risks by collaborating on the development of peer-reviewed tests in lieu of certificates of analysis and the accreditation of suppliers. An illustrative example of the approach to characterization of a novel recombinant cytokine, based on the concepts provided in International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Q6B and other regulatory guidance, is available [11].

Opportunities for flexibility in cell processing

Given the resources required and complexity of manufacturing T-cell−based therapeutic products, identifying similar flexibilities in the cell-processing space would provide significant opportunities for innovation. Although a robust discussion of the kinds of flexibility desired is out of scope for this document, a few examples and the anticipated impacts are offered below. Typically, a T-cell−based therapeutic is engineered using a relatively similar set of manufacturing unit operations: acquisition of patient starting material through apheresis/leukapheresis, possible isolation/purification of the T cells through gradient, magnetic or alternative selection means, activation and retroviral transduction to introduce the CAR or T-cell receptor, expansion of the engineered T cells and final harvest and cryopreservation. Although there are variations on the above approach and a number of different pieces of equipment used in various manufacturing processes, the general process lends itself to some potential flexibilities in the early development space.

Flexibility to permit the use of representative viral vectors in cell product engineering runs

Current approaches to qualify a cellular therapy manufacturing process is often interpreted as requiring the use of GMP-grade viral vector in at least three engineering runs conducted to confirm the adequacy of the cellular product manufacturing process. Clarity that the use of the representative pilot (i.e., development-grade viral vectors

manufactured in accordance with the final manufacturing process) would be acceptable could result in significant time savings because the cellular product engineering runs could be run in parallel with the actual GMP production runs for the viral vector. Additionally, because much of the development work for autologous cell therapies is done at scale, fewer engineering runs (e.g., two) would be reasonable. As such, data from both development runs (e.g., in the process development laboratory) and engineering runs (e.g., in the GMP manufacturing facility) could be combined to demonstrate adequate understanding and control of the process to support an IND submission.

Phase-appropriate release testing

For early-phase exploratory trials, a focus on testing cell product components related to safety can provide flexibility. Safety would be assessed via testing for sterility, mycoplasma (via a rapid testing paradigm), endotoxin, etc., which are each important to demonstrating a lack of contamination of the cell product. Testing the cell product for effectiveness of transduction through assessment of vector integration into the T-cell genome can be performed by determining the average vector copy number via qPCR. Additionally, surrogate measures of viral replication competency can be performed using qPCR with primers against elements of the viral genome (discussed above as part of replication competent retrovirus/replication competent lentivirus testing above). Identity, purity and potency are important release assays used to demonstrate that a particular manufacturing process was able to successfully yield the expected product. While identity and purity tests are generally required even in early phases, due to the complexity of cell products, potency assays may be in development or deferred in very early clinical phases in favor of a well-qualified dose-determining assay with incremental implementation of the potency assay as development proceeds.

Development of a parent-child IND framework to reduce the regulatory burden associated with the clinical testing of multiple potential product candidates

In the setting of small, exploratory early data-intensive clinical studies intended to investigate the safety, feasibility and mechanism of action of several closely related T-cell−based candidates or related manufacturing process alterations (e.g., process alterations to maintain stemness), a more efficient parent-child IND structure and process may be appropriate.

The parent IND submitted by a single sponsor/investigator would contain common sections providing all of the relevant information manufacturing, pre-clinical and clinical information for the prototype candidate to be moved forward in development. Each subsequent child IND submitted by the same sponsor/investigator would cross-reference those sections of the "parent" IND that were in common, while providing additional unique "child"-specific information.

An exploratory parent-child IND is a feasible approach to reducing the regulatory procedural burden associated with evaluating multiple highly related T-cell−based therapeutic constructs or manufacturing alterations in small clinical studies. The parent IND would contain sections providing all of the common information relevant for the to-be-tested initial candidates or manufacturing alterations. For each candidate or manufacturing alteration, a child IND would also be submitted. This child IND would depend on heavy cross-referencing to the common sections in the parent IND while providing only the candidate- or process-specific information (e.g., chemistry, manufacturing, and controls [CMC] or non-clinical data) in separate sections. We note that cross-referencing, for example, by an academic investigator to previously submitted information in another IND, with appropriate authorization, is an accepted current practice.

At the time of initial IND submission, the parent and child IND could be assigned separate IND numbers, to facilitate safety reporting

and so on, but reviewed in parallel within the standard 30-day IND review window. The exploratory IND would include an explicit agreement by the sponsor that once the early testing of a particular construct or process is completed or discontinued, the associated exploratory child IND would be withdrawn. If the sponsor intends to proceed with full development of a candidate or manufacturing process, a new, traditional IND would be submitted for that candidate. Prior clinical experience with the candidate might result in an expedited review of the new product development IND. Subsequent candidates or processes consistent with the common information in the original parent IND could subsequently be added as additional children to the original parent, again relying heavily on cross-references.

Because the time and resource savings associated with the use of parent-child INDs would only be realized in situations where most of the information contained in the parent IND would be relevant to all of the investigational candidates, the use of parent-child INDs would be limited to situations where the commonalities between the early cellular therapy candidates or manufacturing interventions are great enough to produce real gains in efficiency for both the sponsor and the FDA reviewing division. For example, an exploratory IND might be limited to candidates directed at the same target, even if in multiple tumor types (e.g., neurotrophic tyrosine kinase receptor 1 expressing non-small cell lung cancer, or triple negative breast cancer). Whether a parent-child IND is appropriate for a particular set of candidates could be discussed in an INTERACT (Initial Targeted Engagement for Regulatory Advice on Center for Biologics Evaluation and Research products) or pre-IND meeting or the justification could be provided in the IND itself (with an associated risk of delay if the FDA disagrees). Given the potential efficiencies gained, the parent-child IND concept is likely to be embraced by academic- and industry-sponsored researchers moving forward and warrants additional discussions with the FDA in the context of specific development programs.

Pathways to Enable Manufacturing and Testing Evolution During Late-stage Development and Post Licensure

In the case of T-cell−based therapeutics and other cell-based therapies, making these products effective for the greatest number of patients may require adjusting manufacturing parameters for specific patient subsets. The single manufacturing process framework is often chosen for regulatory expediency and a lack of product knowledge to discriminate between patients. At the same time, patient-to-patient variability in the quality of T cells from these patients can lead to suboptimal drug product quality for a subset of patients when a single manufacturing process is used for all patients. It may be more appropriate to adapt the manufacturing process for a subset of patients to increase the efficacy for the specific patient cohort, without impacting safety and efficacy for patients already responding using the original manufacturing process.

These new process parameter combinations for patient subsets are discovered during clinical development as more patients are treated and more clinical, translational and product quality data are collected. As product and process knowledge increases, a regulatory strategy that adjusts a process based on patient or patient-specific raw material information to maximize product quality for all patients will be necessary without conducting extensive costly and lengthy clinical studies. Current regulatory requirements and processes may not readily allow for patient-level modifications, especially when the understanding of the linkage among product quality attributes, manufacturing processes, clinical efficacy and safety continue to evolve late in development or after licensure.

Traditionally, manufacturing processes are locked in advance of late-stage clinical trials to be able to repeatedly measure effect across many patients. As product and process knowledge becomes more comprehensive for cell therapies, it can be anticipated that an adaptive manufacturing process could be developed with the goal of

generating highly similar drug product from variable patient-specific starting materials. The product and process knowledge to enable adaptive manufacturing in most cases will not emerge until a large number of patients are treated because the correlative analysis to discover the relationship is not available until the enrollment of the pivotal trial is already well advanced. An example of this type of relationship includes the frequency of specific T-cell subtypes [15].

Using Post-Approval Lifecycle Management–like plan for making manufacturing and testing changes

As we gain stronger product knowledge and process understanding and are able to correlate their impacts to clinical safety, efficacy and durability results, the insights gained are likely to lead to improvements that can be made to the manufacturing process and/or quality control tests, as anticipated in International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Q12 [16]. For example, based on data gained during clinical development, a process adaptation (e.g., culture medium optimization, culture condition optimization, analytics method improvement) is identified, which modestly improves manufacturing consistency, increases the efficacy or reduces adverse events (i.e., does not impact labeled dose). The magnitude of change in clinical benefit may not be large enough to justify a full clinical development but is still beneficial to patients. For these changes, modifications could be managed via a pre-negotiated plan with health authorities (e.g., Post-Approval Lifecycle Management [PALM] or Comparability Protocol). The filing requirements for the change may include a combination of an analytical comparability assessment, and/or a small clinical study, analogous to a bioequivalence study for a new process. A post-market commitment could be considered to demonstrate/confirm the efficacy of the new process.

Create CMC commercial process change reporting categories for cell-based therapies

The FDA December 2017 draft guidance for CMC changes to an approved application intended to assist manufacturers of biological products in assessing the reporting category for CMC changes. This guidance provides a starting framework that can be further extended to T-cell–based therapies. As the cell-based therapeutic industry accumulates commercial manufacturing experience, sponsors can identify the most frequent manufacturing changes and propose recommended reporting categories based on risk assessment: Annual Reportable (AR), Changes Being Effected–0, Changes Being Effected–30 or Post-Approval Supplement. Consistent with the fundamental guiding principle from the biologics guideline, the reporting category selected should be commensurate with the risk of an unintended outcome resulting from changes involving these elements. When assessing the impact of change on product quality, the historical product and process knowledge including experience gained during commercial manufacturing should be fully leveraged. Developing a best practice guide for cell therapy with specific examples of process and testing changes for the range of categorization would be a beneficial activity, which could be created by an industrial consortium.

However, it should be noted that the overall variability in cell-based therapy processes is influenced by the incoming patient-to-patient variability. Therefore, the traditional process performance qualification approach using three healthy donor batches to qualify each change has limited applicability and, instead, a rigorous, continuous process verification (CPV) plays a larger role in demonstrating process control. Use of healthy donors to characterize process and analytical variability in theory is a good approach, but a significant number of healthy donors are potentially needed to quantify the variability contribution of the process and analytics. This consumes resources and manufacturing capacity that otherwise would be used to produce clinical or commercial

products. Hence, a concurrent qualification approach, where a change is introduced in manufacturing based on small-scale data and is subject to verification through a CPV program during clinical/commercial use, is not only more efficient but would also allow the confirmation of change in the setting of real patients instead of healthy donors. In addition, stand-alone qualification of the specific process or manufacturing change on a risk-based impact assessment without the need for end-to-end full process performance qualification may be sufficient in some cases (e.g., a change in a supplier of raw materials, reagents and solvents that have a minimal potential to affect product quality) provided that the materials' specific use, physicochemical properties, impurity content and acceptance criteria remain comparable could be validated offline and reported as an AR. Additionally, a change from a manual operation to an automated operation that does not change the process parameter set points could be addressed through automation qualification and reported as an AR.

Lastly, in some cases, demonstrating analytical comparability at the appropriate in-process intermediate level may be sufficient. For example, demonstration of comparability for the vector bulk material due to a process change in the vector manufacturing process should not require demonstration of final product comparability post-transduction. Analytical comparability of the bulk viral vector and, if needed, use of a small-scale model to confirm transducibility of the cellular in-process product at the same transduction levels should be considered sufficient. The life cycle plan for process and method changes needs to be carefully sequenced so that potential impact of the changes is seen throughout the CPV program. Changes to process parameters outside of previously validated ranges should be assessed with respect to criticality to process performance and product quality.

The reporting categories and extent of requalification for these changes will be assessed keeping the above considerations in mind. A risk-based approach to determine the extent and approach of qualification should be used that would determine if (i) qualification can be performed using small-scale or whether full-scale confirmation is needed; (ii) qualification exercise can be limited to evaluating product attributes of the impacted intermediate or the final drug product; and (iii) separate qualification is needed or if heightened CPV program for a period of time can be used. Given that many cell therapy companies are focused on early access to the promising therapies, several process improvements are deferred and become part of the post-approval life cycle plan (Box 1).

Box 1. Examples of such deferred changes.

1) New primary packaging components for the final product to simplify ease of administration and enable more clinical sites
2) New activation reagents; introduction of a new media processing system to improve manufacturing robustness
3) A higher-grade of fetal bovine serum to improve reliability
4) Change of buffer manufacturer from in-house to an external manufacturer
5) Automation of manual processing steps
6) Automation of flow cytometry data analysis
7) Increase in vector production scale to meet increasing demand
8) Change to a rapid sterility method
9) Rapid microbiological testing
10) Change of vector manufacturing process to a suspension cell culture process
11) The addition of an identical manufacturing suite to double capacity for both vector and drug product
12) Change in the antibiotic resistance in the vector cell bank/plasmid; improved potency method
13) Change to stability data for expiry extension

Quality standards for ancillary materials used in the manufacturing of cell-based therapy products intended to be developed as commercial products

Currently, sponsor companies are restrained by the limited numbers of GMP producers of these ancillary materials (e.g., recombinant proteins, growth factors, cytokines and small molecules) because of the regulatory requirements associated with choosing novel reagents. For the foreseeable future, the supply chain will be a critical path for product commercialization. The root cause for this limited supply chain is multi-factorial, but some modifications of applicable regulatory guidance could accelerate innovation.

Stakeholders desire more uniform feedback from within health authorities around quality and testing standards for non-GMP ancillary materials. Stronger guidance on how to stratify quality and/or characterization requirements based on whether they are excipients, product contacting (primary) or secondary ancillary material (e.g., plasmids used in viral vector manufacturing) or tertiary ancillary materials would be beneficial to the field. Moreover, greater health authority alignment with the principles published in United States Pharmacopeia <1043> or other guidance documents could result in greater consistency in CMC development across multiple phases.

Conclusion

The "Designing the Future of Cellular Therapies" meeting brought together a group of stakeholders from academia, biotechnology, pharmaceuticals, government, regulatory agencies, patient advocates and non-profits. Several opportunities and strategies that could expedite T-cell−based therapies into first-in-human studies and ensure that T-cell−based therapeutics are impactful for the greatest number of patients were discussed (Table 2). Ongoing dialogue and collaboration with the FDA and other global health authorities to increase the clarity and uniformity of regulatory requirements will be critical to the development of the field. Moreover, efforts to encourage transparency, collaboration and data sharing are needed so changes can be appropriately monitored, which would allow the field to adapt to improvements efficiently. The proposals could be particularly useful in bringing cutting edge biological and genetic approaches forward to enhance the potential of the next generation of cell therapies to demonstrate efficacy in the highly complicated tumor microenvironment and extend these promising new therapies to the many patients with life-threatening solid tumors with no remaining treatment options.

Funding

Declaration of Competing Interest

Jeff Bluestone is employed by Sonoma BioTherapeutics, a Treg cell therapy company, and reports grants and personal fees from Juno, non-financial support from Becton Dickenson, and has several issued and pending patents. Lisa Butterfield received consulting honoraria from Calidi Biotherapeutics, SapVax, NextCure, Replimmune, Western Oncolytics, Torque Therapeutics and Phyxis Oncology and has stock options in Calidi Biotherapeutics, NextCure and Phyxis Oncology. Ramy Ibrahim is an observer on the board of directors for Lyell Immunopharma and serves on the Scientific Advisory Board for Bit Bio. Michael Kalos is an employee of Arsenal Biosciences and has stock in Arsenal Biosciences and Johnson and Johnson. Anne Keane is employed by Lyell Immunopharma. Chin Koerner is employed by Novartis. Ann Lee is employed by Juno Therapeutics, A Bristol Myers Squibb Company. Bruce Levine is on the Scientific Advisory Board for Avectas, Brammer Bio, Incysus, CRC Oncology/Cure Genetics, Vycellix and Ori Biotech, received consultancy fees from Brammer Bio, Incysus and CRC Oncology/Cure Genetics, has patents and royalties with and received research funding from Novartis Pharmaceuticals Corporation and holds equity ownership in and received research funding from Tmunity Therapeutics. Ingrid Markovic is employed by Genentech. Tim Moore is employed by PACT Pharma. Chris Ramsborg is employed by Juno Therapeutics, A Bristol Myers Squibb Company. Bruce Thompson is an employee of Lyell Immunopharma. Yuan Xu is employed by Legend Biotech and currently owns equity in Merck.

Author Contributions

Conception and design of the study: MDS, AK, LHB, BLL, RI, JB, ES, JA. Acquisition of data: MDS, AK LHB, BLL, BT YX, CR, AL, MK, CK, TM, IM. Analysis and interpretation of data: MDS, AK, LHB, BLL, BT, YX, CR, AL, MK, CK, TM, IM, LL, RI, JB, ES, JA. Drafting or revising the manuscript: MDS, AK, LHB, BLL, BT, YX, CR, AL, MK, CK, TM, IM, LL, RI, JB, ES, JA. All authors have approved the final article.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jcyt.2020.01.014.

References

[1] Stevanović S, et al. Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. Science 2017;356(80):200LP–205.

[2] Zacharakis N, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. Nat. Med. 2018;24:724–30.

[3] Tran E, et al. Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient with Epithelial Cancer. Science 2014;344(80):641LP–645.

[4] Angelo SP, et al. Antitumor Activity Associated with Prolonged Persistence of Adoptively Transferred NY-ESO-1 c259 T Cells in Synovial Sarcoma. Cancer Discov. 2018;8:944LP–957.

[5] Brown CE, et al. Regression of Glioblastoma after Chimeric Antigen Receptor T-Cell Therapy. N. Engl. J. Med. 2016;375:2561–9.

[6] NCI SEER. Cancer Stat Facts: Cancer of Any Site. Available at: https://seer.cancer.gov/statfacts/html/all.html. (Accessed: 14th August 2019)

[7] FDA. Guidance for Industry, Investigators, and Reviewers: Exploratory IND Studies. (2006). Available at: https://www.fda.gov/media/72325/download. Accessed: 14th August 2019.

[8] FDA. Master Protocols: Efficient Clinical Trial Design Strategies To Expedite Development of Oncology Drugs and Biologics. (2018). Available at: https://www.fda.gov/media/120721/download. Accessed: 14th August 2019.

[9] FDA. Adaptive Design Clinical Trials for Drugs and Biologics. (2018). Available at: https://www.fda.gov/media/78495/download. Accessed: 14th August 2019.

[10] FDA. Guidance for Industry: CGMP for Phase 1 Investigational Drugs. (2008). Available at: https://www.fda.gov/media/70975/download. (Accessed: 14th August 2019)

[11] Friends of Cancer Research & Parker Institute for Cancer Immunotherapy. Designing the Future of Cell Therapies. (2019). Available at: https://www.focr.org/sites/default/files/pdf/Friends_Cellular_Therapies_White_Paper.pdf. Accessed: 14th August 2019.

[12] Cornetta K, et al. Absence of Replication-Competent Lentivirus in the Clinic: Analysis of Infused T Cell Products. Mol. Ther. 2018;26:280−8.

[13] Cornetta K, et al. Screening Clinical Cell Products for Replication Competent Retrovirus: The National Gene Vector Biorepository Experience. Mol. Ther. - Methods Clin. Dev. 2018;10:371−8.

[14] Skrdlant LM, Armstrong RJ, Keidaisch BM, Lorente MF, DiGiusto DL. Detection of Replication Competent Lentivirus Using a qPCR Assay for VSV-G. Mol. Ther. Methods Clin. Dev. 2017;8:1−7.

[15] Fraietta JA, et al. Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. Nat. Med. 2018;24:563−71.

[16] FDA. Q12 Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle Management Core Guideline Guidance for Industry. (2018). Available at: https://www.fda.gov/media/113483/download. Accessed: 14th August 2019.

Biomarker-driven therapies for previously treated squamous non-small-cell lung cancer (Lung-MAP SWOG S1400): a biomarker-driven master protocol

Mary W Redman, Vassiliki A Papadimitrakopoulou, Katherine Minichiello, Fred R Hirsch, Philip C Mack, Lawrence H Schwartz, Everett Vokes, Suresh Ramalingam, Natasha Leighl, Jeff Bradley, Jieling Miao, James Moon, Louise Highleyman, Crystal Miwa, Michael L LeBlanc, Shakun Malik, Vincent A Miller, Ellen V Sigal, Stacey Adam, David Wholley, Caroline Sigman, Beverly Smolich, Charles D Blanke, Karen Kelly, David R Gandara, Roy S Herbst

Summary

Background The Lung Cancer Master Protocol (Lung-MAP; S1400) is a completed biomarker-driven master protocol designed to address an unmet need for better therapies for squamous non-small-cell lung cancer. Lung-MAP (S1400) was created to establish an infrastructure for biomarker screening and rapid regulatory intent evaluation of targeted therapies and was the first biomarker-driven master protocol initiated with the US National Cancer Institute (NCI).

Methods Lung-MAP (S1400) was done within the National Clinical Trials Network of the NCI using a public–private partnership. Eligible patients were aged 18 years or older, had stage IV or recurrent squamous non-small-cell lung cancer, had previously been treated with platinum-based chemotherapy, and had an Eastern Cooperative Oncology Group (ECOG) performance status of 0–2. The study included a screening component using the FoundationOne assay (Foundation Medicine, Cambridge, MA, USA) for next-generation sequencing, and a clinical trial component with biomarker-driven substudies and non-match substudies for patients who were ineligible for biomarker-driven substudies. Patients were pre-screened and received their substudy assignment upon progression, or they were screened at progression and received their substudy assignment upon completion of testing. Patients could enrol onto additional substudies after progression on a substudy. The study is registered with ClinicalTrials.gov, NCT02154490, and all research related to Lung-MAP (S1400) is completed.

Findings Between June 16, 2014, and Jan 28, 2019, 1864 patients enrolled and 1841 (98·9%) submitted tissue. 1674 (90·9%) of 1841 patients had biomarker results, and 1404 (83·9%) of 1674 patients received a substudy assignment. Of the assigned patients, 655 (46·7%) registered to a substudy. The biomarker-driven substudies evaluated taselisib (targeting *PIK3CA* alterations), palbociclib (cell cycle gene alterations), AZD4547 (*FGFR* alteration), rilotumumab plus erlotinib (MET), talazoparib (homologous recombination repair deficiency), and telisotuzumab vedotin (MET). The non-match substudies evaluated durvalumab, and nivolumab plus ipilimumab for anti-PD-1 or anti-PD-L1-naive disease, and durvalumab plus tremelimumab for anti-PD-1 or anti-PD-L1 relapsed disease. Combining data from the substudies, ten (7·0%) of 143 patients responded to targeted therapy, 53 (16·8%) of 315 patients responded to anti-PD-1 or anti-PD-L1 therapy for immunotherapy-naive disease, and three (5·4%) of 56 responded to docetaxel in the second line of therapy. Median overall survival was 5·9 months (95% CI 4·8–7·8) for the targeted therapy groups, 7·7 months (6·7–9·2) for the docetaxel groups, and 10·8 months (9·4–12·3) for the anti-PD-1 or anti-PD-L1-containing groups. Median progression-free survival was 2·5 months (95% CI 1·7–2·8) for the targeted therapy groups, 2·7 months (1·9–2·9) for the docetaxel groups, and 3·0 months (2·7–3·9) for the anti-PD-1 or anti-PD-L1-containing groups.

Interpretation Lung-MAP (S1400) met its goal to quickly address biomarker-driven therapy questions in squamous non-small-cell lung cancer. In early 2019, a new screening protocol was implemented expanding to all histological types of non-small-cell lung cancer and to add focus on immunotherapy combinations for anti-PD-1 and anti-PD-L1 therapy-relapsed disease. With these changes, Lung-MAP continues to meet its goal to focus on unmet needs in the treatment of advanced lung cancers.

Funding US National Institutes of Health, and AbbVie, Amgen, AstraZeneca, Bristol Myers Squibb, Genentech, and Pfizer through the Foundation for the National Institutes of Health.

Introduction

The Lung Cancer Master Protocol (Lung-MAP; S1400), part of the precision medicine initiative of the US National Cancer Institute (NCI), was created to facilitate the discovery of more effective therapies for patients with previously treated advanced squamous non-small-cell

Operations Office, San Antonio, TX, USA (C Miwa BS); **National Cancer Institute, National Institutes of Health, Bethesda, MD, USA** (S Malik MD); **Foundation Medicine, Cambridge, MA, USA** (V A Miller MD); **Friends of Cancer Research, Washington, DC, USA** (E V Sigal PhD); **Foundation for the National Institutes of Health, Bethesda, MD, USA** (S Adam PhD, D Wholley MS); **CCS Associates, San Jose, CA, USA** (C Sigman PhD, B Smolich PhD); **SWOG Group Chair's Office and Knight Cancer Institute, School of Medicine, Oregon Health and Science University, Portland, OR, USA** (Prof C D Blanke MD); **Department of Internal Medicine, UC Davis Comprehensive Cancer Center, Sacramento, CA, USA** (Prof K Kelly MD, Prof D R Gandara MD); **and Department of Medicine, Yale Cancer Center, New Haven, CT, USA** (Prof R S Herbst MD)

Correspondence to: Prof Mary W Redman, SWOG Statistics and Data Management Center and Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
mredman@fredhutch.org

Research in context

Evidence before this study

A formal literature search was not completed before this study. The concept was based on subject matter knowledge of the medical doctors (specifically, VAP), the reporting of The Cancer Genome Atlas, and a meeting between the US National Cancer Institute, US Food and Drug Administration, and academics and clinicians to launch the idea of a master protocol in lung cancer. S1400, the Lung Cancer Master Protocol (Lung-MAP), was created to address an unmet need to improve treatment options for patients with previously treated stage IV or recurrent squamous non-small-cell lung cancer. When the Lung-MAP (S1400) protocol was initiated in 2014, the standard of care options for patients with squamous non-small-cell lung cancer were platinum-based chemotherapy for first-line treatment and a small choice of chemotherapy monotherapies for previously treated patients, with no targeted therapy options. In 2012, The Cancer Genome Atlas reported on the characterisation of genomic alterations present in squamous non-small-cell lung cancers. The natural resulting question was whether targeted therapies in biomarker-defined subgroups in squamous non-small-cell lung cancer could be discovered in this population. However, a major concern existed regarding the feasibility of screening for sets of potentially rare alterations in an efficient way to minimise delays to begin treatment and to evaluate these targeted therapies efficiently. There also was a need to address care delivery disparities and bring these agents to the community and rural sites. Importantly, many of these mutations are rare, occurring in 5–15% of patients, making it essential to develop a master protocol that would provide a screening umbrella to allow researchers to answer this important clinical question.

Added value of this study

The Lung-MAP study (S1400) was the first master protocol launched within the National Clinical Trials Network (NCTN) of the National Cancer Institute. The trial demonstrated that biomarker-driven master protocols are feasible within an aggressive disease setting, such as previously treated advanced squamous non-small-cell lung cancers, that the infrastructure of the NCTN combined with unique public–private partnership enhanced the functioning and feasibility of conduct of the study, and that the infrastructure of a biomarker-driven master protocol could provide efficient answers to the activity of targeted therapies in rare populations. Of equal importance, the Lung-MAP trial established a roadmap for how to carry out master protocols within public and private settings and the lessons from the study and study conduct will continue to inform other such efforts in various settings, oncology and otherwise.

Implications of all the available evidence

The first chapter of Lung-MAP evaluated therapeutic options for patients with previously treated advanced squamous non-small-cell lung cancers. It demonstrated that the discovery of targeted therapies is a challenging endeavour, but important to pursue. Lung-MAP continues, now evaluating targeted therapies in all histological types of advanced non-small-cell lung cancers. The revised study also includes an additional focus on treatment options for patients with immune checkpoint inhibitor relapsed or refractory disease, an area of high unmet need. The Lung-MAP trial will continue to provide an infrastructure to evaluate investigational therapies and to learn more about how best to treat these types of lung cancers.

lung cancer. To achieve this goal, Lung-MAP established an infrastructure for the conduct of biomarker screening and rapid evaluation of molecularly targeted therapies in biomarker-defined subgroups that could lead to regulatory approval.

The premise for Lung-MAP (S1400) was that progress in squamous non-small-cell lung cancer would follow successes in other subtypes of non-small-cell lung cancer by the identification of effective molecularly targeted therapies for oncogene drivers (eg, EGFR and ALK).[1] This rationale was supported by the identification of potentially actionable or druggable molecular alterations in squamous non-small-cell lung cancer within sequencing efforts such as The Cancer Genome Atlas (TCGA) combined with advances in clinically applicable next-generation sequencing (NGS) technologies.[1–6] Given the lack of progress in the past few decades, there was a need to improve the efficiency of therapeutic drug evaluation in squamous non-small-cell lung cancer.

Evaluation of molecularly targeted therapies in populations with rare mutations using conventional stand-alone clinical trial designs had proved to be challenging logistically or infeasible. Stand-alone trials in rare populations require screening of large numbers of patients to enrol a small number of participants, thus requiring multicentre accrual. Single institutions might only enrol one or two patients at most despite screening many patients, making participation infeasible because of cost or logistical constraints. Additionally, patients might find waiting several weeks with a small chance of study participation unacceptable. The overall impact can be that studies in rare populations might either be infeasible or take a long time to complete.

Thus, the concept of biomarker-driven master protocols, in which multiple investigational therapies matched to biomarkers are tested within a single clinical trial infrastructure, emerged as an approach to accelerate drug testing and approval in biomarker-defined subgroups.[7–10] Biomarker-driven master protocols received immediate and enthusiastic support from the NCI, US Food and Drug Administration (FDA), industry, patient advocacy groups, and academic researchers.[11] Moreover, a natural efficiency strategy was to use the established infrastructure of the NCI's National Clinical

Trials Network (NCTN) to carry out biomarker-driven master protocols. Implemented in June, 2014, Lung-MAP (S1400) was the first master protocol for precision medicine launched within the NCTN, as well as the only one to integrate the potential for regulatory approval of new drug candidates.[12–20]

Here, we describe the Lung-MAP (S1400) protocol, including protocol design, patient eligibility, patients screened, results of study conduct, and some of the lessons learned from the conduct of this innovative trial.

Methods

Protocol design

Lung-MAP consisted of a screening component and an investigational study component. The screening component established common eligibility criteria for every substudy and procedures for specimen submission and analysis. The investigational study component consisted of multiple independently conducted and analysed substudies of two categories: biomarker-driven substudies for patients eligible based on detection of a biomarker or a set of biomarkers and non-match substudies for otherwise eligible patients not meeting the criteria to enrol in a biomarker-driven substudy. New substudies were independently added as new ideas were developed and were closed to accrual as they met the criteria for closure. Lung-MAP had a single investigational new drug application covering the overarching screening study and individual therapeutic substudies.

Partnership structure, study conduct, and accruing sites

Lung-MAP (S1400) was carried out through a public–private partnership including the Cancer Therapy Evaluation Program (CTEP) at the NCI, the adult NCTN groups (Alliance for Clinical Trials in Oncology, Eastern Cooperative Oncology Group-American College of Radiology Imaging Network [ECOG-ACRIN], NRG Oncology, and SWOG Cancer Research Network), the Foundation for the National Institutes of Health, and Friends of Cancer Research. The Lung-MAP (S1400) protocol was reviewed by the CTEP and approved by the institutional review board at each participating site. Sites were responsible for obtaining written, informed consent, which was obtained from all participants. Trial governance comprised representatives from the partners. A drug selection committee evaluated candidate drugs and biomarkers using prespecified criteria based on scientific and practical considerations (appendix pp 4–5). A trial oversight committee provided guidance on study design and conduct. The SWOG data safety and monitoring committee was responsible for oversight of all substudies.

Lung-MAP (S1400) was scientifically and operationally led by SWOG, with clinical and translational leadership representation from all NCTN groups. The Canadian Cancer Trials Group participated between Dec 18, 2015, and July 12, 2018, closing the study due to challenges with drug distribution across the US–Canadian border. Since inception, the FDA has been an active collaborator. Participation in the trial was available at more than 750 NCTN sites, including those in the NCI Community Oncology Research Program (NCORP).

Patient eligibility

Eligibility criteria at inception specified that patients were aged 18 years or older, had pathologically proven stage IV or recurrent squamous non-small-cell lung cancer confirmed by tumour biopsy, fine-needle aspiration, or both without mixed histologies, had no other previous untreated malignancies, had progressive disease in the opinion of the treating physician following one previous treatment with platinum-based chemotherapy, had no *EGFR* mutation or *ALK* fusion, and had sufficient tumour tissue for biomarker analysis. 11 months after the initiation of Lung-MAP, on May 26, 2015, eligibility was modified to allow patients who had previously received second-line or further treatment for stage IV or recurrent disease, with at least one line including platinum-based chemotherapy. At this time, the option to be pre-screened during previous treatment for stage IV or recurrent disease was also added. Patients with previous systemic therapy for earlier stage (I–III) lung cancer were eligible if progression on platinum-based chemotherapy occurred within 1 year from the last date that the patient received that therapy. Patients with an ECOG performance status of 0–2 were initially eligible to participate; however, 18 months after initiation of the trial, on Dec 18, 2015, patients with an ECOG performance status of 2 were disallowed due to the addition of pre-screening and concerns with toxicities of immune checkpoint inhibitors.

Biomarker platform and screening requirements

On-study NGS screening was required and was done with the Foundation One assay (Foundation Medicine, Cambridge, MA, USA).[6,21] This platform was selected based on a request for proposal process, in which a formal announcement and advertisement was created and distributed. Additionally, potential companies were invited to submit a proposal. A committee of experts from both academia and government (NCI) reviewed the proposals. Once all the potential participants had been reviewed, all members of the committee voted to select the single company that would be used for broad biomarker screening in Lung-MAP. Archival formalin-fixed paraffin-embedded (FFPE) tumour specimens in a Clinical Laboratory Improvement Amendments (CLIA)-certified and College of American Pathologists (CAP)-accredited laboratory (Foundation Medicine) were used for mutational analysis. Genomic DNA (≥50 ng) was extracted from FFPE specimens and sonicated[6,21] to fragments about 200 bp in size. Material underwent whole-genome shotgun library construction and hybridisation capture of at least 236 genes and selected introns of 19 genes involved in rearrangements. Using the

Illumina (San Diego, CA, USA) HiSeq 2000, 2500, and 4000 platforms, libraries selected by hybrid capture were sequenced using 49×49 bp paired-end reads to high uniform depth. Sequence data were processed using a customised analysis pipeline designed to accurately detect base substitutions, small insertions and deletions, focal copy number amplifications, homozygous gene deletions, and genomic rearrangements. Additionally, screening for MET expression was intermittently screened by immunohistochemistry. A tumour block or a minimum of 12 unstained slides were required, although up to 20 slides were requested. The treating institution's local pathologist was required to confirm that adequate tissue was available before tissue submission. Adequate tissue was defined as at least 20% tumour cells and a tumour volume of at least 0·2 µL. If the initial tissue submission was inadequate or if there was a sequencing failure, due to insufficient tumour content, DNA, tumour cellularity, or other tumour-related or assay-related reasons, sites could submit additional tissue for screening. Patients for whom biomarker results were not available (either due to inadequate tissue or sequencing failures) were not eligible to register for any of the substudies.

Substudy assignments
Substudy assignments were to be reported within 16 days from tissue submission for patients who were screened at progression and within 1 day of notification of progression for pre-screened patients. If the notice was never submitted, the patient would not have received a substudy assignment. Patients screened at progression received their substudy assignment as soon as the biomarker results were available. Subsequently, detection of one eligibility biomarker resulted in assignment to the associated biomarker-driven substudy. Detection of more than one eligibility biomarker resulted in the patient being randomly assigned to a substudy with a weighted randomisation procedure that favoured substudies with lower prevalence biomarkers. The randomisation ratio was the inverse ratio of expected prevalence of the biomarkers. Finally, patients with successful biomarker analysis and who were not eligible for any of the biomarker-driven substudies were assigned to a non-match substudy. All patients who did not register to a substudy were followed for survival for up to 3 years.

Substudy eligibility
To be eligible for registration for a substudy, patients had to be assigned to the substudy, and had to have measurable disease, no previous systemic therapy 21 days before substudy registration, and recovery (to grade 1 or better) from any side-effects of previous therapy. Localised palliative radiotherapy was allowed if completed 14 days or more before substudy registration; all other radiation treatment must have been completed 28 days or more before substudy registration. A baseline diagnostic scan (CT or MRI) was required within 28 days of substudy registration, and a brain scan (CT or MRI) for evaluation of CNS disease was required within 42 days of substudy registration (no leptomeningeal disease, spinal cord compression, or untreated or uncontrolled brain metastases allowed unless asymptomatic for at least 14 days following treatment, and patient was off corticosteroids for at least 1 day before substudy registration).

Substudy designs
By use of a modular design for the initiation, conduct, database build, and completion of substudies, each substudy was independently conducted and analysed. The statistical designs were selected from a limited set of design templates. The use of design templates allowed for consistency and efficient development, and streamlined the approvals process both at the NCI and FDA. However, use of the template design as a guide provided the flexibility to tailor designs to the specific goals of individual drug–biomarker combination.

Initially, the substudies were designed as randomised phase 2/3 studies, such that, if successful, they could result in regulatory approval of the biomarker-targeted therapy pairs, and this design was reviewed and approved by the FDA. When nivolumab received regulatory approval for the treatment of previously treated squamous non-small-cell lung cancer, replacing docetaxel as the standard of care, the initial substudies were modified to be single-arm, signal-seeking trials. Since Lung-MAP is set up to provide regulatory-level data, the new pathway for these substudies would be to pursue regulatory approval based on single-arm data if the response rates were sufficiently high and durable, or to initiate a randomised phase 3 study against the current standard of care. As each new substudy was developed, the new substudy protocol was submitted to the FDA as an update to the overarching investigational new drug application for Lung-MAP. The specific pathway for registration for each substudy was discussed with the FDA, and review meetings were solicited if the substudy was intended to be submitted to the agency.

Substudies and evaluated investigational therapies
The following biomarker-driven therapies were evaluated: taselisib, a PI3K inhibitor in *PIK3CA* alteration-positive disease; palbociclib, a CDK4/6 inhibitor in cell cycle gene alteration-positive disease; rilotumumab (a monoclonal antibody directed against MET) plus erlotinib (an EGFR inhibitor) in MET-positive disease; AZD4547, an FGFR inhibitor in *FGFR* alteration-positive disease; talazoparib, a PARP inhibitor for homologous recombination repair deficiency-positive disease; and telisotuzumab vedotin, an antibody–drug conjugate against MET, in MET-positive disease. Non-match substudies evaluated durvalumab and nivolumab plus ipilimumab versus nivolumab monotherapy for anti-PD-1 or anti-PD-L1 naive disease; and durvalumab and tremelimumab for

anti-PD-1 or anti-PD-L1 relapsed or refractory disease. From June 16, 2014, to Dec 18, 2015, evaluation of taselisib, palbociclib, and AZD4547 were compared with docetaxel. From June 16, 2014, to May 26, 2015, evaluation of durvalumab was compared with docetaxel.

Each substudy is published independently, and the substudy manuscripts include detail about the statistical designs implemented for that specific substudy. The studies for taselisib, palbociclib, and AZD4547 have been published.[22–27] The remaining substudies are yet to be published. Additionally, the study team is planning a separate manuscript that describes the statistical under-pinnings of the Lung-MAP studies.

Metrics for the master protocol infrastructure

To evaluate the primary objective of the study, which was to establish an infrastructure that can be used to efficiently evaluate targeted therapies in biomarker subgroups, a series of analyses were done, most of which are descriptive. The infrastructure was evaluated by the number of patients who successfully proceeded through the steps of Lung-MAP. Patients who were pre-screened versus those who were screened at progression were evaluated by comparing survival between the two types of screening, and overall outcomes of the study were evaluated by pooling data from treatment types.

To evaluate whether there was a survival difference by screening type, the Kaplan-Meier method was used to estimate the survival distribution for assigned patients alive at least 2 weeks after substudy assignment. The analyses included the subset alive at least 2 weeks from assignment given that this is the typical window needed to carry out the required tests and procedures to establish patient eligibility. Overall survival, defined as the duration from 2 weeks after assignment to death due to any cause or censored at the date of last contact, was compared using a log-rank test and summarised with a hazard ratio (HR) and 95% CI using a Cox model. A significance level of 5% was used.

To summarise clinical outcomes, the response rate (complete, partial, confirmed or unconfirmed by Response Evaluation Criteria in Solid Tumors [RECIST], version 1.1), progression-free survival, and overall survival distributions were evaluated in an exploratory analysis for the targeted therapies, docetaxel, and anti-PD-1 and anti-PD-L1 therapies (durvalumab, nivo-lumab plus ipilimumab, or nivolumab) for anti-PD-1 or anti-PD-L1-naive disease. For these analyses, progression-free survival was defined as the duration from substudy registration to progression per RECIST (version 1.1) or death due to any cause (whichever came first), and overall survival followed the definition above measured from substudy registration. This analysis includes the patients meeting the eligibility criteria (see substudy publications for eligibility information[22–27]) who received targeted therapy as their first line of therapy within Lung-MAP, eligible patients who received docetaxel within

Lung-MAP, and eligible patients who received anti-PD-1 or anti-PD-L1 therapy for naive disease as their first line of therapy within Lung-MAP. Treatments were not statistically compared because these combine data across substudies. Statistical analyses were done with R (version 1.2.5033).

This study is registered with ClinicalTrials.gov, NCT02154490.

Role of the funding source

The NCI participated in the study design, data interpretation, and writing of the report, but had no role in data collection and data analysis. The pharmaceutical collaborators participated in the substudy designs, data interpretation, and review of the report, but had no role in data collection and data analysis. SWOG was responsible for study design, data collection, data analysis, data interpretation, and writing of the report. MWR, KM, JMi, JMo, and MLL had access to the raw data. All authors had full access to all the data in the study and the corresponding author had final responsibility for the decision to submit for publication.

Results

Lung-MAP (S1400) was open to enrolment between June 16, 2014, and Jan 28, 2019; 436 of approximately 750 sites involved enrolled at least one patient. The median number of patients accrued at a site was two (IQR one to six), with a maximum of 45 patients accrued at a site.

1864 patients with stage IV or recurrent squamous non-small-cell lung cancer were enrolled to the screening component (figure 1). After pre-screening was added, 737 (46·5%) of the 1585 total patients screened were enrolled using the pre-screening option. Of the 1864 patients enrolled, 1841 (98·8%) had tissue submitted, and NGS testing was successful for 1592 (86·5%) of 1841 of the initial submissions. Of the 249 specimens that failed initial sequencing, 98 (39·4%) patients had tissue resubmitted (one or more times) for biomarker profiling and of them, repeated biomarker testing was successful for 82 (83·7%); overall, 1674 (90·9%) of 1841 patients had biomarker results. The reasons for failure of testing are detailed in the appendix (p 6). Of 81 patients in the pre-screened group and 109 patients in the screened at progression group without biomarker results, 12 (14·8%) in the pre-screened group and 11 (10·1%) in the screened at progression group did not have tissue submitted and 69 (85·2%) in the pre-screened group and 98 (89·9%) in the screened at progression group had tissue that was either insufficient for biomarker profiling or failed sequencing. Of the 1941 total tissue submissions (among 1841 patients with tissue), 527 (27·2%) were fresh tissue biopsy specimens. Biomarker results were reported within 16 days from tissue submission for 1462 (79·4%) of 1841 patients, and the median time was 14 days (IQR 11–16).
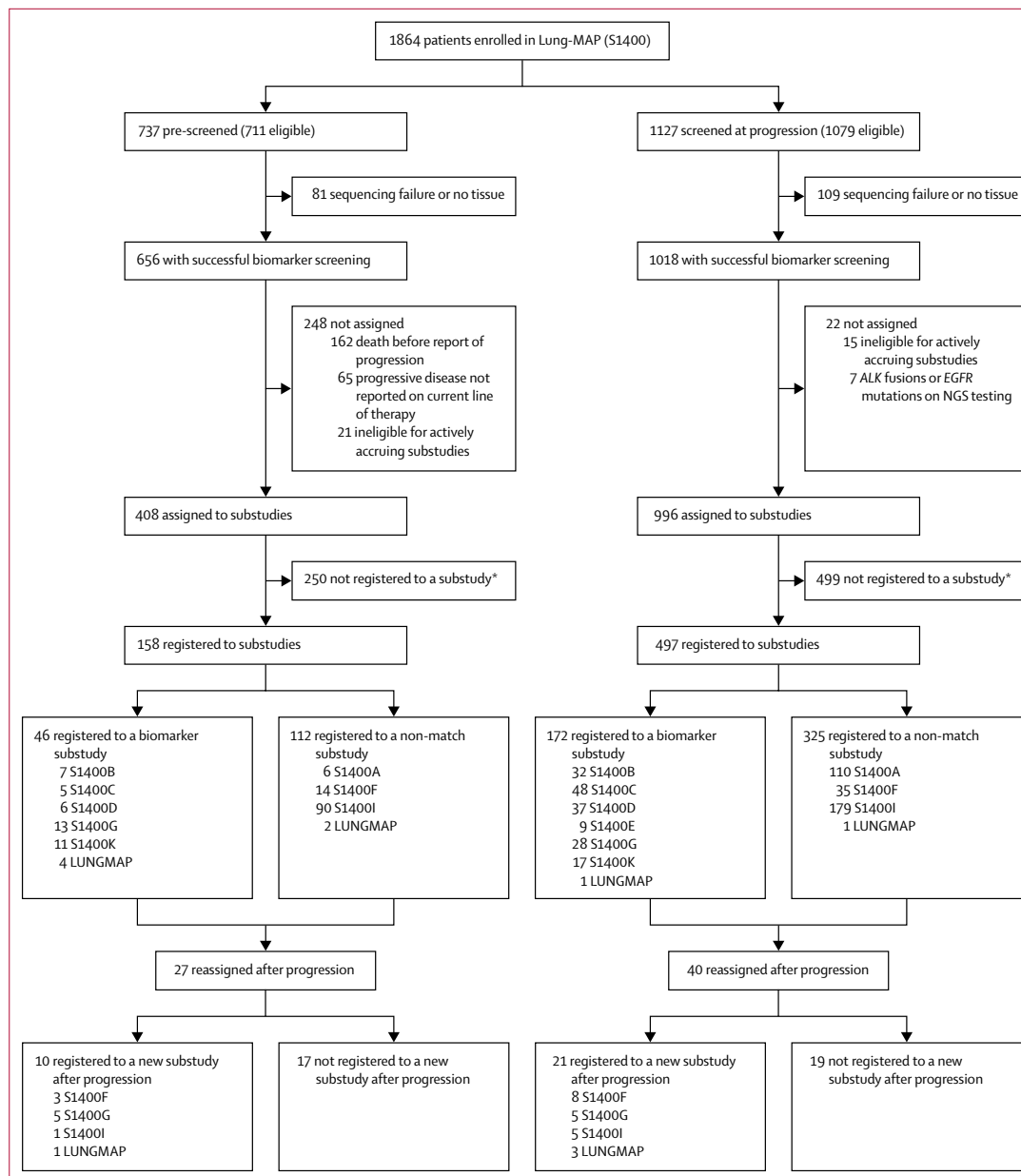
Figure 1: Lung-MAP (S1400) screening registrations, assignments, and substudy registrations summary
The S1400 protocol was active between June 16, 2014, and Jan 28, 2019 (see table 1 and the appendix pp 2–3 for details of the substudies). Simultaneous with the closure of S1400, a new screening protocol called LUNGMAP was opened (on Jan 28, 2019). Accrual to substudies activated under LUNGMAP are listed under that name. HRRD=homologous recombination repair deficiency. Lung-MAP=Lung Cancer Master Protocol. NGS=next-generation sequencing. *See appendix (p 1) for reasons for not registering to a substudy.

Of the 1674 patients with biomarker results, 1404 (83·9%) were assigned to a substudy. For assigned patients, the median time from screening registration to substudy assignment was 3·3 months (IQR 1·4–6·7) in the pre-screened group and 0·5 months (0·4–0·5) in the screened at progression group. Reasons why patients were not assigned to a substudy are shown in figure 1 and the appendix (p 1).

Of the 1404 patients assigned to a substudy, 655 (46·7%) were registered to a substudy (158 [38·7%] of 408 in the pre-screened group and 497 [49·9%] of 996 in the screened at progression group). For patients registered to

	Biomarker or population	Therapies	Design	Sample size goal (number of eligible patients)	Primary endpoint	Study outcome	Dates of activity (duration)
S1400A	Non-match, anti-PD-1 or anti-PD-L1 naive	Durvalumab (investigational) vs docatexel (standard of care)	Originally phase 2/3; modified to single-arm phase 2	Originally 400; modified to 100 total, including 30 PD-L1 high	Originally progression-free survival and overall survival; modified to response*	Administratively closed before completion of accrual	June 16, 2014, to Dec 18, 2015 (18 months); standard of care group closed on May 26, 2015
S1400B	PI3KCA alteration determined by NGS (Foundation One assay [Foundation Medicine, Cambridge, MA, USA])	Taselisib (investigational) vs docetaxel (standard of care)	Originally phase 2/3; modified to single-arm phase 2	Originally 400; modified to 59 total, 40 in primary analysis population	Originally progression-free survival and overall survival; modified to response*	Closed at interim analysis for futility	June 16, 2014, to Dec 12, 2016 (30 months); standard of care group closed on Dec 18, 2015
S1400C	Cell cycle gene alterations determined by NGS (Foundation One assay)	Palbociclib (investigational) vs docetaxel (standard of care)	Originally phase 2/3; modified to single-arm phase 2	Originally 312; modified to 40	Originally progression-free survival and overall survival; modified to response*	Closed at interim analysis for futility	June 16, 2014, to Jan 9, 2016 (27 months); standard of care group closed on Dec 18, 2015
S1400D	FGFR alteration determined by NGS (Foundation One assay)	AZD4547 (investigational) vs docetaxel (standard of care)	Originally phase 2/3; modified to single-arm phase 2	Originally 302; modified to 40	Originally progression-free survival and overall survival; modified to response*	Closed at interim analysis for futility	June 16, 2014, to Oct 31, 2016 (29 months); standard of care group closed on Dec 18, 2015
S1400E	MET determined by immunohistochemistry (Dako [Carpinteria, CA, USA] MET immunohistochemotry PharmDx Kit)	Rilotumumab plus erlotinib (investigational) vs erlotinib (standard of care)	Phase 2/3	326	Progression-free survival and overall survival	Closed due to discontinuation of development of rilotumumab	June 16, 2014, to Nov 25, 2014
S1400F	Non-match, anti-PD-1 or anti-PD-L1 relapsed or refractory	Durvalumab plus tremelimumab	Single-arm phase 2	60 per cohort	Response*	Acquired resistance: closed at interim analysis for futility; primary resistance: passed first interim analysis, closed due to changes in standard of care treatment and feasibility	Acquired resistance: Feb 10, 2017, to Nov 6, 2019 (25 months); primary resistance: Oct 2, 2017, to March 24, 2020 (30 months)
S1400G	Homologous recombination repair deficiency genes determined by NGS (Foundation One assay)	Talazoparib	Single-arm phase 2	60 total, 40 in primary analysis population	Response*	Response*	Feb 7, 2017, to July 23, 2018 (17 months)
S1400I	Non-match, anti-PD-1 or anti-PD-L1 naive	Nivolumab plus ipilimumab (investigational) vs nivolumab (standard of care)	Phase 3	332	Overall survival	Closed at interim analysis for futility	Dec 18, 2015, to April 23, 2018 (28 months)
S1400K	MET determined by immunohistochemistry (Ventana [Tuscon, AZ, USA] Rabbit SP44 Antibody MET assay)	Telisotuzumab vedotin	Single-arm phase 2	40	Response*	Closed at interim analysis for futility	Feb 5, 2018, to Dec 21, 2018 (11 months)

NGS=next-generation sequencing. *Confirmed or unconfirmed complete or partial response by Response Evaluation Criteria in Solid Tumors, version 1.1.

Table 1: S1400 substudy design and implementation details

a substudy, the median time from screening to substudy registration was 3·5 months (IQR 2·0–6·0) in the pre-screened group and 0·9 months (0·7–1·1) in the screened at progression group. Reasons for non-registrations are summarised in the appendix (p 1). The major reason for lack of participation in a substudy was a period when the number of biomarker-driven substudies was small (from Dec 12, 2016, to Oct 2, 2017) and no non-match substudies were available for patients previously treated with an anti-PD-1 or anti-PD-L1 therapy (figure 2). Consequently, 104 (41·6%) of 250 patients in the pre-screened group and 234 (46·9%) of 499 patients in the screened at

progression group did not register to a substudy because there were no available substudies. Less frequent reasons included rapidly progressing disease (ie, symptomatic deterioration or death before registration; 55 [22·0%] of 250 in the pre-screened group and 116 [23·3%] of 499 in the screened at progression group) and patient refusal or investigator decision (51 [20·4%] of 250 in the pre-screened group and 93 [18·6%] of 499 in the screened at progression group).

The characteristics of eligible patients enrolled in the screening component are described in table 2. Of the 1404 patients with substudy assignments, 1317 (93·8%)
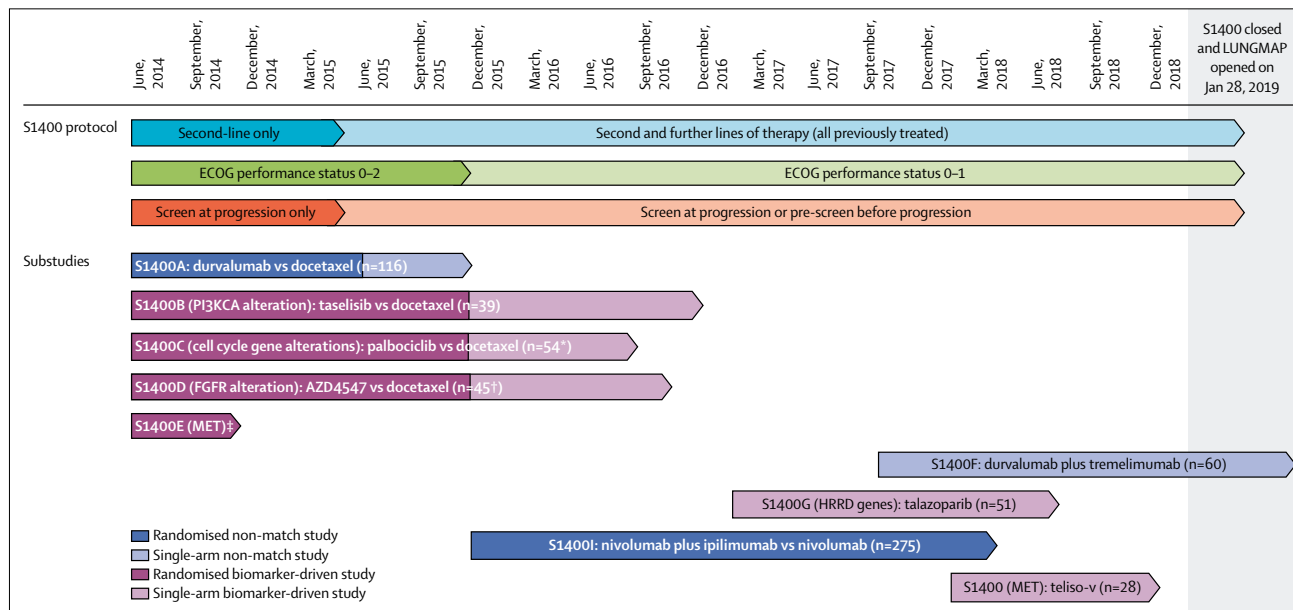
Figure 2: Timelines for substudies, activation, design, and eligibility changes for Lung-MAP (S1400)
See table 1 and the appendix (pp 2–3) for details of the substudies. ECOG=Eastern Cooperative Oncology Group. HRRD=homologous recombination repair deficiency. Lung-MAP=Lung Cancer Master Protocol. Teliso-v=telisotuzumab vedotin. *One patient who was originally randomly assigned to receive docetaxel reregistered to receive palbociclib after that group was closed. †Two patients who were originally randomly assigned to receive docetaxel reregistered to receive AZD4745 after that group was closed. ‡Rilotumumab plus erlotinib vs erlotinib (n=9).

were alive at least 2 weeks from assignment (379 [92·9%] of 408 in the pre-screened and 938 [94·2%] of 996 in the screened at progression group); median follow-up among the subset of 194 patients who were alive at last contact was 27 months (IQR 14–36). Of the 379 patients in the pre-screened group, 308 died, and median overall survival was 8·3 months (95% CI 7·2–9·5). Of the 938 patients in the screened at progression group, 814 died, and median overall survival was 7·0 months (95% CI 6·4–7·6). Patients in the screened at progression group did not have shorter overall survival than those in the pre-screened group (HR 1·10 [95% CI 0·97–1·26], p=0·15; figure 3).

655 patients registered to a substudy (218 to biomarker driven and 437 to non-match), and 31 patients registered to a new substudy after progression on a previous substudy (figure 1). Nine substudies (three non-match and six biomarker driven) were activated, and their timelines are depicted in figure 2. Five biomarker-driven substudies were closed at their interim analysis because of insufficient activity (with approximately 20 patients evaluable for response in the primary analysis population in each substudy), and one biomarker-driven substudy was closed due to discontinuation of development of the drug (rilotumumab; table 1). Of the non-match substudies, the S1400A study of durvalumab was closed to accrual to open the S1400I study of nivolumab plus ipilimumab, which closed at an interim analysis for futility (at 50% maturity for overall survival data); in S1400F, the acquired resistance cohort of durvalumab plus tremelimumab

was closed at the interim analysis for futility (at 20 patients evaluable for response) and the primary resistance cohort was closed early because of feasibility concerns (the study eligibility required single-agent checkpoint inhibitor therapy as the most recent line of therapy preceded by platinum-based chemotherapy). In contrast to when the study was conceived of, very few patients received this sequence of therapies at the time of study closure. A summary of the scientific justification for the substudies is included in the appendix (pp 2–4).

Across the substudies, 143 eligible patients were registered to receive a molecularly targeted treatment as their first line of therapy within Lung-MAP, 56 eligible patients were randomly assigned to docetaxel before closure of those treatment groups, and 315 anti-PD-1 or anti-PD-L1 naive eligible patients were registered to receive an anti-PD-1 or anti-PD-L1 therapy (durvalumab, nivolumab, or nivolumab plus ipilimumab) as their first line of therapy within Lung-MAP. Ten (7·0%) of 143 patients responded to molecularly targeted therapy, three (5·4%) of 56 responded to docetaxel, and 53 (16·8%) of 315 responded to anti-PD-1 or anti-PD-L1 therapy in the immune checkpoint inhibitor-naive setting. Figure 4 depicts progression-free survival and overall survival from substudy registration by treatment group, with a median follow-up among the 81 patients still alive (four in the docetaxel groups, 60 in the immunotherapy groups, and 17 in the targeted therapy groups) of 32 months (IQR 25–37). 126 patients in the molecularly targeted therapy groups, 52 in the docetaxel

	Pre-screened group* (n=711)	Screened at progression group (n=1079)	Total (n=1790)
Age, years	67 (61–73; 39–90)	67 (60–73; 23–92)	67 (61–73; 23–92)
Sex			
Male	466 (66%)	742 (69%)	1208 (67%)
Female	245 (34%)	337 (31%)	582 (33%)
Race			
White	598 (84%)	919 (85%)	1517 (85%)
Black	74 (10%)	95 (9%)	169 (9%)
Asian	15 (2%)	27 (3%)	42 (2%)
Native American	5 (1%)	9 (1%)	14 (1%)
Pacific Islander	4 (1%)	3 (<1%)	7 (<1%)
Multiracial	0	3 (<1%)	3 (<1%)
Unknown	15 (2%)	23 (2%)	38 (2%)
Hispanic	16 (2%)	30 (3%)	46 (3%)
ECOG performance status†			
0	199 (28%)	275 (25%)	474 (26%)
1	499 (70%)	743 (69%)	1242 (69%)
2	13 (2%)	61 (6%)	74 (4%)
Tobacco smoking history			
Current	243 (34%)	375 (35%)	618 (35%)
Former	437 (61%)	665 (62%)	1102 (62%)
Never	31 (4%)	39 (4%)	70 (4%)
Previous lines of treatment for stage IV or recurrent disease‡			
0 or 1	521 (73%)	886 (82%)	1407 (79%)
2	132 (19%)	122 (11%)	254 (14%)
3 or more	58 (8%)	71 (7%)	129 (7%)
Weight loss in past 6 months			
<5% or weight gain	513 (72%)	759 (70%)	1272 (71%)
5–9%	103 (14%)	195 (18%)	298 (17%)
10–19%	81 (11%)	110 (10%)	191 (11%)
≥20%	12 (2%)	10 (1%)	22 (1%)
Unknown	2 (<1%)	5 (<1%)	7 (<1%)
			(Table 2 continues in next column)

	Pre-screened group* (n=711)	Screened at progression group (n=1079)	Total (n=1790)
(Continued from previous column)			
NCTN group§			
SWOG	322 (45%)	567 (53%)	889 (50%)
Alliance	167 (23%)	209 (19%)	376 (21%)
ECOG-ACRIN	144 (20%)	170 (16%)	314 (18%)
NRG Oncology	72 (10%)	121 (11%)	193 (11%)
CCTG	6 (1%)	12 (1%)	18 (1%)
Type of site			
Community (NCORP)	262 (37%)	424 (39%)	686 (38%)
Member	237 (33%)	367 (34%)	604 (34%)
Lead academic	212 (30%)	288 (27%)	500 (28%)

Data are median (IQR; range) or n (%). Percentages might not sum to 100% because of rounding. ACRIN=American College of Radiology Imaging Network. CCTG=Canadian Cancer Trials Group. ECOG=Eastern Cooperative Oncology Group. NCI=US National Cancer Institute. NCORP=NCI Community Oncology Research Program. NCTN=National Clinical Trials Network. *Pre-screening was added to the study on May 26, 2015, 11 months after activation. †Before Dec 18, 2015, a performance status of 0–2 was allowed, but after then only a performance status of 0–1 was allowed. ‡Before May 26, 2015, the study was a second line-only trial, allowing only patients with one line of previous therapy for stage IV or recurrent disease to be screened. §The CCTG participated between Dec 18, 2015, and July 12, 2018, and closed the study because of challenges with drug distribution across the US–Canadian border.

Table 2: Characteristics of eligible patients enrolled to be screened on Lung-MAP (S1400)

groups, and 255 in the anti-PD-1 and anti-PD-L1 groups had died at data cutoff. Median overall survival was 5·9 months (95% CI 4·8–7·8) for the targeted therapy groups, 7·7 months (6·7–9·2) for the docetaxel groups, and 10·8 months (9·4–12·3) for the anti-PD-1 or anti-PD-L1-containing groups. At data cutoff, 139 patients in the molecularly targeted therapy groups, 56 in the docetaxel groups, and 296 in the anti-PD-1 and anti-PD-L1 groups had disease progression or had died. Median progression-free survival was 2·5 months (95% CI 1·7–2·8) for the targeted therapy groups, 2·7 months (1·9–2·9) for the docetaxel groups, and 3·0 months (2·7–3·9) for the anti-PD-1 or anti-PD-L1-containing groups.

Discussion

With Lung-MAP (S1400), we successfully established an infrastructure to evaluate molecularly targeted therapies in genomically defined subgroups of lung cancer and implemented the first biomarker-driven master protocol within the NCI. We demonstrated that comprehensive, centralised genomic screening using a commercially available NGS platform is feasible in a diverse patient population; biomarker-driven master protocols conducted within the NCTN are attractive to industry partners; a series of biomarker-driven studies can be done simultaneously, even in rare genotypes, to efficiently assess drug activity; the non-match option was essential; and this approach is amenable to a drug registration-compliant strategy. Moreover, as part of the precision medicine initiative within the NCI, the NCTN provided far-reaching patient access with the largest percentage of accrual coming from community sites or the veterans affairs system. The public–private partnership that engaged people with broad experience to carry out a complicated study infrastructure was essential. Most importantly, we addressed research questions that might not have otherwise been answered and included patient populations who might not have otherwise been able to participate in investigational studies of biomarker-driven therapies.

Master protocols present both opportunities and challenges. Conduct of a complex biomarker-driven master protocol requires adaptability as well as constant and intensive efforts by all partners. A comprehensive communications plan included biweekly leadership
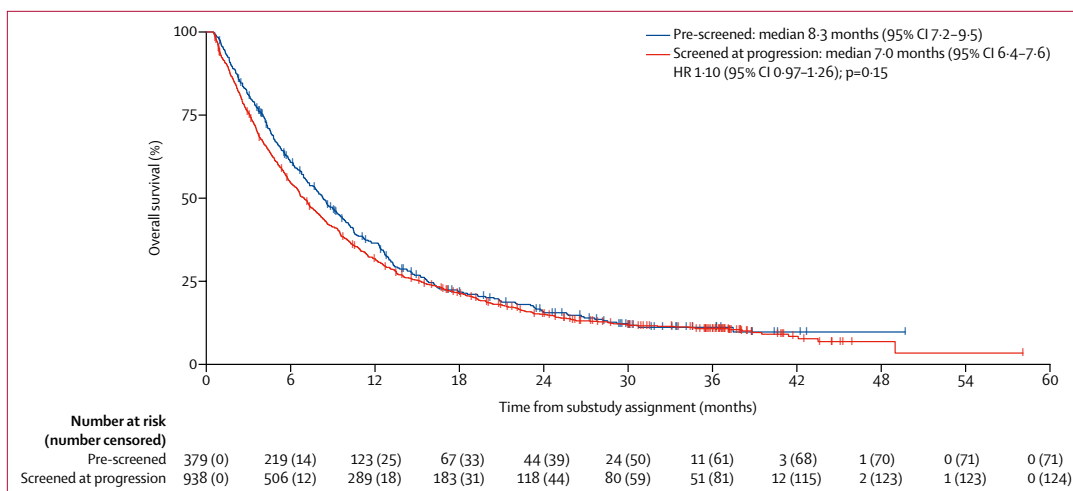
Figure 3: Overall survival from time of substudy assignment by screening type
HR=hazard ratio.

teleconferences and monthly teleconferences for each of the subcommittees (drug selection; substudy chairs; site coordinators; accrual enhancement; statistical, data management, information technology, and protocol operations). Additional oversight, staff support, and operational efficiencies were needed to shepherd substudies through their entire life cycle. Extensive outreach to investigators, site coordinators, and patients through newsletters, study materials, videos, webinars, presentations at scientific conferences, and NCTN group meetings, established and maintained engagement in the study. A site coordinators committee addressed challenges of implementation and informed study conduct. An accrual enhancement committee developed and maintained educational and outreach materials.

Inclusion of the pre-screening option was an attractive option to both patients and physicians, allowing for a patient and their physician to determine the optimal time for that patient to be screened. Notably, the comparison of overall survival from substudy assignment (pre-screening versus screening at progression) does not capture all of the benefit of pre-screening. With pre-screening, sites have ample time to locate and evaluate tissue for adequacy or to order a fresh biopsy, as needed. This approach allows for time to do the necessary tests and procedures to evaluate the patient for eligibility criteria, and it all can be done in a less stressful time when the patient is receiving therapy versus having just learned that their previous therapy was no longer effective and their disease has progressed. In terms of study conduct, it should be noted that proportionally fewer pre-screened patients than those screened at progression enrolled to their assigned substudy, which increases the relative per patient costs of screening this population.

Although the non-match substudies would probably have been more efficiently conducted had they been run independently, the non-match substudy option has been essential to Lung-MAP (S1400). Importantly, this option makes the biomarker-matched substudies feasible. In this aggressive disease setting, it was essential that all screened patients had an option for participation in a substudy even if they did not have one of the matching biomarkers. That said, inclusion of non-match substudies in Lung-MAP (S1400) have additional value independent of their role in facilitating the biomarker-driven studies. A benefit to evaluation of non-match therapies within Lung-MAP (S1400) is that all patients have the full NGS results, allowing for retrospective assessment of potential biomarkers for treatment activity. The ultimate goal is to translate all studies in unselected populations into biomarker-selected subgroups.

A potential challenge for the evaluation of therapies within the non-match setting is the definition of the patient population. By definition, the non-match studies include patients who are not eligible for one of the biomarker-driven substudies. This population can vary during the conduct of the study based on accruing biomarker-driven substudies, but is also probably not a natural definition of a population. If there is no clear reason that the excluded biomarker subgroups would benefit differentially from the non-match therapy, an argument could be made for assuming that the results apply to these populations. However, regulatory agencies might find it challenging to define the label, and conversations are certainly needed upfront to discuss the subtleties and interpretation of the data at the end of study, if being used for regulatory approval.

A major criticism of Lung-MAP (S1400) is that the investigated targeted therapies did not demonstrate
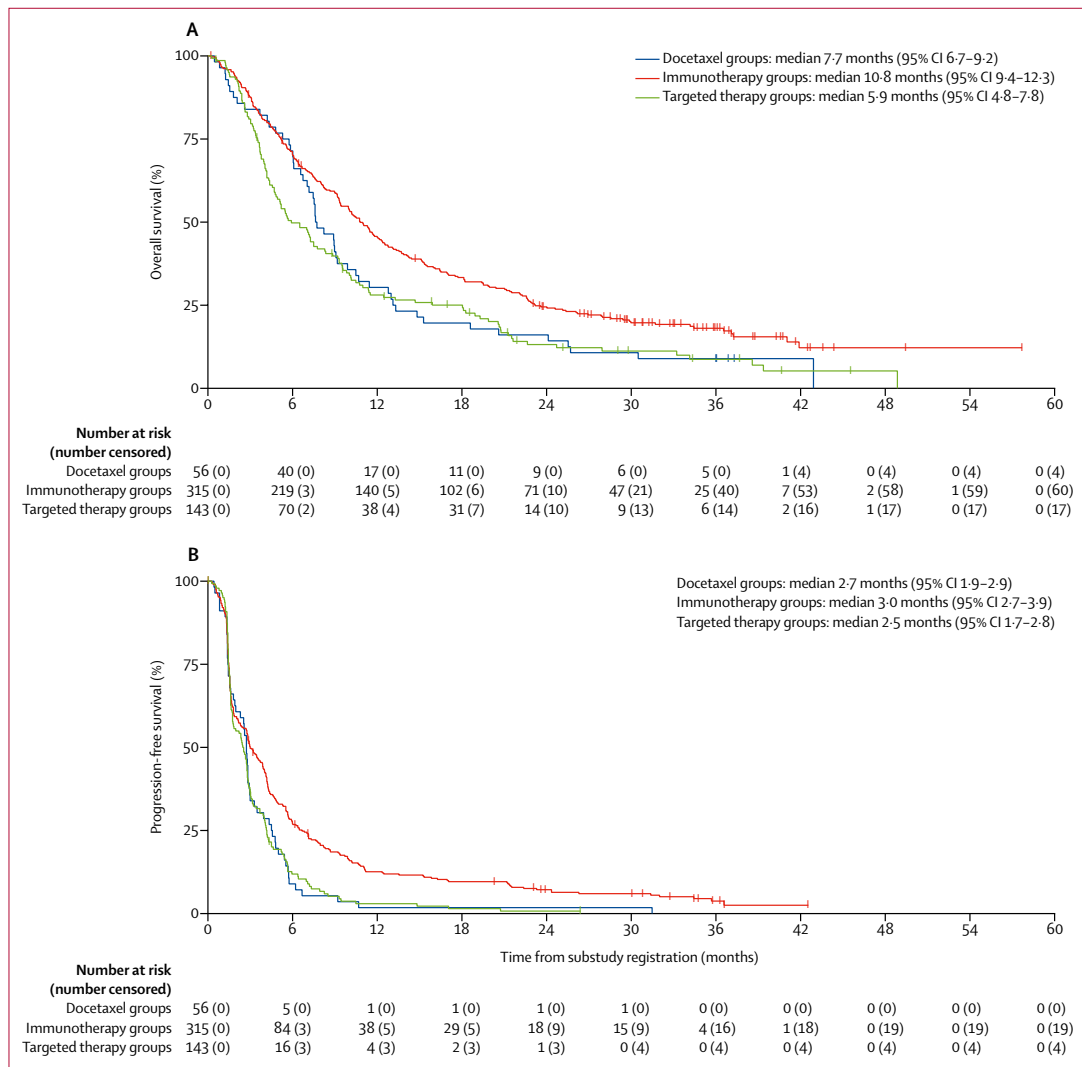
A

Docetaxel groups: median 7·7 months (95% CI 6·7–9·2)
Immunotherapy groups: median 10·8 months (95% CI 9·4–12·3)
Targeted therapy groups: median 5·9 months (95% CI 4·8–7·8)

Overall survival (%)

Number at risk (number censored)										
Docetaxel groups	56 (0)	40 (0)	17 (0)	11 (0)	9 (0)	6 (0)	5 (0)	1 (4)	0 (4)	0 (4)	0 (4)
Immunotherapy groups	315 (0)	219 (3)	140 (5)	102 (6)	71 (10)	47 (21)	25 (40)	7 (53)	2 (58)	1 (59)	0 (60)
Targeted therapy groups	143 (0)	70 (2)	38 (4)	31 (7)	14 (10)	9 (13)	6 (14)	2 (16)	1 (17)	0 (17)	0 (17)

B

Docetaxel groups: median 2·7 months (95% CI 1·9–2·9)
Immunotherapy groups: median 3·0 months (95% CI 2·7–3·9)
Targeted therapy groups: median 2·5 months (95% CI 1·7–2·8)

Progression-free survival (%)

Time from substudy registration (months)

Number at risk (number censored)										
Docetaxel groups	56 (0)	5 (0)	1 (0)	1 (0)	1 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Immunotherapy groups	315 (0)	84 (3)	38 (5)	29 (5)	18 (9)	15 (9)	4 (16)	1 (18)	0 (19)	0 (19)	0 (19)
Targeted therapy groups	143 (0)	16 (3)	4 (3)	2 (3)	1 (3)	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)

Figure 4: Overall survival (A) and progression-free survival (B) from the time of substudy registration by treatment type

activity. We expected that squamous non-small-cell lung cancer would be a particularly challenging disease setting given its genomic complexity and tumour mutational burden.[28] Concurrent with evaluation in Lung-MAP (S1400), all evaluated drug targets were under study in many smaller trials. Additionally, all targets, with the exception of MET, received FDA approval for drugs in class, subsequent to evaluation within Lung-MAP (S1400). Lung-MAP (S1400) allowed for a more definitive statement regarding the use of these drug targets in lung cancer—something that would have been much more challenging without the master protocol. Therefore, although Lung-MAP (S1400) has not been successful in establishing new treatments, it has also prevented prolonged evaluation of ineffective drug targets in lung cancer.

Lung-MAP (S1400) has limitations. The Foundation Medicine NGS platform might not be the best biomarker for all therapies, and bridging studies might be required if a therapy is to be submitted for regulatory approval. Since all biomarker studies are within biomarker-selected populations, the studies cannot differentiate between prognostic and predictive biomarkers, necessitating data from outside the trial to support a predictive biomarker interpretation. By far the greatest challenge encountered in Lung-MAP (S1400) was the rapid approval of anti-PD-1 and anti-PD-L1 therapies across various indications in squamous non-small-cell lung cancer in March, 2015.[29] In response, and crucial to the success of a biomarker-driven master protocol, Lung-MAP (S1400) efficiently implemented major revisions, which included

changes in eligibility, statistical designs, and the addition of an option to pre-screen patients during their first-line treatment. The changes implemented in Lung-MAP (S1400) demonstrated how a biomarker-driven master protocol can be self-sustaining and adaptable to scientific advancements. It follows that the most important lesson learned during the conduct of Lung-MAP (S1400) was that biomarker-driven master protocols must be nimble and incorporate new science quickly. To do so requires a continuous stream of collaborations with the pharmaceutical industry and an active and engaged group of academic investigators for the continual development of new substudies.

Lung-MAP (S1400) is not alone in the master protocol space, although it was among the first. For example, within the NCTN, the ALCHEMIST trial[17] is evaluating adjuvant targeted therapy in biomarker-defined subgroups of lung cancer for therapies known to be effective in the advanced setting, and the NCI-MATCH trial[14,17] evaluated targeted therapy–biomarker pairs in a histology-agnostic approach. The FOCUS4 trial[19] evaluated molecularly targeted therapies in colorectal cancer, and the National Lung Matrix trial[13] is evaluating targeted therapies for non-small-cell lung cancer. The SHIVA[20] and MOSCATO 01[15] studies evaluated the strategy of assigning treatment based on molecular alterations. Each of these master protocols have a slightly different design and set of objectives.[12–20] Relative to other master protocols being used in lung cancer, the Lung-MAP (S1400) trial was created to evaluate signals of activity but also to be a pathway for regulatory approval by the FDA of targeted therapy–biomarker pairs. Our approach for designs for regulatory approval has been US centric; although standard statistical designs were used within Lung-MAP (S1400), other regulatory bodies might view these data differently.

Although the Lung-MAP (S1400) screening protocol and its associated substudies are completed, the Lung-MAP study continues. In early 2019, a new screening protocol (named LUNGMAP) was implemented, expanding eligibility to all histological types of non-small-cell lung cancer. Additionally, the new screening protocol implemented circulating tumour DNA (ctDNA) testing for the subset of patients using a fresh tissue biopsy for screening with the goal to evaluate whether ctDNA could be used as part of the upfront biomarker screening on Lung-MAP. All new substudies include ctDNA evaluation at baseline and on-study timepoints. Lung-MAP (S1400) patients are allowed to participate in our new studies. With this change, the primary pursuit of Lung-MAP (S1400) to evaluate molecularly targeted therapies in biomarker-defined subgroups was augmented to include the pursuit of immunotherapy combinations to overcome resistance to anti-PD-1 or anti-PD-L1 therapy. Motivation for these changes was twofold: first, an infrastructure to evaluate targeted therapies for rare alterations in previously treated non-squamous non-small-cell lung cancer was needed, and second, a clear unmet need in lung cancer is treatment for patients with disease that has relapsed after, or is refractory to, anti-PD-1 or anti-PD-L1 therapy. This new protocol goes beyond histology to evaluate targeted and immunotherapies in non-small-cell lung cancer. With multiple substudies currently accruing and a robust pipeline of regimens in development, Lung-MAP remains true to its original vision to expeditiously improve treatment options for patients with lung cancer.

Contributors

MWR, VAP, FRH, PCM, EV, SR, NL, JB, JMi, JMo, CM, MLL, VAM, EVS, DW, CS, CDB, DRG, and RSH contributed to the study design. MWR, VAP, KM, LHS, SR, NL, JMi, JMo, LH, CM, MLL, DRG, and RSH collected data. MWR, VAP, KM, FRH, PCM, LHS, JMi, JMo, MLL, CS, CDB, KK, DRG, and RSH analysed the data. MWR, VAP, FRH, LH, CM, SM, SA, CDB, DRG, and RSH were responsible for study conduct. MWR, VAP, KM, PCM, LHS, EV, SR, NL, JB, JMi, JMo, MLL, CS, CDB, KK, DRG, and RSH interpreted the data. MWR, VAP, KM, LHS, EV, NL, JB, JM, JM, MLL, CSS, BS, CDB, KK, DRG, and RSH were involved in writing of the report. MWR and KM were responsible for the figures and tables. MWR, NL, VAM, SA, CS, CSS, BS, KK, DRG, and RSH reviewed the manuscript and approved the final version.

References

1. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 2018; **553:** 446–54.

2. Roskoski R Jr. Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacol Res* 2019; **144:** 19–50.

3. Morgensztern D, Waqar S, Subramanian J, Gao F, Govindan R. Improving survival for stage IV non-small cell lung cancer: a surveillance, epidemiology, and end results survey from 1990 to 2005. *J Thorac Oncol* 2009; **4:** 1524–29.

4. Gandara DR, Hammerman PS, Sos ML, Lara PN Jr, Hirsch FR. Squamous cell lung cancer: from tumor genomics to cancer therapeutics. *Clin Cancer Res* 2015; **21:** 2236–43.

5. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012; **489:** 519–25.

6. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* 2013; **31:** 1023–31.

7. Herbst RS, Gandara DR, Hirsch FR, et al. Lung Master Protocol (Lung-MAP)—a biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S1400. *Clin Cancer Res* 2015; **21:** 1514–24.

8. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 2017; **377:** 62–70.

9. Redman MW, Allegra CJ. The master protocol concept. *Semin Oncol* 2015; **42:** 724–30.

10. Cecchini M, Rubin EH, Blumenthal GM, et al. Challenges with novel clinical trial designs: master protocols. *Clin Cancer Res* 2019; **25:** 2049–57.

11. Malik SM, Pazdur R, Abrams JS, et al. Consensus report of a joint NCI thoracic malignancies steering committee: FDA workshop on strategies for integrating biomarkers into clinical development of new therapies for lung cancer leading to the inception of "master protocols" in lung cancer. *J Thorac Oncol* 2014; **9:** 1443–48.

12. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* 2011; **1:** 44–53.

13. Middleton G, Crack LR, Popat S, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. *Ann Oncol* 2015; **26:** 2464–69.

14. Coyne GO, Takebe N, Chen AP. Defining precision: the precision medicine initiative trials NCI-MPACT and NCI-MATCH. *Curr Probl Cancer* 2017; **41:** 182–93.

15. Massard C, Michiels S, Ferté C, et al. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the MOSCATO 01 trial. *Cancer Discov* 2017; **7:** 586–95.

16. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 2009; **86:** 97–100.

17. Chen AP, Eljanne M, Harris L, Malik S, Seibel NL. National Cancer Institute basket/umbrella clinical trials: MATCH, Lung-MAP, and beyond. *Cancer J* 2019; **25:** 272–81.

18. Stockley TL, Oza AM, Berman HK, et al. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med* 2016; **8:** 109.

19. Adams R, Brown E, Brown L, et al. Inhibition of EGFR, HER2, and HER3 signalling in patients with colorectal cancer wild-type for BRAF, PIK3CA, KRAS, and NRAS (FOCUS4-D): a phase 2–3 randomised trial. *Lancet Gastroenterol Hepatol* 2018; **3:** 162–71.

20. Le Tourneau C, Paoletti X, Servant N, et al. Randomised proof-of-concept phase II trial comparing targeted therapy based on tumour molecular profiling vs conventional therapy in patients with refractory cancer: results of the feasibility part of the SHIVA trial. *Br J Cancer* 2014; **111:** 17–24.

21. Foundation Medicine. Foundation One CDx. 2017. https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019C.pdf (accessed Nov 30, 2017).

22. Aggarwal C, Redman MW, Lara PN Jr, et al. SWOG S1400D (NCT02965378), a phase II study of the fibroblast growth factor receptor inhibitor AZD4547 in previously treated patients with fibroblast growth factor pathway-activated stage IV squamous cell lung cancer (Lung-MAP substudy). *J Thorac Oncol* 2019; **14:** 1847–52.

23. Bazhenova L, Redman MW, Gettinger SN, et al. A phase III randomized study of nivolumab plus ipilimumab versus nivolumab for previously treated patients with stage IV squamous cell lung cancer and no matching biomarker (Lung-MAP Sub-Study S1400I, NCT02785952). *Proc Am Soc Clin Oncol* 2019; **37** (suppl): 9014 (abstr).

24. Edelman MJ, Redman MW, Albain KS, et al. SWOG S1400C (NCT02154490)—a phase II study of palbociclib for previously treated cell cycle gene alteration-positive patients with stage IV squamous cell lung cancer (Lung-MAP substudy). *J Thorac Oncol* 2019; **14:** 1853–59.

25. Langer CJ, Redman MW, Wade JL 3rd, et al. SWOG S1400B (NCT02785913), a phase II study of GDC-0032 (taselisib) for previously treated PI3K-positive patients with stage IV squamous cell lung cancer (Lung-MAP sub-study). *J Thorac Oncol* 2019; **14:** 1839–46.

26. Owonikoko TK, Redman MW, Byers LA, et al. A phase II study of talazoparib (BMN 673) in patients with homologous recombination repair deficiency (HRRD) positive stage IV squamous cell lung cancer (Lung-MAP Sub-Study, S1400G). *Proc Am Soc Clin Oncol* 2019; **37** (suppl): 9022 (abstr).

27. Waqar SN, Redman MW, Arnold SM, et al. Phase II study of ABBV-399 (Process II) in patients with C-MET positive stage IV/recurrent lung squamous cell cancer (SCC): LUNG-MAP sub-study S1400K (NCT03574753). *Proc Am Soc Clin Oncol* 2019; **37** (suppl): 9075 (abstr).

28. Willis C, Fiander M, Tran D, et al. Tumor mutational burden in lung cancer: a systematic literature review. *Oncotarget* 2019; **10:** 6604–22.

29. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med* 2015; **373:** 123–35.

FRIENDS of CANCER RESEARCH

A FRIENDS OF CANCER RESEARCH WHITE PAPER

Modernizing Expedited Development Programs

FRIENDS OF CANCER RESEARCH ANNUAL MEETING 2020

Introduction: Expedited development programs and pathways

Advances in our understanding of disease processes, genetics, manufacturing technologies, and innovative trial designs have enabled the development of novel, effective, and greatly improved therapeutic agents. Particularly in oncology, the ability to target a novel agent against a driver oncogene or protective immune checkpoint has led to several therapeutic breakthroughs in diseases with limited or no systemic treatment options. These breakthroughs have established new classes of therapeutics leading to, in some instances, unprecedented efficacy results for serious, life-threatening diseases. In situations where substantial benefit over existing therapies is observed in early clinical studies addressing unmet need, expedited drug development pathways help balance the need to provide individuals with serious diseases or conditions with expedited access to breakthroughs while also maintaining the rigorous standards established for approving drugs.[1,2]

The US Food and Drug Administration (FDA) currently uses several tools to expedite the development of promising new medicines aimed at treating serious disease with unmet needs. These include the following tools: 1) Fast Track; 2) Breakthrough Therapy; 3) Regenerative Medicine Advanced Therapy (RMAT); 4) Priority Review; and 5) Accelerated Approval.[3,4]

Objectives

▼ **Evaluate current use and application of expedited drug development pathways**

▼ **Recommend proposals to clarify and simplify expedited programs that facilitate the development of promising therapies and address emerging drug development challenges**

▼ **Delineate optimal processes and actions that occur following the initial designation of an expedited development program(s)**

Thank You to Our Working Group Members

Julia Beaver
Chief of Medical Oncology, Oncology Center of Excellence, FDA

Anne-Virginie Eggimann
SVP, Regulatory Science, bluebird bio

Scott Korn
VP, Global Regulatory Affairs, Merck

Bea Lavery
VP, Regulatory Affairs, Genentech

Amy McKee
VP, Regulatory Consulting Services, Parexel

Paul Melmeyer
Director, Regulatory Affairs, Muscular Dystrophy Association

Harpreet Singh
Division Director, Division of Oncology 2, FDA

Kathleen Winson
Senior Group Director, Regulatory Affairs, Genentech

1. **Fast Track** is a process designed to facilitate the development and expedite the review of drugs that treat serious diseases and address unmet medical needs. It entails early and frequent communication between the FDA and sponsor throughout the development and review process. Under this program, a sponsor may submit complete sections of a New Drug Application (NDA) or Biologics License Application (BLA) as they are ready ("rolling review"), rather than the standard requirement to submit the complete NDA or BLA application in one submission.

2. **Breakthrough Therapy** designation expedites the development and review of drugs that are intended to treat a serious condition, and preliminary clinical evidence indicates that the drug may demonstrate substantial improvement over available therapy. A drug with Breakthrough Therapy designation is also eligible for all considerations of the Fast Track designation. In addition, Breakthrough Therapy affords intensive FDA drug development guidance with an FDA organizational commitment with early involvement of senior managers and early manufacturing consultation. An NDA/BLA submission will be provided rolling review with potential for priority review.

3. **RMAT** designation includes all the benefits of the Fast Track and Breakthrough Therapy designation programs, including early interactions with FDA. RMAT designation is granted for advanced therapies (which is defined as a cell and gene therapy, therapeutic tissue engineering product, human cell and tissue product, or any combination product using such therapies or products) intended to treat, modify, reverse, or cure a serious or life-threatening disease or condition and preliminary clinical evidence indicates that the drug has the potential to address unmet medical needs for such a disease or condition. RMAT does not require evidence to indicate that the drug may offer a substantial improvement over existing therapies. Due to the definition of regenerative medicine, these products will be reviewed by the Center for Biologics Evaluation and Research (CBER).

4. **Priority Review** is available to drugs that provide a significant improvement in the treatment, prevention, or diagnosis of a disease when compared to standard NDAs or BLAs. It shortens the goal review time from 10 months to 6 months from the 60-day filing date (or from 12 months to 8 months respectively from date of submission of the application). A Priority Review designation directs attention and resources to evaluate drugs that would significantly improve the treatment, diagnosis, or prevention of serious conditions.

5. **Accelerated Approval** allows a drug to receive FDA approval based on an early efficacy endpoint (such as objective response rate) considered reasonably likely to predict a clinical benefit (such as prolonged survival). Accelerated Approval is a critical pathway for expediting access to new therapies in disease settings in which the effect on an intermediate clinical endpoint that predicts the drug's clinical benefit can be shown much sooner than the effect on an endpoint that directly demonstrates clinical benefit. This pathway is reserved for drugs/biologics that seek to treat a serious or life-threatening disease and that provide meaningful therapeutic benefit to patients over existing treatments. Drugs approved via the Accelerated Approval pathway should undergo further clinical testing to confirm the predicted clinical benefit ("confirmatory trial/clinical evidence"). If the confirmatory trial/

evidence does not show that the drug provides clinical benefit for patients, FDA may seek to remove the drug from the market, or remove the indication from the drug's labeling in cases where the drug is approved for other uses.

Some of these pathways are used throughout the development lifecycle of the drug before the NDA/BLA is submitted (Fast Track, Breakthrough Therapy, and RMAT) while other tools are applied once the license application is submitted (Priority Review and Accelerated Approval). **Table 1** provides a comparison of features associated with these pathways. This white paper prioritizes discussions on the programs intended to be utilized prior to NDA/BLA submission; opportunities to optimize accelerated approval are discussed in the companion white paper "Optimizing the Use of Accelerated Approval."

It is also worth noting that some of the elements associated with these FDA expedited pathways are mirrored by health authorities outside of the United States.[5] For example, this is seen with the European Medicines Agency's (EMA) Priority Medicines (PRIME) program and the Pharmaceuticals and Medical Devices Agency's (PMDA) SAKIGAKE designation, which share characteristics of FDA's Breakthrough Therapy designation.[6–8] In addition, the EMA has several approval frameworks (approval under exceptional circumstances and conditional approval), which allow approval using an intermediate endpoint and are similar, in this respect, to FDA's Accelerated Approval. It is important to note that regional differences can add complexity to global clinical drug development for therapies aimed at treating serious or life-threatening conditions.

With the creation of the FDA's Oncology Center of Excellence (OCE), several pilot projects have successfully launched to test novel approaches to regulatory review for oncology drugs, such as Real-Time Oncology Review (RTOR) and the Assessment Aid.[9,10] The RTOR Pilot Program aims to improve the efficiency of the review process for clinical applications through data and analysis standardization and early iterative engagement between the FDA and applicant by allowing for the submission of key efficacy and safety tables/figures and datasets prior to the complete dossier submission. Eligible applications include oncology NDAs and BLAs for drugs or biologics likely to demonstrate substantial improvements over available therapies (e.g., Breakthrough Therapy, Accelerated Approval, and Priority Review eligible indications) and based on clinical trials with straightforward study designs and easily interpretable endpoints. The Assessment Aid is a unified FDA review document that contains an applicant assessment (submitted at the time of (s)NDA/BLA submission) and an FDA assessment and improves the efficiency of the FDA review. The regulatory review process for pharmaceutical drugs is a resource intensive undertaking for both the drug sponsor and the FDA. Therefore, continued evaluation of current pathways is necessary to ensure pathways facilitate the science and make sense for patients.

Landscape Analysis of Expedited Pathways

Expedited programs at the FDA have been highly utilized by sponsors and with increasing frequency for oncology drugs in the US (**Table 2**). Between 2012–2019, 90% of initial oncology drug approvals utilized an expedited program versus only 55% of new non-oncology drug approvals. Accelerated Approval, Breakthrough Therapy designation, and Priority Review are overwhelm-

ingly used more for oncology products than non-oncology products. While this can be partly attributed to the fact that many non-oncology diseases may not meet criteria for expedited programs if they are not deemed serious or life-threatening, it may also highlight differing approaches across review divisions within FDA.

Priority Review and Fast Track appear to have been the most popular tools, followed by Breakthrough Therapy (**Table 2**), which was available after the Food and Drug Administration Safety and Innovation Act (FDASIA) was signed into law in July 2012, with the first products receiving Breakthrough Therapy designation in January 2013. It is interesting to note that approximately half of the programs with Breakthrough Therapy designation followed the Accelerated Approval pathway.

Value of Expedited Programs Across Disease Areas and at Key Points in Development

As depicted in **Table 3**, these programs are not mutually exclusive and can be used in combination with the other expedited programs. Expedited programs can have different utility within different disease settings. The exact pathway a promising therapy might take will depend on several factors including the disease setting and indication sought, endpoint(s) used, as well as the magnitude and durability of the signal relative to the existing standard of care. The ultimate decision-maker for assigning an expedited pathway to a drug development program is the review division. Therefore, consulting the review division before applying for a respective expedited program is highly recommended. Coordinating the added benefits of these programs should be considered to minimize unnecessary administrative work for the Agency and sponsor. For example, it may not be necessary to apply for both Breakthrough Therapy designation and RMAT since they provide similar opportunities to facilitate development of a promising agent. The highest value for sponsors noted to date in using RMAT or Breakthrough Therapy designation has been the ability to meet with the Agency often.

In a cohort of drugs that utilized Fast Track, Breakthrough Therapy, Priority Review, and Accelerated Approval (n=9), **Figure 1** helps depict the utilization of these programs across the development lifecycle of a drug. The use of Fast Track and Breakthrough Therapy designation often occurs later in the life cycle of a drug development program (several years after IND submission) and close to the time of submitting an NDA/BLA, likely indicative of having greater confidence in the clinical data. However, the benefit of these expedited development programs may be most realized earlier in development and could enable more meaningful interactions on other key aspects of a development program (e.g., chemistry, manufacturing, and controls [CMC], co-development of a diagnostic assay).

Learnings from Current Experience with Expedited Pathways

Expedited development programs at the FDA have had a positive impact on ushering new drugs through clinical development to reach patients more quickly. Drugs that qualified for an expedited program are approved on average two years earlier than drugs not under an expedited program (_Friends_ Drug Development Dashboard).[11] This is, in part, due to development and

appropriate identification of promising drugs, increased dialogue with the FDA, and the positive momentum and collaborative mindset created within companies and at the FDA when a drug development program qualifies for an expedited pathway. While these pathways have been quite successful, cataloguing the learnings from these past experiences can help optimize their use moving forward.

Addressing current unmet need is becoming increasingly challenging. At the time many of these expedited development pathways were designed, treatment options in oncology, for example, consisted primarily of surgery, radiotherapy, and cytotoxic chemotherapy. As the treatment paradigm in oncology has shifted to therapies targeted against specific oncogenic proteins or pathways and immunotherapies, patients' lives have been improved and extended. Nonetheless, most of these newer treatments still are not curative; therefore, despite the availability of new anti-cancer therapies, significant unmet need remains, especially in the setting of metastatic disease. Furthermore, while significant advancements have been made in serious and life-threatening non-oncology conditions, most remain without a treatment to significantly alter the course of the disease. Hence, there is still a need for expedited pathways to facilitate development of promising therapies.

It can be difficult, though, to decipher which program/tool has been or will be the most beneficial in accelerating development to bring the right product to the right patient at the right time. *Is there redundancy in terms of benefits from these expedited programs and how could we either simplify or improve them so that their intrinsic value increases?*

To help start answering these questions, the working group extracted several learnings based on the landscape analysis, sponsor/FDA interactions, and the wealth of experience gained over the past decade through drug approvals.

It is important to coordinate the use and timing of expedited pathways with clinical need and appropriate drug development stage. When creating each expedited development program, significant attention was paid to the eligibility criteria necessary for a new treatment to qualify for each program or designation. This has resulted in numerous potential duplicative application and review processes that the same drug may go through when qualifying for each program. Less attention has been devoted to assessing what occurs following a successful designation or delineating the steps applied to optimally expedite development post-designation for all disciplines (CMC, nonclinical and clinical areas of development). As experience is gained with each expedited development program, it is important to identify the subsequent actions that helped foster successful development so that those approaches can be anticipated and replicated as appropriate in a consistent manner.

Streamlining expedited programs where less redundancy exists can lead to more optimal and successful use within the lifecycle of a drug to avoid confusion as to when they can be used during a development program. Informal assessments revealed that recurring reasons for Breakthrough Therapy designation or RMAT denials included that the application was simply submitted too early or included data from an insufficient number of patients, there were issues with durability of response, or manufacturing concerns existed (for example, when early clini-

cal data were generated with a previous manufacturing process that subsequently changed significantly). Critical elements that can impact a program regardless of how good the clinical data or product are, include: non-oncology safety database issues, clinical site/Good Clinical Practice (GCP) concerns, lack of product stability data, and manufacturing site/Good Manufacturing Practice (GMP) concerns. Codifying processes and best practices for expedited programs could result in more impactful use of the expedited pathways to guide drug development programs through these critical stages of drug development (e.g., manufacturing, clinical pharmacology/toxicology, and clinical development). Later stage components such as manufacturing site inspections, diagnostic test development, or design of potential post-market commitments that may occur later in development could be sufficiently planned for through earlier interactions with the FDA.

Delineating the optimal early stage versus late stage development milestones important for expediting development is critical to help coordinate efforts within the sponsor and across the different teams at the FDA. Breakthrough Therapy designation and RMAT are both helpful to accelerate clinical development but challenges remain in accelerating CMC development particularly for novel therapies using emerging manufacturing technologies. There is an opportunity to utilize a more holistic approach where the FDA provides advice that will help synchronize clinical development and CMC development.

Expedited pathways and associated tools may be most needed for emerging therapies or for complex development programs to increase frequency and depth of interactions with the FDA. This can create a paradoxical scenario where comparatively less-novel products in better understood disease areas receive greater research and development (R&D) investment as there is an increased likelihood of qualifying for an expedited pathway. Consequently, greater investments lead to a better understanding of the disease and established class of products.

An important first step to qualifying for expedited pathways is to establish whether there is an unmet need or urgent public health concern. This helps determine the degree of regulatory flexibility to which novel or atypical regulatory pathways may be leveraged. The level of regulatory flexibility can be impacted by the confidence or how much trust is in the package being brought forward for review. This is driven by the development stage where the drug currently is, and what the biological and clinical evidence is to inform safety and efficacy. However, in a novel space it can be hard to be truly confident. For example, in the 1990s when monoclonal antibodies were entirely novel, the regulatory confidence was very low. However, with increasing numbers of monoclonal antibody therapies being developed, approved, and on the market, sponsors had more mature expertise on how to manage the complexity of manufacturing and development of monoclonal antibodies, while Health Authorities had a better understanding of where more stringency or flexibility could be applied in the regulatory process. The same can be said about increased regulatory confidence as there was increasing evidence supporting use of intermediate clinical endpoint (e.g., progression free survival in specific cancer types) that predicts the drug's clinical benefit rather than directly measuring clinical benefit using overall survival.

It will be important to develop mechanisms to ensure expedited development programs can be

used in diseases and classes of products with less certainty and understanding to identify the most important steps to take to enable the use of these expedited pathways. Understanding what constitutes meaningful improvement over standard of care and determining standard of care in a crowded class of drugs or rapidly evolving disease area can become very challenging. Enabling innovative trial designs or approaches incorporating novel elements (e.g., real-world evidence, ctDNA, digital tools, in vitro diagnostics, impact of COVID-19, decentralized trials) to participate in programs designed to accelerate clinical development could help more rapidly advance learnings and harmonize approaches.

Proposals for Modernizing Expedited Pathways at the FDA

Based on the above learnings, the working group recommends several proposals that can translate to actionable opportunities to facilitate drug development.

1. Maximize Intent of and Modernize Expedited Programs in the Pre-NDA/BLA Stage

Reconfiguring expedited development programs at the FDA to utilize a more simplified approach with a common entry point for drugs intended to treat a serious or life-threatening condition and the potential to address an unmet need may make the goal of these programs more apparent and streamline their use. This can also help reduce administrative burden for the agency and sponsor. The redundancy of the various qualification criteria for these pathways can often result in duplicative efforts as sponsors assemble applications and set up meetings while the FDA formally reviews each application.

One approach could be to reimagine expedited development programs utilized in the pre-NDA/BLA space by condensing them into a single pathway where the application requirements associated with Fast Track and RMAT are bundled into a pre-Breakthrough Therapy designation pathway. Any drug that would previously qualify for Fast Track or RMAT would qualify for pre-Breakthrough Therapy designation. This may help efficiently usher drugs through key development stages as clinical evidence is generated to support qualifying for Breakthrough Therapy designation. This can help maximize earlier interaction and iterative rapid feedback between sponsors and FDA.

This simplistic approach should be centered around the conversations or interactions that ought to occur when a development program sees early, promising data and when it sees clear confirmatory data to transition from pre-Breakthrough Therapy designation to qualifying for Breakthrough Therapy designation and eventually approval.

2. Codify a Process for Utilizing Expedited Programs

Much attention is given to whether a product is a breakthrough therapy or not, but little focus is given to the processes that follow a Breakthrough Therapy designation. Identifying scenarios where earlier and more frequent interaction would have benefited a program, especially when it was less successful at expediting development, could help elucidate best practices. A comprehensive effort to assess what happens "Beyond Breakthrough," following a designation, is

needed to delineate the obligations and deliverables for sponsors and the FDA once a program qualifies for an expedited program. This should inform the development of updated FDA guidance documents.

A.   Early Stage Development: Pre-Breakthrough Therapy Designation

This is a key place for intervention—when a company is setting up its manufacturing, characterizing its product, conducting a nonclinical program, and starting to generate data to support a Breakthrough Therapy designation or even planning a pivotal trial. Iterative interactions during this key phase of development when clear trends from clinical data are starting to emerge and when important decisions are being made can be extremely valuable. A structured process should be defined that enables early and frequent feedback/dialogue in a more standardized way with shorter timelines than currently available with formal interactions to address early stage questions in a development program, such as optimal analytical tools, discussion on planned manufacturing changes (improved processes, scale up), design of any additional nonclinical studies, dose finding, proof of concept, design of pivotal studies, and approval pathway.

B.   Seeking Breakthrough Therapy Designation

Table 4 provides an outline of actions within the Agency and best practices for sponsors leading up to and following a Breakthrough Therapy designation.

C.   Late Stage Development: After receiving Breakthrough Therapy designation

Actions associated with manufacturing site inspections, strategies for associated diagnostic test development, or design of potential post-market commitments may need to occur following the development of initial clinical evidence. A cross-disciplinary project lead for Breakthrough Therapy designated/RMAT products should use a holistic multidisciplinary approach to begin to map out various processes and the necessary interactions that should occur with different groups within FDA.

D.   Post Approval

Continued interaction and flexibility may also be necessary post-approval for clinical supplements, long-term follow-up studies including the use of real-world data to provide confirmatory clinical evidence, and prior-approval CMC supplements to sustainably provide Breakthrough Therapy designated products to patients.

3. Facilitate Development of Emerging Therapies and Complex Development Programs

Synchronizing Key Components of Drug Development for Emerging Therapies

Dedicated and more frequent meetings for emerging therapies, such as cell and gene therapies and next generation immunotherapies, in a pre-Breakthrough Therapy designation setting may

be necessary to keep key development components in sync to get these potentially transformative therapies to patients quickly and safely. For example, sponsors and FDA could initiate manufacturing meetings in a pre-Breakthrough Therapy designation space in instances where clinical data is indicative of a "breakthrough product" but duration of follow-up is not at the point to support a designation, but likely will in 6 months or so, if data holds.

As considered in the Expedited Programs for Serious Conditions guidance, "*The sponsor of a product that receives an expedited drug development designation may need to pursue a more rapid manufacturing development program to accommodate the accelerated pace of the clinical program,*" and *"Although sponsors must ensure the availability of quality product at the time of approval, FDA may exercise some flexibility on the type and extent of manufacturing information that is expected at the time of submission and approval for certain component."*[3]

The FDA Expedited Programs for Serious Conditions guidance and the FDA Expedited Programs for Regenerative Medicine Therapies for Serious Diseases guidance should be amended to provide additional recommendations on how a sponsor should consider acceleration and flexibility for CMC development and formalizing extended CMC discussions at critical milestones in development. However, it is acknowledged that granting this flexibility may be challenging for very novel therapeutics with limited precedents, such as gene editing products, and will be determined on a case by case basis, requiring additional CMC-specific dialogue with sponsors as well as robust quality risk assessments.

Development of a pilot program to accelerate CMC for products with complex innovative manufacturing processes should be explored. For example, extending the concept of real-time review to manufacturing for these products could further support improvement of the expedited pathways and support innovation. While "rolling review" allows for submission of individual completed modules one at a time rather than at once all together, "real time review" takes this concept a step further and allows the Agency to start the review of a module before the application is complete and may allow submission of pre-agreed CMC data during the NDA/BLA review.

Complex Development Programs

Current, expedited pathways are for drugs that treat serious illnesses and show promise in early trials. However, to demonstrate initial promise, a clinical program may try to utilize a complex innovative design or require advice earlier on for complex manufacturing to generate the early clinical evidence. Products that are completely novel may require considerably more coordination across disciplines within FDA (Clinical, CMC, in vitro diagnostics). A structured process for iterative, holistic cross-discipline interactions (as early as pre-IND) regarding the development program with promise to qualify for expedited pathway(s) should be defined.

Establishing a dialogue very early in the process (phase 1 or earlier) between the sponsor and the FDA would help sponsors devise an efficient development plan and may incentivize sponsors to establish harmonized strategies more likely to generate meaningful clinical data that would be of potential use for multiple therapeutic products. These early dialogues should also acknowledge the complexity of global development as sponsors will be trying to have early

parallel dialogue outside the US. This is an important aspect for global sponsors as feedback is integrated from multiple health authorities while also reconciling the different development speed/pace of each region due to the constraints or limitations of the respective regions. This is especially important when there may be novel aspects to the development program as well (e.g., rapidly changing science, digital tools/endpoints, CMC complexity, decentralized trial design).

Conclusion

Expedited development programs are highly utilized at the FDA, especially for oncology drugs, and sponsors and the FDA have gained substantial experience in identifying and qualifying drugs for these pathways. However, the processes that occur downstream and the interactions between the sponsor and agency that help expedite drug development should be surveyed and more clearly delineated and codified in FDA guidance documents. Over the past several decades, expedited programs have continued to grow to address current needs and facilitate drug development; however, redundancy in the qualification criteria and benefits across the current programs can make it difficult to understand when to apply for one or all in a particular development program. This white paper outlines proposals to streamline expedited development programs, codify a process for expedited programs that outlines pre and post designation processes, and ensure emerging therapies and complex development programs using innovative trial designs can benefit from expedited development pathways.

References

1. Chabner, B. A. Early Accelerated Approval for Highly Targeted Cancer Drugs. N. Engl. J. Med. 364, 1087–1089 (2011).

2. Sharma, M. R. & Schilsky, R. L. Role of randomized phase III trials in an era of effective targeted therapies. Nat. Rev. Clin. Oncol. 9, 208–214 (2011).

3. FDA. Guidance for Industry: Expedited Programs for Serious Conditions––Drugs and Biologics. (2014). Available at: https://www.fda.gov/media/86377/download. (Accessed: 10th July 2020)

4. FDA. Guidance for Industry: Expedited Programs for Regenerative Medicine Therapies for Serious Conditions. (2019). Available at: https://www.fda.gov/media/120267/download. (Accessed: 10th July 2020)

5. Khera, A. Expediting Drug Development Regulatory Pathways Globally. Clin. Res. 34, (2020).

6. EMA. PRIME: priority medicines. Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/prime-priority-medicines.

7. PMDA. Strategy of SAKIGAKE. Available at: https://www.mhlw.go.jp/english/policy/health-medical/pharmaceuticals/140729-01.html.

8. EMA. Adaptive pathways. (2015). Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/adaptive-pathways. (Accessed: 10th July 2020)

9. FDA. Real-Time Oncology Review Pilot Program. (2020). Available at: https://www.fda.gov/about-fda/oncology-center-excellence/real-time-oncology-review-pilot-program. (Accessed: 10th July 2020)

10. FDA. Assessment Aid. (2020). Available at: https://www.fda.gov/about-fda/oncology-center-excellence/assessment-aid. (Accessed: 10th July 2020)

11. Shea, M. et al. Regulatory Watch: Impact of breakthrough therapy designation on cancer drug development. Nature reviews. Drug discovery 15, 152 (2016).

12. FDA. Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals. Available at: https://www.fda.gov/drugs/drug-approvals-and-databases/compilation-cder-new-molecular-entity-nme-drug-and-new-biologic-approvals.

Table 1: Expedited development and review pathways

	Fast-track Designation	Regenerative Medicine Advanced Therapy	Breakthrough Therapy Designation	Priority Review	Accelerated Approval
Eligibility	1. Treat serious condition 2. Potential to fill an unmet medical need (clinical or nonclinical data)	1. The drug is a regenerative medicine therapy, which is defined as a cell and gene therapy, therapeutic tissue engineering product, human cell and tissue product, or any combination product using such therapies or products 2. Intent to treat, modify, reverse, or cure a serious condition 3. Preliminary clinical evidence indicates potential to address unmet medical need	1. Treat serious condition 2. Preliminary clinical evidence drug may demonstrate substantial improvement over existing therapies	1. An application for a drug that treats a serious condition 2. If approved would provide a significant improvement in safety or effectiveness	1. Treat serious condition 2. Provide meaningful therapeutic benefit over existing therapies 3. Surrogate endpoint reasonably likely to predict clinical benefit
Designation	With IND or after; FDA has 60 days to respond	With IND or after, ideally no later than end-of-phase 2 meeting; FDA has 60 days to respond	With IND or after, ideally no later than end-of-phase 2 meeting; FDA has 60 days to respond	Within 60 days of receipt of original BLA, NDA, or efficacy supplement	Not applicable. Agreed upon during formal meetings (typically Type B meeting)
Features	Earlier and more frequent communication	All Breakthrough Therapy designation features, including early interactions to discuss any potential surrogate or intermediate endpoints	Intensive guidance on efficient drug development; earlier and more frequent communication; delegation of senior reviewers and cross-disciplinary review team	Not applicable	Accelerated Approval granted based on early endpoints
Review Process	Option for rolling NDA/BLA submission; official review clock begins when last module is submitted	Option for rolling BLA submission; potential for shorter review time	Option for rolling NDA/BLA submission; potential for shorter review time	NDA/BLA data submitted in one package; review time shortened to 6 months after filling	Confirmatory post approval clinical evidence part of post-marketing requirement

Table 2: Utilization of current expedited programs from 2012–2019

Expedited Program	Total (n=327)	Oncology (n=88)	Non-Oncology (n=239)
Fast Track	123 (38%)	49 (56%)	84 (35%)
Breakthrough Therapy des-ignation	72 (22%)	36 (41%)	36 (15%)
Priority Review	195 (60%)	71 (80%)	124 (52%)
Accelerated Approval	45 (14%)	35 (40%)	10 (4%)
None	116 (35%)	9 (10%)	107 (45%)

Note: Percentages calculated using totals within each clinical group. Percentages total greater than 100% because multiple programs can be used for a single drug. Data from "Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals."[12]

Table 3: Frequency of use for different combinations of expedited programs from 2012–2019

Expedited Program	Total (n=327)	Oncology (n=88)	Non-Oncology (n=239)
PR only	46 (14%)	14 (16%)	32 (13%)
FT only	13 (4%)	5 (6%)	8 (3%)
PR + AA	3 (1%)	1 (1%)	2 (1%)
BTD + PR	20 (6%)	7 (8%)	13 (5%)
FT + PR	63 (19%)	15 (17%)	48 (20%)
AA + FT	1 (0%)	1 (1%)	0 (0%)
PR + BTD + AA	19 (6%)	18 (20%)	1 (0%)
FT + PR + AA	13 (4%)	7 (8%)	6 (3%)
FT + BTD + PR	24 (7%)	3 (3%)	21 (9%)
FT + BTD + PR + AA	9 (3%)	8 (9%)	1 (0%)

Note: Fast Track, FT; Breakthrough Therapy designation, BTD; Priority Review, PR; Accelerated Approval, AA. Data from "Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals."[12]

Table 4: Key steps and processes within the FDA and drug Sponsor

Steps	FDA Procedures	Sponsor Actions & Best Practices
Preliminary Breakthrough Therapy Designation Request • Summary briefing document • Preliminary Breakthrough Therapy teleconference	Review preliminary Breakthrough Therapy briefing document and discuss internally Share thoughts regarding appropriateness of Breakthrough Therapy submission in teleconference	Sponsor initiates dialogue with review division by preparing and submitting a concise (2-pager) document that summarizes the eligibility of the drug/therapy, the preliminary data supports the promising therapy. Recommend that the pivotal study dose is selected and that there is sufficient data to support the safety profile and preliminary activity (e.g., 30+ patients with at least 6 months follow-up with an established endpoint). The preliminary request is reviewed, and a full Breakthrough Therapy designation application should only be submitted after receiving support from the respective review division.
Breakthrough Therapy Review • 60-day review	Review by Division/Office and review by CDER Medical Policy Counsel (via email or presentation depending on complexity)	A full Breakthrough Therapy designation request will be granted or denied following a 60-day review period.
Post-Breakthrough Therapy Designation Review (when granted) • Multidisciplinary Breakthrough Therapy meeting • Subsequent Type B Meetings • Critical IND milestone meetings • Other communications	All disciplines invited to attend and participate in the multidisciplinary meeting The frequency of subsequent meetings determined by the communication plan established at the initial comprehensive multidisciplinary meeting	Once Breakthrough Therapy has been granted, the Sponsor can request formal multidisciplinary and milestone meetings with the Agency with increased frequency and to ensure real-time collaborative dialogue. Initially, it may be beneficial to have a broad meeting with several FDA disciplines to ensure alignment for the overall program (e.g., clinical, pharmacology, CMC, CDRH). Subsequent meetings may require more detailed and focused discussion with the primary review discipline. However, Sponsors are encouraged to leverage the clinical FDA project manager to ensure consistent communication and dialogue with all FDA review disciplines.
Application Review (NDA/BLA)	Early internal discussions about the appropriateness of an expedited review Consideration of real-time oncology review/ORBIS/Priority Review/Assessment Aid	Sponsors are encouraged to communicate key milestones to FDA in advance, which allows an ongoing dialogue and advice on whether additional tools, pathways, or pilots can be leveraged for the license application review. One such milestone is communicating to the Agency potential pivotal data availability/unblinding approximately 4 months in advance, which would enable both FDA and the Sponsor to prepare for expedited submission and review, if applicable.

Figure 1. Time of key regulatory actions in relation to drug approval
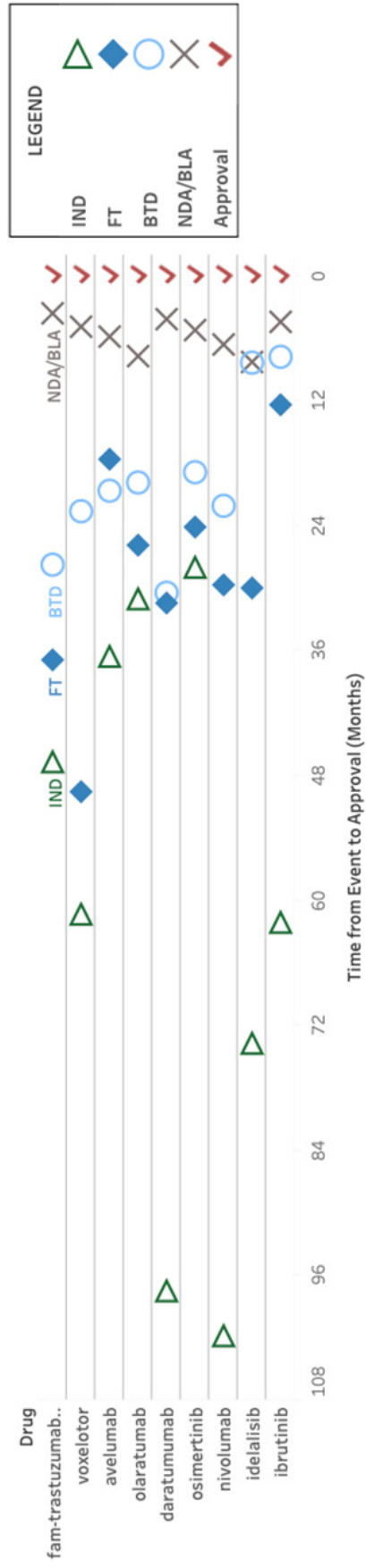
Investigational New Drug, IND; Fast Track, FT; Breakthrough Therapy Designation, BTD; New Drug Application, NDA; Biologics License Application, BLA

Note: These drugs were selected because they utilized Fast Track, Breakthrough Therapy, Priority Review, and Accelerated Approval expedited programs.

FRIENDS
of CANCER
RESEARCH

Optimizing the Use of Accelerated Approval

FRIENDS OF CANCER RESEARCH ANNUAL MEETING 2020

The Accelerated Approval (AA) Program has been an important regulatory mechanism for FDA to allow for earlier approval of drugs that treat serious and life-threatening illnesses than would occur through the traditional approval program. Created in 1992, the AA program was conceived as a direct response to patient therapy during the HIV/AIDS epidemic and in recognition of the urgency of access to new therapy needs faced by patients with life-threatening illnesses. As opposed to traditional approval, which is based upon a direct measure of clinical benefit (Glossary) or a validated surrogate, AA is intended to allow for the initial approval of a drug based on a demonstration of effect on a surrogate endpoint—or an intermediate clinical endpoint—that is reasonably likely to predict a clinical benefit. [1-3] Under FDA regulations, sponsors should conduct post-marketing studies that verify and describe the expected clinical benefit of the drug with a clinical trial design as agreed upon with FDA at the time of AA.[3] The AA statute also establishes provisions for withdrawal of an AA drug where confirmatory trials fail to verify clinical benefit or safety concerns arise.

In 2012, the AA program (Subpart H – drugs and Subpart E - biologics) was amended by the FDA Safety and Innovation Act (FDASIA)[4]:

> *"The Secretary may approve an application for approval of a product for a serious or life-threatening disease or condition, including a fast track product, under section 355(c) of this title or section 351(a) of the Public Health Service Act [42 U.S.C. 262(a)] upon a determination that the product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit, or on a clinical endpoint that can be measured earlier than irreversible morbidity or mortality, that is reasonably likely to predict an effect on irreversible morbidity or mortality or other clinical benefit, taking into account the severity, rarity, or prevalence of the condition and the availability or lack of alternative treatments."*

FDASIA maintained the reliance of an AA on an intermediate endpoint (either surrogate or clinical endpoint that can be measured earlier) that is reasonably likely to predict an effect on clinical benefit but removed the initial requirement for an AA drug to "generally provide meaningful

Thank You to Our Working Group Members

Josh Bilenker
Loxo Oncology at Lilly

Gideon Blumenthal
Merck

R. Angelo de Claro
FDA

George Demetri
Dana-Farber Cancer Institute

Ritesh Jain
EMD Serono, Inc. (Merck KGaA)

Praveen Marapaka
EMD Serono, Inc. (Merck KGaA)

David Mitchell
Multiple myeloma patient

advantage over available therapies." Although FDA's regulations and guidance have not yet been modified to reflect the later language change, the modified language in FDASIA nonetheless reduced some ambiguity regarding which products may qualify for accelerated approval. By explicitly incorporating a more comprehensive benefit-risk assessment in FDA communications regarding an AA, along with outcomes meaured by a surrogate or intermediate clinical endpoint, stakeholder confusion related to AA could be further reduced.

The AA pathway broadly applies to all drug classes and is used across clinical divisions within the FDA. However, AA has been most frequently used in oncology. In the past 10 years (2010-2019), 84% of FDA's accelerated approvals were granted for oncology indications.[5] The robust experience of AA in oncology, which is unique given an extensive infrastructure for conducting research and aggregating data, can be used to inform the use of AA in other indications. This white paper will explore the impact of AA, identify challenges, and pose improvements both broadly and within the context of, and informed by, learnings from applications of AA in oncology. For a discussion of other expedited programs used by the FDA see the companion white paper "Modernizing Expedited Development Programs."

Why is AA Important for Patients?

Since its creation by FDA in 1992, 148 new drugs or biologics to treat serious or life-threatening illnesses have been approved through the AA program.[5] One assessment of oncology treatments concluded that therapies receiving AA were made available a median of 3.4 years earlier than would be achievable if confirmation of clinical benefit based upon a primary endpoint, such as overall survival, was required.[6]

AA has extended or, in certain cases, saved patients' lives by providing earlier access to novel therapies than would have been possible using the traditional FDA approval pathway. Specifically, for multiple myeloma, access to new therapies that are used as single agents, and are now being used in combination, have been accelerated, extending the time of disease stabilization when, in the absence of the AA drug, patients would have experienced debilitating symptoms, progressed, or died.

As more transformative treatments are developed that extend survival by years or even decades, the ability to quantify overall survival will become increasingly difficult or impossible within the context of a clinical trial. Specifically, for patients with a terminal illness or those that lack other treatment options, randomization to a control arm to determine overall survival (OS) is, in many cases, unethical. Further, as treatments become more highly targeted to smaller populations or subsets of diseases, traditional measures of benefit will become more difficult to employ where large numbers of patients are needed to statistically quanitify OS as compared to surrogate measures of clinical benefit such as response rate. Enhancements to the AA pathway will help to ensure continued benefit from this program as medicines and drug development evolve. For example, where a confirmatory trial for an AA therapy would traditionally verify clinical benefit measured by a surrogate endpoint, communication of a preliminary benefit-risk assessment that includes but is not limited to consideration of outcomes of a surrogate endpoint, may better reflect a more holistic consideration of factors that are important to patients

in an approval determination. In other words, the confirmatory trial would be conducted to verify the totality of evidence of a drug, including magnitude and duration of benefit and safety, and whether the benefits received from a drug justified the risk rather than focus only on confirmation of a primary endpoint. A framework to encourage greater patient input on the determination of benefit-risk is important to amplifying the patient voice in drug development.

Finally, although the benefits of a drug may outweigh the risks in a clinical trial population, it is important to characterize the benefit-risk profile in the real-world population. The information provided by clinical trials is based upon a highly selected, homogenous, patient pool (typically younger patients with fewer comorbidities) that is less reflective of the general population. A patient cannot make truly informed decisions regarding treatment choices without adequate data to provide a complete picture of the benefits and risk of a therapy. For this reason, the importance of Phase IV confirmatory studies, which examine the benefit of a therapy and the toxicities in a broader population or in the real world through real-world studies, cannot be overemphasized. The AA pathway could supplement post-approval required trials with such real-world assessments to capitalize on all mechanisms of data generation to produce the most robust benefit-risk assessment possible.

Surrogate Endpoint vs Benefit-Risk Assessment

Since the AA Program was codified in the US, analogous regulatory pathways have been implemented by other regulatory bodies, including the European Medicines Agency (EMA) and Health Canada, with the intent of expediting access to new therapies intended to treat serious diseases in those settings (Table 1). While implemented for a similar purpose, a comparison of each pathway also reveals important differences. For example, Conditional Marketing Authorisation (Conditional Approval), the pathway implemented by the EMA is distinctive with respect to its use of an initial benefit-risk analysis of a drug as a basis for a Conditional Approval as opposed to evaluation of an drug's effect on a surrogate endpoint that is used as the basis for AA by FDA and many of the other programs. Further, Conditional Approval by EMA is valid for only one year with the option of sponsor application for renewal. It is also important to note that regulatory approval in the EU does not necessarily translate to immediate patient access to new drugs as in the US because European countries also require a health technology assessment once a drug is determined to be safe and effective before reimbursement is awarded.

There is some discrepancy, particularly in oncology, regarding the clinical and regulatory context in which a surrogate endpoint is used to grant traditional approval and when a surrogate endpoint is considered reasonably likely to predict benefit to support AA. This can create confusion over distinctions between traditional and accelerated development programs. For example, objective response rate (ORR) is considered a surrogate endpoint used for AA in oncology. However, FDA can also grant tranditional approval based on this surrogate endpoint in single-arm trials when the ORR is substantial and durable. As another example, responses of fungating skin lesions were considered evidence of direct clinical benefit to support traditional approval of vismodegib for advanced basal cell carcinoma.[7] Consequently, the use of the same surrogate in different contexts necessitates greater clarity regarding the level of evidence necessary for and how various endpoints are considered across development programs when a

drug can be granted traditional approval based upon an endpoint that is also used for AA.

Similar to the use of a benefit-risk assessment to support conditional approval by the EMA, drug development through the AA pathway in the US could be enhanced if communication about an AA were shifted away from a focus solely on predictive endpoints and toward a discussion about overall benefit-risk considerations. FDA already uses a standard framework for benefit-risk considerations when making approval decisions. The elements of FDA's benefit-risk framework include Analysis of Condition, Current Treatment Options, Benefit, and Risks and Risk Management.[8] Greater clarity regarding how FDA considers benefit-risk could be helpful, particularly regarding the magnitude of effect and potential toxicities of a drug. For example, when considering magnitude of effect, a substantial outcome in a surrogate endpoint may be a superior outcome vs a less impactful outcome as measured in a traditional endpoint. Additionally, potential toxicities should be considered within the context of importance to a patient's quality of life and may contribute to determining the "availability" of treatments. It may be appropriate to award AA to a drug with a lower ORR if the drug is less toxic or has a positive impact on patient-reported outcomes or function—and a confirmatory trial would aim to verify that benefit vs risk was maintained in the post-market setting.

Challenges and Solutions

Pre-approval setting

Prioritization of the benefit-risk framework for drug review would facilitate a more holistic assessment of new therapies. It is within the context the above considerations, regarding the benefit-risk assessment in regulatory determinations, that we suggest additional considerations to improve AA within the benefit-risk framework.

Defining an "available therapy." The statute for AA, as amended by FDASIA, establishes eligibility for an AA and requires FDA to take "into account the availability or lack of alternative treatments." Patient access to treatments through AA has benefited from the clinical judgement that FDA reviewers have been afforded and the ability to account for confounders when considering an "available therapy." However, challenges remain when interpreting the definition of available therapy in certain situations. First, it is not always clear whether the existence of an FDA approved drug with an FDA approved indication in the disease of interest should necessarily be considered an available therapy. For example, over time and as the standard of care improves, some drugs become less relevant, or not used at all, in clinical practice at the time of a new AA and should not be considered an available therapy when assessing a new drug application in the same indication. Second, the use of published literature to establish an available therapy is highly discretionary but could benefit from additional clarity. For example, FDA has considered drugs for first-line lung cancer as an "available therapy" for lung cancer patients when determining eligibility for AA. When AA was awarded for crizotinib for metastatic non-small cell lung cancer with anaplastic lymphoma kinase rearrangements (ALK+), platinum doublet chemotherapy in first-line and docetaxel in second-line were considered available therapies. The crizotinib AA was based upon two single arm trials compared to published literature of ORR for platinum doublet chemotherapy and docetaxel.[9] However, benefit was confirmed in a randomized confirmatory trial. Last, an emerging consideration is how to define available therapy for molecular

indications. When considering a biomarker positive population, it may not be appropriate to consider an FDA approved drug with an expansive indication, which would include the biomarker positive population, but was never studied in that subpopulation, as an available therapy. A standardized approach to the definition of an "available therapy" in the context of a specific disease setting or population/subpopulation, including biomarker positive and novel refractory disease states (e.g., PD-(L)1-refractory populations), should be considered to provide greater clarity and consistency in application of AA.

Surrogate endpoint. Surrogate or intermediate endpoints such as duration of response or ORR are tumor-based endpoints, and there is no consistency in the magnitude of improvement in response rates that would constitute a change in other endpoints such as overall survival. Further, ORR from historic literature may not be as accurately assessed as compared to ORR in a modern registrational clinical trial, which typically requires blinded independent central radiology review. It is difficult to assess the level of evidence needed to establish that a surrogate endpoint fulfills the requirement of "reasonably likely to predict clinical benefit." Standardization or additional guidance for qualitative metrics of surrogate or intermediate endpoints would be helpful to provide more clarity and predictability for development programs without reducing flexibility for regulatory decision-making. Expectations and transparency in how FDA will consider magnitude of surrogate measures could be further clarified in design of confirmatory trials.

Another consideration in the use of surrogate endpoints is a better understanding of how the intermediate endpoint is weighted in a benefit-risk assessment. Different considerations may need to be taken for response rate vs duration of response and the magnitude of each. For example, tazemetostat was unanimously recommended by an Oncology Drug Advisory Committee (ODAC) based upon a 11-15% ORR for patients with metastatic or locally advanced epithelioid sarcoma with a lack of available therapies being a key consideration.[10] In oncology, a high response rate with a duration of response that lasts more than one year is preferable, but less substantive outcomes will require more nuanced consideration, including the rarity of the patient population.

Heterogeneity in populations. There has emerged a phenomenon of "excellent responders" in the context of immunotherapies, where there may be less than 10% of patients that respond to a therapy but those minority of patients that do respond exhibit dramatic and long term responses. In these cases, the overall trial for the general population may fail to demonstrate a benefit, but treatment may still be appropriate for those "excellent responders." It may be appropriate to award AA in that "excellent responder" subpopulation, despite failure of the trial to demonstrate benefit in the overall population, followed by post-market confirmatory requirements. How FDA considers surrogate endpoints in a benefit-risk assessment could be further clarified in guidance including how to appropriately design a statistically powered trial to identify efficacy in these sub-populations. This may need to include considerations for the objective or definition of a confirmatory trial. For example, a "confirmation of benefit" may be less about demonstrating superior survival of a therapy in the overall population and it may be more important to identify those patients that are "exceptional responders" based upon response measured in a biomarker-positive population.

Development of surrogate endpoints. More research is needed to develop new surrogate endpoints or provide more substantial evidence of likelihood to predict clinical benefit in support of AA. Surrogate endpoints that can clarify benefit in patients who achieve disease stabilization, such as changes in circulating tumor DNA, may be an important tool for drug development and clinical decision making. FDA could create a formal process, or expand upon the Drug Development Tool (DDT) Qualification Program, for sponsors to submit key data variables and patient outcomes from clinical trials used to support accelerated approval and traditional approval to help validate endpoints that predict clinical benefit. FDA considers clinical outcomes assessments to be a DDT and has issued draft guidance to inform the qualification of these metrics.[11] This evidence could be aggregated through a collaborative community or independently of FDA through precompetitive consortia to provide a publicly available database of evidence to support benefit-risk assessments that include evidence based upon a surrogate endpoint.

External control arms to support clinical trials. Clinical trials, from Phase I dose-finding and safety trials to Phase III randomized trials examining efficacy, form the backbone of the drug development pipeline and inform regulatory approvals. Single-arm clinical trials are now used to support regulatory approval, particularly important for AA, in settings where ethical concerns or challenges with feasibility of deploying a concurrent control arm exist, such as rare diseases or populations with unmet needs where randomization to a placebo or active comparator (for refractory settings) would be inappropriate and/or not feasible. While single-arm trials alone may yield important safety and efficacy signals and can be relied on for regulatory decision making in these clinical and regulatory contexts, real-world evidence (RWE), such as external controls (sometimes referred to as synthetic controls) may provide additional context and supplementary evidence. For example, in 2017, avelumab received AA for Merkel cell carcinoma on the basis of an 88-patient single arm Phase II trial. Real-world evidence (RWE), contributed by external data from a registry, was used as supportive evidence, but the regulatory approval was based primarily on data from the Phase II trial.[12] Expanding the use of external controls to other difficult-to-study indications may reduce patient burden where research may be slowed or uninterpretable due to the use of a concurrent randomized control. The latter may be the case with some confirmatory trials of medical products made available through the accelerated approval pathway where the control arm may be compromised by early discontinuation or treatment crossover to the investigational therapy made available by an AA.[13]

Post-approval challenges and solutions
Although awarding of an AA to market a new therapy is contingent upon continued generation of evidence to verify and describe drug effectiveness, enrollment in confirmatory trials once a drug is already on the market may pose challenges. Certainly, once a drug is widely available, the incentive for a patient to participate in a clinical trial, and risk randomization to a placebo or an active control that is perceived as inferior, is reduced. This situation can be further exacerbated where a substantial improvement in overall survival is expected, as in the case of AA and breakthrough therapy designated drugs, and there may be loss of equipoise for conducting a randomized trial with a less effective therapy for confirmation of clinical benefit following AA. Further, it may not be ethical to take advantage of access barriers outside of the US, where the therapy is not yet available, to accrue patients to a trial that otherwise would be unlikely

to accrue patients within the US. The confirmatory evidence deemed necessary by the FDA to assess the benefit-risk of the therapy is nevertheless critical to ensure patient safety and benefit and different approaches to generating this information are needed, along with consideration of how evidence generated from confirmatory trials inform changes to labeled indications.

Initiation of confirmatory studies in pre-market setting. Confirmatory trials could be required to be initiated and have enrolled a pre-determined number of patients when the marketing application (NDA or BLA) is filed. This would require additional and earlier communication between FDA and sponsor to facilitate, including, more real-time access for the FDA to the necessary data that would inform design of a confirmatory trial, including guidance to determine how a "minimum number of patients accrued" would be defined across different drugs and disease settings. Access to data could be provided to FDA on a similar timeframe as a drug manufacturing and formulation program. This could not only inform confirmatory trial design earlier but also facilitate use of more pragmatic trial designs by aligning the patient population pre-approval to the initial study to give greater confidence in results of the confirmatory trial. Further, it could promote the awarding of AA based on an "intermediate data review," in line with a benefit-risk assessment as opposed to a surrogate endpoint.

Consideration of subsequent confirmatory studies. Sponsors can conduct a confirmatory trial in different populations or settings than the initial trial for which AA was awarded for numerous reasons, including low accruing trials and loss of equipoise, and there are examples where this has occurred. Most commonly, this is used for AA based upon substantial response rate in a single-arm trial with a monotherapy in a refractory population, the confirmatory trial, then, utilizes a randomized design in an earlier line setting. For rare populations where randomized trials are not feasible, confirmatory trials, with a single-arm design consisting of more patients and/or longer follow-up for duration of response may be considered.

A randomized design approach, however, can become problematic if the confirmatory trial that utilizes a different population than the initial trial, fails to confirm benefit in the subsequent population. These results are not necessarily a reflection of the effectiveness of the drug but are likely reflective of trial design related issues such as inability to accrue sufficient patients, high drop-out or crossover rates that impact the statistical power of the study, or enrolling the wrong patient population. Gefitinib, approved in non-small-cell lung cancer, is an example in which the confirmatory trials failed and ultimately lead to a withdrawal of AA.[14] However, subsequent trials were able to identify an appropriate patient population, leading to a subsequent approval in EGFR-mutant lung cancer.[15,16] In similar instances, FDA may be hesitant to remove an AA for a therapy due to concerns that the treatment may meet an unmet medical need in a certain subpopulation while still recognizing that additional clinical trials are needed to confirm benefit in that subpopulation. Indeed, the impact of withdrawal of AA for gefitinib on unmet medical need was mitigated by the availability of another therapy, erlotinib, which remained on the market. Without the availability of an alternative therapy, access issues for that particular population would have been of concern. By allowing the sponsor to retain AA for the drug after the initial confirmatory trial failed and conduct additional post-market trials allows the FDA to address both concerns. Further, FDA could host a public discussion at an ODAC to discuss the failed trial and to consider whether other confirmatory trials or withdrawal may be appropriate. This public

format would facilitate a transparent discourse that bolsters patient input in the decision-making process and prioritizes benefit-risk assessments in the post-market setting. FDA currently has a withdrawal process for removing AA, but this is an onerous and time-consuming process, making it an ineffective method for withdrawing AA in a timely manner when the company does not agree with the withdrawal. The withdrawal process will also require improvement to facilitate opportunities for subsequent confirmatory trials, where appropriate and necessary, to ensure a robust AA pathway. The FDA should consider ways in which this pathway can be made more nimble and improve this mechanism as an enforcement mechanism for required confirmatory studies. Additional changes to the AA Program could be considered. For example, analogous programs to AA implemented by other regulatory agencies, such as the EMA, are valid for one year after approval with the option to renew annually. The FDA could be given authority to require an annual update of post-market requirements and review of new data to ensure post-market commitments are met in a timely manner.

Real-world evidence to support confirmation of benefit. RWE is increasingly becoming utilized in drug development, including in the post-market setting. Recent legislative and regulatory policies focused on RWE, such as the 21st Century Cures Act, Prescription Drug User Fee Act (PDUFA) Reauthorization of 2017, and FDA Framework on Real-World Evidence, highlight the interest in using RWE applications across the drug development life cycle.[17-20] While the centrality of clinical trials remains, the homogenous patient populations included to produce rigorous data limit the generalizability of clinical trial-related drug safety and efficacy to its broader use in clinical practice.[21] Real-world datasets, on the other hand, can be assembled that produce robust analyses that complement those of clinical trials. RWE can reflect broader, more diverse patient populations than are typically included in traditional clinical trials and can be applied across multiple use cases, including to answer timely clinical questions, assess endpoints measures, perform comparative effectiveness research, and study long-term drug safety. Within the context of a benefit-risk assessment for AA, additional evidence from RWE could be used to supplement confirmatory trial results and contribute to a more complete understanding of drug efficacy and safety.

Conclusions

The AA pathway has proved to be an extremely important mechanism to promote development of and access to therapies for serious or life-threatening illnesses. It is important to continue to improve on this mechanism to maximize the benefit achievable through this pathway for patients. This white paper provides several possible recommendations to achieve this goal.

Glossary

Clinical benefit: a positive therapeutic effect that is clinically meaningful in the context of a given disease. The clinical benefit must be weighed against a treatment's risks to determine whether there is an overall benefit for patients (i.e., a positive benefit-risk profile).[3, 22]

Clinical endpoint: a characteristic or variable that directly measures a therapeutic effect of a drug——an effect on how a patient feels (e.g., symptom relief), functions (e.g., improved mobility), or survives.[3, 22]

Intermediate clinical endpoint: a measurement of a therapeutic effect that can be measured earlier than an effect on irrecersible morbidity or mortality (IMM) and is considered reasonably likely to predict the drug's effect on IMM or other clinical benefit.[3, 22]

Reasonably likely surrogate endpoint: surrogate endpoint that is supported by strong mechanistic and/or epidemiologic rationale, but the amount of clinical data available is not sufficient to show that they are a validated surrogate endpoint.[22]

Surrogate endpoint: a marker, such as a laboratory measurement, radiographic image, physical sign, or other measure, that is thought to predict clinical benefit, but is not itself a measure of clinical benefit. Depending on the strength of the evidence supporting the ability of a marker to predict clinical benefit, the marker may be a surrogate endpoint that is known to predict clinical benefit (a validated surrogate endpoint that could be used for traditional approval), a surrogate endpoint that is reasonably likely to predict a drug's intended clinical benefit (and that could therefore be used as a basis for accelerated approval), or a marker for which there is insufficient evidence to support reliance on the marker as either kind of surrogate endpoint (and that therefore cannot be used to support traditional or accelerated approval of a marketing application).[3, 22]

Validated surrogate endpoint: surrogate endpoint supported by a clear mechanistic rationale and clinical data providing strong evidence that an effect on the surrogate endpoint predicts a specific clinical benefit.[22]

References

1. 21 CFR part 314, subpart H

2. 21 CFR part 601, subpart E

3. Food and Drug Administration. Guidance for Industry Expedited Programs for Serious Conditions – Drugs and Biologics. May 2014. https://www.fda.gov/media/119293/download. Accessed online 10/7/2020.

4. Section 506(c) of the FD&C Act, as amended by section 901 of FDASIA.

5. US Food and Drug Administration. CDER Drug and Biologic Accelerated Approvals Based on a Surrogate Endpoint As of June 30, 2020. https://www.fda.gov/media/88907/download. Accessed online 10/1/2020.

6. Beaver JA, Howie LJ, Pelosof L, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics: A Review. JAMA Oncol. 2018;4(6):849–856. doi:10.1001/jamaoncol.2017.5618. Accessed online 10/1/2020.

7. Media release. FDA approves Erivedge (vismodegib) capsule, the first medicine for adults with advanced basal cell carcinoma. Roche Group Media Relations. Basel, 30 January 2012. https://www.roche.com/media/releases/med-cor-2012-01-30.htm.   Accessed online 10/1/2020.

8. US Food and Drug Administration. Benefit-Risk Assessment in Drug Regulatory Decision-Making: Draft PDUFA VI Implementation Plan (FY 2018-2022) accessed on 10/27/2020. https://www.fda.gov/media/112570/download

9. Kazandjian D, Blumenthal GM, Chen HY, et al. FDA approval summary: crizotinib for the treatment of metastatic non-small cell lung cancer with anaplastic lymphoma kinase rearrangements. Oncologist. 2014;19(10):e5-e11. doi:10.1634/theoncologist.2014-0241. Accessed online 10/8/2020.

10. Media Release. FDA Approves Tazemetostat for Epithelioid Sarcoma. OncLive. January 23, 2020. tps://www.onclive.com/view/fda-approves-tazemetostat-for-epithelioid-sarcoma. Accessed online 10/1/2020.

11. Food and Drug Administration. Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff. DRAFT GUIDANCE. December 2019. https://www.fda.gov/media/133511/download. Accessed online 10/7/2020.

12. Cowey, C. L. et al. Real-world treatment outcomes in patients with metastatic Merkel cell carcinoma treated with chemotherapy in the USA. Futur. Oncol. 13, 1699–1710 (2017).

13. Friends of Cancer Research. Exploring Whether a Synthetic Control Arm can be Deriving from Historical Clinical Trials that Match Baseline Characteristics and Overall Survival Outcome of a Randomized Control Arm. (2018). Available at: https://www.focr.org/sites/ default/files/SCA White Paper.pdf.

14. Food and Drug Administration. AstraZeneca Pharmaceuticals LP; Withdrawal of Approval of a New Drug Application for IRESSA. (77 FR 24723, April 25, 2012).

15. Pircher A, Ploner F, Popper H, Hilbe W. Rationale of a relaunch of gefitinib in Caucasian non-small cell lung cancer patients. Lung Cancer. 2010 Sep;69(3):265-71. doi: 10.1016/j.lungcan.2010.01.017. Epub 2010 Feb 18. PMID: 20167388. Accessed online 10/7/2020.

16. Kazandjian D, Blumenthal G, Yuan W, He K, Keegan P, Pazdur R. FDA Approval of Gefitinib for the Treatment of Patients with Metastatic EGFR Mutation–Positive Non–Small Cell Lung Cancer. Clin Cancer Res 2016;22:1307-1312. doi: 10.1158/1078-0432.CCR-15-2266. Accessed online 10/7/2020.

17. US Food and Drug Administration: PDUFA reauthorization performance goals and procedures fiscal years 2018 through 2022. https://www.fda.gov/media/99140/download

18. US Congress: 21st Century Cures Act. P.L. 114-255, 2016.

19. US Congress: Prescription Drug User Fee Act. P.L. 102-571, 1992.

20. Food and Drug Administration. Framework for FDA's Real-world Evidence Framework. December 2018. https://www.fda.gov/media/120060/download. Accessed online 10/26/2020.

21. Elting L, Cooksley C, Bekele B, Frumovitz M, Avritscher E, Sun C, Bodurka D. Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. Cancer 2006; 106(11):2452-2458. DOI 10.1002/cncr.21907

22. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. Glossary. 2016 Jan 28 [Updated 2020 Aug 4]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338448/ Co-published by National Institutes of Health (US), Bethesda (MD).

Table 1: Comparison of AA Programs Across Nations

Regulatory Agency	Program Name	Date Initiated	How do you qualify?	How is drug evaluated?	Post-marketing Requirements	References/Notes
FDA – US	Accelerated Approval	1992	A drug that treats a serious condition **AND** generally provides a meaningful advantage over available therapies **AND** demonstrates an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit or on a clinical endpoint that can be measured earlier than irreversible morbidity or mortality (IMM) that is reasonably likely to predict an effect on IMM or other clinical benefit (i.e., an intermediate clinical endpoint)	- **Surrogate Endpoint:** a marker that is thought to predict clinical benefit, but is not itself a measure of clinical benefit **- Intermediate Clinical Endpoint:** a measurement of a therapeutic effect that can be measured earlier than an effect on IMM and is considered reasonably likely to predict the drug's effect on IMM	- Postmarketing confirmatory trial that evaluates a clinical endpoint that directly measures clinical benefit **OR** - A confirmatory trial conducted in a different but related population that is capable of verifying the predicted clinical benefit	• 21 CFR part 314, sub-part H • 21 CFR part 601, sub-part E • Section 506(c) of the FD&C Act, as amended by section 901 of FDASIA

EMA – EU	Conditional Marketing Authorisation	2006	Medicinal products that aim at the treatment, prevention or medical diagnosis of seriously debilitating or life-threatening diseases, or medicinal products to be used in emergency situations in response to public health threats	(a) the risk-benefit balance of the medicinal product, as defined in Article 1(28a) of Directive 2001/83/EC, is positive; (b) it is likely that the applicant will be in a position to provide the comprehensive clinical data; (c) unmet medical needs will be fulfilled; (d) the benefit to public health of the immediate availability on the market of the medicinal product concerned outweighs the risk inherent in the fact that additional data are still required	– The holder will be required to complete specific obligations (ongoing or new studies, and in some cases additional activities) with a view to providing comprehensive data confirming that the benefit–risk balance is positive – Valid for one year, and renewable afterwards	Regulation (EC) No 507/2006
PMDA – Japan	Conditional Early Approval System	2017	– No standard existing therapy or superior clinical usefulness as compared with the existing products in terms of quality of life of patients, efficacy, or safety – Applicable to serious diseases or orphan drug designation – Confirmatory clinical trials are difficult to conduct or take a long time due to a limited number of patients – Clinical trials other than confirmatory trials have shown a certain degree of efficacy and safety	– Clinical trials that use a justified surrogate endpoint – Show the safety and efficacy of the drug in some other way	– Post-marketing surveillance period extended – Surveillance or clinical studies must be conducted as a post-marketing requirement (recent examples indicate that post-marketing comparative studies are not necessary and post-marketing surveillance is acceptable)	Pharmaceutical and Medical Device Act

NMPA – China	The Green Channel	2014	Fills 1 of 8 criteria: – Drug applications for products with "significant clinical value," such as innovative drugs and those with advanced formulations; – Clinical trial applications for generic drugs submitted three years before the patent expires on the reference product; – Clinical trial applications for new drugs undergoing parallel review in the U.S. and EU; – Registration applications for drugs undergoing parallel review in the U.S. and EU using the same production line; – Registration applications for new drugs developed under Chinese national research programs; – Registration applications for drugs to treat AIDS, tuberculosis, viral hepatitis, rare diseases and cancer, as well as pediatric and geriatric drugs; – Registration applications for drugs in short supply approved by China's National Health and Family Planning Commission and the Ministry of Industry and Information Technology; and – Registration applications for pediatric drugs that have been approved in the U.S., EU and "surrounding areas" of China, backed by compelling clinical data	– Phase 1 and Phase 2 data if reviewers can reasonably predict or determine the clinical benefit has a significant advantage versus existing treatments – Trial applicants with less convincing Phase 1 and 2 data still may request an abbreviated Phase 3 trial to speed the drug to market	– Will still need to complete Phase III trials that show clinical benefit	CFDA Order [2014] No. 13 https://www.fdanews.com/PRM0325162 Updated in 2019. https://regulatory.usc.edu/regulatory-updates-special-report-china-september-2019/

TGA – Australia	Provisional Approval	2018	- New prescription medicine or new indications medicine - For treating a serious condition - Favorable comparison against existing therapeutic goods - Major therapeutic advance - Evidence of a plan to submit comprehensive clinical data	- Less comprehensive clinical data - Surrogate endpoints with justification: the ability to predict benefit based on evidence, the strength of the evidence, the certainty of the prediction, and why remaining uncertainties are considered acceptable	- Company must agree to continue clinical trials and submit comprehensive evidence for review - The validity of approved provisional determinations will lapse six months after the determination is granted. Sponsors may apply to the TGA for an extension of determination validity for a further six months	https://www.tga.gov.au/provisional-approval-pathway-prescription-medicines
Health Canada	Notice of Compliance with Conditions	1998	Provide access to promising new drugs for patients suffering from serious, life-threatening or severely debilitating diseases or conditions for which no drug is presently marketed in Canada or for which a significant increase in efficacy or a significant decrease in risk is demonstrated in relation to an existing drug marketed in Canada	- Trials with surrogate markers that require validation; - Phase II trials that would require confirmation with Phase III trials consistent with the normal course of development of a therapeutic entity; or - Phase III trials where a single small to moderately sized trial would require confirmation of either the efficacy or safety of the agent under question. These trials should be replicates of the pivotal trials or trials of different design where the outcomes are congruent with, and complimentary to, those of the original trial	- Confirmatory trials that verify the clinical benefit of the drug - Annual progress reports of confirmatory and other ongoing trials	Guidance on the notice of compliance with conditions

NMPA – China	The Green Channel	2014	Fills 1 of 8 criteria: – Drug applications for products with "significant clinical value," such as innovative drugs and those with advanced formulations; – Clinical trial applications for generic drugs submitted three years before the patent expires on the reference product; – Clinical trial applications for new drugs undergoing parallel review in the U.S. and EU; – Registration applications for drugs undergoing parallel review in the U.S. and EU using the same production line; – Registration applications for new drugs developed under Chinese national research programs; – Registration applications for drugs to treat AIDS, tuberculosis, viral hepatitis, rare diseases and cancer, as well as pediatric and geriatric drugs; – Registration applications for drugs in short supply approved by China's National Health and Family Planning Commission and the Ministry of Industry and Information Technology; and – Registration applications for pediatric drugs that have been approved in the U.S., EU and "surrounding areas" of China, backed by compelling clinical data	– Phase 1 and Phase 2 data if reviewers can reasonably predict or determine the clinical benefit has a significant advantage versus existing treatments – Trial applicants with less convincing Phase 1 and 2 data still may request an abbreviated Phase 3 trial to speed the drug to market	– Will still need to complete Phase III trials that show clinical benefit	CFDA Order [2014] No. 13 https://www.fdanews.com/IPRM0325162 Updated in 2019. https://regulatory.usc.edu/regulatory-updates-special-report-china-september-2019/

VIEWPOINT

New drug approvals in oncology

Razelle Kurzrock, Hagop M. Kantarjian, Aaron S. Kesselheim and Ellen V. Sigal

Abstract | The traditional regulatory drug approval paradigm comprising discrete phases of clinical testing that culminate in a large randomized superiority trial has historically been predominant in oncology. However, this approach has evolved in the current era of drug development, with multiple other development pathways now being utilized. Indeed, treatment approaches designed on the basis of an improved understanding of cancer biology have led to unprecedented responses in early phase trials, sometimes resulting in drug approvals in the absence of large-scale trials. At the same time, improved molecular diagnostic technologies have led to the identification of ever-smaller patient subgroups for molecularly targeted therapy. Moreover, new FDA regulatory paradigms have enabled the rapid review and accelerated approval of certain drugs in the absence of survival data. Regulatory approvals based on large-cohort trials with surrogate or intermediate clinical end points or on non-inferiority trials, as well as new tumour-agnostic indications, also set important precedents in the field. In this Viewpoint, we asked two leading oncologists involved in clinical drug development, an expert in regulatory science and prescription drug policy and a prominent patient advocate, to provide their opinions on the implications of these changes in regulatory practices for patient care.

Q *Are new approvals of antitumour agents over the past 3 years finding the right balance between drug access, safety and efficacy?*

Razelle Kurzrock. Yes.

In recent years, the FDA has recognized that patients suffering from lethal cancers cannot wait years for results from randomized controlled trials (RCTs) in order to gain access to drugs that have already demonstrated high response rates with reassuring safety profiles in phase I/II studies. Even so, objections to this rapid approval pathway exist. The argument that the age-old, gold-standard phase III RCT must always be performed to prevent later medical reversals reflects a not uncommon belief in the logical fallacy that there is a perfect solution to a particular problem and, hence, compares a realistic solution with an idealized one.

In oncology, this idealized strategy misconception spawns the argument that treatment approaches based on an in-depth understanding of cancer biology

that have led to unprecedented responses in early phase trials should not result in regulatory authorization because they are not supported by data from RCTs with survival end points, the latter being the only source of robust data on safety and efficacy. In reality, however, all solutions have advantages and disadvantages, and one can always find individual situations where the solution failed or succeeded. A highly touted example of medical reversal is provided by the experience with autologous bone marrow transplantation for patients with metastatic breast cancer, which was used in thousands of women before this intervention was subjected to a RCT, which failed to demonstrate any survival benefit[1]. On the other hand, delays in the regulatory approval of effective treatments can also cost lives — the experience with trastuzumab emtansine (T-DM1) for patients with breast cancer is a pertinent example. Phase II studies involving 108 patients with metastatic breast cancer showed a remarkable response rate of 34% among patients with HER2-positive tumours, despite these patients having failed

a median of eight prior therapies, including trastuzumab. Yet the FDA declared that the T-DM1 trials did not meet the standard for accelerated approval; it took more than two additional years for this agent to gain approval (in 2013), with an estimated 8,000 patient life-years lost owing to this delay[2]. Therefore, rather than pointing out individual situations in which one approach to regulatory endorsement failed or succeeded, we must weigh the relative pros and cons of different strategies. Indeed, before the current era of faster approvals, we have estimated that, overall, regulatory delays in the development of beneficial cancer remedies resulted in tens to hundreds of thousands of life-years lost, and that the estimated cost per life-year extended by stricter regulations was US$2,700,000 (which is considerably higher than the costs associated with other health measures)[3]. Therefore, the germane question, when discussing anticancer drugs with noteworthy response rates in non-randomized trials, is which approach is likely to save the most lives — is it waiting for the results of RCTs or early regulatory approval without RCTs?

To further address this question, we have previously analysed the long-term consequences for 31 drugs approved by the FDA without an RCT between 1976 and 2006 (REF.[4]). A median of two clinical trials involving a median total of 79 patients (range 40–413) were required per approval. Response rate was the most common surrogate end point (median response rate 33%, range 11–90%). At the time of analysis, 30 drugs were still approved. Ironically, the one drug that lost marketing authorization (gefitinib) did so because of misleading results from a post-marketing RCT, and has since been re-approved for the same indication, albeit with the presence of an *EGFR* mutation as a biomarker. Of the 31 drugs approved, 19 have since gained additional indications and no drug has shown serious safety concerns, indicating that agents that were FDA-approved based on non-randomized trials demonstrated long-term safety and efficacy.

Hagop M. Kantarjian. After more than 38 years as a leukaemia researcher, I find the FDA process for drug approvals improved but inconsistent. This inconsistency, in my view,

using different standards of clinical review judgment, sometimes leading to inconsistent decisions. These decisions can be confusing to both researchers and the pharmaceutical industry, and might reflect an insufficient understanding of the risk that the disease itself poses. For example, in the setting of refractory and/or relapsed acute myeloid leukaemia (AML), patients frequently die of their disease early; in this context, both patients and cancer experts are likely to accept greater levels of risk. In fact, the FDA Code of Federal Regulations (CFR), Section 312.80 of Subpart E–Drugs Intended to Treat Life-Threatening and Severely-debilitating Illnesses, states that "…The Food and Drug Administration (FDA) has determined that it is appropriate to exercise the broadest flexibility in applying the statutory standards, while preserving appropriate guarantees for safety and effectiveness. These procedures reflect the recognition that physicians and patients are generally willing to accept greater risks or side effects from products that treat life-threatening and severely debilitating illnesses…"[5]. Today, I find it unclear how the FDA applies this statute in its decisions on drug approvals for acute leukaemias.

Let me give you two notable examples from the past that I believe should have been different: the 2010 decision to withdraw gemtuzumab ozogamicin, an antibody–drug conjugate (ADC) originally approved in 2000 for patients with CD33-positive AML,

abine as a frontline treatment for elderly patients (defined as those >65 years of age) with AML who are not candidates for intensive chemotherapy[6–10]. Gemtuzumab ozogamicin was re-approved in 2017 based on additional data. Decitabine was approved by the European Medicines Agency (EMA) as a first-line treatment option for elderly patients with AML based on the same clinical data[10]. Another more recent surprising decision was the Oncology Drugs Advisory Committee (ODAC) recommendation that the FDA not approve quizartinib, an effective FLT3 inhibitor, for patients with relapsed and/or refractory *FLT3* internal tandem duplication-positive AML despite the registration RCT meeting the overall survival (OS) end point. I am not aware of any other precedent in oncology where an agent met the primary survival end point of the registration study but was not approved. The ODAC group might have been influenced by some late comments by FDA representatives after closing the discussion.

In general, I find that most of the important research in leukaemia happens in studies of drugs conducted after FDA approval. For example, even at 30–40 years after the approvals of cytarabine and anthracyclines for AML, we are still investigating the optimization of both the doses and administration schedules of these agents. Similarly, most of the BCR–ABL1 tyrosine kinase inhibitors (TKIs) approved for chronic myeloid leukaemia (CML) have

not have happened had these agents not been available to investigators outside the context of industry-sponsored studies. The same is true for the ADC inotuzumab ozogamicin and the bispecific T cell-engager antibody blinatumomab, approved in 2017 and 2014, respectively, as single agents for relapsed and/or refractory acute lymphocytic leukaemia (ALL); investigators are now reconsidering the doses and schedules for these agents and combining them with chemotherapy, resulting in substantially better results in both the salvage and frontline settings[12–16]. The IDH1 inhibitor ivosidenib, the IDH2 inhibitor enasidenib and the FLT3 inhibitor gilteritinib are all also approved as single agents for the treatment of relapsed and/or refractory AML in patients with the respective mutations. Nonetheless, post-approval studies hopefully will soon demonstrate their optimal use in combinations with intensive chemotherapy and with low-intensity epigenetic therapy for patients with relapsed and previously untreated AML, respectively[17,18].

Aaron S. Kesselheim. The need for prescription drugs to be tested for efficacy and safety before they can be widely sold to patients was established over the early part of the twentieth century in response to widespread promotion and use of prescription drugs that lacked any evidence that they improved patient outcomes while having adverse effects. Since 1962, the basic requirement in the USA that prescription drugs show substantial evidence of efficacy arising from adequate and well-controlled trials completed before they can be sold to patients helped ingrain scientifically rigorous practices in drug development, principally the widespread use of RCTs. Attempts to frame this regulatory system as a trade-off between drug access on the one hand and evidence on the other, ignores the reality that these are complementary and not competing interests: most patients do not want access to a product for which no clear evidence exists that it works and that might also be dangerously unsafe.

In the decades since 1962, the application of these basic rules has evolved. One principal change has been in the development of expedited drug approval pathways that reduce clinical testing times. In a review published in 2018, we examined the 58 new cancer drugs approved by the FDA between 2012 and 2017. Nearly all of them (55 of 58, 95%) qualified for an expedited development or approval pathway, including

The contributors

R.K. is the Director of the Center for Personalized Cancer Therapy and founder/director of the Rare Tumor Clinic (as well as Associate Director of the Clinical Science division and Leader of the Experimental Therapeutics Program) at the University of California San Diego Moores Cancer Center. Previously, at MD Anderson Cancer Center, she founded, built up and chaired one of the world's largest clinical trials departments. Her expertise and interests are in patient-focused precision medicine.

H.M.K. is the Chair of the Leukaemia Department at MD Anderson Cancer Center. He joined the institution in 1981 as a fellow and over the past 38 years has continued his clinical-translational research in leukaemia. He is also a Special Fellow on Healthcare Policies at the Rice Baker Institute and advocates for health-care issues pertinent to patients with cancer and leukaemia.

A.S.K. is a Professor of Medicine at Harvard Medical School and Director of the Program On Regulation, Therapeutics, And Law (PORTAL) in the Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's Hospital. PORTAL is an interdisciplinary research centre focusing on intersections among prescription drugs and medical devices, patient health outcomes, and regulatory practices and the law.

E.V.S. is the Chairperson and Founder of Friends of Cancer Research, and in addition is the Chair of the Board of Directors at the Reagan–Udall Foundation for the FDA, board member of the Foundation for the National Institute of Health, where she chairs the Portfolio Oversight Committee, and is on the Board of Governors of the Patient-Centered Outcomes Research Institute (PCORI). She also holds leadership positions with a broad range of cancer advocacy, public policy organizations and academic health centres including: MD Anderson Cancer Center External Advisory Board, the Duke University Cancer Center Board of Overseers, and The Sidney Kimmel Comprehensive Cancer Center Advisory Council.

priority review (46, 79%), accelerated approval (26, 45%), fast-track approval (28, 48%), and breakthrough therapy status (25, 43%). Qualifying for expedited approval pathways substantially shortened the median time from designation as an Investigational New Drug (IND) to approval, which is the period marking the initiation of clinical trials and regulatory review. For example, the 11 oncology drugs qualifying for priority review, accelerated approval and breakthrough therapy designation had a median trial and review time of 4.8 years (interquartile range (IQR) 3.9–7.3 years) while the six that qualified for priority review, accelerated approval, fast-track, and breakthrough therapy designation also had a median trial and review time of 4.8 years (IQR 2.7–8.1 years), compared with a median of about 7 years for the whole sample[19].

Some studies of drugs approved after expedited development or FDA review have found an association with increased risks of reports of spontaneous adverse events, while others have not[20,21]. For example, in a matched retrospective cohort study involving all new drug approvals in the USA from 1997 to 2016, we found that expedited pathway drugs had a 48% higher rate of change to boxed warning and contraindications, the two most clinically important categories of safety warnings[22]. Of course, patients might accept such an increased risk for drugs designed to treat life-threatening conditions for which no other better-studied options exist, but other research has shown that the increased application of expedited development and review programmes in the past few decades has largely come from their application to drugs that are not first-in-class agents[23].

Ellen V. Sigal. For the FDA to approve any drug, it must be deemed safe and effective. The FDA's primary focus is on serving the needs of patients. As such, the FDA has found an appropriate balance that ensures that new drugs are approved and available to critically ill patients who often have few or even no other treatment options available. Furthermore, the FDA has worked diligently over the past decade to incorporate the patient voice throughout the regulatory process and actively engages with patients and advocacy groups.

Regulatory pathways have been adapted to evolve with and to facilitate the rapid developments in science, trial designs and drug products without compromising FDA standards of safety and efficacy. The FDA's

Oncology Center of Excellence is a prime example of how the FDA is taking an integrated and hands-on approach to the advancement of drug development and regulatory policy.

The FDA has been tremendously successful at leveraging available regulatory pathways, such as the breakthrough therapy designation and accelerated approval, to ensure timely access to treatments and address unmet medical needs for patients faced with serious or life-threatening illnesses. We have seen remarkable progress in the types of therapies and the diseases they treat. Over the past 3 years, the FDA has noted the accelerated approval of 14 novel oncology drugs, including the immune checkpoint inhibitor (ICI) atezolizumab as the first new treatment for urothelial carcinoma in more than several decades, the approval of another ICI, pembrolizumab, as a tissue-agnostic drug targeting a rare alteration present across several different types of solid tumour, and a new ALK targeted therapy, lorlatinib, for patients with non-small-cell lung cancer with resistance to previously approved TKIs[24–26]. The FDA has also ushered in a new era of treatments with the expedited approvals of several cell and gene therapies that have curative potential for a variety of life-threatening diseases.

> **Q** *In what situations are approvals based on non-inferiority and/or surrogate end points appropriate?*

R.K. The most clear-cut condition for approval on the basis of non-inferiority is when the new compound or intervention provides some other benefit that might not be easy to quantify in a RCT. For example, if one compares surgery versus a specific drug, non-inferiority approval of the drug provides an option with very distinct morbidity risks, as compared with surgery. On the other hand, it can be argued that no two drugs have precisely the same toxicity profile and, hence, non-inferiority almost always provides additional options that might each be preferable for different subsets of patients. Furthermore, even if a drug does not differ in terms of clinical efficacy from that of an existing product, the approval of both agents can foster competition that, theoretically, should lower drug prices.

Surrogate end points are also a matter of debate. Use of response rate or progression-free survival (PFS) as an alternative to OS reduces the time needed to develop a drug by ~1–1.5 years[27]. Once again, however, one must determine the trade-off between lives saved by greater access (in addition to

reduced drug development costs spurring innovation) and the lives lost by misguided early approvals. Response rates in early phase trials do not necessarily predict a successful RCT, although an extensive body of literature exists suggesting that response rate and PFS can be useful surrogates, especially when used with threshold values[28,29].

H.M.K. The answer depends on the context. For novel drugs with equal levels of efficacy but better safety than the FDA-approved equivalent, a non-inferiority trial is reasonable. For example, having more than one BCR–ABL1 TKI targeting *T315I*-mutated CML is important, both to offer a less-toxic alternative to ponatinib and to break the market monopoly. On the other hand, targeted novel agents measured against standard-of-care chemotherapy might need to demonstrate superior clinical efficacy owing to the higher levels of uncertainty regarding specific toxicities, and greater costs.

The sum of improvements in several surrogate markers is important to expedite approval, particularly in areas of greatest need and in rare tumours, for which large-cohort RCTs might take a decade to accrue and complete. OS should not be the only end point that matters.

A.S.K. In addition to shorter development times, many oncology drug approvals that have moved through expedited approval pathways have used surrogate outcome measures, such as response rates and PFS (of course, oncology drugs that do not qualify for expedited pathways also frequently use surrogate measures in pivotal trials for regulatory approval). However, these measures are often not well correlated with the actual clinical end points, such as OS or quality of life, for which they are supposed to be surrogates[30]. When this reliance on surrogate measures is combined with the frequency of non-blinded, non-randomized trials among studies of new cancer drugs, this raises concerns about the robustness of the findings of such trials and makes clinical improvement less certain for patients[31].

FDA statements and a National Academy of Medicine panel have provided three criteria for when surrogate measures are most appropriate to be used as the primary end point in a clinical trial designed to support regulatory approval: (1) for a chronic condition, in which direct clinical outcomes cannot be measured in the time frame of a trial; (2) for a serious or life-threatening condition; and (3) for trials in which the experimental therapy is expected

to offer some substantial benefit over available therapy[32,33]. Cancer is certainly a serious or life-threatening condition and new treatments are often tested in the context of metastatic disease, a scenario in which using a surrogate measure might not substantially reduce the drug development time. A recent study found that the use of PFS as the primary end point in cancer drug trials shaved only 11 months (12%) off the average development time of 7.3 years, while the use of response rate reduced the development time by 19 months[27]. Surrogate measures are also frequently used in cancer drug trials involving agents that might not be expected to offer substantial benefit over available therapies. For example, we found that trials designed to test second-generation cancer drugs — a drug with the same mechanism of action developed in the same disease category as an original innovative product — commonly used surrogate measures as end points[34].

Non-inferiority trials can be another controversial technique when used as the basis for cancer drug approvals. Clinical trials using non-inferiority hypotheses are designed to evaluate whether test interventions are 'non-inferior' in terms of efficacy compared to control interventions, usually the current standard of care, by a certain predefined margin. Such trials might provide the manufacturer with a greater level of assurance that testing will lead to an affirmative result than a trial with a superiority design, but these trials are most clinically and ethically appropriate when somewhat decreased levels of effectiveness would not result in serious patient harm and when new interventions might offer benefits to patients in terms of outcomes other than improved effectiveness, such as fewer adverse events or improved dosing convenience[35]. However, in the case of the TKI lenvatinib, a non-inferiority trial was used to approve this drug for patients with hepatocellular carcinoma, despite no clear advantage in terms of ease of administration, safety profile or even price that would justify a potential reduction in efficacy[36]. In such cases, only accelerated approval should be acceptable, to ensure that a post-approval study confirming the efficacy of lenvatinib is required.

E.V.S. The FDA is tasked with ensuring no undue harm is brought to patients, and they ultimately must answer the question: is the American public better or worse off with access to a specific drug? While many have criticized the FDA's approach to drug approvals, we must ask ourselves how many patients might die waiting for one more

chance because we required a statistically significant *P* value for OS for a potentially transformative drug that could have been assessed earlier using available intermediate clinical end points. Accelerated approval pathways exist to avoid these scenarios by capitalizing on our increasing understanding of surrogate end points and trial analyses that enable earlier access to drugs designed to treat serious or life-threatening diseases while additional confirmatory evidence is generated.

Critically ill patients, particularly those with relapsed and/or refractory disease, often have very limited or even no choices with evidence of prolonged survival after exhausting currently available therapies. We know that not every new drug will be the silver bullet. But in many cases, even drugs that provide small improvements in survival might extend life long enough for the next breakthrough that might be even more effective. The gold standard for the approval of new drugs is a RCT in which the new intervention is demonstrated to provide clinical benefit compared with an existing standard of care, if one exists. Metrics designed to assess clinical benefit should be tailored to individual cancers and drugs based on the specific needs faced by each patient population and, when appropriate, could rely on non-inferiority trials or surrogate end points. Patients, along with their physicians and families, must assess the available evidence supporting the use of a drug in light of the individual preferences of the patient when making treatment choices.

Q *What should be the minimum requirements for post-marketing studies?*

R.K. The first requirement is to further assess the risks of serious adverse effects, since all possible drug toxicities cannot be anticipated based on pre-approval studies involving only small numbers of patients. Further information about an already known risk of serious adverse events from the approved agent might also need to be acquired. Currently, physicians and other health-care professionals, patients and their families, and manufacturers can voluntarily submit reports to the FDA Adverse Event Reporting System (FAERS), which is a database designed to document adverse events, medication errors and product quality complaints. Any manufacturer that receives a report is mandated to direct the report to the FDA — this system should continue.

In addition, clinical trials often have eligibility exclusions, and more data on

the performance of the drug in non-ideal patients, who are more likely to have comorbidities, might be helpful, regardless of whether or not a RCT led to approval. A real-world registry-based and/or observational study examining safety and efficacy could be the most efficient way to collect these data. Post-marketing surveillance should also outline adverse events during longer follow-up and/or new analyses of data from existing clinical trials.

H.M.K. Post-marketing studies designed to satisfy particular FDA requirements should demonstrate some level of benefit for the patient, but this should be taken together with the entire body of evidence available, which in many instances extends beyond one post-marketing FDA-sanctioned study being evaluated. For example, the pivotal negative Southwest Oncology Group (SWOG) RCT of gemtuzumab ozogamicin for patients with AML led to its withdrawal from the market in 2010, but data from multiple more recent studies suggested benefit, resulting in its re-approval in 2017.

Post-marketing studies designed to enable expanded utilization of an FDA-approved agent beyond the specifics of its approval are, I would argue, as important if not more important than the studies that led to the original approval (see above for examples relating to inotuzumab ozogamicin and blinatumomab). In addition, such studies do not need to be of the stringent, randomized, large-scale designs required by the FDA. In this context, well-designed, sizeable, single-arm or 'other-design' trials, conducted by reputable investigators or research groups, published in peer-reviewed journals and reproduced by other groups using a similar concept (rather than simply conducting an identical study) would be sufficient, in my view, to establish new standards of care.

A.S.K. Post-approval confirmatory testing should be routine among all new cancer drugs that are approved based on unvalidated surrogate measures, atypical non-inferiority trials or other expedited testing pathways that use data from non-randomized prospective trials. However, under current rules, such confirmatory trials can only be mandated under the accelerated approval pathway. We have found that even with agents approved via this pathway, post-approval confirmatory studies can be delayed and are often designed to test, as trial end points, the same surrogate measures as the pre-approval studies[37,38]. At a minimum, post-approval confirmatory trials should be organized and have begun to recruit

patients by the time the drug is approved, and continued unfettered marketing of the drug should be contingent on their timely completion. The price of the drug should also be discounted to reflect the lack of firm evidence on the product and to provide an additional incentive to complete the trials (it can then be raised to the commensurate level when such evidence emerges)[39].

E.V.S. As defined by statute, accelerated approval of a drug is "subject to the requirement that the applicant study the biological product further, to verify and describe its clinical benefit, where there is uncertainty as to the relation of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome."[40]

Sponsors should be held to their post-market requirements when accelerated approval is granted; however, challenges can occur when conducting these studies over time. Despite thoughtful trial designs and the selection of meaningful end points, scenarios in which a drug is available for off-label use, drugs with similar mechanisms of action are made available, or a newer drug is approved that is perceived to be better, can all affect the ability to accrue or retain patients on post-marketing confirmatory studies. When options are limited or non-existent for critically ill patients or patients with rare diseases, access to an experimental intervention might be the only remaining option. Patients and physicians might in particular seek trials with patient-friendly designs, such as trials that allow crossover to the experimental drug at the time of progression.

Together, these real-life scenarios can make randomization difficult, or essentially impossible, and can lead to challenges to patient recruitment that have real consequences for the interpretation of survival end points and final study results. Minimum requirements are likely to differ by disease setting and drug type; therefore, we should be thinking about innovative ways to maximize the amount and quality of the information that is gleaned from these studies and how best to maximize the relevance of the data to patients receiving the therapy.

Q *What future steps should investigators and regulators take to address these challenges?*

R.K. (1) Genomic alterations and failed immune responses underlie cancer. The organ of origin and tumour histology provide a secondary context to these abnormalities. Therefore, genomic and/or immune biomarker-based, organ-of-origin and/or histology-agnostic studies for tissue-agnostic indications should be pursued, rather than just traditional histology-based, genomic and/or immune biomarker-agnostic studies and indications[41,42]. (2) Clinical trials should have limited exclusion criteria so that participants better reflect the type and characteristics of patients who will receive the drug or regimen after approval. (3) Approaches involving real-world data and artificial intelligence should be developed to help elucidate the emergence of toxicities after marketing and better validate the effectiveness of experimental compounds, as well as provide a foundation for a new pathway towards drug approval[43].

H.M.K. Investigators and regulators should take three critically important steps: (1) simplify the regulatory burdens; (2) simplify the stringency of the research requirements, which make trials substantially more expensive without adding to the quality of research or improving patient safety; and (3) eliminate the need for clinical research organizations (CROs) outside pivotal phase III approval trials.

The current level of research bureaucracy is paralysing cancer research and slowing the pace of much-needed discoveries. This extra administration is burdening the research process (such as with expensive CROs), discouraging and penalizing investigators and increasing the costs of clinical research (and, consequently, the price of drugs and health care), without enhancing patient safety or research quality. After all, the International Council for Harmonisation (ICH) guidelines for Good Clinical Practice (GCP)[44] and the CFR guidelines concerning the conduct of human research have not changed substantially in 50 years (with the exception of the recent Common Rule changes)[45]. What has changed is the interpretation of these guidelines by different regulatory bodies, including the FDA, pharmaceutical industry, academia and CROs, each adding layer upon layer of unnecessarily stifling bureaucracy and paperwork.

A prime example is the requirement under the CFR, Section 312.64 of Subpart D–Responsibilities of Sponsors and Investigators, that "an investigator must immediately report to the sponsor any serious adverse event whether or not considered drug related, including those listed in the protocol or investigator brochure and must include an assessment of whether there is a reasonable possibility that the drug caused the event."[46] Also, GCP 4.11.1 states that "All serious adverse events (SAEs) should be reported immediately to the sponsor except for those SAEs that the protocol or other document (e.g., Investigator's Brochure) identifies as not needing immediate reporting."[47] Almost every patient with acute leukaemia is hospitalized with neutropenic fever, as many begin therapy with neutropenia and others develop it with treatment; however, because the definition of a SAE includes any hospitalization, we must report these hospitalizations within 24 hours even though they are expected events. This stipulation does not enhance patient safety. Reporting an SAE 25 hours after its occurrence results in a protocol violation, which then must be reported to both the sponsor and the Institutional Review Board. In other words, a vicious cycle of reporting exists that is punitive, time-consuming and wasteful. In my department alone, we have hired ten extra research staff to manage this regulatory reporting burden. For investigator-initiated studies in which the University of Texas MD Anderson Cancer Center is the sponsor, the interpretation of CFR Section 312.64 is vastly different from that of trials with pharmaceutical industry sponsors and those involving CROs, and is more reasonable. CROs sometimes charge upward of US$3,000 each to process an SAE, which means that for them, regulatory compliance is a money-making venture. Equally important is the concern among cancer researchers that monitoring regulatory compliance has become more of a game of 'gotcha' than a reasonable form of oversight designed to protect patient safety and ensure rigorous conduct of research. No common-sense consideration of regulations seems to exist, only 'regulatory fundamentalism' with the principal investigator as the target.

The fact that many drugs have become the standard of care in cancers for which they were not initially approved attests to the need for a 'reality check' and re-evaluation of the approval and regulatory processes. Examples include decitabine and azacitidine, both of which were originally approved for one indication, higher-risk myelodysplastic syndrome, but have also become the standards of care for many years in elderly patients with AML, and sorafenib, which was initially approved for several types of solid tumour, but is now also a standard of care for patients with *FLT3*-mutated AML.

The current research paradigm, which has shifted the fulcrum of cancer research into the hands of the drug companies,

producing the 'one drug, one disease' approach, is cumbersome, slow, extremely expensive, produces a low success rate and reduces the length of time that a patented drug is available after approval. A better strategy, proposed in 2018, is to shift clinical research and development back into the hands of experts in the form of 'research/academia–drug industry alliances'[48]. These initiatives have successfully shortened the duration of research programmes, reduced costs and produced better success rates, thus potentially prolonging the drug patent times, constituting a win–win situation for all involved: patients, cancer researchers and drug companies. Examples of drugs developed in this way include inotuzumab ozogamicin and venetoclax.

A.S.K. Congress and the FDA should enact legislative and regulatory changes that better ensure that post-approval confirmatory trials are completed in a timely fashion and are designed to collect data that are optimally useful to patients and their physicians. The regulatory system should ensure that investigational products are tested to a reasonable standard and then made available to patients, and should be flexible enough to prioritize the approval of particularly promising products that are designed to address unmet medical needs. However, the regulatory system should also be empowered to mandate that all such products are eventually tested in high-quality trials so that patients and physicians can learn which products offer them the best chance of success. Currently, only drugs approved under the accelerated approval framework are required to be tested in such trials, and even those drugs have not consistently met that goal.

E.V.S. Despite the best intentions of drug sponsors and investigators, challenges exist that can affect the ability to optimally assess clinical benefit in post-marketing studies. However, opportunities exist to maximize the availability of information on safety and efficacy throughout the course of a drug's life cycle to help guide treatment decisions and ultimately ensure that patients receive the right drug at the right time, based on their needs.

The FDA recently released a series of draft guidance documents recommending the broadening of trial eligibility criteria to enable more patients to participate in trials and improve the generalizability of findings to those most likely to receive the drug[49]. By including a more representative patient population or conducting subset analyses in these early clinical trials that could serve as the basis for an accelerated approval, important information can be made available sooner, in order to guide important patient–physician discussions while confirmatory studies are underway. Additionally, leveraging additional data sources, such as real-world evidence or historical clinical trial data, to complement data from RCTs can help in rapidly expanding the available evidence and filling in important information gaps[50].

We will never eliminate all risks associated with receiving a drug. However, there are important steps we can take as a community, such as optimizing the amount of information available early in the life cycle of a drug, in order to help identify patients who are most likely to benefit and to meaningfully examine and characterize the role of complementary data sources to support RCTs and long-term outcome analyses.

Razelle Kurzrock[1]*, Hagop M. Kantarjian[2]*, Aaron S. Kesselheim[3]* and Ellen V. Sigal[4]*

[1]*Center for Personalized Cancer Therapy and the Division of Hematology and Oncology, Department of Medicine, University of California San Diego Moores Cancer Center, La Jolla, CA, USA.*

[2]*Department of Leukemia, University of Texas MD Anderson Cancer Center, Houston, TX, USA.*

[3]*Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.*

[4]*Friends of Cancer Research, Washington, DC, USA.*

**e-mail: rkurzrock@ucsd.edu; hkantarjian@mdanderson.org; akesselheim@bwh.harvard.edu; esigal@focr.org*

1. Stadtmauer, E. A. et al. Conventional-dose chemotherapy compared with high dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. Philadelphia Bone Marrow Transplant Group. *N. Engl. J. Med.* **342**, 1069–1076 (2000).
2. Kurzrock, R., Kantarjian, H. & Stewart, D. J. A cancer trial scandal and its regulatory backlash. *Nat. Biotechnol.* **32**, 27–31 (2014).
3. Stewart, D. J., Whitney, S. N. & Kurzrock, R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. *J. Clin. Oncol.* **28**, 2925–2935 (2010).
4. Tsimberidou, A. M., Braiteh, F., Stewart, D. J. & Kurzrock, R. Ultimate fate of oncology drugs approved by the US Food and Drug Administration without a randomized trial. *J. Clin. Oncol.* **27**, 6243–6250 (2009).
5. FDA. *CFR - Code of Federal Regulations Title 21* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=312&showFR=1&subpartNode=21:5.0.1.1.3.5 (2019).
6. Ravandi, F. & Kantarjian, H. Gemtuzumab ozogamicin in acute myeloid leukaemia. *Nat. Rev. Clin. Oncol.* **9**, 310–311 (2012).
7. Ravandi, F. et al. Gemtuzumab ozogamicin: time to resurrect? *J. Clin. Oncol.* **30**, 3921–3923 (2012).
8. Kantarjian, H. et al. Decitabine in older adults with acute myeloid leukemia: why was the dream broken? *J. Clin. Oncol.* **31**, 1795–1796 (2013).
9. Kantarjian, H. & Estey, E. Drug approvals in acute myeloid leukemia: can we do better? *The ASCO Post* https://www.ascopost.com/issues/may-15-2013/drug-approvals-in-acute-myeloid-leukemia-can-we-do-better/ (2013).
10. Kantarjian, H. M. et al. Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia. *J. Clin. Oncol.* **30**, 2670–2677 (2012).
11. Naqvi, K. et al. Early results of lower dose dasatinib (50 mg daily) as frontline therapy for newly diagnosed chronic-phase chronic myeloid leukemia. *Cancer* **124**, 2740–2747 (2018).
12. Kantarjian, H. M. et al. Inotuzumab ozogamicin versus standard therapy for acute lymphoblastic leukemia. *N. Engl. J. Med.* **375**, 740–753 (2016).
13. Kantarjian, H. et al. Inotuzumab ozogamicin in combination with low-intensity chemotherapy for older patients with Philadelphia chromosome-negative acute lymphoblastic leukaemia: a single-arm, phase 2 study. *Lancet Oncol.* **19**, 240–248 (2018).
14. Jabbour, E. et al. Salvage chemoimmunotherapy with inotuzumab ozogamicin combined with mini-hyper-CVD for patients with relapsed or refractory Philadelphia chromosome-negative acute lymphoblastic leukemia: a phase 2 clinical trial. *JAMA Oncol.* **4**, 230–234 (2018).
15. Jabbour, E. et al. Chemoimmunotherapy with inotuzumab ozogamicin combined with mini-hyper-CVD, with or without blinatumomab, is highly effective in patients with Philadelphia chromosome-negative acute lymphoblastic leukemia in first salvage. *Cancer* **124**, 4044–4055 (2018).
16. Kantarjian, H. & O'Brien, S. Poor penetration of existing effective chemoimmunotherapy for the treatment of older patients with newly diagnosed acute lymphocytic leukemia (ALL) or refractory/relapsed ALL. *J. Oncol. Pract.* **15**, 77–79 (2019).
17. Stein, E. et al. Ivosidenib or enasidenib combined with induction and consolidation chemotherapy in patients with newly diagnosed AML with an IDH1 or IDH2 mutation is safe, effective, and leads to MRD-negative complete remissions. *Blood* **132**, 560 (2018).
18. DiNardo, C. D. et al. Mutant IDH (MIDH) inhibitors, ivosidenib or enasidenib, with azacitidine (AZA) in patients with acute myeloid leukemia (AML). *J. Clin. Oncol.* **36** (suppl.), 7042–7042 (2018).
19. Hwang, T. J. et al. Efficacy, safety, and regulatory approval of Food and Drug Administration-designated breakthrough and nonbreakthrough cancer medicines. *J. Clin. Oncol.* **36**, 1805–1812 (2018).
20. Olson, M. K. The risk we bear: the effects of review speed and industry user fees on new drug safety. *J. Health Econ.* **27**, 175–200 (2008).
21. Grabowski, H. & Wang, Y. R. Do faster Food and Drug Administration drug reviews adversely affect patient safety? An analysis of the 1992 Prescription Drug User Fee Act. *J. Law Econ.* **51**, 377–406 (2008).
22. Mostaghim, S., Gagne, J. J. & Kesselheim, A. S. Safety-related label changes for new drugs after approval in the US through expedited regulatory pathways: retrospective cohort study. *BMJ* **358**, j3837 (2017).
23. Kesselheim, A. S., Wang, B., Franklin, J. M. & Darrow, J. J. Trends in utilization of FDA expedited drug development and approval programs, 1987–2014: cohort study. *BMJ* **351**, h4633 (2015).
24. FDA. *2016 novel drugs summary* https://www.fda.gov/media/102618/download (2017).
25. FDA. *Advancing health through innovation: 2017 new drug therapy approvals* https://www.fda.gov/media/110526/download (2018).
26. FDA. *Advancing health through innovation: 2018 new drug therapy approvals* https://www.fda.gov/media/120357/download (2019).
27. Chen, E. Y., Joshi, S. K., Tran, A. & Prasad, V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. *JAMA Intern. Med.* **179**, 642–647 (2019).
28. Johnson, K. R. et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol.* **7**, 741–746 (2006).
29. Kurzrock, R. & Stewart, D. J. Equipoise abandoned? Randomization and clinical trials. *Ann. Oncol.* **24**, 2471–2474 (2013).

30. Kim, C. & Prasad, V. Strength of validation for surrogate end points used in the US Food and Drug Administration's approval of oncology drugs. *Mayo Clin. Proc.* **91**, 713–725 (2016).

31. Kesselheim, A. S., Myers, J. A. & Avorn, J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. *JAMA* **305**, 2320–2326 (2011).

32. [No authors listed]. New drug, antibiotic, and biological drug product regulations; accelerated approval--FDA. *Fed. Reg.* **57**, 13234–13242 (1992).

33. Micheel, C. M. & Ball, J. R. (eds) *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* (National Academies Press, 2010).

34. Vokinger, K. N. & Kesselheim, A. S. Characteristics of trials and regulatory pathways leading to US approval of innovative vs. non-innovative oncology drugs. *Health Policy* **123**, 721–727 (2019).

35. Powers, J. H., Evans, S. R. & Kesselheim, A. S. Studying new antibiotics for multidrug resistant infections: are today's patients paying for unproved future benefits? *BMJ* **360**, k587 (2018).

36. Gyawali, B. & Kesselheim, A. S. FDA approval of new drugs based on non-inferiority trials in oncology: a dangerous precedent? *JAMA Oncol.* **5**, 607–608 (2019).

37. Naci, H., Smalley, K. R. & Kesselheim, A. S. Characteristics of preapproval and postapproval studies for drugs granted accelerated approval by the US Food and Drug Administration. *JAMA* **318**, 626–636 (2017).

38. Gyawali, B., Hey, S. P. & Kesselheim, A. S. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. *JAMA Intern. Med.* **179**, 906–913 (2019).

39. Gellad, W. F. & Kesselheim, A. S. Accelerated approval and expensive drugs — a challenging combination. *N. Engl. J. Med.* **376**, 2001–2004 (2017).

40. FDA. *CFR - Code of Federal Regulations Title 21* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=601.41 (2019).

41. Stewart, D. J. & Kurzrock, R. Cancer: the road to Amiens. *J. Clin. Oncol.* **27**, 328–333 (2009).

42. Subbiah, V. & Kurzrock, R. The marriage between genomics and immunotherapy: mismatch meets its match. *Oncologist* **24**, 1–3 (2019).

43. Shah, P. et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit. Med.* **2**, 69 (2019).

44. European Medicines Agency. *ICH E6 (R2) Good clinical practice* https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice (2016).

45. FDA. *CFR - Code of Federal Regulations Title 21* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm (2019).

46. FDA. *CFR - Code of Federal Regulations Title 21* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=312.64 (2019).

47. Good clinical practice network. *4. Investigator* https://ichgcp.net/4-investigator/ (2016).

48. Kantarjian, H. M. et al. Cancer research in the United States: a critical review of current status and proposal for alternative models. *Cancer* **124**, 2881–2889 (2018).

49. Kim, E. S. et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research joint research statement. *J. Clin. Oncol.* **35**, 3737–3744 (2017).

50. Sherman, R. E. et al. Real-world evidence — what is it and what can it tell us? *N. Engl. J. Med.* **375**, 2293–2297 (2016).

Disclaimer

All comments by Hagop Kantarjian are his personal opinions and do not reflect the opinions of his institutional affiliation.
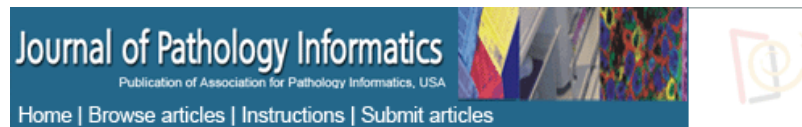
Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS

FDA Adverse Event Reporting System: https://open.fda.gov/data/faers/

A Regulatory Science Initiative to Harmonize and Standardize Digital Pathology and Machine Learning Processes to Speed up Clinical Innovation to Patients

Hetal Desai Marble,[1] Richard Huang,[1] Sarah Nixon Dudgeon,[2] Amanda Lowe,[3] Markus D. Herrmann,[4] Scott Blakely,[5] Matthew O. Leavitt,[6] Mike Isaacs,[7] Matthew G. Hanna,[8] Ashish Sharma,[9] Jithesh Veetil,[10] Pamela Goldberg,[10] Joachim H. Schmid,[11] Laura Lasiter,[12] Brandon D. Gallas,[2] Esther Abels,[13] and Jochen K. Lennerz[1]

[1]Department of Pathology, Center for Integrated Diagnostics, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA

[2]Division of Imaging, Diagnostics, and Software Reliability, Center for Devices and Radiological Health, Food and Drug Administration, Office of Science and Engineering Laboratories, Silver Spring, MD, USA

[3]Visiopharm Americas, Westminster, CO, USA

[4]Department of Pathology, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA

[5]Hamamatsu Corporation, Pittsburgh, PA, USA

[6]LUMEA, Salt Lake City, UT, USA

[7]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA

[8]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[9]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA

[10]Medical Device Innovation Consortium, Arlington, VA, USA

[11]Roche Tissue Diagnostics, Santa Clara, USA

[12]Friends of Cancer Research, Washington, DC, USA

[13]PathAI, Boston, MA, USA

Address for correspondence: Dr. Jochen K. Lennerz, Department of Pathology, Center for Integrated Diagnostics, Harvard Medical School, Massachusetts General Hospital, 55 Fruit Street, GRJ1015 Boston, MA 02114, USA. E-mail: JLennerz@partners.org

Abstract

Unlocking the full potential of pathology data by gaining computational access to histological pixel data and metadata (digital pathology) is one of the key promises of computational pathology. Despite scientific progress and several regulatory approvals for primary diagnosis using whole-slide imaging, true clinical adoption at scale is slower than anticipated. In the U.S., advances in digital pathology are often siloed pursuits by individual stakeholders, and to our knowledge, there has not been a systematic approach to

OPTIMIZING DRUG DEVELOPMENT: IDENTIFYING OPPORTUNITIES FOR MODERNIZATION AND INNOVATION

OPTIMIZING DRUG DEVELOPMENT: IDENTIFYING OPPORTUNITIES FOR MODERNIZATION AND INNOVATION

3/11/2021   A Regulatory Science Initiative to Harmonize and Standardize Digital Pathology and Machine Learning Processes to Speed up Clinical In…

advance the field through a regulatory science initiative. The Alliance for Digital Pathology *(the Alliance)* is a recently established, volunteer, collaborative, regulatory science initiative to standardize digital pathology processes to speed up innovation to patients. The purpose is: (1) to account for the patient perspective by including patient advocacy; (2) to investigate and develop methods and tools for the evaluation of effectiveness, safety, and quality to specify risks and benefits in the precompetitive phase; (3) to help strategize the sequence of clinically meaningful deliverables; (4) to encourage and streamline the development of ground-truth data sets for machine learning model development and validation; and (5) to clarify regulatory pathways by investigating relevant regulatory science questions. The *Alliance* accepts participation from all stakeholders, and we solicit clinically relevant proposals that will benefit the field at large. The initiative will dissolve once a clinical, interoperable, modularized, integrated solution (from tissue acquisition to diagnostic algorithm) has been implemented. In times of rapidly evolving discoveries, scientific input from subject-matter experts is one essential element to inform regulatory guidance and decision-making. The *Alliance* aims to establish and promote synergistic regulatory science efforts that will leverage diverse inputs to move digital pathology forward and ultimately improve patient care.

Keywords: Artificial intelligence, digital pathology, machine learning, regulatory science, slide scanning

INTRODUCTION

"The scientist and science provide the means, the politician and politics decide the ends."

-Alvin M. Weinberg[1]

Regulatory science is an established discipline that entails the application of the scientific method to support regulatory and other policy objectives.[2] Simply put, when medical research provides a novel solution to a health need, regulatory science applies the scientific method to assess benefits and risks before marketing for clinical use. To assess benefits and risks, regulatory scientists develop new tools, standards, and approaches to evaluate the effectiveness, safety, and quality of medical products. A primary challenge in the field of digital pathology is the lack of understanding that strong relationships between regulatory, basic, and translational scientists can substantially improve clinical innovation.[3,4,5,6] For example, regulatory science is not restricted to regulatory agencies.[2,4,5,6] As a scientific discipline, regulatory science challenges current concepts of benefit and risk assessments, submission and approval strategies, patient involvement, and various ethical aspects. Regulatory science includes the creation of a scientific dialog for launching new ideas – not only derived from industry and regulatory authorities but also by, for example, academics, clinicians, and patients.[7] It has been recognized that regulatory science can have a significant impact in bringing new devices to patients in need.[7]

Here, we outline a recently established, volunteer, collaborative regulatory science initiative termed the Alliance for Digital Pathology (the *Alliance*). To prevent confusion, our intent is to familiarize the community with the aims, scope, and rationale of the *Alliance*. The *Alliance* aims to move the field of digital pathology forward by systematically assessing relevant aspects and providing publicly available resources (e.g., data, tools, and methods) to inform and improve the relevant regulatory guidance landscape.[8] Our premise (thesis) is that the *Alliance* promotes regulatory science as a bridge between digital pathology (the means) and moving the field of diagnostic pathology forward (the ends). By promoting regulatory science, the *Alliance* helps to unlock the potential of new technologies and thereby overcomes the dichotomy illustrated in the epigraph by Dr. Weinberg.[1]

TOWARD AN OPERATIONAL DEFINITION OF A CLINICAL, INTEROPERABLE, INTEGRATED SOLUTION FOR DIGITAL PATHOLOGY

workflows into interoperable, digitally enhanced solutions by contributing regulatory science deliverables that can be used to inform and improve the applicable regulatory guidance landscape. Numerous groups have attempted to specify the relevant components of digital pathology solutions; [9,10,11,12,13,14,15,16,17,18] however, given the modularized nature of diagnostic pathology, defining the specific scope of a digital pathology solution is highly context dependent. For example, the variability of a stain (e.g., hematoxylin and eosin across or within laboratories) may influence the performance of a downstream mutation prediction algorithm.[19,20,21] In this example, one may consider drawing an arbitrary boundary before the staining step; however, the fixation and processing method (e.g., formalin fixed, paraffin embedded) or even the tissue acquisition, handling, or image acquisition[22] may influence the performance of the predictor as well. Thus, for the purpose of the *Alliance*, we considered three descriptors for the solution. First, we aim toward a clinical (as opposed to a research-based) solution. Second, due to the modularized nature of the various subprocesses within the main workflows in pathology, we aim for interoperability of systems. Third, to account for the various and arbitrary boundaries of workflow steps (modules) and technologies relevant for a given task (intended use), we consider every step, from the medical procedure acquiring the cell or tissue sample all the way to the fully integrated diagnostic output (e.g., report or model output), as relevant. As opposed to an end-to-end solution, where the supplier of an application or system will provide all the hardware and/or software to meet specific requirements, we are aiming for modularized solutions within the main workflow. We refer to these three solution descriptors (clinical, interoperable, and modularized) as an "integrated solution" for digital pathology. We acknowledge that this definition is operational and arguably incomplete yet represents a technique that enables flexible modeling to solve challenging problems.[23,24,25,26]

THE MULTIFACETED NATURE OF DIGITAL PATHOLOGY NEEDS INCREASED REGULATORY CLARITY

Digital pathology has grown into a multimillion-dollar vendor landscape,[27] and the application of machine learning algorithms holds big promise for improving diagnostics in numerous ways.[28,29,30] Despite this active and promising research, the Food and Drug Administration (FDA) has only recently authorized two digital pathology whole-slide imaging (WSI) systems for primary diagnosis.[3,9,11,31,32] Even with the authorization of two WSI systems and numerous use cases,[12,13,14,18,33,34,35,36,37,38] in the U.S., we see few hospitals changing their daily clinical operations to integrate WSI for primary diagnosis.[39,40,41,42,43] Clinical laboratories face additional challenges when implementing high complexity and/or high-risk medical devices coupled with software solutions as laboratory-developed tests (LDTs).[44,45,46] For example, even when using an FDA-authorized whole-slide imaging device, the approval or clearance does not eliminate the need for an individual laboratory to verify the performance of these systems for the specific intended diagnostic purpose. Specifically, Clinical Laboratory Improvement Amendments of 1988 or CLIA '88 in the US requires at least verification[47] and substantial adaptation to implement.[48,49,50,51,52]

One value proposition for digital pathology is to take advantage of the digital nature of WSI and use artificial intelligence/machine learning (AI/ML) algorithms to support clinical decisions.[11,53] In fact, several groups have proposed that AI/ML will unlock the full potential of digital pathology.[53,54]

To examine the current regulatory guidance landscape related to digital pathology and AI, four authors (HDM, RH, EA, and JKL) performed a review of pertinent documents from the FDA. We noted the official release dates and assigned each document to one of five dimensions [Figure 1 and Supplemental Table 1]. By plotting these documents and dimensions over time, we show how the regulatory guidance landscape evolves. A novice in the field may look for one comprehensive guidance document for digital pathology and may be discouraged by the initial complexity; however, we hope that Figure 1 provides a reasonable starting point for learning the current regulatory guidance landscape. As we show Figure 1, arrows], the regulatory guidance landscape adapts over time as technologies and the

associated regulatory science matures. One key element in the multistep process to improve the regulatory guidance landscape is critical scientific input from subject-matter experts.[3,4,5,10,11,15,53] We strongly believe that "watching and waiting" will not help the case of digital pathology. Similarly, workarounds[84,85,86,87,88,89] turn into long and winding roads that ultimately end at the FDA and within the FDA's regulatory framework.[83] The *Alliance* intends to organize subject-matter experts and provide scientific input.

Simply put, the practical dilemma in digital pathology is that developers are challenged to create an FDA submission following the evolving and complex regulatory guidance landscape, and the adoption of WSI by pathologists is slowed because they cannot realize the full potential and utility of digital pathology and AI/ML without full clinical integration. The field of digital pathology is looking for broader guidance, practical advice, and streamlined regulatory pathways to help navigate this uncharted and exciting territory.

REGULATORY SCIENCE, THE PRECOMPETITIVE SPACE, AND REAL-WORLD EVIDENCE

FDA clearance of a medical device offers a vendor market access. Once introduced, market forces tend not to encourage the vendor to make the device or its subsystems interoperable.[55,56,57,58,59,60,61] We like to emphasize that routine diagnostic pathology is highly modularized and the practice does not lend itself easily to nonmodular, locked down solutions.[3,9,10,11,27,50,51,54,62] The *Alliance* believes that it can promote interoperability and innovation by launching initiatives and creating deliverables (data, standards, tools, and methods) in the precompetitive space. Organizing industry to work collaboratively in the precompetitive space will eliminate unnecessary or duplicative (proprietary) efforts and thereby save all parties' time, money, and resources when pursuing device authorizations.[63] The *Alliance* initiatives and deliverables will speed clinical integration and carry mutual benefit to all stakeholders, including regulators, clinicians, manufacturers, and most importantly, patients.

Real-world evidence (RWE) comes from the competitive, postmarket space. RWE can identify trends in adverse events, summarize where resources are being spent, and track the impact of a new diagnostic device or therapy in terms of patient outcomes. RWE can support clinical practice guidelines and decisions about reimbursement and policy. Furthermore, RWE can inform regulatory decision making, as effectively demonstrated by the Medical Device Innovation Consortium,[64,65] the National Evaluation System for health Technology Coordinating Center,[66] the Patient-Centered Outcomes Research Institute,[67,68] Friends of Cancer Research,[69,70] and others.[3,5,6,9,71,72,73,74]

FROM KEY MISSION ELEMENTS TO A DELIVERY PROCESS

Accomplishing mutual benefit to multiple stakeholders is a daunting value proposition that requires a unique regulatory science approach and stakeholder involvement for selection and prioritization of deliverables. The approach of the *Alliance* [Figure 2a] is to deliver tools by harnessing existing, precompetitive FDA programs and use the gained experience to inform effective regulation. The approach thereby aims to streamline precompetitive and eventually competitive submissions that enable faster time to market to improve patient care. Regulatory science deliverables, including tools and the experience from precompetitive submissions, will be shared, and when one integrated solution has been enabled, the *Alliance* can dissolve [Figure 2a]. The key mission elements of the *Alliance* are summarized in Table 1. [75]

To align stakeholder interests, initiatives and deliverables need to be prioritized and prioritization requires a process. We conceptualized an approach that is composed of synergistic review, project components, and resource allocation [Figure 2b]. The process starts with synergizing various stakeholder interests into concise individual projects. An *Alliance* project may consist of a clinically relevant intended use case, a data set (e.g., pixel and metadata), and an applicable regulatory science pathway [e.g., Figure 2b, triangle].

The *Alliance* membership, composed of subject-matter experts from various domains, will have the opportunity to review, contribute, and potentially modify these projects through free and voluntary feedback to the project owner. Over time, individual effort and maturation of ideas will result in optimized projects ("big ideas"). To help realize the proposed deliverables and/or allocate additional resources, we established the *Alliance* Steering Committee, a flexible organizational structure, and a code of conduct [Supplemental Table 2].

An example project is illustrated in Figure 2c. A subset of members in the *Alliance* are studying the relevance of tumor-infiltrating lymphocytes (TILs) as a prognostic and predictive biomarker.[76,77] The interest in this clinical use case led to a collaborative project that includes members from the FDA, academic medical centers (AMCs), and industry. The project, referred to as the high-throughput truthing (HTT) project, aims to demonstrate the collection and use of pathologist annotations for the purpose of evaluating AI/ML algorithms and other digital pathology initiatives. The project also aims to qualify the glass slides, whole-slide images, and pathologist annotations for evaluating AI/ML algorithms through the precompetitive FDA's Medical Device Development Tools (MDDT) program.[78] If qualified, the "ground-truth" materials can serve as a publicly available, standardized evaluation "tool" for algorithm evaluation that can be used in submissions to the FDA.

In relation to the *Alliance*, the HTT project was submitted to the *Alliance* and discussed in November 2019. The *Alliance* can contribute in multiple ways to accelerate the realization of this and similar projects. First, the *Alliance* confirmed that the aims of the project could benefit many stakeholders.

The discussions provided useful feedback from subject-matter experts regarding the clinical use case, sourcing slides from multiple sites, agreements for sharing materials within the project, and issues related to sharing materials publicly. The discussions also identified future work that could build on the lessons, methods, infrastructure, and relationships created while pursuing the current aims. Important future work identified in the discussions included scaling the effort to address generalizability across sites and generalizability across use cases.

The *Alliance* has since provided help with the project [Figure 2b, triangle 01, relevant intended use case; Figure 2c, 01] by disseminating the project needs. This networking through the *Alliance* has yielded volunteers for sourcing and scanning slides, pathologists to annotate slides and images, and opportunities to collect data. Connections have been created that are expected to help in the development of the statistical analyses and the future hosting of slides, images, and annotations. Currently, the project is developing the strategy and materials for the FDA's MDDT program [Figure 2b, triangle, MDDT; Figure 2c, 03]. The development is a learning experience for all involved, with contributions from project and *Alliance* subject-matter and regulatory affairs experts. The learning experience is expected to continue through official interactions with the FDA related to the MDDT submission. Thus, aside from helping to create the ground-truth data set, the *Alliance* aims to understand regulatory issues and processes for future streamlining of other projects and submissions. As demonstrated here, a qualified data set may result in time-savings when preparing submissions, generating additional tools, and streamlining regulatory review, resulting in faster time to market and improved patient care.

Who Is the Alliance?

The *Alliance* is composed of a diverse and interdisciplinary group of stakeholders who contribute to various aspects of diagnostic pathology, from tissue acquisition to reporting and data analytics. When deconstructing the clinical digital pathology and AI/ML pipeline into its component parts, numerous workflow steps have to function in unison [Figure 3a]. Aside from the modular nature and operational complexity, these components emphasize the importance of involving various stakeholders with each module. Given the novelty of pursuing a collaborative regulatory science effort to solve the challenge of clinical adoption of digital pathology, we noted a lack of concrete data on interested stakeholders and their

priorities. In September 2019, we conducted an internal survey [n = 42; Supplemental Table 3]. At that time, the survey respondents stated that the top 3 deliverables/workflow steps to focus on should be the DICOM standard, AI/ML test validation, and pixel and metadata capture [Figure 3b]. By self-reported primary affiliation, the *Alliance* encompasses representation from academia (32%), industry (50%), government regulators and nongovernment organizations (12%), and patient advocacy groups (6%) [Figure 3c].

Meetings, Growth, and Working Groups

Since its inception in May 2019, the *Alliance* hosted numerous teleconferences, web meetings, and three, in-person, national meetings [Figure 4a]. Over this period (May 2019–January 2020), the *Alliance* membership grew from an initial *n* = 37 (July 2019) to *n* = 322 individuals [May 2020; Figure 4a]. Each of these in-person meetings solicited collaborative input from stakeholders toward execution of concrete regulatory science deliverables. Figure 4a also includes the number of participants and frequency of steering committee web meetings. By July 2019, it became clear that various stakeholders worked on or had interest in distinct topics that the *Alliance* subsequently organized into 8 working groups by autumn 2019 [Figure 4b]. These group topics are intended to align stakeholders with subject-matter expertise and interest. Clearly, some functional requirements are relevant for multiple groups. However, we hope to minimize such redundancies by providing clear documentation of projects through appropriate project management and frequent content updates. The names of the founding and current working group leaders are provided in Figure 4b. One example of a regulatory science deliverable is also provided per group [Figure 4b]. For further updates or details on the various topics, please visit the *Alliance* website[8] or to become a member and get involved.

The *Alliance* Facilitates Regulatory Submissions

As a first key regulatory science deliverable, in late 2019, members of the *Alliance* submitted an MDDT proposal to the FDA for review (HTT project described above). The experience gained through this submission will create a starting point and testing ground for the proposed approach of the *Alliance*. In contrast to the largely confidential submission owned by the submitting entity (typically represented through a consulting firm and/or a regulatory affairs division), gaining and sharing the submission experience may inform subsequent submissions, and *Alliance* members can draw from the experience of these submissions. This particular concept is new to digital pathology. Similarly, we consider several precompetitive submission programs by the FDA[78,79] a paradigm shift that enables different ways to engage with regulatory entities. Importantly, the *Alliance* intends to create a repository of submission documents as a resource to bolster subsequent submissions with the collective experience of previous submitters. We propose that the field, and in particular patients,[80] will ultimately benefit from sharing the experiences of *Alliance* members who have submitted to regulatory agencies.

Conclusion

In the current environment of sparse and dispersed regulatory guidance for digital pathology and AI/ML, with siloed pursuits by diverse stakeholders, the *Alliance* saw an opportunity to establish an important missing element: a precompetitive regulatory science collaboration. We believe that for patients to benefit from highly complex new technologies, benefit and risk assessments are essential.[81,82] The *Alliance* helps tackle this daunting task (i. e., benefit and risk assessment for digital pathology and AI/ML) through regulatory sciences with the hope of successful clinical integration and improved patient care. That said, there are numerous issues that we need to address. For example, we want to investigate and develop protocols and definitions for continuous performance assessments of continuously learning ML algorithms. Similarly, approaching financial sustainability will require clear demonstration of clinical utility. However, the fact that numerous unanswered questions persist represents an opportunity for other agencies, regulatory entities, professional groups, and collaborative movements (like the *Alliance*) to step

up and drive developments toward comprehensive risk and safety assessments. It is important to emphasize the crucial importance of funding for regulatory and implementation science projects, in particular those that aim to inform technically appropriate and efficient science-based regulatory decision-making processes. Such funding is needed to advance cutting-edge innovations into clinical practice. In summary, the *Alliance* aims to advance the field of digital pathology and we hope that synergistic efforts between various stakeholders and regulatory scientists will ultimately speed the improvement of patient care. This begs the question: Who, if not us?

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

Acknowledgments

Footnotes

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2020/11/1/22/291538

REFERENCES

1. Weinberg AM. Science and trans-science. *Science.* 1972;177:211. [PubMed: 17815612]

2. FDA. *Advancing Regulatory Science.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/science-research/science-and-research-special-topics/advancing-regulatory-science .

3. Abels E, Pantanowitz L. Current state of the regulatory trajectory for whole slide imaging devices in the USA. *J Pathol Inform.* 2017;8:23. [PMCID: PMC5450449] [PubMed: 28584684]

4. Spindler P, Bach KF, Schmiegelow M, Bedlington N, Eichler HG. Innovation of medical products: the evolution of regulatory science, research, and education. *Ther Innov Regul Sci.* 2016;50:44–8. [PubMed: 30236007]

5. FDA. *Advancing Regulatory Science for Public Health.* 2010. [Last accessed on Jul 7]. Available from: https://wwwfdagov/media/123792/download .

6. FDA. *Regulatory Science in FDA's Center for Devices and Radiological Health: A Vital Framework for Protecting and Promoting Public Health.* 2018

7. FDA. *Collaborative Communities Toolkit.* 2019. [Last accessed on Jul 7]. Available from: https://wwwfdagov/media/116467/download .

8. MGH. MDIC DF. *MDIC DF The Alliance for Digital Pathology.* 2019. [Last accessed on Jul 7]. Available from: https://digitalpathology Allianceorg/

9. Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: A white paper from the Digital Pathology Association. *J Pathol.* 2019;249:286–94. [PMCID: PMC6852275] [PubMed: 31355445]

10. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. *J Pathol Inform.* 2019;10:9. [PMCID: PMC6437786] [PubMed: 30984469]

11. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology New tools for diagnosis and precision oncology. *Nat Rev Clin Oncol.* 2019;16:703–15. [PMCID: PMC6880861] [PubMed: 31399699]

12. García-Rojo M. International Clinical guidelines for the adoption of digital pathology: A review of technical aspects. *Pathobiology.* 2016;83:99–109. [PubMed: 27100834]

13. Guo H, Birsa J, Farahani N, Hartman DJ, Piccoli A, O'Leary M, et al. Digital pathology and anatomic pathology laboratory information system integration to support digital pathology sign-out. *J Pathol Inform.* 2016;7:23. [PMCID: PMC4872480] [PubMed: 27217973]

14. Hosseini MS, Brawley-Hayes JA, Zhang Y, Chan L, Plataniotis K, Damaskinos S. Focus quality assessment of high-throughput whole slide imaging in digital pathology. *IEEE Trans Med Imaging.* 2020;39:62–74. [PubMed: 31150339]

15. Lange H. Digital Pathology: A regulatory overview. *Lab Med.* 2011;42:587–91.

16. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study) *Am J Surg Pathol.* 2018;42:39–52. [PMCID: PMC5737464] [PubMed: 28961557]

17. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol.* 2019;20:e253–61. [PubMed: 31044723]

18. Williams BJ, Bottoms D, Clark D, Treanor D. Future-proofing pathology part 2: Building a business case for digital pathology. *J Clin Pathol.* 2019;72:198–205. [PubMed: 29549217]

19. Azar JC, Busch C, Carlbom IB. Histological stain evaluation for machine learning applications. *J Pathol Inform.* 2013;4:S11. [PMCID: PMC3678749] [PubMed: 23766933]

20. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Comput Struct Biotechnol J.* 2018;16:34–42. [PMCID: PMC6158771] [PubMed: 30275936]

21. Ström P, Kartasalo K, Olsson H, Saolorzano L, Delahunt B, Berney DM, et al. Pathologist-level grading of prostate biopsies with artificial intelligence. *arXiv preprint arXiv.* 2019:1907.01368.

22. Rivenson Y, Liu T, Wei Z, Zhang Y, de Haan K, Ozcan A. PhaseStain: The digital staining of label.free quantitative phase microscopy images using deep learning. *Light Sci Appl.* 2019;8:23. [PMCID: PMC6363787] [PubMed: 30728961]

23. Alba E. In: *Optimization techniques for solving complex problems.* Hoboken NJ, editor. Wiley: Telecom; 2009. p. 476.

24. Frensch PA, Funke J. Hoboken NJ. Complex problem solving: The European perspective. *L. Erlbaum Associates: Psychology Press.* (1 edition) 1995:340.

25. Malakooti B. *Operations and Production Systems with Multiple Objectives Hoboken.* New Jersey: John Wiley & Sons Inc; 2014. p. 1075.

26. Lyles MA, Marjorie A, Mitroff II. Organizational problem formulation: An empirical study. *Administrative Sci Quarterly.* 1980;25:102–19.

27. Cucoranu IC, Parwani AV, Vepa S, Weinstein RS, Pantanowitz L. Digital pathology: A systematic evaluation of the patent landscape. *J Pathol Inform.* 2014;5:16. [PMCID: PMC4060404] [PubMed: 25057430]

28. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med.* 2018;24:1559–67. [PubMed: 30224757]

29. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019;25:1301–9. [PMCID: PMC7418463] [PubMed: 31308507]

30. Campanella G, Rajanna AR, Corsale L, Schüffler PJ, Yagi Y, Fuchs TJ. Towards machine learned quality control: A benchmark for sharpness quantification in digital pathology. *Comput Med Imaging Graph.* 2018;65:142–51. [PubMed: 29241972]

31. FDA. *510(k) Premarket Notification Aperio AT2 DX System.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwaccessdatafdagov/cdrh_docs/pdf19/K190332pdf .

32. FDA. *FDA Allows Marketing of First Whole Slide Imaging System for Digital Pathology.* 2017. [Last accessed on 2020 Jul 07]. https://wwwfdagov/news-events/press-announcements/fda-allows-marketing-first-whole-slide-imaging-system- digital-pathology .

33. Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. *Methods.* 2014;70:59–73. [PubMed: 25034370]

34. Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL, et al. Can digital pathology result in cost savings.A financial projection for digital pathology implementation at a large integrated health care organization? *J Pathol Inform.* 2014;5:33. [PMCID: PMC4168664] [PubMed: 25250191]

35. Isaacs M, Lennerz JK, Yates S, Clermont W, Rossi J, Pfeifer JD. Implementation of whole slide imaging in surgical pathology: A value added approach. *J Pathol Inform.* 2011;2:39. [PMCID: PMC3169923] [PubMed: 21969920]

36. Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, et al. Diagnostic efficiency in digital pathology: A comparison of optical versus digital assessment in 510 surgical pathology cases. *Am J Surg Pathol.* 2018;42:53–9. [PubMed: 28877052]

37. Pantanowitz L, Dickinson K, Evans AJ, Hassell LA, Henricks WH, Lennerz JK, et al. American Telemedicine Association clinical guidelines for telepathology. *J Pathol Inform.* 2014;5:39. [PMCID: PMC4221880] [PubMed: 25379345]

38. Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: The case for clinical adoption of digital pathology. *J Clin Pathol.* 2017;70:1010–8. [PubMed: 28780514]

39. Baidoshvili A, Bucur A, van Leeuwen J, van der Laak J, Kluin P, van Diest PJ. Evaluating the benefits of digital pathology implementation: Time savings in laboratory logistics. *Histopathology.* 2018;73:784–94. [PubMed: 29924891]

40. Fraggetta F, Garozzo S, Zannoni GF, Pantanowitz L, Rossi ED. Routine Digital Pathology Workflow: The Catania Experience. *J Pathol Inform.* 2017;8:51. [PMCID: PMC5760840] [PubMed: 29416914]

41. Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, et al. Whole slide imaging equivalency and efficiency study: Experience at a large academic center. *Mod Pathol.* 2019;32:916–28. [PubMed: 30778169]

42. Hanna MG, Reuter VE, Samboy J, England C, Corsale L, Fine SW, et al. Implementation of digital pathology offers clinical and operational increase in efficiency and cost savings. *Arch Pathol Lab Med.* 2019;143:1545–55. [PMCID: PMC7448534] [PubMed: 31173528]

43. Sirintrapun SJ, Lopez AM. Telemedicine in cancer care. *Am Soc Clin Oncol Educ Book.* 2018;38:540–5. [PubMed: 30231354]

44. FDA. *Laboratory Developed Tests.* 2018

45. Schreier J, Feeney R, Keeling P. Diagnostics reform and harmonization of clinical laboratory testing. *J Mol Diagn.* 2019;21:737–45. [PubMed: 31075512]

46. Zehnbauer B. Integration standardization and diagnostics oversight of laboratory testing. *J Mol Diagn.* 2019;21:735–6. [PubMed: 31443843]

47. Halling KC, Schrijver I, Persons DL. Test verification and validation for molecular diagnostic assays. *Arch Pathol Lab Med.* 2012;136:11–3. [PubMed: 22208481]

48. Clinical Laboratory Improvement Amendments of 1988. USA: Public Law; pp. 100–578.

49. CAP. *Digital Pathology Resource Guide.* 2017. [Last accessed on 2020 Jul 07]. Available from: https://estorecaporg/OA_HTML/xxCAPibeCCtpItmDspRtejspsection=10045&item=247859&sitex=10020:22372:US .

50. Cucoranu IC, Parwani AV, Pantanowitz L. Digital whole slide imaging in cytology. *Arch Pathol Lab Med.* 2014;138:300. [PubMed: 24576021]

51. Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, et al. A practical guide to whole slide imaging: A white paper from the digital pathology association. *Arch Pathol Lab Med.* 2019;143:222–34. [PubMed: 30307746]

52. Hartman DJ, Ahmad F, Ferris RL, Rimm DL, Pantanowitz L. Utility of CD8 score by automated quantitative image analysis in head and neck squamous cell carcinoma. *Oral Oncol.* 2018;86:278–87. [PMCID: PMC6260977] [PubMed: 30409313]

53. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: The present landscape of supervised methods. *Acad Pathol.* 2019;6:2374289519873088. [PMCID: PMC6727099] [PubMed: 31523704]

54. Herrmann MD, Clunie DA, Fedorov A, Doyle SW, Pieper S, Klepeis V, et al. Implementing the DICOM standard for digital pathology. *J Pathol Inform.* 2018;9:37. [PMCID: PMC6236926] [PubMed: 30533276]

55. D'Angelo R, Weiss R, Wolfe D, Chinnam R, Murat A, Gluesing J, et al. Facing the inevitable: being prepared for regulatory requirements for laboratory developed tests. *Am J Clin Pathol.* 2018;149:484–98. [PubMed: 29659675]

56. Genzen JR. Regulation of laboratory-developed tests. *Am J Clin Pathol.* 2019;152:122–31. [PMCID: PMC6610067] [PubMed: 31242284]

57. Genzen JR, Mohlman JS, Lynch JL, Squires MW, Weiss RL. Laboratory-developed tests: A legislative and regulatory review. *Clin Chem.* 2017;63:1575–84. [PubMed: 28687634]

58. Kim AS, Bartley AN, Bridge JA, Kamel-Reid S, Lazar AJ, Lindeman NI, et al. Comparison of laboratory-developed tests and FDA-approved assays for BRAF, EGFR, and KRAS Testing. *JAMA Oncol.* 2018;4:838–41. [PMCID: PMC6145687] [PubMed: 29242895]

59. Walker Q. Designing a standard of proof: The case for professional standards in next-generation sequencing laboratory-developed tests. *J Law Biosci.* 2017;4:216–26. [PMCID: PMC5570716] [PubMed: 28852565]

60. Khullar D, Ohn JA, Trusheim M, Bach PB. Understanding the Rewards of Successful Drug Development-Thinking Inside the Box. *N Engl J Med.* 2020;382:473–80. [PubMed: 31995697]

61. Sim I. Mobile Devices and Health. *N Engl J Med.* 2019;381:956–68. [PubMed: 31483966]

62. Clunie D, Hosseinzadeh D, Wintell M, De Mena D, Lajara N, Garcia-Rojo M, et al. Digital imaging and communications in medicine whole slide imaging connectathon at digital pathology association pathology visions 2017. *J Pathol Inform.* 2018;9:6. [PMCID: PMC5869966] [PubMed: 29619278]

63. Research GV. *Digital Pathology Market Size, Share & Trends Analysis Report By Application (Drug Discovery & Development, By Product, Academic Research, Diagnosis), By End Use, And Segment Forecasts, 2019 – 2026.* 2019. Available from: https://wwwgrandviewresearchcom/industry-analysis/digital-pathology-systems-market .

64. MDIC. *Press Release: Patient Preference Collaboration.* 2017

65. MDIC. *Framework for Developing Clinical Evidence for Regulatory and Coverage Assessments in In vitro Diagnostics (IVDs)* 2019. Available from: https://mdicorg/wp-content/uploads/2019/08/Clinical-Evidence-IVD-Framework-FINALpdf .

66. NEST. *National Evaluation System for Health Technology Coordinating Center (NESTcc)* 2019. [Last accessed on 2020 Jul 07]. Available from: https://nestccorg/about/what-we-do/

67. PCORI. *Patient-Centered Research for Standards of Outcomes in Diagnostic Tests (PROD)* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwpcoriorg/research-results/2015/patient-centered-research-standards-outcomes-diagnostic-tests-prod .

68. PCORI. *The Role of Breast Density in Decision Making for Breast Cancer Screening and Diagnostic Testing-The BCSC-ADVANCE Study.* 2016. [Last accessed on 2020 Jul 07]. Available from: https://wwwpcoriorg/research-results/2016/role-breast-density-decision-making-breast-cancer-screening -and-diagnostic .

69. FOCR. *Tumor Mutational Burden (TMB)* 2018. [Last accessed on 2020 Jul 07]. Available from: https://wwwfocrorg/tmb .

70. FOCR. *ctDNA for Monitoring Treatment Response (ctMoniTR) Project.* 2019

71. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, et al. CM-Path AI in Histopathology Working Group. Artificial intelligence in digital pathology: A roadmap to routine use in clinical practice. *J Pathol.* 2019;249:143–50. [PubMed: 31144302]

72. Turner OC, Aeffner F, Bangari DS, High W, Knight B, Forest T, et al. Society of toxicologic pathology digital pathology and image analysis special interest group article*: Opinion on the application of artificial intelligence and machine learning to digital toxicologic pathology. *Toxicol Pathol.* 2020;48:277–94. [PubMed: 31645203]

73. FDA. *Real-World Evidence.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/science-research/science-and-research-special-topics/real-world-evidence .

74. Research FoC. *Pilot Project 20: Establishing the Utility of Real-World Endpoints.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfocrorg/rwe .

75. DPA. *A Brief Explanation of the Differences Between the DPA and the Alliance.* 2019

76. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin Oncol.* 2015;33:983–91. [PubMed: 25534375]

77. TILS. *International Immuno-Oncology Biomarker Working Group on Breast Cancer.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwtilsinbreastcancerorg/

78. FDA. *Medical Device Development Tools (MDDT)* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt .

79. Myers KJ. *Mock Submissions to FDA/CDRH: History and Lessons Learned; April, 2019.* Available from: https://mdicorg/wp-content/uploads/2014/05/CMS-Summit-Myerspdf.pdf .

80. FDA. *Patient Engagement in the Design and Conduct of Medical Device Clinical Investigations.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/media/130917/download .

81. FDA. *Factors to Consider Regarding Benefit-Risk in Medical Device Product Availability, Compliance, and Enforcement Decisions.* 2018. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/regulatory-information/search-fda-guidance-documents/factors-consider-regarding-benefit-risk-medical-device-product- availability-compliance-and .

82. FDA, CDRH. *Digital Health.* 2020. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/medical-devices/digital-health .

83. andMe. *23andMe And The FDA.* 2019. https://customercare23andmecom/hc/en-us/articles/211831908-23andMe-and-the-FDA .

84. Wicklung E. Down on DIY mHealth Platforms for Diabetes Care Management. *Health Intelligence.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://mhealthintelligencecom/news/fda- cracks-down-on-diy-mhealth-platforms-for-diabetes-care-management .

85. FDA. *FDA Warns People with Diabetes and Health Care Providers Against the Use of Devices for Diabetes Management Not Authorized for Sale in the United States: FDA Safety Communication.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwfdagov/medical-devices/safety-communications/fda-warns-people-diabetes-and -health-care-providers-against-use-devices-diabetes-management-not .

86. Hensley S. 23andMe Bows To FDA's Demands, Drops Health Claims. *NPR 2013.* [Last accessed on 2020 Jul 07]. Available from: https://wwwnprorg/sections/health-shots/2013/12/06/249231236/23andme-bows-to-fdas-demands- drops-health-claims .

87. Kelly S. *Software Workarounds put Patient data at Risk: OIG Medtechdive.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwmedtechdivecom/news/va-software-workarounds-put-patient-data-at-risk-oig/560204/

88. Team I. *The Future Of Voice AI In Patient Care Forbes.* 2019. Available from: https://wwwforbescom/sites/insights-intelai/2019/02/11/the-future-of-voice-ai-in-patient-care/#2c8efa87309c .

89. Eddy N. *Overload often lead to staff workarounds Healthcare IT News.* 2019. [Last accessed on 2020 Jul 07]. Available from: https://wwwhealthcareitnewscom/news/ehr-challenges-information-overload-often-lead-staff-workarounds .

Figures and Tables

Figure 1
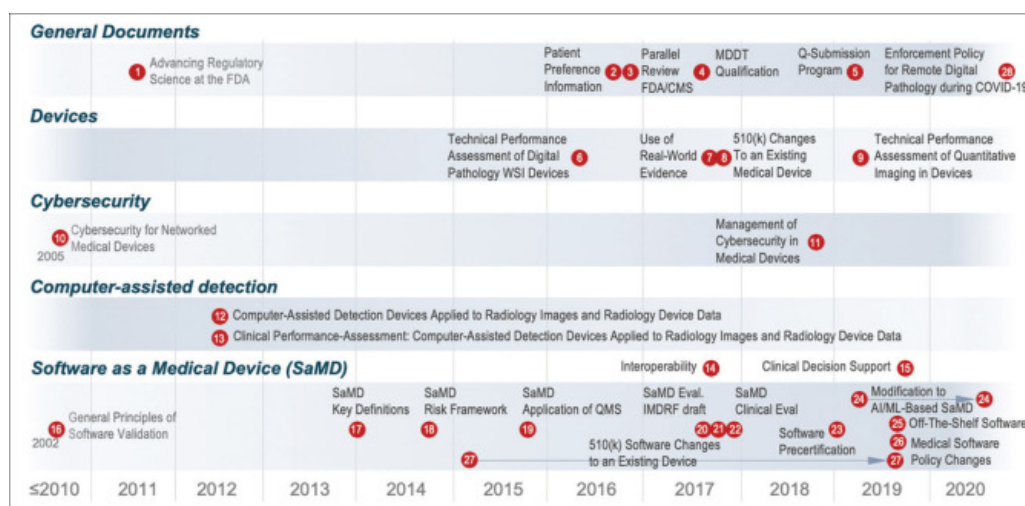
Overview of selected FDA guidance documents. Four of the authors (HM, RH, EA, and JKL) performed a meta-review of selected FDA guidance documents relevant to the scope and aims of the *Alliance*. The figure shows grouping of these guidance documents across five dimensions over time. Please note: the numbers refer to the order of review during the meta-review process; Supplemental Table 1 provides the original release dates, the official FDA guidance title, and the issuer. AI/ML: Artificial intelligence/machine learning; CMS: Centers for Medicare and Medicaid Services; FDA: Food and Drug Administration; IMDRF: International Medical Device Regulators Forum; MDDT: Medical Device Development Tools; SaMD: Software as a Medical Device; QMS: Quality management system; WSI: Whole-slide imaging

Supplemental Table 1

Meta-review of pertinent Food and Drug Administration documents

Date	n*	Title	Issuer
January 11, 2002	16	General Principles of Software Validation https://www.fda.gov/media/73141/download	CDRH and OPEQ
January 14, 2005	10	Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software https://www.fda.gov/media/72154/download	CDRH and OPEQ
August 17, 2011	1	Advancing Regulatory Science at FDA https://www.fda.gov/media/81109/download	FDA
July 02, 2012	12	Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions https://www.fda.gov/media/77635/download	CDRH, OSEL, and OPEQ
July 02, 2012	13	Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions https://www.fda.gov/media/77642/download	CDRH, OSEL, and OPEQ
December 09, 2013	17	Software as a Medical Device (SaMD): Key Definitions http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf	IMDRF and SaMD WG
September 18, 2014	18	Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf	IMDRF and SaMD WG
February 09, 2015	27[a]	Medical Device Data Systems, Medical Image Storage Devices, and Medical Image Communications Devices https://www.fda.gov/media/88572/download	CDRH and CBER
October 02, 2015	19	Software as a Medical Device (SaMD): Application of Quality Management System http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.pdf	IMDRF and SaMD WG
April 20, 2016	6	Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices https://www.fda.gov/media/90791/download	CDRH, OPEQ,

No* refers to numbering in main Figure 1; a,bRefers to updated guidance documents. CBER: Center for Biologics Evaluation and Research; CDRH: Center for Devices and Radiological Health; CMS: Centers for Medicare and Medicaid Services; DDH: Division of Digital Health; DMGP: Division of Molecular Genetics and Pathology; DRP1: Division of Submission Support; FDA: Food and Drug Administration; IMDRF: International Medical Device Regulators Forum; OCD: Office of the Center Director; OHT7: Office of Health Technology 7; OPEQ: Office of Product Evaluation and Quality; ORP: Office of Regulatory Programs; OSEL: Office of Science and Engineering Laboratories; OSPTI: Office of Strategic Partnerships and Technology Innovation; SaMD WG: Software as a Medical Device Working Group

Figure 2
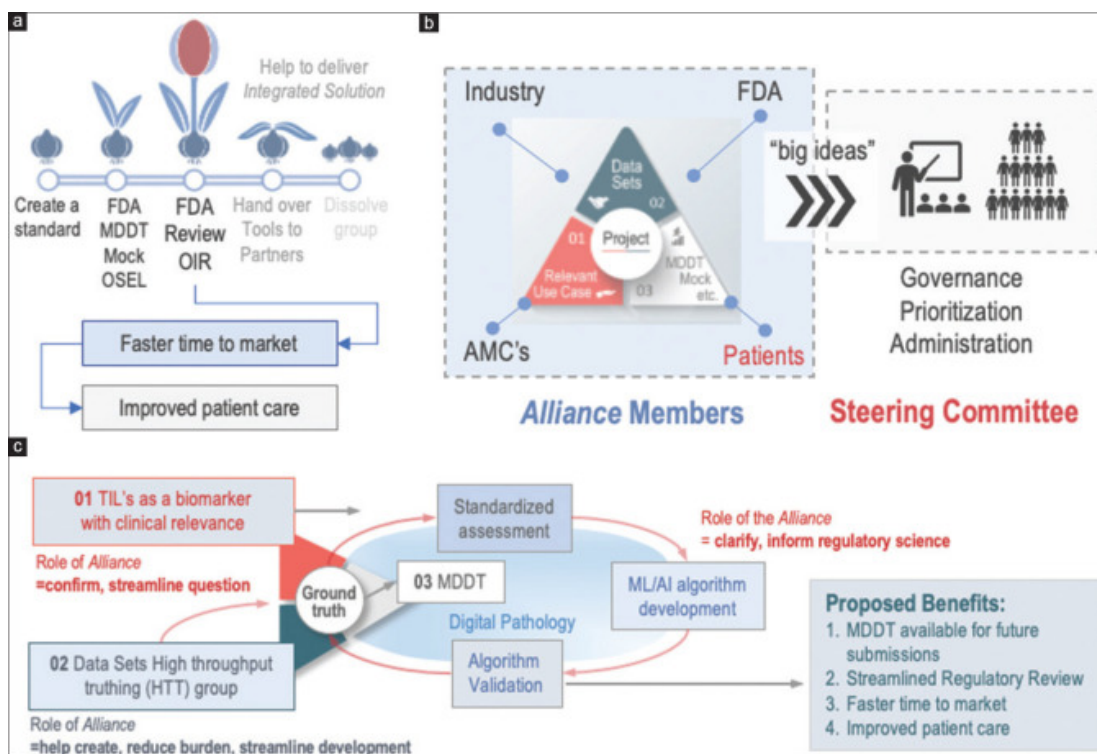
Concept, process, role, and proposed benefits of the *Alliance*. (a) The approach of the *Alliance* is to deliver tools via precompetitive FDA programs and use the gained experience to support effective FDA review. The concept also includes a predetermined exit strategy (i.e., one fully integrated solution for digital pathology). (b) The process of moving *Alliance* projects forward is essentially a two-step, multidisciplinary peer review by subject-matter experts. First, projects are reviewed, and after a multidisciplinary selection process that emphasizes the patient perspective and relevance for patient care, the steering committee (jointly with relevant partners) attempts to allocate resources. (c) Role and proposed benefits of the *Alliance* exemplified using the high-throughput truthing project for tumor-infiltrating lymphocytes as a biomarker in breast cancer. AMCs: Academic medical centers; MDDT: Medical Device Development Tools (precompetitive FDA submission program); Mock: mock submission program (precompetitive FDA submission program); OIR: Office of *In vitro* Diagnostics and Radiological Health; OPEQ: Office of Product Evaluation and Quality; OSEL: Office of Science and Engineering Laboratories; FDA: Food and Drug Administration

Table 1

Key mission elements of the *Alliance*

Definition	Explanation
Aim	To move the field of digital pathology, AI/ML and computational pathology, forward
Focus	Key emphasis on regulatory science ("how to get to the next step"); inform regulatory guidance and decision-making; explore new regulatory programs
Deliverables	The *Alliance* focuses on concrete practical deliverables, such as projects or practical guidelines, that can be used to inform and improve the regulatory guidance landscape (regulatory science)
Collaboration	We seek participation from all stakeholders
Participatory	We aim to sustain and expand the existing collaborative infrastructure of the *Alliance*
Market strategy	Focus on the precompetitive space with an emphasis on clinical deliverables towards financial sustainability for all stakeholders
Patient perspective	Make the patient perspective and clinical relevance an integral part of the deliverables
Temporary	Exit strategy: Once an end-to-end solution has been clinically integrated, the *Alliance* ends
Free	No membership fees

AI: Artificial intelligence; ML: Machine learning

Supplemental Table 2

The *Alliance* Steering Committee and Membership by Sector

Founders	Affiliation	Sector
Jochen K. Lennerz, MD, PhD	Medical Director, center for Integrated Diagnostics, Massachusetts General Hospital/Harvard Medical School	Academia
Esther Abels, MSc	Vice President of Regulatory Affairs, Clinical Affairs and Strategic Business Development, PathAI	Industry
Brandon D. Gallas, PhD	Mathematician, FDA/CDRH/OSEL/Division of Imaging, Diagnostics, and Software Reliability	Government

Steering Committee	Affiliation	Sector
Alain C. Borczuk, MD	Professor of Pathology and Laboratory Medicine, Weill Cornell Medicine	Academia
Amanda Lowe	Managing Director of Americas, Visiopharm Corporation	Industry
Ashish Sharma, PhD	Associate Professor, Department of Biomedical Informatics, Emory University School of Medicine	Academia
Clive R. Taylor, MD, DPhil	Professor Emeritus, University Southern California	Academia
David A. Clunie, MBBS	Owner, PixelMed Publishing, LLC	Industry
Frank R. Dookie, MBA	CEO and President, Sales Management Operations Consulting, Inc.; Strategic Consultant, JAV Advisors Corp.	Industry
Gina Giannini, MS	Manager of Regulatory Affairs, Digital Pathology, Roche Tissue Diagnostics	Industry
Hetal D. Marble, PhD	Program Manage of Biomarker Development and CDx, Left for Integrated Diagnostics, Massachusetts General Hospital/Harvard Medical School	Academia
Jithesh Veetil, PhD	Program Director of Data Science and Technology, Medical Device Innovation Consortium	Nonprofit
Joachim H. Schmid, PhD	Vice President of Research and Development, Digital Pathology, Roche Tissue Diagnostics	Industry

[Open in a separate window](#)

CDRH: Center for Devices and Radiological Health; OSEL: Office of Science and Engineering Laboratories; FDA: Food and Drug Administration

Figure 3

Workflow steps and *Alliance* survey results. (a) Digital pathology workflows include preanalytical, retrieval, scan (image acquisition), clinical data, metadata, machine learning algorithm development, clinical integration, clinical utility, and financial sustainability considerations; all dependent on the specific use case/application. These workflow steps correspond to the axis labels in b. (b) The *Alliance* conducted a survey among the members in September 2019. Bar graphs show the workflow steps that survey respondents felt the *Alliance* should focus on. These steps are reflected in a workflow diagram in a. (c) Survey results from September 2019. DICOM: Digital Imaging and Communications in Medicine (here referring to an interoperable file format for digital pathology); EHR: Electronic health record; H&E: Hematoxylin and eosin stain; IHC: Immunohistochemistry; LIMS: Laboratory information management system; MDIC: Medical Device Innovation Consortium

Supplemental Table 3

Survey questions and answer choices sent to the *Alliance* for Digital Pathology membership

Question number	Question	Answer choices
1	How long have you been involved with digital pathology?	<1 year
		1-5 years
		5-10 years
		>10 years
2	How many papers have you published about digital pathology?	Open ended
3	What sector do you represent?	Academia
		Industry
		Government
		Nongovernmental organization
		Other
4	Are you familiar with the MDIC?	Yes
		No
5	Should patient advocacy groups be a part of the *Alliance*?	Yes
		No
6	FDA regulatory oversight of digital pathology is:	Too simple
		Adequate
		Too complex
7	Should the *Alliance* focus on slide generation as a preanalytical factor?	Yes
		No
8	Should the *Alliance* focus on metadata capture?	Yes
		No
9	Which workflow steps should the *Alliance* focus on?	Archive retrieval
		Preanalytics
		Slide scan
		Pixel data
		Electronic health record
		Laboratory inventory

[Open in a separate window](#)

DICOM: Digital Imaging and Communications in Medicine (here referring to an interoperable file format for digital pathology); FDA: Food and Drug Administration; MDIC: Medical Device Innovation Consortium

Figure 4

Roadmap and working groups. (a) Roadmap of in-person events (status May 2020). In addition to the date, the roadmap shows hosting organization, key developments, and location of the meetings. The graph shows the membership number over time along with the number and frequency of the steering committee meetings as well as the high-throughput truthing working group. (b) The *Alliance* proposed to tackle regulatory science deliverables in digital pathology by splitting up the topic into eight distinct working groups. Each workgroup is provided with the steering committee member (s) and at least one key regulatory science deliverable. The steering committee is also responsible for minimizing redundancy between the workgroups. AI: Artificial intelligence; DPA: Digital Pathology Association; FDA: Food and Drug Administration; HTT: High-throughput truthing (an independent workgroup); MDIC: Medical Device Innovation Consortium; ML: Machine learning; USCAP: USCAP stands for United States and Canadian Academy of Pathology

---

Marble HD., et al. A regulatory science initiative to harmonize and standardize digital pathology and machine learning processes to speed up clinical innovation to patients. J Path Info. (2020)11(1):22

https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/33042601/