2018 SCIENTIFIC REPORT

Regulatory Advancements for Patients

FRIENDS of CANCER RESEARCH

FRIENDS of CANCER RESEARCH

INTRODUCTION

Throughout the year, Friends of Cancer Research (*Friends*) develops and writes white papers and publications that address leading-edge science and regulatory challenges. *Friends* convenes multi-stakeholder working groups, hosts scientific conferences and roundtables, and conducts original research to promote innovative and meaningful improvements in drug development and patient care.

Friends' white papers and publications stem from expert working groups and discussions at conferences with thought leaders and serve as resources for federal officials, regulators, drug sponsors, diagnostic companies, academics, and patient advocates. These publications help inform policy makers and other key stakeholders and catalyze the development of innovative strategies and regulatory policies for oncology drug development.

In 2018, *Friends*' white papers and publications focused on several key themes:

- Promoting new opportunities to advance pediatric drug development and research
- Ensuring optimal development, oversight, and reimbursement of diagnostic tests
- Characterizing the use of real-world evidence
- Demonstrating the value of the patient voice in oncology drug development

This journal contains a collection of the full text of the *Friends* 2018 white papers and publications. We hope our 2018 Scientific Report will be a resource to inform ongoing discussions within drug development and the regulatory space and is informative for those interested in science and regulatory issues in oncology.

FRIENDS of CANCER RESEARCH

A Friends of Cancer Research White Paper

BROADENING THE DEFINITION OF TOLERABILITY IN CANCER CLINICAL TRIALS TO BETTER MEASURE THE PATIENT EXPERIENCE

OBJECTIVE

Robust safety and tolerability data are essential in cancer therapeutic studies, and some trials are specifically designed with a key objective of demonstrating improved safety and tolerability. The development of a clinical trial framework and data elements to demonstrate comparative safety and tolerability requires a suite of endpoints and approaches to enable meaningful interpretation of results for regulatory and clinical decision-making. Identification of data elements suitable for a comparative tolerability trial design would be useful across cancer clinical trial settings where a comprehensive characterization of safety and tolerability is a critical component in the evaluation of individual and collective patient benefit.

A multi-stakeholder working group was convened, including drug sponsors, regulators from the US and Europe, researchers, and patients, to develop a contemporary definition of tolerability that better encompasses the patient experience receiving a given treatment; to identify a broader array of data elements and methodologies that more fully characterize tolerability; and to consider a trial design framework that includes patient-reported outcome (PRO) endpoints and other clinical outcomes to support patient treatment choice, regulatory and clinician decision-making, and direct patient communication in U.S. Food and Drug Administration (FDA) labeling. The concepts outlined in this whitepaper were conceived to foster patient focused drug development. In particular, this whitepaper presents opportunities to enhance the collection of the patient's perspective on symptomatic adverse events including their impacts on work and daily activities and overall side effect burden. Advancing the use of clinical outcome assessments, including PRO measurement, can complement our understanding of safety and tolerability, and the principles discussed in this whitepaper may extend into the broader cancer clinical trial setting.

CONTRIBUTORS

Ethan Basch
University of North Carolina at Chapel Hill

Alicyn Campbell
Genentech, A Member of the Roche Group

Stacie Hudgens
Clinical Outcomes Solutions

Lee Jones
Research and Patient Advocate

Bellinda King-Kallimanis
U.S. Food and Drug Administration

Paul Kluetz
U.S. Food and Drug Administration

Daniel O'Connor
The Medicines and Healthcare Products Regulatory Agency (MHRA)

Oliver Rosen
Deciphera Pharmaceuticals

INTRODUCTION

Focused efforts in clinical cancer research have led to treatment options with novel therapeutic modalities for a given cancer target. These drugs are often associated with unique safety profiles and are more frequently administered over prolonged periods of time. They can, for example, be associated with low-grade toxicities that in the short-term may be tolerable but can become burdensome over the course of treatment leading to dose reduction or treatment discontinuation despite promising treatment effects. Therefore, longitudinal assessment of patient-reported symptomatic adverse events can help better describe the tolerability of a drug and inform patient decision-making.

Tolerability is a complex concept defined by the International Conference on Harmonization (ICH) as "the degree to which overt adverse effects can be tolerated by the subject" (ICH E9).[i] Information currently used in oncology trials to assess tolerability includes clinician-reported safety data using the Common Terminology Criteria for Adverse Events (CTCAE), and other trial data including dose modifications, dose discontinuations, and hospitalizations (Figure 1). Many symptomatic adverse events are unobservable (e.g., nausea, fatigue), and how adverse events may interfere with a patient's life is best known and reported by the patient. It is known that these treatment-related symptoms impacting a patient's daily activities and quality of life frequently go undetected by investigators.[ii, iii] Therefore, integration of patient-reported data is critical to fully understand the tolerability of a therapy and provide complementary information to clinician-reported safety that identifies which symptomatic adverse events are most burdensome to patients.[iv] This is particularly important in diseases with multiple therapeutic choices, where there is a poor overall prognosis and where an optimal treatment algorithm has not yet been established. Data characterizing tolerability can also provide important additional information in a non-inferiority trial, so that better tolerated regimens with similar clinical efficacy can be more easily identified.

The current ICH definition of tolerability does not emphasize the patient experience* while on treatment and lacks focus on how adverse events associated with a treatment can be best evaluated from the patient's perspective.

*Patient experience data: Defined in Title III, section 3001 of the 21st Century Cures Act, as amended by section 605 of the FDA Reauthorization Act of 2017 (FDARA), and includes data that are collected by any persons and are intended to provide informa-tion about patients' experiences with a disease or condition. Patient experience data can be interpreted as information that cap-tures patients' experiences, perspectives, needs, and priorities related to (but not limited to): 1) the symptoms of their condition and its natural history; 2) the impact of the conditions on their functioning and quality of life; 3) their experience with treatments; 4) input on which outcomes are important to them; 5) patient preferences for outcomes and treatments; and 6) the relative importance of any issue as defined by patients." https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm610317.htm

Thus, a new working definition is proposed that incorporates the patient experience by measuring treatment burden and patient-reported symptomatic toxicity and function.

> *The tolerability of a medical product is the degree to which symptomatic and non-symptomatic adverse events associated with the product's administration affect the ability or desire of the patient to adhere to the dose or intensity of therapy. A complete understanding of tolerability should include direct measurement from the patient on how they are feeling and functioning while on treatment.*

This whitepaper will focus on data elements that can be used to assess tolerability based on the new definition above in cancer product development. It is expected that these elements could be used to generate evidence to evaluate treatment tolerability as part of a comparative tolerability trial design.

Figure 1. Components that Inform Tolerability. Clinician-reported outcomes and case report data are routinely collected to assess the safety and tolerability of a therapy. Although this information is important, it provides a limited understanding of the full scope of tolerability from a patient's perspective. Routine systematic collection of patient-reported outcomes and gaining the patient's view on treatment burden can provide important information regarding how patients experience treatment and which symptoms and adverse events might impact treatment decisions.

CONTRIBUTION OF PATIENT-REPORTED OUTCOMES TO UNDERSTANDING TOLERABILITY

PRO assessments provide important supportive data in oncology trials and are becoming more commonly used to assess both treatment benefits and adverse events (toxicities) to fully evaluate the impact of the treatment and disease on the patient. Regulatory recommendations exist: the FDA released guidance in 2009 for drug manufacturers seeking PRO claims of treatment benefit and the European Medicines Agency (EMA) released an Appendix 2 to the "Guideline on the Evaluation of Anticancer Medicinal Products in Man" on the use of PROs in 2016. The FDA is currently developing additional and updated guidance to further encourage the development and use of PROs in clinical trials.

Table 1 outlines a proposed list of key conceptual data elements that should be considered in a trial design to measure tolerability based on the expanded definition. Importantly, the incorporation of PRO measures allows for the characterization of tolerability based on direct patient experience. Several categories of PROs for tolerability assessment are identified:

- Patient-reported symptomatic adverse events
- Patient-reported overall burden of adverse events
- Patient-reported physical functioning
- Other types of functional assessments

We acknowledge there are other types of PRO measures such as patient preference and satisfaction that support optimal patient decision-making, but for the purposes of this whitepaper we intend to focus on symptomatic adverse events and functional concepts most proximal to the effects of the therapy in the clinical trial setting. We also encourage exploring other existing and emerging sources of data for physical function such as wearable devices and performance outcome measures that can be used to support and complement PRO measurement of physical function. The outcomes of such trials may support patient choice based on better overall understanding of the treatment experience for a particular therapy.

Patient-reported data must be obtained from PRO instruments that are well-defined and reliable and that are fit-for-purpose. The FDA defines fit-for-purpose as "a conclusion that the level of validation associated with a medical product development tool is sufficient to support its context of use."[v] There are available PRO measurement systems that can be used to generate this data including item libraries like the National Cancer Institute's PRO-CTCAE (Patient-Reported Outcomes version of the Common Toxicity Criteria for Adverse Events), which was developed specifically for the assessment of symptomatic adverse events and is mapped to the CTCAE and MedDRA. Other single item questions on symptom severity and well-defined functional scales could be selected from existing measurement systems, although the acceptability of various approaches should be discussed with regulatory agencies in scientific advice.

Table 1: Key Data Elements for an Oncology Clinical Trial Assessment of Tolerability Alongside Traditional Measures of Efficacy

Source	Clinical Outcome	Utility of Elements
Efficacy Data	Response rate (RR)	Need to demonstrate efficacy using well recognized endpoints
	Progression free survival (PFS)	
	Overall survival (OS)	
Clinician-Derived Safety/Tolerability	Common Terminology Criteria for Adverse Events (CTCAE)	Traditionally used signals of tolerability are reported by the clinician/healthcare professional and should continue to be routinely captured
	Dose interruption	
	Dose modification	
	Dose discontinuation not due to progressive disease or death	
Patient-Derived Adverse Event Data	Symptomatic adverse events	The importance of the patient experience of the treatment is emphasized in the new definition of tolerability 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
	Global side effect impact/bother/burden	
Additional Supportive Patient-Derived Data	Physical function	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
	Role function (ability to work and carry out daily activities)	
	Other well-defined functional domains (e.g., emotional, social, cognitive)	
	Specific key symptoms (e.g., pain, fatigue, anorexia)	
	Disease symptom scale (if applicable)	
Healthcare Utilization	Hospitalization rates/duration	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
	Emergency department visits	
	Supportive care medication use	

TRANSLATING DATA FROM MULTIPLE SOURCES TO COMMUNICATE TOLERABILITY

Through the new definition of tolerability, four key components for measuring tolerability have been identified. In addition to efficacy endpoints, each data element described in Table 1 brings a unique quality to characterizing tolerability, providing a more patient-centric view of a treatment regimen.

In order to bring all the elements together in a way that can help inform tolerability, it is suggested that a descriptive analysis is provided in a table format where the key aspects of the data from each component can be considered, summarized, and their impact noted. Any uncertainties can also be recorded. It is therefore not envisaged that a binary statement of tolerability will be made (i.e., on treatment and tolerating versus discontinued due to AE and not tolerating), but rather a more complete picture of the patient experience obtained from the various data elements.

Quantitatively, a large component of tolerability is likely the overall impact of the side effects of treatment. It would be useful to quantify a range of overall side effect burden on the patient, and in this case the data element of overall side effect impact or burden could be used as a key endpoint. This side effect impact or treatment burden endpoint would be informed and interpreted by the other tolerability data elements including symptomatic AEs and the potential impact the side effect burden would have on other patient-reported functional domains (e.g., physical and role function). Some statistical methods that could be applied to the evaluation of this data could include but are not limited to:

- Proportion of patients experiencing the worst magnitude of each response level of each elicited symptomatic AE PRO item, by treatment, each time point of measurement, and for the total period of study participation
- Proportion of patients with each response level of an item eliciting overall perceived burden of adverse events
- Qualitative inquiry with patients on relevant PRO items contributing to tolerability (e.g., end of treatment questionnaire)
- Impact of frequent or high-grade symptomatic AEs on physical function
- Impact of frequent or high-grade AEs on other functional measures and HRQoL
- Comprehensive description of global side effect impact

BENEFITS OF IMPROVED TOLERABILITY INFORMATION

Including PRO measures and other tolerability data elements throughout drug development can have numerous benefits. Tolerability data can provide information for clinical dose selection early in development and allow more precise dose-finding by balancing the biologically optimal dose with the dose that has the most favorable tolerability profile. For example, when deciding between possible dose regimens of similar clinical activity, PRO measurement of symptomatic AEs and overall treatment burden can provide evidence beyond the traditional clinical AE reporting on the impact of cumulative symptomatic toxicity (e.g., the

impact of frequent and prolonged symptomatic Grade 1 AEs may be more burdensome to patients than less frequent and potentially asymptomatic Grade 3 adverse events). Identifying a more tolerable dose can maximize patient adherence to the selected late stage or approved dose, rather than ad hoc dose modification in the registration trial or post-marketing setting, which can lead to unnecessary patient burden and suboptimal dose intensity potentially affecting efficacy. In addition, exploration of tolerability in early dose-finding trials can identify candidates for later-phase comparative tolerability trial designs.

Tolerability data elements should be used in late stage settings to support clinical benefit by complementing standard safety data. Where the objective is comparative improvement in safety or tolerability, one trial design that could be considered is a superiority trial design against an active comparator considered standard-of-care. Such a trial would have an efficacy-based primary endpoint. The results of the primary endpoint could either be enhanced or diminished by an added 'comparative tolerability' endpoint. Another example of a trial design that would benefit from tolerability data elements is the non-inferiority trial design. Non-inferiority trials have an efficacy primary endpoint and typically do not prioritize tolerability assessment. In some cases, there may be a similarly effective drug amongst available therapies that appears to have improved safety and tolerability. In this setting, one or more elements of 'tolerability' could be a co-primary endpoint with efficacy data (unless comparative efficacy has been previously assessed in a head to head study). Regardless of whether data elements for tolerability are used as primary, secondary, or descriptive exploratory data, there is a benefit for improved characterization of tolerability through the inclusion of patient-reported symptomatic adverse events and function, across early and late stage drug development.

The advantages of collecting rigorous patient relevant evidence also creates new challenges. PRO assessments are commonly incorporated into registration trials, but best practices for incorporating PRO data with the objective of demonstrating improved safety and tolerability will require careful clinical trial design.[vi] While item libraries such as the PRO-CTCAE can provide the needed flexibility to adapt to different toxicity profiles, an objective method to select which symptomatic adverse events to assess will be important to ensure an unbiased selection is obtained. Identifying an appropriate PRO assessment frequency will be important as well as monitoring for completion rates to mitigate missing data. In addition, standard methods to analyze and present PRO data and other tolerability data have been initially developed and are being further advanced.[vii, viii] Several international efforts have been undertaken in these areas.[iv,v]

COMMUNICATING TOLERABILITY INFORMATION

Tolerability data elements including PRO assessments and healthcare utilization data can further inform a product's clinical benefit and form part of the totality of the evidence evaluated by the FDA and other regulatory agencies when determining benefit:risk. Regardless of whether tolerability data are included in product labeling, all PRO and other tolerability results can be reviewed as part of the totality of data to support a benefit:risk determination. Tolerability as currently communicated in product labels and other data sources (e.g., dose modifications) can be further characterized by PRO data assessing symptom severity/occurrence and impact on function in addition to overall treatment burden as previously described. This data could be descrip-

tively analyzed and presented in product labeling, provided the assessments are well-defined and fit-for-purpose, there is an acceptable level of PRO completion in the trial, and the data add information that informs safety and tolerability (as is done with CTCAE data in section 6 of FDA labeling).

Importantly, communicating tolerability data must be balanced by describing both the positive and negative effects of the therapy. While descriptive data can be labeled to inform safety and tolerability, where a marketing claim of treatment benefit is the objective, a hypothesis must be stated, and this requires an endpoint to be constructed and statistically tested, including adjustment for multiple testing (i.e., multiplicity). This is no different than an efficacy marketing claim of improved progression-free survival. A claim of improved safety or tolerability such as "drug A causes less overall side effect burden than drug B" or "drug A causes less diarrhea than drug B" will need to be supported by substantial evidence in a well-controlled trial using well-defined a priori described methodology and reliable assessments. Space limitations and layout of product labels will necessitate concise, accurate, and non-misleading presentation of the data that will be interpretable and meaningful to providers.

Multiple other forms of communication such as guidelines and clinical pathways should also be explored and have potential as additional communication vehicles. Standard analytic and presentation methods will also be useful for other communication avenues such as published literature. Initial analytics are likely to be directed to scientific and policy audiences, however it is acknowledged that technical presentations may not be intuitive for patients. For example, common scientific descriptions of data that include p-values, means, and hazard ratios can be challenging to understand and translate into meaningful decision making for patients. Patients may find bar charts, arrays, and graphs easier to interpret, whereas forest plots that display relative risks may be more intuitive for physicians. More work will need to be carried out with both clinical and patient groups to test various data displays and layouts, identifying the most interpretable visualization for the target audience. The ability to include valid, understandable, and reliable data in communication materials beyond the product label, such as patient and clinician facing educational materials, can provide another opportunity for better and more informed decision-making in a more flexible format.

CONCLUSION

We have proposed a new working definition for tolerability that incorporates the patient experience by including patient-reported data elements that measure symptomatic adverse events, overall impact of adverse events, and function as well as elements of healthcare utilization. These tolerability elements can impact the ability and desire of the patient to adhere to the dose or intensity of therapy. Incorporating patient-reported symptomatic adverse events and impacts into early and late stage drug development holds promise to improve dose selection, provide additional information on the side effect profile of a therapy, and support informed therapeutic decision-making for patients. Sponsors should engage regulators early in the drug development program to discuss concepts and trial designs. Measurement tools such as the PRO-CTCAE and measures of patient-reported impacts are available, and efforts to identify best practices for using PRO assessment to meet the objective of safety and tolerability are underway. Standard analytic and visualization methods will need to be tested with various stakeholders who will use this information for

policy decisions as well as patient and clinical decision making. Communicating tolerability data via multiple venues is important to provide valid and reliable information to guide treatment decisions. In conclusion, as cancer treatment evolves and two, three, and four drug regimens become more common in oncology, more systematic and rigorous assessment of tolerability is key for patients, providers, regulators, and payers when assessing the impact of new treatments.

NEXT STEPS

The authors encourage comments and reactions to the perspectives presented in this whitepaper. To further develop the conclusions and concepts presented in this whitepaper, we propose these next steps:

1. Encourage the integration of patient-reported symptomatic adverse events, overall side effect impact, and functional endpoints into oncology clinical trials to provide improved understanding of tolerability.

2. Explore methodology and analytical methods to quantify tolerability data elements (Table 1) to ensure each aspect can be considered and summarized, and their impact understood.

3. Develop a case study to demonstrate how to operationalize the concepts in this whitepaper.

4. Understand how tolerability data can be better disseminated and communicated in a variety of formats, with an initial focus on patient-centric healthcare professional material.

5. Engage payers and international regulators to discuss and identify how tolerability endpoints and improved patient experience data will impact decision-making.

REFERENCES

[i] ICH Harmonised Tripartite Guideline "Statistical Principles for Clinical Trials" http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

[ii] Atkinson, et al. 2016. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. Support Care Cancer 24(8):3669-76.

[iii] Basch, et al. 2009. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 101(23):1624-32.

[iv] Kluetz, et al. 2017. Informing the Tolerability of Cancer Treatments Using Patient-Reported Outcome Measures: Summary of an FDA and Critical Path Institute Workshop. Value in Health 21(6):742-747.

[v] FDA-NIH Biomarker Working Group. BEST (Biomarkers, Endpoints, and other Tools) Resource [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK326791/

[vi] Calvert, et al. 2018. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. JAMA 319(5):483-494.

[vii] Bottomley et al. 2016 Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. Lancet Oncol. 17(11):e510-e514.

[viii] Basch, et al. 2016. Methods for Implementing and Reporting Patient-Reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clin Ther. 38(4):821-30.

Commentary

DIA

Therapeutic Innovation & Regulatory Science
1-5

DOI: 10.1177/2168479018790533
tirs.sagepub.com

Incorporating Patient Advocates in Oncology Clinical Development: Lessons Learned From a Novel Pilot Program

Wendy Selig, MSJ[1], Ian Banks, MD[2], Anjelica Davis, MPPA[3], Gissoo DeCotiis, MBA[4], Ryan Hohman, JD[5], and Lisa Schlager[6]

Abstract

The advent of patient-focused drug development (PFDD) has underscored the priority of engaging the "voice of the patient" in therapy development. Industry sponsors are working to enhance engagement of patients early, particularly within decision making for design and execution of clinical trials. This trend is especially significant within oncology, as industry leaders partner with patient advocacy organizations, individual patients, and clinicians to enhance patient-centricity. These partnerships often require a willingness to change attitudes, approaches, and processes to reshape traditional models of drug development. In 2016, Bayer Oncology launched a pilot program called the Patient Advocate Advisory Council (PAAC), to design and execute a program whereby patients join clinical development teams. The PAAC, composed of experienced patient advocates from the US and Europe, worked closely with company leaders to design and execute a pilot in an ongoing clinical development program. The PAAC and Bayer teams have identified important learnings from the first phase of the program, emphasizing earlier engagement of patient advisors, launching the enhanced training platform, and recruiting additional PAAC members to expand the initiative's reach across the cancer community. A critical success factor is having champions for patient engagement within the company to ensure that activities are streamlined and standardized as patient engagement becomes more common. This is particularly important given that patient engagement should be a long-term investment with sufficient and sustained resources. PAAC members and Bayer have committed to sharing learnings, to advance opportunities for successful patient engagement in drug development throughout the oncology therapeutic landscape.

Keywords

patient engagement, patient advocacy, patient centricity, drug development, oncology

Background

The advent of patient-focused drug development (PFDD)[1] has brought into sharp focus the priority of establishing a systematic approach to engaging the "patient voice" in therapy development. This is especially true in the US and Europe, where legislative and regulatory activities are defining, measuring, and advancing patient centricity and empowerment.[2-4] Increasingly, stakeholders are developing processes and systems to embed the voice of the patient into therapy development and treatment decision making.[i]

These trends reflect an environment where patients are seeking a more proactive, dynamic, and central role in their health care, as they move from the role of passenger to one of "co-pilot."[5]

Industry sponsors have worked to enhance patient engagement in drug discovery, particularly within clinical trial planning and decision making.[6-8] This trend is especially significant within oncology, as company leaders partner with patient advocacy organizations, individual patients, and leading clinicians to enhance patient-centricity. These partnerships require a willingness to change attitudes and processes, reshaping models of drug development that have traditionally been somewhat disconnected from patients.[9]

In 2016, Bayer Oncology launched a pilot program called the Patient Advocate Advisory Council (PAAC), whose purpose was to bring patients into clinical development teams. The

[1] WS Collaborative, McLean, VA, USA
[2] European Men's Health Forum, London, United Kingdom
[3] Fight Colorectal Cancer (FightCRC), Springfield, MO, USA
[4] Bayer Oncology, Whippany, NJ, USA
[5] Friends of Cancer Research (Friends), Washington, DC, USA
[6] Facing Our Risk of Cancer Empowered (FORCE), Washington, DC, USA

Submitted 8-Apr-2018; accepted 3-Jul-2018

Corresponding Author:
Wendy Selig, MSJ, WS Collaborative, 934 Bellview Road, McLean, VA 22102, USA.
Email: wendy@wscollaborative.com

PAAC, comprised of experienced patient advocates in the US and Europe,[ii] worked with company leaders to implement a pilot patient engagement program, taking advantage of a planned Phase 2 oncology trial.

This paper summarizes the PAAC initiative and key learnings, as the PAAC model is integrated across the company's oncology pipeline and its leaders share best practices.

Methods

In 2016, for its annual Patient Advocacy Summit, Bayer leadership challenged patient advocacy organizations to present proposals to help integrate the patient voice earlier in clinical development processes. Diverse teams of patient advocates representing national advocacy organizations brought specific program ideas for review by company leaders and prominent patient advocates.

The review panel assessed the ideas for innovation and creativity, direct benefits to patients, effectiveness of a roll-out plan, ease of implementation, and benefit to the company's portfolio and processes. The winning proposal, to launch a pilot project for systematic integration of patient perspectives into the company's clinical trial process, was granted a 1-year budget for a 5-member Patient Advocacy Advisory Council (PAAC).

Results

In its first phase of work, the PAAC developed a blueprint to systematically engage patient advisors (also known as expert patients or patient advocates) and built an execution strategy using an ongoing Bayer development program as its pilot.

The 2016-2017 PAAC project deliverables included

- defining and developing Patient Selection Criteria and a "role description" for patient advisors chosen to participate in clinical development activities;
- recruiting suitable patient advisors;
- developing a training curriculum framework and platform for patient advisors and relevant audiences inside the company;
- implementing seamless training, communication, and mutual learning with identified patient advisors, including assisting Bayer in establishing appropriate "rules of engagement" between patients and clinical drug development teams; and
- piloting the new framework and approach through an ongoing company development program, including evaluating and adapting the model based on learnings from the pilot.

In late 2016, the PAAC conducted a series of discussions with Bayer leaders in advocacy relations and clinical development to plan the pilot. With access to information about one of the company's clinical development programs under confidentiality agreement, PAAC members provided feedback to the clinical team about specific elements of a trial protocol from a patient perspective and identified models for integration and training to achieve successful engagement of a patient advisor into the program. Discussions focused on ensuring stakeholder agreement about appropriate expectations for the patient advisor's level of experience, time commitment, and compensation. Recommendations recently published by the Clinical Trials Transformation Initiative (CTTI) for stakeholder collaborations in clinical trials were incorporated.[10]

PAAC members worked along parallel tracks to define general patient advisor selection criteria and role description documents (see Figure 1), develop training curricula modules for both patient advisors and audiences within the company, while identifying and onboarding a relevant cancer patient advisor for engagement with the clinical development team working on the study selected for the PAAC pilot. The selected patient advisor began engagement with the project in the spring of 2017, meeting by phone with key members of the study team and providing feedback on the protocol and development program.

The PAAC developed materials to populate an interactive, expandable training platform for patient advisors and company staff. Content was compiled and reviewed by company stakeholders, including legal and compliance, and then incorporated into a web-based architecture built by an external consultant and tested by the PAAC. This platform utilizes the Friends of Cancer Research ProgressforPatients.org regulatory education online learning program.[11] The training platform is flexible, allowing for easy updates and augmentation as needed to incorporate additional content and tools.

Discussion

The PAAC met with company leaders during the 2017 ASCO Innovation Summit to review the project and evaluate its success. All stakeholders were generally pleased with the results of the pilot and agreed to leverage its learnings and expand its reach within the company's development programs. A key learning from the pilot was that patient advisors can and should be engaged with a development program *even earlier* than was possible with the pilot study. PAAC members and their company colleagues agree that, despite the risk that therapy development programs in the earliest stages may not advance to later development, it is nonetheless important to ensure these early programs are designed with direct patient input.

As summed up by Bayer's President of Pharmaceuticals, Americas: "The collective team netted out that the recruitment of patient advisors who can consult on clinical trial design should begin as soon as program discussions around drug administration and target indication are approved."[12]

Additional learnings from the pilot PAAC project included: tackling challenges within existing corporate culture; appropriately defining and clarifying roles and responsibilities for patient advisors and company study team members; anticipating and meeting resource needs; and addressing logistical and communications issues (see Figure 2).

PAAC Recommended Patient Advisor Roles & Responsibilities

When possible, Patient Advisor should be engaged in:

- Discussion of unmet need (definition of ideal drug properties, including IV vs. PO, side effect profile, administration frequency)
- Input into formulation options (ex: coated caplet, dispersible tablets, tablets)
- Draft TPP (Target Product Profile)
- Advisory Board member when appropriate
- PRO development and insight into patient perspectives (including wearables, and apps)
- Draft protocol review (including inclusion/exclusion criteria; frequency of visits, assessment details)
- Site feasibility questionnaire
- Patient recruitment materials (feedback on patient facing documents)
- Informed Consent review
- Study collateral materials (feedback on patient facing documents)
- Development of patient journey (@ time of registration)
- Patient Education materials (@ time of registration) (feedback on patient facing documents and treatment management / shared decision making)
- Evaluation of risk/benefit, QoL and PRO
- Regulatory body interactions if applicable
- Input into design of real life experience studies (RLE)
- Support education of call centers to better understand patient's journey
- Input on 2nd generation drug improvements

Figure 1. PAAC recommended patient advisor roles & responsibilities. A key objective of the Patient Advocate Advisory Council (PAAC) pilot program was to define a "role description" for patient advisors chosen to participate in clinical development activities. The full range of advisors' roles and responsibilities are listed here.

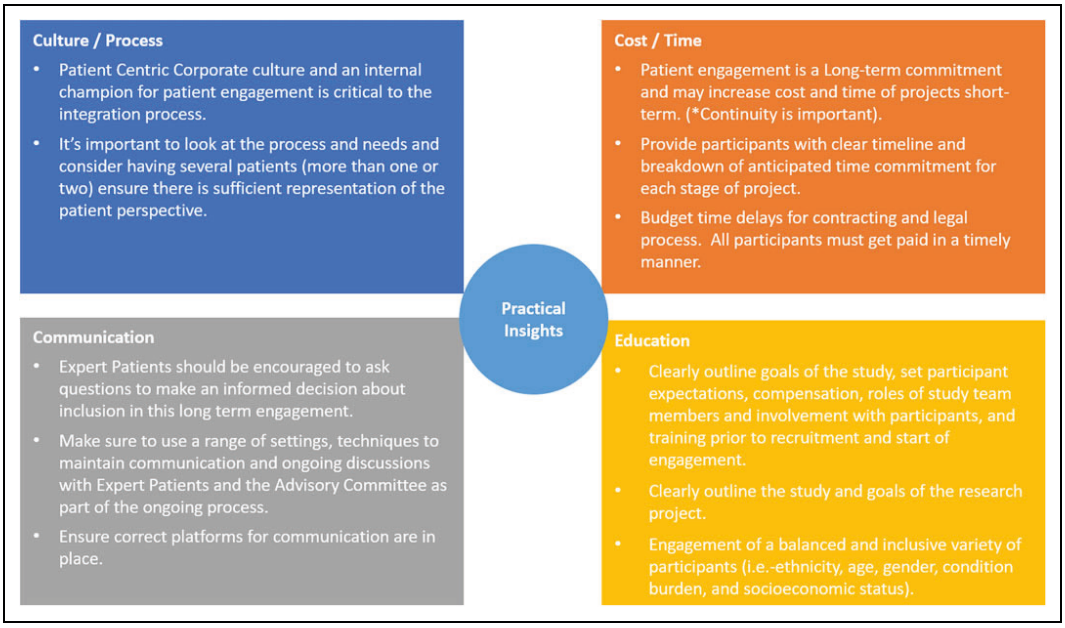


Figure 2. Lessons learned from PAAC Pilot. Key lessons drawn from the PAAC initiative encompassed several thematic areas, including tackling challenges within existing corporate culture; appropriately defining and clarifying roles and responsibilities for patient advisors and company study team members; anticipating and meeting resource needs; and addressing logistical and communications issues.

Specific challenges and solutions identified by the PAAC and company stakeholders were:

Challenge: Ensure seamless and sufficient, ongoing two-way communication, and prioritize efforts to coordinate necessary discussions involving the patient advisor.

Solutions: Identify and onboard patient advisor early in the trial process to allow for regular communication touch-points and appropriate inclusion in key discussions. Use a range of settings and techniques to facilitate appropriate inclusion of the patient advisor and PAAC members.

Challenge: Address barriers within existing corporate culture that impede patient engagement.

Solutions: Develop an appropriate compensation structure for patient advisors, streamline necessary legal and contractual processes, and establish common expectations between the team and the patient advisor.

Challenge: Carefully select the right patient advisors.

Solution: Consider involving more than one patient in a study team to enhance representativeness of perspectives and skill sets.

A critical success factor is having champions for patient engagement within the company to mitigate obstacles and streamline activities as patient engagement becomes more commonplace within the company. This is particularly important given that patient engagement should be viewed as a long-term investment with sufficient and sustained resources.

Patient engagement in drug development programs is expanding quickly as the field evolves and best practices are developed. Stakeholders are actively exploring novel methods for measurement of success and returns on investment. The PAAC and Bayer team are committed to evaluating relevant metrics from the pilot program, including reviewing the number of amendments to the study, quantity and type of suggestions from the expert patient included in the study design, number of additional study teams that adopt this approach, and ability to engage patients even earlier in the process by those study teams.

While the pilot is still ongoing, there have already been positive effects from the program. For example, there were important insights from the expert patient relating to crossover design considerations and type of information provided to patients. In addition, the expert patient was involved in discussions about feasibility for capturing regular data from patients and in formulating informed consent materials.

The study team will ultimately evaluate how these insights impact protocol execution or result in savings (in time, money patient resources etc.). While this process continues, there has been significant interest expressed within Bayer among other study teams (and even other therapeutic areas beyond oncology) to adopt the program.

The pilot program has enabled creation of a robust training platform for expert patients and internal company audiences, providing an opportunity to seamlessly integrate patients into Bayer's clinical development program, rapidly expand the program to additional study teams, and elevate understanding of best practices for patient engagement within Bayer.

Conclusion

Based on successful execution of the Pilot program, Bayer has committed to continuing the PAAC initiative and expanding its application. The PAAC and Bayer teams have identified important learnings from the first phase of the program and are incorporating them into the program's next phase, emphasizing earlier engagement of patient advisors, launching the enhanced training platform, and recruiting additional PAAC members to expand the initiative's reach across the cancer community.

Acknowledgment

The authors wish to thank the European Men's Health Forum, Fight Colorectal Cancer, Friends of Cancer Research, Facing our Risk of Cancer Empowered and Bayer Oncology for their support of the PAAC initiative.

Declaration of Conflicting Interests

No potential conflicts were declared.

Funding

Development of this manuscript was made possible by a grant from Bayer Oncology.

ORCID iD

Wendy Selig, MSJ http://orcid.org/0000-0002-9252-7170

Supplemental Material

Supplemental material for this article is available online.

Notes

i. For example, the ECCO summit in Sept 2018 (Vienna) has the chair of its patient advisory committee as co-chair of the meeting. More to the point, there is *no* patient track, the patient voice is embedded in every single facet of the event with speakers, panellists, and chairs. The importance of this cannot be overstated as the summit will be a decision-making process based on recommendations formulated by the summit discussion and work preceding the event to ensure that patients are no longer a passive recipient of decisions made in their absence.

ii. Ian Banks, MD, European Men's Health Forum; Anjelica Davis, MPPA, Fight Colorectal Cancer (FightCRC); Ryan Hohman, JD, Friends of Cancer Research (Friends); Lisa Schlager, Facing Our Risk of Cancer Empowered (FORCE); Wendy Selig, MSJ, WSCollaborative (former CEO, Melanoma Research Alliance).

References

1. Food and Drug Administration. Plan for issuance of patient focused drug development guidance. https://www.fda.gov/down

loads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM563618.pdf. Published May 2017.
2. EUPATI European Patients' Academy. Guidance for patient involvement in industry-led medicines R&D. https://www.eupati.eu/patient-involvement/guidance-for-patient-involvement-in-industry-led-medicines-rd/.
3. European Medicines Agency. Patients & consumers. http://www.ema.europa.eu/ema/index.jsp?curl=pages/partners_and_networks/general/general_content_000317.jsp&.
4. Food and Drug Administration. Enhancing benefit risk assessment in regulatory decision-making. http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm. Published June 2017.
5. Anderson M, McCleary K. From passengers to co-pilots: patient roles expand. *Sci Transl Med*. 2015;7(291):291fs25.
6. DIA. DIA insights: patient engagement. http://www.diaglobal.org/en/resources/how-we-think/patient-engagement.
7. BIO. BIO-PPMD Whitepaper: key considerations in developing and integrating patient perspectives in drug development: examination of the Duchenne case study. https://www.bio.org/sites/default/files/BIO_PPMD_whitepaper_web.pdf. Published June 2016.
8. National Health Council. Patient engagement resources. http://www.nationalhealthcouncil.org/public-policy/patient-engagement.
9. Parikh S. Study insights: patient engagement in drug development. *Appl Clin Trials*. http://www.appliedclinicaltrialsonline.com/study-insights-patient-engagement-drug-development. Published April 2017.
10. Clinical Trials Transformation Initiative. Patient groups and Clinical Trials: CTTI Recommendations. https://www.ctti-clinicaltrials.org/files/pgct-executive-summary-2015-09-01.pdf; https://www.ctti-clinicaltrials.org/files/pgctrecs.pdf. Published 2015.
11. Friends of Cancer Research. Progress for patients. Advocacy alliance and education. https://www.progressforpatients.org/education.
12. Bayer Oncology Brunn C. Making real progress on infusing the patient voice into oncology clinical development. *Medium*. https://medium.com/@bayerus/making-real-progress-on-infusing-the-patient-voice-into-oncology-clinical-development-18b3cf6f7fe4. Published August 2017.

DIA

Therapeutic Innovation
& Regulatory Science
1-9

DOI: 10.1177/2168479018774533
tirs.sagepub.com

Accelerating Pediatric Cancer Drug Development: Challenges and Opportunities for Pediatric Master Protocols

Tahira Khan, PhD[1], Mark Stewart, PhD[2], Samuel Blackman, MD[3], Raphaël Rousseau, MD, PhD[1], Martha Donoghue, MD[4], Kenneth Cohen, MD, MBA[5], Nita Seibel, MD[6], Mark Fleury, PhD[7], Bouchra Benettaib, MD[8], Raleigh Malik, PhD[9], Gilles Vassal, MD, PhD[10], and Gregory Reaman, MD[4]

Abstract

Although outcomes for children with cancer have significantly improved over the past 40 years, there has been little progress in the treatment of some pediatric cancers, particularly when advanced. Additionally, clinical trial options and availability are often insufficient. Improved genomic and immunologic understanding of pediatric cancers, combined with innovative clinical trial designs, may provide an enhanced opportunity to study childhood cancers. Master protocols, which incorporate the use of precision medicine approaches, coupled with the ability to quickly assess the safety and effectiveness of new therapies, have the potential to accelerate early-phase clinical testing of novel therapeutics and which may result in more rapid approval of new drugs for children with cancer. Designing and conducting master protocols for children requires addressing similar principles and requirements as traditional adult oncology trials, but there are also unique considerations for master protocols conducted in children with cancer. The purpose of this paper is to define the key challenges and opportunities associated with this approach in order to ensure that master protocols can be adapted to benefit children and adolescents and ensure that adequate data are captured to advance, in parallel, the clinical development of investigational agents for children with cancer.

Keywords

pediatric cancer, clinical trial design, master protocol, regulatory policy, drug development

Introduction

Despite substantial progress over the past several decades, challenges remain with the pace and breadth of pediatric oncology drug development.[1,2] Advancements in the scientific and clinical understanding of the molecular pathogenesis underlying pediatric tumor initiation and progression provide new therapeutic opportunities that could leverage these findings to better inform clinical development strategies and the design of pediatric clinical trials. Compared to adult cancers, the relative rarity and differences in etiology and natural history of childhood cancers present logistical challenges for drug discovery, pre-clinical and clinical development, and clinical trial design and conduct.[3] Historically, pediatric patients have generally not been included in adult trials. Moreover, most pediatric clinical trials for new therapeutics are initiated after their adult counterparts are completed, often years after the new therapeutic is already approved for use in adults. At that point, off-label use may be prevalent, especially in cancers that span

[1] Genentech Inc, a member of the Roche Group, South San Francisco, CA, USA
[2] Friends of Cancer Research, Washington, DC, USA
[3] Silverback Therapeutics, Seattle, WA, USA
[4] Food and Drug Administration, Silver Spring, MD, USA
[5] Department of Pediatrics and Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD
[6] USA National Cancer Institute, Bethesda, MD, USA
[7] American Cancer Society Cancer Action Network Inc, Washington, DC, USA
[8] Celgene Corporation, Summit, NJ, USA
[9] Drug Information Association, Washington, DC, USA
[10] Department of Clinical Research, Institut Gustave Roussy, Paris-Sud University, Paris, France

Submitted 21-Dec-2017; accepted 9-Apr-2018

Corresponding Author:
Mark Stewart, Senior Science Policy Analyst, Friends of Cancer Research, 1800 M St NW Ste. 1050s, Washington, DC 20036, USA.
Email: mstewart@focr.org

the adult and pediatric age ranges. This may hinder accrual to trials and decrease the opportunity to collect meaningful data about safety and tolerability, as well as effectiveness in pediatric cancer patients.[4] In response, incentives and regulatory requirements have been created in both the United States (US) and Europe (eg, Pediatric Research Equity Act [PREA] and the Best Pharmaceuticals for Children Act [BPCA], European Medicines Agency [EMA] Pediatric Rule) to help promote pediatric drug development, and stakeholders are working together to improve drug development for pediatric cancers.

One of the key challenges in conducting pediatric cancer trials is the rarity of pediatric cancers and the substantial logistical challenges in designing and executing these trials. A master protocol is a clinical trial model consisting of multiple investigational treatment cohorts that uses a molecular screening approach to assign patients to receive a targeted therapy in one of these cohorts based on their unique tumor profiles and the target or mechanism of action of the drug.[5-7] Master protocols can employ an umbrella, basket, or platform design to study a single targeted therapy in multiple diseases, multiple targeted therapies in a single disease, or multiple targeted therapies in diseases spanning multiple histologic subtypes harboring one or more molecular features.[7] In platform trials, arms testing various agents are added or removed from these types of trials through a well-defined clinical prioritization process; the overall clinical trial infrastructure can be maintained for rapid inclusion of new agents without starting or writing new protocols. Master protocols have the ability to allow multiple stakeholders and sponsors to work together in order to conserve resources, improve efficiency, safeguard patient safety, and streamline regulatory review.[7] Master protocols may help address some challenges in conducting clinical trials in pediatric and adolescent patient populations. These protocols may accelerate pediatric drug development by matching targeted therapies to children by facilitating efficient and effective testing of one or more targeted agents in relatively small patient subpopulations. This allows for development of new therapeutics with greater efficiency, while limiting exposure of children to potentially ineffective therapies.

In cooperation with key stakeholders from industry, academia, regulators, and patient and disease advocacy groups, Friends of Cancer Research convened a half-day forum on February 21, 2017, to assess the feasibility, designs, and potential limitations of master protocols and platform trials for the investigation of new therapies for childhood cancers. The findings and recommendations of this forum are outlined in this paper.

Considerations for Master Protocols for Pediatric Cancer Trial Design

Selecting the appropriate trial design is key to ensure studies are feasible for small populations and efficiently leverage existing resources. Master protocols can be designed to include a specific histologically defined tumor type or can be histology agnostic by including various tumor types harboring a specific molecular alteration or target.[7] If the tumor biology is not well understood, an "all-comer" strategy may be employed with patient eligibility being subsequently adapted to any observed correlations between the molecular biomarker and efficacy. This trial design serves as an efficient method for determining if there is promise in the treatment for either the disease or molecular alteration or target being explored (depending on which approach is employed). As is true for any clinical trial studying a targeted therapy in a biomarker-defined population, understanding the underlying pathophysiology, validation of biomarkers for enrollment and/or treatment assignment, qualification and utilization of in vitro diagnostic tests for patient selection that have sufficient sensitivity and performance characteristics for their context of use, and selection of appropriate endpoints are needed to maximize the efficiencies master protocols can afford. Clearly defined decision making and a robust clinical trial infrastructure should be in place to allow for the selection and inclusion of new agents for various clinical trial phases, and a process to enable the screening and assigning patients to new treatment arms will be critical.[8] Design features of select pediatric trials are summarized in Table 1.

A thoughtful and clear trial design is critical for successful multiagent and multitumor type platform trials. As such, the definition and use of common elements of the trial should be a focus during the development of a master protocol template. These elements include screening and inclusion/exclusion criteria, biomarker and enrichment strategies, statistical analysis plans, dose-limiting toxicity and/or stopping rules, study endpoints, and the potential use of the data for regulatory submission and registration.

Master Protocol Screening

Many childhood cancers contain genomic alterations that may predict response to targeted therapies; however, childhood cancers generally have lower mutation rates than adult cancers and the driver mutations of most childhood cancers are distinct from those tumor types more commonly seen in adults.[9,10] Nonetheless, molecular aberrations known to cause adult tumors may be important for rare subpopulations of pediatric tumors. Generating information from genomic and biomarker analysis of pediatric tumors as well as the development of preclinical models, including pediatric models, that can establish mechanistic proof-of-concept for targeted therapies is crucial to improve drug development and increase the likelihood of identifying potential effective treatments. The importance of this approach is highlighted by recent discoveries in the underlying pathophysiology of diffuse intrinsic pontine glioma (DIPG), a brainstem tumor that has historically been one where biopsies have not been frequently performed due to the perceived risk to the child. More recently, protocols in which stereotactic biopsy were performed at diagnosis has led to the recent discovery of specific oncohistones not previously recognized in this, or any other, tumor.[11] The development and

Table 1. Comparison and Description of Select Pediatric Trials with a Master Protocol.

Features	AcSé-ESMART	iMATRIX[a]	Pediatric MATCH	TAPUR
Clinicaltrials.gov identifier	NCT02813135	NCT02541604; NCT02639546	NCT03155620	NCT02693535
Study type	Phase 1/2 Dose finding, safety assessing, and activity estimating	Phase 1/2 Dose finding, safety assessing, and activity estimating	Phase 2 Dose finding, safety assessing, and activity estimating	Phase 2 Safety assessing and activity estimating
Eligibility criteria	Children, adolescents, and young adults (up to 18 y old; patients 18 y and older may be considered for inclusion) with refractory or recurrent malignancies (solid tumors, brain tumors, and hematologic malignancies)	Children and adolescents (up to 30 y old) with recurrent or refractory solid tumors, brain tumors and lymphomas, with plans to expand to liquid tumors	Children and adolescents (up to 21 y old) with recurrent or refractory solid tumors, non-Hodgkin lymphomas, or histiocytoses with measurable disease	Lymphoma, non-Hodgkin multiple myeloma Advanced solid tumors (proposed amendment to include adolescents [12-18 y old] when scientifically justified)
Screening method	Advanced tumor molecular profiling (eg, whole-exome and RNA sequencing) is an inclusion criteria. Performed in specific matching trials before inclusion into the trial	Specific to each treatment and based on the mechanism of action of the molecule	Biomarker profiling protocol	Identified genomic variation in patients' tumors will be matched to drugs on trial. If there is no match, the Molecular Tumor Board can help identify other treatment options
Primary study endpoint	Objective response rate	Objective response rate	Objective response rate	Objective response rate
Treatment assignment	In each therapeutic arm (single agent or in combination) either 100% of patients receive a matched agent or only 50% (enrichment) depending on the biomarker	Treatment allocation decisions will be informed by data from completed gate assessments	Computerized algorithm based on levels of evidence for the target and the drugs for the specific target	Agents matched to identified genomic variant in patient tumor
Sponsor	Academic	Industry	Academic-government	Academic-nonprofit organization
Location	Europe	Europe and United States	United States	United States
Governance structure	A steering committee that consists of the coordinating sponsor, investigators, statistician, funding partners, and if trial progresses to phase III, the participating pharmaceutical company	A steering committee that consists of external experts	The NCI-COG Pediatric MATCH Steering Committee consisting of members from NCI, COG, and FDA	TAPUR Data and Safety Monitoring Board

Abbreviations: AcSé-ESMART, European Proof-of-Concept Therapeutic Stratification Trial of Molecular Anomalies in Relapsed or Refractory Tumors; COG, Children's Oncology Group; NCI, National Cancer Institute; Pediatric MATCH, Pediatric Molecular Analysis for Therapeutic Choice; TAPUR, Targeted Agent and Profiling Utilization Registry.

[a]The iMATRIX trial currently consists of 2 standalone clinical protocols built as modules of a master protocol while discussions with regulatory authorities evaluate the feasibility of a single clinical trial application.

validation of liquid biopsies may also facilitate accruing molecular knowledge as these methods allow for repeated sampling of tumor DNA over time in children using blood draws—a far less invasive method. It is important to mention that generating meaningful data from preclinical models, including pediatric tumor models, if available, is crucial to determining target actionability, safety of the test agent and for predicting responsiveness of the tumor to target inhibition. Preclinical testing consortia or public private partnerships (eg, the NCI Pediatric Preclinical Testing Consortium in the US and Innovative Therapies for Children with Cancer Pediatric Preclinical Proof-of-Concept Platform in the European Union [EU]) may serve as mechanisms to help identify molecular agents that are most likely to benefit children with cancer.

Pediatric oncologists and other experts (eg, genomic scientists, pathologists, statisticians, patient advocates) should be involved throughout the design and execution of the clinical trial to optimize patient safety, trial feasibility, and a

scientifically rigorous development strategy and study design. Because children and adolescents with cancer are often evaluated and treated at specialized institutions, there is increased opportunity for the appropriate collection of specimens and sharing of specimens and clinical information to help better inform the development of new therapies and screening protocols. However, to be screened, there must be tissue available from the time of recurrence or progression. Depending on the type of cancer, physicians may be hesitant to perform a biopsy knowing that a molecular match is likely to be low. Thus, opportunities for obtaining genomic information from tumors without tissue biopsies to expand potential treatment populations should be explored further, while also ensuring appropriate enrichment strategies and data integrity for analytical and clinical interpretation. Therefore, physicians, patients, and families will need to determine which study approach (eg, biomarker-driven or all-comers approach) is most optimal.

Biomarker and Enrichment Strategies

During master protocol development, investigators and sponsors need an optimal biomarker strategy for each agent in the protocol, and as such, some trials may have multiple strategies for patient selection within the same trial. Ideally, a strategy for utilization of validated biomarkers should be well defined in advance of protocol design. Some protocols may follow a biomarker-driven enrichment strategy, in which patients are screened for the presence or absence of a biomarker, and only patients with (or without) the specified biomarker are included in the study. Other protocols may utilize an "all-comer" approach, in which all patients meeting eligibility criteria, regardless of biomarker status, enter the study, with an intent to ascertain the biological features underlying response in retrospect.[12] The latter approach may be considered when there is an insufficient understanding of the tumor biology, biomarker, or assay to confidently exclude or include patient subpopulations at the time of trial initiation. Retrospective analysis of an "all-comer" population, though, may be hypothesis generating and help inform enrichment strategies for future trial phases. Studies with enrichment strategies can help both patients and sponsors. For pediatric patients, these studies increase the possibility of direct treatment benefit in early phase clinical trials. For sponsors, these studies enable a better understanding of the safety and activity of a new agent or combination therapy in targeted and molecularly characterized subpopulations to better inform future drug development. The success of this approach is exemplified in recent clinical trials for entrectinib and larotractinib for tumors harboring tropomyosin receptor kinase (TRK) fusions, and dabrafenib for low-grade gliomas that have a BRAF V600E mutation.[13]

Ultimately, the enrichment strategy selected will depend on the validity of the biomarker and the clinical tractability of utilizing it in the setting of a clinical protocol. When determining which type of strategy to use, sponsors and investigators should consider if the trial is for discovery or exploratory purposes, or eventual product registration. Trials run with hypothesis-generating or exploratory intent should be sized adequately to provide confidence around retrospective analysis. Master protocols seeking regulatory approval require specific considerations with respect to study design, data analysis, and validation and qualification of the biomarker assay to accumulate the substantial evidence of safety and efficacy required for regulatory approval.[14] Finally, it is important to recognize that many factors—a patient's clinical status, the molecular characteristic(s) of the tumor, the availability of clinic-ready assay, and the availability of a specific targeted agent, antitumor activity, and acceptable safety profile—need to align for patients to potentially benefit from genomic profiling–directed clinical studies.

Statistical Analysis Plans

The current generation of planned and open master protocols for pediatric cancers generally do not have control arms as they are primarily designed to screen for antitumor activity or for estimation of antitumor activity of therapies that could be developed in subsequent, larger trials. Comparison with a control therapy may not be necessary depending on the objective of the trial, such as when the objective of the master protocol is to rapidly identify which new targeted agents have the greatest potential for therapeutic efficacy for a specific molecularly defined subgroup of tumor(s). In these instances, simply predefining a treatment magnitude that would warrant future study in a randomized trial may be appropriate; however, such master protocols should prespecify the rules for response assessment and determination of the levels of antitumor activity that would warrant future study. To limit potential exposure of patients to ineffective therapies, gating strategies that assess safety, pharmacodynamics, and efficacy at prespecified times may be utilized to guide new therapeutic agents through the trial efficiently. For example, the iMATRIX master protocol proposal, which is similar to the Simon 2-stage design, has three gated assessments: PK and initial safety assessment, initial response assessment, and additional response assessment.[14,15] These gated assessments take place once a prespecified number of patients have been treated with the molecule.

Eligibility Criteria

Although establishing appropriate eligibility criteria is crucial for any clinical trial, it is particularly important for pediatric oncology master protocols, given that many will aim to enroll pediatric patients with a small molecularly defined subset of an already rare cancer. Therefore, overly restrictive eligibility criteria that are not scientifically justified may prevent efficient and successful enrollment of pediatric patients with cancer. Eligibility requirements for pediatric studies should reflect the safety profile of the targeted agents in the study and the unique considerations of pediatric populations, such as the potential risk of developmental toxicities and metabolic differences between age groups that can result in differential drug exposure

levels. It is recommended that sponsors and investigators seek input from regulators to ensure patient eligibility criteria are scientifically appropriate and provide the necessary balance and preservation of patient safety and trial accessibility. It is also important to consider opportunities for inclusion of adolescents and young adults in adult trials when it is scientifically appropriate, as these older pediatric patients are less susceptible to adverse events that impact growth and development.[4,16]

Stopping Rules

As with all early-phase trials, pediatric oncology master protocols should include stopping rules to appropriately mitigate the risk of serious adverse reactions and limit unnecessary exposure of patients to an investigational agent if little antitumor activity is observed in a predefined number of patients. Early-phase trials must also include appropriate monitoring and stopping rules for lack of efficacy or disproportionate risk-benefit profiles to protect the safety of pediatric patients. Gating strategies employed by master protocols can help identify the most promising agents to move forward, while also identifying poorly performing agents and stopping enrollment before too many patients are exposed to an inappropriate treatment or creating unnecessary competition for enrolling children with rare cancers into other clinical trials.

Study Endpoints

Defining meaningful, valuable, and valid endpoints for pediatric oncology master protocols necessitates collaboration between sponsors, investigators, and regulators. The majority of early phase (Phase 1 or Phase 2) pediatric trials have objective response rate as the primary endpoint. Duration of response is an important secondary measure for all studies, where applicable. When developing the study and determining endpoints, it is recommended that sponsors and investigators consider what is most clinically meaningful to pediatric patients and their parents (or guardians), especially if the study transitions to a trial that may support drug approval. Endpoint considerations for future regulatory approval should also be discussed among collaborators; although the goal should ultimately be to develop curative approaches for pediatric cancer patients. As with all studies, sponsors and investigators should "begin with the end in mind."

Registrational Intent

Sponsors need to be prepared to address licensing considerations and timing of registration studies, as there may be potential for accelerated approval from the results of an expansion cohort of a signal-seeking master protocol trial, particularly when treatment effects are especially robust and durable. Sponsors should consider how the data derived from a master protocol fit into the context of the overall pediatric development plan of a specific agent or combination. It is possible that the demonstration of a durable and large antitumor effect in a small molecularly defined subset of patients may serve as the basis of an accelerated approval in that patient population.

A master protocol development strategy should factor in pediatric laws and regulations to ensure that each molecule in an ongoing study meets its pediatric obligations and/or qualifies for incentives under the pediatric study/investigation plans. If there is registrational intent for agents in the study, the sponsor should be prepared to have early conversations with health authorities and request joint discussion and consideration of development plans by FDA and the EMA prior to initiation of the clinical trial to gain alignment in the number and types of studies required to fulfill regulatory obligations for informing labeling and qualifying for incentives, and to maximize the efficiencies of a master protocol design. Attention must also be paid to potential regulatory requirements for development of a companion in vitro diagnostic device that reliably identifies patients with a specific molecular alteration that can benefit from the drug.

Logistical and Operational Considerations

Master protocols have the potential to increase operational efficiency of clinical trials and accelerate the availability of new treatment options for children with cancer. For patients, these trials potentially provide earlier access to novel targeted agents that may exert antitumor activity resulting in improved outcomes. Benefits for sponsors and investigators include standardized protocols for subarms, shared costs and resources between partners, and consistency in the collection, analysis, and interpretation of data for health authorities. Despite the efficiencies that can be gained through master protocols, they do present unique challenges.[10] Master protocols necessitate some level of standardization among substudies, and industry partners involved in collaborative master protocols need to be willing to relinquish some control or specificity related to clinical operations that they typically implement in their own sponsored trials. Providing the necessary training and education to trial site participants may be necessary to ensure uniformity in data collection and patient enrollment. In addition, consideration should be given to trial site locations to guarantee patient access. Working with sites in multiple countries can also create new logistical challenges associated with differing review boards, ethics committees, and regulations. National and international cooperative groups may be one way to bring these stakeholders together to promote a coordinated effort, provide the necessary infrastructure, and evaluate the clinical readiness of potential trial sites.

Master protocols will require more extensive planning than traditional trials because multiple stakeholders are often involved and due to the complexity of infrastructure required for initiation and efficient implementation of these trials. Collaboration typically entails a learning period for partners who are not used to working together and may have differing priorities, and it is important for stakeholders to be aware and prepared for this educational period. Early interactions with

regulatory authorities during the master protocol planning process are necessary, especially if the master protocols are complex, involve in vitro diagnostics, or have potential for registrational intent.

Implementing a Target and Agent Prioritization Criteria in Multisponsor Trials

Because pediatric cancers are relatively rare, testing similar agents from different sponsors simultaneously in an uncoordinated effort can be difficult or impossible as they progress through the clinical trial phases. During the study design development, sponsors and researchers should establish appropriate methods and a prospective mechanism for prioritizing agents for inclusion in trials. For example, The Children's Oncology Group (COG)-NCI Pediatric MATCH (Pediatric Molecular Analysis for Therapeutic Choice) trials created a committee composed of representatives from the COG, NCI, and FDA to evaluate the strength of evidence that a particular target and corresponding therapeutic product was relevant to pediatric cancer, as well as evaluate any nonclinical or early adult data of activity and safety.[17] When prioritizing agents for inclusion in pediatric cancer trials, trial sponsors should consider the mechanism of action of the agent, rationale for target and agent selection, evaluation of relevant testing platforms, formulation for use in pediatric subpopulations, safety profile, regulatory requirements, and existence of ongoing trials that may have overlap or compete for patient enrollment and completion of the trial. One strategy being implemented in the EU by the ACCELERATE platform and the EMA is the organization of Pediatric Strategy Forums. A Pediatric Strategy Forum is a scientific meeting to share information and advance learning among all stakeholders (eg, academia, industry, parents, regulators) in a precompetitive setting, which will inform a pediatric drug development strategy and subsequent decisions that will greatly facilitate the introduction of innovative treatments into the standard of care of children with rare cancers. The first Pediatric Strategy Forum was held in January 2017, and the Forum discussed the role of ALK inhibition in childhood malignancies and reviewed available data regarding 6 ALK inhibitors, which are approved or in development for adults.[18] Indeed, not all 6 ALK inhibitors can be developed in pediatric cancers such as inflammatory myofibroblastic tumor, anaplastic lager cell lymphoma, and neuroblastoma, which are considered rare or even extremely rare. Priority should be given to agents that have a mechanism of action that has been shown to target a driver responsible for the growth or progression of one or more pediatric cancers; sufficient nonclinical data on antitumor activity, including comparison with agents of the same class on pediatric tumor models; and nonclinical and clinical data in adults to inform an appropriate starting dose, safety information, and monitoring plan to mitigate patient risk. For oral compounds, availability of an age-appropriate formulation for the pediatric population is an added value. Utilizing preclinical and clinical data to identify targets and pathways to inform initial testing in children in a concerted effort may also help alleviate some issues related to prioritization. It is essential that pediatric oncologists are involved early in discussions prioritizing the pipeline of available drugs followed by the development and implementation of master protocols to leverage their expertise in pediatric populations. It should also be noted that prioritization can introduce additional regulatory complexities that will need further discussion, such as whether there would be opportunities for a molecular agent that is not given high priority to fulfill regulatory requirements or qualify for exclusivity.

The Role of a Multistakeholder Decision-Making Body

Given the considerable number of agents that may have activity in pediatric cancers, the rarity of pediatric cancer cases, and the rarity of targetable alterations in these cancers, increased communication and collaboration among stakeholders is essential to ensure master protocols are adequately designed to efficiently develop drugs that address the needs of the pediatric oncology community. Master protocol platform trials may involve multiple sponsors coming together to test agents in a single trial. Thus, an establishment of a multistakeholder decision-making body, or governance body, can help identify appropriate diseases to study, prioritize drugs for inclusion in master protocols, assign treatment arms, and review the clinical data generated in the study. However, it may be necessary to have two separate decision-making bodies: one that oversees prioritization decisions regarding which agents are incorporated into the trial and another one that provides trial oversight and analysis. In the context of a master protocol, it is essential that there is close alignment among health authorities regarding molecule prioritization decisions and that such prioritization decisions will consider regional regulatory requirements that sponsors and health authorities can prospectively agree upon prior to inclusion of a molecule in the master protocol. The pediatric community is cognizant of the need for collaborative efforts to study, treat, and cure pediatric cancers. Several cooperative infrastructures currently exist that may provide the necessary infrastructure for study conduct and management for these types of trials. Examples of such cooperative groups include COG, Pediatric Oncology Experimental Therapeutics Investigators' Consortium (POETIC), the European consortium for Innovative Therapies for Children with Cancer (ITCC), and many other academic clinical pediatric oncology or industry consortia to expedite the evaluation of drugs for children with cancer. It would be ideal for these consortia to partner with each other to develop a cohesive strategic oversight plan.[19]

It is essential that decision-making bodies encourage and engage in transparent communication and oversight among all collaborators that can build on the synergies between the differing organizations. Data shared among collaborators can be used to inform future research decisions. For example,

academic collaborators can use positive and negative outcome data analyses to determine where to prioritize their research efforts and inform the development of new agents. The governance body may need to provide guidance to clinical operations structure to meet some of the unique demands of a multistakeholder clinical trial. Invoking a successful culture change requires a commitment and understanding of the primary end goal from those involved.

Regulatory and Policy Considerations for Pediatric Drug Development

Both the US and the EU have pediatric laws and regulations in place that mandate and incentivize pediatric studies in some instances and provide waivers in others. Designing a master protocol to encompass regional pediatric regulatory requirements can be challenging. However, these potential barriers can be overcome with a coordinated global drug development strategy and collaboration among key stakeholders including health authorities. More recently, additional legislative or regulatory measures have been introduced in the US and EU to speed development of better drugs for children with cancer. In addition, utilization of novel trial designs such as master protocols can help increase efficiencies in the development of novel targeted drugs for rare pediatric cancer patients and also enable companies to comply with pediatric study requirements or qualify for incentives. Thus, in the US, pediatric data derived from master protocols such as the NCI pediatric cancer MATCH study may be used to fulfill part or all of the regulatory requirements and/or to qualify for incentives.

Global Clinical Trial Processes

Implementation of master protocols in pediatrics can also become complicated because of regional differences in clinical trial implementation processes. A master protocol that has several different treatment arms may be implemented in the EU under a single Clinical Trial Application (CTA) and in the US under a single master Investigational New Drug (IND) application. However, the addition of new investigational agents in an ongoing basis to a single overarching master protocol requires that regulatory agencies and clinical trial implementation groups in different regions consider further simplification and/or harmonization of master protocol review and implementation procedures. This will encourage sponsors to participate in global multiagent, multisponsor trials with master protocols.

In the United States, the FDA Pediatric Subcommittee of the Oncologic Drugs Advisory Committee Consensus statement indicates, "Pediatric oncology drug development should generally be coordinated with oncology drug development for adults, as part of an overall development plan." However, there are some challenges associated with the execution of pediatric oncology trials, as previously described. Regulators, investigators, and sponsors should continue to collaborate and consider opportunities for coordinated pathways to streamline pediatric cancer drug development. Regulators also need to engage industry sponsors and key academic and community opinion leaders on the challenges associated with pediatric oncology drug development so this is taken into consideration when working within the legislative framework that is intended to promote the expedited development of new drugs for pediatric cancers. There are opportunities for sponsors to request collaborative discussion and input on pediatric development from international health authorities; for example, parallel scientific advice between the FDA and EMA can be requested. Sharing of Common Commentary from joint FDA and EMA pediatric cluster calls with sponsors also helps inform the scientific and regulatory strategy of pediatric study plans for optimal implementation of pediatric clinical trials. In addition, indicating on http://clinicaltrials.gov which trials are being performed to meet a written request in the US or a pediatric investigation plan (PIP) in the EU may also help sponsors coordinate the timing of similar trials. Engaging stakeholders during the policy development process will provide unique perspectives on the impact of current and future regulations on pediatric drug development. Together, patients, regulators, industry, and academia can identify solutions for increasing the efficiency of pediatric drug approvals.

Global Regulatory Policies

Two legislatives acts in the US—Pediatric Research Equity Act of 2003 (PREA) and Best Pharmaceuticals for Children Act of 2002 (BPCA), made permanent by The Food and Drug Administration Safety and Innovation Act (FDASIA) in 2012—provide a requirement and incentive model, respectively, to perform pediatric studies on drugs being developed for use in adults.[20,21] PREA mandates that pediatric studies be performed only in the overlapping adult indication under review, if a pediatric population exists, and exempts drugs with orphan designation from pediatric study requirements. However, in 2017, US Congress passed the FDA Reauthorization Act of 2017 (FDARA), which included Title V, also known as the Research to Accelerate Cures and Equity (RACE) for Children Act. Title V [Sec 504 (a)(3)(A)] of FDARA amended PREA to support the early evaluation of potentially effective drugs by requiring pediatric investigation of appropriate new drugs intended for the treatment of an adult cancer, and if the corresponding molecular target is substantially relevant to the growth or progression of a pediatric cancer(s). FDARA also removes the exemption from pediatric studies for certain oncology drugs with orphan designation. These provisions under FDARA are aimed at addressing the unmet medical need for children with cancer, particularly pediatric cancers that have a shared molecular target with an adult cancer. For financial incentives, BPCA encourages sponsors to conduct clinical trials in pediatric diseases by providing, upon completion and submission of pediatric studies to the FDA, additional 6 months period of market protection at the end of listed patents and/or data exclusivity for the agent being tested. For biologics, the 6 months of pediatric exclusivity only extends the data

exclusivity period, and therefore only provides market protection if the data exclusivity is longer than the patent. Another incentive for rare pediatric diseases includes the rare pediatric priority review voucher program, which can be used to speed the review of new drug applications by four months.[22]

Since the initiation of the EU Pediatric Regulation in 2007, pediatric development is obligatory in the region.[23,24] The EU Pediatric Regulation unifies the incentive and requirement for pediatric drug development under one regulation. The Pediatric Regulation applies to new products, new indications, new routes of administration, and new pharmaceutical forms of existing products that are protected by a Supplementary Protection Certificate or a patent that qualifies for it. Rare diseases and orphan-designated products are not exempt. Important issues have been identified in the field of pediatric oncology that delay or unjustifiably waive the development of therapeutic innovations. In 2015, in recognition of the existence of certain cancers in children, the EMA revoked pediatric class waivers for all medicines for the treatment of kidney and renal pelvis carcinoma and for liver and intrahepatic bile duct carcinoma and a revised list of class waivers will come into effect in July 2018. Differences do exist between EU and US pediatric drug development, but both have the shared goal of encouraging drug development and assessing safety and efficacy for pediatric patients.

Conclusions

As pediatric trials look to replicate the successes seen and efficiencies gained in adult targeted therapy trials, it is vital to bring the global pediatric community together to ensure trials are adequately designed and that specific genomic characteristics can be adequately assessed to maximize the potential for therapeutic benefit. Genomic profiling might reveal new therapy options for pediatric patients, such as treatment with an investigational agent in a clinical trial or use of a targeted therapy drug that has been approved by FDA or EMA for a different cancer. Master protocols provide a path for mechanism of action-based approaches and utilizing genomic profiling to guide molecularly targeted drug choice in different pediatric cancers in a single study, as well as the opportunity to speed up the development of therapies in a rational and scientifically-driven manner. The development of successful pediatric master protocols will require the cooperation and engagement of investigators, cooperative groups, industry, and regulatory agencies to overcome barriers and unify efforts. Developing master protocols that incorporate the needs of all stakeholders will help incentivize participation in innovative trials that can help provide timely evaluation of new and potentially more effective therapeutic options for children with cancer.

Author Note

Manuscript topic and outline was discussed on February 21, 2016, at the Friends of Cancer Research Meeting "Accelerating Pediatric Drug Development."

Tahira Khan's and Raphaël Rousseau's affiliations have changed since the time this article was written. Tahira Khan is now with Nektar Therapeutics (San Francisco, CA), and Raphaël Rousseau is now with Gritstone Oncology (Emeryville, CA).

Acknowledgments

This article was developed based on public discussions at the Friends of Cancer Research Public Forum on Accelerating Pediatric Drug Development, held on February 21, 2017. The authors would like to thank the participants of this public meeting, and the panelists who facilitated the discussions and contributed to the information provided in this article: Peter Adamson (Children's Hospital of Philadelphia), Bouchra Benettaib (Celgene), Kenneth Cohen (Johns Hopkins University), Martha Donoghue (FDA), Mark Fleury (ACS), Katherine Janeway (Dana-Farber Cancer Institute), Tahira Khan (Nektar Therapeutics), Shakuntala Malik (NCI), Pam Mangat (ASCO), Kenan Onel (Feinstein Institute for Medical Research), Gregory Reaman (FDA), Giles Robinson (St Jude Children's Research Hospital), Raymond Rodriguez-Torres (Live Like Bella), Raphaël Rousseau (Gritstone Oncology), Eric Rubin (Merck), Nita Seibel (NCI), and Gilles Vassal (Institut Gustave Roussy).

Declaration of Conflicting Interests

Dr Tahira Khan was an employee of Genentech, Inc, a member of the Roche group at the time of the preparation of the manuscript. Dr Khan is currently an employee of Nektar Therapeutics. Dr Raphaël Rousseau was an employee and shareholder of Genentech, Inc, a member of the Roche group at the time of the preparation of the manuscript. Dr Rousseau is currently an employee and shareholder of Gritstone Oncology, Inc. Dr Samuel Blackman is an employee of Silverback Therapeutics. Dr Bouchra Benettaib is an employee of Celgene Corporation. Dr Gilles Vassal receives funding or other support from Bristol Myer Squibb, Genentech, Inc, AstraZeneca, Novartis Pharmaceuticals, Celgene, Inc, and Bayer Corp, outside the submitted work.

Funding

No financial support of the research, authorship, and/or publication of this article was declared.

ORCID iD

Mark Stewart, PhD http://orcid.org/0000-0002-4847-0736

References

1. Schmid I, Schweinitz DV. Pediatric hepatocellular carcinoma: challenges and solutions. *J Hepatocell Carcinoma*. 2017;4:15-21.
2. Pui CH, Gajjar AJ, Kane JR, Qaddoumi IA, Pappo AS. Challenging issues in pediatric oncology. *Nat Rev Clin Oncol*. 2011;8:540-549.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2017;67:7-30.
4. Chuk MK, Mulugeta Y, Roth-Cline M, Mehrotra N, Reaman GH. Enrolling adolescents in disease/target-appropriate adult oncology clinical trials of investigational agents. *Clin Cancer Res*. 2017;1:9-12.
5. Herbst R, Gandara D, Hirsch F, 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;21:1514-1524.
6. Sleijfer S, Bogaerts J, Siu LL. Designing transformative clinical trials in the cancer genome era. *J Clin Oncol*. 2013;31: 1834-1841.
7. Woodcock J, Lavange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377: 62-70.
8. Lawrence M, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499:214-218.
9. Mody R, Wu Y, Lonigro R, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. *J Am Med Assoc*. 2015;314:913-925.
10. Redman MW, Allegra CJ. The master protocol concept. *Semin Oncol*. 2015;42:724-730.
11. Khuong-Quang D, Buczkowicz P, Rakopoulos P, et al. K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol*. 2012;124:439-47.
12. Khan T, Karski E, Caron H, et al. Mechanism-of-action-based master trial approach in pediatric oncology. 2017. https://www.focr.org/sites/default/files/iMatrix%20Concept%20Paper.pdf. Accessed September 2017.
13. Khotskaya Y, Holla V, Farago A, Mills Shaw K, Meric-Bernstam F, Hong D. Targeting TRK family proteins in cancer. *Pharmacol Ther*. 2017;173:58-66.
14. US Food and Drug Administration. Guidance for industry. Enrichment strategies for clinical trials to support approval of human drugs and biological products. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf. Published 2012.
15. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10:1-10.
16. Gore L, Ivy P, Balis F, et al. Modernizing clinical trial eligibility: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Minimum Age Working Group. *J Clin Oncol*. 2017;35:3781-3787.
17. Allen C, Laetsch T, Mody R, et al. Target and agent prioritization for the Children's Oncology Group–National Cancer Institute Pediatric MATCH Trial. *J Natl Cancer Inst*. 2017;109.
18. European Medicines Agency. Report: Paediatric strategy forum for anaplastic lymphoma kinase (ALK) inhibition in paediatric malignancies. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2017/06/WC500228940.pdf. Published 2017.
19. Vassal G, Rousseau R, Blanc P, et al. Creating a unique, multi-stakeholder Paediatric Oncology Platform to improve drug development for children and adolescents with cancer. *Eur J Cancer*. 2015;51:218-224.
20. US Food and Drug Administration. Safety and Innovation Act of 2012, Pub L No. 112-144, 126 Stat 993. https://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf. Accessed September 2017.
21. US Food and Drug Administration. Pediatric Research Equity Act of 2003. http://www.fda.gov/downloads/drugs/developmentapprovalprocess/developmentresources/ucm077853.pdf. Accessed September 2017.
22. Rare pediatric disease priority review vouchers, draft guidance for industry. https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM423325.pdf. Accessed September 2017.
23. Regulation (EC) No 1901/2006 of the European Parliament and of the Council of 12 December 2006 on medicinal products for pediatric use and amending Regulation (EEC) No 1768/92, Directive 2001/20/EC, Directive 2001/83/EC and Regulation (EC) No 726/2004. Official Journal of the European Communities L 378/1.
24. Regulation (EC) No 1902/2006 of the European Parliament and of the Council of 12 December 2006 amending Regulation 1901/2006 on medicinal products for pediatric use. Official Journal of the European Communities L 378/20.

Early Evaluation of Molecularly Targeted Therapies for Childhood and Adolescent Cancer

Discussion Document for use during Friends of Cancer Research's meeting on February 20, 2018

Disclaimer:
Friends of Cancer Research prepared this discussion document with input from a multi-stakeholder working group representing a broad cross-section of the pediatric cancer community. This document does not reflect consensus views, opinions, or positions of the working group members or the institutions they represent. This document is meant to facilitate open discussion among different stakeholders.

Background

The Pediatric Research Equity Act (PREA) requires sponsors of new drug applications (NDA) or biologics license applications (BLA) (or supplements to applications) for a new active ingredient, indication, dosage form or dosing regimen, or route of administration, to submit *assessments** (Federal Food, Drug and Cosmetic Act (FD&C Act) Sec. 505B (a)(2), 21 USC 355c (a)(2), amended in the Food and Drug Administration Safety and Innovation Act (FDASIA) Public Law 112-144). The assessment consists of data gathered using appropriate formulations for each age group that are adequate to assess the safety and effectiveness of the drug or biological product for that indication, and support dosing and administration for each pediatric subpopulation for which the drug or biological product is safe and effective. When sponsors have data to demonstrate that assessments in the pediatric population are not feasible, they can obtain a waiver or deferral for completing the assessments in all or some of the *pediatric age groups*.

Until the passage of Title V of the FDA Reauthorization Act (FDARA) of 2017 (FD&C Act Sec. 505B (a)(3), 21 USC 355c (a)(3), Public Law 115-52), PREA had not been an effective mechanism to establish a requirement for the development of drugs for *pediatric cancers*, as most of the oncology drugs approved have been for treating cancers that occur in adults and not in children (e.g., cancers of the lung, prostate and breast). Therefore, drug sponsors would obtain waivers for conducting assessments of these drugs in pediatric patients. Additionally, drugs developed for rare cancer indications, which may occur in both adult and pediatric populations, are granted orphan drug designation and are thus exempted from PREA required studies.

While there was limited obligation to study investigational therapies in pediatric oncology, incentives have been put into place to promote the development of oncology products for pediatric cancer when these agents are in development or already approved for adult use. The Best Pharmaceuticals for Children Act (BPCA) is a voluntary mechanism which provides incentives in the form of six months of exclusivity for marketing to sponsors upon the completion and submission of pediatric studies that meet the terms of a written request from FDA (FD&C Act Sec. 505A, 21USC 355a, reauthorized in the FDA Amendments Act (FDAAA) Public Law 110-85). BPCA also allows FDA to request studies that cannot be realized under PREA because the applications or supplements are not subject to the requirements of PREA. To date, BPCA has been the primary mechanism used to develop oncology products for the treatment of malignancies in children and adolescents. Recent legislation (FDARA) has created a mechanism to further encourage the development of novel medicines that address the high unmet need in the pediatric population.

Molecularly targeted agents developed for adult cancers have greatly advanced the concept of precision medicine in oncology. As malignancies occurring in children and adolescents can harbor molecular abnormalities similar to those found in adult cancers, these agents may be relevant to the treatment of pediatric patients with cancer. Although large scale sequencing efforts, such as TARGET[1] and the Pediatric Cancer Genome Project (PCGP)[2] have provided evidence that the genetic and epigenetic repertoires of driver gene aberrations often differ between adult and pediatric cancers, a growing body of evidence suggests that genetic and other molecular biological vulnerabilities of certain adult cancers may recapitulate opportunities for the use of targeted therapies in select pediatric tumors[3,4].

Therefore, timely investigation of signals of activity of potentially useful targeted drugs and biologic agents under development and of their toxicities relative to the unique growth and developmental considerations of pediatric patients is often warranted for pediatric populations with cancer.

Children and adolescents are not smaller adults, and the efficacy, dosage form, and dosing regimen of targeted drugs developed for malignant disease, which develop in adults, cannot simply be extrapolated to different indications in the pediatric population. Thus, there is a need to more expediently identify and evaluate new anti-cancer agents which may be appropriate for investigation in pediatric cancers earlier in drug development programs. Title V of FDARA amended PREA to support the early evaluation of potentially effective drugs by requiring pediatric investigation of appropriate new drugs intended for adults with cancer. The investigations that FDA may require by statute are referred to as *molecularly targeted pediatric cancer investigations*. These investigations may include clinical studies designed to yield clinically meaningful pediatric study data, gathered using appropriate *formulations* for each age group for which the study is required, regarding dosing, safety and preliminary efficacy to inform potential pediatric labeling [FDARA Title V Sec 504 (a)(3)(A), FD&C Act Sec. 505B (a)(3)(A), 21 USC 355c(a)(3)(A)]. Importantly, Title V of FDARA also specifies that the requirement for early pediatric investigations of drugs directed at *molecular targets* considered substantially relevant to the growth or progression of a pediatric cancer be applied, even when the adult indication

has received an orphan designation, or when the adult cancer indication does not occur or is biologically different in the pediatric population (e.g., breast cancer).

The law directs the FDA, in collaboration with the NCI, to establish, publish, and regularly update a list of molecular targets considered on the basis of data the Agency determines to be adequate, to be substantially relevant to the growth or progression of pediatric cancers, and that may trigger the requirement for pediatric investigations [21 USC 355c (m)(1)(A)]. Molecular targets that are considered "not relevant" to the growth or progression of pediatric cancers will be placed on a second list [21 USC 355c (m)(1)(B)].

The FDA is mandated to convene a public meeting no later than one year after the date of the enactment of FDARA, to solicit views of physicians, academic researchers (including pediatric oncologists and rare disease specialists), patient advocates, industry, and other stakeholders for the establishment of the molecular targets lists [21 USC 355c (m)(2)(1)]. Future public meetings are planned for later in the year at the meeting of the Pediatric Subcommittee of the Oncologic Drugs Advisory Committee (ODAC).

In order to facilitate an early, informal opportunity for stakeholders to discuss the molecular targets list, Friends of Cancer Research (*Friends*) will convene a public meeting to discuss approaches for developing, updating, and applying the molecular target list. Friends has invited several stakeholders, including FDA, NCI, industry, academic researchers, clinical investigators, and patient advocates to discuss the implementation of the FDARA provisions. The objective of this document, which will be presented at the *Friends* meeting, is to discuss numerous ways to develop transparent and scientifically sound processes that address the following provisions:

1. Developing the molecular target lists: Forming frameworks of factors that may guide the definition of molecular targets as substantially relevant or not relevant to the growth or progression of one or more pediatric cancers
2. Updating the lists of molecular targets: Defining mechanisms and timelines by which such updates may occur
3. Applying the molecular target lists: Addressing key considerations in the application of the lists to pediatric cancer drug development

Framework of factors to consider for the development of a pediatric cancer molecular target list

Although there may be variations in the way "molecular target" is defined, for the purposes of this discussion document, we refer to a molecular target as a molecule in human cells that is intrinsically associated with a particular disease process, such as etiology, progression, and/or drug resistance. To be referred to as a target, there must be evidence that by engaging the target, either with a targeted small molecule, biologic product, or other treatment intervention, a desired therapeutic effect is produced that results in the alteration of the disease process. In other words, a molecule would not be referred to as a molecular target if there is no evidence to inform the hypothesis that its modulation (i.e. inhibition or activation) alters the disease.

In this discussion document, we are focusing on molecular targets in cancer, which can be further classified by subtype. One set of targets can be classified by whether they represent the result of specific gene abnormalities, are present in a critical biologically-related pathway of a gene abnormality, or exhibit a synthetic lethal relationship to a gene abnormality (*gene abnormality-based targets)*. Targets can also be intrinsic to the cancer cell lineage or developmental stage (*cancer cell lineage-based targets*), or they may be identified in non-cancer cells, such as normal immune cells or supporting cells contributing to the tumor micro-environment (*non-cancer cell targets*). A final category is targets present in the cancer cells as well as non-cancer cells that do not show cancer-specific genetic alterations, such as tubulin or heat-shock proteins (*other targets*).

When there is evidence of effectiveness for a drug or biologic directed at a molecular target in an adult cancer, and the target has been identified as substantially relevant for the growth or progression of a pediatric cancer, there may be a rationale for the agent's evaluation in the pediatric cancer population, regardless of similarity to the histologically-defined cancer found in the adult. Although not an absolute requirement, it is beneficial for sponsors of an agent such as this to have associated *in vitro* and *in vivo* data using pediatric non-clinical models to provide increased confidence for the role of the target in growth or progression of specific cancers. These data may help guide pediatric clinical development of the agent.

Here, we propose two frameworks, one that outlines factors that may be useful when determining whether a target is substantially relevant in pediatric cancer and may trigger the requirement for pediatric investigations. The second framework outlines factors to consider when assessing the available data that may help determine there is insufficient evidence of relevance, and that the target is hence "not relevant."

Factors to consider for defining a target as substantially relevant for the growth or progression of pediatric cancer

The FDA in collaboration with the NCI is tasked with determining whether a molecular target is considered substantially relevant to the growth or progression of pediatric cancer, or whether there is evidence that the target is not relevant to pediatric cancer. It is solely the prerogative of the FDA to determine whether adequate evidence is available to define a target as substantially relevant that triggers a requirement for pediatric investigations. Thus, defining "adequate evidence" is beyond the scope of this document. However, several factors may support a scientifically-based and data-driven decision-making approach. These factors are not meant to be either inclusive or prescriptive, as there may be additional factors for some specific targets and some of the listed factors may not be required for all targets within a class. Indeed, specific considerations related to the framework factors may have different applicability depending upon the target class. The framework (Table 1) is meant to guide discussion on the types of evidence available that will support the determination of whether a molecular target is substantially relevant to the growth or progression of pediatric cancers. The framework is not meant to be read as a checklist. It is important to note that the totality of evidence available may

be considered when guiding discussions to determine target relevance. The presence of a single factor or a particular combination of factors may not be sufficient to trigger relevance.

Table 1: Framework of factors and characteristics that may guide the determination of whether molecular targets are **substantially relevant** in the growth or progression of pediatric cancer

Factors		**Considerations**
Presence of target		The target has been identified in at least one case of a pediatric cancer
	Target class: Gene abnormality	The gene abnormality has been identified in at least one case of a pediatric cancer
	Target class: Cancer cell lineage	The target is intrinsically and differentially expressed in the cancer of interest compared to normal site-specific tissues
Function/Mechanism		The biological function of the target is relevant to the etiology and growth of the childhood cancer
	Target class: Gene abnormality	Modulation of the affected gene product or of a critical downstream pathway or correction/deletion of the affected gene defect adversely affects cancer cells
		The presence of the gene abnormality creates a synthetic lethal relationship with another cellular pathway
	Target class: Cancer cell lineage	The target is associated to cancer cell development, growth, and survival
Non-clinical evidence		Non-clinical evidence supports relevance of the target in one or more pediatric cancers
	In vitro activity	Target modulation shows *in vitro* selectivity for cancer cell lines containing/expressing the molecular target (pediatric or adult cell lines if target is known to be shared by multiple cancer types regardless of patient population) compared to the sensitivity of cell lines not containing/expressing the target
	In vivo activity*	Target modulation shows *in vivo* activity manifested as tumor stabilization or regression in models of pediatric cancers with the molecular target of interest (or adult cancer models containing/expressing the target)
	Lack of *in vitro* or *in vivo* activity	For targets for which target modulation does not show *in vivo* or *in vitro* activity, support for relevance may be found in evidence for supra-additive or synergistic activity when target modulation is used in biologically rational combinations
Adult clinical experience		Target modulation by investigational agents known to affect the target, shows clinical activity in specific cancers in adults
Predictive biomarkers		Biomarkers that predict responses to target modulation may be useful in the selection of appropriate pediatric study populations
Location		For immunotherapy targets, the target is expressed on the cell surface (excepting immunotherapies that target intracellular antigens that are displayed as peptides by MHC proteins on the cell surface)
Agent under development		There is an agent in development or proceeding to development that addresses the specific target

*The *in vivo* activity should be observed at drug exposures that are relevant to the clinical setting if there is clinical experience with the agent. Prolonged stable disease may be relevant, particularly for agents that induce their anticancer effect through mechanisms other than cancer cell apoptosis.

Because of the importance of non-clinical evaluation in determining relevance of molecular targets, every effort should be made to ensure sponsors expedite early non-clinical

investigation, which could be in collaboration with academic research teams with pediatric expertise in non-clinical testing. The creation of these collaborations and/or partnerships should be explored further as they will be crucial for early testing of non-clinical models, such as xenograft mouse models.

Biomarkers that are identified as predictive for the activity of adult cancer targeted agents should also be evaluated for distribution and potential utility across pediatric cancers. Sponsors are strongly encouraged to test samples from pediatric cancers to determine relevance, especially when an assay to identify a target is developed in conjunction with the investigational agent and is not available for use on patients by investigators other than the sponsors.

Factors to consider that will help identify targets that are not relevant to the growth or progression of pediatric cancer

There may be evidence available that demonstrates a molecular target is not relevant in pediatric cancers that would prevent it from being added to the substantially relevant molecular target list. The factors listed in Table 2 highlight considerations that may guide the determination of whether a molecular target is not relevant to the growth or progression of pediatric cancer. Again, it is solely the FDA's responsibility to determine what is the evidence necessary to determine whether a molecular target is considered not relevant in pediatric cancer, and thus this document does not attempt to define what "adequate evidence" refers to in this context.

Table 2: Framework of factors and characteristics to consider that may guide the determination of whether molecular targets are **not relevant** to the growth or progression of pediatric cancer[#]

Factors	Considerations
Biologically implausible	Molecular targets for which available evidence supports no role for the targets in pediatric cancers (e.g. endocrine/autocrine sex steroid hormonal pathways that are known to be drivers of specific adult cancer types but are very rarely to never observed in pediatric cancers)
Non-clinical evidence	Evidence of lack of activity of an agent in development against a specific target in non-clinical systems could be a component of the evidence base used to determine that a specific molecular target may not be relevant to the growth or progression of a pediatric cancer
Adult clinical evidence	Evidence of lack of clinical activity of an agent in development against a specific target could be a component of the evidence base used to determine that a specific molecular target may not be relevant to the growth or progression of a pediatric cancer

[#]There may be agents that are relevant to the growth or progression of disease but that would not be considered for development because of their association with developmental processes such that their inhibition would raise concerns about irreversibly deleterious developmental effects and subsequent growth-related toxicities (see *Additional Considerations* section below).

Targets with insufficient evidence

Molecular targets for which sufficient evidence to make a determination of "substantially relevant" or "not relevant" may not yet be available and will not be included in either list.

Decisions regarding relevance of these targets to the growth or progression of pediatric cancers could be made when there is an adequate evidence base to make such a determination. Sponsors and investigators are strongly encouraged to investigate the potential relevance of new and currently unlisted targets as expeditiously as possible, especially when there are early non-clinical or clinical signals of activity.

Mechanisms to update the molecular targets lists

To ensure the molecular targets lists are updated with the most relevant evidence available in light of the rapid pace at which scientific advances occur, three distinct opportunities are discussed.

The first opportunity includes an annual public workshop at which all stakeholders, including but not limited to members of the FDA, NCI, industry, academic and clinical investigators and patient advocates, will discuss potential changes to the molecular targets lists. The FDA will be responsible for convening and presiding over this annual meeting, which may occur following a national or international scientific meeting. This meeting will seek input from individual stakeholders on advances in relevant scientific evidence that may impact the inclusion of one or more molecular targets on the current published lists, including potential relevance of unlisted targets. Final decisions related to the lists will require input from the Pediatric Subcommittee of the ODAC.

The second opportunity consists of a transparent nomination mechanism to occur during or prior to meetings of the Pediatric Subcommittee of the ODAC. This mechanism could include, but is not limited to, clinical investigators as well as researchers in academia and industry, who will have the opportunity to suggest targets to be added to or removed from the list based on substantial scientific evidence that demonstrate emerging relevant targets, or that demonstrate no relevance in pediatric disease, respectively.

The third opportunity would create a transparent process for clinical investigators or sponsors to request a meeting at any time with the FDA to discuss new scientific data related to a new or existing molecular target, which may warrant a change in that target's status as substantially relevant or non-relevant and could result in changes to the lists.

Data gathered from any and all sources could then be assessed by the FDA with input from the Pediatric Subcommittee of the ODAC in order to determine whether there is substantial new evidence to change the status of the target of interest. It is important to note that even if agents under development addressing molecular targets for adult indications are added to the "substantially relevant in pediatric cancer" list late in the development paradigm for the adult indication, those targets will not be exempted from the requirement for pediatric investigation.

Continuous review of nominations for potential targets of relevance obtained through any of the opportunities listed may be accomplished by a transparent mechanism where members of the Pediatric Subcommittee of the ODAC review nominations on an ad hoc basis to inform the

FDA as to a target's potential relevance. Changes made to the list after nomination review could be made immediately and not wait for the next meeting of the Pediatric Subcommittee of the ODAC. As mandated by law the resulting lists will be published on the internet website of the FDA [21 USC 355c (m)(1)].

Additional considerations for the potential application of the molecular target list

Additional considerations may potentially arise when seeking to apply the list of molecular targets. In this section we will highlight a few factors that could influence the application of the list, such as balancing clinical benefit and risk, the availability of pediatric formulations, and the size of the patient population when conducting clinical trials. These factors will play different roles in each scenario but discussing and brainstorming potential approaches with several key stakeholders in the pediatric cancer community is imperative to help accelerate the availability of life-saving therapies for children and adolescents with cancer.

1. Clinical benefit: risk analysis

   As with any clinical study, investigations in pediatric patients must be scientific and ethically justified, taking into consideration the prospect of direct benefit to individual children and adolescents with cancer. Regulatory requirements for pediatric clinical research are provided in 45 CFR part 46, with subpart D specifically addressing the categories of allowable research involving children as subjects.
   As per FDA's guideline, "E11(R1) Addendum: Clinical Investigation of Medicinal Products in the Pediatric Population", clinical studies will assess the balance of risk and anticipated clinical benefit.

   > *"Experimental interventions or procedures that present greater than low risk must offer a sufficient prospect of clinical benefit to justify exposure of a pediatric population to such risk. Likewise, the balance of risk and anticipated clinical benefit must be at least comparable to the available alternative treatments. There should be a reasonable expectation that a clinical benefit resulting from the clinical study can be made available to this population in the future."*

   Therefore, in addition to the factors in the framework outlined in this document, the requirement for sponsors to study a molecularly targeted therapy in pediatric cancer patients must be supported by the prospect of direct clinical benefit. A reasonable balance needs to be identified in a case-by-case scenario and all data need to be considered in identifying the right balance.

   For example, when a target is considered to be substantially relevant to the growth or progression of pediatric cancer, yet the toxicity profile of a new agent modulating this target is known to cause irreversible adverse effects of sufficient magnitude, including those associated with a vital developmental pathway, conducting a pediatric investigation using this agent may not be justified and further development may be precluded.

2. Pediatric formulation requirements
   Drugs and biologics need to be formulated to best suit a pediatric patient's age, size, physiologic condition, and treatment requirements to be studied in children and adolescents. To facilitate the availability of these *pediatric formulation*s needed for the pediatric investigations outlined in Title V of FDARA, sponsors are encouraged to begin establishing a pediatric formulation early in the adult drug development process. This will help sponsors meet the requirements outlined in FDARA and provide an initial pediatric study plan (iPSP) at the conclusion of the adult Phase 2 study, which includes plans for the development of a potentially marketable pediatric appropriate formulation.

3. Patient population
   A molecularly targeted pediatric cancer investigation, as required by Title V of FDARA, is designed to yield clinically meaningful pediatric study data, gathered using appropriate formulations for each age group for which the study is required, regarding dosing, safety, and preliminary efficacy to inform potential pediatric labeling. Thus, a sufficient patient cohort needs to be accrued to identify proper dosing, safety concerns, and signals of preliminary efficacy. However, due to the rarity of some pediatric cancers, accruing an adequate number of pediatric patients with cancer for early clinical studies conducted at single centers may not be feasible. Collaborations among different clinical centers and between strong pediatric trial networks are encouraged in order to conduct pediatric investigations that will yield robust findings. Moreover, international collaborations for clinical trials involving rare forms of pediatric cancer may be considered to improve accrual rates. Collaborative drug development efforts are logistically, operationally, and legally complex, and as such, require increased and more transparent communication among regulatory organizations, industry, and other stakeholders.

Questions:
These questions may guide the discussion during the meeting:

1. Should the term "molecular target" only be used for molecules that already have an agent that modulates its activity and that is either fully developed or in the process of development?
2. What is considered an optimal level of evidence required for the factors presented in the framework that may guide the determination of substantially relevant to the growth or progression of pediatric cancer?
3. What level of evidence is necessary and would be considered substantial to predict direct benefit for institutional review boards to approve protocols for pediatric patients?
4. When a potential target is identified in one case of a pediatric cancer, how could a drug development strategy be defined and what are the responsibilities of each stakeholder?
5. What types of evidence inform preliminary efficacy in molecularly targeted pediatric cancer investigations and in what phase of the clinical study would these data be collected?
6. Does the validation of the drug-target relationship have to be established in pediatric non-clinical models to be considered substantially relevant to the growth or progression of a pediatric cancer?
7. The European Innovative Therapies for Children with Cancer (ITPCC)-P4 (Paediatric Preclinical Proof of Concept Platform) program and the NCI Pediatric Preclinical Testing Consortium are public-private partnerships to advance non-clinical science and enable rational drug development in pediatrics. How do these partnerships and others work, are they effective, and is there a need for additional efforts to expedite non-clinical research?
8. Would it be helpful to have a private effort that aims to create and encourage an open-access crowd-sourcing approach for the updating and maintenance of the list of relevant molecular targets?
    i. How could this crowd-sourcing effort inform the FDA's mandate to update the molecular targets list?
9. Could relevant data generated by international agencies and institutions be used in determining whether a molecular target is substantially relevant in pediatric cancer?
10. What considerations should be explored to facilitate international collaboration and coordination that addresses work in small patient populations?
11. Should there be a mechanism in place whereby waivers granted by the FDA are published to avert unnecessary trials for agents sharing a similar MOA?

Glossary of terms (in order of appearance):

Assessment refers to an evaluation of data gathered using appropriate formulations for each age group for which the assessment is required and that are adequate to assess the safety and effectiveness of the drug or biological product for the claimed indication in all relevant pediatric subpopulations, and support dosing and administration for each pediatric subpopulation for which the drug or biological product is safe and effective.

Pediatric age groups, according to the FDA, refers to neonates (newborns up to one month of age), infants (one month to two years of age), children (two to twelve years of age), and adolescents (twelve to sixteen years of age) (*see* FDA Draft Guidance "Pediatric Information Incorporated Into Human Prescription Drug and Biological Products Labeling," Feb. 2013). NIH policy defines "child" as individuals under 18 years old. For informed consent purposes, in clinical studies "children" refers to those under the legal age of consent.

Pediatric cancer refers to cancers arising in the pediatric population, which includes neonates, infants, children and adolescents.

Molecularly targeted pediatric cancer investigation refers to studies designed to yield clinically meaningful pediatric study data, gathered using appropriate formulations for each age group for which the study is required, regarding dosing, safety, and preliminary efficacy to inform potential pediatric labeling.

Pediatric formulations refer to drugs and biologics that are formulated to best suit a pediatric patient's age, size, physiologic condition, and treatment requirements, taking into consideration the differences between adult and pediatric patients with regard to pharmacotherapy, including capabilities for drug administration, medicine-related toxicity, and taste preferences.

Molecular target refers to a molecule in human cells that is intrinsically associated with a particular disease process, such as etiology, progression, and/or drug resistance, and for which there is evidence that the resulting disease process might be addressed by a targeted, small molecule, biologic product, or other treatment intervention to produce a desired therapeutic effect.

Gene abnormality-based targets refer to targets that are the result of specific gene abnormalities or that are present in a critical biologically-related pathway of a gene abnormality or that are in a synthetic lethal relationship to a gene abnormality.

Cancer cell lineage-based targets refer to targets intrinsic to the cancer cell lineage (e.g., CD19 for B-ALL, estrogen receptor for breast cancer, and GD2 for neuroblastoma) or developmental stage.

Non-cancer cell targets refer to targets identified in non-cancer cells, such as normal immune cells or supporting cells, contributing to the tumor micro-environment

Other targets refer to targets present in cancer cells but that do not have specific genetic alterations (e.g., tubulin, HSP90, proteasome, etc.).

References:

1. St. Jude Children's Research Hospital - Washington University. Pediatric Cancer Genome Project. Accessed February 2018: http://explore.pediatriccancergenomeproject.org/;jsessionid=D171CBEC77362D9DE9802E9BD65ACD42
2. National Cancer Institute, Office of Cancer Genomics. TARGET: Therapeutically Applicable Research To Generate Effective Treatments. Accessed February 2018: https://ocg.cancer.gov/programs/target
3. Allen CE, Laetsch TW, Mody R, et al. (2017) Target and Agent Prioritization for the Children's Oncology Group—National Cancer Institute Pediatric MATCH Trial, JNCI: Journal of the National Cancer Institute, 109:5
4. Mack SC, Pajtler KW, Chavez L, et al. (2018) Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling, Nature, 553:101-5

FRIENDS of CANCER RESEARCH

A Friends of Cancer Research White Paper

REAL-TIME ONCOLOGY REVIEW AND THE ASSESSMENT AID:

INCREASING REVIEW EFFICIENCY THROUGH STANDARDIZATION AND EARLIER DATA ACCESS

CONTRIBUTORS

Gideon Blumenthal
U.S. Food and Drug Administration

Jonathan Cheng
Merck & Co., Inc.

Katherine Couvillon
Patient Advocate

Jennifer Gao
U.S. Food and Drug Administration

Karen Jones
Genentech,
A Member of the Roche Group

Robert Kester
Merck & Co., Inc.

Scott Korn
Merck & Co., Inc.

Qi Liu
U.S. Food and Drug Administration

Jiten Rana
Novartis Pharmaceuticals

Giuseppe Randazzo
U.S. Food and Drug Administration

Samantha Roberts
Genentech
A Member of the Roche Group

Mahesh Ramanadham
U.S. Food and Drug Administration

Donna Roscoe
U.S. Food and Drug Administration

INTRODUCTION*

The regulatory review process for pharmaceutical drugs is a resource intensive undertaking for both the drug sponsor and the United States Food and Drug Administration (FDA) that assesses the drug's benefit and risk. Improvements in the efficiency of the process can have significant impact on the resources and time required to complete a drug review, consequently, bringing new therapies or new therapy indications to patients more quickly. There are currently several tools that the FDA can employ to expedite certain applications, including fast track designation, breakthrough therapy designation (BTD), accelerated approval, and priority review designation, Table 1. The FDA Oncology Center of Excellence (OCE) has established two new pilot projects with voluntary participation to test novel approaches to regulatory review for oncology drugs, the Real-Time Oncology Review (RTOR) and the Assessment Aid (AAid).

The RTOR Pilot Program aims to improve the efficiency of the review process for supplemental 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 supplements for drugs or biologics likely to demonstrate substantial improvements over available therapies (e.g. BTD, accelerated approval, and priority review designation-eligible indications) and based on clinical trials with straightforward study designs and easily interpretable endpoints

* The considerations for possible future expansion of the Real-Time Oncology Review and Assessment Aid pilots presented in this whitepaper should not be construed as final FDA policy.

(for example, overall survival in a randomized trial), as defined by the Review Division. The pilot will include applications to be reviewed by Division of Oncology Products 1, Division of Oncology Products 2, and Division of Hematology Products. The first two RTOR approvals were supplemental approvals for KISQALI® (ribociclib)[1] for two new indications, based upon two randomized, placebo-controlled, phase III trials with progression free survival endpoints, and KEYTRUDA® (pembrolizumab)[2] based upon a randomized phase III trial compared to chemotherapy with progression free survival and overall survival endpoints. Both applications showed unequivocal efficacy results. Both KISQALI® (Kisqali) and KEYTRUDA® (Keytruda) had previously received BTD and priority review designation, whereas Keytruda had also previously been granted accelerated approval for other indications.

The second OCE pilot is the AAid for new drug applications (NDA) and biologic license applications (BLA) submissions or supplements (sNDA/sBLA). The AAid can improve review quality and efficiency by providing a shared document into which the applicant can insert their positions and the FDA review team can subsequently layer in their assessment, which reflects their critical evaluation. Participation in the AAid pilot can occur in conjunction with the RTOR pilot or independently.

Both pilot programs may ultimately be converted to permanent programs (there is no definitive timeline for the pilot); however, the full value of the pilots will be realized from expansion beyond their initial, limited scope. The FDA will need to accumulate more experience with the pilots to fully assess their success, but should consider priorities of the various drug development stakeholders, Table 2, when determining metrics for success to inform expansion. It will be important to consider how the review phase is defined in the RTOR context and implications to statutory obligations. Metrics for success should not be limited; however, to the review phase but should reflect benefits of RTOR as it may extend to other phases of the drug development pathway, including clinical development and the post-marketing phase. Although the pilots are still in their early stages and have not defined specific timelines, the ultimate benefit of this novel approach to regulatory review will likely be demonstrated through earlier patient access to important therapies if it is expanded to NDAs/BLAs.

Table 1: Regulatory Review Mechanisms

	Accelerated Approval	Fast-track Designation	Priority Review	Breakthrough Therapy Designation	Summary Level Review
Eligibility	1. Treat serious or life-threatening diseases 2. Provide meaningful therapeutic benefit over existing therapies 3. Surrogate endpoint reasonably likely to predict clinical benefit	1. Intent to treat broad range of serious diseases 2. Potential to fill an unmet medical need	1. Offer major advances in treatment over existing therapies	1. Treat serious or life-threatening diseases 2. Early clinical evidence of substantial improvement over existing therapies	Supplemental applications that: 1. The FDA determines the existing data is acceptable to demonstrate safety, and 2. The data used to develop the qualified data summary is submitted to the FDA 3. Not eligible for use with RTOR
Designation	No formal process	Can be requested by sponsor at any time; FDA has 60 days to respond	Requested by sponsor at time of NDA/BLA submission; FDA has 45 days to respond	Can be requested by sponsor at any time after IND submission; FDA has 60 days to respond	Supplemental applications for a qualified indication for a drug that the FDA determines to be appropriate for summary level review
Clinical Development	Conditional approval granted using surrogate endpoints from phase II trials or interim phase III data; controlled trials with hard clinical endpoints required to confirm clinical benefit	Earlier and more frequent communication	Not applicable	Abbreviated or condensed development; earlier and more frequent communication; delegation of senior reviewers and cross-disciplinary review team	Not applicable
Review Process	NDA/BLA data submitted in one package; standard 10-month review	Option for rolling NDA/BLA submission; official review clock begins when last module is submitted	NDA/BLA data submitted in one package; review time shortened to 6 months	NDA/BLA data submitted as they are accumulated; review time shortened	The FDA may rely upon qualified data summaries submitted as part of a sNDA/sBLA to support the approval of a supplemental application, with respect to a qualified indication

Table 2: Potential Impact of Real-time Oncology Review on Drug Development Programs

Stakeholder	Review Phase	Clinical and Post-Approval Programs
Regulatory Authority	• Pinpoint areas for focused review • Improved review quality	• Earlier access to trial data and supportive documents • Identify opportunities and concerns sooner
Sponsor/ Applicant	• Interactive/iterative process • Earlier feedback from FDA before dossier submission on data and review focus	• Increased predictability • Ability to address concerns sooner • Opportunity to further develop clinical program, data submission, etc. with collaborative feedback from the FDA
Patients	• Increased confidence in safety and efficacy data	• Earlier access to therapies

CASE STUDY: **NOVARTIS**

The first approval made through the RTOR pilot was ribociclib (trade name: Kisqali). On July 18, 2018, the FDA expanded the indication for ribociclib combination with an aromatase inhibitor for pre/perimenopausal women with hormone receptor (HR) positive, human epidermal growth factor receptor 2 (HER2) negative advanced or metastatic breast cancer, as initial endocrine-based therapy. FDA also approved ribociclib in combination with fulvestrant for the treatment of postmenopausal women with HR-positive, HER2-negative advanced or metastatic breast cancer as initial endocrine therapy or following disease progression on endocrine therapy. Ribociclib was previously approved for postmenoposal women with HR-positive, HER2-negative advanced or metastatic breast cancer in combination with an aromatase inhibitor as initial endocrine therapy and received BTD. The ribociclib sNDA was submitted to expand the indication based upon results of two phase III studies, one to support each indication change. Under the RTOR pilot, many components of the submission dossier were submitted as pre-submission materials, on a periodic basis (Table 3). The early submission from Novartis not only included efficacy and safety data, but also a clinical pharmacology package including pharmacokinetic and drug-drug interaction data. Once these components were received, the FDA review team analyzed the data for quality and integrity and verified the sponsor's results and conclusions. In addition, the FDA also conducted their own analyses. Novartis and FDA scheduled regular, bi-weekly teleconference meetings, eliminating the need for typical applicant orientation and mid-cycle meetings. The FDA approved the sNDA in less than one month following final dossier submission. The Novartis sNDA was also the first to use the AAid, discussed later in this whitepaper.

Table 3: Novartis RTOR Timeline

Event Date	Action	Notes
January 2018	Pre-NDA meeting held with FDA	
April 6, 2018	Novartis/FDA RTOR discussion	
April 24, 2018	Pre-submission packages start to be sent to FDA	• Safety and Efficacy datasets • Draft labeling • Module 2 summary documents and safety reports • Module 4 components • Clinical pharmacology package • Clinical study reports • 90-day safety update datasets
April-June 2018	FDA issues multiple IRs	
June 28, 2018	Full dossier submission	• Financial disclosures and BIMO information • Annotated USPI
July 18, 2018	sNDA for Kisqali approved	

CASE STUDY: **MERCK**

Pembrolizumab (Trade name: Keytruda) has been granted 13 BTDs including two for pembrolizumab monotherapy for non-small cell lung cancer (NSCLC). Merck was granted accelerated approval under a priority review timeline in May 2017 for pembrolizumab for first-line treatment of patients with metastatic NSCLC in combination with pemetrexed and carboplatin. Accelerated approval was based upon the KEYNOTE-021 trial cohort G and KEYNOTE-189 was identified as the confirmatory trial. Full approval was granted for pembrolizumab, based on KEYNOTE-189, for first-line treatment of metastatic NSCLC with no epidermal growth factor receptor (EGFR) or anaplastic lymphoma kinase (ALK) genomic tumor aberrations in combination with pemetrexed and platinum chemotherapy through the RTOR pilot. The approval was granted approximately 1 month prior to the Prescription Drug User Fee Act (PDUFA) assigned action date for a priority review designation. Merck and the FDA determined the components of a pre-submission package as part of a meeting to discuss the RTOR pilot (Table 4). The pilot was a collaborative process that included more frequent contact between the FDA project manager and Merck regulatory contact.

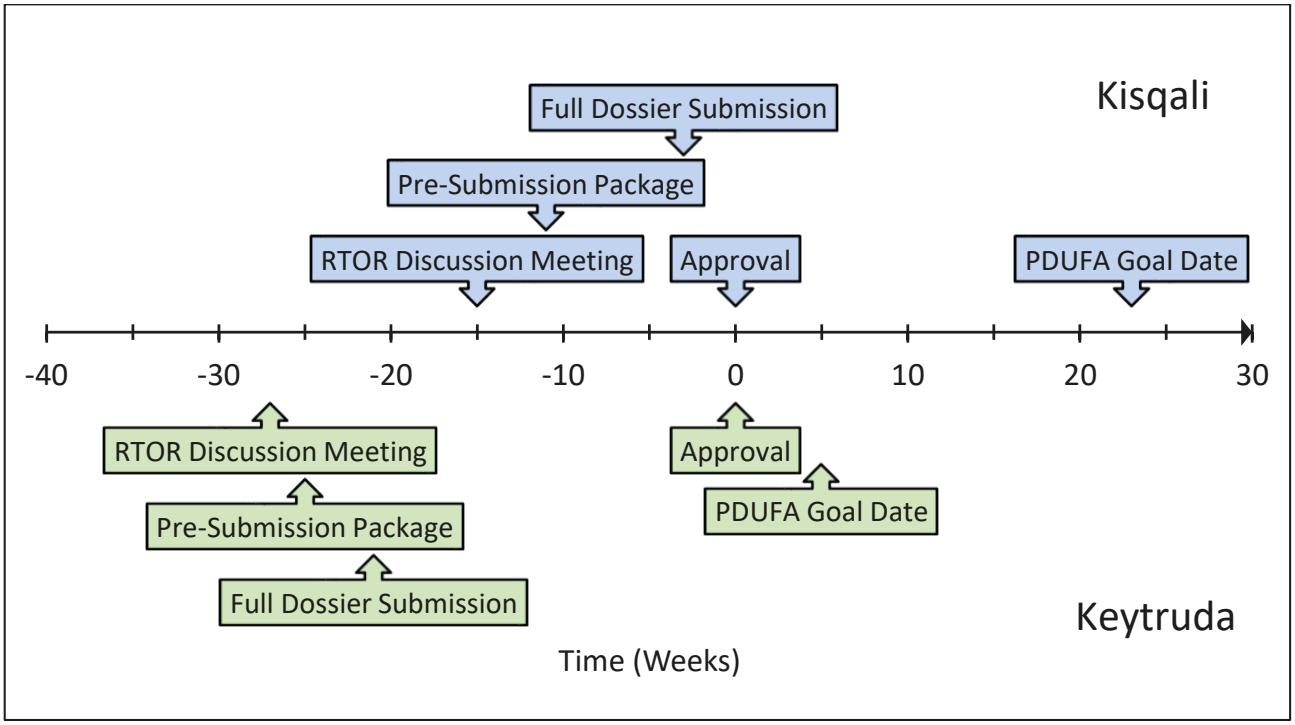
Table 4: Merck RTOR Timeline

Event Date	Action	Notes
January 2018	Informed FDA of topline results from KEYNOTE-189	Indicated intent of sBLA submission based on KEYNOTE-189
February 2018	Merck/FDA RTOR discussion	Determined data components and contents of the pre-submission package
February 27, 2018	Pre-submission package sent to FDA	• Key efficacy and safety tables and figures • SDTM dataset package and supporting documentation • Draft USPI • ADaM datasets and SAS programs • Protocol including all amendments and the SAP • DMC meeting minutes • Case report forms
February - March 2018	FDA issued IR regarding USPI and request for PMC	
March 23, 2018	Full dossier submission	• Included annotated USPI • OSI and Financial Disclosure information • Module 2 documents and CSRs
August 20, 2018	sBLA approved	

LESSONS LEARNED

The early examples that have informed this RTOR pilot have allowed for data and document submissions prior to final dossier submission, providing the FDA with additional time to begin evaluating results as they were submitted (Figure 1). Access to the SDTM and ADaM datasets, and draft USPI were important components of the pre-submission package. The agency was then able to submit IRs to the sponsor to fill in the gaps as the data was reviewed, allowing for ongoing communication between the agency and sponsor and quick response/submission of information and additional analyses by the sponsor as requested by the FDA. The IRs that were issued to both sponsors primarily related to the USPI and study datasets. Both Merck and Novartis indicated that submitting general comments/information related to data derivation, such as grouped terms provided in the draft USPI, earlier in the pre-submission communication along with additional documentation accompanying datasets may have helped to facilitate FDA review. Additionally, earlier submission of the data definition files would be desirable, where possible.

Figure 1: Timeline for RTOR submissions*



* Similar timelines are not guaranteed for all RTOR pilot submissions

PILOT EXPANSION

Expansion of the RTOR pilot should be approached in a stepwise fashion as the FDA and industry gain more experience with a "front-loaded" application process. Initially, the FDA could consider expanding eligibility from supplemental applications with straightforward clinical trial designs and easily interpretable endpoints to more complex supplemental applications such as those that include more complex clinical trial designs, more challenging endpoints, or a companion diagnostic claim. With more experience troubleshooting complex supplemental applications, the FDA could consider expanding eligibility to simple Breakthrough-designated New Molecular Entity (NME) NDAs/BLAs and, eventually, increasingly complex NDAs/BLAs (Table 5). If the RTOR pilot is ultimately expanded to NME applications, impact to other aspects of the drug development pipeline, including clinical trial design, will need to be addressed.[3-4] Eventually, as we gain more scientific knowledge and achieve more data standardization, other evidence such as patient-experience data (e.g., collected through patient-reported outcomes) and other real-world data, could be integrated to the RTOR to foster a comprehensive benefit-risk assessment of a product.

A key to efficient expansion of the RTOR pilot project will be to capitalize on the successes and lessons learned from pilot submissions to identify potential barriers to expansion and recommend policy to address those barriers.

Table 5. Mock Plan for RTOR Expansion

RTOR	Pilot 1 Scope	Pilot 2 Scope	Pilot 3, etc. Scope	Final Pilot Scope
Pilot Timeframe	2018	2019		
Criteria for Inclusion	• sNDA/sBLA for drugs likely to demonstrate substantial improvements over available therapy (e.g., BTD, priority review designation, or accelerated approval-eligible designations) with: ○ Straight forward study designs, as determined by the review division and the OCE, and ○ Endpoints that can be easily interpreted (for example, overall survival in a randomized trial)	• Drugs likely to demonstrate substantial improvements over available therapy (e.g., BTD, priority review designation, or accelerated approval-eligible designations) with: ○ Complex study designs or simple diagnostic scenarios based upon a prospective trial that demonstrates efficacy of a drug in a biomarker defined population using an approved CDx test measuring the same marker and tissue type (e.g., Pfizer's dacomitinib approved with Qiagen therascreen EGFR test for NSCLC), or ○ Simple diagnostic scenarios based upon an approved therapy in a new indication (line extension) via a prospective trial that demonstrates efficacy in a biomarker defined population using a new diagnostic test	• Establish criteria for increasing complexity	• NMEs with BTD • Complex study designs • Single arm study designs • RWE and PROs
Exclusions	• Diagnostics • RWE • CMC supplements • NME	• CMC supplements • RWE • NMEs	• NMEs	
Considerations		• Align CDRH processes	• Align manufacturing processes	• Align manufacturing processes and inspections • Align clinical site inspection processes • Align OPDP activities to ensure earlier submission and review of first 120-day marketing materials if single arm studies will lead to accelerated approval

Aligning Manufacturing Processes with Real-Time Oncology Review

In an expedited approval setting such as BTD, it is important to align and synchronize product development with commercialization in order to successfully realize needed acceleration. Aligning product development and commercialization can be challenging in the setting of an expedited approval pathway such as RTOR when data is requested earlier than during a traditional development and review program. Lessons from the implementation of BTD could inform policies for expedient alignment and implementation (Box 1).

Box 1. Recommendations for Manufacturing Processes for Expedited Pathway [Dye et al. AAPS PharmSciTech (2016) 17(3)]

1. Encourage more flexible approaches to ensuring information exchange and understanding to facilitate expediting development and review
2. Agree upon schedule of important review milestones and turnaround timeframes for information requests
3. Discuss approach to submit agreed upon data packages during the review:
   a. Submission of the dissolution method development report and dissolution specification setting strategy for early review by FDA Biopharmaceutics reviewers
   b. Additional real-time stability data on commercial product
   c. Additional batch data to support validation
4. Initiate discussions to enable more rapid access to CMC and facility data to facilitate pre-approval inspection scheduling and conduct

Additionally, it is important to note that, dependent upon the potential pilot candidate (i.e., NDA or BLA), the request for certain information may vary. These potential differences would be discussed in meetings with the FDA prior to submission of the pilot candidate. Table 6 outlines manufacturing components and readiness to consider during the different phases of a drug development program.

There are two key issues for early communication of manufacturing data: 1. early agreement upon an appropriate timeline for submission of manufacturing data, which will necessitate prioritization of product stability and batch data and facility inspections; and 2. identification of components that can be addressed in post-approval commitments.

Table 6. Manufacturing Components and Readiness

Phase	Component
Pre-submission	• Analytical method development and validation* • Commercial to-be-marketed formulation • Container/closure system for commercially marketed product • Product specification • Stability and degradation studies • Representative batch data • Manufacturing process development, description of intended initial processes and controls • Facility information for assessment
Submission	• Submit comparability strategy/protocol for post-approval site changes • At least one executed batch record • Demonstration of successful manufacturing using processes and controls representative of intended initial commercial operations • Updated primary stability data • Rolling submission of process validation information
Post-Approval	• Process and formulation optimization • Concurrent release of process performance qualification lots[5]

* Control strategy, acceptance criteria, and methods may still be evolving at this stage.

To accommodate the accelerated submission timeline described, sponsors will need to prospectively design CMC development such that process and product improvement and optimization require minimal comparability assessment while keeping the following aspects in mind:

- **Optimize candidate selection**
  - Physical-chemical properties and pharmacokinetic profile of small molecule drugs
  - Screening and engineering out hot spots for degradation or undesired modifications for biologic drugs

- **Leverage platform knowledge** - Ensure fit of candidate molecules into manufacturer's platform for drug substance and drug product and related processes to improve speed and robustness

- Consider additional in-process and specification tests in the control strategy to balance uncertainty driven by accelerated product/process development. It is envisioned that additional controls could be removed post approval when adequate product/process knowledge has been accumulated and its evaluation indicates a stable and capable process and control strategy

- **Leverage use of Physiologically Based Pharmacokinetic (PBPK) models to enable rapid development of drug products with optimal performance** – The models can be applied to support formulation optimization and other changes required during fast moving development programs (e.g. PSD, manufacturing process, scale-up, etc.)

- **Initiate key activities early:**
  - Activities needed to address non-platform behavior and/or unusual product and process characteristics
  - Assessment of CQAs
  - Identification of launch sites for drug substance and drug product or consider launch from R&D facilities while ensuring product quality and patient safety with reliable supply and pre-approval inspection readiness

- **Focus on reliable supply of quality product at launch**

Aligning Inspections Processes with RTOR for Pilot Expansion to NMEs

An additional area of focus in the aim of removing barriers in getting products to patients are the BIMO Good Clinical Practice (GCP) and manufacturing site pre-approval inspections that currently occur on the critical path to approval (manufacturing pre-approval inspection readiness was addressed in the previous section). GCP pre-approval inspections involve retrospective evaluation by the FDA of the sponsor study monitoring practices and procedures post-submission (typically 3-5 months to organize and execute) to determine compliance with applicable regulations.

Proactive information sharing with the FDA (earlier submission of site level datasets, inclusion of sponsor GCP quality assurance briefing as part of submission, and sharing of quality assurance data output in real time during pivotal study conduct) to enable faster assessment of GCP compliance, could save on resources for both the sponsor and the FDA, while further expediting timelines.

Aligning CDRH Processes with RTOR for Pilot Expansion to NMEs

A great deal of work has already been undertaken to align CDRH processes for BTD. CDRH review mechanisms such as modular Premarket Approvals (PMAs), which enable review and acceptance of submission components in advance of the clinical validation data, are successful for aligning review of companion diagnostics with drug approvals and will continue to be valuable for RTORs. However, development and market-ready distribution of a diagnostic at the time of approval may not be feasible. Given the increasing number of targeted therapies in development in oncology, it bears considering how drug/diagnostic co-development, review, and approval can be coordinated within the RTOR pilots and eventually be established as practice.

- Use of previously approved tests will enable swift review of new therapeutic indications. To this end, pharmaceutical companies can reach out to key diagnostic companies and clinical laboratories to bring tests in as follow-on companion diagnostics. This will increase the number of readily available diagnostic partners for development of new CDx indications.[7]

- Post-market commitments may extend the opportunity to bring a validated test to market. Points to consider for planning post-market device validation would include:

    - Adequately banking specimens from patients eligible for the trial to enable swift validation of the final *in vitro* diagnostic (IVD).

    - In the case of very rare biomarkers (e.g. ROS1), increasing availability of well-annotated specimen biobanks will enable improved access to tissues with rare biomarkers needed for analytical validation studies to support the diagnostic. Where specimen banking is not feasible, use of clinical specimens from an equivalent patient population may be feasible.

    - Move toward study designs that stratify patients based on the biomarker using an analytically validated test. Development of study designs that can be implemented would focus on

complementary device claims rather than companion diagnostic claims. Complementary device claims can then be supported in the post-market setting with specimens from subjects in the trial. This would enable line extension based on retrospective analysis and/or RWE that shows increased efficacy for patients with a certain biomarker or genomic profile (e.g., micro satellite instability (MSI) or tumor mutational burden (TMB)). Study designs should remain consistent with CDER review; issues include the target population, sample size, endpoints, and statistical analyses for missing samples.

- Ideally, understanding the value of the biomarker to patient management with the therapeutic by pre-planning clinical trials that stratify patients on the biomarker would also enable FDA to evaluate the magnitude of relative treatment benefit through an interaction effect between the biomarker and drug efficacy. Such a study design would allow identification of a clinically meaningful threshold, which could allow faster contemporaneous co-approval of companion tests with the therapeutic that support the efficacy. Such an approach may be able to pave the way for obtaining additional robust analytical validation in the post-market setting because the clinical utility and cut-off of the test for the biomarker is well supported.

Data Standardization

Facilitation of a more efficient submission and review pathway for expansion of the RTOR pilot should be accompanied by better data standardization and a more iterative submission process for improved communications between the sponsor and agency. For example, an iterative process for updating drafts of the USPI may be necessary in the pre-submission setting. Using CDISC data format for key datasets, such as adverse events, demographics, treatment response, exposure, etc., while allowing legacy data format for other datasets may facilitate a smoother transition during data standardization. In the future, data standardization that would be beneficial to realizing the full potential of the RTOR might include universal protocols, electronic case report forms, and data formats in an effort to streamline processes for better clinical trial design, data submission, and review. Development and adoption of dynamic interactive analysis tools will be essential to facilitating data standardization for efficient communications. Such tools could aid the agency's review of the data and analyses more efficiently with an option to extract the programming codes for understanding of the data derivation and statistical methodology applied in the analyses. Encouraging companies to include the interactive analysis tools, such as R-Shiny in the RTOR pilot will ultimately lead to the development of industry-wide interactive analysis application.

Adequate preparation will be necessary on the part of both the FDA and sponsor to efficiently expand RTOR. For the FDA to review greater volumes of pre-submission data, earlier engagement with the Office of Pharmaceutical Quality will be necessary and the agency will need to facilitate earlier international inspections of clinical trial sites and manufacturing facilities. Also, the FDA will need to address how to expand beyond supplemental applications where agency reviewers are already familiar with efficacy data, safety signal identification, clinical trial design, and data structure and format for approved drugs.

Similarly, drug/biologic sponsors will need to identify process improvements necessary to enable earlier dataset preparation for pre-submission data sharing and the type of data, particularly manufacturing data, that would be feasible to share during pre-submission. Finally, sponsors will need to consider the implications of pre-submission data-sharing on clinical trial design and whether adjustments will need to be made in trial design to enable earlier formatting of clinical data.

ASSESSMENT AID

In addition to the RTOR pilot program, the Novartis sNDA submission for Kisqali was also the first approval using AAid pilot**(Table 7).

Table 7. Novartis Assessment Aid Timeline

Date	Action	RTOR Action
Early April, 2018		Pre-submission package sent to FDA
April 24, 2018	Novartis received Assessment Aid template	
June 5, 2018	Novartis completed Assessment Aid	
June 28, 2018	FDA returned agency feedback on Assessment Aid to Novartis*	Full dossier submission
July 6, 2018	Novartis submitted Assessment Aid to FDA with final updates*	
July 18, 2018		sNDA for Kisqali approved

* This is a special case because both the agency and applicant were exploring the best practice for use of the AAid. Once submitted, the applicant would generally not have the opportunity to revise their portion of the AAid.

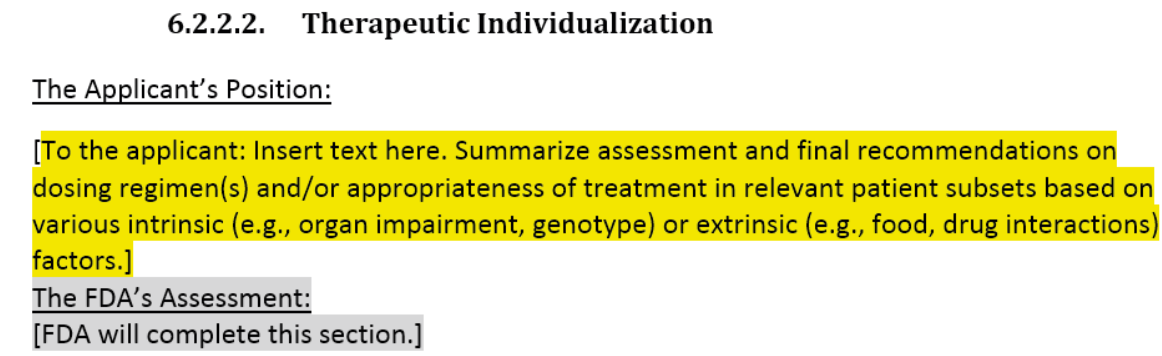
The AAid is a form, developed based on the FDA Multidisciplinary Review template, which covers the critical regulatory components that need to be evaluated to make approval decisions and labeling recommendations. Most sections of the template are divided into two parts, clearly delineated to emphasize the ownership of each position:

1. The Applicant's Position
2. The FDA's Assessment

** The Merck sBLA was part of the RTOR pilot but not the AAid. The AAid was not developed at the time that Merck entered the pilot for RTOR and the regulatory review was well under way when the AAid pilot became available.

The separation of the applicant's positions and FDA's assessment is intended to clarify (1) the ownership of each statement and (2) agreement/disagreement between the applicant and the FDA's position (Figure 2). The AAid template is sent to the applicant during the Investigational New Drug (IND) phase (for example, around the pre-NDA/BLA meeting). The applicant then adds their position to the template in preparation for the NDA/BLA or sNDA/sBLA submission. When the AAid is used in conjunction with the RTOR pilot, the applicant can submit the document before the formal sNDA/sBLA submission. Otherwise, the document is submitted at the time of the NDA/BLA or sNDA/sBLA submission or shortly thereafter. The FDA review team, after conducting their scientific analysis, then inputs their assessment into the same document, expounding upon areas of disagreement and additional findings in the FDA's analyses. The AAid can help focus the FDA review on critical assessment, rather than repeating the applicant's data analyses for improved review efficiency and consistency.

Figure 2: Section 6.2.2.2: Therapuetic Individualization



6.2.2.2. Therapeutic Individualization

The Applicant's Position:

[To the applicant: Insert text here. Summarize assessment and final recommendations on dosing regimen(s) and/or appropriateness of treatment in relevant patient subsets based on various intrinsic (e.g., organ impairment, genotype) or extrinsic (e.g., food, drug interactions) factors.]

The FDA's Assessment:

[FDA will complete this section.]

Figure 2: Section 6.2.2.2 in the Assessment Aid Template. Sections of the AAid are divided into two parts (The Applicant's Position and The FDA's Assessment) and are clearly delineated to emphasize the ownership of each position. Instructions to the applicant are provided in some sections to clarify the FDA's expectations of what should be included.

While successful at increasing review efficiency in this initial case study, the maximum benefit from inclusion in future submissions will be from expanded uses. For example, a potential application of the AAid could be to consolidate documents submitted by FDA and the sponsor to the Oncology Drug Advisory Committee (ODAC) to provide more streamlined briefing document materials for ODAC members and the public. Further, the AAid could be expanded to incorporate additional analyses of patient reported outcomes to inform benefit-risk assessments of NDAs/BLAs. Future considerations for the AAid will be how IRs and updates to the company position will be addressed and how the completed AAid will be communicated to the sponsor after regulatory action has been taken.

CONCLUSION: PATHWAY FORWARD FOR PILOT EXPANSION

Great strides have been made in regulatory policy with the implementation of expedited programs, such as accelerated approval, breakthrough therapy designation, priority review, and fast track designation, to streamline the development and review of new therapies, but further optimization can still be achieved. Results from the first two RTOR supplemental application approvals and first use of the Assessment Aid garner optimism regarding the utility of both pilots to furthering this goal. However, the greatest value from both pilots will be gained from expansion into new settings where patients can achieve the greatest benefit from improvements in drug development and clinical trial designs for sustained efficiency. By expanding the complexity of the RTOR pilot in a robust and step wise approach, both FDA and the sponsor can gain valuable understanding and practice to ensure increased efficiency gains can be maintained, while also ensuring the quality of the review and risk-benefit decision. Ultimately, successes from the expansion of these programs can be an example for optimization by other health authorities and global harmonization to enable a greater number of patients to benefit from earlier access to important new drugs likely to demonstrate improvements over existing therapies.

TABLE GLOSSARY

ADaM – analysis data model
ASAP - Administrative Systems Automation Project
BIMO – bioresearch monitoring
CDRH – Center for Devices and Radiological Health
CDx – companion diagnostic
CMC – chemistry, manufacturing, and control
CQA – critical quality attribute
CSR – clinical study report
DMC – data monitoring committee
IR- information request
OSI – Office of Scientific Investigation
PAS – prior approval supplement
PDUFA – Prescription Drug User Fee Act
PMC – post-marketing commitment
SAP – statistical analysis plan
SAS – statistical analysis software
sBLA – supplemental BLA
SDTM – study data tabulation model
sNDA – supplemental NDA
USPI – US Prescribing Information

REFERENCES

[1] KISQALI® is a registered trademark of Novartis Pharmaceuticals Corporation (Novartis)

[2] KEYTRUDA® is a registered trademark of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc. (Merck)

[3] US Food and Drug Administration. (2018). *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics Guidance for Industry - DRAFT GUIDANCE*. Retrieved from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM621817.pdf?elqTrackId=1B8C5F95655CFB4E33D74C16AA376B41&elq=2f5ae25cb11a4c4594af5320a932d7af&elqaid=5289&elqat=1&elqCampaignId=4225

[4] US Food and Drug Administration. (2018). Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics Guidance for Industry - DRAFT GUIDANCE. Retrieved from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM616325.pdf

[5] FDA Guidance for Industry. Process Validation: General Principles and Practices, 2011

[6] FDA Guidance for Industry, Comparability Protocols for Human Drugs and Biologics: Chemistry, Manufacturing, and Controls Information, 2016

[7] US Food and Drug Administration. (2018). *Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product: Draft Guidance for Industry and Food and Drug Administration Staff - DRAFT GUIDANCE*. Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM510824.pdf

FRIENDS of CANCER RESEARCH

A Friends of Cancer Research White Paper

EXPLORING WHETHER A SYNTHETIC CONTROL ARM CAN BE DERIVED FROM HISTORICAL CLINICAL TRIALS THAT MATCH BASELINE CHARACTERISTICS AND OVERALL SURVIVAL OUTCOME OF A RANDOMIZED CONTROL ARM:

CASE STUDY IN NON-SMALL CELL LUNG CANCER

INTRODUCTION

The U.S. Food and Drug Administration (FDA) aims to expedite the development and review of products intended to address an unmet medical need in the treatment of serious life-threatening conditions through the breakthrough therapy designation (BTD) as well as fast track, accelerated approval (AA), and priority review mechanisms.[1] In the case of AA, randomized trials meant to establish clinical benefit normally conducted before approval, may be conducted after AA, to confirm clinical benefit. For drugs and biologics intended to treat a serious or life-threatening condition, the FDA may grant BTD if preliminary clinical evidence indicates the product may provide substantial improvement over existing therapies, on ≥ 1 clinically significant endpoint.[2] Many products with BTD are approved through the AA pathway. Although AA may allow patients access to therapies that have demonstrated a substantial treatment effect, this introduces loss of clinical equipoise that may interfere with continued drug development. For example, patients may be reluctant to enroll in trials where they may be randomized to receive a perceived inferior therapy, or they may discontinue from ongoing clinical trials once the product is accessible through AA. FDA guidance states, "If it is clear during development that a product is intended to be approved under accelerated approval… confirmatory trial(s) should be underway at the time the marketing application is submitted."[1] However, recruitment and conduct of the confirmatory trial must continue after AA. Data from the control arm may be compromised by early discontinuation or "crossover" to the investigational therapy made available by AA, resulting in an inability to interpret the confirmatory clinical trial results. Finally, there are some clinical settings (e.g., rare diseases) where scarcity of patients or ethical concerns have demonstrated that a randomized control is difficult or not feasible. These indications are often studied using single arm trials in which all enrolled patients receive the investigational agent.

CONTRIBUTORS

Ruthie Davi
Medidata Solutions

Andrea Ferris
LUNGevity Foundation

Andrew Howland
Medidata Solutions

Dominic Labriola
Bristol-Myers Squibb

Michael LeBlanc
Fred Hutchinson Cancer Research Center, University of Washington

David Lee
Medidata Solutions

Antara Majumdar
Medidata Solutions

Pallavi Mishra-Kalyani
U.S. Food and Drug Administration

Zhenming Shun
Daiichi Sankyo

Rajeshwari Sridhara
U.S. Food and Drug Administration

Larry Strianese
Medidata Solutions

Elizabeth Stuart
Johns Hopkins University

Joohee Sul
U.S. Food and Drug Administration

Xiang Yin
Medidata Solutions

Antoine Yver
Daiichi Sankyo

The same impact on patient recruitment and retention may occur in circumstances where the drug is approved and available for off-label use, or when drugs with similar mechanisms of action, in the same drug class are approved. Interpretation of study results, such as overall survival (OS), are compromised when patients use alternate treatments (whether off-label use of the product under investigation or a newly marketed alternate treatment). This phenomenon has been coined "cross-over" or "treatment switch-over" and while some drugs have demonstrated benefits in OS even after cross over, "better methods to capture and summarize the OS benefit are needed" to address confounding bias introduced by this practice.[3]

Consider the example of the large, randomized trial (BRAVO study) assessing the PARP inhibitor niraparib in patients with breast cancer and germline BRCA mutation carriers.[4] The sponsor of the trial announced, "A large number of patients in the chemotherapy control arm did not continue in the trial long enough to receive their first radiological scan, which is required to assess disease progression, resulting in an unusually high rate of censoring in the control arm." While the early discontinuation of these patients could be related to a toxicity of the drug, the sponsor conjectures that "this is likely associated with the desire of patients who carry germline BRCA mutations to be treated with a PARP inhibitor rather than chemotherapy and the increased availability of PARP inhibitors." The trial sponsor concluded that the study is, "unlikely to produce data that is interpretable."

An example of the consequences of treatment cross-over are seen in a trial in patients with gastrointestinal stromal tumor (GIST) described in the labeling of sunitinib.[5] This trial was a double-blind, randomized study comparing sunitinib malate to placebo and appears to have been designed and conducted in accordance with very high standards typical of pharmaceutical drug development. After an interim analysis demonstrated a large effect on progression free survival in favor of the sunitinib arm (HR 0.33, 95% CI (0.24, 0.47)), patients on the placebo arm were offered open label sunitinib malate; 99 of the 118 patients (84%) assigned to placebo elected to receive sunitinib malate. At the protocol specified final analysis, there was no difference observed in OS (median OS 72.7 weeks for the sunitinib malate arm and 64.9 weeks for the placebo arm, HR 0.88, 95% CI (0.68, 1.1)) by the original randomized arms. The absence of an effect at the final analysis time point is likely a result of the treatment "cross-over" in the placebo arm.

One approach to circumvent these challenges introduced by loss of equipoise is to consider the use of historical data to facilitate the conduct of clinical trials. Historical patient level data generally has been gathered before the experimental product or similar products are available on the market and while effects of other rescue therapy after progression cannot be ruled out, effects due to the pure treatment "cross-over" to the experimental therapy or very similar therapy generally are not present.

The potential use of historical clinical data in the context of randomized clinical trials was first discussed in the literature by Pocock (1976).[6] More recently, Lim et al. (2018) provided a comprehensive review of well-known frequentist and Bayesian methodologies for leveraging histor-

ical clinical trial data in a regulatory setting.[7] Use of historical clinical trials data to enhance current research has some precedent. For instance, historical clinical trials data and propensity score methods were used to construct a reference response rate for a single-arm study of Blinatumomab for relapsed/refractory acute lymphoblastic leukemia, a rare disease.[8] Lim et al. cite five drugs that incorporated historical control data in differing capacity, as part of a confirmatory clinical trial to obtain regulatory approvals between 2005 and 2015.[5] None of those approvals; however, involved a direct comparison of the historical control arm to that of the treatment arm through a standard hypothesis testing procedure. The research proposed in this document aims to fill that gap. By choosing to retrospectively evaluate a carefully constructed synthetic control arm, not only against the actual control arm, but in future work, also against the treatment arm, we aim to understand the extent to which a synthetic control arm could be used for pragmatic purposes in cancer drug development.

An example of the use of historical control data for internal drug development decision making at a pharmaceutical company is presented in Neuenschwander et al.[9] The discussion in that paper relates to non-confirmatory trials but can also be potentially used in a confirmatory trial setting. Rosmalen et al. present a comparative study of Bayesian methods to include historical data in the analysis of clinical trials data and stress the need to estimate the heterogeneity among trials and to satisfy criteria for comparability between the historical and current controls.[10] Hobbs et al. investigate an adaptive randomization procedure that makes assignment to experimental therapy more likely when there is an absence of evidence for heterogeneity among the concurrent and historical controls.[11]

Like any novel research initiative, the proposed use of historical control data to build a Synthetic Control Arm (SCA) has some associated risks. Selection bias and historical time effect are obvious risk factors. However, careful statistical planning and designing, along with a thorough understanding of the characteristics of the target population of interest, can help circumvent some of those risks. Pocock (1976) proposed a formal statistical plan for methodological inclusion of historical data in a randomized clinical trial.[6] Appropriate statistical inference procedures for the context are also discussed. In addition, simulation studies can aid in understanding the bias-variance trade off and more generally, the influence of the historical control data.

This project is a unique collaboration of multiple stakeholders including contributions from

- Bristol-Myers Squibb
- Daiichi Sankyo
- Fred Hutchinson Cancer Research Center
- Friends of Cancer Research
- Johns Hopkins University
- LUNGevity Foundation
- Medidata Solutions
- Project Data Sphere
- U.S. Food and Drug Administration

We are grateful for the data, expertise, and resources each party has provided.

DATA SOURCES

Data from two sources will be utilized in this project. Project Data Sphere[a] has provided patient level data from the control arms of three large randomized trials in non-small cell lung cancer (NSCLC). Medidata Solutions has provided patient level data from multiple clinical trials in NSCLC conducted by the pharmaceutical industry for purposes of drug development and are available in the Medidata Enterprise Data Store (MEDS). All patients in these trials presented at baseline with previously treated advanced NSCLC and were assigned to receive docetaxel in the control arm.

MEDS is a collection of thousands of previous clinical trials conducted by the pharmaceutical industry for drug or medical product development with patient level data recorded through the Medidata electronic data capture system. Per the legal agreements with the sponsors of these historical clinical trials and Medidata, these data are available for use in deidentified (i.e., patients and original sponsor of the trial cannot be identified) and aggregated (i.e., every analysis must include data from two or more sponsors) form.

ANALYSIS OBJECTIVES

The scope of this work is to explore the potential applications of historical clinical trials data in randomized clinical trials, with the aim of minimizing the number of patients required to be assigned to the control arm and providing a better understanding of the effects of the experimental therapy independent of the effect of treatment "cross-over" assuming the historical clinical trials data has been generated at a time when the current experimental therapy was not available.

This project will explore whether a *synthetic control arm* (SCA) can mimic the results of a traditional randomized control. This will be investigated with a case study in previously treated advanced NSCLC as follows.

- First, one of the three historical trials provided by Project Data Sphere will be selected and designated as the 'Target Trial A'. This selection is limited to Project Data Sphere studies since MEDS studies may not be displayed individually. Legal restrictions governing MEDS data require analyses to be aggregate, that is including data from two or more sponsors.

- Next, a SCA will be constructed using patient-level data from all other available historical data in NSCLC. Patients in the SCA will be selected to match the control patients in the Target Trial A based on important baseline characteristics and prognostic factors and with a propensity score matching approach.

[a] Project Data Sphere is a platform where the research community can share historical patient level data from academic and industry phase 3 cancer clinical trials. The analyses in this case study are at least partially based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owner(s) of any information from the website have contributed to, approved, or are in any way responsible for the contents of this work.

- Third, we will evaluate whether this matching has been successful by examining differences in baseline characteristics and prognostic scores in the target trial control arm and the SCA, as well as by exploring whether OS results observed for the target trial control arm are replicated in the SCA.

- Finally, additional evidence will be gained by repeating this process for a second Project Data Sphere trial designated as 'Target Trial B'. The process will not be repeated for the third Project Data Sphere trial since this trial is smaller than the others and fewer baseline variables are available for the matching processes.

Future research may be undertaken to explore whether a SCA can be used to mimic the treatment effect from a traditional randomized controlled trial. In that case, a SCA will be created to match the experimentally treated patients in the target trial and comparisons of the treatment effect using the randomized control and the same using the SCA will be made.

KEY FEATURES OF HISTORICAL DATA AND SCA ELIGIBILITY CRITERIA

Key features of the historical data and SCA eligibility criteria are described in this section. These studies were selected, and eligibility criteria were defined, based on clinical importance, balancing the need to identify a fairly homogenous set of historical clinical trial participants representative of a typical single indication in drug development and the desire to identify the largest volume of applicable historical data as possible.

As shown in Table 1, the historical data originated from open label or blinded phase 2 or 3 multinational trials, which began between 2004 and 2013. Enrollment in Target Trial A began in February of 2004 and the study reached its primary efficacy analysis time point in March 2007. Target Trial B began enrollment in May of 2006 and reached its primary efficacy analysis timepoint in August 2008. All patients were previously treated and presented at baseline with locally advanced or metastatic NSCLC. All patients were included in study arms that assigned treatment with docetaxel. Overall survival was measured as a key endpoint in all trials. One thousand three hundred ninety-nine (1,399) historical patients are available for this case study.

Table 1: Features of Historical Data

Table 1: Features of Historical Data							
	Design	**Region**	**Start/End of Trial(s)**	**Baseline Characteristics**	**Endpoints**	**Number of Patients**	**Control regimen**
Historical data (from multiple trials)	Open label or blind-ed, phases 2 or 3	Multi-	Began between 2004 and 2013. Ended btwn 2007 and 2016.	Previously treated local-ly advanced or metastatic non-small cell lung cancer	Overall survival measured	1399	Docetaxel

Eligibility criteria for the SCA are shown in Table 2. All patients in this set of 1,399 met these requirements at baseline. Historical patient level data, including assessments of eligibility criteria and other screening measurements from source historical trials were used to make these assessments.

Table 2: SCA Patient Eligibility Criteria

Table 2: SCA Patient Eligibility Criteria
1. Inclusion in a historical clinical trial accessible within this project
2. NSCLC stage III or IV at baseline
3. Received prior platinum-based chemotherapy
4. Men and women ≥ 18 years of age
5. Eastern Cooperative Oncology Group (ECOG) performance status of ≤ 2
6. Had measurable disease
7. Assigned to receive docetaxel as study treatment

ENDPOINTS AND COVARIATES

Because the historical data in this case study come from trials that have been conducted as part of clinical development programs and because methods for investigation of many indications in a regulatory setting are somewhat standardized by precedent, the populations, study design, data collection methods, and endpoints utilized in these trials are similar across trials. Overall survival is the endpoint of interest for this case study and was measured as a key outcome in all historical trials. Differences across studies in covariate definitions were present but have been reconciled as part of the data standardization process. Clinically important baseline covariates available across studies and to be used in the propensity score matching process are shown in Table 3.

Table 3: Clinically Important Baseline Covariates Utilized in Propensity Score Matching

Table 3: Clinically Important Baseline Covariates Utilized in Propensity Score Matching
1. Age at baseline (continuous)
2. Years from cancer diagnosis (continuous)
3. Race (White vs Others)
4. Sex (Female vs Male)
5. Smoking (Current vs Former vs Never)
6. Histology (Squamous vs Non-squamous)
7. Stage (III vs IV)
8. ECOG (0 vs 1 vs 2)
9. Prior surgery (Yes/Maybe vs No)
10. EGFR/KRAS mutation (Positive vs No/Unknown)

MATCHING METHODS AND EFFICACY ANALYSES

Propensity score matching is commonly used to analyze observational data to reduce bias due to confounding variables that are unbalanced between groups of interest (e.g., patients that received the treatment and those that did not). In the context of randomized clinical trials, the presumption is that the treatment groups will be generally balanced in terms of baseline covariates due to randomization and so differences between treatment and control can be reliably attributed to the treatment assignment. The intent of this project is to explore whether historical clinical trials data and matching procedures can stand-in for prospective patients and random assignment to treatment with standard of care in indications where there may be loss of equipoise.

Rubin and Thomas (1992) derive analytic expressions for the effect of matching using linear propensity scores with normally distributed covariates and find that substantial reductions in bias and variance are possible when these conditions are met.[12] Rubin and Thomas (1996) extend these results to covariates with a symmetrical ellipsoidal distribution, such as t-distributions.[13] Using Monte Carlo simulations, they confirm the accuracy of analytical approximations under normal and non-normal ellipsoidal distributions. Ho, et. al. (2007) further demonstrate that nonparametric matching using estimated propensity scores reduces the degree of model dependence, resulting in estimated treatment effects that exhibit greater robustness to researchers' parametric assumptions relative to analytic methods without data preprocessing by matching procedures.[14]

Using the guidelines proposed in Ho, et. al. (2007),[14] the following procedures will be used to carry out the propensity score matching:

Step 1: Estimate propensity scores. The propensity score is the probability of assignment of target trial control therapy conditional on the baseline characteristics (i.e., potential confounders) using logistic regression

$$p(\boldsymbol{x}) = P(T = 1|\boldsymbol{X} = \boldsymbol{x})$$

where T denotes the control in the target trial (T=1) / historical control (T=0) and $\boldsymbol{X}$ is a vector representing the covariates to be included in the propensity score model (see Box 1 for an additional explanation of propensity score matching). The predictors included in the propensity score model are all available baseline characteristics described in Table 3. These baseline covariates will be utilized without further variable selection or trimming to obtain optimal balance between the matched subjects. Using a large set of covariates is recommended, even if some of the covariates are only related to self-selection and other covariates, and not necessarily to the outcome of interest.[15,16] Some researchers recommend using all available baseline covariates in the analysis if the sample size permits.[7]

Step 2: Create SCA by selecting historical patients to match control patients in the target trial using the estimated propensity scores. We will use greedy nearest-neighbor matching without replacement and a fixed 1-to-1 matching ratio, which aligns with the commonly used 1:1 randomization ratio in NSCLC historical trials. The control patients in the target trial will be randomly ordered. We will start from the first control patient in the target trial and will match the patient to a historical patient whose propensity score is closest to that of the control patient from the target trial and within a prespecified maximum distance (i.e., caliper). A caliper width equal to 0.25 of the pooled standard deviation of logit of the propensity score from the 2 groups, a widely utilized rule of thumb, will be used.[17] We will conduct matching without replacement, that is, the matched historical patients will be removed from consideration for further matching and next target trial control patients will be selected. This process will be repeated sequentially for all control patients in the target trial. The matched patients from the historical group are the components of SCA.

Step 3: Post-matching evaluation of covariate balance. The true propensity score should be a balancing score. We will examine whether the distribution of measured baseline covariates is similar between the matched target trial control arm and historical SCA subjects. Baseline demographic and disease characteristics will be summarized with descriptive statistics for the target trial control arm and SCA both before and after matching. Standardized difference in covariate means before matching and after matching will be computed and compared.

For a continuous covariate, the standardized difference is:

$$d = \frac{\overline{x_t} - \overline{x_c}}{\sqrt{(s_t^2 + s_c^2)/2}}$$

Where $\overline{x_t}$ and $\overline{x_c}$ denote the sample mean of the covariate for the target trial control and historical control groups, respectively; s_t^2 and s_c^2 denote the sample variance of the covariate for the target trial control and historical control groups, respectively.

For dichotomous (or categorical) variables, the standardized difference is defined as:

$$d = \frac{\hat{p}_t - \hat{p}_c}{\sqrt{\{\hat{p}_t(1-\hat{p}_t) + \hat{p}_c(1-\hat{p}_c)\}/2}}$$

Where $\hat{p}_t$ and $\hat{p}_c$ denote the prevalence of covariate (or a category of covariate) for the target trial control and historical control groups, respectively. For covariates with more than 2 categories, the standardized difference for each level of the categorical variable will be calculated.

The absolute standardized differences should generally be less than 0.25.[15] An absolute standardized difference of less than 0.10 has been taken to indicate a negligible difference in the mean or prevalence of a covariate between treatment groups.18 In addition, the matching process will be evaluated by examining the distribution of propensity scores, as well as individual baseline characteristics, including prognostic factors between the target trial control arm and SCA using graphical methods such as cloud plots, box plots, and quantile-quantile (Q-Q) plots. For continuous covariates, we will also summarize the mean and maximum deviation between the two empirical distributions in the Q-Q plots on the scale of the variables being measured.

Step 4: To explore whether OS observed in the control arm of the target trial is replicated by SCA, we will examine the similarity of OS between the SCA and target trial with the hazard ratio and associated 95% confidence interval for both before and after matching. Kaplan-Meier curves will be presented along with estimates of the median and other percentiles of survival times and 95% confidence intervals both before and after matching. Commonly used tests for differences in survival curves (i.e., log-rank test, Wilcoxon test, and likelihood ratio test) will also be presented both before and after matching.

Box 1. A non-technical description of propensity score matching and its possible effects

For illustration of the nature of propensity score matching, first consider a simplistic example where the number of important baseline characteristics is quite small, say age and ECOG score alone. Then for each patient in the target, we seek a patient from the historical pool with the same age and ECOG score. Assuming the amount of historical data is plentiful, this would lead to certain balance between the SCA and the target arm in terms of important baseline characteristics, age, and ECOG. However, the number of important baseline characteristics is rarely small and the scarcity of patients with exactly the same covariate pattern becomes problematic when the number of important covariates is larger. The propensity score can be thought of as a summarization of all the important baseline characteristics and their relationship to whether a patient is eligible to receive the therapy being studied. A key advantage of the propensity score approach is the reduction in dimension (i.e., many important baseline covariates) to a single value (i.e., propensity score). Achieving a match for most or all target patients on their propensity score is much more likely to be successful than requiring a direct match on many covariates at once. Matching on the propensity score likely will not provide exact balance between groups on all important baseline characteristics; rather, it will provide approximate balance for many baseline characteristics. Even with a propensity score approach there are some patients for whom an appropriate match will not be present in the available historical pool. In these cases, it is common practice to exclude these patients from the target matched set and proceed. To many accustomed to analyzing clinical trials, this practice may seem alarming and in direct contradiction to the intent-to-treat principle normally relied upon in clinical trials to preserve the balance between treatment groups afforded by random treatment assignment. However, in this setting, randomization is not utilized and removing patients from the target improves balance between groups rather than threatens it. This practice of removing patients from the target could restrict the matched target patients to a set of patients with baseline characteristics that are not as wide ranging as is present in the overall disease population and so the appropriateness of extrapolating the analysis of this precise set and applying it to a more varied population should be considered.

RESULTS

PERFORMANCE OF SCA MATCHING PROCESS

Baseline Characteristics

The control arm in Target Trial A included 459 patients. As shown in Table 4, most patients were white (65%), male (63%), and current or former smokers (16% and 60%, respectively). Prior surgery was reported in 35% of patients and the rate of known EGFR or KRAS mutation was 7%. Patients commonly had non-squamous type NSCLC (78%), ECOG scores of 0 or 1 (24% and 67%, respectively), and disease stage 4 (87%).

The pool of historical clinical trial subjects available for possible inclusion in the SCA included 940 patients. As shown in Table 4, these patients were similar to the Target Trial A control arm in terms of age, years since cancer diagnosis, race, gender, ECOG score, and EGFR/KRAS mutation. Differences between the historical pool and Target Trial A control were evident though in the rate of current smokers (28% vs. 16%) and former smokers (46% vs. 60%), non-squamous disease (87% vs. 78%), disease stage 4 (77% vs. 87%), and prior surgery (9% vs. 35%).

Table 4. Baseline Characteristics by Arm Before and After Matching – Target Trial A

Baseline Characteristic	Before Matching		Matched		Unmatched
	Historical Pool (N=940)	Control in Target Trl A (N=459)	SCA (N=366)	Control in Target Trial A (N=366)	Control in Target Trial A (N=93)
Age at baseline, mean (std)	57.6 (10.5)	56.8 (11.0)	57.4 (11.0)	57.0 (10.7)	56.1 (12.1)
Years from cancer diagnosis, median (Q1, Q3)	0.7 (0.5, 1.0)	0.8 (0.5, 1.3)	0.7 (0.5, 1.0)	0.7 (0.5, 1.1)	1.3 (0.7, 1.9)
Race – White n (%)	645 (69%)	299 (65%)	239 (65%)	239 (65%)	60 (65%)
Sex – Female n (%)	316 (34%)	172 (37%)	128 (35%)	133 (36%)	39 (42%)
Smoking, n (%) Current Former Never	 267 (28%) 436 (46%) 237 (25%)	 74 (16%) 276 (60%) 109 (24%)	 66 (18%) 211 (58%) 89 (24%)	 71 (19%) 208 (57%) 87 (24%)	 3 (3%) 68 (73%) 22 (24%)
Histology – Squamous, n (%)	120 (13%)	100 (22%)	65 (18%)	67 (18%)	33 (35%)
Stage – III, n (%)	213 (23%)	58 (13%)	54 (15%)	54 (15%)	4 (4%)
ECOG, n (%) 0 1 2	 334 (36%) 545 (58%) 61 (7%)	 112 (24%) 306 (67%) 41 (9%)	 85 (23%) 254 (69%) 27 (7%)	 100 (27%) 233 (64%) 33 (9%)	 12 (13%) 73 (78%) 8 (9%)
Prior surgery – Yes/Maybe, n (%)	83 (9%)	162 (35%)	66 (18%)	69 (19%)	93 (100%)
EGFR/KRAS mutation – Positive, n(%)	33 (4%)	33 (7%)	13 (4%)	16 (4%)	17 (18%)

The control arm in Target Trial B included 542 patients. As shown in Table 5, most patients were white (54%), male (67%), and current or former smokers (34% and 39%, respectively). Prior surgery was reported in 1% of patients and the rate of known EGFR or KRAS mutation was 6%. Patients commonly had non-squamous type NSCLC (79%), ECOG scores of 0 or 1 (33% and 64%, respectively), and disease stage 4 (84%).

The pool of historical clinical trial subjects available for possible inclusion in the SCA included 857 patients. As shown in Table 5, these patients were similar to the Target Trial B control arm in terms of age, years since cancer diagnosis, gender, ECOG score, and EGFR/KRAS mutation. Differences between the historical pool and Target Trial B control were evident though in the rate of white patients (76% vs. 54%), the rate of current smokers (18% vs. 34%) and former smokers (59% vs. 39%), non-squamous type NSCLC (88% vs. 79%), disease stage 4 (78% vs. 84%), and prior surgery (28% vs. 1%).

Table 5. Baseline Characteristics by Arm Before and After Matching – Target Trial B

Baseline Characteristic	Before Matching	Matched		Unmatched	
	Historical Pool (N=857)	Control in Target Trial B (N=542)	SCA (N=417)	Control in Target Trial B (N=417)	Control in Target Trial B (N=125)
Age at baseline, mean (std)	58.0 (10.3)	56.2 (11.1)	57.1 (10.5)	57.0 (11.0)	53.6 (11.1)
Years from cancer diagnosis, median (Q1, Q3)	0.7 (0.5, 1.1)	0.7 (0.4, 1.0)	0.7 (0.5, 1.0)	0.7 (0.4, 1.0)	0.6 (0.4, 1.0)
Race – White n (%)	653 (76%)	291 (54%)	276 (66%)	270 (65%)	21 (17%)
Sex – Female n (%)	308 (36%)	180 (33%)	143 (34%)	140 (34%)	40 (32%)
Smoking, n (%) Current Former Never	 157 (18%) 503 (59%) 197 (23%)	 184 (34%) 209 (39%) 149 (28%)	 109 (26%) 199 (48%) 109 (26%)	 106 (25%) 196 (47%) 115 (28%)	 78 (62%) 13 (10%) 34 (27%)
Histology – Squamous, n (%)	104 (12%)	116 (21%)	65 (16%)	67 (16%)	49 (39%)
Stage – III, n (%)	186 (22%)	85 (16%)	78 (19%)	69 (17%)	16 (13%)
ECOG, n (%) 0 1 2	 266 (31%) 503 (59%) 88 (10%)	 180 (33%) 348 (64%) 14 (3%)	 142 (34%) 258 (62%) 17 (4%)	 134 (32%) 269 (65%) 14 (3%)	 46 (37%) 79 (63%) 0 (0%)
Prior surgery – Yes/ Maybe, n (%)	241 (28%)	4 (1%)	4 (1%)	4 (1%)	0 (0%)
EGFR/KRAS mutation – Positive, n (%)	33 (4%)	33 (6%)	14 (3%)	12 (3%)	21 17%)

Propensity Score Matching

As specified in the analysis plan, propensity score matching was utilized to attempt to select the appropriate patients from the historical pool for inclusion in the SCA so that the distribution of baseline characteristics would be well balanced between the SCA and the control from the target trial. This section details evidence that leads to the conclusion that indeed the matched groups are well balanced in terms of all observed baseline characteristics. The same conclusion is reached for both Target Trial A and Target Trial B.

The Cloud Plot in Figure 1A shows the distribution of propensity scores for the control arm of Target Trial A and the pool of historical patients available for inclusion in the SCA and the degree to which these distributions overlap. Green dots represent patients who are successfully matched with a patient in the opposite group with a similar propensity score. Red circles and blue x's represent patients for whom a match is not available. These are generally in the tails of the distribution of the target trial and visually we can see that there are no analogous patients available in this region of the historical pool. Patients in the target trial control arm who cannot be matched with a patient from the historical pool are excluded from further analysis.

Excluding unmatched target trial patients from further analysis is a common practice when utilizing matching methods. To many accustomed to analyzing clinical trials, this practice may seem alarming and in direct contradiction to the intent-to-treat principle normally relied upon in clinical trials to preserve the balance between treatment groups afforded by random treatment assignment. However, in this setting, randomization is not utilized and removing patients from the target improves balance between groups rather than threatens it (in essence, prioritizing internal validity over external validity). This practice of removing patients from the target could restrict the matched patients to a set of patients with baseline characteristics that are not as wide ranging as is present in the target or overall disease setting and so the appropriateness of extrapolating the analysis of this precise set and applying it to a more varied population should be considered.

A similar display is shown for Target Trial B in Figure 1B.

Figure 1A. Cloud Plot for Target Trial A Illustrating Distribution of Propensity Scores

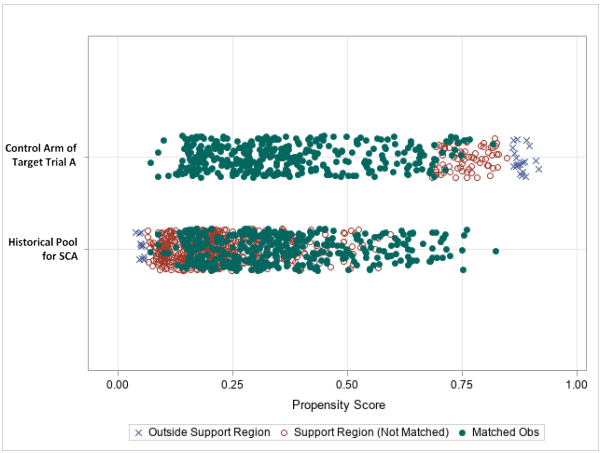


Figure 1B. Cloud Plot for Target Trial B Illustrating Distribution of Propensity Scores

The control arm in Target Trial A included 459 patients. Overlap in the distribution of propensity scores for the control arm of Target Trial A and the historical pool was significant but not complete. Three hundred sixty-six (80%) of the Target Trial A patients were successfully matched. The remaining 93 patients (20%) were not matched and were removed from further analysis. The baseline characteristics of the matched patients as well as the set of excluded unmatched patients from the target are described in Table 4. Baseline characteristics for the SCA and control arm in Target Trial A after matching now appear to be well balanced between groups, even for characteristics where differences were observed between the historical pool and target trial before matching. The most notable characteristic of the set of target patients who are not matched and are excluded from further analysis is the rate of patients with prior surgery. Attention should be given to the question of whether an analysis of patients with low rates of prior surgery can be extrapolated to the overall population, including patients with prior surgery.

The control arm in Target Trial B included 542 patients. Overlap in the distribution of propensity scores for the control arm of Target Trial B and the historical pool was significant but not complete. Four hundred seventeen (77%) of the target trial patients were successfully matched. The remaining 175 patients (23%) were not matched and were removed from further analysis. The baseline characteristics of the matched patients as well as the set of excluded unmatched patients from the target are described in Table 5. Baseline characteristics for the SCA and control arm in Target Trial B after matching now appear to be well balanced between groups, even for characteristics where differences were observed between the historical pool and target trial before matching. The most notable characteristics of the set of target patients who are not matched and are excluded from further analysis is the rate of white patients and rate of current smokers. Attention should be given to the question of whether an analysis of patients with differences in these characteristics be extrapolated to the overall population.

The balance between groups noted by numerical examination of the baseline characteristics can be explored further through graphical displays commonly used for the evaluation of the degree of success of the propensity score matching approach. Figures 2A and 2B provide a box plot and Q-Q plot respectively of the distribution of the propensity score before and after matching for Target Trial A. Figures 3A and 3B provide the same for Target Trial B. In all cases, significant gains in the comparability of the groups after matching are evident.

The distributions of the propensity score for the target trial and historical pool including all patients before matching are shown in the lower set of boxplots in Figures 2A and 3A. The analogous distributions after matching are shown in the upper region of these figures. There is considerable discordance between the target and historical pool before matching. In the case of Target Trial A, the median for the control is higher than that of the historical pool and the variability in scores is larger in the control than the historical pool. However, after matching, both the median and variability of the groups are very similar as evidenced by the similar placement of the median line and width of the 'box' in the boxplots for the groups. In the case of Target Trial B, the median for the control is higher than that of the historical pool and the variability in scores is smaller in the control than the historical pool. However, after matching, both the median and variability of the groups are very similar.

Q-Q plots are scatterplots created by plotting the quantiles for one group of data against another. Quantiles are cut points that divide the range of a probability distribution into continuous intervals with equal probabilities. For example, a commonly used set of quantiles are 'quartiles', and they divide the distribution into quarters. The first quartile is defined as the middle number between the smallest number and the median of the data set. The second quartile is the median of the data. The third quartile is the middle value between the median and the highest value of the data set. Although this may seem a complex derivation, the Q-Q plot provides a straightforward interpretation for assessing similarity between groups. If both sets of quantiles come from the equal distributions, we will see the points forming a line that's roughly straight from the origin at 45^0. The blue dots in the Q-Q plots in Figures 2B and 3B are a comparison of the quantiles in the historical pool to that of the Target Trial A control before matching. The red dots are the analogous comparison after matching. As evidenced by the red dots falling right along the 45^0 reference line and the blue dots not forming a straight line and being some distance from the reference line, we conclude that the degree of similarity in the distributions after matching is better than before matching. The mean (standard deviation) of deviation in propensity score between the two groups in the Q-Q plots changed from 0.121 (0.065) before matching to 0.001 (0.003) after matching. A similar result holds for Target Trial B.

Figure 2A. Distribution of Propensity Scores Before and After Matching – Target Trial A

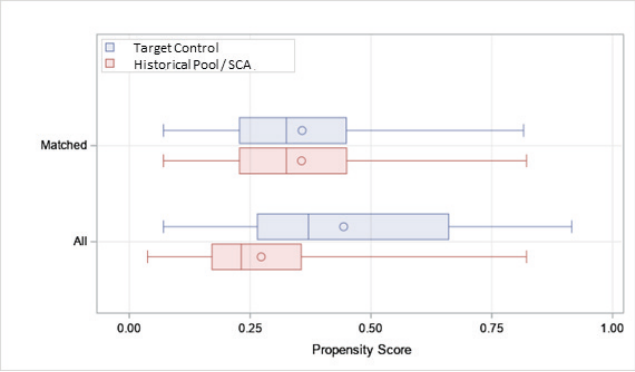


Figure 2B. Q-Q of Propensity Scores Before and After Matching – Target Trial A

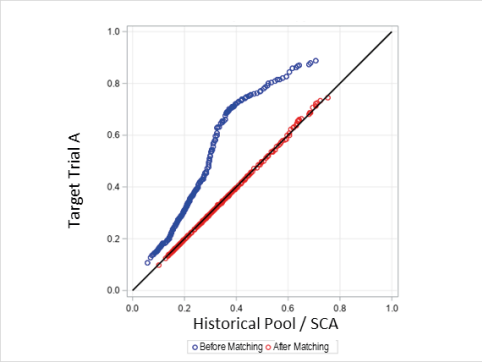


Figure 3A. Distribution of Propensity Scores Before and After Matching – Target Trial B

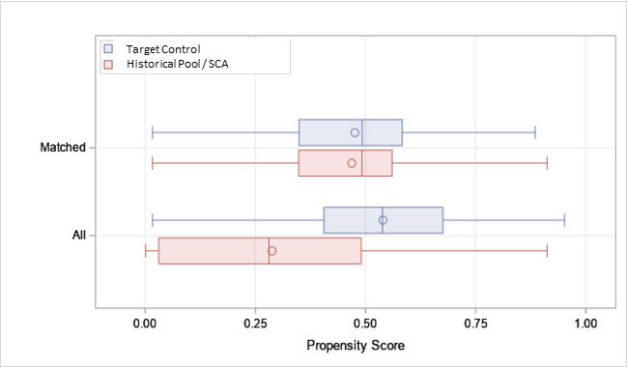
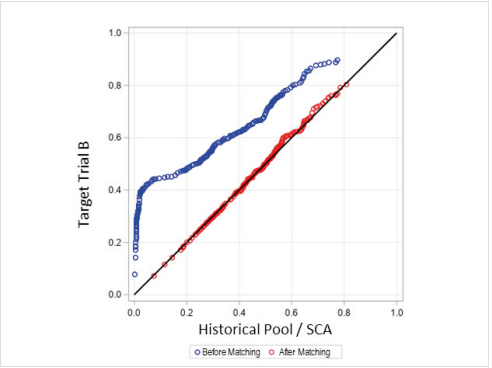


Figure 3B. Q-Q of Propensity Scores Before and After Matching – Target Trial B



Assessment of balance in terms of individual baseline covariates yields observations consistent with the conclusions afforded above by examination of the propensity scores. Figure 4A illustrates the standardized difference between the target trial and historical pool (before matching)/SCA (after matching) for each important baseline characteristic for Target Trial A. Figure 4B provides the same for Target Trial B. In all cases, reductions in the absolute standardized difference between groups for each variable are observed and the absolute standardized differences after matching are well below 0.10, the pre-specified threshold for designating a negligible difference in the mean or prevalence of a covariate between groups, for all but one instance.

Figure 4A. Plot of Standardized Difference of Important Baseline Covariates Before and After Matching – Target Trial A

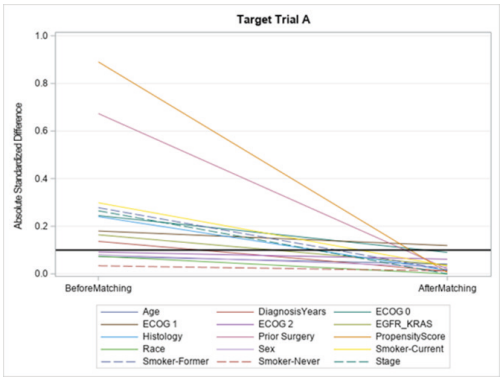
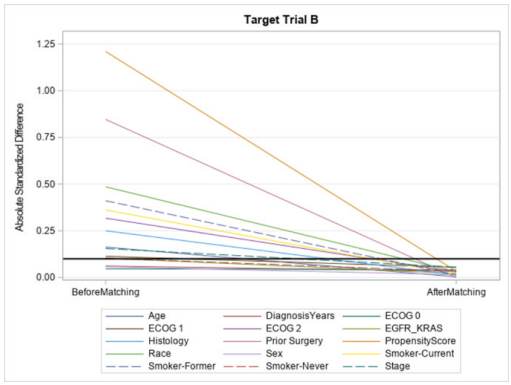


Figure 4B. Plot of Standardized Difference of Important Baseline Covariates Before and After Matching – Target Trial B



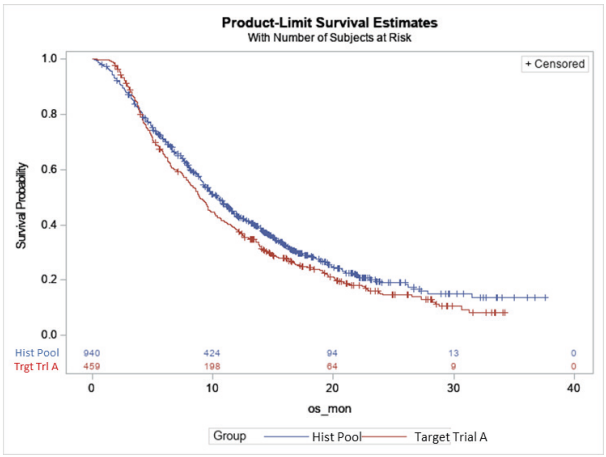
ASSESSMENT OF OVERALL SURVIVAL REPLICATION WITH SCA

In previous sections, we have demonstrated that the propensity score matching successfully balanced the distribution of baseline characteristics between the SCA and the control from the target trial. The main objective of this case study though is to explore whether the outcome of the randomized control arm from the target trial can be replicated using the SCA. This section details evidence that leads to the conclusion that indeed the OS for the SCA is very similar to that of the target trial. The same conclusion is reached for both Target Trial A and Target Trial B.

Figures 5A and 5B provide a comparison of the OS between the control arm of Target Trial A and the historical pool (before matching)/SCA (after matching), respectively. Before matching, there is a suggestion that the curves differ, as evidenced by little overlap of the Kaplan-Meier curves and space present between the curves suggesting that the OS for the Target Trial A is worse than that of the historical pool. The median survival was 8.9 months in the target versus 10.4 in the historical pool. The hazard ratio for the target relative to the historical pool was 1.16 with confidence interval that excludes 1 (95% CI 1.02, 1.32). This difference between groups is further supported by the log rank, Wilcoxon, and likelihood ratio tests comparing the difference in these curves (p=0.03, 0.07, and 0.04, respectively). After matching; however, there is significant overlap in the Kaplan-Meier curves for the target and SCA. The median survival was

8.8 months in the target versus 9.2 months in the SCA. The hazard ratio for the target relative to the SCA was 1.04 with confidence interval that includes 1 and indicates the plausible range for the HR is between 0.88 and 1.23, suggesting similarity of the SCA and target trial control arm in terms of OS. This similarity between groups is further supported by the log rank, Wilcoxon, and likelihood ratio tests comparing the difference in these curves (p=0.65, 0.97, and 0.66, respectively).

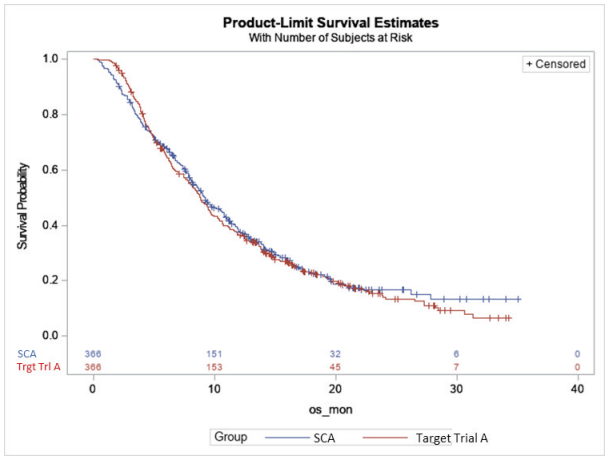
Figure 5A. Comparison of Overall Survival in Control Arm of Target Trial A versus Historical Pool (Before Matching)



Quartile	Hist Pool	Target
75	19.8 (18.4, 22.1)	17.4 (14.9, 20.1)
50	10.4 (9.6, 11.1)	8.9 (8.2, 9.6)
25	5.1 (4.4, 5.6)	4.6 (4.1, 5.0)

Log-Rank p-value: 0.03
Wilcoxon p-value: 0.07
-2Log(LR) p-value: 0.04
HR (Target vs Hist Pool), 95% CI: 1.16 (1.02, 1.32)

Figure 5B. Comparison of Overall Survival in Control Arm of Target Trial A versus SCA (After Matching)

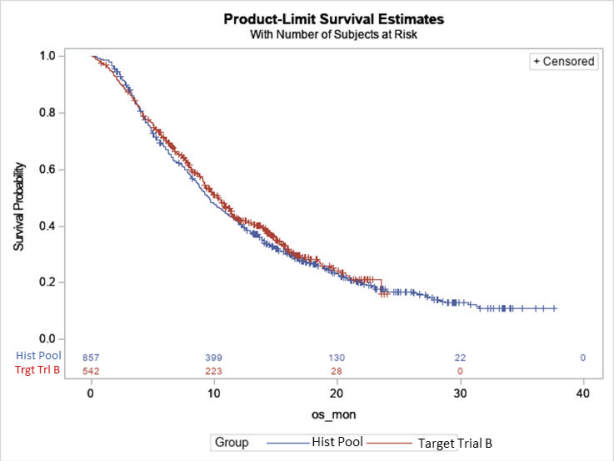


Quartile	SCA	Target
75	17.0 (14.9, 19.6)	16.6 (14.3, 19.6)
50	9.2 (8.2, 10.7)	8.8 (7.9, 9.6)
25	4.4 (3.6, 5.3)	4.6 (4.1, 5.0)

Log-Rank p-value: 0.65
Wilcoxon p-value: 0.97
-2Log(LR) p-value: 0.66
HR (Target vs SCA), 95% CI: 1.04 (0.88, 1.23)

Similar results are observed for Target Trial B (Figures 6A and 6B). Although the difference in OS between the control in Target Trial B and historical pool before matching is not clear, as it was with Target Trial A, there is still evidence that the similarity in OS is enhanced by the propensity score matching. After matching, the median survival was 9.9 years in the target versus 9.6 years in the SCA. The hazard ratio for the target relative to SCA was 1.01 with confidence interval that includes 1 and indicates the plausible range for the HR is between 0.85 and 1.19, suggesting similarity of the SCA and target control. This similarity between groups is further supported by the log rank, Wilcoxon, and likelihood ratio tests comparing the difference in these curves (p=0.91, 0.98, and 0.94, respectively).

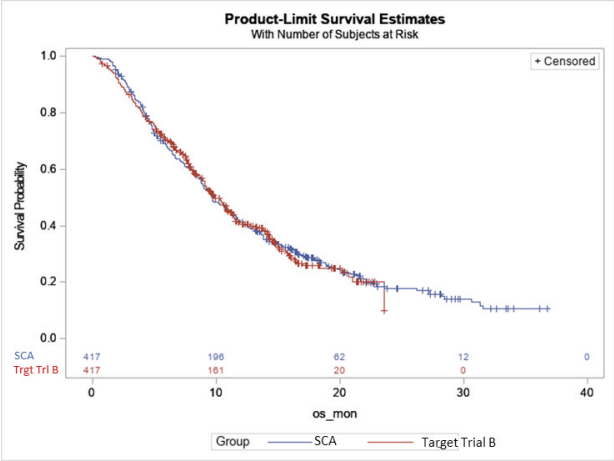
Figure 6A. Comparison of Overall Survival in Control Arm of Target Trial B versus Historical Pool (Before Matching)



Quartile	Hist Pool	Target
75	19.1 (16.9, 20.5)	19.7 (16.5, NE)
50	9.5 (8.9, 10.3)	10.4 (9.3, 11.3)
25	4.8 (4.2, 5.1)	5.1 (4.3, 5.9)

Log-Rank p-value: 0.35
Wilcoxon p-value: 0.37
-2Log(LR) p-value: 0.39
HR (Target vs Hist Pool), 95% CI: 0.94 (0.82, 1.07)

Figure 6B. Comparison of Overall Survival in Control Arm of Target Trial B versus SCA (After Matching)



Quartile	SCA	Target
75	19.6 (17.0, 22.1)	18.4 (15.8, NE)
50	9.6 (8.8, 11.0)	9.9 (9.0, 10.9)
25	4.8 (4.3, 5.3)	5.0 (4.1, 5.9)

Log-Rank p-value: 0.91
Wilcoxon p-value: 0.98
-2Log(LR) p-value: 0.94
HR (Target vs SCA), 95% CI: 1.01 (0.85, 1.19)

CONCLUSIONS

With this case study in NSCLC, we have demonstrated that it is possible to produce "matched" cohorts of patients from historical clinical trials using propensity scores derived from observed baseline characteristics. In these examples, the OS for the SCA was observed to be very similar to that of the randomized control. Further research is needed to build a broader body of experience and to identify the circumstances under which this approach is feasible and appropriate. An assessment of whether a synthetic control can be used to replicate the treatment effect (difference between arms) of a randomized controlled trial, as well as an assessment of sensitivity to unknown or unobserved confounders is planned by this working group. Exploration of alternative matching methods, in addition to the 1-1 nearest neighbor caliper matching without replacement used in this case study, may make it possible to reduce the proportion of unmatched patients and resolve extrapolation concerns.

Overall, the results of this case study are promising and represent an important step toward understanding whether the use of SCA can inform the design of a randomized trial, potentially minimizing the number of patients required to be assigned to a control arm. This approach may mitigate many of the challenges faced when enrolling or maintaining a concurrent control arm is difficult due to rarity of the disease, or availability of the investigational agent outside the study.

REFERENCES

[1] FDA Guidance for Industry Expedited Programs for Serious Conditions – Drugs and Biologics. 2014. Available online: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm358301.pdf

[2] Food and Drug Safety Innovation Act. 2012. Available online: https://www.congress.gov/bill/112th-congress/senate-bill/3187.

[3] Sridhara R., He K., Nie L., Shen Y, Tang S. 2015. Current Statistical Challenges in Oncology Clinical Trials in the Era of Targeted Therapy. *Statistics in Biopharmaceutical Research*. 7:348-356.

[4] Tesaro press release. 2017. Available online: http://ir.tesarobio.com/news-releases/news-release-details/tesaro-announces-expanded-development-program-niraparib-focused

[5] Sunitinib malate capsule prescribing information. 2006. Available online: http://labeling.pfizer.com/ShowLabeling.aspx?id=607

[6] Pocock, S. 1976. The combination of randomized and historical controls in clinical trials. J Chron Dis. 29:175-188.

[7] Lim J., Walley R., Yuan J., Liu J., Dabral A., Best N., Grieve A., Hampson L., Wolfram J., Woodward P., Yong F., Zhang X., Bowen E. 2018. Minimizing patient burden through the use of historical subject-level data in innovative confirmatory clinical trials: review of methods and opportunities. DIA Therapeutic Innovation and Regulatory Science.

[8] Gökbuget, N, Kelsh, M, Chia, V, Advani, A, Bassan, R, Dombret, H, Doubek, M, Fielding, AK, Giebel, S, Haddad, V, Hoelzer, D, Holland, C, Ifrah, N, Katz, A, Maniar, T, Martinelli, G, Morgades, M, O'Brien, S, Ribera, JM, Rowe, JM, Stein, A, Topp, M, Wadleigh, M, Kantarjian, H. 2016. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood cancer journal, 6(9), e473.

[9] Neuenschwander, B., Capkun-Niggli, G., Branson, M., Spiegelhalter, D. 2010. Summarizing historical information on controls in clinical trials. *Clinical Trials*. 7:5-18.

[10] Rosmalen J., Dejardin D., van Norden Y., Lowenberg B., Lesaffre E. 2017. Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical Methods in Medical Research*.

[11] Hobbs, B, Carlin, B., Sargent, D. 2013. Adaptive adjustment of the randomization ratio using historical control data. *Clin Trials*. 10:430-440.

[12] Rubin, D., Thomas, N. 1992. Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika*. 79:797–809.

[13] Rubin, D., Thomas, N. 1996. Matching using estimated propensity scores, relating theory to practice. *Biometrics*. 52:249–64.

[14] Ho, DE, Imai, K, King, G, Stuart, E. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*. 15(3):199–236.

[15] Stuart, E., Rubin, D. 2007. Best Practices in Quasi-Experimental Designs: Matching methods for causal inference. Chapter 11 in *Best Practices in Quantitative Social Science*. J. Osborne (Ed.). Thousand Oaks, CA: Sage Publications. 11:155-176.

[16] Harris, H., Horst, S. 2016. A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process. *Practical Assessment, Research & Evaluation*. 21(4). Available online: http://pareonline.net/getvn.asp?v=21&n=4

[17] Rosenbaum, P., Rubin, D. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*. 39:33–38.

[18] Normand, S., Landrum, M., Guadagnoli, E., Ayanian, J., Ryan, T., Cleary, P., McNeil, B. 2001. Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology*. 54:387–398.

ADDITIONAL REFERENCES

Borghaei H., Paz-Ares L., Horn L., Spigel D., Steins M., Ready N., Chow L., Vokes E., Felip E., Holgado E., Bariesi F., Kohlhaufl M., Arrieta O., Burgio M., Fayette J., Lena H., Poddubskaya E., Gerber D., Gettinger S., Rudin C., Rizvi N., Crino L., Blumenschein G., Antonia S., Dorange C., Harbison C., Finckenstein F., Brahmer J. 2015. Nivolumab versus Docetaxel in Advanced Non-Squamous Non-small Cell Lung Cancer. *N Engl J Med*. 373(17): 1627-1639.

Lin D, Psaty B, Kronmal R. 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*. 54:948–963.

Liu, W., Kuramoto, S., Stuart, E.A. 2013. An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research. *Prevention Science* 14(6): 570-580.

Establishing a Framework to Evaluate Real-World Endpoints

July 10, 2018
Washington, DC

Introduction

Advances in data analytics and data capture through electronic health records (EHRs) and medical/pharmacy claims have brought the opportunities and challenges associated with using real-world evidence (RWE) to the forefront of the US healthcare industry. Increasingly, the promise of RWE to contribute to a more complete picture of the benefits and risks associated with therapies, when paired with results from randomized, controlled clinical trials, is being realized. RWE provides an opportunity to collect data rapidly on a broader patient population outside of a strict clinical trial protocol to help identify new indications or rare safety events, provide more generalizability of clinical trial results, and confirm clinical benefit in the post-market setting. Further, integration of the various sources of real-world data (RWD), including EHRs, clinical decision and support and hospital-based systems, administrative billing and claims databases, patient registries, longitudinal cohort studies, and patient reported outcomes tools, will yield a more robust dataset of RWE. However, the methods to aggregate data and the implications of integrating these multiple data sets as they evolve (especially in often dynamic post-approval settings) needs to be validated.

Applications for RWE extend the spectrum of therapeutics development from regulatory decision-making, to clinical use, to coverage and payment decisions. In the regulatory space, RWE has been utilized most frequently to evaluate drug safety through pharmacovigilance and adverse event monitoring in pre- and post-approval settings. However, RWE has increasingly been used to support effectiveness studies, in the form of historical data, as a surrogate for control arms in clinical trials (in the rare disease setting, for instance). Beyond regulatory decisions, RWE is frequently used to support clinical trial design, development of clinical practice guidelines, confirmation of population/subgroup size, and payment decisions including formulary placement.

These current applications of RWE in healthcare are quite limited with respect to the potential uses once appropriate standards and guardrails are implemented. Indeed, the pharmaceutical industry, FDA, and Congress recognize the importance of further developing this resource as evidenced by numerous recent publications by the FDA, passage of the 21st Century Cures Act (Cures Act), and the Prescription Drug User Fee Act (PDUFA) VI reauthorization. The Cures Act, passed in December of 2016, requires the FDA to develop a framework and issue guidance regarding the use of RWE to support a new indication for an already approved drug or post-market studies as a requirement for regulatory approval. Interestingly, FDA has already issued similar guidance regarding use of RWE for medical devices which includes supporting new indications and in post-approval studies. PDUFA VI builds upon the requirements of the Cures Act by instructing the FDA to consider stakeholder input through hosting of

public workshops as it develops its guidance for use of RWE. Other uses of RWE that could be imagined for future pharmaceutical approvals include expanded labels, pragmatic clinical trial design, and confirming benefit in the case of converting an Accelerated Approval to full approval status. In addition to potential regulatory uses, RWE could provide helpful information about the long-term value of a product and could inform future value assessments. For example, long-term efficacy endpoints that may not have been incorporated in pre-market clinical trial might be able to be captured using RWE, which requires increased understanding of how time-on-treatment or treatment discontinuation rates correlate to overall survival.

Significant progress has been made in data collection efforts to support use of RWE in regulatory settings, however challenges remain, chiefly with combining, organizing, and analyzing data from various information sources. Friends of Cancer Research proposes a pilot project, comprised of six leading healthcare data organizations, to develop a dataset curation process and validation framework to operationalize RWD collection and explore potential real-world endpoints that may be fit for regulatory purposes as well as assessing long-term benefits of a product.

Pilot Project Overview

Immunotherapies are being used to treat patients with cancers that have historically had few treatment options, which has generated high level of interest in their use and development. While immunotherapies have resulted in significant improvements in some patients, many other patients do not respond or only respond for a limited time. This has raised questions about the value of these new drugs. Applying current value frameworks to immune checkpoint inhibitors has proved difficult as they tend to underestimate the benefits of long-term survival and treatment-free survival.[1] This is likely due to the reliance on pivotal trial data, and in the setting of expedited approvals, assessments of the full clinical endpoints have not been completed. Thus, conclusions are often based on surrogate efficacy endpoints. At the initiation of this pilot project, three immune checkpoint inhibitors were approved for use in non-small cell lung cancer (NSCLC), which presented an opportunity to collect a robust amount of data for analysis from the post-market setting.

This pilot project was initiated to help determine whether RWD can be used to develop an early perspective on real-world outcomes, as defined by real-world endpoints from EHR and claims data, and whether these data correlate to overall survival (OS) in the context of randomized control trials (RCTs) for patients treated with novel therapies. The pilot project evaluates the performance of real-world endpoints across multiple data sets by focusing on a common question: ***What outcomes can be evaluated for aNSCLC patients treated with immune checkpoint inhibitors?***

To answer this question, a framework of necessary data elements, characteristics, and internal validation processes were proposed along with a set of definitions for real-world endpoints in the

[1] Ben-Aharon O, Magnezi R, Leshno M, Goldstein DA. Association of Immunotherapy with Durable Survival as Defined by Value Frameworks for Cancer Care. JAMA Oncol. 2018, 4(3):326–332.

context of their use in RCTs, FDA's regulatory framework, and data availability in EHR and claims systems. The pilot project will help evaluate whether the various data sets included in this study can achieve a similar level of correlation and statistical significance using a common framework.

Pilot Project Study Design and Objectives

This is a retrospective observational analysis of data derived from EHR and claims data. The data sets generated for the study include all relevant, retrospective patient-level data available for eligible individuals up to the data cutoff date, pending approval by a third-party de-identification.

Objective 1: Describe the demographic and clinical characteristics of aNSCLC patients treated with immune checkpoint inhibitors (Table 1)

Objective 2: Assess ability to generate real-world endpoints (OS, rwPFS, rwTTP, TTNT, TTD) in aNSCLC patients treated with immune checkpoint inhibitors, and segmented by clinical and demographic characteristics (Tables 2, 3, and 4)

Objective 3: Assess performance of real-world endpoints (rwPFS, rwTTP, TTNT, TTD) as surrogate endpoints for OS (Table 5)

Methods

Project Details	
Cohort and inclusion / exclusion criteria	aNSCLC patients treated with an immune checkpoint inhibitor (i.e., nivolumab, pembrolizumab, atezolizumab) Inclusion: • At least two documented clinical visits on or after January 1, 2011 until data cutoff date • Pathology consistent with NSCLC[2] • Has evidence of IIIB or IV NSCLC or has early stage NSCLC with a recurrence or progression described/documented in the EHR or claims • Treatment with immune checkpoint inhibitor, as documented by a structured medication order or claim as evidence of having received the treatment Exclusion: • Incomplete historical treatment data available within the database (i.e., patients whose advanced diagnosis date is more than 90 days before first activity date)

[2] For claims data, to minimize misclassification of aNSCLC, treatment with an IO agent following diagnosis of lung cancer was required. During the timeframe of this project, coverage for IO agents required evidence of advanced disease defined as either stage IIIB or IV NSCLC at initial diagnosis or early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression.

EHR and Claims-derived endpoints definition and analytical guidance	**Overall survival (OS)** • *Data definition / computation*: length of time from the date the patient initiates the study treatment to the date of death or proxied by time to disenrollment. Patients without a date of death will be censored at their last known activity or date of disenrollment from the health plan identified and categorized as "due to death" if the date of death captured by SSA DMF was within 30 days prior or 60 days following. **Time to Next Treatment (TTNT)** • *Data definition / computation:* length of time from the date the patient initiates the study treatment to the date the patient initiates their next systemic treatment. When subsequent treatment is not received (e.g., continuing current treatment or disenrollment not due to confirmed death), patients will be censored at their last known activity. • Start date of regimen immediately after PD-(L)1 line (i.e., the subsequent systemic therapy after the initial PD-(L)1-containing regimen) **Time to Treatment Discontinuation (TTD)** • *Data definition / computation*: length of time from the date the patient initiates the PD-(L)1 regimen to the date the patient discontinues treatment. Patients still on treatment will be censored at their last known activity. • Event Date: Date of PD-(L)1 regimen discontinuation defined as last administration or non-cancelled order of a drug contained within the PD-(L)1 line regimen (between the line's start and end date) among patients that discontinued their immune checkpoint inhibitor therapy. Permanent discontinuation is defined as meeting one of the following conditions: ○ Having a subsequent systemic therapy after the initial PD-(L)1-containing regimen ○ Having a date of death while on the PD-(L)1-containing regimen Having a gap of more than 120 days between the last administration or non-cancelled order of the PD-(L)1 line and the patient's last visit or medication administration if there is no other systemic therapy after the PD-(L)1-containing regimen • Censor date: Patients without a discontinuation will be censored at their last known PD-(L)1 usage defined as the last administration or non-cancelled order of a drug contained within the PD-(L)1 regimen **Progression Event** • *Data definition / computation*: distinct episode in which the treating clinician concludes that there has been growth or worsening in the aNSCLC. The progression event (and date) is based on review of the patient chart. **Real-world Progression Free Survival (rwPFS)** • *Data definition / computation*: length of time from the date the patient initiates the PD-(L)1 regimen to the date that a progression event as evident in the EHR is documented in the patient's chart or the patient passes away.

	Patients without a progression event or date of death will be censored at the end of the patient's chart. **Real-world Time to Progression (rwTTP)** • *Data definition / computation*: length of time from the date the patient initiated the PD-(L)1 regimen to the date that a progression event is documented in the patient's EHR (excludes death as an event). Patients without a progression event will be censored at the end of the patient's chart. • Event date: Patient's first progression date more than 14 days after PD-(L)1 initiation as described in the index date definition. Death will not be considered a progression event in TTP • Censor date: Patients without a progression date more than 14 days after the index date or date of death (for PFS) will be censored at the last date the patient could have been assessed for progression (e.g., last clinic note date) **Index Date** • *Data definition / computation*: the earliest PD-(L)1 inhibitor initiation in the advanced setting anchored to start (e.g., first administration or non-cancelled order) of the immune-checkpoint inhibitor-containing regimen (nivolumab, pembrolizumab, atezolizumab).
Analyses	<u>Table 1:</u> • Assess ability to identify aNSCLC patients treated with immune checkpoint inhibitors • Description of demographic and clinical characteristics of aNSCLC patients treated with immune checkpoint inhibitors, example characteristics include: ○ Demographic: gender, age, SES, region ○ Clinical: histology, smoking status, group stage at time of initial diagnosis, follow up, biomarker status (e.g., ALK, EGFR, PD-L1), hepatic and renal function • Description of population characteristics for overall population and by treatment setting / line of therapy (e.g., 1st line metastatic, 2nd line, 3rd line plus) <u>Table 2, Table 3, and Table 4:</u> • Assess ability to generate real-world endpoints (OS, TTNT, TTD) for aNSCLC patients treated with immune checkpoint inhibitors within the advanced treatment setting (range and median figures) • Evaluate these endpoints when patient cohort is segmented by treatment setting and demographic /clinical characteristics <u>Table 5</u>: • Assess correlation of real-world endpoints (TTNT, TTD) to overall survival (OS)

Summary of Data Sources for Pilot Project Study

Cancer Research Network

The Cancer Research Network originated as an NCI-funded consortium of research groups affiliated with a dozen integrated health care systems across the US, and among whom the Health Care Systems Research Network was formed. In the early 2000's, the CRN created the Virtual Data Warehouse (VDW), a common data model to facilitate collaborative research. Data in the VDW are extracted from multiple source databases and maintained by each research group with the possibility of pooling data under specific IRB-approved research protocols. For most participating institutions, the VDW has essentially complete information on care dating back to 1996 or earlier for most data domains. Domains include health plan enrollment periods, cancer registries, encounters including diagnoses and procedures, prescription and infusion medications, laboratory results, and other areas. The data provided are results from one of the participating CRN organizations.

Cota

The Cota Real-World Evidence (RWE) database is a HIPAA-compliant, de-identified data source drawn from the electronic health records (EHR) of contributing academic, for-profit, and community oncologist provider sites and hospital systems. The database includes detailed demographic, diagnostic, molecular and genomic testing, treatment, and outcome data. As of 2018, Cota's RWE is comprised of rich longitudinal patient records collected from over 40 unique locations across North America. For the purposes of this pilot study, patient data was sourced from a predominantly community setting (98%).

Flatiron Health

Flatiron Health is a longitudinal, demographically and geographically diverse database derived from electronic health record (EHR) data from over 265 cancer clinics (~800 sites of care) including more than 2 million active US cancer patients available for analysis. The patient-level data in the EHRs includes structured data (e.g., laboratory values, and prescribed drugs) in addition to unstructured data collected via technology-enabled chart abstraction from physician's notes and other unstructured documents (e.g., biomarker reports).

IQVIA™

IQVIA™ is a leading global provider of information, innovative technology solutions and contract research services focused on using data and science to help healthcare clients find better solutions for their patients. For this engagement, IQVIA provided data sourced through Oncology Electronic Medical Records (EMR) from multiple partners, including TransMed. The data are comprised predominately of community practices (90%+). The integrated EMR platform includes activity from all payer types and all practice sizes across the United States. Results for this analysis were calculated primarily based on structured EMR fields.

Mayo Clinic Analysis using OptumLabs® Data Warehouse

OptumLabs® is an open, collaborative research and innovation center founded in 2013 as a partnership between Optum and Mayo Clinic. Its core linked data assets include de-identified claims data for privately insured and Medicare Advantage enrollees and de-identified electronic health record (EHR)

data from a nationwide network of provider groups. This pilot project was a retrospective analysis of claims data from the OptumLabs® Data Warehouse (OLDW), which includes de-identified claims data for privately insured and Medicare Advantage enrollees in a large, private, U.S. health plan. The database contains longitudinal health information on enrollees, representing a diverse mixture of ages, ethnicities and geographical regions across the United States. The health plan provides comprehensive full insurance coverage for physician, hospital, and prescription drug services.

PCORnet Sites

This pilot project included 11 PCORnet partner sites who had previously participated in a PCORnet Rapid Cycle Project. The 11 sites are based in healthcare systems within three PCORnet networks across 10 US states and include 10 academic medical centers. These sites were selected from 80 PCORnet partner sites because they could rapidly provide tumor registry data and linked electronic health records in PCORnet Common Data Model (CDM) format. The pooled database contributed to the RWE Endpoints Pilot Project consisted of tumor registry data from each site and linked CDM diagnosis, procedures, prescribing, dispensing, medication administration, and death data tables. Data sources for the CDM include institutional billing and electronic health record data. The study cohort includes patients with a single primary advanced stage non-small cell lung cancer (NSCLC) diagnosis who were either diagnosed at stage 3b or 4 or who had an ICD9/10 diagnosis code for secondary metastasis.

Table 1. Description of demographic and clinical characteristics of aNSCLC patients treated with PD-(L)1 checkpoint inhibitors

Demographics	Data Set A PD-(L)1-treated N=2595	Data Set B PD-(L)1-treated N=557	Data Set C PD-(L)1-treated N=435	Data Set D PD-(L)1-treated N=6924	Data Set E PD-(L)1-treated N=2860	Data Set F PD-(L)1-treated N=269
Age at advanced diagnosis (years), median [IQR]	68 [15]	64 [14]	66 [14]	69 [14]	68 [14]	70 [14]
Age at PD-(L)1 inhibitor initiation (years), median [IQR]	69 [14]	65 [14]	68 [14]	69 [14]	69 [14]	71 [14]
Age categories at PD-(L)1 inhibitor initiation (categorical):						
≤49 years	120 (5%)	24 (4%)	21 (5%)	219 (3%)	80 (3%)	8 (3%)
50-64 years	888 (34%)	251 (45%)	129 (30%)	2048 (30%)	863 (30%)	65 (24%)
65-74 years	866 (33%)	198 (36%)	169 (39%)	2504 (36%)	1047 (37%)	94 (35%)
75+ years	721 (28%)	84 (15%)	116 (27%)	2153 (31%)	870 (30%)	102 (38%)
Age categories at PD-(L)1 inhibitor initiation (binary):						
<75 years	1874 (72%)	473 (85%)	319 (73%)	4771 (69%)	1990 (70%)	167 (62%)
75+ years	721 (28%)	84 (15%)	116 (27%)	2153 (31%)	870 (30%)	102 (38)
Gender:						
Female	1147 (44%)	276 (50%)	212 (49%)	3172 (46%)	1351 (47%)	125 (46%)
Male	1448 (56%)	281 (50%)	222 (51%)	3752 (54%)	1509 (53%)	143 (53%)
Unknown/Missing	0	0	≤5	0	0	1
Race/ethnicity:						
White	1704 (78%)	478 (86%)	284 (65%)	4969 (79%)	676 (87%)	160 (87%)
Black or African American	282 (13%)	67 (12%)	37 (9%)	594 (9%)	44 (6%)	14 (8%)
Asian	52 (2%)	6 (1%)	83 (19%)	155 (3%)	13 (2%)	9 (5%)
Other Race	142 (7%)	6 (1%)	31 (7%)	580 (9%)	42 (5%)	1 (1%)
Unknown/Missing	415	0	0	626	2085	85
Median household income (zip-level):						
1 (lowest median household income)			103 (24%)	1003 (15%)		
2			105 (24%)	1539 (22%)		
3			114 (26%)	1833 (27%)		
4 (highest median household income)			112 (26%)	2525 (37%)		
Unknown			≤5	24		
CLINICAL CHARACTERISTICS						
Group stage at initial diagnosis:						
Stage 0 / Occult		0		2 (0%)		
Stage I		23 (6%)		496 (7%)		18 (7%)
Stage II		22 (6%)		426 (6%)		17 (7%)
Stage III		88 (23%)	39 (9%)	1494 (22%)		17 (7%)
Stage IV		248 (65%)	396 (91%)	4335 (64%)		161 (62%)
Group stage is not reported		176		171		10
Histology:						
Non-squamous cell carcinoma		370 (66%)	320 (74%)	4679 (70%)	1981 (69%)	194 (73%)
Squamous cell carcinoma		147 (26%)	73 (17%)	1983 (30%)	659 (23%)	61 (23%)

NSCLC histology not otherwise specified (NOS)		40 (7%)	42 (10%)	262 (3%)	220 (8%)	10 (4%)
Missing						4
Smoking status:						
History of smoking			340 (78%)	6185 (90%)	448 (92%)	182 (87%)
No history of smoking			94 (22%)	717 (10%)	38 (8%)	28 (13%)
Unknown/Not documented			≤5	22	2374	210
PD-L1 tested on or prior to PD-(L)1 inhibitor start	326 (13%)			2384 (34%)	96	80/96 (83%)
PD-L1 expression status (among those tested):						
PD-L1 positive				512 (22%)	45 (50%)	65 (68%)
PD-L1 negative/not detected				691 (29%)	45 (50%)	29 (30%)
Unsuccessful/indeterminate test				1012 (42%)	0	2 (2%)
Results pending/unknown				169 (7%)	6	173
ALK tested on or prior to PD-(L)1 inhibitor start	258 (10%)			4513 (65%)	582	143/173 (83%)
ALK status (among those tested):						
Rearrangement present				57 (1%)	8 (1%)	1 (1%)
Rearrangement not present				4145 (92%)	570 (99%)	170 (98%)
Results pending/unknown				68 (2%)	0	2 (1%)
Unsuccessful/indeterminate test				243 (5%)	4	96
EGFR tested on or prior to PD-(L)1 inhibitor start	543 (21%)		171(39%)	4684 (68%)	953	115/142 (81%)
EGFR status (among those tested)[2,3]:						
Mutation positive				305 (7%)	68 (11%)	6/142 (4%)
Mutation negative				4161 (89%)	525 (89%)	135/142 (95%)
Results pending/unknown				60 (1%)	358	1/142 (1%)
Unsuccessful/indeterminate test				158 (3%)	2	127
No prior therapy received	690 (27%)		80 (18%)	2074 (30%)	777 (27%)	77 (29%)
Line number of first PD-(L)1 inhibitor in advanced setting:						
1	690 (27%)		80 (18%)	2074 (30%)	777 (27%)	77 (29%)
2	1440 (56%)		205 (47%)	3357 (49%)	1414 (49%)	87 (32%)
3	380 (15%)		85 (20%)	1012 (15%)	448 (16%)	51 (19%)
4+	85 (3%)		65 (15%)	481 (7%)	221 (8%)	54 (20%)
Patients receiving a second PD-(L)1 inhibitor in a subsequent line:						
No			402 (92%)	1740 (25%)		
No subsequent therapy received				4879 (71%)		
Yes	93 (4%)		33 (8%)	305 (4%)	112	14
Line number of second PD-(L)1 inhibitor in advanced setting:						
2	28 (30%)		11 (33%)	99 (33%)	9 (8%)	5 (36%)
3	45 (48%)		10 (30%)	134 (44%)	51 (46%)	4 (29%)
4+	20 (22%)		12 (36%)	72 (24%)	52 (46%)	5 (36%)
N/A			402			
Time from advanced diagnosis to first PD-(L)1 inhibitor initiation (months), median [IQR]		7 [11]	8 [11]	6 [11]	8 [14]	7 [12]

Structured follow up time[3]					
Structured follow-up time from advanced diagnosis (months), median [IQR]	18 [18]	18 [21]	14 [17]	18 [20]	18 [18]
Structured follow-up time from PD-(L)1 inhibitor initiation (months), median [IQR]	8 [13]	9 [13]	6 [10]	8 [11]	8 [9]

[3] Structured follow-up time is calculated from the relevant time-point for each patient until their last structured activity (i.e., most recent visit or administration)

Table 2. Median time and 95% confidence interval for real-world extracted endpoints

Data Set	rwOS	rwTTNT	rwTTD	rwTTP	rwPFS
Data Set A	13.50 [12.80, 14.50][4]	22.50 [NA]	7.03 [6.27, 9.97]		
Data Set B	15.78 [12.2, 24.59]; 8.58 [7.56, 10.26][5]		3.25[2.76, 3.75]		
Data Set C	8.67 [6.83, 10.02]	11.60 [8.80, 16.10]	4.70 [3.68, 5.52]		
Data Set D	9.15 [8.82, 9.51]	14.03 [12.89, 15.15]	3.21 [3.21, 3.44]	5.41 [5.18, 5.67]	3.28 [3.18, 3.41]
Data Set E	12.69 [11.7, 13.87]	12.07 [11.24, 13.48]	3.63 [3.40, 3.87]		
Data Set F	12.30 [9.61, 16.94]	12.50 [9.29, NA]	4.60 [3.71, 6.32]	9.37 [7.42, 11.93]	9.37 [7.42, 11.93]

Data sets measured median time for real-world extracted endpoints utilizing a common definition as described in the pilot project methods

Table 3. One-year real-world overall survival landmark analysis post PD-(L)1 initiation

Data Set	One-year rwOS Landmark Analysis
Data Set A	0.57 [0.52, 0.57][4]
Data Set B	0.54 [0.48, 0.57]; 0.41 [0.34, 0.47][5]
Data Set C	0.40 [0.35, 0.46]
Data Set D	0.42 [0.41, 0.43]
Data Set E	0.51 [0.49, 0.53]
Data Set F	0.40 [0.34, 0.48]

[4] OS was calculated as days between I/O initiation and disenrollment.

[5] Sites with social security or state death data, censored at estimated earliest date such data should be available if no death was observed

Table 4. Median times and 95% confidence interval (indexed to initial PD-(L)1 inhibitor line start in advanced setting) segmented by treatment setting and demographic characteristics as described in Table 1

Demographics		Data Set A		Data Set B		Data Set C		Data Set D		Data Set E		Data Set F
	N	rwTTD (Months) Median [95% CI]	N	rwTTD (Months) Median [95% CI]	N	rwTTD (Months) Median [95% CI]	N	rwTTD (Months) Median [95% CI]	N	rwTTD (Months) Median [95% CI]	N	rwTTD (Months) Median [95% CI]
		rwOS (Months) Median [95% CI][6]	N	rwOS (Months) Median [95% CI] [7]		rwOS (Months) Median [95% CI]		rwOS (Months) Median [95% CI]		rwOS (Months) Median [95% CI]		rwOS (Months) Median [95% CI]
Age categories at PD-(L)1 inhibitor initiation:												
≤49 years	100	5.87 [3.97, 9.80]	24	3.97 [1.84, 17.03]	21	6.44 [1.28, 16.29]	219	2.89 [2.30, 3.64]	80	2.33 [1.43, 4.40]	8	8.52 [2.77, NA)
		18.1 [11.87, 21.63]	16	9.07 [2.66, NA]		12.02 [4.27, NA)		9.28 [7.77, 12.07]		10.20 [7.96, 17.36]		NA
50-64 years	723	6.23 [5.20, 7.27]	251	3.62 [2.96, 4.14]	129	5.35 [3.35, 9.23]	2047	3.21 [2.92, 3.44]	863	3.53 [3.17, 4.03]	65	6.16 [3.71, 12.26]
		13.60 [11.83,14.6]	163	8.78 [6.81, 11.41]		9.33 [6.73, 13.27]		9.34 [8.43, 10.26]		13.84 [11.80, 15.32]		16.94 [9.37, NA]
65-74 years	728	7.40 [6.10, 9.60]	198	2.76 [2.27, 3.75]	169	4.63 [3.45, 6.44]	2504	3.41 [3.21, 3.67]	1047	3.77 [3.30, 4.23]	94	5.38 [3.42, 6.90]
		13.40[12.07,14.93]	135	8.52 [6.54, 9.70]		8.97 [5.78, 11.14]		9.34 [8.79, 10.26]		12.16 [10.49, 14.33]		12.30 [7.55, 21.58]
75+ years	593	7.70 [6.37, 9.37]	84	2.47 [1.45, 4.64]	116	3.57 [1.84, 5.26]	2153	3.25 [3.18, 3.61]	870	3.77 [3.30, 4.23]	102	3.67 [2.82, 5.74]
		13.22 [11.83,14.61]	49	10.65 [5.03, NA]		6.83 [4.24, 9.13]		8.79 [8.23, 9.28]		13.02 [10.62, 14.79]		10.00 [8.71, 15.55]
Gender:												
Female	950	6.80 [5.90, 8.23]	276	3.53 [2.76, 3.98]	212	4.76 [3.45, 7.72]	3751	2.98 [2.75, 3.21]	1351	3.83 [3.53, 4.27]	125	5.03 [3.77,7.03]
		13.70 [12.8, 15.23]	188	8.94 [6.67, 12.33]		9.33 [7.42, 13.44]		8.43 [7.93, 8.98]		14.79 [13.15, 16.87]		13.23 [8.77, 23.33]
Male	1194	7.23 [6.10, 8.43]	281	3.16 [2.37, 3.81]	222	4.62 [3.35, 5.52]	3172	3.51 [3.21, 3.70]	1509	3.30 [2.87, 3.77]	143	3.87 [2.93, 6.40]
		13.20 [12.13, 14.5]	175	7.92 [6.77, 10.06]		7.49 [6.34, 10.02]		9.84 [9.38, 10.72]		11.08 [10.16, 12.39]		11.87 [8.40, 20.40]
CLINICAL CHARACTERISTICS												
Group stage at initial diagnosis:												
Stage 0 / 1			23	3.45 [0.92, 5.69]			498	4.36 [3.67, 5.28]			18	5.18 [3.43, 15.65]
			9	5.69 [0.49, 9.70]				12.07 [10.69, 14.03]				7.03 [4.87, NA]

[6] rwOS was calculated as time between I/O initiation and disenrollment

[7] rwOS estimates include sites with social security or state death data available; excluded are sites with only local/EHR death data available

Stage II	22	3.68 [1.41, 6.38]			426	3.90 [3.28, 4.95]			17	4.10 [2.80, NA]
	13	6.38 [1.48, NA]				11.84 [10.59, 13.28]				13.90 [4.03, NA]
Stage III	88	4.14 [2.76, 3.88]	39	4.43 [1.38, 10.35]	1494	3.67 [3.44, 4.13]			63	5.73 [2.90, 9.67]
	62	9.60 [6.67, NA]		8.97 [2.73, 13.44]		9.84 [9.18, 10.79]				14.87 [9.37, 22.10]
Stage IV	248	3.35 [2.76, 3.88]	396	4.70 [3.68, 5.52]	4334	2.89 [2.75, 3.18]			161	4.32 [3.19, 6.47]
	187	8.77 [6.77, 10.85]		8.67 [6.83, 10.02]		8.26 [7.80, 8.79]				12.10 [8.30, 20.73]
Unknown	176	5.79 [1.87, 3.85]							10	8.83 [3.74, NA]
	92	7.92 [6.15, 10.55]								NA
Histology:										
Non-squamous cell carcinoma	370	3.16 [2.53, 3.75]	320	4.76 [3.81, 5.81]	4678	3.34 [3.21, 3.51]	1981	3.60 [3.30, 3.90]	194	4.15 [3.23, 5.97]
	240	9.69 [7.56, 12.33]		8.67 [6.7, 10.18]		9.61 [9.11, 10.30]		14.14 [12.69, 15.81]		11.87 [8.77, 21.80]
Squamous cell carcinoma	147	3.25 [2.47, 4.01]	73	4.14 [1.68, 7.33]	1983	3.21 [2.98, 3.54]	659	3.77 [3.30, 4.30]	61	7.55 [3.67, 12.97]
	93	6.80 [4.87, 8.78]		8.38 [4.8, 12.06]		8.66 [7.84, 9.25]		10.36 [9.24, 11.77]		13.90 [8.57, NA]
NSCLC histology not otherwise specified (NOS)	40	3.88 [1.94, 5.10]	42	3.80 [2.07, 9.13]			220	3.57 [2.83, 5.17]	10	5.07 [2.29, NA]
	30	10.26 [5.13, 13.32]		7.92 [3.58, 17.15]				11.84 [7.89, 16.08]		20.40 [6.13, NA]
Smoking status:										
History of smoking			340	4.53 [3.45, 5.35]	6185	3.28 [3.21, 3.48]	448	5.17 [3.90, 6.53]	182	4.27 [3.43, 6.19]
				8.67 [6.7, 10.18]		9.21 [8.85. 9.64]		19.17 [14.30, 24.46]		11.43 [8.40, 20.73]
No history of smoking			94	5.52 [3.22, 9.20]	716	2.75 [2.49, 3.11]	38	3.30 [2.37, 9.07]	28	6.08 [3.23, 8.71]
				8.34 [6.04, 12.88]		8.69 [8.03, 9.77]		14.50 [4.57, NA]		12.10 [8.71, NA]
Unknown/Not documented							2374	3.53 [3.30, 3.77]	59	5.97 [2.77, 10.00]
								12.00 [10.82, 13.08]		13.21 [9.37, NA]
PD-L1 expression status (among those tested):										
PD-L1 positive					512	4.10 [3.38, 4.82]	45	5.63 [2.83, 18.23]	65	3.68 [2.82, 5.97]
						10.79 [9.05, 13.28]		NA		9.63 [7.55, NA]
PD-L1 negative/not detected					690	2.75 [2.49, 3.11]	45	9.63 [NA]	29	6.32 [4.19, NA]
						8.69 [7.48, 9.84]		NA		20.80 [11.43, NA]
Line number of first PD-(L)1 inhibitor in advanced setting:										
1	592	9.10 [7.97, 12.40]	80	5.26 [2.63, 6.64]	2074	3.90 [3.67, 4.23]	777	5.90 [4.93, 6.80]	77	5.03 [3.67, 8.83]
		19.83 [17.23, 22.23]		9.17 [5.68, 17.05]		10.36 [9.48, 11.18]		20.78 [14.79, 25.12]		15.87 [9.87, NA]

2	1174	6.57 [6.03, 7.50] 11.7 [10.97, 12.83]	205	4.53 [3.22, 5.81] 7.75 [5.75, 10.28]	3357	3.21 [2.82, 3.21] 8.66 [8.13, 9.15]	1414	3.00 [2.83, 3.30] 10.68 [9.83, 11.93]	87	4.81 [2.74, 7.00] 9.07 [7.39, 20.40]
3	304	4.47 [3.8, 6.1] 12.8 [10.7, 14.67]	85	4.63 [2.30, 8.05] 9.33 [6.04, 12.02]	1011	2.98 [2.75, 3.44] 9.02 [7.80, 9.97]	448	3.53 [3.00, 4.23] 10.72 [9.04, 13.87]	51	5.67 [3.43, 8.71] 15.29 [9.63, NA]
4+	74	3.83 [2.83, 5.47] 14.2 [10.1, 17.17]	65	4.76 [3.94, 11.10] 8.67 [5.26, 13.27]	481	2.59 [2.30, 3.18] 8.52 [6.89, 10.46]	221	3.30 [2.60, 4.23] 12.00 [8.25, 15.58]	54	3.47 [2.77, 7.16] 10.43 [6.90, 21.80]

Table 5. Correlation between real-world overall survival and real-world extracted endpoints using Spearman's rank correlation coefficient

Data Set	Comparison	N	Correlation (95% CI)
Data Set A	rwOS vs rwTTNT	83	0.36
	rwOS vs rwTTD	254	0.63
Data Set B	rwOS vs rwTTNT		
	rwOS vs rwTTD	225	0.62 (0.54, 0.69)
Data Set C	rwOS vs rwTTNT	96	0.70 (0.58, 0.79)
	rwOS vs rwTTD	295	0.89 (0.86, 0.91)
Data Set D	rwOS vs rwTTNT	1203	0.61 (0.57, 0.64)
	rwOS vs rwTTD	4337	0.80 (0.79, 0.81)
	rwOS vs rwPFS	4337	0.75 (0.74, 0.76)
	rwOS vs rwTTP	2286	0.60 (0.57, 0.63)
Data Set E	rwOS vs rwTTNT	358	0.62 (0.54, 0.68)
	rwOS vs rwTTD	1456	0.77 (0.75, 0.79)
Data Set F	rwOS vs rwTTNT	39	0.46 (0.33, 0.81)
	rwOS vs rwTTD	142	0.80 (0.66, 0.85)
	rwOS vs rwPFS	142	0.84 (0.62, 0.86)
	rwOS vs rwTTP	55	0.56 (0.21, 0.71)

The correlation analysis is restricted to patients with a death date and documented event as described in the definitions and algorithms

Conclusions from Pilot Project Study

1. There is a high level of shared characteristics among the varying data sets despite varying sample sizes, data capture processes, and data sources demonstrating the feasibility of identifying aNSCLC patients treated with immune checkpoint inhibitors from diverse RWD sources.
2. The pilot project demonstrated that several extractable endpoints from EHR and claims data correlate with OS. Further validation is required to determine whether these endpoints are reliable surrogates for OS outside of a traditional clinical trial and whether they can support regulatory and payer decision-making.
3. Survival among patients as assessed through EHR and claims data fall within the range of median OS values observed in several immune checkpoint inhibitor trials.[8]
4. Assessment of extracted endpoints from EHR and claims data demonstrate that efficacy of immune checkpoint inhibitors is relatively consistent across a variety of patient characteristics, such as age and sex.

Assumptions and Limitations of Pilot Project Data Sets

- Ability to collect reliable data will vary across data providers
- Approaches to analysis may vary even when using a common protocol; A careful review and collaboration is needed to align on a consistent and reliable approach
- Verified diagnosis and diagnosis date, clinical stage and cell type, planned chemotherapy regimen (dose and schedule) and other clinical and socioeconomic factors cannot always be determined from the available EHR and claims data
- Verifying and determining date of death may also prove challenging. Although discharge status and some diagnosis codes may be a source of mortality information, but some data partners rely on linkage to the public SSA death master file (DMF). The public DMF has been shown to under identify deaths[9]
- For claims-based data, some patients with advanced disease may enroll in clinical trials and some or all the care received in a clinical trial setting may not generate insurance claims, thus, data for these patients may not be fully captured or captured at all
- Approaches to the analyses may vary even when using a common protocol and careful review and collaboration is needed to align on a consistent and reliable approach
- Some biomarkers may not routinely be assessed in the real-world setting, but more would have been included in this analysis if a chart review had been conducted or the use of natural language processing (NLP)
- Provider data (EHR) may not identify all chemotherapy as patients may seek care inside and outside a provider group that contributes to the EHR data (e.g., chemotherapy at an academic center then move to a community setting). This may or may not be a source of missing information in the advance NSCLC setting

[8] Huang G, Sun X, Liu D, et al. The efficacy and safety of anti-PD-1/PD-L1 antibody therapy versus docetaxel for pretreated advanced NSCLC: a meta-analysis. Oncotarget, 4239-4248

[9] Jones B, V. D. (2015, March). Measuring Mortality Information in Clinical Data Warehouses. AMIA Jt Summits Transl Sci Proc, 450-5

Discussion Questions

These questions may help guide the discussion during the meeting:

1. Are there processes to handle challenges associated with the availability and consistency of data across provider types and settings?

2. How to overcome difficulties associated with determining events like death?

3. What opportunities or incentives exist to help improve the format, quality, and validity of RWE?

4. Are there lessons from clinical trials, or registration trials, that need to be considered for RW data?

5. What opportunities exist for FDA decision-making to be supported by RWE?

6. What opportunities exist to expand to other endpoints such as patient reported outcomes (PROs) and patient-generated health data?

7. Are there other extractable endpoints for EHR- or claims-based algorithms that should be validated?

8. What is the role and use of real-world endpoints, such as TTD, TTNT, or PFS, for payer decision-making, particularly in the context of accelerated approval or breakthrough therapy designation?

9. How important is RWE in the development of new payment designs, such as value-based payment, risk-sharing arrangements, and outcomes-based agreements?

10. How timely does the data have to be for regulatory or reimbursement? How quickly must the data be analyzed/reported?

11. For reimbursement/value-based payment/risk sharing, are data from all data sets (A-F) available to payers? Manufacturers?

FRIENDS of CANCER RESEARCH

A Friends of Cancer Research White Paper

EXPLORING THE USE OF CIRCULATING TUMOR DNA AS A MONITORING TOOL FOR DRUG DEVELOPMENT

CONTRIBUTORS

Hesham Abdullah
AstraZeneca

Jeff Allen
Friends of Cancer Research

J. Carl Barrett
AstraZeneca

Julia Beaver
U.S. Food and Drug Administration

Gideon Blumenthal
U.S. Food and Drug Administration

Darya Chudova
Guardant Health

Leena Das-Young
Guardant Health

Bruno Gomes
Genentech, A Member of the Roche Group

Jamie Holloway
Patient Advocate

Diana Merino
Friends of Cancer Research

Geoffrey Oxnard
Dana Farber Cancer Institute

Reena Philip
U.S. Food and Drug Administration

David Shames
Genentech, A Member of the Roche Group

Jean-Charles Soria
MedImmune

Mark Stewart
Friends of Cancer Research

INTRODUCTION

*Cell-free DNA (cfDNA)**, defined as free extra-nucleic acid circulating in plasma, was first described in the blood of healthy and diseased individuals in 1948.[1] Most cfDNA in blood is derived from ruptured nonmalignant cells arising from normal physiological tissue remodeling events and originates from the germline. However, in patients with cancer, a fraction of this cfDNA is made up of nucleic acids that are shed from primary or metastatic lesions undergoing tumor cell apoptosis and necrosis and are referred to as *circulating tumor DNA* (ctDNA). ctDNA is composed of small fragments of nucleic acid that are not associated with cells or cell fragments, thus differentiating it from circulating tumor cells (CTCs).

The greatest proportion of DNA fragments in circulation measure between 180-200 nucleotides in size, suggesting they are a result of cellular apoptosis; however, much smaller fragments have been reported in some tumor types, such as hepatocellular carcinoma, as well as much larger fragments consisting of thousands of base pairs that may be a result of tumor necrosis.[2] The amount of ctDNA in circulation is very small ranging between <0.1-10% of total cfDNA detectable in human blood. This value varies according to tumor burden or size, inflammatory status, cellular turnover, and proximity of cancer cells to blood vessels.[3]

The ability to detect small amounts of ctDNA in fluids has given rise to the use of liquid biopsies, a minimally invasive test done on a blood sample, or other fluids, that provide an alternative to surgical biopsies of solid tissues.[4] The recent development of large-scale genomics and bioinformatics approaches has facilitated the use of highly sensitive molecular assays that can detect tumor-specific alterations present in at least 5% of the cells

**Terms in italics are defined at the end of this documentin the "List of Definitions."*

analyzed and at frequencies as low as 0.05%.[5] Classical methods for ctDNA analysis include hotspot assays that detect specific known somatic variants at very low levels found in a single gene or small number of genes, and typically use polymerase chain reaction (PCR)-based strategies such as *digital droplet PCR (ddPCR),* or *real-time PCR (RT-PCR)*. More recent strategies use *next-generation sequencing (NGS)* approaches for the detection of somatic and germline (heritable) variants in more than one gene target and are capable of detecting a larger number of variants in multiple genes. Deep sequencing can typically detect tumor-specific alterations in the whole genome or exome, and most recently, in gene panels that have been especially designed to incorporate relevant genes associated to cancer growth and progression. The most common ctDNA genomic alterations identified include point mutations, deletions, amplifications and translocations, and gene fusions.[5,6] Measuring these alterations in the ctDNA isolated from cancer patients' blood can considerably facilitate the clinical management of patients diagnosed with blood cancers and solid tumors. Although assessing disease burden using blood samples is already a common practice for patients with blood malignancies, investigating blood for traces of solid tumor cells and DNA is a more recent practice, and has the potential to facilitate clinical cancer care and benefit more patients.

First, because drawing blood for a *liquid biopsy* is minimally invasive and significantly less risky than conducting a tissue biopsy, especially for tumors that are not easily accessible, using *ctDNA assays* to conduct repeated assessments that monitor a patient's tumor response over time poses less risk to the patient. Moreover, because the test can be conducted at a central site, patients don't need access to technical molecular pathology labs, which are rarely found in the community setting. Analysis of ctDNA is more convenient and logistically feasible than traditional biopsies, and it is a tool that can democratize access to powerful diagnostics and targeted therapies regardless of where a patient receives care.

The assessment of ctDNA requires powerful technology that is highly sensitive and dynamic and enables the detection of very small amounts of tumor DNA from very early to well advanced stages of disease. Additionally, the multiplex assays used in ctDNA analyses capture a broad array of somatic, or tumor-derived, genetic alterations found in numerous genes from ctDNA in blood *(i.e., genotyping)*.

Leveraging the advent of new technologies with these remarkable features, ctDNA may be used in a way that goes beyond simply identifying the presence or absence of tumor DNA in blood but could be potentially used for (1) *cancer detection*, (2) prognosis determination, and (3) molecular characterization of a patient's tumor.

As a powerful cancer screening tool, ctDNA could be used for early cancer detection. This could mean detecting cancer prior to a cancer diagnosis in an asymptomatic population, or detecting early recurrence, or the degree or burden of disease in patients that have already been diagnosed with cancer.[5,7] ctDNA could also help assess patient prognosis. This could mean catego-

rizing patients into different risk groups by examining the presence of specific somatic genetic alterations associated to patient outcomes.[2, 8-11] Lastly, identifying somatic genetic alterations in ctDNA would enable the molecular characterization of the tumor, which could guide targeted therapy selection and identify potential mechanisms of tumor resistance.[12-13]

Although using ctDNA for cancer detection, prognosis determination, and molecular characterization of a patient's tumor are very important and becoming more common in clinical practice, this white paper will focus on recommending best practices for the use of ctDNA for disease *monitoring* in cancer patients and will investigate the feasibility of operationalizing this tool in drug development. Additionally, even though much effort has been given to the definition and study of *minimum residual disease* as a way to *monitor disease response* and progression in patients with blood cancers, this white paper will concentrate on the use of ctDNA in patients with solid tumors.

Given the rapid advancement of technologies that have promoted the use of ctDNA in drug development and the growing number of studies that seek to use liquid biopsies as a tool to assess tumor response; Friends of Cancer Research *(Friends)* has convened a multi-stakeholder group of experts to examine the state of ctDNA in tumor monitoring, recommend best practices, and propose initiatives that would directly demonstrate how data derived from ctDNA could be used to facilitate cancer drug development.

OBJECTIVES

The objectives of this white paper are to assess the current state of ctDNA as a monitoring tool used to evaluate clinical response through the description of relevant case studies, suggest best practices for the use of ctDNA as a potential monitoring tool for drug development in clinical research, and propose two potential opportunities that promote the operationalization of ctDNA in drug development.

CASE STUDIES

Various studies have investigated the use of cfDNA or ctDNA to monitor tumor response. Many of these studies have been retrospective using previously collected data and consisting of a few samples. The working group identified three prospective clinical trials where serial analyses of cfDNA was used to gain insight into treatment effect **(Table 1)**. These prospective studies demonstrate the diverse ways investigators are using cfDNA to monitor clinical outcomes, highlighting the promising potential of this accessible biomarker in clinical trials, but also unraveling the difficulties that lie in seeking to compare data from all three studies given the different methods, units, and outcomes assessed in each study.

1. Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy, 2015 [14]

Mok and colleagues describe the findings of a multicenter, randomized, placebo-controlled, double-blind, phase III study of intercalated erlotinib or placebo with gemcitabine plus platinum followed by maintenance erlotinib or placebo as first-line treatment in patients with stage IIIB/IV NSCLC (FASTACT-2). The primary objective of this study was to define the diagnostic utility of a RT-PCR based blood test that detects activating mutations in *EGFR* in cfDNA. The secondary objective was to examine the predictive value of cfDNA *EGFR* at baseline and the changes in mutation status during therapy in relation to patient outcomes. This study found very high concordance between tissue and blood tests, that *EGFR* mutation status defined by blood-based cfDNA analysis appears to produce similar results to tissue-based assessment in terms of predicting outcomes, and that dynamic changes in cfDNA *EGFR* mutation status correlate with disease progression, ORR, and survival.

Blood from 305 patients was extracted according to standard procedures at baseline, at day 1 of cycle 3 (C3, mid-protocol), and at the time of progression, while tumor tissue samples were obtained at initial diagnosis, diagnosis of advanced disease, or biopsy 14 days before first study dose. The cobas 4800 blood test by Roche Molecular Systems Inc. was used to detect 41 different *EGFR* activating mutations. The number of *EGFR* mutant copies (copy/mL of blood) were measured across the three timepoints (baseline, C3, and PD) and correlated with ORR, PFS, and OS.

This study found that generally, total *EGFR* mutation-specific cfDNA levels decreased at C3 and returned at time of PD, which may reflect changes in tumor volume or increased metastases. For patients with detectable *EGFR* mutations at baseline, ORR was lower in patients whose cfDNA analysis showed detectable *EGFR* mutations at C3 (mid-protocol) compared

Table 1: Case studies and study parameters

Parameters/Study	Mok et al., Clinical Cancer Research, 2015[14]	Yu et al., Clinical Cancer Research, 2017[15]	Raja et al., Clinical Cancer Research, 2018[16]
Histology	Stage IIIB and IV NSCLC	Advanced NSCLC patients with disease progression after EGFR TKI treatment	NSCLC and UC
# of patients	305	93	100 (28 discovery, 72 validation) and 29 (validation) from 2 different studies
Clinical trial	FASTACT-2 study	NCT02113813	ATLANTIC and Study 1108
ctDNA/cfDNA	cfDNA	cfDNA	ctDNA
Technology	Semi-quantitative—Cobas 4800 blood test (RT-PCR)	Quantitative—BEAMing PCR	Quantitative—NGS, targeted panel (Guardant 360)
Gene	EGFR	EGFR	Gene panel (73 genes)
Units	Copy/mL	% mutant EGFR cfDNA	Mean VAF
Timepoints	Baseline, cycle 3 (~12 weeks) and progression (PD)	Baseline, cycle 2	Baseline and 6 weeks-prior to 4th treatment
Median follow up time	Not specified	Not specified	Ranged between 9-15 months depending on study
Drug(s) being tested	Erlotinib (after gemcitabine/platinum)	ASP8273 (3rd generation EGFR TKI)	Durvalumab (anti PD-L1)
Clinical Response/ Outcome	ORR, PFS, OS	ORR	Tumor volume, PFS, OS
Tube	"collected according to standard procedures"	n/a	K2-EDTA
Timing of processing	"collected according to standard procedures"	n/a	n/a

Abbreviations: cfDNA, cell-free DNA; ctDNA, circulating tumor DNA; EDTA, ethylenediaminetetraacetic acid; *EGFR*, epidermal growth factor receptor; NGS, next generation sequencing; NSCLC, non-small cell lung cancer; ORR, objective response rate; OS, overall survival; PD, progressive disease; PD-L1, programmed death-ligand 1; PFS, progression free survival; RT-PCR, real time- polymerase chain reaction; TKI, tyrosine kinase inhibitor; UC, urothelial carcinoma; VAF, variant allele fraction.

with patients whose cfDNA analysis showed undetectable *EGFR* mutations at C3. Likewise, the PFS and OS of patients whose cfDNA samples remained positive for *EGFR* mutations at mid-protocol were also lower than in patients whose cfDNA samples became negative for *EGFR* mutations.

Authors concluded that assessing *EGFR* mutation status mid-protocol, in this case at C3, approximately 12 weeks after the start of the first study dose, may predict clinical outcomes and that the serial quantitative measurement of *EGFR* cfDNA could serve to assess tumor progression. Moreover, because of the good correlation between tumor and blood tests, the authors identified cfDNA *EGFR* mutation analysis as a potential reliable alternative method for patients from whom a tumor tissue sample cannot be obtained.

2. A phase 1, dose-escalation/response-expansion study of oral ASP8273 in patients with non-1 small cell lung cancers with epidermal growth factor receptor mutations, 2017[15]

Yu and colleagues describe the results of a prospective, open-label, multicenter dose escalation phase I study (NCT02113813) testing the third-generation *EGFR* TKI, ASP8273 in patients with advanced NSCLC harboring *EGFR* activating mutations and previous *EGFR* TKI treatment. Exploratory endpoints of this study included the evaluation of potential biomarkers in cfDNA and their association with treatment effects. This study found for patients who achieved partial response and stable disease as best overall response, *EGFR* activating and T790M mutations in cfDNA were generally reduced to near or below level of detection after 1 cycle of treatment. Additionally, in patients who developed acquired resistance to ASP8273, *EGFR* activating and T790M mutations reemerged in the plasma of 5 out of 9 patients.

110 patients from the study met the criteria for the study and were assigned to dose-escalation cohorts where ASP8273 was administered orally in a single-dose period lasting 2 days and followed by repeat-dose cycles consisting of once-daily treatment over 21 days. Of the 110 patients, 93 were eligible for biomarker analysis of cfDNA, and 46 out of 93 had sufficient plasma samples for longitudinal analysis. Mutations in *EGFR* were examined in cfDNA isolated from blood serially collected prior to study start and at each treatment cycle, using beads, emulsification, amplification, and magnetics (BEAMing) digital PCR. Additionally, *EGFR* mutation status was also assessed centrally by RT-PCR. Percentage mutant *EGFR* cfDNA (%) was observed at baseline and at cycle 2 in patients with *EGFR* T790M positive metastatic NSCLC treated with ASP8273. Patients were grouped by best response to ASP8273, including partial response, stable disease, or progressive disease.

The authors concluded that the presence of *EGFR* T790M mutations in cfDNA predicted response to ASP8273 and that using cfDNA to identify mutation patterns of progression throughout treatment, such as the emergence of new mutations in *EGFR*, or the reemergence of mutations initially identified at baseline may be potentially useful in the clinic. Reductions in *EGFR* levels in cfDNA were seen across a broad range of doses in this phase I study (100mg-500mg), which suggests activity of the agent at a range of doses. Due to the high concordance observed with tumor tissue, the authors recommended that further studies to understand the relationship between cfDNA and tumor burden, as well as other clinical parameters, be conducted.

3. ctDNA changes in advanced lung and bladder cancer patients receiving PD-L1 inhibitor (durvalumab) as a potential response biomarker, 2018[16]

This study investigated changes in *variant allele frequencies (VAF)* of somatic mutations in ctDNA from the blood of patients with advanced NSCLC and urothelial cancer (UC) and their association with patient outcomes after treatment with PD-L1 inhibitor durvalumab. The study found that a reduction in ctDNA VAF at 6 weeks is associated to tumor shrinkage and improved progression-free and overall survival.

Patient blood was extracted at baseline (pre-dose) and six weeks after the first dose (post-dose). ctDNA was tested using the Guardant 360 gene panel comprising of 73 genes. Somatic variants, including single nucleotide variants (SNVs), insertions/deletions, and fusions were summarized for each patient by calculating the mean allele frequency of all genes with a VAF ≥0.3% at pre-dose. Both synonymous and non-synonymous mutations were included in the VAF calculation. Change in mean VAF was calculated when mean VAF at pre-dose was subtracted from VAF at 6 weeks. Mean VAF was compared across timepoints (pre-dose and post-dose) and correlated with objective response rate (ORR), time on study, tumor volume, and survival **(Table 1)**.

Patients from two different clinical trials were included in this analysis. Study 1108 (NCT01693562) was a phase 1/2, first-in-human, multicenter, open-label dose-escalation, and dose-expansion study. Eligible patients were ≥18 years of age with histologically or cytologically confirmed inoperable or metastatic transitional-cell UC or NSCLC and who had progressed on, been ineligible for, or refused any number of prior therapies. The second clinical trial, ATLANTIC (NCT02087423), was a multicenter, phase 2 open-label study enrolling patients with Stage IIIB/IV NSCLC with disease progression following two or more systemic treatments, including one platinum-based chemotherapy and one tyrosine kinase inhibitor (TKI) for *EGFR* mut/ALK+ patients.

This study also observed the emergence of new *EGFR* mutations in patients with progressive disease at week 6. These mutations have been previously associated with resistance to immunotherapies. Thus, the use of liquid biopsies throughout the course of therapy will enable longitudinal monitoring of changes in tumor burden, and the identification of new mutations that are associated with patient outcomes that may facilitate the development of combination therapies in immuno-oncology.

BEST PRACTICES FOR THE USE OF ctDNA AS A POTENTIAL MONITORING TOOL

The case studies described above demonstrate the potential clinical impact ctDNA may have on disease monitoring and the potential utility liquid biopsies may have to help assess drug efficacy early during a clinical trial. Given the convenience of ctDNA analysis and its ability to quantify mutations in ctDNA throughout treatment and identify new mutations that arise during treatment that may confer resistance to ongoing therapies, identifying a consistent way to use ctDNA as a monitoring tool is imperative. Outcomes of studies to date have been variable, and this variability is explained by different technologies used, the lack of standardization, and the absence of prospective clinical and biomarker data.

Drawing from studies performed to date, their methods and the limitations of those methodologies, as well as from a wealth of personal experience, the working group has generated a list of best practices and recommendations that have been classified into the following categories: material collection, detection platform technology, and analysis **(Table 2)**.

While not the primary focus of this white paper, the need for rigorous analytical validation parameters of ctDNA assays should also be acknowledged. Ongoing efforts being led by other organizations, such as the Blood Profiling Atlas in Cancer (BloodPAC), are determining best practice principles for validating liquid biopsy tools for ctDNA assessment.

Generally, prior to using a ctDNA assay as a tool for drug development in a clinical trial, the assay should be analytically validated, and the cutoffs should be pre-specified and locked down. Some of the key analytical studies include, but not limited to, limit of blank (LoB), limit of quantitation (LoQ, only for quantitative assay i.e., the assay has continuous output), limit of detection (LoD), linearity (only for quantitative assays), analytical accuracy, and precision/reproducibility, should be evaluated to establish optimal assay performance. Since these assays will likely have a quantitative output, it is expected that the analytical and clinical studies are consistent with the assay's intended use. In order to report underlying continuous measures (e.g. MAF, bTMB, circulating tumor fraction), analytical validation studies or analyses should be done to demonstrate that those continuous values can be accurately and reliably measured.

Additionally, for monitoring purposes, the proposed assay should continuously assess a subject's status over a period of time or monitor at intermittent times, and the study duration should be long enough to capture the range of variation in the assay measurements and clinical status from the assay's intended use population. The time interval at which the data is collected and how many and how often data points per patient are collected should be clinically acceptable. How change in the assay result or patient clinical status is defined and determined should be clearly prespecified.

Table 2: Best practices for the use of ctDNA in disease monitoring

Best Practice	Recommendations
Material collection	
Timing	1. Collection at cycle 1, day 1 (screening sample may not be representative) 2. Early collection after 2-4 weeks 3. Collection at the time of restaging scans 4. Collection at or after progression (prior to next therapy)
Amount of material	• One 10ml tube is usually adequate for analysis • Recommend collection of a second 10mL tube for future bridging studies • Recommend saving the cell pellet to allow study of white blood cells if needed.
Tube type	• If site has capacity to spin down tubes locally within a few hours after collection, EDTA tubes would be adequate. Otherwise tubes including a DNA stabilization agent (e.g. Streck tubes) are preferred to allow delayed spinning of specimens
Detection platform technology	• Should be able to measure ctDNA changes quantitatively • Recommend quantification of variant allelic fraction, which can be calculated across various assays (e.g. ddPCR, NGS) • Platform should be validated to show optimal commutability against other assays (orthogonal approaches)
Analysis	• Consider calculation of percent change from baseline, similar to approach used for tumor measurements in imaging • Analysis should account for the possibility of mutations derived from clonal hematopoiesis. Sequencing of white blood cells can be useful for distinguishing this

ESTABLISHMENT OF A MULTI-STAKEHOLDER CONSORTIUM TO OPERATIONALIZE ctDNA IN DRUG DEVELOPMENT

As demonstrated by the case studies above, several studies have examined the association between ctDNA and clinical outcomes in patients with advanced cancer. However, different analytical approaches are currently used in each study, which make it challenging to generate broad learnings across cancer types and treatment settings. Through conversations with multiple stakeholders, this working group has identified two potential opportunities to better understand the relationship between changes in ctDNA levels in plasma and treatment outcomes and promote the operationalization of ctDNA in drug development: a prospective collection of ctDNA from ongoing clinical trials, which will implement standard practices for plasma collection and analyses of plasma response, and the collection of existing datasets from past clinical trials and studies from which to learn how to best use ctDNA in drug development.

ctDNA Pilot Project: Monitoring therapeutic effect of immune checkpoint inhibitors

The variability observed across studies and existing datasets demonstrates the need for the prospective validation of ctDNA in rigorous cohorts. Achieving this will require standardization of data processing, collection, and analysis.

There is a need for the development of standard practices that may promote the integration of ctDNA into clinical trials and facilitate the aggregation and analysis of resulting data. Moreover, it is important to understand how optimal, feasible, and reproducible these practices are, and whether the data collected could be easily aggregated from large trial studies.

A unified prospective pilot could allow us to rigorously address a key clinical question: Do changes in ctDNA levels accurately reflect the therapeutic effect of immune checkpoint inhibitors?

To address this important question, this working group proposes the creation of a pilot project where a standardized add-on study framework is adopted for the collection of a core set of ctDNA measurements and clinical endpoints as part of ongoing or new clinical trials.

The pilot project would assess the feasibility of bringing together data from several clinical trials that are investigating same in-class agents in a specific population and determine the minimum amount of data that sponsors would be willing and able to share to evaluate outcomes based on ctDNA measurements.

Table 3: *Friends* ctDNA pilot project framework

Parameter	Proposed Pilot
Patient population	Patients with advanced/metastatic disease
Population size	As determined by the clinical trial or drug sponsor
Drug class	Immune checkpoint inhibitors
Trial phase	All phases
Technology for ctDNA assessment	ddPCR or NGS gene panel
Minimum Limit of Detection	0.2-0.25% VAF
Test tubes	If site has capacity to spin down tubes locally within a few hours after collection: EDTA. Otherwise tubes including a DNA stabilization agent (e.g. Steck tubes)
Timepoints	1. Collection at cycle 1, day 1 (screening sample may not be representative) 2. Early collection after 2-4 weeks 3. Collection at the time of restaging scans 4. Collection at or after progression (prior to next therapy)
Median follow up	6 months
Diagnostic endpoints	Relative percent change from baseline
Alterations (definition)	Mutations, insertions, deletions, amplifications, and fusions
Clinical endpoints	Raw tumor size/volume, ORR and PFS and/or OS, if applicable (trial dependent)
Adjustment factors	Age, gender, smoking status, baseline ECOG score, previous line of therapy, and histology

Table 3 describes a framework proposed by the working group that could be added on to an ongoing trial. This framework outlines a few key elements that will delineate how ctDNA and clinical data could be collected during clinical trials and proposes methods for assessing the correlation between differences in ctDNA dynamics and response. The working group hopes the framework is reasonable and feasible for participating sponsors to readily incorporate into ongoing or planned trials, without compromising or interrupting their primary trial objectives.

If the right clinical trials are identified and the pilot project framework is well implemented, the preliminary evidence collected would increase our understanding on the feasibility and effectiveness of using ctDNA as a monitoring tool in clinical trials that investigate the efficacy and safety of immune checkpoint inhibitors either used as monotherapy or in combination.

Virtual ctDNA data repository

ctDNA has been and is currently being collected in clinical trials. These rich datasets are currently stored in isolated silos, which preclude powerful and robust analyses that measure the association between plasma response and therapeutic effect. Aggregating these existing datasets in a central virtual repository would allow for datasets to be analyzed together, enabling researchers to draw more significant conclusions and promoting a more refined understanding of plasma response to various therapies, such as chemotherapy, targeted therapies, and immunotherapies.

The working group proposes to explore the creation of a central virtual ctDNA data repository by bringing different stakeholders across academia and industry together to discuss how already-generated data from individual studies could be brought together in a pre-competitive environment. The overarching goal of this initiative would be to discuss how these data could be brought together, what data could be shared across studies, and how these data would be used to derive more insightful conclusions than isolated and smaller studies with limited sample sizes.

A multi-stakeholder virtual data repository offers potential to generate broad learnings in a pre-competitive fashion to facilitate our understanding of ctDNA changes as a measure of drug effect.

Clinical trials use a range of ctDNA analytical approaches and technologies, but most studies have a common core set of data elements and offer means to calculate the *allelic fraction (AF)* of key cancer-associated genes like *EGFR*, *KRAS*, and *TP53*. A combined analysis of existing datasets offers the potential for several learnings:

1) **What magnitude of change in AF portends a better response rate, PFS, or overall survival on therapy (e.g., any change, 50% change, 90% change, or 100% change?)**
2) **How does the relationship between change in AF differ in patients treated with chemotherapy, targeted therapy, or immunotherapy?**
3) **What minimum baseline "measurable" AF is needed to be able to accurately detect a response in plasma ctDNA?**

These learnings will be helpful in furthering our understanding of plasma response and the use of ctDNA in drug development, but a proper framework that will foster collaborations is critical to ensure such a repository is a successful collaborative tool. The working group has put together a list of considerations and questions that begins to explore the potential design and implementation of a virtual data repository that would host ctDNA data to explore plasma response **(Table 4)**.

This type of repository would be beneficial for understanding how best ctDNA could be used in drug development and would help inform future initiatives that seek to operationalize ctDNA in drug development.

Table 4: Considerations for a virtual data repository

Issues	Questions
Core dataset	• What is the minimum core set of data elements that s ponsors would feel comfortable sharing as part of a pilot project? • Should raw or analyzed data be uploaded to the repository? • What kind of case report data on clinical response is necessary?
Legal, ethical, and privacy concerns	Are there any legal, ethical, and/or privacy concerns for contributing data to a virtual repository?
Logistical concerns	
Data storage	Where would the data be stored? Would there be a maximum data storage value? Could this data be hosted on a cloud?
Data transfer	How would data be transferred/uploaded?
Blinding	Does the data need to be blinded?
Analytical opportunities	Will the data be analyzed as a meta-analysis, or could the data be combined and analyzed together?

NEXT STEPS

This white paper lays out best practices for ctDNA use in disease monitoring and proposes two collaborative initiatives that could help elucidate how ctDNA may be used in drug development across cancer types and treatment settings. The members of the working group encourage comments and reactions to the best practices and the collaborative initiatives proposed in this whitepaper.

Future steps will include the following:

1. ***Friends* will seek to develop a multi-stakeholder consortium: interested members of the academic, diagnostics, government, pharmaceutical, and patient advocacy communities should request to join the ctDNA multi-stakeholder consortium;**
2. **The consortium will meet to discuss the feasibility of the initiatives discussed in this white paper; and**
3. **The consortium will implement the optimal approach to advance our understanding of ctDNA use in drug development**

LIST OF DEFINITIONS

- **Allelic fraction (AF):** refers to the percentage of a sample represented by an allele. Thus, a mutant allele fraction refers to the fraction of alleles (DNA molecules) at a locus that carry a mutation.
- **Cell-free DNA (cfDNA):** total amount of cell-free DNA in plasma or serum, which can be derived from multiple sources, including tumor cells.
- **Circulating tumor DNA (ctDNA):** the fraction of cell-free DNA that originates from tumor cells. The presence of ctDNA in cell-free DNA is generally inferred by the detection of somatic variants, consequently, the presence of ctDNA in cell-free DNA is usually not confirmed until after a ctDNA assay is performed.
- **ctDNA assay:** a clinical test designed to detect somatic variants in cell-free DNA. These encompass a single variant in a gene or broad assays that may interrogate numerous variants in various genes. Other terms to describe ctDNA assays include circulating cell-free plasma DNA assays and plasma genotyping assays.
- **Digital droplet PCR (ddPCR):** a refinement of conventional polymerase chain reaction (PCR) methods where the PCR solution is divided into smaller reactions contained in droplets created through a water oil emulsion technique. Each droplet runs individual PCR reactions independently to directly quantify and clonally amplify nucleic acids in a more accurate and sensitive manner.
- **Liquid biopsy:** a broad category for a minimally invasive test done in a sample of blood to look for cancer cells from a tumor that are circulating in the blood or for fragments of tumor-derived DNA that are in the blood. Tumor genetics or genomics from ctDNA assays are one example.
- **Genotyping (uses for ctDNA):** detection of targetable biomarkers or resistance mutations to guide treatment selection.
- **Monitoring (uses for ctDNA):** repeat assessment to evaluate quantitatively or qualitatively for treatment effect.
- **Cancer detection (uses for ctDNA):** detection of hallmarks of cancer either for initial diagnosis of cancer or for detection of residual cancer at a single high-risk timepoint (e.g. minimal residual disease).
- **Minimum residual disease (MRD):** residual cancer burden persisting in patients considered to be in morphologic remission. Commonly used term in the treatment of blood cancers.
- **Molecular/Plasma response:** changes in ctDNA as a result of a therapeutic intervention.

- **MRD assay:** assay that is tested at some early high impact time to help determine whether a patient is cured or not. Such an assay could also be used at intervals to monitor for recurrence. But the statistical characteristics (and development path, and cost/benefit implications) for a single-timepoint detection assay is quite different than for a multi-timepoint monitoring assay.
- **Next-generation sequencing (NGS):** next-generation sequencing (NGS), also known as high-throughput sequencing, is a term used to describe a number of different modern sequencing technologies that allow us to sequence DNA and RNA much more quickly and cheaply, and as such have revolutionized the study of genomics and molecular biology
- **Real-time PCR (RT-PCR):** real-time polymerase chain reaction (PCR) monitors the amplification of a targeted DNA molecule during the PCR in real-time, and not at its end, as in conventional PCR. RT-PCR can be used quantitatively or semi-quantitatively.
- **Recurrence:** cancer that has recurred usually after a period of time during which the cancer could not be detected. The cancer may come back to the same place as the original (primary) malignancy (local recurrence) or to another place in the body (distant recurrence, or metastasis).
- **Variant allele fraction (VAF):** the fraction of alleles in a specimen that contain the variant, or mutation.

REFERENCES

[1] Mandel P, Metais P. Les acides nucleiques du plasma sanguin chez l'Homme. C R Acad Sci Paris 1948;142:241–3.

[2] Kitahara M, Hazama S, Tsunedomi R, et al. Prediction of the efficacy of immunotherapy by measuring the integrity of cell-free DNA in plasma in colorectal cancer. Cancer Sci 2016;107(12):1825–9.

[3] Castro-Giner F, Gkountela S, Donato C, et al. Cancer Diagnosis Using a Liquid Biopsy: Challenges and Expectations. Diagnostics 2018;8(31):1–18.

[4] Husain H, Velculescu VE. Cancer DNA in the circulation: The liquid biopsy. JAMA - J Am Med Assoc 2017;318(13):1272–3.

[5] Phallen J, Sausen M, Adleff V, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017;9(403).

[6] Whole-Genome Sequencing. Sci Transl Med [Internet] 2012 [cited 2012 Nov 29];4(162):162ra154-162ra154. Available from: http://stm.sciencemag.org/cgi/doi/10.1126/scitranslmed.3004742

[7] Bettegowda C, Sausen M, Leary RJ, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci Transl Med [Internet] 2014;6(224):224ra24. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4017867/

[8] Tissot C, Toffart AC, Villar S, et al. Circulating free DNA concentration is an independent prognostic biomarker in lung cancer. Eur Respir J [Internet] 2015;46(6):1773–80. Available from: http://dx.doi.org/10.1183/13993003.00676-2015

[9] Lee JH, Long G V., Boyd S, et al. Circulating tumour DNA predicts response to anti-PD1 antibodies in metastatic melanoma. Ann Oncol Off J Eur Soc Med Oncol 2017;28(5):1130–6.

[10] Toledo RA, Cubillo A, Vega E, et al. Clinical validation of prospective liquid biopsy monitoring in patients with wild-type RAS metastatic colorectal cancer treated with FOLFIRI-cetuximab. Oncotarget [Internet] 2015;8(21):35289–300. Available from: http://www.oncotarget.com/abstract/13311

[11] Dawson S-J, Tsui DWY, Murtaza M, et al. Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. N Engl J Med [Internet] 2013 [cited 2013 Mar 13];130313140010009. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1213261

[12] Gray ES, Rizos H, Reid AL, et al. Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. Oncotarget [Internet] 2015;6(39):42008–18. Available from: http://www.oncotarget.com/fulltext/5788

[13] Oxnard GR, Paweletz CP, Kuang Y, et al. Noninvasive detection of response and resistance in *egfr*mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin Cancer Res 2014;20(6):1698–705.

[14] Mok T, Wu YL, Lee JS, et al. Detection and dynamic changes of *EGFR* mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC Patients treated with first-line intercalated erlotinib and chemotherapy. Clin Cancer Res 2015;21(14):3196–203.

[15] Yu HA, Spira A, Horn L, et al. A phase I, Dose escalation study of oral ASP8273 in patients with non–small cell lung cancers with epidermal growth factor receptor mutations. Clin Cancer Res 2017;23(24):7467–73.

[16] Raja R, Kuziora M, Brohawn P, et al. Early reduction in ctDNA predicts survival in lung and bladder cancer patients treated with durvalumab. Clin Cancer Res [Internet] 2018;clincanres.0386.2018. Available from: http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-18-0386

ALEXANDRIA®
Building the Future of Life-Changing Innovation™

FRIENDS of CANCER RESEARCH

BLUEPRINT FOR BREAKTHROUGH FORUM

RESEARCH AND REIMBURSEMENT IN THE AGE OF PRECISION MEDICINE

GOAL

This whitepaper addresses the need to establish minimum analytical and clinical data elements to improve transparency in test performance and expand sources of evidence collection that could ensure patient and provider confidence. This whitepaper focuses specifically on next generation sequencing-based tests intended to detect somatic mutations in clinically actionable genes in solid tumors.

INTRODUCTION

Next-generation sequencing (NGS) is becoming a commonly used tool in cancer treatment to provide essential information about a patient's diagnosis and treatment options. These tests are widely available as laboratory developed tests (LDT) and, in recent months, the Food & Drug Administration (FDA) approved several new diagnostic tools that utilize NGS technologies as well. Further, the Center for Medicare & Medicaid Services (CMS) issued a national coverage decision to support coverage for certain NGS-based tests.

These advancements in diagnostic technology and regulatory and coverage policy present new opportunities to gain information about hundreds of genomic alterations at once. Providing adequate information about tests to patients and physicians is critical to ensuring the appropriate clinical use and interpretation of test results. Transparency regarding the clinical performance and utility of different NGS-based tests available will aid in

CONTRIBUTORS

Stayce Beck
U.S. FDA

Mike Berger
Memorial Sloan Kettering Cancer Center

Gideon Blumenthal
U.S. FDA

Pam Bradley
Personal Genome Diagnostics

Asia Cheng
Thermo Fisher Scientific

Jonathan Hirsch
Syapse

Florence Kurttila
Patient advocate

Alexander Lazar
MD Anderson Cancer Center

Kavita Patel
Brookings Institution

Reena Philip
U.S. FDA

Girish Putcha
Palmetto GBA MolDX

Katherine Szarama
CMS

ABOUT FRIENDS OF CANCER RESEARCH

Friends of Cancer Research drives collaboration among partners from every healthcare sector to power advances in science, policy, and regulation that speed life-saving treatments to patients.

ABOUT ALEXANDRIA REAL ESTATE EQUITIES, INC.

Alexandria Real Estate Equities, Inc., an S&P 500® company with a $18B total market cap, is an urban office REIT uniquely focused on collaborative life science and technology campuses and passionately committed to curing disease and ending hunger. Through our four core business verticals — real estate, venture investments, thought leadership, and corporate responsibility — Alexandria builds dynamic ecosystems that enable the world's most innovative people, organizations, and companies to develop life-changing breakthroughs that improve the human condition.

Alexandria's Thought Leadership platform convenes its global life science network for highly unique and interactive programming and events, including the renowned Alexandria Summit and annual *Friends*-Alexandria Blueprint Forum, to drive discussions and collaborations that will shape the future of healthcare policy and the future of human health.

clinical decision-making and facilitate improvements in patient care. Furthermore, this information could help inform reimbursement decisions by private and public payors. However, the types of evidence and mechanisms to communicate this information are an area of continual debate. Demonstrating analytical and clinical validity and clinical utility of diagnostic tests requires time and money. Innovative reimbursement mechanisms can help facilitate and encourage the development of evidence over time with the ultimate goal of ensuring maximum benefit to patients and the healthcare system overall.

This whitepaper will address three key questions regarding reimbursement mechanisms designed to facilitate more transparency and robust evidence development for diagnostic tests with the intent of establishing consensus on best practices and next steps.

1. **What is the minimum core dataset that should be made publicly available for NGS-based diagnostic tests? What information is important to patients, providers, and payors? How can this be updated over time based upon changes to the test or clinical knowledge?**

2. **What mechanisms exist to support the collection of this data in a real-world setting? What are the standards needed to ensure collection of high-quality data?**

3. **How should the reporting of this data be formatted to make it readily informative to patients and providers in diagnostic and treatment decision-making?**

Establishing a core dataset for informing patients, providers, and payors of optimal use of NGS-based diagnostic tests.

The technology upon which diagnostic tests are based is becoming increasingly sophisticated, making it more difficult, and simultaneously more imperative, to validate them and accurately and transparently communicate their performance specifications. Patients, providers, and payors require greater transparency regarding the analytical and clinical validity and clinical utility of diagnostic tests to ensure public confidence and support their use. However, appropriate levels of transparency are difficult to achieve for these complex tests as the type and depth of information that should be shared varies according to the specific consumer of that information. Certainly, while payors require a wide range of detailed analytical and clinical data to support reimbursement decisions, patients and providers may desire access to more clinically relevant information conveyed in a meaningful manner to ensure that patients receive the most appropriate diagnostic test for them.

One approach to addressing transparency could be for laboratories to provide test performance characteristics in a standardized format available in a public database, on company websites, or

on third party sites (e.g., NIH, ASCO, AMP, CAP, etc.). This transparency would allow physicians and patients the opportunity to assess the potential advantages and disadvantages of individual tests. A second approach would be to provide a publicly available list of individual tests that meet certain analytical, and possibly clinical, performance characteristics using properly qualified reference samples and/or materials. This would provide patients and their physicians with assurance that the test being used to guide their care is accurate and reliable, without placing the potential burden of test evaluation on the patient or treating physician. Processes for certifying the test performance and updating the list of tests would require additional discussion. Ultimately, the goals are to ensure maximum benefit for patients and to incentivize clinically beneficial innovation by providing reimbursement commensurate with the quality and transparency of data provided.

In addition, one must also balance between the availability of such information and the administrative burdens of reporting it. Communicating adequate information in an appropriate format for each of the various stakeholders (patient, provider, and payor) will necessitate agreement upon a minimum set of validation elements that should be made public concerning each test and a standardized template for communicating these specifications in the least burdensome manner. For example, a standardized questionnaire could be adopted for reporting test validation elements to payors and a similar but simplified questionnaire could be adopted for making data publicly available for providers. Reports containing the data elements outlined in **Table 1** could be generated and provided to patients, either as part of patient education materials concerning their specific test or as part of their laboratory test report.

Table 1: Data elements for public availability*

Validation Element	Validation Element Detail
Accuracy	
Method Comparison(s)[1, 2]	Compare new test to "standard of care" reference method
Specimen Types[1]	List all specimen types and how they were validated
Matrix Comparison(s)[1]	Indicate all validated sample matrices and how they were validated
Analytical Sensitivity	
Limit of Blank (LOB)	If applicable
Limit of Detection (LOD)	
Limits of Quantitation[1]	Include descriptions of analytically measurable range and clinically reportable range, if applicable
Linearity and Reportable Range[1]	If applicable
Minimum Input Quantity and Quality[1]	
Minimum Tumor Content[1]	
Precision	
Repeatability	Single operator, instrument, lot, day, and run
Intermediate Precision[1, 2]	Multiple operators, instruments, days, and runs within a lab
Reproducibility	Multiple labs/sites, if applicable
Lot-to-lot Reproducibility	Multiple reagent, calibrator, and control lots, as applicable
Reference Intervals	If applicable
Sample Stability	
Primary Sample	
Clinical Performance Characteristics	
Positive Percent Agreement (PPA)	Reported with respect to each variant type and LOD for that variant
Negative Percent Agreement (NPA)	
Overall Percent Agreement (OPA)[1, 2]	
Clinical Utility[1, 2]	
Intended Use Population(s)[1, 2]	
Clinical Outcomes Data[1, 2]	Summaries of studies supporting clinical outcomes of the specific test

*All validation elements should be reported with confidence intervals.
Note: All information above should be provided to payors, while only certain subsets may be appropriate and relevant for providers([1]) and patients([2]).

Furthermore, particularly for NGS-based gene panels, it may not be necessary to provide this information for all genes and variants on a panel but only for "clinically actionable" genes to reduce the administrative burden associated with reporting and provide predictability as to when reporting is appropriate. While the definition of "clinically actionable" can be controversial, one approach is to make publicly available a test's performance on FDA-approved biomarkers linked to the prescribing of an FDA-approved drug **(Table 2)**.

Table 2: Representative clinically actionable gene targets relevant to oncology*

Gene	Disease	Indicated Drug(s)
BRAF	Non-Langerhans Cell Histiocytosis/Erdheim-Chester Disease, Anaplastic Thyroid Cancer, Melanoma, Non-Small Cell Lung Cancer	Vemurafenib, Dabrafenib + Trametinib, Dabrafenib, Vemurafenib, Binimetinib + Encorafenib, Cobimetinib + Vemurafenib, Trametinib
BRCA1	Ovarian Cancer	Niraparib, Rucaparib
BRCA2	Ovarian Cancer	Niraparib, Rucaparib
EGFR	Non-Small Cell Lung Cancer	Afitinib, Erlotinib, Gefitinib, Osimertinib
ERBB2	Breast Cancer, Esophagogastric Cancer	Ado-Trastuzumab Emtansine, Lapatinib, Lapatinib + Trastuzumab, Neratinib, Pertuzumab + Trastuzumab, Trastuzumab
KIT	Gastrointestinal Stromal Tumor	Regorafenib, Imatinib, Sunitinib
KRAS	Colorectal Cancer	Cetuximab, Panitumumab, Regorafenib
PDGFRA	Gastrointestinal Stromal Tumor	Imatinib
TSC1	CNS Cancer	Everolimus
TSC2	CNS Cancer	Everolimus

*This is not a comprehensive list. This list was limited to include single nucleotide variants and insertion/deletion events, initially, and could eventually be expanded to include other events relevant to oncology, including rearrangements with companion diagnostic claims such as ALK and ROS1 in non-small cell lung cancer or PDGFRB in dermatofibrosarcoma protuberans.

Many laboratories have also begun reporting gene signatures relevant to oncology, such as microsatellite instability and tumor mutational burden, which add further complexity to validation and should also be considered for NGS-based test reporting.

Further refinement of reporting could be achieved if different validation elements could be identified for public availability based upon different uses of a test. For example, limited public information, such as summary analytical validity, may be desired for lower tier tests since they are likely to be largely utilized for research purposes and the evidence base is still being established. However, it should be made clear to patients what is known and not known about the test being performed on them. For clinical uses to make treatment decisions, it may be desired to have components of analytical and clinical validity data available, and ultimately for the highest tiered tests that are used as companion diagnostics, clinical outcomes data would be important to be made readily available for different stakeholders.

Equally as critical as determining the appropriate metrics by which to assess a test's performance, is the source of data and the entity that validates the data. Evaluation of analytical and clinical performance may require access to appropriate clinical samples and/or reference materials. The availability of clinical samples, especially with clinical outcomes, is limited, so other sources and types of evidence should be explored, and the limitations understood. Specifically, the below sources of evidence would not be used to support clinical utility, see section *"Identifying innovative methods and standards for data collection on evolving uses in the real-world setting"* for exploration of the use of real-world data to support evidence of clinical utility.

Table 3: Sources of evidence to assess test performance

Sources of Evidence	Evidence Supporting
Clinical samples (with outcomes)	Analytical and clinical validity
Clinical samples (with known biomarker status but no clinical outcomes)	Analytical validity only
Reference materials (RMs)	Analytical validity only
Reference samples (as distinct from RMs)	Analytical validity only
Published literature	Clinical validity only

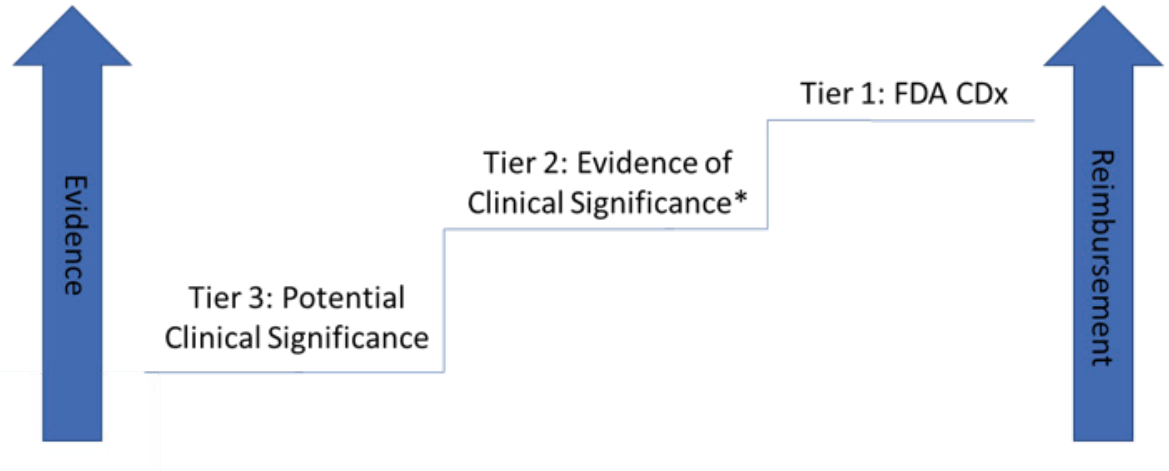
Appropriate third-party reviewers and frameworks for reporting validated data will be discussed in *"Identifying mechanisms to readily communicate data to patients and providers for diagnostic and treatment decision-making."*

Identifying innovative methods and standards for data collection on evolving uses in the real-world setting.

The extensive efforts of test developers that have demonstrated analytical and clinical validity and clinical utility of their diagnostic tests should be recognized in some way such that it provides an incentive for test developers to pursue evidence generation (e.g., differential reimbursement, **Figure 1**). Mechanisms to establish clinical utility without a randomized clinical trial and assess changes in patient outcomes to justify payment and the role of evidence from the real-world setting were explored. Vehicles and standards for data collection in the real-world should be explored, including identifying real-world endpoints that can establish the clinical utility of molecular tests; defining a pathway to validate real-world endpoints; and a framework for the potential use of real-world evidence to support reimbursement of molecular tests.

Figure 1. Tiered reimbursement scheme. A potential model to incentivize test developers to pursue additional evidence generation.

Figure 1: Tiered reimbursement scheme



*with reporting of analytical and clinical validity

Table 4 lists possible real-world evidence that can be collected in order to support the use of a diagnostic test based upon the clinical outcome data elements and clinically actionable genes identified in **Tables 1** and **2**.

Table 4. Types of evidence to collect through real-world sources

	Component	**Description**	**Purpose and Utilization for Decision-making**
What is the context of use of the molecular test?	Disease Characteristics	Primary Cancer Type Stage at Diagnosis Current Clinical Stage/ Metastatic Disease Status Prior Line of Therapy	Clinical characteristics of patient population and impact on clinical endpoints
	Diagnostic Test	Test Vendor Test Type Genes and Variant Types Tested Genomic Results Quality Measures (as defined in Table 1)	Understand the testing performed
Does the molecular test impact clinical care decisions?	Change in Care	*Following physician receiving molecular test results…* Intent to Change Treatment (including stop and start of treatment; inclusive of targeted therapies, immunotherapies, and clinical trials) Change in Treatment (as measured by successful fill/ administration) Difference between Intent and Change (assess whether obstacles in therapy procurement or trials enrollment effected molecular test impact)	Understand whether testing led to change in care decisions

Identifying mechanisms to readily communicate data to patients and providers for diagnostic and treatment decision-making.

While access to adequate and high-quality information regarding diagnostic tests for providers, patients, and payors is imperative, it is equally important that this information is made available in a format that is tailored to meet the needs of the intended audience.

PROVIDERS

A mechanism that identifies the appropriate information to convey expectations and capabilities of each test to providers is needed to support decision making. The CMS Appropriate Use Criteria (AUC) Program for advanced diagnostic imaging tests could be adapted for communicating information concerning quality and appropriateness of prescribing specific diagnostic tests. As with the existing AUC Program, entities with expertise in diagnostic assessments could be identified for certification as provider led entities (PLE). These PLEs would be qualified to develop, modify, and endorse AUC based on the submissions of a minimum set of validation elements (**Table 1**) by diagnostic test manufacturers or clinical labs. AUC would then be incorporated into a qualified electronic clinical decision support mechanism (CDSM) to be referenced by providers. This process would enable physicians to order diagnostic tests on a patient-specific basis according to the test analytical and clinical validity and clinical outcomes information provided through the mechanism in a user-friendly format.

PATIENTS

Patients may not be aware of concerns with the specifics of a test's analytical validation, such as comparisons of minimum input quality or limits of detection, but an overall assurance that the test has been adequately validated is necessary to ensure confidence. A general grading scale of A, B, or C administered through the Appropriate Use Criteria program and reported to patients by providers could be used to convey the level and quality of data reported for tests to enable patients to become more informed and increase patient confidence in test outcomes. Overall performance results from organizations administering proficiency testing could also be provided for inclusion as a metric in the AUC grading scheme to provide a better understanding of the comparability of analytical performance across platforms and laboratories (**Table 5**). This grading scheme and reporting will be essential for standardizing the information reported to patients and physicians and ensuring the interpretability of lab report information. However, appropriate confidentiality mechanisms would be needed when implementing such a framework to avoid use of the framework as a marketing tool, which could undermine the true intent of the grading system. Further, patient and provider groups could make available a standardized questionnaire (**Supplemental Table 1**) to guide patient discussions with their healthcare team concerning their diagnostic tests to enable more informed patients and providers.

Table 5. Elements for consideration in a diagnostic test grading checklist

	Grade		
Validation or Proficiency Element	A	B	C
Accuracy			
Analytical Sensitivity			
Precision			
Sample Stability			
Gene Coverage			
Clinical Performance Characteristics (PPA, NPA, OPA)			
Clinical Validity (Quality and quantity of data)			
Clinical Utility (Impact on clinical care and outcomes)			

PAYORS

Consistent with the existing Appropriate Use Criteria program, Medicare reimbursement decisions could be tied to provider consultation of AUC through qualified CDSMs during their diagnostic test decision making process. As the AUC program currently specifies, ordering providers would be required to consult CDSMs and report this consultation information to furnishing providers. Furnishing providers would then be responsible for including on the Medicare claim information about the ordering professional's consultation with a CDSM.

Questions for consideration

1. For a clinically well characterized biomarker with an existing companion diagnostic test, what is required to establish confidence in that test by physicians? Patients? For reimbursement? Is a clinical trial always necessary?

2. Could data collected from clinical experience with an NGS test be used to identify a targeted population? If so, what would the desired data elements be?

3. Are there scenarios in which an NGS test could be eligible for reimbursement without being contemporaneously developed with a drug (if so, when is a prospective demonstration of clinical outcomes the only acceptable approach)?

4. When a new companion diagnostic/drug pair becomes approved for a "new" variant or for a "new" indication, what evidence should existing tests provide in order to qualify for regulatory approval? For reimbursement?

5. Should "higher" levels of evidence support higher levels of reimbursement from payors? What are the "tiers" of evidence that warrant higher levels of reimbursement? Is this feasible given the existing billing codes?

6. What incentives or legal protections would need to be in place to promote data sharing and development of an evidence base (either for reimbursement purposes or regulatory decisions?

7. Is it possible to promote sharing into research-grade databases, using the established metrics, such that these could be elevated to regulatory-grade with improved evidence base?

SUPPLEMENTAL TABLE 1

Patient Questionnaire
Important Questions to Ask My Healthcare Team Before My Procedure

PATIENT NEEDS	BASIC QUESTIONS	YES	NO	NOTES FOR HEALTHCARE TEAM
Transparency	Has it been explained to me why I need this test?			
	Have the benefits of the test been explained to me?			
	Have the risks associated with the test been explained to me?			
	Has the accuracy of this test been explained to me, as compared to other, similar tests?			
	Who will be performing the diagnostic test? (Doctor, Technician, Nurse, Clinician, etc.?)			
Ongoing Communication with my Healthcare Team	Have the diagnostic test, procedure, and expected outcomes been explained to me in a way I understand?			
	Are other similar diagnostic tests available and have they been explained to me? (Why do I need this test; could another test help me more?)			
	Has the intent of the test been explained to me (what will it confirm or rule out)?			
	Has my informed consent been explained to me and do I understand what I am signing?			
	Have I been told what the test involves?			
Cost, Co-Pays, Financial Responsibilities	Has the actual cost of the test, co-pays, and other out-of-pocket expenses been explained to me?			
	Will my private insurance pay for this test?			
	Will Medicare or Medicaid pay for this test?			
	Do I need prior authorization for this test?			

This questionnaire was developed as a guideline to assist patients and caregivers with specific questions to ask their healthcare team in the event of the necessity of a diagnostic test. It is not all inclusive. Each patient has a different story with different treatments and care plans for their disease, as well as other concerns. This is meant to initiate a good foundation and obtain information that is very basic to the needs and questions of a patient undergoing diagnostic procedures.

PATIENT NEED	BASIC QUESTIONS	YES	NO	NOTES FOR HEALTHCARE TEAM
Procedure / Test Description	Do I understand the actual procedure (has it been explained to me in a way I understand)?			
	Will I have pain? (Will I be anesthetized?)			
	Has the length of the procedure been explained?			
	Has the prep (if any) for the procedure been explained?			
	Have I been told how soon the procedure will be scheduled?			
	Is there a video / handout / or other resource available that I can research the procedure to be better prepared?			
	Have medications used in the procedure been explained to me?			
	Have they explained to me how long it will take to receive the results?			
	Have possible medications been explained to me due to the results of the procedure?			
	Have any potential interactions with my current treatment plan been explained to me?			
Understanding Terminology	Have medical terms, abbreviations, or acronyms been explained to me?			
	Do I understand them fully?			
	Do I have further questions on anything relative to the procedure?			
Resources, Research & Other Questions	What genes does this test identify and are they relevant for my cancer and possible treatment decisions?			
	Where can I obtain more information on my specific test (FDA approved? Lab Developed Test?, etc.)			
	What information is available regarding the clinical outcomes of the test that was ordered for me?			
	If I have other specific questions, who do I ask?			
	Can I change my mind about receiving the test?			

Patient Date

33rd Annual Meeting & Pre-Conference Programs of the Society for Immunotherapy of Cancer (SITC 2018): Late-Breaking Abstracts

Washington, D.C., USA. 7-11 November 2018

Published: 6 December 2018

O48

In silico assessment of variation in TMB quantification across diagnostic platforms: Phase 1 of the Friends of Cancer Research Harmonization Project

David Fabrizio, PhD[1], Shu-Jen Chen, PhD[2], Mingchao Xie, PhD[3], Wangjuh (Sting) Chen, PhD[4], Katie J. Quinn, PhD[5], Chen Zhao, PhD[6], Ahmet Zehir, PhD[7], Vincent Funari, PhD (CGMBS)[8], Jennifer S. Dickey, PhD[9], Vikas Gupta, PhD[10], Dinesh Cyanam, PhD[11], Lisa M. McShane, PhD[12], Naiyer A. Rizvi, MD[13], Matthew Hellmann, MD[7], Mark Stewart, PhD[14], Diana M. Merino, PhD[14], Jeff Allen, PhD[14], Friends of Cancer Research TMB Harmonization Team[14]

[1]Foundation Medicine, Cambridge, MA, USA; [2]ACT Genomics, Taipei City, Taiwan; [3]AstraZeneca, Waltham, MA, USA; [4]Caris Life Sciences, Phoenix, AZ, USA; [5]Guardant Health, Redwood City, CA, USA; [6]Illumina, San Diego, CA, USA; [7]Memorial Sloan Kettering Cancer Center, New York, NY, USA; [8]NeoGenomics Laboratories, Inc., Aliso Viego, CA, USA; [9]Personal Genome Diagnotics, Baltimore, MD, USA; [10]Qiagen, Aarhus, Denmark; [11]Thermo Fisher Scientific, Ann Arbor, MI, USA; [12]National Cancer Institute, Bethesda, MD, USA; [13]Columbia University Medical Center, New York, NY, USA; [14]Friends of Cancer Research, Washington, DC, USA

Correspondence: Jeff Allen (jallen@focr.org)

Journal for ImmunoTherapy of Cancer 2018, **6(Suppl 2):**O48

Background

Tumor mutational burden (TMB) is a measure of the number of somatic mutations and a predictive biomarker of response to immune checkpoint inhibitors (ICI) across several cancers. TMB can be estimated using targeted next-generation sequencing (NGS), but differences in quantification can arise based on platform differences, testing panel size and composition, and bioinformatic algorithms. Harmonization of methods to quantify TMB will facilitate biomarker development and optimize clinical utilization and treatment decision-making. Friends of Cancer Research (Friends) convened a group of leading diagnostic partners to assess and identify sources of TMB variability and determine best practices for harmonizing TMB estimation to ensure consistent clinical interpretation in the future.

Method

Eleven diagnostic members of the Friends TMB Harmonization Team used whole exome sequencing (WES) data from The Cancer Genome Atlas (TCGA) MC3 samples, comprising 32 cancer types. Each diagnostic partner calculated TMB from the subset of the exome restricted to the genes covered by their targeted panel and using their own bioinformatics pipeline (panel-derived TMB). A "gold-standard" TMB estimate was calculated from the entire exome using a uniform bioinformatics pipeline that all members agreed upon (WES-derived TMB). Linear regression analyses were performed to investigate relationship between WES-derived TMB and each panel-derived TMB. Exploratory analyses by cancer type were also performed. Bias and variability in TMB estimates across panel-derived TMB values were assessed.

Results

In silico quantification of TMB is relatively consistent between panels across a wide range of TMB values (0-40 mut/Mb). Panel-derived TMB strongly correlated with WES-derived TMB (regression R2 values range across panels 0.85-0.93, with slopes ranging from 0.82-1.37). Variation in TMB quantification was attributable to unique composition and technical specifications of each panel, as well as differences in the underlying algorithms used to estimate TMB from observed somatic mutations. Exploratory analyses suggested possible cancer type dependence for the relationship of panel vs WES-derived TMB, meriting further investigation.

Conclusions

In this in silico analysis, panel-derived TMB was strongly correlated with WES-derived TMB. Some variation in TMB quantification across panel-based diagnostic platforms exists. Identifying factors that contribute to variation will facilitate harmonization and help ensure appropriate use and implementation of tests results in the clinic. Subsequent steps will assess the effect of biologic factors (e.g. specimen type, cancer type, treatment setting), the impact of variation on clinical outcomes, align standards, and define best practices for quantification of TMB.

& Regulatory Science
1-7

DOI: 10.1177/2168479018759662
tirs.sagepub.com

Outdated Prescription Drug Labeling: How FDA-Approved Prescribing Information Lags Behind Real-World Clinical Practice

Michael B. Shea, BA[1], Mark Stewart, PhD[1], Hugo Van Dyke, MS[2], Linda Ostermann, BA[1], Jeff Allen, PhD[1], and Ellen Sigal, PhD[1]

Abstract

Background: Prescription drug labeling is an authoritative source of information that guides the safe and effective use of approved medications. In many instances, however, labeling may fail to be updated as new information about drug efficacy emerges in the postmarket setting. When labeling becomes outdated, it loses its value for prescribers and undermines a core part of the FDA's mission to communicate accurate and reliable information to patients and physicians. *Methods:* We compared the number of drug uses indicated on product labels to the number of uses contained in a leading drug compendium for 43 cancer drugs approved between 1999 and 2011. We defined a "well-accepted off-label use" of a drug as one that was not approved by the FDA and received a category 1 or 2A evidence grade. *Results:* Of the 43 drugs reviewed in this study, 34 (79%) had at least one well-accepted off-label use. In total, 253 off-label uses were identified; 91% were well accepted, and 65% were in cancer types not previously represented on labeling. Off-patent drugs had more well-accepted off-label uses than brand-name drugs, on average (mean 13.7 vs 3.8, $P = .018$). *Conclusions:* The labeling for many cancer drugs, particularly for older drugs, is outdated. Although FDA-approved labeling can never be fully aligned with real-world clinical practice, steps should be taken to better align the two when high-quality data exist. Such steps, if taken, will assist patients and prescribers in discerning which uses of drugs are supported by the highest quality evidence.

Keywords

FDA, labeling, off-label use, compendia, postmarket evidence

Introduction

Each time a new drug is approved for marketing in the United States, an accompanying collection of drug-related information, called "labeling," is made available to health care practitioners to inform safe and effective prescribing. Federal regulations state that labeling must contain a summary of the essential scientific information about a drug, and that the information contained therein must be informative and accurate.[1] The content of labeling is written by drug manufacturers, but must be approved by the Food and Drug Administration (FDA) to ensure that it meets standards laid out in regulations.[2]

Labeling is a crucial source of trusted information about prescription drugs, but it can easily become outdated if new evidence of drug effectiveness is not submitted to the FDA in a timely manner. Most often, labeling becomes outdated when high-quality scientific evidence is generated that supports a new use of a drug, but the drug's manufacturer does not file a supplemental application requesting the new use be added to the drug's labeling. This may occur because the manufacturer did not sponsor the research investigating the new use, or because the manufacturer lacked sufficient incentives to pursue a labeling expansion. Drug manufacturers are not required by law to update their products' labeling with new uses, though they may choose to do so voluntarily when they wish to market their products in new settings.[3]

Uses of drugs in patient populations or for indications that differ from those prescribed on labeling are referred to as "off-label" uses. Off-label use in oncology is common: it has been estimated that more than half of all uses of cancer drugs are beyond the scope of approved labeling.[4,5] The fact that a particular use is off-label does not preclude it from being incorporated into routine practice and covered by insurers. A policy dating back to 1993 requires Medicare to cover off-label cancer drug uses that have been deemed medically accepted by at least

[1] Friends of Cancer Research, Washington, DC, USA
[2] American University, Washington, DC, USA

Submitted 19-Sep-2017; accepted 12-Jan-2018

Corresponding Author:
Michael B. Shea, BA, Regulatory Policy Analyst, Friends of Cancer Research, 1800 M St NW Ste. 1050 S Washington, DC 20036, USA.
Email: mshea@focr.org

one federally designated drug compendium.[6] The National Comprehensive Cancer Network Drugs & Biologics Compendium (NCCN Compendium) is the most widely used compendium of oncology drugs and is used not just by Medicare but by most private insurers to guide coverage decisions.[7] The NCCN Compendium contains a collection of drug uses that have been identified based on an evaluation of the scientific literature and expert judgment and includes both on- and off-label uses.[8]

In this study, we investigate the extent to which the recommendations of medical experts who crafted the NCCN Compendium align with approved uses of drugs on FDA labeling. Although a wide disparity between labeling and the Compendium is to be expected, given the high rate of off-label use in oncology, comparing the 2 sources allows us to quantify the extent to which the labeling of individual drug products diverges from high-quality clinical practice. Furthermore, because NCCN assigns an evidence grade to each off-label use it recommends, it allows us to analyze the quality of evidence supporting off-label indications, and the diversity of the indications themselves.

Methods

Sample Construction and Data Collection

We identified all new molecular entities and new biological products approved by the FDA between January 1, 1999, and December 31, 2011, for anticancer indications. For each drug in our sample, we recorded the approved uses that were listed on labeling, which are contained in the "Indications and Usage" portion of the physician package insert and are marked with a unique numerical listing or a separate bulleted entry. We then accessed entries for the sampled drugs in the NCCN Compendium and recorded the description, disease setting, ICD-10 code, and NCCN evidence category for each recommended use.

Uses in the NCCN Compendium were divided into 2 groups based on NCCN-designated evidence categories. Uses graded category 1 or 2A were deemed to be "well-accepted" because of NCCN's assertion that these uses are supported by "uniform" consensus, meaning a majority panel vote of at least 85% is required.[9] Uses graded category 2B or 3 were not deemed to be well accepted because they lack uniform consensus from NCCN committees. Uses in the Compendium that were both well accepted and not FDA approved were assigned the category of "well-accepted off-label use."

Comparison of FDA-Approved Labels and the NCCN Compendium

We conducted a comparison of uses listed in the NCCN Compendium with uses listed on approved labeling. An NCCN-recommended use was classified as "on-label" if the following criteria were met: (1) the use was indicated for a cancer type listed on approved labeling or a subtype of a broader cancer type listed on approved labeling; *and* (2) all conditions of use mentioned on the label (eg, line of therapy, drug combinations, prior treatments, biomarker selection criteria) did not differ between NCCN's description of the recommended use and the description of the use on labeling. We then identified which products had outdated labels, defining the term "outdated label" to mean a label that was missing at least one well-accepted off-label use (ie, one use that NCCN graded as category 1 or 2A).

Classification of NCCN-Recommended Off-Label Uses

We grouped the off-label uses in the Compendium into 3 mutually exclusive categories adapted from an existing classification system.[10] The categories were (1) new disease indication; (2) modified disease indication; and (3) expanded patient population. New indications were uses in separate disease settings than those listed on the FDA label; modified indications were uses that represented a new line of therapy, a new drug combination, or a new purpose (eg, adjuvant therapy vs symptom palliation); expanded patient populations were new uses that represented closely related subtypes to already-approved indications, new age groups, and new biomarker selection criteria. Disease subtypes were clarified and terminological differences reconciled using the World Health Organization's (WHO's) ICD-10 online browser.

Statistical Analysis

We ran a series of paired and 2-sample t tests as well as a Mann-Whitney U test to evaluate differences between the number of FDA-labeled uses and NCCN-recommended uses, as well as differences between NCCN-recommended uses of different categories. For additional detail on our methods and statistical analysis, see Supplemental Information.[i] This article does not contain any studies with human or animal subjects performed by any of the authors.

Results

We identified 43 anticancer agents approved by the FDA between 1999 and 2011 (Figure 1). A total of 99 FDA-labeled uses were identified, compared to 451 NCCN-recommended uses. The average difference between the number of NCCN-recommended and FDA-labeled uses for each drug was 8.16 ($P < .001$). All FDA-labeled uses were also recommended in the Compendium, with the exception of 2 non-oncology indications for imatinib. Among the 451 NCCN-recommended uses, 198 (43.9%) were classified as on-label uses and 253 (56.1%) were classified as off-label uses. Of the 253 off-label uses in the NCCN Compendium, 26 (10.3%) were graded category 1, and 205 (81%) were graded category 2A (Table 1). Thus, 231 (91%) of uses were deemed a "well-accepted off-label use" according to our definition of the term.[ii] There was evidence that the proportion of drugs with well-accepted off-label uses is greater than the proportion of drugs with no well-accepted off-label uses ($P < .001$). Additionally, of the 253 off-label uses, 165 (65.2%) were

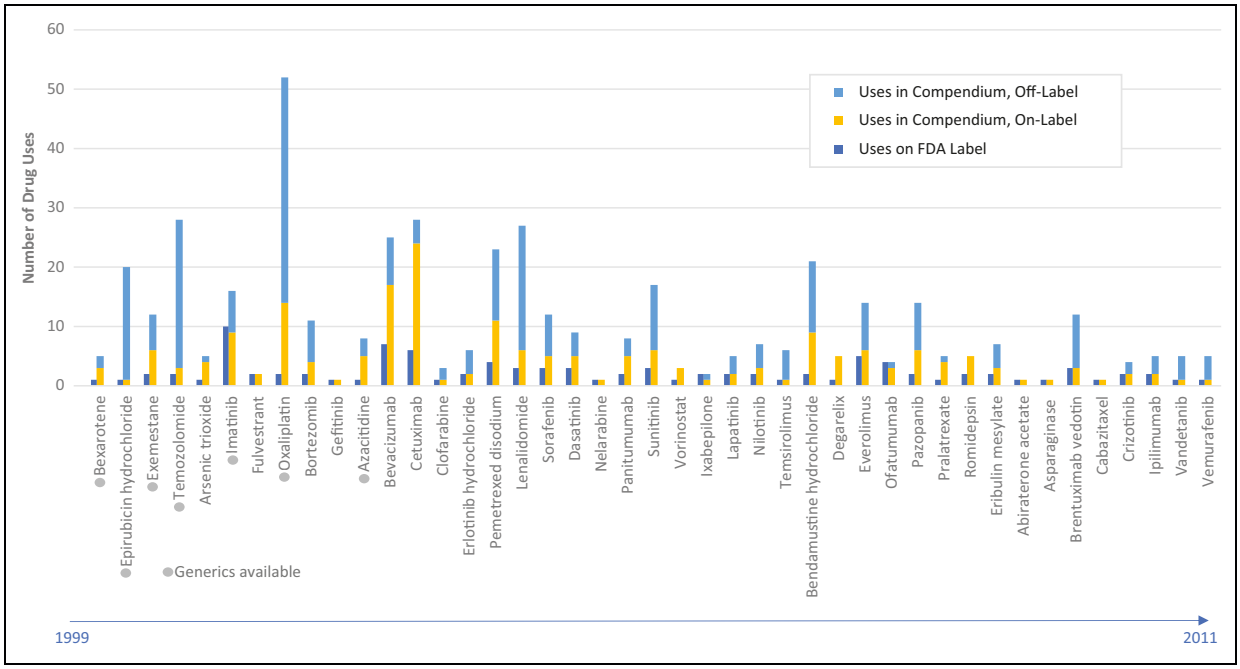


Figure 1. Oncology drug uses listed on FDA-approved labeling vs the NCCN Compendium, 1999-2011. The figure shows a comparison of FDA-approved labeling and the NCCN Compendium for 43 cancer drugs approved between 1999 and 2011. Drug uses listed in approved labeling were counted from the Indications and Usage section of physician package inserts. Drug uses listed in the NCCN Compendium were categorized as either within or outside the scope of labeling (ie, "on-label" or "off-label") through a direct comparison with uses listed on labeling. The average difference between the number of NCCN-recommended and FDA-labeled uses for each drug was 8.16 ($P < .001$). The total number of uses supported by the NCCN Compendium also differed for drugs with and without generic competition ($P = .018$).

Table 1. Characteristics of Drug Uses Included in the NCCN Compendium.[a]

	Total Uses (n = 451)		On-Label Uses (n = 198)		Off-Label Uses (n = 253)	
Characteristic	Number	Percent	Number	Percent	Number	Percent
Use category						
On-label	198	43.90	198	100	0	0
Off-label: New indication	165	36.59	0	0	165	65.22
Off-label: Modified indication	32	7.10	0	0	32	12.65
Off-label: Expanded population	56	12.42	0	0	56	22.13
NCCN evidence grade						
Category 1	81	17.96	55	27.78	26	10.28
Category 2A	339	75.17	134	67.68	205	81.03
Category 2B	25	5.54	8	4.04	17	6.72
Category 3	6	1.33	1	0.51	5	1.98

[a]The table shows the breakdown of uses recommended on the NCCN Compendium for 43 cancer drugs approved between 1999 and 2011, stratified by use category and evidence grade. Use categories were assigned to each NCCN-recommended use by the authors using a process described in the article. Evidence grades are assigned to each recommended use in the Compendium by NCCN panels. Evidence grades are defined by NCCN as follows: category 1—based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate; category 2A—based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate; category 2B—based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate; category 3—based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

categorized as "new indications," meaning they were in disease settings not represented on labels (Table 1).

Of the 43 drugs in the analysis, 34 (79.1%) had at least one well-accepted off-label use; 34.8% had at least 5 well-accepted off-label uses; and the mean number of well-accepted off-label uses was 5.4. The mean number of well-accepted off-label uses in the NCCN Compendium also differed for drugs with and without generic competition (mean 13.7 vs 3.8, $P = .018$). The difference between FDA labeling and the NCCN Compendium is further illustrated by a case

Table 2. Diversity of Disease Settings Represented in FDA Labeling and the NCCN Compendium: 3 Case Studies.[a]

	✓ = At least 1 use in disease setting					
	Eloxatin (oxaliplatin)		Avastin (bevacizumab)		Erbitux (cetuximab)	
Disease setting[b]	FDA	NCCN	FDA	NCCN	FDA	NCCN
Breast cancer				✓		
Central nervous system cancers			✓	✓		
Cervical cancer			✓	✓		
Chronic lymphocytic leukemia		✓				
Colorectal cancer	✓	✓	✓	✓	✓	✓
Esophageal cancer		✓				
Gastric cancer		✓				
Head and neck cancers					✓	✓
Hepatobiliary cancers		✓				
Kidney cancer			✓	✓		
Malignant pleural mesothelioma				✓		
Neuroendocrine tumors		✓				
Non-Hodgkin lymphoma		✓				
Non-melanoma skin cancers						✓
Non–small cell lung cancer			✓	✓		✓
Occult primary		✓				
Ovarian cancer		✓	✓	✓		
Pancreatic cancer		✓				
Penile cancer						✓
Soft tissue sarcoma				✓		
Testicular cancer		✓				
Uterine neoplasms				✓		

[a]The 3 drugs listed were initially approved by the FDA for colorectal cancer indications. Two of 3 (bevacizumab and cetuximab) were subsequently approved in additional disease settings, but for all 3 drugs, the number of disease settings represented in the NCCN Compendium is greater than what is represented on approved labeling. This chart illustrates the variety of supplemental uses that are recommended by NCCN but are not contained on approved labeling.
[b]Disease categories listed reflect NCCN's grouping of cancer types.

study of 3 drugs initially approved for colorectal cancer indications (Table 2).

A review of 5 of the largest private payers' coverage policies identified 80% (4 of 5) with policies that explicitly use the NCCN Compendium to support coverage decisions at the time of this writing. Medical and pharmacy coverage policies containing explicit reference to NCCN evidence categories accepted for coverage were obtained for all but Humana. Aetna, Cigna and United Healthcare policies accepted categories 1, 2A, and 2B; and Anthem's accepted categories 1 and 2A.[11-14] The percentage of off-label uses in the Compendium that were category 1, 2A, or 2B, and thus accepted by 3 of the 5 largest payers, was 98% (248 of 253) (Supplemental Table 1).

Discussion

Our analysis of the NCCN Compendium and FDA drug labels for 43 cancer drugs approved between 1999 and 2011 identified hundreds of off-label uses, most of which were strongly supported by NCCN expert panels. Ninety-one percent of off-label uses were "well accepted" (defined in this study as receiving a category 1 or 2A evidence grade), and 65% were for cancers not currently represented in labeling. Drugs that had gone off patent had the most well-accepted off-label uses associated with them. From these findings, we infer that the labeling of many cancer drugs is out of date, and this is especially true for older, generic products.

A review of commercial payer coverage policies further illustrates the divergence between labeling and high-quality clinical practice. We found that 4 of the 5 largest private payers, as well as Medicare, cover over 90% of uses listed on the NCCN Compendium (uses graded 1 and 2A), suggesting widespread acceptance of these uses by diverse stakeholders . While standards for FDA approval differ from standards for coverage determinations, these findings indicate that the gulf between labeled uses and covered uses may be needlessly wide.

The absence from approved labeling of many well-accepted drug uses presents a significant public health concern. Labeling is the FDA's primary means of communicating information about drugs, and as such it contains a rich supply of information about drug safety and effectiveness. But as labels fall out of date, their status as useful resources may decline, causing prescribers to rely instead on other sources of information. Labeling has already been shown to be of limited interest to many physicians, many of whom cannot accurately identify labeled indications of the medications they commonly prescribe.[15] Inattention to labeling can cause patient harm, as was seen in

the case of cisapride, when a revised label warning of life-threatening adverse events did not change prescribing behavior.[16] By the same token, overreliance on sources other than labeling, such as compendia, may result in misplaced confidence in some off-label uses. While compendia recommend many strongly supported uses of drugs, they have also been shown to recommend uses that are supported by far less rigorous evidence.[17] Therefore, unforeseen consequences for patients may arise from both the disregard of labeling and the overreliance on other sources, such as compendia.

Given that the prevalence of off-label use in oncology is well known, the existence of outdated labeling will likely not come as a surprise to many observers. However, these findings demonstrate the extent to which individual drugs are strongly recommended for many (sometimes dozens) off-label uses, and that the diversity of these uses themselves is often striking. The case studies presented in Table 2 further illustrate this point. In the case of the drug Eloxatin (oxaliplatin), the disparity between the uses recommended by NCCN and those approved by the FDA is especially stark. Eloxatin was initially approved in 2002 for relapsed metastatic colorectal cancer, and an additional use was added in 2004 for adjuvant treatment of stage III colon cancer. Since then, no new indications were added to the drug's labeling. In contrast, at the time of this analysis, the NCCN Compendium included 38 off-label uses of the drug, representing 10 additional disease settings beyond those that are approved by the FDA. This is not just true of oxaliplatin: over half of the drugs in our sample had well-accepted off-label uses in disease settings not currently represented on labeling.

Restoring the relevance of approved labeling is an important public health goal. While other high-quality sources of clinical prescribing information exist, labeling is the sole source of information that reflects the scientific and methodological rigor of the FDA approval process. Patients and prescribers can have the assurance that the use of medicines in conformity with drug labeling is supported by a positive benefit-risk assessment. The inclusion of new uses in product labeling, as appropriate, will provide patients and prescribers with these assurances of safety and effectiveness on a more frequent basis.

However, it is equally important to consider the critical role of off-label use to safe and effective prescribing. As a former editor of the *Journal of the American Medical Association* put it, "There are too many variations in clinical circumstances and too much time delay in regulations to allow the government to impede the physician's ability to [prescribe off-label] . . . when it is medically appropriate."[18] Thus, while restoring the relevance of approved labeling would foster greater trust in medical products, it should not come at the expense of lowering access to important off-label uses.

Congress recognized the importance of off-label prescribing in 1997 with the passage of the Food and Drug Administration Modernization Act (FDAMA), which described ways in which manufacturers could disseminate medical and scientific information about unapproved uses without violating the legal prohibition against off-label promotion. These "safe harbors" have been reinforced in subsequent FDA guidance documents.[19] However, the FDA has noted in these guidance documents that allowing the dissemination of information about unapproved uses is predicated on the assumption that a manufacturer would soon seek FDA approval for such unapproved uses. As such, permitting the dissemination of information about off-label uses is not intended to be a substitute to the eventual inclusion of such uses onto approved labeling.

Owing to its desire to communicate effectively with prescribers through labeling, the FDA has attempted at several points in the past 20 years to maximize labels' accessibility and usability. In 1998, the FDA issued proposed regulations aimed at helping speed the incorporation of "new uses" of approved products onto labeling.[20] Then in 2006, the FDA altered the structure of labeling to make it more user-friendly.[21] Most recently, in 2013, FDA launched the Prescription Drug Labeling Improvement and Enhancement Initiative to "enhance the safe and effective use of prescription drugs by facilitating optimal communication through labeling."[22] In total, these actions represent a concerted effort on the part of FDA to make labeling a more valuable source of prescribing information, but they have not had their desired effect.

The FDA's past attempts to achieve more up-to-date labels have not succeeded in part because responsibility to update labeling largely falls on drug manufacturers, not the FDA. Under current law, drug manufacturers can request that additional uses of their products be added to labeling by submitting supplemental new drug applications. This is a voluntary process; manufacturers are not required to update labeling with new information about drug effectiveness. Thus, manufacturers typically submit new efficacy data about previously approved drugs only if they wish to market their products for additional uses. In 2007, the Food and Drug Administration Amendments Act added new authority for FDA to require safety-related labeling changes when new safety information becomes available after approval, but no such requirement currently exists for the addition of efficacy-related information.[23]

To ensure that labeling is updated in a timely manner, drug manufacturers should be encouraged to submit more frequent supplemental applications to the FDA. Progress has recently been made on this front: the sixth reauthorization of Prescription Drug User Fee Act, passed in August 2017, eliminated user fees for supplemental applications.[24] However, since there may be scenarios in which manufactures lack any incentive to submit efficacy supplements, such as when a drug has gone off patent, the FDA may need to play a more proactive role in promoting drug labeling that is up-to-date and accurate. One method of accomplishing this would involve a collaboration between the FDA and the developers of clinical guidelines and drug compendia. The latter, who aggregate and synthesize postmarket evidence, could work with the FDA to evaluate existing evidence about approved drugs and suggest updates to labeling. Manufacturers would then be able to reference this material in supplemental applications, thus

lowering the barriers associated with the submission of such applications.

The collaboration envisioned between the FDA and clinical experts would be far less resource-intensive than a program requiring FDA to update labeling on its own. Many professional societies and guideline developers have already spent much time evaluating postmarket evidence supporting off-label drug use. Moreover, such a collaboration would result in labeling that includes new uses of drugs that are supported by strong evidence. Thus, not all the off-label uses currently recommended by NCCN should be incorporated into labeling, but rather only those that are supported by "substantial evidence" of effectiveness, a term that is defined in Section 505 of the Food, Drug and Cosmetic Act and expanded upon in federal regulations.[25,26] It is likely that many of the off-label uses recommended by NCCN would in fact meet existing evidentiary standards, given the widespread acceptance of these uses by physicians and payers, as well as frequent assertions by the FDA and others that many off-label uses have become standard of care.[27-30] The method outlined above, which would seek to encourage more frequent labeling updates by drug sponsors, may not adequately facilitate label extensions when a brand-name product has been withdrawn from the market and generic versions remain available. Existing laws requiring that generic product labels be the "same" as brand-name reference product labels, as well as ongoing concerns over product liability, complicate the initiation of labeling changes by generic firms.[31] Our analysis of NCCN guidelines has some limitations. First, it was limited to oncology drugs, although the issue of outdated labeling extends beyond this disease setting. In fact, outdated labeling may pose an even greater risk in settings where well-curated compendia do not exist, or where reimbursement is tied to the contents of labeling, as sometimes takes place in rheumatology.[32] Additionally, we did not conduct an analysis of changes to labeling or the NCCN Compendium over time. Further research into the evolution of these resources following the approval of a new drug would help illustrate how postmarket evidence is developed and identify additional opportunities to incorporate it into labeling.

Conclusions

This study provides evidence that FDA-approved labeling is missing a large amount of important and clinically relevant information about the effectiveness of cancer drugs. Labeling can be a valuable resource for prescribers, but can easily lose its utility if it becomes outdated. Over time, the presence of outdated labeling erodes the FDA's ability to communicate important prescribing information to physicians, which is a core part of the Agency's mission. Facilitating the timely addition of new drug uses to approved labeling will enable patients and prescribers to discern which uses of drugs are supported by the highest quality evidence.

Declaration of Conflicting Interests

No potential conflicts were declared.

Funding

No financial support of the research, authorship, and/or publication of this article was declared.

ORCID iD

Mark Stewart, PhD http://orcid.org/0000-0002-4847-0736

Supplemental Material

Supplementary material for this article is available online.

Notes

i. See Supplemental Information.
ii. See Methods. Uses graded category 1 or 2A were deemed to be "well-accepted" due to NCCN's assertion that these uses are supported by "uniform" consensus, meaning a majority panel vote of at least 85% is required.

References

1. US Department of Health and Human Services. Requirements on content and format of labeling for human prescription drug and biological products. 21 CFR §201.56. https://www.gpo.gov/fdsys/granule/CFR-2012-title21-vol4/CFR-2012-title21-vol4-sec201-56. Effective April 1, 2012. Accessed February 20, 2018.
2. Lindstrom J. Sources of drug information: FDA-approved labeling and other official FDA sources. *Dermatol Ther*. 2009;22:246-256.
3. Greene SM, Noah L. Debate: Off-label drug promotion and the first amendment. *Univ PA Law Rev*. 2015;62:239-267.
4. Pfister DG. Off-label use of oncology drugs: the need for more data and then some. *J Clin Oncol*. 2012;30:584-586.
5. Soares M. "Off-label" indications for oncology drug use and drug compendia: history and current status. *J Oncol Pract*. 2005;1:102-105.
6. Abernethy AP, Raman G, Balk EM, et al. Systematic review: reliability of compendia methods for off-label oncology indications. *Ann Intern Med*. 2009;150:336-343.
7. American Society of Clinical Oncology. Reimbursement for cancer treatment: coverage of off-label drug indications. *J Clin Oncol*. 2006;24:3206-3208.
8. National Comprehensive Cancer Network. About the NCCN Drugs & Biologics Compendium (NCCN Compendium®). Accessed December 5, 2017.
9. National Comprehensive Cancer Network. NCCN Guidelines® & Clinical Resources: development and update of the NCCN Guidelines®. https://www.nccn.org/professionals/development.aspx. Accessed March 23, 2017.
10. Wang B, Kesselheim AS. Characteristics of efficacy evidence supporting approval of supplemental indications for prescription drugs in United States, 2005-14: systematic review. *BMJ*. 2015;351:h4679.

11. National Comprehensive Cancer Network. NCCN Recognition Program. https://www.nccn.org/professionals/recognition.asp. Accessed May 2017.
12. Anthem. Clinical UM Guideline: Oncology drug treatment regimens for adults. https://www11.anthem.com/ca/medicalpolicies/guidelines/gl_pw_c166610.htm. Accessed February 2017.
13. Cigna. Cigna drug and biologic coverage policy: oncology medications. https://cignaforhcp.cigna.com/public/content/pdf/coveragePolicies/pharmacy/ph_1403_coveragepositioncriteria_oncology.pdf. Accessed February 2017.
14. United Healthcare. Oncology clinical medication coverage policy. https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Tools%20and%20Resources/Policies%20and%20Protocols/Medical%20Policies/Drug%20Policies/Oncology_Medication_Clinical_Coverage_Policy.pdf. Accessed February 2017.
15. Chen DT, Wynia MK, Moloney RM, et al. U.S. physician knowledge of the FDA-approved indications and evidence base for commonly prescribed drugs: results of a national survey. *Pharmacoepidemiol Drug Saf*. 2009;18:1094-1100.
16. Smalley W, Shatin D, Wysowski DK, et al. Contraindicated use of Cisapride: impact of food and drug administration regulatory action. *JAMA*. 2000;284:3036-3039.
17. Green AK, Wood WA, Basch EM. Time to reassess the cancer compendia for off-label drug coverage in oncology. *JAMA*. 2016; 316:1541-1542.
18. Klein D, Tabarrok A. Do off-label drug practices argue against FDA efficacy requirements? A critical analysis of physicians' argumentation for initial efficacy requirements. *Am J Econ Sociol*. 2008;67:743-776.
19. United States Food and Drug Administration. Guidance for industry: distributing scientific and medical publications on unapproved new uses—recommended practices. Rockville, MD: FDA, 2014.
20. Dissemination of information on unapproved/new uses for marketed drugs, biologics, and devices. *Fed Regist*. 1998;63:31143.
21. United States Food and Drug Administration. Guidance for industry: labeling for human prescription drug and biological products—implementing the PLR content and format requirements. Rockville, MD: FDA, 2013.
22. Prescription drug labeling improvement and enhancement initiative; request for comments and information. *Fed Regist*. 2013;78:8446.
23. US Department of Health and Human Services. Food and Drug Administration Amendments Act of 2007. Public Law 110–85. Effective September 27, 2007. https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf. Accessed February 20, 2018.
24. United States Food and Drug Administration. Guidance for industry: assessing user fees under the Prescription Drug User Fee Amendments of 2017. Rockville, MD: FDA, 2017.
25. Food and Drugs, 21 USC §355.
26. US Department of Health and Human Services. Adequate and well-controlled studies. 21 CFR §314.126. https://www.gpo.gov/fdsys/granule/CFR-2010-title21-vol5/CFR-2010-title21-vol5-sec314-126. Effective April 1, 2010. Accessed February 20, 2018.
27. United States Food and Drug Administration. Guidance for industry: responding to unsolicited requests for off-label information about prescription drugs and medical devices. Rockville, MD: FDA, 2011.
28. United States Food and Drug Administration. Guidance for industry: good reprint practices for the distribution of medical journal articles and medical or scientific reference publications on unapproved new uses of approved drugs and approved or cleared medical devices. Rockville, MD: FDA, 2009.
29. Cranston JW, Williams MA, Nielsen NH, et al; for the Council on Scientific Affairs. Report of the Council on Scientific Affairs: unlabeled indications of Food and Drug Administration-approved drugs. *Drug Inf J*. 1998;32:1049-1061.
30. Radley DC, Finkelstein SN, Stafford RS. Off-label prescribing among office-based physicians. *Arch Intern Med*. 2006;166: 1021-1026.
31. Silverman E. FDA again delays rule to allow generic drug makers to change labels. *Stat News*. https://www.statnews.com/pharmalot/2016/05/19/fda-generic-drugs-safety/. Published May 19, 2016. Accessed December 2017.
32. American College of Rheumatologists. Position statement: FDA labels. https://www.rheumatology.org/Portals/0/Files/Position-Statement-FDA-Labeling.pdf. Accessed September 2017.

FRIENDS of CANCER RESEARCH

FRIENDS OF CANCER RESEARCH

1800 M Street NW, Suite 1050 South
Washington, DC 20036

Tel: 202.944.6700 *Email:* info@focr.org

@CancerResrch
@FriendsofCancerResearch
www.focr.org