

2022 SCIENTIFIC REPORT

Regulatory Advancements for Patients

FRIENDS
of CANCER
RESEARCH

3 INTRODUCTION

4 2022 BY THE NUMBERS

**15 PATIENT-FOCUSED DRUG DEVELOPMENT: ENSURING
PATIENT-CENTERED TRIAL DESIGNS**

- 16 Supporting a Patient-Centric Approach to Dose Optimization in Oncology: The Essential Role of Patient-Reported Outcomes (PROs)
- 28 Need for Aligning the Definition and Reporting of Cytokine Release Syndrome (CRS) in Immuno-oncology Clinical Trials
- 36 Accelerating Investigation of New Therapies in Earlier Metastatic Treatment Settings

**55 REAL-WORLD EVIDENCE: LEVERAGING RWD TO ADVANCE
RESEARCH**

- 56 Multi-Cancer Early Detection Screening Tests: Considerations for Use of Real-World Data
- 79 Exploring the Potential of External Control Arms created from Patient Level Data: A case study in non-small cell lung cancer

**95 INNOVATIVE DRUG DEVELOPMENT: EVALUATING LESSONS
LEARNED TO INFORM CONTINUED PROGRESS**

- 96 Evaluating the impact of the RACE act: An interim analysis
- 97 Breakthrough Therapy Designation Criteria Identify Drugs that Improve Clinical Outcomes for Patients: A Case for More Streamlined Coverage of Promising Therapies
- 101 Impact of the COVID-19 Pandemic Mitigation Strategies on Cancer Clinical Trials: Preliminary Findings of a *Friends*-ASCO Study
- 119 Expedited Development of Diagnostics for Therapies Targeting Rare Biomarkers or Indications

**129 COMPLEX BIOMARKERS: GENERATING EVIDENCE TO SUPPORT
ALIGNMENT IN DRUG DEVELOPMENT**

- 130 Homologous Recombination Deficiency: Concepts, Definitions, and Assays
- 138 Assessing Variability Across HRD Assays: Findings from The *Friends*' HRD Harmonization Project
- 139 Changes in ctDNA Reflect Clinical Benefit Across Multiple Studies of Patients with Advanced Non-Small Cell Lung Cancer Treated with Immune Checkpoint Inhibitors
- 152 Circulating Tumor DNA in Development of Therapies For Cancer: An Evidentiary Roadmap to an Early Endpoint for Regulatory Decision Making

Introduction

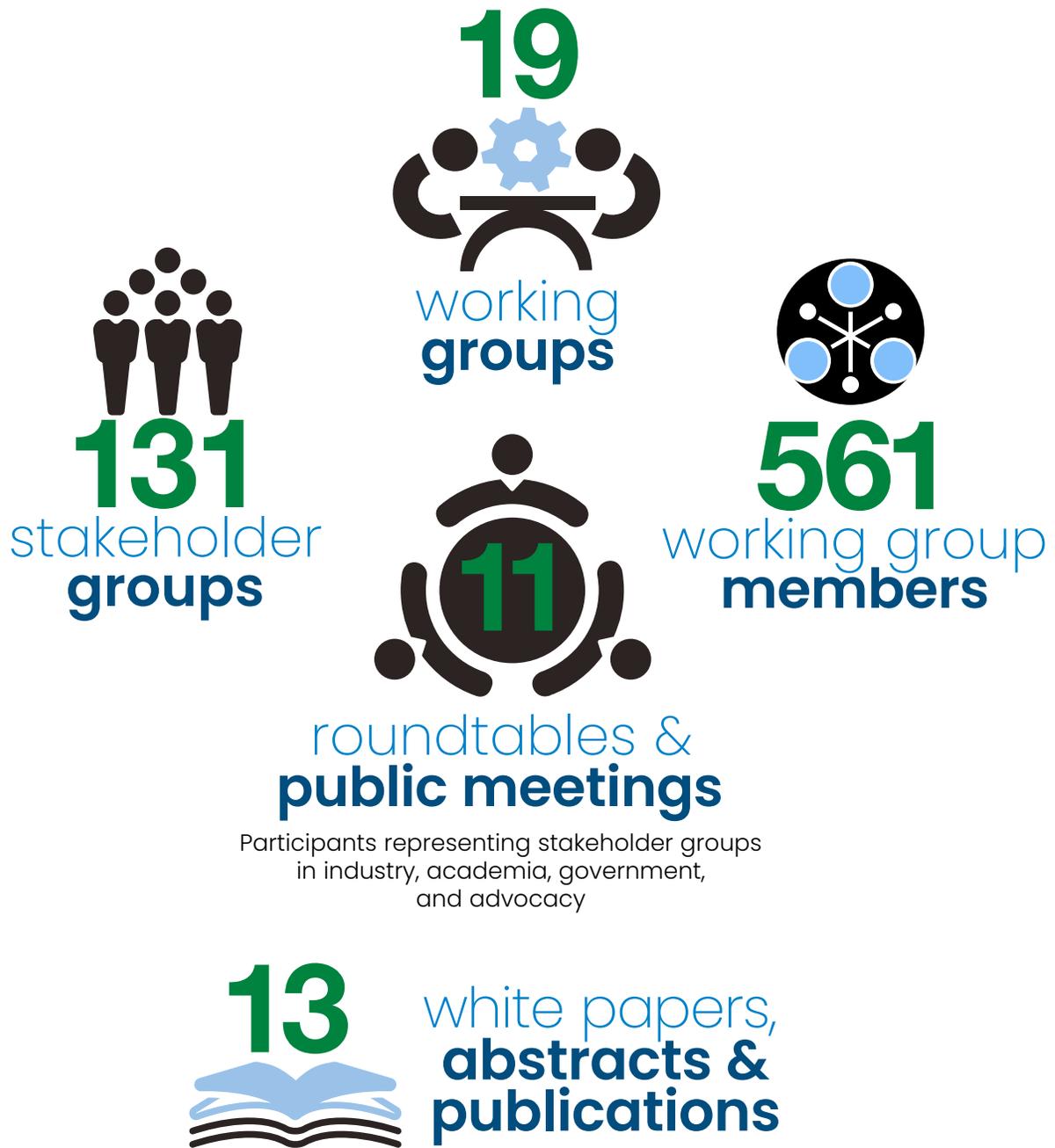
In 2022, Friends of Cancer Research (*Friends*) led critical work in science, policy, and regulation to catalyze meaningful improvements in oncology drug development, legislative and regulatory policy, and patient care. *Friends* unites scientists, experts, advocates, and patients throughout the year to collaboratively generate timely solutions through working groups, roundtables, and scientific conferences.

This year, the U.S. Food and Drug Administration (FDA) released several guidance documents, including guidance on the use of circulating tumor DNA (ctDNA), real-world data and evidence (RWD/E), and a tissue agnostic approach to drug development in oncology, demonstrating the importance of alignment on strategies and evidentiary needs in these areas. Several of *Friends*' recently launched research partnerships enable *Friends* to play a leading role in this work by generating novel data to inform solutions to current challenges. Notably, *Friends*' ctMoniTR project, which investigates the use of ctDNA as an endpoint to measure treatment response, fulfills the evidentiary needs for using ctDNA as an early endpoint described in FDA guidance. This year, *Friends* developed a roadmap to use ctDNA as an early endpoint in drug development and highlight these efforts in our Project Spotlight (page 12), which includes the launch of a new collaborative project comparing ctDNA baseline levels across cancer types, disease stages, and ctDNA assays.

Other research partnerships include projects that identify clinically useful endpoints in real-world data (RWE Pilot) and harmonize complex biomarkers to optimize test reliability and accuracy (HRD Harmonization). Outputs from these research partnerships, in addition to our working groups, roundtables, and scientific conferences, are captured in this scientific report. The 2022 Scientific Report contains the full text of our 2022 white papers and publications to serve as a resource for those in the drug development and regulatory space [and can be found online using the QR code on the cover]. Our white papers and publications focused on several key themes:

- ❶ **PATIENT-FOCUSED DRUG DEVELOPMENT:** Ensuring patient-centered trial designs
- ❷ **REAL-WORLD EVIDENCE:** Leveraging RWD to advance research
- ❸ **INNOVATIVE DRUG DEVELOPMENT:** Evaluating lessons learned to inform continued progress
- ❹ **COMPLEX BIOMARKERS:** Generating evidence to support alignment in drug development

2022 By the Numbers



Participants representing stakeholder groups
in industry, academia, government,
and advocacy

Patient-Focused Drug Development

Introduction

Over the past 10 years, FDA's patient-focused drug development program has increased awareness of the need to incorporate the patient voice into drug development and clinical trial designs. One way of doing so is through the collection of patient experience data, including patient reported outcomes (PROs), which are data collected directly from the patient about their experience with their disease and treatment. In 2015, *Friends* convened stakeholders to discuss the need for alignment and use of tools to collect PROs and identified the National Cancer Institute (NCI)'s PROs Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) symptom library as a resource to collect symptomatic side effects. Then, in 2018, *Friends* gathered key stakeholders to develop a more holistic definition of tolerability that incorporates the patient experience by collecting PROs. At the *Friends'* 2022 Annual Meeting, discussions highlighted the need for including PROs in early phase studies to understand tolerability and inform dosage selection. Future work will focus on opportunities for implementing and interpreting PRO data in early phase clinical trials to inform dosage selection.

In addition to PROs, consistent identification, characterization, and reporting of clinically relevant toxicities contribute to a comprehensive understanding of a therapy's risk/benefit profile. Greater availability and use of immuno-oncology therapies has helped identify unique toxicities associated with their use such as cytokine release syndrome (CRS). In 2022, *Friends* proposed several evidence-based strategies for capturing and reporting CRS and associated adverse events (AEs) which will support more consistent identification, monitoring, and management of CRS and enable accurate comparisons of these toxicities across therapies.

172% growth
in cancer cell
immunotherapy
research and
development from
2019 – 2022.
Therefore, it
is important to
enhance safety
data collection
for optimal
patient care and
outcomes.

Cytokine Release Syndrome

is the most
common toxicity
associated with
use of CAR T-cell
therapies with
an incidence
ranging between

17%-94%
of patients.

>61%

of novel oncology drugs approved since 2013 were for patients with metastatic disease who had received prior therapy.

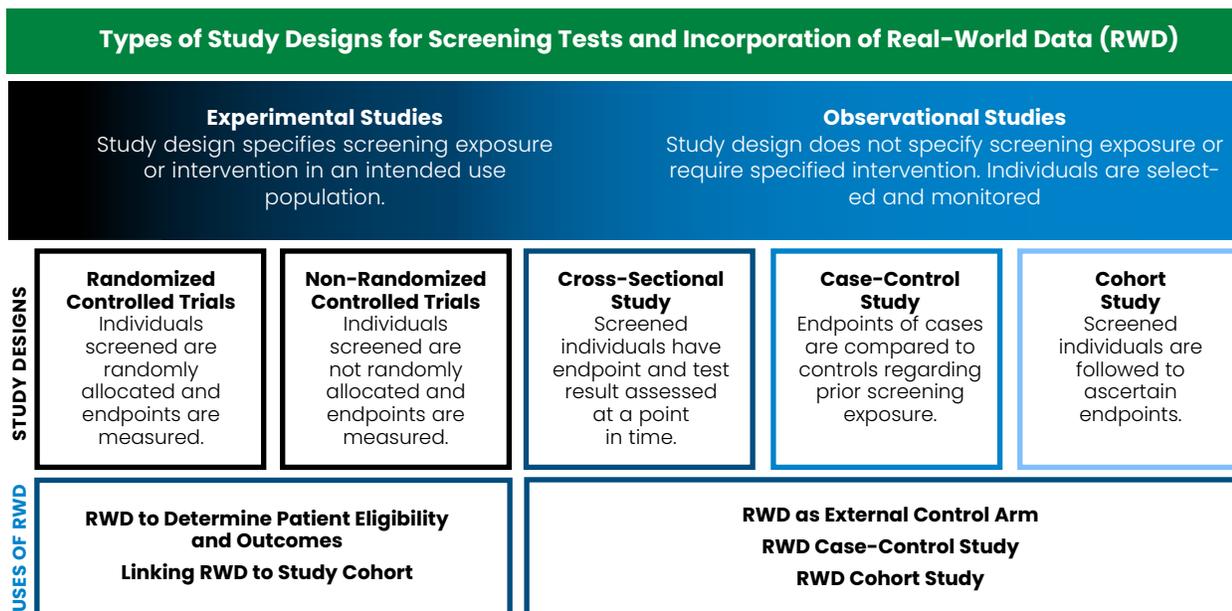
Another area focused on moving therapies into earlier lines of treatment. Traditionally, oncology therapeutics are investigated and approved in late line metastatic settings, however, moving treatments into earlier lines of therapy earlier in clinical development would allow for more patients to be treated and at a time when cancer outcomes are better. As such, FDA established Project FrontRunner, which proposes opportunities for studying therapies in early metastatic settings. At the *Friends'* Annual Meeting in 2022, *Friends* convened key stakeholders to identify opportunities and challenges related to Project FrontRunner.

2022 GUIDANCE RELEASED BY FDA RELATED TO PATIENT-FOCUSED DRUG DEVELOPMENT

- Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments, Draft Guidance, June 2022
- Characterizing, Collecting, & Reporting Immune-Mediated Adverse Reactions in Cancer Immunotherapeutic Clinical Trials, Draft Guidance, October 2022
- Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials, Draft Guidance, April 2022

Real-World Evidence

Randomized controlled trials (RCTs) are the gold standard for generating data and evaluating the safety and efficacy of medical products. However, in certain disease settings or populations, performing a RCT may not be feasible or ethical. Real-world data (RWD) provides one opportunity to supplement clinical trial data to support clinical development programs and inform the safety and efficacy of a product. Additionally, as clinical trials often enroll a well-defined and more narrow patient population than the patients who will likely receive



the product in the real world, RWD can generate additional evidence that reflects a larger and more diverse patient population than is included in clinical trials.

In 2022, *Friends* shared the findings from a multi-stakeholder collaborative initiative to identify opportunities and highlight key considerations for use of RWD to generate supplemental evidence for regulatory decision-making for multi-cancer early detection (MCED) screening tests. MCED screening tests hold promise in detecting cancer-associated signals at early stages of cancer, including for cancers without standard of care screening modalities. Another opportunity for use of RWD to augment RCTs explored by *Friends* was the use of external control arms (ECAs). *Friends* worked in a multi-stakeholder group to publish a manuscript highlighting the feasibility of using an ECA, generated from patient-level data from previously conducted clinical trials or RWD, to mirror the overall survival results from a randomized control in the same indication. Our work supports the thoughtful and robust use of RWD to generate clinically meaningful evidence to support the evaluation of novel therapies and tests for patients with cancer.

2022 GUIDANCE RELEASED BY FDA RELATED TO REAL-WORLD EVIDENCE

- Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drug and Biological Products, Final Guidance, September 2022

In 2022,
FDA approved
12
novel drugs
for treatment of
10 different cancers.

>50%
(6/12) of novel
oncology
drugs approved
in 2022 were
first-in-class
drugs.

Innovative Drug Development

As innovation in oncology drug development continues, opportunities to get safe and effective treatments to patients efficiently requires thoughtful approaches to clinical trial design. In 2022, *Friends* shared findings from multiple collaborative initiatives focused on how regulatory flexibilities and policy changes could improve clinical trial designs and expand patient access.

Friends conducted several analyses to evaluate how policy changes over the past decade shaped the current landscape of oncology clinical trials, therapies, and drug development paradigms, while identifying further opportunities for advancements. For the 10-year anniversary of the Breakthrough Therapy Designation (BTD), *Friends* demonstrated that products with BTD improved outcomes for patients with cancer and identified an opportunity to establish a timely process for initiating coverage and reimbursement decision-making. In 2022, *Friends* also assessed how the Research to Accelerate Cures and Equity (RACE) Act has impacted the investigation of therapies in pediatric patients one-year post-implementation and found an increase in approved cancer therapeutics that require pediatric investigation. These assessments demonstrate the value of new legislation in improving patient access to innovative treatments.

As innovative drugs are developed, regulatory flexibility and novel approaches to clinical trial design are critical. The COVID-19 pandemic allowed for greater flexibility of clinical trial designs, and *Friends* established a project with the American Society of Clinical Oncology (ASCO) and a task force of thought leaders to characterize the flexibilities that clinical trial sponsors incorporated during the pandemic to work towards determining whether these changes impacted data integrity. The hope is that the findings from this project will support an understanding of which flexibilities are most important to carry forward.

31% vs. 15%

BTD Drugs

Non-BTD Drugs

Improvement in overall survival (OS)
over standard of care

2 years **prior** to
the RACE Act,

0%

**of cancer
drugs**

were required
to conduct
pediatric studies.

2 years **after**
the RACE Act,

50%

**of cancer
drugs**

were required
to conduct
pediatric studies.

2022 GUIDANCE RELEASED BY FDA RELATED TO INNOVATIVE DRUG DEVELOPMENT

- Considerations for Rescinding a Breakthrough Therapy Designation, Draft Guidance, June 2022
- Ethical Considerations for Clinical Investigations of Medical Products Involving Children, Draft Guidance, September 2022
- Human Gene Therapy Products Incorporating Human Genome Editing, Draft Guidance, March 2022

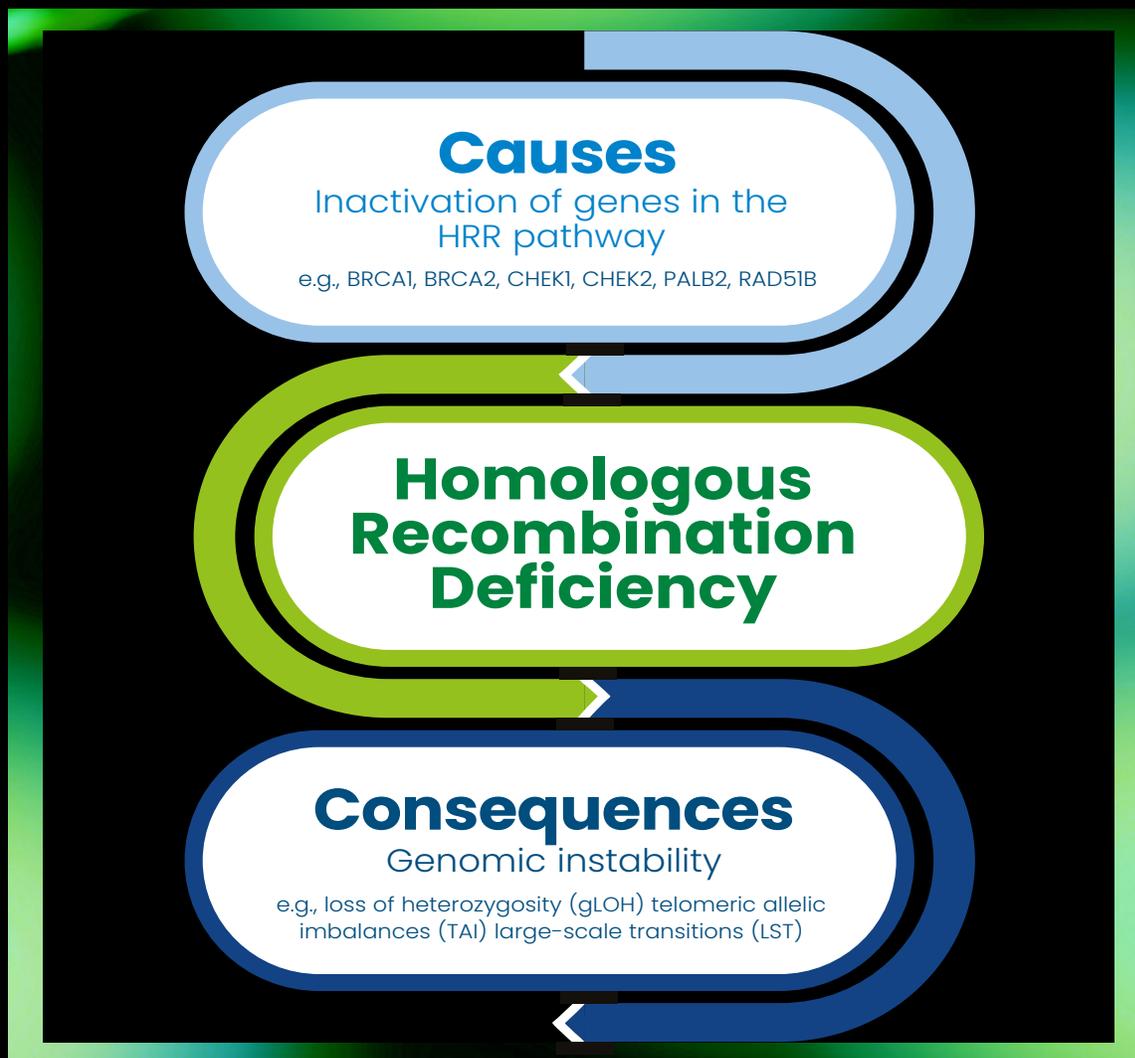
Complex Biomarkers

Innovative drug development has led to the advent of targeted therapies, which work most effectively when the patient's tumor has a specific biomarker present. Biomarkers can also be used to understand how a tumor responds to treatment over time. To identify biomarkers, doctors use diagnostic tests, which may have variable test performance and discordance results from test to test due to fragmented regulatory oversight system.

Homologous recombination deficiency (HRD) is a biomarker used to identify patients with certain cancers who would benefit from certain treatments. The *Friends'* HRD Harmonization Project kicked off in 2022 with a goal of examining sources of variability across HRD tests and identifying opportunities for alignment while proposing solutions to improve agreement in the future. *Friends* published a landscape assessment of HRD tests, then used that foundation to build a project focused on assay alignment with over 20 diagnostic developers and other key stakeholders. Initial findings from the project were presented at the Association for Molecular Pathology conference in November 2022. Additional results are forthcoming in 2023.

Circulating tumor DNA (ctDNA) is fragments of DNA shed from cancer cells, found in the bloodstream, and collected using a blood draw. Emerging science demonstrates ctDNA may be useful for diagnosing and tracking a patient's cancer. However, before ctDNA measurement is used in clinical practice and for regulatory decisions, additional evidence needs to be generated. The ctDNA for Monitoring Treatment Response (ctMoniTR) Project is a meta-analysis to understand whether changes in ctDNA are predictive of outcomes. Findings from Step 1 of ctMoniTR were published in 2022 and *Friends* brought together key stakeholders to align on a roadmap to identify key data necessary to support the use of ctDNA as an early endpoint in clinical trials.

Friends brought together key stakeholders to perform a landscape assessment of homologous recombination deficiency (HRD) and its use as a biomarker in cancer. A key output was clarifying the approach to defining HRD. Diagnostics developers identify causes and consequences of HRD by analyzing genetic material (i.e., DNA, RNA).



2022 GUIDANCE RELEASED BY FDA RELATED TO COMPLEX BIOMARKERS

- Tissue Agnostic Drug Development in Oncology, Draft Guidance, October 2022
- Use of Circulating Tumor DNA for Early-Stage Solid Tumor Drug Development, Draft Guidance, May 2022

Project Spotlight: ctDNA Portfolio Development and Milestones

Goal

ctDNA holds promise for measuring treatment efficacy in clinical trials. Friends of Cancer Research (*Friends*) is working to establish an aligned strategy for developing the necessary data to support the use of ctDNA as an early endpoint for treatment response for regulatory decision-making and leading a multi-stakeholder group to generate this data. Validating the use of ctDNA as an endpoint will accelerate research by enabling rapid identification of effective new cancer therapies and ultimately allow them to reach patients sooner.

Background

The introduction of novel therapeutics, especially targeted therapies, has changed the paradigm for treating solid tumors. While these new therapies provide increased clinical benefit for patients, the concomitant increase in survival time creates a unique challenge in the expedient evaluation of new therapies. Traditional clinical trial designs using long-term clinical outcome endpoints such as progression-free survival (PFS) or overall survival (OS) may not allow for an efficacy determination in a timely manner. The use of ctDNA levels as an early endpoint represents an emerging opportunity to assess efficacy earlier. However, it is critical to obtain robust data to fully qualify and validate ctDNA as an early endpoint for long-term clinical outcomes in solid tumors.

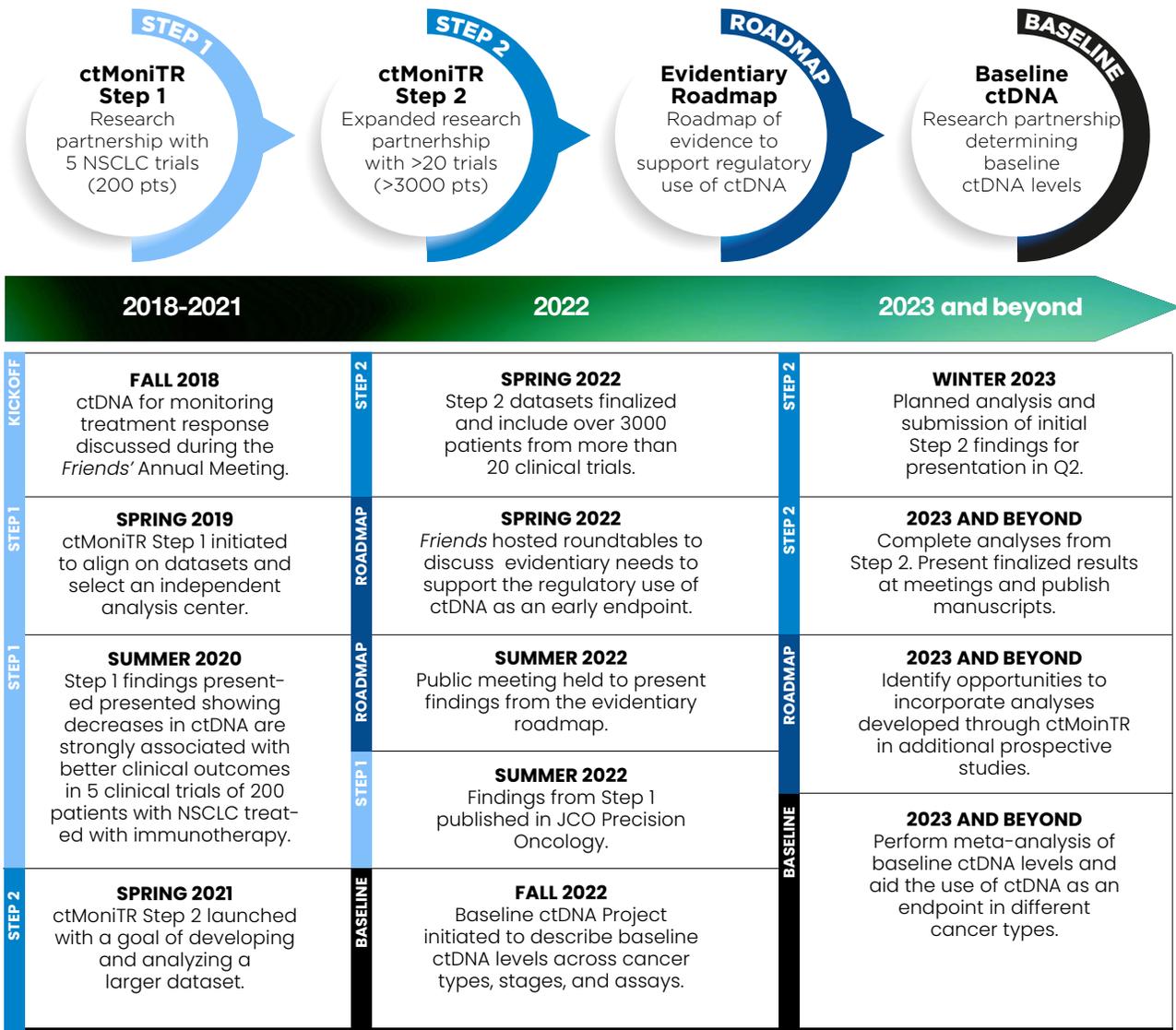
Approach

Establishing the necessary evidence to support the use of ctDNA as an early endpoint requires a multiprong approach:

- **CTDNA EVIDENTIARY ROADMAP:** In 2022, *Friends* coordinated a group of stakeholders to develop an aligned strategy for generating data and evidence. Findings demonstrate there are multiple technical and clinical characteristics contributing to variability in ctDNA measurements that should be adequately accounted for when conducting validation studies.
- **THE CTDNA TO MONITOR TREATMENT RESPONSE (CTMONITR) PROJECT:** This first of its kind partnership seeks to answer the important question: Do changes in ctDNA reflect response to treatment? Step 1 of the project kicked off in early 2019 and included data from 5 clinical trials representing 200 patients with advanced non-small cell lung cancer treated with PD(L)-1 inhibitors. *Friends* worked with stakeholders to establish and implement an analysis approach conducted by Cancer Research And Biostatistics (CRAB). Findings from the study published in the summer of 2022 demonstrate that changes in ctDNA levels associate with treatment outcomes: increases in ctDNA levels associate with poor outcomes while decreases in ctDNA associate with better outcomes. Step 1 showed that harmonizing data across trials with different assays and time points is feasible and set the stage for the ongoing Step 2 project, which expands the approach to rather than into more patients, trials, additional cancer types, and treatments.

- **BASELINE CTDNA LEVELS PROJECT:** Findings from the ctDNA Evidentiary Roadmap highlight a need to evaluate the landscape of ctDNA detection in different cancer types and stages to provide insights into the extent to which findings can be generalized across early- and late-stage cancer settings, as well as across assay technologies. Through a collaborative effort involving multiple diagnostic developers, *Friends* seeks to establish evidence regarding baseline (i.e., pre-treatment) sensitivity metrics for ctDNA detection across cancer types, stages, and assays. This greater understanding of the biological landscape of baseline ctDNA levels will help inform a conceptual framework for the use of ctDNA as an early endpoint predictive of long-term outcomes.

Findings from our continued work in this space will be consolidated and presented in public meetings and peer-reviewed literature in the future. Our hope is that ctDNA will ultimately be used to support regulatory decisions to provide safe and effective treatments to patients faster.





Patient-Focused Drug Development: Ensuring patient-centered trial designs

Supporting a Patient-Centric Approach to Dose Optimization in Oncology: The Essential Role of Patient-Reported Outcomes (PROs)

FRIENDS OF CANCER RESEARCH ANNUAL MEETING 2022

Introduction

Patient experience data (PED) in the context of drug regulation is a growing part of the totality of evidence to understand the safety and efficacy of a cancer therapeutic. PED intends to provide information about patients' experiences with a disease or condition.¹ One type of PED, patient-reported outcomes (PROs), is a clinical outcome assessment based on information directly reported by the patient about the status of their own health condition. Patients are uniquely positioned to report their own quality of life, symptoms, and function, and several studies support that patients are a highly reliable reporting source of such information that adds value to the traditional clinician assessment.² For example, clinicians, including oncologists, may overestimate functional status and underestimate patient symptoms, supporting the clinical and scientific value of PROs for quantifying symptomatic adverse events (AEs) with the greatest impact on patient health-related quality of life while on the therapy.³

Recent US Food and Drug Administration (FDA) draft guidance highlights the need for benefit-risk planning when developing new oncology drug and biologic products, including collecting appropriate data to inform the dose exposure response for efficacy and safety/tolerability.⁴ Oncology clinical trials commonly consider clinician-reported safety data, dose modifications, dose discontinuations, and severe AEs including hospitalizations to determine tolerability. In 2018, Friends of Cancer Research (*Friends*) gathered key stakeholders to develop a new working definition of treatment tolerability that incorporates the patient experience by collecting rigorously developed PRO data to inform symptomatic toxicity and functional information.⁵ The group aligned on the position that a complete understanding of tolerability should include direct patient measurement on how they are feeling and functioning while on treatment. Integrating PROs early in drug development, alongside traditional measures, can support a more comprehensive understanding of the benefits and risks of a therapeutic, including the perception of the patient on the tolerability of the therapy and their ability or desire to adhere to the dose or intensity of therapy for prolonged periods. This can add unique data to inform dose optimization, or dose range, for new oncology drugs.

Thank You to Our Contributors

Vishal Bhatnagar, MD, FDA

Corina Dutcus, MD, Eisai Inc.

Serban Ghiorghiu, MD, AstraZeneca

Paul Kluetz, MD, FDA

Lee Jones, MBA, Patient Advocate

Kirstin R. McJunkins, M.Ed., Patient Advocate

Daniel O'Connor, MBChB, PhD, Medicines and Healthcare products

Regulatory Agency (MHRA)

Devin Peipert, PhD, Northwestern University Feinberg School of Medicine

Ashley F. Slagle, MS, PhD, Aspen Consulting, LLC

Hillary Stires, PhD, Friends of Cancer Research

Peter C. Trask, PhD, MPH, Genentech

To adequately implement this expanded definition of tolerability, drug sponsors need to select and deploy fit-for-purpose PROs using well-defined and reliable tools at an assessment frequency appropriate for the drug.⁶ Several symptom libraries are available, and one tool developed specifically to capture symptomatic side effects is the National Cancer Institute (NCI)'s Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) symptom library.⁷ *Friends* identified this tool in 2015 as a potential resource, and many of the earlier operational challenges delaying its commercial use have been addressed since then. The PRO-CTCAE can be used across various trial contexts in multinational settings to provide patient-reported symptomatic toxicity information that complements standard clinician-reported CTCAE safety data. In addition to individual symptom data, tolerability can be informed by other widely available PRO measurement systems to evaluate side effect impacts including patient-reported overall side effect bother, physical function, and ability to work and carry out leisure activities. Physical function can be measured in a variety of ways, including using the Patient-Reported Outcomes Measurement Information System (PROMIS)[®] physical function item bank or the functional scales from the European Organization for the Research and Treatment of Cancer item library (EORTC QLQ-F17).

While PRO use is common in randomized registrational trials, PRO collection in early phase trials is rare.⁸⁻¹² There is increasing interest across stakeholders involved in early phase cancer trials in using PRO data as valuable complementary information to inform tolerability and later phase trial design.¹³ FDA has emphasized the need to collect PROs in cancer clinical trials by offering suggestions for core PROs and how to measure them in draft guidance.⁶ Regulatory authorities in other jurisdictions are also placing increasing emphasis on the collection of PRO data in a variety of settings.¹⁴ While there are limited examples of use of PROs in dose finding trials, there is interest in identifying feasible approaches for the collection and use of these data in early phase clinical trials, particularly to inform dose selection.^{6,15}

FDA announced Project Optimus in 2021, which seeks to place a greater emphasis on dose optimization and dose selection in early phases of oncology drug development towards doses that maximize the efficacy, safety, and tolerability of a drug.¹⁶ Considering the expanded definition of tolerability for all therapies, current paradigms that focus on identifying the *maximum* tolerated dose (MTD) are not appropriate with newer targeted therapies that show relevant efficacy across a range of doses, allowing for better tolerability and potential adherence at doses lower than the MTD. PROs should be included in early phase studies to provide a foundational understanding of short- and longer-term symptom and functional impacts to optimize dosing and facilitate more informed development in later phases. Better understanding of the tolerability of different doses throughout early drug development could inform selection of a dose or doses for approval that patients are more likely to be able to take following approval. Collection of PROs for dose optimization encompasses both the first in human Dose Escalation trials, as well as the Dose-Expansion trials.¹⁷

Friends convened industry, academic, regulatory, and patient advocate representatives to discuss opportunities and challenges for using PROs in early phase clinical trials, specifically focused on measuring tolerability to inform dosage optimization. Open questions regarding PRO inclusion in early phase trials include feasibility, trial design, impacts of sample size, optimal PRO

selection, and PROs influencing results (e.g., over/under reporting clinician AEs).¹³ Including PRO assessment in early phase trials requires careful thought and consideration to fully realize the value of symptom and functional data while overcoming operational challenges.

To identify opportunities, challenges, and solutions for using PROs to inform dose optimization in cancer clinical trials, the working group focused on the following objectives:

- Highlight ways PROs can characterize tolerability and support dosage optimization
- Provide a clinical trial design framework for incorporating PROs into dosage optimization studies
- Discuss opportunities for using PRO findings from early phase studies to inform later phase study designs and to complement traditional safety data for regulatory decision-making

Using PROs to Complement Commonly Collected Data to Inform Dose Selection

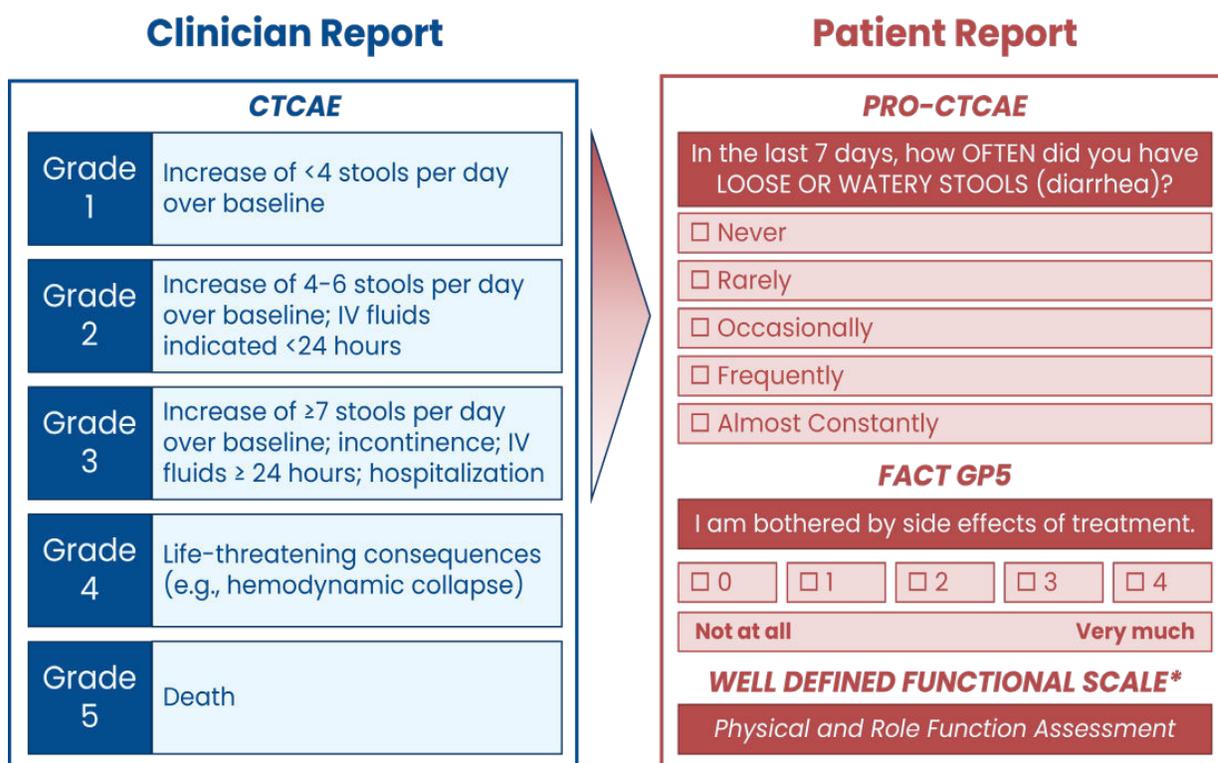
Using PROs to Complement Commonly Collected Data to Inform Dose Selection

PROs provide unique information characterizing specific symptoms, overall side effect burden, and their impact on a patient's ability to function. The systematic nature of PRO collection informs onset, duration, severity, and resolution of side effects and their impacts. Many oncology drugs require long-term administration of therapy to maintain tumor response and control. As such, the tolerability of the drug may shift as patients transition from the immediate to the long-term phase of treatment. For example, even lower grade (Grades 1 and 2) symptomatic AEs may become more burdensome than infrequent Grade 3 toxicity, especially when multiple prolonged lower grade symptomatic toxicities are experienced simultaneously.⁵ While standard clinician-reported AE data provide a rate of worst grade AEs experienced at any time during the clinical trial to characterize tolerability, additional granularity about the severity, frequency, duration, and impacts of side effects can be elucidated from PROs.

Figure 1 highlights an example of how systematically captured PRO data assessing symptoms and functional impacts can complement clinician reporting. In this example, clinician-reported Grade 1-2 diarrhea is considered "low grade," which can minimize the patient's perceived impact of the symptom and may not be considered when defining safety and tolerability of a product in a clinical trial. For diarrhea, the clinician report for lower grade AEs may obscure a wider range of symptom severity and impact on function, which is true of other symptomatic toxicities (e.g., visual, cutaneous, and mucosal side effects). The PRO-CTCAE data of the patient report of diarrhea can provide additional insight, which can be further expanded by assessing how bothered the patient is by side effects overall using an item such as the Functional Assessment of Cancer Therapy (FACT) GP5 item ("I am bothered by side effects of treatment") and by how they report their physical function and ability to work and perform leisure activities using patient-reported functional assessments. Studies have shown that patients with more frequent diarrhea also have worse physical function, which is seen consistently at time intervals over the course of treatment.¹⁸ The additional, longitudinal information provided by PROs can

add granularity and trajectory for the individual symptom as well as potential impacts to how bothered patients are, their function, and medication adherence.¹⁹ This information is not discernable by clinician report alone, particularly for Grade 1-2 clinician AEs.

Figure 1: Example of clinician-reported data and patient-reported data. Patient-reported data can add information to expand on lower grade clinician rated CTCAE side effects (i.e., Grade 1-3). For instance, a CTCAE grade 2 diarrhea event could be considered by a patient as almost constant diarrhea that results in high side effect bother and adverse impacts on physical function and role function (i.e., ability to work or carry out leisure activities).

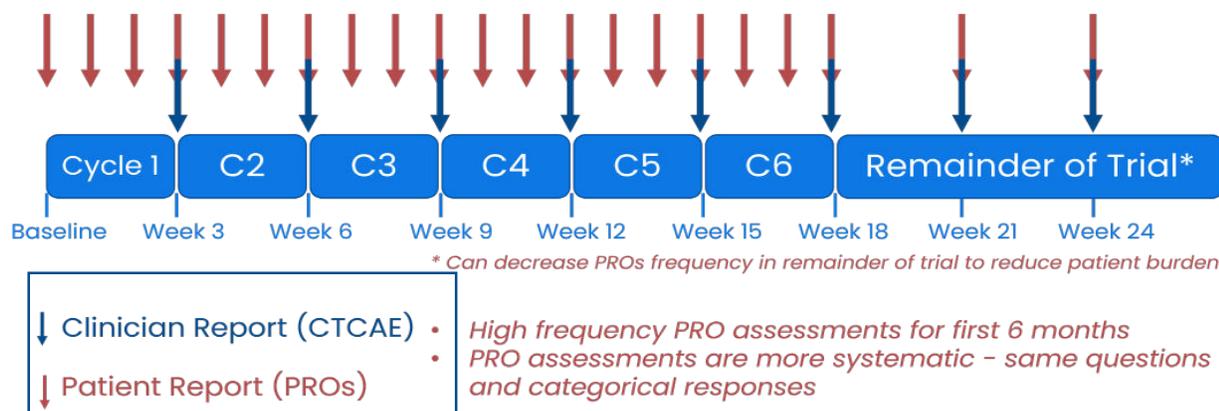


*Two examples of well-defined PRO scales for physical and/or role function are available from EORTC and PROMIS® measurement systems.

In addition to providing data on outcomes not currently assessed (i.e., side effect bother and functional outcomes), the frequency and systematic assessment of PRO data can reveal smaller but potentially important differences in a symptomatic side effect. Several studies note patients report a higher incidence and severity of symptomatic side effects than clinicians' CTCAE evaluation.^{3,20,21} The added ability for PROs to inform safety and tolerability is most clear for unobservable symptoms such as neuropathy, headache, pruritis, nausea, and constipation, where assessments most differ when a patient reports their experience compared with their provider.²² Additionally, providers measure side effects only during clinic visits, which may lead a patient to provide an assessment that does not represent their full experience of side effect

intensity over the course of the 3–4 week cycle (**Figure 2**).²³ This is particularly problematic for intermittently administered therapies, like chemotherapies administered once every 3 weeks, where maximal side effect intensity typically occurs between clinic visits. In these situations, an investigational oral drug administered daily may appear to be tolerated worse than a once every 3 week chemotherapy if assessments are being performed only once per cycle/ office visit where chemotherapy side effects have begun to resolve. More frequent systematic assessment can be valuable for exposure–response analyses related to a cardinal toxicity, and this added power becomes particularly important with the smaller cohorts evaluated in Dose Escalation trials.

Figure 2: Frequency of PRO assessments. PRO symptom data provides more consistent data capture by asking the same question with categorical response options at a higher frequency. This data source can add power to exposure–response analyses during Dose Escalation study. High frequency PRO assessment can be reduced later in trial by asking a comprehensive PRO assessment at several longer-term cross-sectional time points (e.g., 1 year, 2 year, etc.). Adapted from figure courtesy of Zirkelbach, Bhatnagar, and Kluetz.



A Framework to Incorporate PROs into Dose Optimization Studies

PRO data should be collected with approaches that reduce bias, with well-defined and reliable measures, and in ways that the results can be easily interpreted to complement findings from clinician-reported outcomes. FDA’s draft guidance recommends collecting and analyzing five core PROs: disease-related symptoms, symptomatic adverse events, overall side effect impact summary measure, physical function, and role function.⁶ Because there are differing expected toxicities across therapeutic drug classes, individualized symptom item lists should be selected from an item library rather than using an off-the-shelf static questionnaire. For example, NCI’s PRO-CTCAE is a robust item library that contains pertinent patient-reported symptoms. Other item libraries exist including both the EORTC item library and Functional Assessment of Chronic Illness Therapy (FACIT) item library.^{24,25} Single item and summary measures such as overall side effect burden (e.g., FACT GP5) can complement patient-reported symptomatic AEs and

have the advantage of capturing the impact of multiple different toxicities in a single score comparable across groups of patients and different treatments. Functional impacts can be assessed by EORTC or PROMIS®.²⁶

Dose Escalation Trials

Identification of an appropriate dose starts with first in human trials, often called Dose Escalation trials. PROs should be incorporated into Dose Escalation trials to gather a holistic understanding of the patient's toxicity profile. Dose Escalation trials often include a single agent, allowing for analysis of the PROs from the effect of the new agent, rather than impacts from combination therapies in later trials impacting interpretation. Descriptive trends in the severity and duration of symptomatic AEs should be evaluated among patients in these trials to help sponsors understand whether symptomatic toxicities increase with increased doses, new symptoms emerge at higher doses, or if the frequency of lower grade symptoms increases. This information can be used to inform the Dose Expansion study, including the potential of moving forward with a recommended dosage range.

Given sample size is lowest in Dose Escalation trials, high frequency systematic assessment of PRO symptom data can add power to pharmacokinetic exposure-response analyses conducted during dose-escalation. In addition, comprehensive knowledge of the patient population can clarify which PROs support an understanding of the impacts of the treatment to inform later phase PRO selection. Because many of the patients have disease that is refractory, baseline PROs should be taken before treatment initiation to normalize for symptoms of disease or side effects from prior treatments. Selecting a series of PROs related to expected AEs when there are fewer patients included in Dose Escalation studies can help narrow in on key PROs to measure in subsequent studies.

When selecting PROs for Dose Escalation studies, sponsors should consider data from pre-clinical studies, as well as side effect evidence from other drugs in the same drug class when available. A free text PRO item could be included in Dose Escalation trials when the potential treatment related symptoms are not fully known and can inform which PROs to include in subsequent trials. A single item side effect impact question like FACT GP5 can add additional information with little additional patient burden. While some of these measures, including overall side effect burden, function, and key expected symptomatic AEs, should be measured in all patients, trial designs that adapt PRO measures in later cohorts based on symptoms identified in early cohorts could be considered.

Dose Expansion Studies

The Dose Expansion study has historically focused on whether the efficacy signal warrants an additional study by using a single-arm trial with a single dose (often the MTD).²⁷ It is increasingly important to use Dose Expansion trials to optimize the dose, ideally by conducting a randomized evaluation of two or more doses. These early randomized evaluations of dose would not need to strongly control Type I error, but rather be sufficiently sized to make assessments regarding the activity/efficacy, safety, and tolerability of the different doses.^{17,28} Inclusion of PROs would be instrumental in the assessment of tolerability and describing differences in tolerability among the candidate doses. To support dosing decisions in these studies, PRO data including

descriptive trends in severity and duration of symptomatic side effects should be evaluated. Systematic high frequency PRO data adds unique clinical outcome data and additional evidence to describe differences in candidate doses being compared.

Sponsors should be thoughtful when designing Dose Expansion studies and consider PRO question selection that ensures relevance and minimizes duplication. It is generally recommended to focus on relevant treatment-related symptoms. The core outcomes recommended by FDA, including symptoms, overall side effect impact, and physical and role function, should be assessed with available tools in 30 or less questions, and ideally patients should not spend significant time to complete PROs at each assessment (e.g., no more than 10 minutes per assessment timepoint).²⁹ Sponsors should consider including a free text item in Dose Expansion studies to allow additional patient feedback on side effects and support an understanding of optimal PRO selection for subsequent trials.³⁰ Per FDA draft guidance, assessments made more frequently in the first few treatment cycles would be suitable across most drug development programs but this should be tailored to the treatment schedule.⁶ It is recommended to consult with appropriate regulatory authorities early for advice on the PRO strategy including assessment frequency for a specific drug development context.

Additional Considerations for Early Phase Dose Optimization Studies

The patient population and the treatment regimen should be considered when deciding about PRO inclusion in Dose Escalation and Dose Expansion studies. In situations when a randomized approach will not be used but safety and tolerability remain important objectives (e.g., a Dose Escalation or Dose Expansion Trial), PROs should be thoughtfully included.³¹

An important trial design decision will be whether to allow clinicians access to PRO data during the trial. While use of PROs to inform clinical care is an active area of research, their incorporation into clinical workflow is challenging and there is no regulatory requirement that PRO data be reviewed during the trial to inform patient care. Therefore, PROs collected during dose optimization studies as part of the study protocol to inform trial results may not be shared with the trial clinician in real-time to impact care. This should be clearly explained to patients and strategies should be identified to share study-level PRO data with the community once the trial is complete. In addition, there is no requirement to compare clinician and patient reports for the same or similar side effect. Patient-reported symptom data is assessed and quantified differently than CTCAE data, and differences are expected between patient and clinician report. These and other regulatory considerations for use of PROs to inform tolerability have been previously described.³² Currently, Project Patient Voice has focused on presenting data from registrational trials comparing two trial arms, but analysis and visualization techniques from the approach provide suggestions of how to do so for dose optimization studies and other trial designs to provide patients and providers with information about tolerability of different doses.³³

Conclusions

PROs should be included in early phase clinical trials to better understand tolerability and to inform dose selection for future clinical trials and clinical use as well as aiding with the selection of most appropriate PRO instruments for the late phase trials with registration intent. The use of PROs in Dose Escalation and Dose Expansion studies is a newer concept and sponsors should continue to refine approaches for incorporating PROs as methodologies and analyses are improved over time. It will be critical to educate stakeholders about the value of and approach to including PROs. Additionally, careful consideration should be given to which PROs to include based on tumor type, stage of disease, evidence from similar treatments, previously collected data, and trial goals.

While major progress over the last decade has provided the necessary tools to measure PROs in clinical trials, there are still some limitations to using PRO data in clinical trials. For instance, many clinical trials are multinational, and it is critical to ensure culturally validated translations are available. This may not be a significant challenge given many early phase trials are conducted in the US and almost all widely used PRO measures have English and Spanish translations. Additionally, item libraries are being iteratively improved and may not include or have optimized all novel symptomatic side effects of interest. For instance, some of the side effects included in the PRO-CTCAE do not have measures for more than one attribute (i.e., severity, frequency, and interference), and the PRO-CTCAE does not include an overall side effect bother item. This makes it likely that multiple PRO measurement systems may need to be deployed within a single trial.

PRO results should be considered in the context of other data to establish a totality of evidence alongside clinician reports of safety and efficacy, pharmacokinetics, pharmacodynamics, and biomarker studies. PROs can be of particular value as a high-quality data stream to support exposure response relationships in dose optimization decision-making. It is acknowledged that addition of PROs in early drug development is new and will add some cost to early development. However, PRO data are uniquely positioned to add value to the characterization of tolerability which is a critical study objective in dose escalation and optimization. Additional information to better optimize dose may lead to a more tolerable marketed drug with an optimum benefit-risk profile, with better patient-reported information on side effects and positive impacts that can have advantages in adherence and provide important information for decision making by the patient and treating clinician.

References

1. U.S. Food and Drug Administration. Patient-Focused Drug Development Glossary. <https://www.fda.gov/drugs/development-approval-process-drugs/patient-focused-drug-development-glossary>
2. Atkinson TM, Ryan SJ, Bennett A, et al. The Association Between Clinician-Based Common Terminology Criteria for Adverse Events (CTCAE) and Patient-Reported Outcomes (PRO): A Systematic Review. *Supportive Care in Cancer* 2016 24:8. 2016;24(8):3669–3676. [doi:10.1007/s00520-016-3297-9](https://doi.org/10.1007/s00520-016-3297-9)
3. Laugsand EA, Sprangers MAG, Bjordal K, et al. Health Care Providers Underestimate Symptom Intensities of Cancer Patients: A Multicenter European Study. *Health Qual Life Outcomes*. 2010;8(1):1–13. [doi:10.1186/1477-7525-8-104/TABLES/5](https://doi.org/10.1186/1477-7525-8-104/TABLES/5)
4. U.S. Food and Drug Administration. Benefit-Risk Assessment for New Drug and Biological Products Guidance for Industry. September 2021. <https://www.fda.gov/media/152544/download>
5. Basch E, Campbell A, Hudgens S, et al. A Friends of Cancer Research White Paper: Broadening the Definition of Tolerability in Cancer Clinical Trials to Better Measure the Patient Experience. 2018. https://friendsofcancerresearch.org/wp-content/uploads/Comparative-Tolerability-Whitepaper_FINAL.pdf
6. U.S. Food and Drug Administration. Core Patient-Reported Outcomes in Cancer Clinical Trials Draft Guidance for Industry. June 2021. <https://www.fda.gov/media/149994/download>
7. Basch E., Campbell A, Globe, D, et al. Capturing Symptomatic Adverse Events from the Patients' Perspective: The Potential Role of the National Cancer Institute's PRO-CTCAE Measurement System. Issue Brief from the Conference on Clinical Cancer Research. November 2015. <https://friendsofcancerresearch.org/wp-content/uploads/FINAL-PRO-CTCAE-pre-conference-draft-1.pdf>
8. Lai-Kwon J, Yin Z, Minchom A, Yap C. Trends in Patient-Reported Outcome Use in Early Phase Dose-Finding Oncology Trials – An Analysis Of ClinicalTrials.gov. *Cancer Med*. 2021;10(22):7943. [doi:10.1002/CAM4.4307](https://doi.org/10.1002/CAM4.4307)
9. Coleman RL, Beck JT, Baranda JC, et al. The Use of Patient-Reported Outcome Measures in Phase I Oncology Clinical Trials. *Oncology*. 2021;99(7):444–453. [doi:10.1159/000514874](https://doi.org/10.1159/000514874)
10. Bergerot CD, Pal SK, Tripathi A. Patient-Reported Outcomes in Early Phase Clinical Trials: An Opportunity to Actively Promote Patient-Centered Care. *Oncologist*. 2022;27(9):714–715. [doi:10.1093/ONCOLO/OYAC122](https://doi.org/10.1093/ONCOLO/OYAC122)

11. Fiteni F, Le Ray I, Ousmen A, et al. Health-Related Quality of Life As an Endpoint In Oncology Phase I Trials: A Systematic Review. *BMC Cancer*. 2019;19(1):1-8. [doi:10.1186/s12885-019-5579-3/TABLES/2](https://doi.org/10.1186/s12885-019-5579-3/TABLES/2)
12. Watson GA, Veitch ZW, Shepshelovich D, et al. Evaluation of the patient experience of symptomatic adverse events on Phase I clinical trials using PRO-CTCAE. *British Journal of Cancer* 2022 127:9. 2022;127(9):1629-1635. [doi:10.1038/s41416-022-01926-z](https://doi.org/10.1038/s41416-022-01926-z)
13. Lai-Kwon J, Vanderbeek AM, Minchom A, et al. Using Patient-Reported Outcomes in Dose-Finding Oncology Trials: Surveys of Key Stakeholders and the National Cancer Research Institute Consumer Forum. *Oncologist*. 2022;27(9):768-777. [doi:10.1093/ONCOLO/OYAC117](https://doi.org/10.1093/ONCOLO/OYAC117)
14. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the Patient Experience Into Regulatory Decision Making in the USA, Europe, and Canada. *Lancet Oncol*. 2018;19(5):e267-e274. [doi:10.1016/S1470-2045\(18\)30097-4](https://doi.org/10.1016/S1470-2045(18)30097-4)
15. Retzer A, Aiyegbusi OL, Rowe A, et al. The Value of Patient-Reported Outcomes in Early-Phase Clinical Trials. *Nature Medicine* 2022 28:1. 2022;28(1):18-20. [doi:10.1038/s41591-021-01648-4](https://doi.org/10.1038/s41591-021-01648-4)
16. U.S. Food and Drug Administration. FDA's Project Optimus. <https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus>
17. Jain L, Pithavala YK, Rahman A, et al. A Friends of Cancer Research White Paper: Optimizing Dosing in Oncology Drug Development. 2021. https://friendsofcancerresearch.org/wp-content/uploads/Optimizing_Dosing_in_Oncology_Drug_Development.pdf
18. Chen TY, King-Kallimanis BL, Merzoug L, et al. US Food and Drug Administration Analysis of Patient-Reported Diarrhea and Its Impact on Function and Quality of Life in Patients Receiving Treatment for Breast Cancer. *Value in Health*. 2022;25(4):566-570. [doi:10.1016/J.JVAL.2021.09.007](https://doi.org/10.1016/J.JVAL.2021.09.007)
19. Kluetz PG, Kanapuru B, Lemery S, et al. Informing the Tolerability of Cancer Treatments Using Patient-Reported Outcome Measures: Summary of an FDA and Critical Path Institute Workshop. *Value in Health*. 2018;21:742-747. [doi:10.1016/j.jval.2017.09.009](https://doi.org/10.1016/j.jval.2017.09.009)
20. Montemurro F, Mittica G, Cagnazzo C, et al. Self-evaluation of Adjuvant Chemotherapy-Related Adverse Effects by Patients with Breast Cancer. *JAMA Oncol*. 2016;2(4):445-452. [doi:10.1001/JAMAONCOL.2015.4720](https://doi.org/10.1001/JAMAONCOL.2015.4720)
21. Mavić MP, Šeparović R, Vuger AT, Vazdar L. Difference in Estimation of Side Effects of Chemotherapy between Physicians and Patients with Early-Stage Breast Cancer: The Use of Patient Reported Outcomes (PROs) in the Evaluation of Toxicity in Everyday Clinical Practice. *Cancers (Basel)*. 2021;13(23). [doi:10.3390/CANCERS13235922](https://doi.org/10.3390/CANCERS13235922)

22. di Maio M, Gallo C, Leighl NB, et al. Symptomatic Toxicities Experienced During Anticancer Treatment: Agreement Between Patient and Physician Reporting in Three Randomized Trials. *Journal of Clinical Oncology*. 2015;33(8):910–915. [doi:10.1200/JCO.2014.57.9334](https://doi.org/10.1200/JCO.2014.57.9334)
23. Thanarajasingam G, Hubbard JM, Sloan JA, Grothey A. The Imperative for a New Approach to Toxicity Analysis in Oncology Clinical Trials. *JNCI: Journal of the National Cancer Institute*. 2015;107(10). [doi:10.1093/JNCI/DJV216](https://doi.org/10.1093/JNCI/DJV216)
24. EORTC – Quality of Life: Item Library. <https://qol.eortc.org/item-library/>
25. FACIT Searchable Library. <https://www.facit.org/facit-searchable-library>
26. Griffiths P, Peipert JD, Leith A, et al. Validity of a Single-Item Indicator of Treatment Side Effect Bother in a Diverse Sample of Cancer Patients. *Supportive Care in Cancer*. 2022;30(4):3613–3623. [doi:10.1007/s00520-022-06802-3/TABLES/5](https://doi.org/10.1007/s00520-022-06802-3/TABLES/5)
27. Torres-Saavedra PA, Winter KA. An Overview of Phase 2 Clinical Trial Designs. *Int J Radiat Oncol Biol Phys*. 2022;112:22–29. [doi:10.1016/j.ijrobp.2021.07.1700](https://doi.org/10.1016/j.ijrobp.2021.07.1700)
28. Shah M, Rahman A, Theoret MR, Pazdur R. The Drug-Dosing Conundrum in Oncology –When Less Is More. *New England Journal of Medicine*. 2021;385(16):1445–1447. [doi:10.1056/NEJMp2109826](https://doi.org/10.1056/NEJMp2109826)
29. Albaba H, Barnes TA, Veitch Z, et al. Acceptability of Routine Evaluations Using Patient-Reported Outcomes of Common Terminology Criteria for Adverse Events and Other Patient-Reported Symptom Outcome Tools in Cancer Outpatients: Princess Margaret Cancer Centre Experience. *Oncologist*. 2019;24(11):e1219–e1227. [doi:10.1634/theoncologist.2018-0830](https://doi.org/10.1634/theoncologist.2018-0830)
30. Trask PC, Dueck AC, Piau E, Campbell A. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events: Methods for item selection in industry-sponsored oncology clinical trials. *Clinical Trials*. 2018;15(6):616–623. [doi:10.1177/1740774518799985](https://doi.org/10.1177/1740774518799985)
31. Bhatnagar V, Kluetz PG. Encouraging Rigorous Patient-Generated Data All Along the Drug Development Continuum. *JNCI: Journal of the National Cancer Institute*. 2022; 114(10):1313–1314. [doi:10.1093/jnci/djac129](https://doi.org/10.1093/jnci/djac129)
32. Kim J, Singh H, Ayalew K, et al. Use of PRO Measures to Inform Tolerability in Oncology Trials: Implications for Clinical Review, IND Safety Reporting, and Clinical Site Inspections. *Clinical Cancer Research*. 2018;24(8):1780–1784. [doi:10.1158/1078-0432.CCR-17-2555](https://doi.org/10.1158/1078-0432.CCR-17-2555)
33. U.S. Food and Drug Administration. Project Patient Voice <https://www.fda.gov/about-fda/oncology-center-excellence/project-patient-voice>



Full Length Article
Translational Research

Need for aligning the definition and reporting of cytokine release syndrome (CRS) in immuno-oncology clinical trials

Mark D. Stewart^{1,*}, Bruce McCall², Marcelo Pasquini³, Allen S. Yang⁴, Carolyn D. Britten⁵, Meredith Chuk⁶, R Angelo De Claro⁶, Bindu George⁶, Nicole Gormley⁶, Mary M. Horowitz³, Eric Kowack⁴, Candice McCoy⁷, Phuong Khanh Morrow⁵, Emmanuel Okoye⁸, Rosanna Ricafort⁷, John Rossi⁹, Elad Sharon¹⁰, Marc Theoret⁶, Ferdinando Vegni⁷, Tai Yu⁵, Jeff Allen¹

¹ Friends of Cancer Research

² Genentech, A Member of the Roche Group

³ Medical College of Wisconsin, Center for International Blood and Marrow Transplant Research

⁴ Xencor

⁵ Amgen

⁶ Food and Drug Administration

⁷ Bristol Myers Squibb

⁸ Regeneron Pharmaceuticals

⁹ CERO Therapeutics

¹⁰ National Cancer Institute

ARTICLE INFO

Article History:

Received 18 October 2021

Accepted 21 January 2022

Keywords:

Cytokine Release Syndrome (CRS)

Immunotherapies

T cell receptors

Drug development

ABSTRACT

As cancer immunotherapies continue to expand across all areas of oncology, it is imperative to establish a standardized approach for defining and capturing clinically important toxicities, such as cytokine release syndrome (CRS). In this paper, we provide considerations for categorizing the variety of adverse events that may accompany CRS and for recognizing that presentations of CRS may differ among various immunotherapies (e.g., monoclonal antibodies, CAR T cell therapies and T cell engagers, which can include bispecific antibodies and other constructs). The goals of this paper are to ensure accurate and consistent identification of CRS in patients receiving immunotherapies in clinical studies to aid in reporting; enable more precise evaluation of the therapeutic risk–benefit profile and cross-study analyses; support evidence-based monitoring and management of important toxicities related to cancer immunotherapies; and improve patient care and outcomes. These efforts will become more important as the number and variety of molecular targets for immunotherapies broaden and as therapies with novel mechanisms continue to be developed.

© 2022 Published by Elsevier Inc. on behalf of International Society for Cell & Gene Therapy. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

The emergence of cancer immunotherapies has led to transformational advances across solid and hematologic malignancies, bringing new hope to patients with serious, life-threatening diseases. Cancer immunotherapies provide clinically beneficial alternatives and additions to traditional cytotoxic treatments. Recent U.S. Food and Drug Administration (FDA) approvals of immunotherapies and the rapid

expansion into new indications for existing agents are enabling broader availability of immunotherapies to cancer patients.

The immuno-oncology (IO) drug development pipeline continues to grow, and cancer immunotherapies are quickly being integrated into the standard of care for many cancers [1]. Importantly, our increasing clinical experience with these immunotherapeutic agents has brought greater awareness to several toxicities unique to immunotherapies that are not typically observed with traditional cytotoxic agents. With the success of newer immunotherapies such as T cell engagers and chimeric antigen receptor (CAR) T cells in several hematologic malignancies, there has been growing recognition of cytokine release

* Corresponding author:

E-mail address: mstewart@focr.org (M.D. Stewart).

syndrome (CRS) as a distinct clinical entity. Cytokine release syndrome represents one of the most common toxicities of these therapies and occurs with varying frequency, severity and presentation among immunotherapeutic agents [2]. The incidence of CRS is relatively low for conventional monoclonal antibodies, but there is a higher risk of CRS with CAR T cell therapies and T cell engagers (incidence ranging from 17% to 94% for all grades) [3]. Early in the development of immunotherapies, the term CRS was used more generally to describe a syndrome with a dramatic presentation requiring intensive care, but we now understand that CRS presents with a spectrum of severities, ranging from a self-limited low-grade fever to serious multiorgan collapse.

Although CRS is increasingly recognized as an on-target effect associated with CAR T cells and T cell engagers, the full extent of this syndrome, including pathophysiology and effects on end-organ function, has not been fully characterized. A standardized approach is needed for diagnosing CRS and its manifestations in clinical trials and for reporting CRS in both prescribing information and published literature. In addition, with the advent of T cell engagers and other IO agents, there is an increasing need to distinguish CRS from other clinical entities, such as acute infusion-related reactions (IRR), septic shock, or hemophagocytic lymphohistiocytosis (HLH). For instance, acute IRRs and CRS can have overlapping symptoms and temporality, but likely have different pathophysiology and differ in management and prognosis. Our current mechanistic understanding of these overlapping clinical entities continues to evolve, and concerted efforts to harmonize data capture will help better characterize these events to uncover key differential features and inform development of individualized mitigation strategies, as appropriate. The importance of capturing adverse events (AEs) in a systematic and harmonized manner has been highlighted by the emergence and growing recognition of immune-effector cell-related AEs observed with immune checkpoint inhibitors and their management, which has been a focus in recent clinical guidelines published by the Society for Immunotherapy in Cancer and the American Society for Clinical Oncology [4,5]. Inconsistent or inadequate characterization of these toxicities in clinical trials impact how data are presented in publications and prescribing information, potentially resulting in suboptimal representation of these clinical events. This, in turn, can put patients at risk if their treatment side effects are not appropriately managed.

Growing Clinical Experience of Infusion Reactions and CRS

Adverse events known broadly as IRRs have long been defined, diagnosed and reported in an ambiguous and inconsistent manner [6]. This arises, in part, from the fact that the term IRR came into use at a time when few biological therapies were available and acute reactions to an infusion of a biologic agent were starting to be reported. Additionally, little was known about the exact mediators involved in these reactions. Since the introduction of therapeutic monoclonal antibodies and other biologics into clinical practice, IRR has been used as a broad term to encompass acute findings during or shortly after an infusion that may include hypersensitivity/anaphylaxis, complement activation-related pseudoallergy (CARPA), CRS, or more nonspecific signs and symptoms [6]. During clinical development, IRRs are generally defined as AEs occurring within the first 24 h after infusion of a therapy, with causality deemed by the investigator to be related to the therapy. This operational definition has resulted in the term IRR being used to define a wide array of symptoms with potentially disparate pathophysiology whose main commonality is occurrence within 24 h of infusion. The majority of IRRs reported with therapeutic monoclonal antibodies are self-limited and treated symptomatically [7–10]. Infusion-related reactions after CAR T cell administration are infrequent and generally mild. Nevertheless, with the emergence of T cell-engaging therapeutics, in particular T cell engagers and other IO agents, distinguishing CRS from

IRR has been a challenge, in that the signs and symptoms may partially overlap.

CRS is a supraphysiologic response driven by the immune system. It has been commonly observed in sepsis and other infections, including those related to COVID-19, and as an on-target AE of T cell-mediated therapies or in response to other therapies such as COVID-19 mRNA vaccines [11]. CRS is initiated by the activation of T cells and mediated by cytokines produced by macrophages and other myeloid cells. CRS can occur within several hours to days after infusion of an immunotherapeutic, but typically does not present beyond 14 days after initiation of therapy [12]. CRS can be short-lived, but often lasts for several days. Because symptoms of CRS can overlap with other toxicities that have generally been classified as IRRs, and because both CRS and IRR can occur within a day after infusion, careful examination of the signs and symptoms, their attribution, and the response to therapy is important. The presentation of CRS may differ depending on the immunotherapeutic and the clinical and biological status of the patient. Factors (therapy- and patient-dependent) include tumor antigen target, location of tumor (*i.e.*, blood vs. solid tumor) and target antigen or T cell binding potency. In addition, the timing of the onset of CRS can coincide closely with infusion of T cell engagers. However, for cellular products, T cell expansion precedes the onset of CRS, and there is therefore a lag between infusion and CRS symptom onset [13].

CRS typically presents with a fever and may progress to hypotension or hypoxia. Flushing and rash may accompany both anaphylactic reactions and CRS, although specific skin and mucosal changes such as hives and mucosal swelling predominate in anaphylactic reactions, occurring in 80% of cases [14]. An underlying hallmark associated with CRS is the release of cytokines, and this has been identified as a differentiating criterion in the Common Terminology Criteria for Adverse Events (CTCAE) v.5 definition for CRS and IRRs. However, the measurement of cytokines is not yet a routine element in clinical practice, nor are there reliable cytokine thresholds for CRS diagnosis. Thus, this distinction alone may not yet be helpful to clinicians at the bedside, and emergent interventions are still largely based on the clinical manifestations and severity of CRS as well as response to therapeutic interventions. For example, the role of the interleukin (IL)-6 pathway in CAR T cell therapy has been characterized, and therefore use of IL-6 blocking agents has become a mainstay interventional treatment of CRS [15,16].

CRS Definition and Severity

In light of our evolving clinical experience with emerging immunotherapeutics, several efforts have been made to update and harmonize grading criteria for CRS in clinical trials (Table 1). Additionally, the elements described in each grading system offer information on what defines severity.

Fever is a CRS-defining characteristic but does not dictate the severity of CRS. Therefore, the Memorial Sloan Kettering Cancer Center (MSKCC) grading system initially relied on the availability of cytokine levels measured from patients in real time to distinguish severe versus nonsevere CRS [23]. However, real-time cytokine testing may be limited to specific health care research settings, and there is currently poor correlation between pre-/posttreatment cytokine levels and the severity of CRS signs and symptoms. Thus, the presence and severity of hypotension and hypoxia are most commonly used to assign the grade of severity for CRS, as these two events typically drive the need for higher level of care (*e.g.*, intensive care) and clinically relevant sequelae. One unique aspect of CRS grading is that the severity is often based on the type and/or level of practitioner intervention. For example, the utilization of one versus more than one vasopressor agent to treat hypotension, or the use of supplemental oxygen alone versus mechanical ventilation for hypoxia, determines the CRS severity grade in several of the currently used CRS grading

Table 1
Evolving definitions and criteria for grading and managing CRS.

Lee criteria [17]
Grade 1: Symptoms are not life-threatening and require symptomatic treatment only (e.g., fever, nausea, fatigue, headache, myalgias, malaise) Grade 2: Symptoms require and respond to moderate intervention; oxygen requirement <40% or hypotension responsive to IV fluids or low-dose single vasopressor or grade 2 organ toxicity Grade 3: Symptoms require and respond to aggressive intervention; oxygen requirement of ≥40% or hypotension requiring high-dose or multiple vasopressors or grade 3 organ toxicity or grade 4 transaminitis Grade 4: Life-threatening symptoms; requirements for ventilator support or grade 4 toxicity (excluding transaminitis)
CTCAE v5.0 [18]
Grade 1: Fever with or without constitutional symptoms Grade 2: Hypotension responding to fluids; hypoxia responding to <40% oxygen Grade 3: Hypotension managed with one vasopressor; hypoxia requiring ≥40% oxygen Grade 4: Life-threatening consequences; urgent intervention indicated
Memorial Sloan Kettering Cancer Center (MSKCC) [19]
Grade 1: Mild symptoms requiring observation or supportive care only (e.g., antipyretics, antiemetics, pain medication) Grade 2: Hypotension requiring any vasopressors <24 h; hypoxia or dyspnea requiring supplemental oxygen <40% Grade 3: Hypotension requiring any vasopressors ≥24 h; hypoxia or dyspnea requiring supplemental oxygen ≥40% Grade 4: Life-threatening symptoms; Hypotension refractory to high-dose vasopressors; hypoxia or dyspnea requiring mechanical ventilation
Chimeric Antigen Receptor Toxicity (CARTOX) [20]
Grade 1: Temperature ≥38°C; grade 1 organ toxicity Grade 2: Hypotension responds to intravenous fluids or low-dose vasopressor; hypoxia requiring oxygen <40%; grade 2 organ toxicity Grade 3: Hypotension needing high-dose or multiple vasopressors; hypoxia requiring oxygen ≥40%; grade 3 organ toxicity or grade 4 transaminitis Grade 4: Life-threatening hypotension; needing ventilator support; grade 4 organ toxicity except grade 4 transaminitis
Penn Criteria [21]
Grade 1: Mild reaction: treated with supportive care, such as antipyretics, antiemetics Grade 2: Moderate reaction: some signs of organ dysfunction (grade 2 creatinine or grade 3 liver function tests [LFTs]) related to CRS and not attributable to any other condition; hospitalization for management of CRS-related symptoms, including neutropenic fever and need for IV therapies (not including fluid resuscitation for hypotension) Grade 3: More severe reaction: hospitalization required for management of symptoms related to organ dysfunction, including grade 4 LFTs or grade 3 creatinine, related to CRS and not attributable to any other condition; hypotension treated with multiple fluid boluses or low-dose vasopressors; coagulopathy requiring fresh frozen plasma, cryoprecipitate or fibrinogen concentrate; hypoxia requiring supplemental oxygen (nasal cannula oxygen, high-flow oxygen, CPAP or BiPAP) Grade 4: Life-threatening complications, such as hypotension requiring high-dose vasopressors; hypoxia requiring mechanical ventilation
American Society for Transplantation and Cellular Therapy (ASTCT) [22]
Grade 1: Temperature ≥38°C Grade 2: Temperature ≥38°C; hypotension not requiring vasopressor; hypoxia requiring low-flow nasal cannula or oxygen blow-by Grade 3: Temperature ≥38°C; hypotension requiring one vasopressor with or without vasopressin; hypoxia requiring high-flow nasal cannula, facemask, nonrebreather mask or Venturi mask Grade 4: Temperature ≥38°C; hypotension requiring multiple vasopressors (excluding vasopressin); hypoxia requiring positive pressure ventilatory support (CPAP, BiPAP, intubation or mechanical ventilation)

criteria (see Table 1). This is important, as the use of vasopressors or respiratory support is based on the clinical judgment of the physician, which may vary and thus lead to individual bias in CRS grading.

The presence of other organ function abnormalities is included in some, but not all, CRS grading systems. Other organ abnormalities could be reported either as separate AEs with no relationship to CRS or as preferred terms encompassing CRS. Therefore, it is important to clarify whether the definition of CRS should consider including these abnormalities to capture the full extent of CRS and minimize the risk of underdocumenting or underreporting. Additionally, if a therapeutic modality has the potential to cause clinically severe CRS that requires treatment with fluids, vasopressors, supplemental oxygen and anti-cytokine therapy, then initial low-grade events related to these manifestations should be assumed to be part of that spectrum and defined as CRS. Although there are a variety of published manuscripts, descriptions and adapted grading criteria and management strategies for CRS [22], it is noted that published definitions and grading criteria do not readily articulate the distinctions among CRS and

other clinical entities that may have overlapping symptoms and temporality (e.g., IRR, macrophage activation syndrome [MAS]/HLH).

Given the current variations in defining and reporting CRS, the working group feels an urgent need to harmonize the grading, collecting and reporting of CRS. Below are the working group proposals.

Alignment on Defining and Grading CRS

The American Society for Transplantation and Cellular Therapy (ASTCT) defines CRS as “a supraphysiologic response following any immune therapy that results in the activation or engagement of endogenous or infused T cells and/or other immune effector cells. Signs and symptoms can be progressive, must include fever at the onset, and may include hypotension, capillary leak (hypoxia), and end organ dysfunction” [22]. ASTCT’s definition for CRS represents an opportunity for alignment and prioritization of grading of clinically relevant events and can be inclusive of currently available and

Table 2
Principle components for defining CRS.

Principles	Considerations
Therapeutic modality	The spectrum of CRS and symptoms may change as different antigen targets and the methods to engage the immune system evolve; therefore, the definition of CRS may evolve.
Therapeutic schedule	The onset of CRS and severity can differ based on treatment administration (<i>i.e.</i> , one-time infusion vs. multiple infusions). Kinetics of CRS may differ by both disease state and therapeutic platform (<i>e.g.</i> , cellular products vs. T cell engagers).
Temporal association	The timing of development of CRS depends on patient-, disease- and treatment-related factors. In the setting of CAR T cells, <i>in vivo</i> expansion of CAR T cells is associated with the onset and maximum severity of CRS. A reasonable temporal relationship to the therapy must be present.
Sign and symptom manifestation	A suspected diagnosis of CRS should be made based on clinical signs and symptoms. Hallmarks of CRS are fever with or without hypotension and hypoxia; however, symptoms of CRS are not unique and overlap with other toxicities. Careful evaluation is required to ensure that the symptoms are associated with the cancer therapy, and other information such as blood cultures, fever workup, <i>etc.</i> , should be collected to assist in the differential.
Laboratory evaluation	Baseline assessment of inflammatory markers can assist in comparing with increased levels after treatment. Laboratory evaluation including C-reactive protein and ferritin are routinely available. Other cytokine level assessments (IL-6, IL-1, IL-8, TNF α and IFN γ), if available, can be helpful in further characterizing this syndrome retrospectively (unless available in real time).
Interventional care	CRS implies the toxicity may be effectively treated with anti-IL-6 therapy or other cytokine-directed therapies given in conjunction with corticosteroids, depending on the type of immunotherapy.

emerging immunotherapies, with some considerations as noted below.

Although each CRS grading scale in Table 1 has advantages and limitations, the working group supports the use of a harmonized definition and grading scale as well as collection of common data elements within and across development programs. An informal sponsor survey indicated that of eight sponsors, seven are using/planning to use ASTCT criteria for new protocols (see Supplement for Survey Summary). Several sponsors indicated that some development programs have been underway before the release of the ASTCT 2019 grading criteria, and CTCAE and Lee Criteria 2014 were predominantly being used to grade CRS. This is likely driven by efforts to simplify the characterization and categorization of the severity of CRS in the ASTCT criteria. Some limitations exist, such as the overlapping nature of oxygen requirements between grade 1 and grade 2 hypoxia due to the reliance on the oxygen delivery method and exclusion of end organ toxicities that result from CRS (*e.g.*, renal or hepatic injury). Furthermore, the use of proactive premedication (*e.g.*, corticosteroids) may limit or minimize the cascade of signs and symptoms of CRS, such as fever, hypotension and hypoxia, which is used as a defining characteristic of CRS in the ASTCT 2019 definition.

Because these guidelines have primarily been developed based on the clinical experience with CAR T cell therapy, they may prove, with additional clinical experience, to be incomplete for all cancer immunotherapies and may need to be revised as new data become available from existing and novel therapies [22]. As such, it is important that data collection is aimed at more than meeting the requirements of any one grading system. Therefore, establishing core principles for defining CRS that consider the therapeutic modality, symptom manifestation, timing and response to intervention will be important to enable flexibility and maximize utility of a harmonized definition for CRS to adequately assess safety profiles of therapeutics being offered to patients (Table 2).

Strategy for Assessing CRS over the Course of a Clinical Development Program

The characterization of CRS for a given experimental therapeutic in the course of a clinical development program is crucial to ensure the correct diagnosis and management of toxicity to help maximize treatment benefit. During the development of protocols for safety data collection and monitoring strategies as they relate to CRS, consideration should be given to how toxicities will be identified and managed in routine clinical care. Recognizing the association between the immunotherapeutic agent and CRS will inform the framework on how best to collect these data.

The collection of a broad dataset for characterizing CRS is resource intensive for both sponsors and investigators; however, assessing the risk of an IRR or CRS during preclinical and early clinical development of a new therapy will help gauge the robustness of data collection required during development to characterize the potential risk of CRS (Figure 1). The robustness of data collection can be assessed using a decision tree approach, which includes (1) an initial assessment of the risk of IRR or CRS based on mechanistic models and preclinical assessment; (2) biomarker and clinical data collection; and (3) iterative review of aggregate data to make an informed decision regarding CRS designation.

If there is a low risk or no risk of IRR or CRS based on mechanistic models, known class effects and nonclinical data, “LOW/NO” guidelines would be followed (Figure 1). In this instance, standard AE reporting and no upfront cytokine or other biomarker data collection would be recommended initially. With ongoing frequent safety data review and consideration for inclusion of cytokine and biomarker data collection, the data collection plan should be adapted if the clinical data are suggestive of potential IRR or CRS toxicity.

For therapeutic classes that are known to be associated with CRS or at particularly high risk for inducing CRS based on mechanism of action or preclinical data, the implementation of a dedicated clinical and safety monitoring plan may be required from the onset. The potential risks of IRR and CRS should be defined in the Investigator Brochure and protocol for the first-in-human trial, with a dedicated case report form (CRF) for IRR and/or CRS that collects the associated signs and symptoms. In addition, special preparation may be warranted as part of the protocol such as specific site training on CRS and the requirement of certain clinical interventions (*e.g.*, inpatient monitoring, intensive care unit [ICU] availability, and readily available tocilizumab). In most circumstances, it is recommended that physicians report either IRR or CRS as the Medical Dictionary for Regulatory Activities (MedDRA) preferred term until human data at the population level (*e.g.*, aggregate data in the clinical trial) are available. If there is evidence at the population level of cytokine-driven clinical signs and symptoms, increase in CRS biomarkers such as IL-6 or responsiveness to tocilizumab or other cytokine-directed therapies, it would be concluded that CRS is an identified risk and can then be characterized accordingly and allow proper clinical management. Lack of such evidence (*e.g.*, response to IL-6 directed agents) may suggest that the reaction is a manifestation of IRR or hypersensitivity but should not exclude the possibility of CRS based on further exploration and clinical assessments.

As more data are collected in a harmonized fashion, the field can better decide at which point and with which factors an event is determined to be a high-grade IRR versus a low-grade CRS. Understanding if there are implications on patient management will be important.

Table 3
Harmonized collection of discrete data elements.

Parameter	Data collection
Signs and symptoms	Minimum signs and symptoms to collect include fever, nausea, chills, vomiting, diarrhea, confusion, dizziness, dyspnea, tachycardia, headache, hypotension, hypoxia, lymphadenopathy, hepatosplenomegaly; but the eCRF should allow an investigator to enter any symptom thought to be a CRS symptom Date/time onset (e.g., x hour[s] after infusion of dose); initial grade; maximum grade; date/time resolution; intervention; outcome
Hypotension management	No intervention required, blood pressure values, intravenous fluids, use of vasopressors and dose, start/stop date of treatment, duration of treatment
Hypoxia management	No oxygen supplementation required, regular flow nasal cannula, high-flow nasal cannula, facemask, nonrebreather mask, or Venturi mask; positive pressure ventilatory support (CPAP, BiPAP, intubation, mechanical ventilation)
Organ toxicity	Liver function tests, creatinine, amylase, lipase, rash, neurotoxicity, cardiac, pulmonary, renal, hepatic toxicities
Cytokines	IL-6, IL-1, IL-8, TNF α and IFN γ are recommended as a core cytokine panel, if available and considered in a research setting
Other laboratory assessments	Routine hematology analysis, including complete blood count and differential, serum chemistries, coagulation factors, ferritin, C-reactive protein
Care setting	Admitted to hospital or ICU; duration, including distinction between ICU and non-ICU duration
Intervention for management	Tocilizumab or other cytokine-directed therapy administered for management, as well as corticosteroids or other supportive care, such as antipyretics, and type of prophylaxis, if any; if applicable, permanent discontinuation of immunotherapy or ability to rechallenge and administer therapy

symptoms, it is important to capture the laboratory value information, including cytokine biomarkers, as laboratory values and cytokines help improve our understanding of the pathophysiology and may inform future development of management guidelines. Although there are currently no commercially available assays to determine expansion and persistence of CAR T cells, and real-time cytokine analysis is also not typically available, correlative analyses in the context of clinical trials may allow retrospective analyses to interrogate CRS cases and direct future guidelines for toxicity management.

With CAR T cell therapy, routine CRS assessment may range from daily CRS assessments immediately after infusion, to two to three times a week for the first 30 days after infusion, to help characterize the evolution of symptoms, development of additional toxicities, treatment and response to treatment. The timing and frequency of CRS assessments for T cell engagers may vary and depend on the pharmacokinetics and pharmacodynamics of the particular molecule and dosing schedule. Timing of sampling should be adapted to accommodate treatment cycles and protocol-defined scheduled visits.

Consistent Method for Recording and Reporting CRS Events

As described in Figure 1, identification and characterization of CRS early in a new experimental therapeutic's clinical development can

be challenging owing to the heterogeneity in signs and symptoms and similarity of CRS signs and symptoms to those of other AEs, such as IRR or infection, as noted earlier. A hypothetical case is shown in Figure 2. A patient treated with a T cell engager experiences several AEs. Initially, the patient presents with a fever of 40.1°C lasting 6 h that is accompanied by hypotension responsive to a 1-liter fluid bolus. The fever and hypotension are CTCAE grade 3. The next morning, liver function test (LFT) increases are noted (grade 4), and later that day, the patient has a brief generalized seizure that is self-limited, lasting less than a minute (grade 2). Grade 2 CRS is diagnosed [22]. It is important for all these AEs to be captured into the CRF and independently reported for characterization of the range and severity of signs and symptoms constituting the grade 2 CRS. Although all of these may precede the investigator diagnosing CRS, all AEs should be captured into the CRF and independently reported. Independent AE reporting is critical, and clinical investigators should carefully describe all events that are suggestive of potential CRS. This will enable pharmacovigilance experts to evaluate relevant, linked events and code them as CRS, as appropriate.

In the example case, the event of fever precedes the diagnosis of CRS and would be captured into the AE database independently and graded independently, as the differential diagnosis for the fever could include not only CRS, but other potential etiologies such as IRR and infection. Additional events such as the increase in LFTs and seizure are attributable to the CRS, but could also be recorded as independent AEs into the database. Once CRS is diagnosed and recorded as an AE in the trial database, the signs and symptoms indicative of CRS ideally would be linked to the CRS event.

We propose a comprehensive method to capture all the events and link those AEs that are signs and symptoms of CRS to the CRS event, such that CRS is the AE, but the symptoms (fever, LFT increase, seizure) that are AEs in themselves are attributable to CRS and are linked to the CRS event (Figure 2, "Link Events to CRS"). For instance, one way is to flag each AE that is related to CRS and link it to the specific CRS event. This will allow a more qualitative analysis of CRS, as CRS can manifest in a variety of organs including hepatic, renal and neural system. This method would also allow the optionality of reporting all AEs, CRS and the specific organ toxicity separately or allow collapsing of the CRS-related events to a single AE. Given the importance of central nervous system (CNS)-related toxicity with T cell therapies, it is recommended that ICANS (immune effector cell-associated neurotoxicity syndrome) events be captured and scored separately. In the case described, any seizure would be captured as a grade ≥ 3 ICANS.

Without data collection standards, several outcomes in terms of data capture may arise. For instance, one possible method is that all the signs and symptoms that are attributable to CRS could be collapsed into the AE preferred term of CRS. Once the investigator identifies CRS, as part of data cleaning, the fever, LFT increase and even seizure events could be accounted for by CRS, and only the CRS event is reported (Figure 2, "Collapse Events to CRS"). However, this method would lead to the loss of actionable information that may be useful for retrospective application of future CRS diagnostic or grading criteria and for physicians and patients.

As described in Table 3, additional information would be captured including use of concomitant medications (e.g., tocilizumab or other cytokine-directed therapy, oxygen, vasopressors, corticosteroids) and specific interventions (e.g., method of oxygen delivery, mechanical intervention, intravenous fluids). In our example case, the use of intravenous fluids and not vasopressors define a grade 2 CRS event. Although these items may be collected in other parts of the electronic data capture record, it is important that they are easily linked to a specific CRS event, as CRS grading is dependent on these interventions in most classification systems. In addition, some grading systems can lead to downgrading of events. As an example, liver function laboratory values may increase

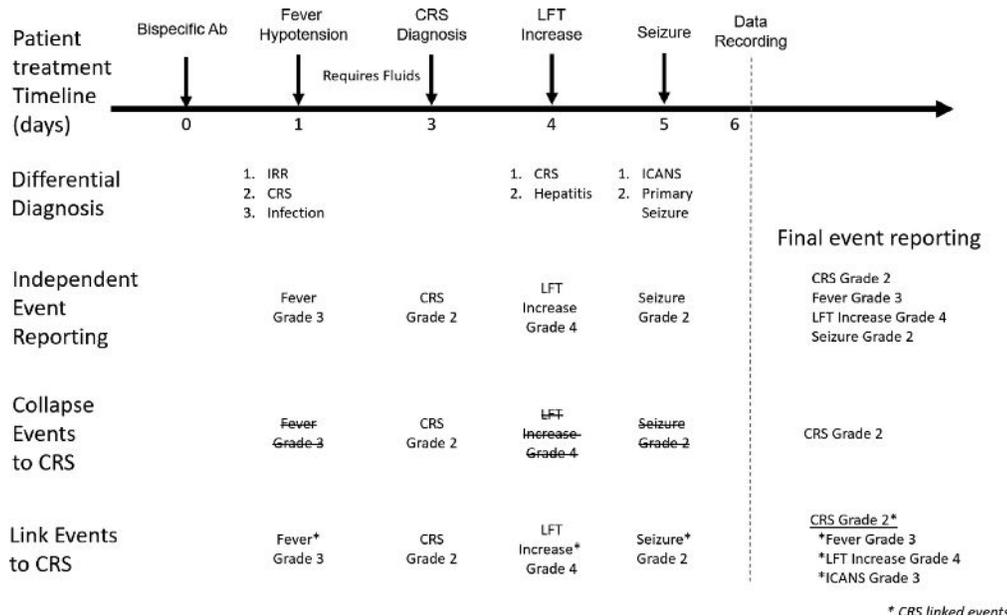


Figure 2. Patient treatment timeline and event reporting after immunotherapy administration.

transiently and meet the criteria for CTCAE grade 4 CRS based on these laboratory changes; however, this increase will only meet the definition of a grade 2 CRS by ASTCT criteria if it is not accompanied by clinically significant changes in blood pressure or oxygen requirement.

Conclusions

Cytokine release syndrome is commonly seen with newer immunotherapies, such as T cell engagers and CAR T cells, and presents as a range of signs and symptoms, most commonly fever, hypotension and hypoxia. There are currently several different scales used to grade the severity of CRS, and therefore, it is important to standardize the grading to ensure consistency in how data are collected and presented and to better distinguish CRS from other clinical entities with overlapping symptoms. Ideally, all investigators would commit to a harmonized data collection approach using a dedicated CRS eCRF with data elements identified in Table 3 as a guide. Moreover, this working group outlined several actionable proposals for deployment in early clinical development programs of emerging immunotherapies. To improve alignment on defining CRS, clinical programs could establish core principles that consider the therapeutic modality, symptom manifestation, timing, and response to intervention as part of a harmonized definition of CRS. A strategy for assessing CRS over the course of a clinical development program that takes into account an initial assessment of the risk of IRR or CRS, biomarker and clinical data collection, and an ongoing review of data is needed to make an informed decision of CRS designation. In regard to data, harmonized data elements for characterizing CRS need to be determined, as comprehensive data capture will allow for easier adaptation of one grading scale to another. Lastly, a consistent method for recording and reporting CRS events is necessary to simultaneously delineate the individual signs and symptoms of CRS as well as to characterize the CRS event as its own entity. In particular, the previous experience with the study of immune-related AEs due to checkpoint inhibitors emphasized that, when evaluating CRS events, alignment of

important data elements and a more comprehensive understanding of AEs early in clinical development can support ongoing pharmacovigilance and real-world data collection to enable further characterization of these events in the postmarketing setting.

As our clinical understanding of CRS and other clinical entities associated with these types of therapies evolves, a harmonized approach for defining, characterizing and reporting CRS in patients receiving immunotherapies is necessary to support evidence-based monitoring and management of novel toxicities; facilitate and harmonize the assessment and communication of risk-benefit profiles with regulatory agencies, the clinical community, and the public; and improve patient care and outcomes. Furthermore, such an approach can also support retrospective analyses to compose new iterations of grading criteria and clinical guidelines, to ensure the safe monitoring and administration of T cell engaging immunotherapies.

Disclaimer

This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

Conflicts of Interest

B.M. is an employee of Genentech and owner of Roche stock. A.S.Y. is an employee of Xencor and owner of Xencor stock. C.D.B. is an employee of Amgen and owner of Amgen stock. M.M.H. has received grant funding for research from Allovir, Amgen, Astellas, Gamida Cell Ltd., Genentech Inc., Magenta Therapeutics, Medac GmbH, Onclimmune, and Vor Biopharma. C.M. is an employee of BMS and owner of BMS and Amgen stocks. P.K.M. is an employee of Amgen and owner of Amgen stock. E.O. is an employee of Regeneron. R.R. is an employee of Bristol Myers Squibb. F.V. is an employee of Bristol Myers Squibb. T. Y. is an employee of Amgen and owner of Amgen stock. All other authors have no disclosures to report.

Acknowledgment

The authors thank Yifang Eva Pan for writing support and editorial assistance and Joe Patterson (Friends of Cancer Research) for graphical assistance.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jcyt.2022.01.004.

References

- [1] Xin Yu J, Hubbard-Lucey VM, Tang J. Immuno-oncology drug development goes global. *Nature reviews. Drug discovery* 2019;18:899–900.
- [2] Khadka RH, Sakemura R, Kenderian SS, Johnson AJ. Management of cytokine release syndrome: an update on emerging antigen-specific T cell engaging immunotherapies. *Immunotherapy* 2019;11:851–7.
- [3] Murthy H, Iqbal M, Chavez JC, Kharfan-Dabaja MA. Cytokine Release Syndrome: Current Perspectives. *ImmunoTargets Ther* 2019;8:43–52.
- [4] Maus MV, et al. Society for Immunotherapy of Cancer (SITC) clinical practice guideline on immune effector cell-related adverse events. *J. Immunother. cancer* 2020;8.
- [5] Santomasso BD, et al. Management of Immune-Related Adverse Events in Patients Treated With Chimeric Antigen Receptor T-Cell Therapy: ASCO Guideline. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 2021;39:3978–92.
- [6] Doessegger L, Banholzer ML. Clinical development methodology for infusion-related reactions with monoclonal antibodies. *Clin. Transl. Immunol.* 2015;4:e39.
- [7] Roselló S, et al. Management of infusion reactions to systemic anticancer therapy: ESMO Clinical Practice Guidelines. *Ann. Oncol.* 2017;28:iv100–18.
- [8] Bupathi M, et al. Incidence of infusion reactions to anti-neoplastic agents in early phase clinical trials: The MD Anderson Cancer Center experience. *Invest. New Drugs* 2017;35:59–67.
- [9] Bernardez, B. et al. Real world of infusion related reactions in an oncology outpatient clinic. (2018).
- [10] Chung CH, et al. Cetuximab-induced anaphylaxis and IgE specific for galactose- α -1, 3-galactose. *N. Engl. J. Med.* 2008;358:1109–17.
- [11] Farshi, E. Cytokine Storm Response to COVID-19 Vaccinations. (2021).
- [12] Giavridis T, et al. CAR T cell-induced cytokine release syndrome is mediated by macrophages and abated by IL-1 blockade. *Nat. Med.* 2018;24:731–8.
- [13] Shimabukuro-Vornhagen A, et al. Cytokine release syndrome. *J. Immunother. cancer* 2018;6:56.
- [14] Sampson HA, et al. Second symposium on the definition and management of anaphylaxis: summary report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *The Journal of allergy and clinical immunology* 2006;117:391–7.
- [15] Maude SL, Barrett D, Teachey DT, Grupp SA. Managing cytokine release syndrome associated with novel T cell-engaging therapies. *Cancer J* 2014;20:119.
- [16] Le RQ, et al. FDA Approval Summary: Tocilizumab for Treatment of Chimeric Antigen Receptor T Cell-Induced Severe or Life-Threatening Cytokine Release Syndrome. *Oncologist* 2018;23:943–7.
- [17] Lee DW, et al. Current concepts in the diagnosis and management of cytokine release syndrome. *Blood* 2014;124:188–95.
- [18] National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf (2017).
- [19] Park JH, et al. Long-Term Follow-up of CD19 CAR Therapy in Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* 2018;378:449–59.
- [20] Neelapu SS, et al. Chimeric antigen receptor T-cell therapy - assessment and management of toxicities. *Nat. Rev. Clin. Oncol.* 2018;15:47–62.
- [21] Porter D, Frey N, Wood PA, Weng Y, Grupp SA. Grading of cytokine release syndrome associated with the CAR T cell therapy tisagenlecleucel. *J. Hematol. Oncol.* 2018;11:35.
- [22] Lee DW, et al. ASTCT Consensus Grading for Cytokine Release Syndrome and Neurologic Toxicity Associated with Immune Effector Cells. *Biol. blood marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* 2019;25:625–38.
- [23] Davila ML, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci. Transl. Med.* 2014;6. 224ra25.
- [24] Wei J, et al. The model of cytokine release syndrome in CAR T-cell treatment for B-cell non-Hodgkin lymphoma. *Signal Transduct. Target. Ther.* 2020;5:134.

Accelerating Investigation of New Therapies in Earlier Metastatic Treatment Settings

FRIENDS OF CANCER RESEARCH ANNUAL MEETING 2022

Introduction

Over the past decade, an increasing number of breakthroughs in cancer research have translated into novel and highly effective therapies for patients. Investigational therapies are often first studied in patients with relapsed or refractory (r/r) disease and who have received multiple prior lines of therapy or have exhausted all available treatment options (i.e., later treatment setting or disease setting with lack of available treatments). Between January 2013 and July 2022, over 61% of oncology approvals for novel molecular entities were for patients with metastatic disease who had received prior therapies.¹ Studies of new investigational therapies are often conducted in the r/r patient population due to the unmet need for treatment options, ethical concerns about exposing newly diagnosed patients to therapies that may be ineffective, and potential earlier market access through the accelerated approval (AA) pathway.² Designing trials in the r/r setting yields important insights for investigational agents (e.g., dosing, tolerability, etc.) and provides access to investigational therapies for patients with r/r disease who may not have other acceptable options.

Recently, concerns have increased regarding the limitations of using single-arm trials to support AA, failure and delays in confirming benefit for drugs granted AA, and the inherent challenges of confirming clinical benefit in the r/r setting when trials are initiated after the AA has been granted.² Conducting trials in earlier metastatic settings (including but not limited to first-line therapy) as a strategy to support initial approval may address some of these limitations and has the potential to maximize the benefit of innovative treatments and expand access to more patients with metastatic disease more quickly.

The Oncology Center of Excellence (OCE) at the U.S. Food and Drug Administration (FDA), launched Project FrontRunner to initiate a discussion among stakeholders in oncology drug development for considerations on shifting the historical drug development paradigm which has focused on first developing and seeking approval of new therapies in the r/r metastatic setting.³ As part of

Thank You to Our Contributors

Hesham Abdullah, MD, GSK

Kenneth Anderson, MD, Dana-Farber Cancer Institute

Elizabeth Barksdale, PhD, LUNGevity

Jonathan Cheng, MD, BMS

'Lola Fashoyin-Aje, MD, MPH, U.S. Food and Drug Administration

Keith Flaherty, MD, Massachusetts General Hospital Cancer Center

Linda Gustavson, PhD, Pfizer

Jane Healy, MD, PhD, Merck

Ariadna Holynskyj, MD, Pfizer

David Hyman, MD, Loxo@Lilly

Ed S. Kim, MD, MBA, City of Hope

Meijuan Li, PhD, Eisai

Stacie Lindsey, BA, Cholangiocarcinoma Foundation

Brittany McKelvey, PhD, Friends of Cancer Research

Colleen Mockbee, MBA, OncXerna Therapeutics

Cassadie Moravek, BS, Pancreatic Cancer Action Network

Anne Quinn Young, MPH, Multiple Myeloma Research Foundation

Elmar Schmitt Dr, MDRA, EMD Serono

Greg Taylor, PharmD, Incyte

Craig Tendler, MD, Janssen

Patricia Thomas, PhD, Takeda

Darshan Wariabharaj, BS, Janssen

this initiative, OCE aims to propose, for use by pharmaceutical sponsors on a voluntary basis, a framework that helps identify clinical development programs that may benefit more patients earlier in the course of their disease and improve the data available at the time of approval to facilitate a benefit-risk assessment. Friends of Cancer Research (*Friends*) convened a multi-stakeholder group of experts including the OCE, drug developers (Sponsors), patient advocates, and academic clinicians to identify key opportunities and challenges for designing studies that support approval in earlier metastatic treatment settings and initiating such studies earlier in the overall drug development program.

Objectives

- Identify opportunities, challenges, and potential strategies to accelerate the study of investigational therapies in the earlier metastatic treatment setting.
- Develop a framework to facilitate determining when it is appropriate to initiate the study of investigational therapies in the earlier metastatic treatment setting, informed by important clinical, statistical, and regulatory considerations.
- Identify the critical components of a comprehensive development strategy to support accelerated clinical development and regulatory approvals.

At the outset, there was broad recognition that the Project FrontRunner paradigm may not be appropriate for every clinical setting or investigational drug. As such, this paper provides a proposed framework for considering whether the Project FrontRunner approach is appropriate in a given context and, outlines some important considerations for implementing this approach in the appropriate setting. Importantly, the framework and clinical development considerations may be subject to further revisions based on additional input and experience.

Rationale for Advancing Investigation of Novel Therapies Earlier in the Course of Metastatic Disease

Provide Greater Clinical Benefit to More Patients

Therapeutic investigations earlier in the course of metastatic disease have the potential to provide a greater benefit to patients with cancer since there are more patients with earlier metastatic disease and the absolute effect size of investigational therapies on endpoints such as progression-free survival (PFS) and overall survival (OS) tends to be greatest.⁴⁻⁷ In the r/r setting, patients may have disease-related factors or complications or may have residual side effects from prior treatments that may confound the evaluation of an investigational therapy for safety and efficacy. In some cases, the effects of prior therapy or disease progression may preclude patient participation in clinical trials.⁸ Investigation in early line metastatic disease increases the clinical trial opportunities for more patients with metastatic disease.

Accelerate Addressing Unmet Need in Earlier Metastatic Treatment Setting

While unmet need is not a regulatory requirement for AA, the intent of regulatory mechanisms and flexibilities that allow for earlier approval of drugs to treat serious conditions is to address

unmet medical needs.¹⁰ It is important to note that investigational therapies in the earlier metastatic setting have the potential to address unmet need by providing therapeutic options, including potentially through the AA pathway, when no standard of care (SOC) treatment exists. This can help provide alternatives to or replace less effective or more toxic therapies in the earlier treatment setting, and/or enhance current SOC efficacy through a combination approach. Initiating investigations in early metastatic settings using the Project FrontRunner paradigm allows for comparison of the investigational therapy to established and approved therapies for enhanced benefit–risk assessments.

Lessons Learned from Past Drug Development Programs

A review of several past drug development programs informed learnings from conducting clinical trials in metastatic disease and strategies for future trial designs that may align with the goals of Project FrontRunner (see **Appendix 1** for case study reviews). Insights were gleaned from the Sponsors and publicly available FDA review summaries.¹¹ Summary key findings from these case studies include:

- Certain clinical scenarios and therapeutic regimens (e.g., when an investigational drug is not expected to be effective as monotherapy or requires a combination therapy that includes current SOC) may require initiation of registrational or pivotal studies in the frontline setting.
- Robust statistical approaches are necessary to address challenges associated with interim analyses. Conducting an interim analysis based on events with limited follow-up may result in an inaccurate estimation of clinical benefit. Statistical considerations for the required hazard ratios and alpha spending for interim analyses will be important.
- Endpoints used for interim analyses (e.g., overall response rate (ORR)) should be established to be reasonably likely to predict clinical benefit for the disease and the therapy being studied.
- The randomized controlled trial (RCT) planned to verify benefit observed in a single-arm trial should be nearly or fully enrolled at time of submission of the single-arm trial for AA.

Key Considerations for Initiating Clinical Development in Earlier Metastatic Settings

Studying investigational therapies earlier in the course of advanced/metastatic disease (e.g., first or second-line setting) may be appropriate for a subset of clinical and drug development scenarios. The proposed considerations for selecting a clinical development scenario appropriate for the Project FrontRunner setting are shown in **Table 1**.

Table 1: Considerations for Selecting a Project FrontRunner Clinical Development Paradigm

Factor	Characteristics	Considerations
Disease Characteristics	<ul style="list-style-type: none"> Natural history of disease (e.g., long or short natural survival) Size of eligible population (e.g., rare or more common) 	<ul style="list-style-type: none"> Natural history of the disease can impact the length of time it takes for data to mature to demonstrate treatment benefit. An earlier readout of a well-established intermediate endpoint could form the basis for AA ahead of clinical benefit outcomes. The size of the eligible patient population is an important consideration because it can impact enrollment rate and ultimately the time it takes for trial results (interim and final) to be available.
Investigational Treatment Characteristics	<ul style="list-style-type: none"> Novelty of mechanism of action (e.g., first-in-class or 3rd/4th in class) Approval status (e.g., new molecular entity or expanding indication of approved drug) Level of toxicity (e.g., high or low) 	<ul style="list-style-type: none"> Data (e.g., efficacy, dosing, toxicity profile) may be leveraged for drugs that have been previously approved in other indications or for investigational agents within an existing drug class, which can help de-risk the approach. Investigational agents with high toxicity may not be amenable to study in early lines with existing, less toxic SOC options.
Other Available Therapies	<ul style="list-style-type: none"> Efficacy, safety of approved available therapies in early metastatic setting (i.e., 1st/2nd line setting) Efficacy of available therapies Tolerability of available therapies 	<ul style="list-style-type: none"> Settings where established SOC is associated with modest-to-moderate outcomes or poor toxicity offers opportunities to demonstrate convincing and clinically meaningful improvement.
Clinical Endpoints	<ul style="list-style-type: none"> Intermediate endpoints available and acceptable for regulatory use 	<ul style="list-style-type: none"> Disease settings that have well-established intermediate endpoints (e.g., correlation to long-term clinical endpoints) to support interim analyses may be most appropriate.

Some disease characteristics are more amenable to the FrontRunner paradigm than others, including those for which there is evidence to support the use of an intermediate endpoint, such as ORR, that is reasonably likely to predict clinical benefit and could support interim analyses evaluating treatment efficacy. These include diseases with long natural histories such as indolent non-Hodgkin lymphoma or multiple myeloma in the frontline setting. Alternatively, diseases where the natural history is short, but OS (rather than PFS) is the endpoint of interest for regulatory approval, such as second-line non-small cell lung cancer, would also be candidates for the FrontRunner approach. In this setting, the established SOC is associated with low-to-moderate outcomes (15-20% ORR) and an investigational therapy may demonstrate

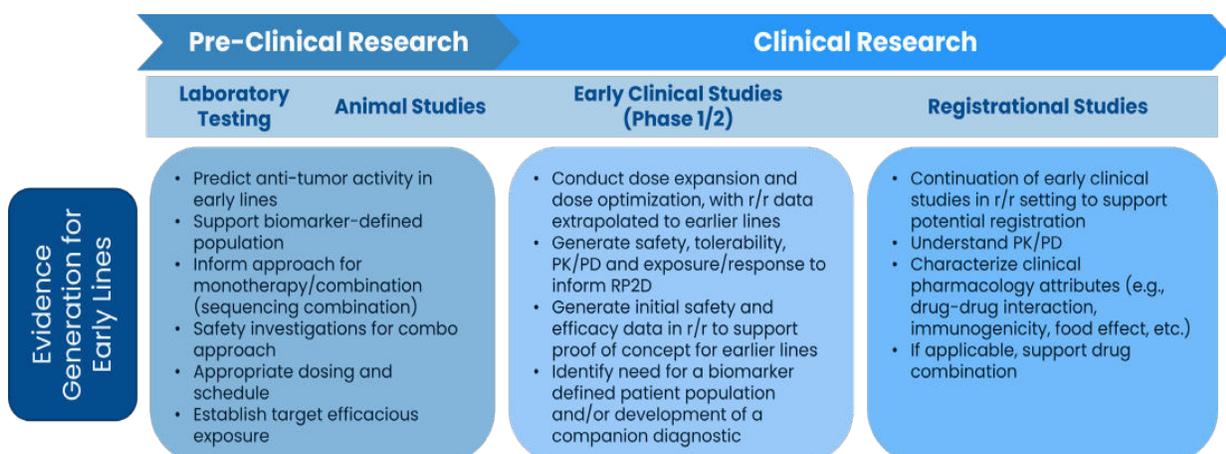
meaningful improvement. More common diseases in which trial enrollment can be completed expeditiously will likely also benefit from a FrontRunner approach, as accrual of a sufficient sample size will likely not be rate-limiting. These considerations are based on the use of ORR for interim analyses. However, in the future if other intermediate endpoints, such as those based on circulating tumor DNA (ctDNA) or minimal residual disease (MRD), are robustly characterized and accepted as reasonably likely to predict clinical benefit, additional clinical scenarios may become amenable.

Operationalizing a Project FrontRunner Approach During Clinical Development

Factors Influencing the Clinical Development

When establishing the development plan for an investigational drug, Sponsors should make plans beyond the initial indication, including establishing a holistic clinical and registration plan. This plan should determine the feasibility and applicability of investigating the agent in earlier settings, including establishing set decision points to determine when to initiate studies and what evidence is needed to support the move to earlier lines. These decisions should be based on early clinical and scientific evidence, further discussed below and highlighted in **Figure 1**, but may also be driven by factors such as the level of risk a sponsor and regulatory bodies are willing to accept, development timeline to regulatory approval, market opportunity (including an assessment of competition and potential changes in the treatment landscape), size of target population, relevance of the target in the earlier setting, and market access considerations of the portfolio. A Project FrontRunner approach would not preclude these decisions from being made, but rather it would clarify opportunities and encourage trial designs in earlier metastatic disease sooner in the development of an investigational agent.

Figure 1: Evidence to Support Investigation in Earlier Metastatic Treatment Settings. Evidence that may be leveraged to support initiating clinical development in earlier metastatic settings, obtained through pre-clinical research and/or clinical research in the r/r setting.



Treatment Landscape in the Earlier Metastatic Treatment Setting

Understanding the current treatment landscape in the early metastatic setting is key to informing clinical trial designs when moving treatments into earlier lines. The primary endpoint(s) used to support previous regulatory decisions in the specific cancer type as well as the magnitude of improvement observed in clinical trial readouts may help guide trial design considerations, including the level of evidence needed for AA decisions and the endpoints used to determine clinical benefit. These considerations may change if the proposed study is intended to replace current SOC as opposed to combining the investigational therapy with SOC. Additionally, the current SOC will impact selection of the control arm of a RCT. In some instances, it may be challenging to identify a single control arm if there are multiple treatments available in the earlier setting and a physician's choice may be most appropriate. Providers and patients may need additional education about the value of enrolling in a clinical trial in an earlier line if a well-established SOC and/or existing therapies for early line treatment already exists.

Pre-clinical and Early Clinical Studies

Figure 1 highlights data from pre-clinical research and early phase clinical studies that can be used to support investigation in earlier metastatic treatment settings. Robust pre-clinical disease models predict the potential anti-tumor activity in the intended tumor type(s) to support the selected patient population and provide insight into investigations in earlier lines. Pre-clinical data are also critical to inform the approach for studying an investigational therapy as a monotherapy or in combination with other therapies. If pursuing a combination therapy approach, early inclusion of the combination partner in nonclinical investigations is beneficial to begin to understand the drug-drug interaction profile as well as inform the dosing regimen. Enrolling patients with early metastatic disease in dose finding studies may be challenging and reaching time to event endpoints may take a long time. Therefore, extrapolating data from dosing studies in the *r/r* setting may be appropriate to aide dose selection for early metastatic settings. Alternatively, it may be beneficial to use earlier endpoints, such as ORR, for dose finding studies, if required, in the early metastatic disease setting.

Clinical Trial Design

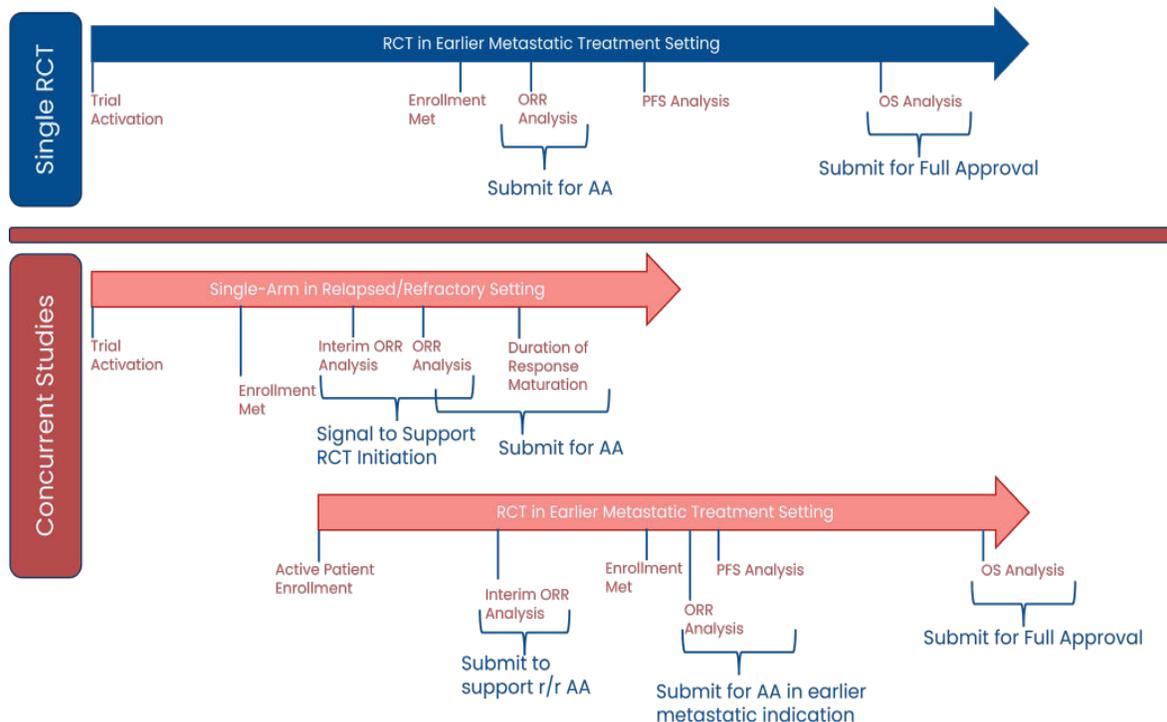
Two strategies can be considered to facilitate regulatory review in earlier metastatic lines: a single RCT in the early treatment setting to support accelerated and full approval, and two concurrent studies that overlap, one of which is a single-arm trial in a *r/r* setting and the other a RCT in an earlier treatment setting (**Figure 2**). **Table 2** highlights advantages and disadvantages of these approaches. In addition to these designs, in rare patient populations with limited therapeutic options and feasibility challenges to enrolling a sufficient number of patients for a RCT, a single-arm trial may be acceptable. In this case, assuming data are available, there may be opportunities to use real-world data (RWD) on the natural disease history as supportive information, if proactively discussed and aligned with the Agency, to contextualize the effect seen in the single-arm trial or for use as an external control arm.

Table 2: Possible Strategies to Support Accelerated and Full Approval in Earlier Settings

	Single Randomized Trial	Two Concurrent Studies
Approach	<ul style="list-style-type: none"> The same study supports AA and subsequently verifies clinical benefit with guarantee of timely confirmatory readout <ul style="list-style-type: none"> AA granted on planned analysis of ORR (potentially in a subset of patients) Traditional approval granted on clinical benefit (e.g., PFS, OS) 	<ul style="list-style-type: none"> Single-arm trial examining ORR in r/r setting, allowing for collection of data that supports an earlier initiation of a RCT in patients in earlier metastatic lines and AA in the r/r setting RCT in patients in earlier treatment setting to support AA and subsequently confirm clinical benefit
Advantages	<ul style="list-style-type: none"> More thorough safety assessment than single-arm trial Definitive evidence of benefit-risk from single trial in same patient population May reduce risk of prematurely halting drug with limited increment to ORR that may still improve OS, depending on characteristics of alpha spending approach 	<ul style="list-style-type: none"> Able to generate evidence to support investigation in earlier setting (if biology is similar) Provides data to support indication in r/r setting Potential to address unmet need in more expeditious manner in both the r/r setting and earlier setting Interim analysis of safety and ORR in confirmatory trial could provide support for an AA in single-arm trial indication ORR in this interim could support an earlier, additional, AA indication in the RCT population
Disadvantages	<ul style="list-style-type: none"> Greater risk/higher investment with less clinical experience from r/r setting to inform study Possible statistical concerns with more stringent alpha control with multiple endpoints 	<ul style="list-style-type: none"> Confirmation of clinical benefit (e.g., PFS, OS) in RCT is not the same patient population as single-arm trial for AA conversion Timing of endpoint readouts in r/r may impact start of RCT (need full enrollment of RCT at submission of r/r single-arm trial for AA)

Adapted from Fashoyin-Aje, et al.²

Figure 2: Possible Clinical Development Approaches to Support Accelerating Investigation in Earlier Settings. Two possible strategies to facilitate regulatory review in earlier metastatic lines, including the submission to support AA and full approval through a single RCT in the early setting, or two concurrent studies of a single-arm trial in the r/r setting and a RCT in the early setting.



Single RCT

A single RCT could be initiated in an earlier metastatic setting with prospectively defined intermediate endpoints to support an AA (e.g., ORR) and traditional long-term clinical endpoints to confirm benefit. The ORR analysis could be conducted in a subset of the enrolled patient population for an initial signal of clinical benefit, triggering an increase in enrollment to confirm the benefit. Resulting data from the RCT are more robust than a single-arm trial due to evaluable clinical efficacy and safety data, however, RCTs may incur greater risk given the limited clinical experience from the r/r setting to inform RCT study design.

Two Concurrent Studies

A single-arm trial in the r/r setting can generate evidence to support investigation in the earlier treatment setting. These data may support an AA in the r/r setting as well as establish a proof of concept for clinical efficacy (based on the ORR and durability of response) and an understanding of the PK/PD including the potential drug-drug interaction profile, to inform the design of a RCT in an earlier setting. Additionally, continued clinical evaluation in the r/r setting supports development of therapies for this population in parallel while the randomized study in the earlier setting begins.

The RCT can serve four purposes:

1. An interim analysis of ORR to support AA in the r/r setting,
2. An analysis of ORR to support an AA in the early metastatic setting,
3. Confirmation of clinical benefit with PFS/OS to support full approval in the early metastatic setting and conversion of AA in the r/r setting to full approval, and
4. A clearer understanding of safety assessment.

In this approach, it is critical that the RCT is ongoing, with enrollment complete or near complete, prior to submission of data from the r/r single-arm trial to support a regulatory decision for AA.

Statistical Considerations for Endpoint Analyses in RCTs

When deciding planned endpoint analyses, multiple factors may influence the choice of endpoints: effect size, effect duration, depth of response, available therapy, disease setting, and risk-benefit relationship.¹³ When analyzing multiple endpoints within a single trial, the optimal alpha spending and multiplicity control strategy must be considered. Allocation of alpha could be initially split between the AA endpoint and confirmatory endpoint and subsequently recycled upon successful demonstration of effects in corresponding endpoints, such that regardless of the ORR result, PFS/OS could potentially still reach significance. This may require a more stringent boundary for the AA endpoint, longer follow-up, and a higher event rate for the confirmatory endpoint, or both. FDA released final guidance to support the use of multiple endpoints in clinical trials which outlines key statistical considerations.¹⁴ Various scenarios and assumptions that impact timing of data readouts and other endpoint considerations are described in **Appendix 2**.

Clinical Equipoise Considerations

There are ethical considerations for conducting randomized studies depending on the expected magnitude of effect based on early clinical signals and pre-clinical evidence. If a high magnitude of benefit is observed in either the r/r setting or as part of the intermediate endpoint analysis in the RCT, it may be challenging to enroll patients onto the control arm, demonstrating the importance of fully enrolling the RCT prior to submission of interim analyses for AA. The trial design and statistical analysis plan could incorporate unblinding during follow-up as data continue to accrue for long-term endpoints or consider challenges and opportunities of cross-over. However, this is similar to the current AA paradigm with ongoing studies to confirm benefit in earlier line settings.

Biomarker-Driven Development Considerations

There may be additional clinical development considerations for indications in biomarker-defined populations (for the purposes of this white paper, a biomarker is a predictive biomarker that is predictive of the efficacy of a specific therapy). Previously validated biomarkers can be utilized more quickly than novel biomarkers, which require more coordination for co-developing a drug and diagnostic. An in vitro diagnostic investigational device exemption (IDE) may be required if a novel biomarker is used in the early treatment setting where available approved therapies exist, as the study may be deemed a significant risk to patients if they are foregoing approved therapies. Establishing an IDE for multiple local tests may be burdensome and Sponsors should align with the Agency when using multiple tests for enrollment. While FDA has

approved therapeutic products when a companion diagnostic (CDx) device is not approved or cleared contemporaneously, this may be unlikely in the earlier treatment setting given the increased risk posed to patients if they receive a potentially ineffective treatment.

To support evidence generation for rare settings (e.g., a rare biomarker in a common disease setting or a rare disease with a common biomarker), RWD can support an understanding of biomarker prevalence, as well as identify high-risk populations who could be tested first, with supportive pre-clinical evidence (that may have the highest magnitude of effect). To determine whether the biomarker is predictive, biomarker-positive patients should be randomized into the investigational and SOC arms. If both biomarker-positive and -negative patients are included in the control, then one can only assess the prognostic role of the biomarker. This approach of first studying the investigational therapy in a pre-defined population with established medical need could also be employed to establish entry cohorts for special patient populations based on clinical characteristics that may be lacking in the clinical trial population but representative of the overall patient population with the disease.⁹ Data from this entry cohort could then be appropriately extrapolated to the broader patient population.

Conclusions and Future Directions

This white paper provides parameters to help Sponsors identify candidates for a Project FrontRunner approach and outlines a framework for operationalizing this approach within the overall drug development plan. This approach is intended to be an adjunct to, but not replacement of, existing paradigms for accelerated approval of oncology products. Early interactions between Sponsors and the Agency are recommended to discuss and confirm the comprehensive development plan. Initiating discussions with international health authorities regarding clinical development plans is important as the acceptability of endpoints, such as ORR, may differ across health authorities and health technology assessment (HTA) bodies for approval and/or reimbursement, which may present challenges to broader adoption. While Project Orbis provides a framework for concurrent submission and review of oncology applications, expanding this framework to earlier phases of development could be beneficial.¹⁵

OCE's Project FrontRunner and the proposed framework for advancing the investigation of therapies in earlier treatment settings holds great promise to extend clinical benefit into broader patient populations. However, this paradigm will not be appropriate for all clinical and drug development scenarios. It will therefore be important to identify the scenarios amenable to this approach, as discussed herein, and hold disease-focused drug development workshops to further operationalize these concepts. Additionally, identifying and validating other novel early endpoints like ctDNA can help expand the application of this approach to other disease settings and therapies. Lastly, operationalizing the considerations and concepts of the framework to support the goals of the initiative at FDA would be beneficial, such as encouraged synergy with CDx development for biomarker-defined populations and designing dosing studies within the FrontRunner paradigm.

References

1. Collins G. Friends of Cancer Research Internal Analysis.
2. Fashoyin-Aje LA, Mehta GU, Beaver JA, Pazdur R. The On- and Off-Ramps of Oncology Accelerated Approval. *New England Journal of Medicine*. 2022;387(16):1439-1442. Published online September 21, 2022. [doi:10.1056/NEJMP2208954](https://doi.org/10.1056/NEJMP2208954)
3. Project FrontRunner | FDA. Accessed October 18, 2022. <https://www.fda.gov/about-fda/oncology-center-excellence/project-fronrunner>
4. Colombo N, Dubot C, Lorusso D, et al. Pembrolizumab for Persistent, Recurrent, or Metastatic Cervical Cancer. *New England Journal of Medicine*. 2021;385(20):1856-1867. [doi:10.1056/NEJMOA2112435/](https://doi.org/10.1056/NEJMOA2112435/)
5. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *Journal of Clinical Oncology*. 2020;38(1):1. [doi:10.1200/JCO.19.02105](https://doi.org/10.1200/JCO.19.02105)
6. Motzer R, Alekseev B, Rha SY, et al. Lenvatinib plus Pembrolizumab or Everolimus for Advanced Renal Cell Carcinoma. *New England Journal of Medicine*. 2021;384(14):1289-1300. [doi:10.1056/NEJMOA2035716/](https://doi.org/10.1056/NEJMOA2035716/)
7. Motzer RJ, Hutson TE, Glen H, et al. Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: a randomised, phase 2, open-label, multicentre trial. *Lancet Oncol*. 2015;16(15):1473-1482. [doi:10.1016/S1470-2045\(15\)00290-9](https://doi.org/10.1016/S1470-2045(15)00290-9)
8. Nelson KA, Walsh D, Abdullah O, et al. Common complications of advanced cancer. *Semin Oncol*. 2000;27(1):34-44.
9. FDA. Enhancing the Diversity of Clinical Trial Populations-Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry. Published online 2020. Accessed October 17, 2022. <https://www.fda.gov/media/127712/download>
10. CFR - Code of Federal Regulations Title 21. Accessed October 30, 2022. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?CFRPart=314&showFR=1&subpartNode=21:5.0.1.1.4.8>
11. About Drugs@FDA | FDA. Accessed November 3, 2022. <https://www.fda.gov/drugs/drug-approvals-and-databases/about-drugsfda>
12. Pokrzywa CJ, Abbott DE, Matkowskyj KA, et al. Natural History and Treatment Trends in Pancreatic Cancer Subtypes. *J Gastrointest Surg*. 2019;23(4):768-778. [doi:10.1007/S11605-019-04113-3](https://doi.org/10.1007/S11605-019-04113-3)

13. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. Published online 2018. Accessed October 23, 2022. <https://www.fda.gov/media/71195/download>
14. FDA. Multiple Endpoints in Clinical Trials Guidance for Industry Biostatistics Multiple Endpoints in Clinical Trials Guidance for Industry. Published online 2022. Accessed October 23, 2022. <https://www.fda.gov/media/162416/download>
15. Project Orbis | FDA. Accessed October 19, 2022. <https://www.fda.gov/about-fda/oncology-center-excellence/project-orbis>
16. CENTER FOR DRUG EVALUATION AND RESEARCH APPLICATION NUMBER: 208716Orig1s000 MULTI-DISCIPLINE REVIEW Summary Review Office Director Cross Discipline Team Leader Review Clinical Review Non-Clinical Review Statistical Review Clinical Pharmacology Review.
17. CENTER FOR DRUG EVALUATION AND RESEARCH APPLICATION NUMBER: 761234Orig1s000 MULTI-DISCIPLINE REVIEW Summary Review Office Director Cross Discipline Team Leader Review Clinical Review Non-Clinical Review Statistical Review Clinical Pharmacology Review.
18. FDA Briefing Document Oncologic Drugs Advisory Committee Meeting. Published online 2021.

Appendix 1

Lessons Learned from Recent Drug Development Programs

Case Study 1: Abemaciclib¹⁶ – Potential challenges associated with performing interim PFS analyses.

Indication: Abemaciclib + fulvestrant for adult patients with HR+/HER2- advanced or metastatic breast cancer with disease progression following endocrine therapy.

Clinical Development Plan: Originally a single-arm Phase 2 trial designed to support an AA based on ORR in previously treated patients, with an initiated Phase 3 RCT. Submission on single-arm trial was discouraged. The Sponsor continued the Phase 3 trial with PFS as the primary endpoint. Regular approval was based on the final PFS analysis, as the originally planned PFS interim analysis did not meet the defined threshold.

Challenges Highlighted: FDA discouragement of a single-arm trial to support AA with a separate RCT to confirm benefit. Potential challenges associated with proposing submissions based on interim analysis for PFS. Statistical considerations, such as the required hazard ratios and alpha spending, may limit the ability to conduct interim analyses.

Key Learnings: In certain scenarios, such as abemaciclib + fulvestrant, with an approximately 50% overall response rate (ORR), assessing ORR and duration of response in a subset of the cohort of an ongoing RCT may be more informative as data to potentially support an AA, than a single-arm trial to support AA followed by a Phase 3 RCT. In certain disease settings, there may be benefit from endpoints other than ORR (e.g., pCR, ctDNA, MRD). However, additional evidence is needed to validate and evaluate the correlation of these endpoints, such as changes in ctDNA, with long-term outcomes to justify use as an alternative early endpoint to support regulatory approval.

Case Study 2: Relatlimab-rmbw¹⁷ – Trial design to support an approval for a first-in-class drug in the early metastatic setting.

Indication: Relatlimab-rmbw + nivolumab for adult and pediatric patients 12 years of age or older with unresectable or metastatic melanoma.

Clinical Development Plan: An adaptive trial design for a first-in-class therapy in early metastatic disease. Patients were randomized to either relatlimab+nivolumab or nivolumab in a Phase 2 study and enrollment paused for a pre-planned PFS interim analysis. The interim analysis demonstrated benefit, thus triggering enrollment of additional patients into the Phase 3 RCT, which had a primary efficacy endpoint of PFS and secondary OS and ORR (with hierarchical testing) for full approval. In melanoma, there is a well-established correlation between PFS and OS, supporting PFS as an adequate endpoint for full approval. The adaptive trial design allowed for integration of Phase 2 with Phase 3 efficacy data.

Challenges Highlighted: The level of evidence to support an approval for a first-in-class drug in the early metastatic setting may be higher than in other settings where the mechanism of action is well known. Additional considerations for trial design include the magnitude of benefit of efficacy and use as a combination therapy.

Key Learnings: This scenario highlights when it may be necessary to investigate a new regimen in earlier settings. As relatlimab was studied as a combination therapy with nivolumab, targeting patients in the early treatment setting was necessary, as many r/r patients had received

nivolumab (the control arm) in prior lines, and rechallenge with nivolumab was considered inappropriate. Scenarios evaluating an add-on therapy may necessitate investigation in earlier settings, given the need to demonstrate contribution of components and that the add-on therapy provides additional benefit compared to monotherapy alone. Here, given the mechanism of action of the drug, a significant ORR was not expected and therefore assessing ORR in an attempt for AA was not likely to be successful. Additionally, it may not be advantageous to aim for an AA if a full approval can be supported in the early setting with an RCT. There is potential for an interim PFS analysis and an earlier readout to support AA, although the Agency has discouraged this as interim PFS analyses may overestimate the true PFS. An earlier PFS interim analysis may avoid exposing too many patients to a potentially inferior therapy compared to SOC.

Case Study 3: Retifanlimab-dlwr¹⁸ – Utility of an ongoing RCT, as single-arm study data may not provide sufficient evidence to justify a regulatory decision when there is a low response rate.

Indication: Retifanlimab for adults with locally advanced or metastatic squamous carcinoma of the anal canal who have progressed on or who are intolerant of platinum.

Clinical Development Plan: A Phase 2 single-arm trial with ORR as the primary endpoint was submitted to the FDA for AA with an ongoing randomized Phase 3 trial in an earlier setting to provide confirmatory evidence.

Challenges Highlighted: A complete response letter (CRL) was issued identifying general concerns with using the data from the single-arm trial for regulatory decision-making due to the low response rate (e.g., 13.8% for retifanlimab). Further, given the high prevalence of potentially confounding factors in the intended population, determination of the safety and efficacy to inform the benefit: risk assessment was challenging in the absence of a control arm.

Key Learnings: There is significant concern with submissions based on preliminary evidence of benefit, particularly when the response rate is considered to be low. This scenario highlights the need for a RCT to be ongoing, with enrollment complete or near complete, prior to any analyses of the single-arm trial to support a regulatory submission for AA. If the drug receives AA, time to confirmation of clinical benefit is faster, and if the ORR analysis is not supportive of an AA, the RCT trial is already ongoing. Resulting data from the RCT will be more robust than a single-arm trial with clinical efficacy and safety data evaluable. However, there may be concerns raised by investigators about the ethics of initiating large Phase 3 trials when there is insufficient preliminary evidence to support the hypothesis of benefit for a new therapy. Pre-planned interim analyses for futility may be considered as one possible solution to this latter concern.

Appendix 2: Endpoint Considerations

Variability in Timing of Endpoint Readouts Based on Different Assumptions

The scenarios are not meant to be exhaustive but representative of what may occur in oncology to show, via these archetypes, how nuances arise amongst endpoints and how enrollment and event rates can come together and impact timings.

Table 1: Timing of RCT Endpoints Given Various Scenarios

Identity	Archetype	Enrollment (N)	mPFS Control (months)	ORR % (Control vs. Investigational)	ORR (N)	Enrollment Duration (months)	ORR Analysis (months)	PFS Interim (75% Info. Frac.) (months)	PFS Final (months)
A	Short PFS/ Slow Enrollment	430	10	55% vs. 70%	430	24	27	24	29
B	Longer PFS/ Slow Enrollment			50% vs. 70%	240				
C	Longer PFS/ Slow Enrollment	450	20	65% vs. 80%	360	12	24	34	44
D	Enrollment Short PFS/ Fast Enrollment			60% vs. 80%	210				
E	Longer PFS/ Fast Enrollment	660	45	55% vs. 70%	430	24	16	17	22
F	Enrollment Short PFS/ Fast Enrollment			50% vs. 70%	240				
G	Longer PFS/ Fast Enrollment	660	45	65% vs. 80%	360	24	14	26	36
H	Enrollment Short PFS/ Fast Enrollment			60% vs. 80%	210				
I	Very Long PFS	660	45	75% vs. 85%	660	24	28	41	53
J									

ORR analysis triggered 4mos after enrollment of 'ORR N' achieved
PFS analyses powered at 80% for true PFS HR = 0.70 throughout (~250 final PFS events required)
ORR analyses powered at 90% for differences displayed in column 3, with 4 mos follow-up
Assume fixed testing sequence (ORR → PFS) with overall $\alpha = 0.025$ (1-sided)

Table 2: Timing of RCT Endpoints Given Various Scenarios, Powered for OS

Identity	Archetype	Enrollment (N)	mPFS (Cntrl)	mOS (Cntrl)	ORR %	ORR N	Enrollment Duration	ORR Analysis	PFS Interim (75% Info. Frac.)	PFS Final	OS Final
A	Short PFS / Slow Enrollment	685	10 mos	20 mos	55% vs. 70%	430	34	27	24	28	49
B					50% vs. 70%	240		19			
C	Longer PFS / Slow Enrollment	685	20 mos	35 mos	65% vs. 80%	360	34	24	31	36	72
D					60% vs. 80%	210		18			

OS powered 80% at HR = 0.75 (~390 OS events required) to be tested in a gated fashion after ORR and/or PFS
 N = 685 was chosen to yield ≤ 72 mos duration to OS for the 'Longer PFS' archetype

Further Endpoint Considerations for Clinical Trials

ORR

In settings where ORR is an established endpoint reasonably likely to predict clinical benefit, it may be appropriate to consider an initial, well-designed statistical comparison of ORR in treatment arms to support accelerated approval. This comparison should be powered to demonstrate a clear clinical benefit, and what constitutes a clinically meaningful benefit in ORR depends on the disease setting and should be agreed upon upfront, along with an agreed timeframe for establishing durability of response. Analyses of ORR may require fewer patients than required to support subsequent analysis of PFS or OS in the RCT. In general, it is preferred that enrollment be mostly complete prior to ORR analysis and be conducted at a timepoint that allows adequate characterization of durability of response, the latter being dependent on the disease setting. When evaluating whether ORR should be included as an interim analysis to support AA, the timing of the readout in relation to other endpoints should be considered in addition to the appropriateness of the endpoint to predict clinical benefit in the specific disease and therapy setting.

PFS

In some cases, based on disease setting and regulatory precedence considerations, a PFS analysis may best support an initial submission for accelerated or regular approval. Additionally, with some exceptions, the trial will need to be ultimately powered for OS. In these cases, enrollment timelines may influence the timing of PFS readout, with PFS readout often occurring shortly after enrollment completes. In scenarios where PFS interim analyses read out around the same time as ORR, and the investigational therapy (alone or in combination) is not expected to increase ORR, a PFS interim analysis to support AA may be considered. However, it is acknowledged that the appropriateness of an interim PFS analysis is situationally dependent and must take into consideration the relationship of PFS to OS to determine if the proposed analysis is fit for purpose. The appropriateness of conducting interim PFS analyses in any specific clinical trial should be discussed with the FDA, as interim PFS analyses may overestimate the true PFS. Additionally, in certain disease settings with a long natural history, the time to OS analysis may be considerably long, and PFS may be appropriate to confirm clinical benefit, depending on the disease, mechanism of action of the drug, and market access considerations. If PFS is used as the primary clinical endpoint for traditional approval, studies can be smaller, and ORR may be more feasible to support AA.

OS

Depending on the mechanism of action of the drug and disease setting, confirmation of clinical benefit through OS analysis may be required. Powering a trial for OS, compared to PFS, can increase the target enrollment size as well as enrollment duration. Given this, analyses for ORR and PFS may reach maturity prior to full enrollment. It is important that clinical trials meant to verify clinical benefit be substantially enrolled, as once results are public, it can be challenging to enroll a sufficient number of US patients in the confirmatory study with enrollment often completed based on non-US patients. If the study comprises a largely non-US population, it may be more challenging to equate to US populations given differing SOC. To mitigate against such challenges, the ORR analysis could be conducted for a pre-specified subset of enrolled patients that has been agreed to with regulators, and while durability of response data are maturing, enrollment continues such that at the time of submission of the application, enrollment is almost complete. Another consideration in the early metastatic setting is the confounding of a patient receiving subsequent therapies on the final OS analysis. This is a challenge in many frontline settings and is a consideration when confirming benefit for an AA.



Real-World Evidence: Leveraging RWD to advance research

Multi-Cancer Early Detection Screening Tests: Considerations for Use of Real-World Data

Introduction

Cancers that are detected in late stages generally have a worse prognosis compared to cancers detected in earlier stages, when tumors are more amenable to effective and even curative interventions.¹ Currently there are a limited number of cancer types with available minimally-invasive standard of care (SOC) screening approaches to detect cancer earlier, and they are designed to detect only a single cancer type.² As a result, many cancers may go undetected or may be detected at later stages when treatment may not be as effective and outcomes are worse. The observed mortality benefit for screened cancers³⁻⁶ raises the possibility that safe and effective screening tests for currently unscreened cancers may reduce cancer mortality for those cancer types.

Recent innovations enable the emergence of technologies that detect the presence of multiple types of cancer from a sample of blood, i.e., a liquid biopsy. Multi-cancer early detection (MCED) screening tests are a type of liquid biopsy intended to detect cancer-associated signals at early stages, including cancers with and cancers without SOC screening modalities. Given the novel nature and the unique challenges in clinical validation associated with multi-cancer screening approaches⁷, there is an opportunity to explore innovative strategies for generating and assessing evidence to robustly characterize the safety and effectiveness of MCED screening tests.

The safety and effectiveness of cancer screening tests are usually demonstrated through evidence generation by clinical screening studies which use traditional data capture methods (e.g., electronic data capture, case report forms, patient reported outcomes) and occur in a pre-specified, selected population. To date, most screening studies designed to evaluate safety and effectiveness of FDA-approved single cancer screening devices have been prospective and observational studies.^{8,9} Data for some long-term clinical outcome endpoints, such as overall survival and cancer-specific mortality, have been generated from prospective randomized controlled trials (RCTs) and rigorous epidemiologic studies in the post-market setting (**Figure 1**).³⁻⁶

Thank You to Our Contributors

This paper reflects discussions that occurred among stakeholder groups on various challenges and opportunities related to multi-cancer early detection screening tests. The topics covered in the paper, including recommendations, are intended to capture key discussion points and should not be interpreted to reflect alignment on all topics included in the white paper by all the contributors.

Carolyn Aldige, Prevent Cancer Foundation

Seema Singh Bhan, Exact Sciences Corporation

Deepshikha Bhandari, GRAIL

Christina A. Clarke, GRAIL

Teresa Coleman, IQVIA

Michael del Aguila, GRAIL

Ruth Etzioni, University of Washington and the Fred Hutchinson Cancer Research Center

Joanne Hackett, IQVIA

Ernest Hawk, The University of Texas MD Anderson Cancer Center

Daniel F. Hayes, University of Michigan Rogel Cancer Center

Robertino Mera-Giler, Freenome

Anna Pugh, Exact Sciences Corporation

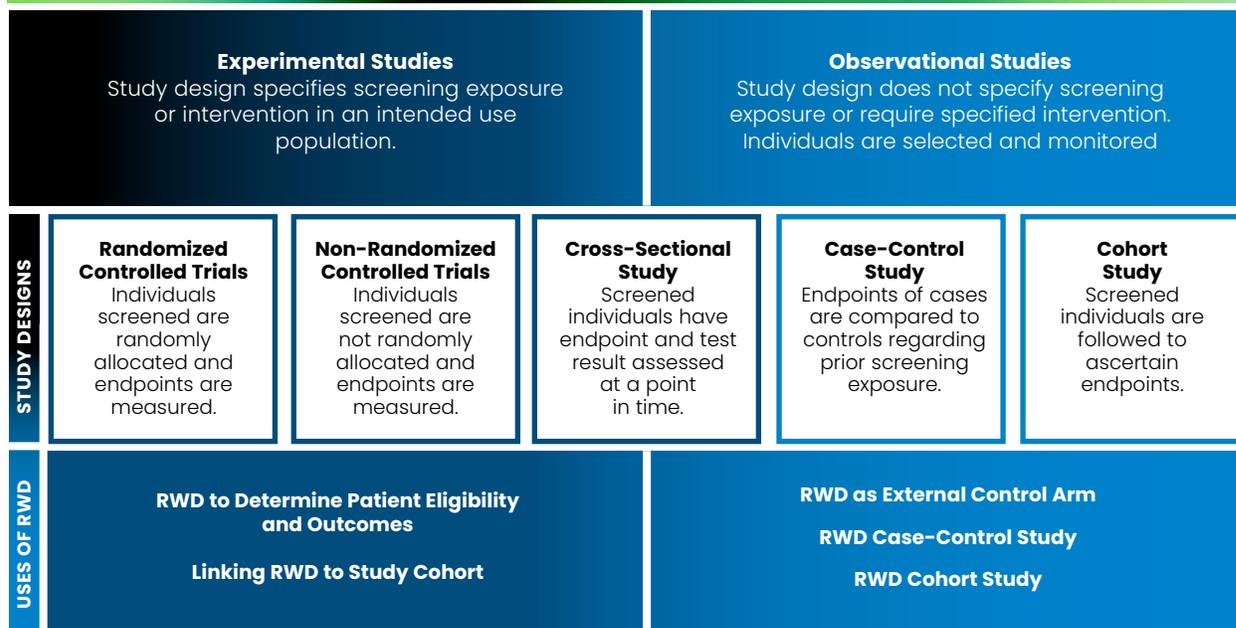
Girish Putcha, Freenome

Natalie Chavez Lau, Freenome

Sam Roosz, Crescendo Health

Sean Tunis, Rubix Health

Fig 1: Types of Study Designs for Screening Tests and Incorporation of Real-World Data (RWD)



Evidence generation for screening tests can occur through experimental and observational studies, where RWD may be incorporated in a variety of ways, including hybridized methods. This figure, which provides examples for use of RWD, is meant to be directional and not intended to be a comprehensive list of study designs and objectives.

Conducting clinical screening studies, such as RCTs, to generate the appropriate evidence of the clinical validity and utility of MCED screening tests may be logistically challenging. Appropriately powering studies for each cancer type, particularly for rare cancers, requires large enrollment numbers (i.e., on the order of tens of thousands of participants), extensive resourcing, and one or more decades of longitudinal follow-up to demonstrate a cancer-specific mortality benefit for individual cancer types across the large set of cancer types in the intended use population.¹⁰ Additionally, highly-controlled clinical studies with protocol screening and follow-up procedures (including diagnostic procedures) may not reflect the real-world screening, adherence, and clinical practice, which also may evolve over time. To help overcome these challenges with clinical screening studies, real-world data (RWD) may be able to supplement data generated by clinical screening studies to assess MCED screening tests. RWD are data collected during the course of usual patient care and can be used to generate real-world evidence (RWE). For the purposes of this white paper, the group focused on RWD as defined by FDA: Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources, including electronic health records (EHRs), claims and billing data, and product and disease registries.¹¹

The use of RWD to assess MCED screening tests to support regulatory decision-making requires careful forethought to ensure the data collected can address key assessment questions, while also acknowledging and planning for data necessary to support the test's clinical utility, as designing studies that include endpoints addressing both clinical validity and utility can

support multiple purposes (e.g., regulatory decision making, reimbursement decisions, etc.). To provide overarching considerations for generating evidence about MCED screening tests using RWD, Friends of Cancer Research (*Friends*) assembled a multi-stakeholder group of experts including government officials, MCED screening test developers, academic clinicians and researchers, patient advocacy groups, and RWD partners and vendors. We first identified endpoints to consider capturing in RWD and then reviewed opportunities for using RWD study designs to support an understanding of MCED screening test safety and effectiveness. This work complements that of others focused on assessment of MCED screening tests, exploring platform trial designs, and identifying novel endpoints for evidence generation about clinical validity and utility.

Objectives

When captured and analyzed appropriately, RWD can be used to generate RWE to evaluate the safety and effectiveness of a medical product.¹¹ The group focused on identifying opportunities to generate meaningful RWE to supplement evidence for regulatory decision-making for MCED screening tests, while also considering opportunities for data collection over the continuum of evidence generation. Within the context of current study designs, RWE is likely to serve a supplementary role and be part of the totality of evidence in an initial premarket application. However, as our understanding of these novel tests evolves and the robustness of RWD is better understood, the use of RWD may expand. This should ultimately be informed by conversations between regulators and sponsors.

The group's objectives were to:

- Identify potential endpoints (including performance metrics and clinical outcomes) that could be captured from RWD sources to assess the clinical validity and utility of MCED screening tests,
- Characterize opportunities and challenges associated with using RWD to support assessment of MCED screening tests, and
- Highlight key considerations for using RWD to generate RWE to support assessment of MCED screening tests.

Every assay may have unique characteristics that are not covered by this document. Further, MCED screening technology is an evolving area, and as evidence continues to build, the optimal approach for assessment of MCED screening tests may also evolve and adapt. MCED test developers are strongly encouraged to submit a pre-submission to FDA to discuss the details about their specific test.

RWE Generation for Assessment of MCED Screening Tests

MCED screening tests use various technologies to detect cancer signals, therefore evaluation approaches may differ both across MCED screening tests and when compared to current screening tests. Some MCED screening tests provide a likelihood score for the tissue of origin (TOO), sometimes referred to as the cancer signal origin (CSO), while other MCED screening tests prompt clinical follow up of positive test results using imaging modalities like whole body PET-CT to identify the TOO. Additionally, analytic approaches to determine safety and effectiveness in multiple cancers are different from a focus on a single cancer, as seen with

currently available screening tests. These differences suggest a need for a review of the current regulatory and evidence development paradigms to assess clinical validity and utility to inform potential solutions.

RWD studies may offer some logistical advantages over clinical screening study designs. RWD studies have the potential to provide data over a lengthy follow-up period to generate evidence about long-term outcomes encompassing a large number of subjects in the intended use population, including those with rare cancers. RWD also provides information reflective of the real-world population setting about diagnoses, screening frequencies, biopsy compliance, and treatment patterns, including as these may evolve over time. However, RWD is subject to its own limitations due to the observational setting and the generation of RWD for administrative and billing, rather than research purposes. These limitations can lead to issues with non-random missing data, mismeasured data, and selection bias. Despite these limitations, RWD represents an opportunity to explore and propose additional, pragmatic solutions to assess MCED screening tests.

While clinical screening studies continue to be a key component and the foundational source of evidence for in vitro devices, there may be opportunities for RWD studies to inform regulatory decisions for MCED screening tests. Previously published FDA guidance notes that RWD of sufficient quality may potentially be used to inform or support a particular regulatory decision for medical devices and diagnostics, with the specific use determined by the specific type of technology¹¹, including use of RWD as:

- Generating hypotheses to be tested in a prospective clinical study,
- A historical control, a prior in a Bayesian trial, or as one source of data in a hierarchical model or a hybrid data synthesis,
- A concurrent control group, or as a mechanism for collecting data in a setting where a registry or some other systematic data collection mechanism exists,
- Evidence to identify, demonstrate, or support the clinical validity of a biomarker,
- Evidence to support FDA approval or authorization,
- Support for a petition for reclassification of a medical device,
- Evidence for expanding the label to include additional indications for use or evidence to update the labeling to include new information on safety and effectiveness,
- Public health surveillance efforts,
- To conduct post-approval studies that are imposed as a condition of device approval or to potentially preclude the need for postmarket surveillance studies ordered under section 522 of the FD&C Act,
- In certain circumstances, for use in generating summary reports of Medical Device Reports (MDRs), and
- To provide postmarket data in lieu of some premarket data.¹¹

Key Questions for Assessment of MCED Screening Tests

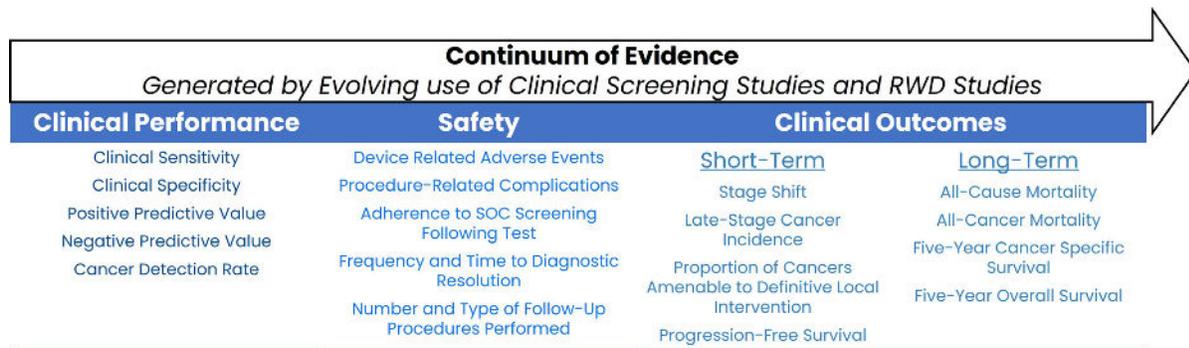
To assess MCED screening tests throughout the product life cycle (e.g., premarket, post-market data collection, benefit-risk determinations), the working group identified key questions to frame necessary evidence generation, endpoints, and data:

1. **Performance Characteristics:** How well does the MCED screening test detect cancer? How early does the test detect cancer?
2. **Safety:** What are the health burdens/harms of MCED screening tests, including the diagnostic confirmation process?
3. **Clinical Outcomes and Utility:** How does an MCED screening test impact cancer outcomes?

Continuum of Evidence Generation

The working group mapped out these key questions in the context of the continuum of evidence generation, which can be supported by data from both prospective clinical screening studies and RWD sources (Figure 2). To help answer these questions, we identified a list of possible endpoints to consider.

Figure 2: The Continuum of Evidence Generation and Proposed Endpoints to Help Answer Key Questions



Evidence can be generated to assess MCED screening tests by clinical screening studies using traditional data capture methods and/or RWD sources, with the types of studies evolving over time and purpose of evidence generation. Possible endpoints to consider for generating evidence are highlighted along the continuum of clinical validity and utility. The specific endpoints necessary to establish clinical validity and utility may vary depending on the technology. The safety and clinical outcomes endpoints require the use of the test in patient management.

Intended Use Considerations for Evidence Generation

Evidence generation should be conducted in the intended use population. Designing the study plan and identifying appropriate endpoints will be influenced by the intended use of the test, as well as the interval of MCED test screening, with considerations including:

- The TOO component of the test. While MCED screening tests detect cancer-associated signals generally, regulatory expectations are for the TOO to be identified, either with TOO

ascertainment built within the assay capabilities and followed by diagnostic confirmation, or by a follow-up methodology (e.g., PET-CT) after a cancer signal is detected.

- The intended use population on the label. The MCED screening test's intended use population defined on the assay label may differ among tests, including age cut-offs, specific types of cancers detected, point of use in the clinical care pathway (e.g., complement of the test to SOC screening tests), and risk profile of individuals eligible for the test. For example, the test may be intended for only high-risk populations, including those with a genetic predisposition, occupational or environmental exposure, a history of cancer, or specific lifestyle factors.

Potential Endpoints to Evaluate MCED Screening Tests

To understand what data are needed to assess MCED screening tests in the various phases of evidence generation, the working group defined possible endpoints that assess clinical validity and clinical utility in the context of the key questions that were asked. Clinical validity is the ability of the test to accurately identify cancer, as well as identify TOO, while clinical utility is the likelihood that patients managed in accordance with test results will demonstrate improved health outcomes, such as a reduction in late-stage cancer diagnoses and mortality.¹² Many of these endpoints encompass both clinical validity and utility. Analytical validity, which confirms that the test accurately measures the target analytes in the blood, is assumed to have been established as part of product development and is not included in the scope of this work. As indicated above, specific endpoints may vary depending on the intended use population of the MCED screening test. **Appendix Table 1** provides aligned definitions for each of the proposed endpoints and is not meant to be a comprehensive list of endpoints.

Performance Characteristics: How well does the MCED screening test detect cancer? How early does the test detect cancer?

It is critical to determine that an MCED screening test detects cancer, including at an earlier stage than it would otherwise be clinically diagnosed. The evidence should show that the MCED screening test returns a positive result in individuals who have cancer (sensitivity), while also providing a negative result for individuals who do not have cancer (specificity).

Cancers vary in preclinical latency, and MCED screening tests will vary in sensitivity per cancer, and across stages, based on a variety of factors unique to each test. Therefore, performance should be reported both in the aggregated form for all cancer detection, as well as for individual cancer types, with performance stratified by stage. In general, screening test sensitivity and specificity are initially assessed via retrospective evaluations, such as case-control studies, in which pathologically confirmed cases and suspected non-cases are examined.^{13,14}

The observed sensitivity and specificity in the intended use population will depend on the diagnostic accuracy of the confirmatory diagnostic test (e.g., PET-CT), which may differ for different cancers. Therefore, it is important to document the evaluation workflow, and develop methods to address the imperfect accuracy of the confirmatory tests.

In the prospective screening setting, sensitivity and specificity are challenging to assess. One method to establish true sensitivity and specificity would be to require full body imaging and pathological confirmation of all individuals for all cancers included in the test, but this method would be impractical because of an undue burden for patients. An approximation of screening test sensitivity can be given by the ratio of screen-detected cancers to the sum of screen- and interval-detected cancers at a given point in time.¹⁵⁻¹⁷ This estimate of screening test sensitivity may be affected by multiple factors including overdiagnosis, preclinical latency, previous screening history, and the time interval chosen. This estimate may deviate from the true sensitivity and will not necessarily match estimates obtained from already-diagnosed cases.^{18,19}

Just as an MCED screening test may exhibit variability in its ability to detect different cancers, there may also be variability in TOO accuracy. The same is true for TOO assessment by PET-CT, in which the accuracy differs for different cancer types. Therefore, performance should be reported in aggregate form for all cancers, and on a per-cancer basis based on TOO assessment.

Additional measures of diagnostic performance under prospective screening are the positive and negative predictive values (PPV and NPV). A high PPV implies a low rate of unnecessary confirmation tests or biopsies, but alone is not a reliable indicator of the likely benefit of the test. Time elapsed without a confirmed cancer diagnosis after a negative test result could be used to determine the NPV; in this case a long interval without cancer following a negative test would indicate that the test was a true negative. The necessary monitoring time for individuals with a negative test result will depend on the cancer type, its given natural history, the effectiveness of the related diagnostic workup, and the interval for any established SOC screening for the cancer type.

Performance Characteristics Endpoints Include (Calculated based on detection of cancer signal and cancer signal detection +TOO):

- *Clinical Sensitivity*
- *Clinical Specificity*
- *Positive Predictive Value*
- *Negative Predictive Value*
- *Cancer Detection Rate*

Safety: What are the health burdens/harms of MCED screening tests, including the diagnostic confirmation process?

Although not specific to MCED screening tests, FDA has released general guidance that details considerations for the assessment of probable benefit and risks/harms of a device, including the risks of adverse events directly related to the test as well as those related to the follow-up diagnostic procedures after a positive test result. Adverse events include both physical and psychological negative occurrences. Additional evidence generated from patient reported outcomes regarding quality of life and anxiety may support an understanding of these adverse events.²⁰ Although not specific to MCED screening tests, based on this guidance, the timing of an assessment of safety should include the interval from administration of the MCED screening test until the determination of cancer status is complete.

There are many facets of MCED screening tests' safety that will factor into the benefit-risk assessment. The first safety concern is how a positive MCED screening test result impacts an individual's health care journey due to follow-up procedures to establish a definitive diagnosis. It will be important to analyze the number and type of follow-up procedures performed, any complications, and the frequency and time to diagnostic resolution. Lack of a diagnostic resolution following a positive MCED test could lead to adverse effects on an individual's quality of life. Theoretically, if MCED screening tests detect cancers that would otherwise go undetected, in the short-term, more surgeries and procedures may occur leading to more safety concerns; however, over a longer term, the net safety profile may improve since the individual may avoid complications and costs associated with diagnosis of (and treatment for) their cancer at later stages. Stratifying the safety outcomes by cancer type will also be important, as the benefit-risk profiles for the diagnostic resolution will vary across cancer types.

The second safety concern is how MCED screening might impact SOC screening. Tests currently in development are expected to have multiple intended uses, including complementing SOC screening. For such tests, whether individuals tested adhere to SOC screening may inform their impact and implications, so SOC screening among individuals who have such an MCED screening test should be recorded to determine if there are changes.

Safety Endpoints Include:

- *Device Related Adverse Events (Physical and Psychological)*
- *Procedure-Related Complications (Physical and Psychological)*
- *Adherence to SOC Screening Following Test*
- *Frequency of Confirmation Diagnostic Tests and Time to Diagnostic Resolution*
- *Number and Type of Follow-Up Procedures Performed*

Clinical Outcomes and Utility: How does an MCED screening test impact clinical cancer outcomes?

To demonstrate that MCED screening tests improve clinical outcomes, evidence must show the test detects cancers that are otherwise undetected before symptoms appear (i.e., at earlier stages) and reduces morbidity and mortality associated with cancer and its treatment. It is important to evaluate endpoints that measure both short- and long-term outcomes.

Short-term endpoints ascertained shortly after the determination of cancer status can support evidence that the test detects cancer earlier and may ultimately translate into improvements in cancer-specific morbidity or mortality (e.g., reductions in late-stage cancer diagnosis). Defining early-stage cancer will likely be cancer type specific but can be considered to mean cancers generally amenable to local intervention for curative intent, whereas late-stage cancers usually cannot be cured via localized treatments. Stage shift is a possible surrogate for the impact of an MCED screening test on disease mortality. There is a concern that increasing the proportion of early-stage cancers may lead to overdiagnosis without any effect on late-stage detection²¹; therefore, a decrease in incidence of late-stage disease is more informative to support an understanding of the likely implications of stage shift.

Long-term endpoints require data capture over multiple years, such as survival and mortality, and are necessary because short-term endpoints may not translate into longer term reductions in morbidity and mortality. Disease-specific mortality is the primary measure of clinical utility for cancer screening trials and the most reliable indicator of whether a cancer screening test reduces deaths from cancer.²² Other long-term and short-term endpoints can complement this primary endpoint but can be difficult to interpret on their own. All-cause mortality has been discussed as an endpoint in single-cancer screening studies but may not be sensitive enough to discern screening benefit.²³ Survival endpoints can also be difficult to interpret due to lead-time and length bias.²⁴ The limitations and potential biases impacting interpretability of each of these endpoints should be carefully noted. As with the clinical validity endpoints described above, assessing the clinical utility by cancer type, in addition to all-cancer, will be important.

Clinical Outcomes Endpoints Include:

- *Short-Term Endpoints*
 - *Stage Shift*
 - *Late-Stage Cancer Incidence*
 - *Proportion of Cancers Amenable to Definitive Local Intervention*
 - *Progression-Free Survival*
- *Long-Term Endpoints*
 - *All-Cancer Mortality*
 - *All-Cause Mortality*
 - *Five-Year Cancer Specific Survival*
 - *Five-Year Overall Survival*

Proposed Data Elements Necessary to Generate Evidence

Generating evidence for the endpoints described above will require rigorous data capture with appropriate ontologies and validated definitions, including specific data elements (suggestions included in **Table 1**). It will also be critical to identify and capture the selection factors that characterize the individual receiving an MCED screening test to understand the representativeness of this population and generate a list of potential confounding or selection variables for comparative and causal studies. Further, it will be important to capture any comorbidities or risk factors for cancer that the individuals have, as comorbidities may influence long-term outcomes, and risk factors for cancer can provide additional information about the risk profile of the population receiving the MCED screening test. Previous FDA guidance, not specific to MCED screening tests, for selecting the study population recommends including individuals across the entire range of disease states, with relevant confounding medical conditions, and across different demographic groups to prevent bias in estimates of test performance.²⁵ An obstacle to unbiased clinical utility analyses includes potential differences in the post-diagnosis treatment pathways for those who receive the test and those who do not. Therefore, collecting treatment information to understand treatment pathways following diagnosis will be valuable.

Table 1: Suggested Data Elements to Consider for Evidence Generation to Support Assessment of MCED Screening Tests

Category	Data Elements	Endpoint Category
Patient Characteristics	<ul style="list-style-type: none"> • Age at Time of Test • Gender • Race/Ethnicity • Socioeconomic Status • Insurance Status • Access to Care for Diagnosis and Treatment • Comorbidities 	Demographics/Intended Use
Cancer Risk Factors	<ul style="list-style-type: none"> • Family History • Smoking History • Alcohol Use • Obesity, Diet, and Exercise • Genetic Predisposition • Prior Cancer History (Cancer Type, Diagnosis Date, Previous Treatments) • Other Risk Factors 	Demographics/Intended Use
MCED Screening Test Administration	<ul style="list-style-type: none"> • Reason for Test Administration • Test Administered • Test Result (Positive/Negative and TOO) • Adverse Events with Administration 	Clinical Performance, Safety
SOC Screening	<ul style="list-style-type: none"> • Adherence to Appropriate SOC Screening Methods • SOC Screening Results 	Clinical Performance, Safety

<p>Clinical Confirmation</p>	<ul style="list-style-type: none"> • Imaging Recommended and Performed • Biopsy Recommended and Performed • Other Procedures Recommended and Performed for Definitive Diagnosis • Results from Definitive Diagnostic Procedures (Cancer Present/Absent) • Time to Diagnostic Resolution • Adverse Events with Confirmatory Procedures • Other Cancer(s) Detected that were Not Tested for or were a Negative Result using the MCED Test 	<p>Clinical Performance, Safety</p>
<p>Cancer Characteristics</p>	<ul style="list-style-type: none"> • TOO • Stage • Histology • Subtype • Method of Detection (if cancer is not detected by the MCED screening test, such as clinical findings, symptoms, etc.) 	<p>Clinical Performance, Clinical Outcomes</p>
<p>Cancer Treatment</p>	<ul style="list-style-type: none"> • Relevant Treatments for Cancer, Including Doses and Duration 	<p>Safety, Clinical Outcomes</p>
<p>Clinical Outcomes</p>	<ul style="list-style-type: none"> • Living Status (Dead/Alive) • Duration of Clinical Follow-Up • Cause of Death, if applicable • Progression or Metastasis (and time) • Disease-Free Survival • Morbidity 	<p>Clinical Outcomes</p>

Adapted and modified.²⁶

RWD Study Design Informed by Characteristics of Cancer Types

One key factor to consider regarding the benefit-risk profile is the characteristics of the specific cancers reported by the MCED screening test. While MCED screening tests detect the presence of a cancer signal in general, it will be important to consider specific cancer characteristics including the incidence and availability of SOC screening modalities, and aggressiveness (or natural history) of the cancer types being evaluated.

Incidence

Cancers with low incidence may be difficult to assess using non-RWD clinical screening studies, and therefore are more likely to have limited evidence to understand the benefit-risk profile for these cancers. RWD enables an analysis of the performance of MCED screening tests on a scale (tens to hundreds of thousands) that is difficult to achieve in a time- and cost-effective manner with traditional clinical screening studies, allowing for evidence generation for rare cancers. Moreover, RWD is also valuable to use in studies that evaluate test performance for cancers that have moderately high incidence and may require large numbers to sufficiently power the analysis.

Existence of Recommended Standard of Care Screening

Cancers with SOC screening recommendations may have more standardized pathways that allow for aligned RWD capture compared to those without such recommendations. For example, cancers with existing United States Preventive Services Task Force (USPSTF) A or B recommendations (breast, cervical, colorectal, and lung cancer) have standard diagnostic pathways which can be captured routinely in RWD. As a result, the follow-up to clinically confirm cancer is more aligned across settings, which can help standardize collection and assessment of RWD for endpoint measurements. In cancers without SOC screening, there may be higher variability in the workup for diagnostic confirmation, creating challenges for the use of RWD to ascertain cancer diagnoses.

Natural History

Variations in the natural history, or aggressiveness, of different cancer types may also affect data capture and evidence generation. Indolent cancers grow slowly and rarely metastasize or contribute to cancer-related death, resulting in better clinical outcomes. Conversely, highly aggressive cancers usually form, grow, or spread quickly, generally resulting in worse morbidity and mortality outcomes.¹ Therefore, the time frame for RWD data captures will be influenced by the natural history and aggressiveness of the cancer (e.g., time to ascertain false negative).

Types of RWD Study Design for MCED Screening Test Assessment

RWD may be incorporated into study designs in a variety of ways, with varying levels of reliance on the RWD in the overall study design, and may include hybrid methods incorporating RWD with traditional study data (Figure 1).²⁷ At one end of the spectrum, traditional RCTs may use RWD elements, such as selected outcomes identified using EHR or claims data. In the middle are trials in clinical practice settings that may be RCTs with pragmatic designs or single arm studies using a RWD external control arm. Lastly, studies may be designed to collect data following a

'usual care' model that is not mandated by study protocol, captured completely through RWD, either with data collection designed prospectively or using existing data infrastructure. One potential strategy to improve data quality, consistency, and completeness is to prospectively design data capture, such as the use of a registry specifically designed for assessment of MCED screening tests. Determining the best study design to support the assessment of a specific MCED screening test will require discussions between the test developer and FDA. Examples of possible use cases for RWD are highlighted in **Table 2**, illustrating the advantages and challenges associated with use of RWD. The possible use cases provided are suggestions and should not be viewed as prescriptive.

The value of RWD depends on the data quality, consistency, and completeness. FDA has previously and generally outlined how to determine whether RWD is fit-for-purpose (not for MCED screening tests).²⁷ FDA does not endorse a particular RWD source but assesses the relevance and reliability of the source and its elements for appropriate use. If RWE is generated from multiple RWD sources, each RWD source must be evaluated individually as well as in aggregate to determine appropriate use.¹¹

Further guidance may be helpful to clarify the appropriate RWD sources, types of data important to capture, and considerations for capture specific to MCED screening tests. For example, evaluation of clinical performance measures in the RWD setting may be subject to selection bias, as patients who receive an MCED test may be systematically different from those who do not, in terms of patient characteristics and disease risk. Further, patients who select MCED testing, and those who receive a positive versus negative test result, may receive different follow-up imaging tests and treatments than those who do not. Accounting for these differences while determining appropriate comparison groups and study designs will be critical. Further, as RWD sources have increased in availability and accessibility, the comparative effectiveness community has generated a host of analytic methods designed to address these challenges to be able to validly draw inferences about the risk and benefit of interventions based on RWD.²⁸⁻³⁰

Table 2: Possible Use Cases of RWD for Generating Evidence for MCED Screening Tests

Use Case	Description	Advantages of RWD	Challenges to Use of RWD
<p>RWD External Control Arm</p>	<p>The control arm is fully comprised of RWD, reflecting the intended use population but without the use of MCED screening tests</p>	<ul style="list-style-type: none"> Reduces the need for patients participating in control arms, for which they may stand to gain no potential clinical benefit Potential to reduce cohort size necessary to demonstrate clinical validity and utility Potential to establish a platform study with the same concurrent comparator across MCED screening tests Potential to achieve adequate comparison by using advanced matching methodology and causal modeling 	<ul style="list-style-type: none"> Historic data may be less suitable than concurrent collection due to variability in cancer incidence, SOC screening adherence, and exposure to risk factors over time Characteristics of RWD cohort may be quite different than those of study cohort, creating challenges for propensity matching A nonrandomized control may introduce bias into detection rate comparison
<p>Participant-Consented RWD Collection</p>	<p>Approach that supports patients in exercising their rights to access their own data to contribute to the study (via record requests or APIs)</p>	<ul style="list-style-type: none"> Data gathered from SOC of study participants can serve as salvage pathway to adjudicate outcomes for patients who otherwise would be lost to follow-up For some cancer types it may be sufficient as primary means of adjudicating study outcomes (e.g., through ascertaining cancer diagnoses in EHR or claims data) 	<ul style="list-style-type: none"> Requires consent and involvement of the patient, making it more suited to prospective than retrospective studies Logistically and technically complex workflows

<p>Linking of De-identified Data to Study Cohort</p>	<p>Data from aggregated de-identified sources is linked to study cohort to expand available data for analysis</p>	<ul style="list-style-type: none"> • Opportunity to more deeply profile study population • In some cases can address gaps in data through linking to external sources • May be performed retroactively in some instances 	<ul style="list-style-type: none"> • In most cases, available comprehensive data will only overlap with a small subset of the studied population. Larger overlap may be possible, but at the expense of comprehensiveness (e.g., a participant may be found in the external data, but the dataset lacks the relevant info) • Subject to potential data quality issues such as misclassification bias
<p>Post-Marketing RWD Studies</p>	<p>Study that aggregates RWD for patients who have received a commercially available MCED test</p>	<ul style="list-style-type: none"> • Ability to assess the impact of various factors that might not reach statistical significance in pivotal study • Potential to support clinical validation for expanded set of cancer types, including rare cancers • Creates opportunity to collect healthcare resource utilization data to support considerations for guideline inclusion and reimbursement • Creates opportunity to amass information on signals currently of unknown clinical significance 	<p>Similar limitations to study types outlined above depending on approach used</p>

Conclusions

This white paper helps identify possible endpoints for assessing MCED screening test performance and characterizes opportunities for capturing these endpoints from RWD. RWE generated from the application of MCED screening tests in the real-world intended use population may help supplement data generated in non-RWD clinical screening studies and may be used to inform regulatory decisions. It is critical that the MCED test developer and regulator align on a plan for the types of data and evidence generation necessary to support regulatory decision-making about the MCED screening test, including approval and post-approval studies to update or expand the label or provide additional supportive evidence.

In addition to this working group, there are many ongoing efforts surrounding the development, assessment, and use of MCED screening tests to support robust characterization of the safety and effectiveness of MCED screening tests while facilitating development and continued innovation for these technologies in a timely manner. Discussions with our working group highlighted the need for alignment on terminology used in MCED screening test development, validation, and evaluation, and an effort by BloodPAC is underway to develop a lexicon for the field. Further work is also needed for the design of clinical screening studies evaluating the clinical validity and utility of MCED screening tests. The National Cancer Institute (NCI) is evaluating the landscape of study designs and seeking to potentially launch a multi-arm, multi-stage, pivotal RCT to evaluate multiple MCED screening tests in the years ahead.³¹

While we have suggested possible data elements and endpoints to capture in RWD, work is still needed to federate data into a common model for ease of future analyses. Another area that needs attention is the use of machine learning and artificial intelligence by many of the MCED screening tests to determine cancer status. RWE can enable real-world learning and evaluation of these technologies as they enter clinical practice, helping to achieve the goal of a learning healthcare system. There may be a role for RWE in periodic (post-market) re-evaluation of MCED screening tests utilizing machine learning to assess the real-world performance of initial and future versions of these tests.

We hope this document supports efforts to collect robust, consistent, and relevant data from various studies and helps optimize evidence generation to facilitate development of MCED screening tests and integration into clinical care.

Disclaimer: This paper reflects discussions that occurred among stakeholder groups, including governmental agencies, on various topics. The topics covered in the paper, including recommendations, therefore, are intended to capture key discussion points. The paper should not be interpreted to reflect alignment on the different topics by the participants, and the recommendations provided should not be used in lieu of FDA published guidance or direct conversations with the Agency about a specific development program. This paper should not be construed to represent views or policies of the US Federal Government.

Glossary

Cancer: A disease in which cells grow and proliferate uncontrollably, not including pre-cancerous lesions.

Clinical Screening Studies: Prospectively designed studies in the intended use population using traditional data capture methods (e.g., electronic data capture, case report forms, patient reported outcomes).

Clinical Utility: The likelihood that use of a test will change the management of patients and, by doing so, improve health outcomes, including, for example, safety, morbidity, quality of life, resource utilization, or survival and mortality.¹²

Clinical Validity: The accuracy with which a test identifies, measures, or predicts the presence or absence of a clinical condition in a patient (e.g., the likelihood that someone with a positive test actually has the specified cancer).¹²

Early-Stage Cancer: Specific TNM stage will vary depending on cancer type but is generally a localized cancer amenable to local intervention for curative intent.

Late-Stage Cancer: Specific TNM stage will vary depending on cancer type but is generally a cancer that has metastasized, and is not amenable to localized intervention.

Liquid Biopsy: The detection of biomarkers using only a blood or fluid sample.

Multi-Cancer Early Detection (MCED) Screening Test: Assays using different technologies to detect cancer-associated biomarkers, such as circulating tumor cells, tumor DNA, and other analytes, to screen for cancers in a defined patient population.

Real-World Data (RWD): Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources, including electronic health records (EHRs), claims and billing data, and product and disease registries.¹¹

Real-World Evidence (RWE): The clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of RWD.¹¹

Tissue of Origin (TOO): The tissue source of the primary cancer (e.g., breast, lung, etc.); also, sometimes referred to as Cancer Signal Origin (CSO).

References

1. Noone AM, Howlader N, Krapcho M et al., ed. SEER Cancer Statistics Review, 1975–2018 – SEER Statistics. National Cancer Institute Accessed November 30, 2021. https://seer.cancer.gov/csr/1975_2018/
2. Hackshaw A, Cohen SS, Reichert H, Kansal AR, Chung KC, Ofman JJ. Estimating the population health impact of a multi-cancer early detection genomic blood test to complement existing screening in the US and UK. *Br J Cancer*. 2021;125(10). doi:10.1038/s41416-021-01498-4
3. Chudgar NP, Bucciarelli PR, Jeffries EM, et al. Results of the National Lung Cancer Screening Trial: Where Are We Now? *Thorac Surg Clin*. 2015;25(2):145. doi:10.1016/J.THORSURG.2014.11.002
4. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to Fourteen-Year Effect of Screening on Breast Cancer Mortality. *JNCI J Natl Cancer Inst*. 1982;69(2):349–355. doi:10.1093/JNCI/69.2.349
5. Schröder FH, Hugosson J, Roobol MJ, et al. The European Randomized Study of Screening for Prostate Cancer – Prostate Cancer Mortality at 13 Years of Follow-up. *Lancet*. 2014;384(9959):2027. doi:10.1016/S0140-6736(14)60525-0
6. Chiu HM, Chen SLS, Yen AMF, et al. Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the One Million Taiwanese Screening Program. *Cancer*. 2015;121(18):3221–3229. doi:10.1002/CNCR.29462
7. Putcha G, Gutierrez A, Skates S. Multicancer Screening: One Size Does Not Fit All. *JCO Precis Oncol*. 2021;(5):574–576. doi:10.1200/PO.20.00488
8. SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED): Stool DNA-Based Colorectal Cancer Screening Test. Accessed February 24, 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf13/P130017b.pdf
9. SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED): Human Papillomavirus DNA detection kit. Accessed February 24, 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf10/P100020b.pdf
10. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials*. 2000;21(6 Suppl). doi:10.1016/S0197-2456(00)00098-2
11. CBER, CDRH. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices Guidance for Industry.; 2016. Accessed May 18, 2021. <https://www.fda.gov/media/99447/download>
12. NCI Dictionary of Genetics Terms – National Cancer Institute. Accessed February 15, 2022. <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>
13. Liu MC, Oxnard GR, Klein EA, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020;31(6):745–759. doi:10.1016/J.ANNONC.2020.02.011/ATTACHMENT/6577D31A-6339-4982-95C7-B370542D40E5/MMC4.PDF
14. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926–930. doi:10.1126/SCIENCE.AAR3247
15. Toyoda Y, Nakayama T, Kusunoki Y, Iso H, Suzuki T. Sensitivity and specificity of lung cancer screening using chest low-dose computed tomography. *Br J Cancer*. 2008;98(10):1602–1607. doi:10.1038/SJ.BJC.6604351
16. Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. *J Med Screen*. 2007;14(4):174–177. doi:10.1258/096914107782912077
17. Lehman CD, Arao RF, Sprague BL, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*. 2017;283(1):49–58. doi:10.1148/RADIOL.2016161174

18. Ten Haaf K, Van Rosmalen J, De Koning HJ. Lung cancer detectability by test, histology, stage and gender: estimates from the NLST and the PLCO trials. *Cancer Epidemiol Biomarkers Prev*. 2015;24(1):154. doi:10.1158/1055-9965.EPI-14-0745
19. Hamashima C. Overdiagnosis of gastric cancer by endoscopic screening. *World J Gastrointest Endosc*. 2017;9(2):55. doi:10.4253/WJGE.V9.I2.55
20. CDRH. Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and De Novo Classifications Guidance for Industry and Food and Drug Administration Staff Preface Public Comment. Accessed January 21, 2022. <https://www.fda.gov/media/99769/download>
21. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol*. 2014;15(6):e234. doi:10.1016/S1470-2045(13)70598-9
22. What Cancer Screening Statistics Really Tell Us – National Cancer Institute. Accessed March 10, 2022. <https://www.cancer.gov/about-cancer/screening/research/what-screening-statistics-mean>
23. Weiss NS, Koepsell TD. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11). doi:10.1093/JNCI/94.11.864
24. Duffy SW, Nagtegaal ID, Wallis M, et al. Correcting for Lead Time and Length Bias in Estimating the Effect of Screen Detection on Cancer Survival. *Am J Epidemiol*. 2008;168(1):98-104. doi:10.1093/AJE/KWN120
25. CDRH. Guidance for Industry and FDA Staff – Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. Published online 2007. Accessed February 12, 2022. <https://www.fda.gov/media/71147/download>
26. Etzioni R, Gulati R, Weiss NS. Multicancer Early Detection: Learning From the Past to Meet the Future. *JNCI J Natl Cancer Inst*. 2021;(0):168. doi:10.1093/JNCI/DJAB168
27. Concato J. FDA Clinical Investigator Training Course Real-World Evidence. Published online 2021.
28. Hernán MA. Methods of Public Health Research – Strengthening Causal Inference from Observational Data. *N Engl J Med*. 2021;385(15):1345-1348. doi:10.1056/NEJMP2113319/SUPPL_FILE/NEJMP2113319_DISCLOSURES.PDF
29. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1033-1039. doi:10.1002/PDS.4297
30. Haneuse S, Daniels M. A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why? *EGEMS (Washington, DC)*. 2016;4(1):16. doi:10.13063/2327-9214.1203
31. NOT-CA-22-033: Request for Information (RFI): Seeking Input from Multi-Cancer Early Detection Test Developers on Readiness for Participation in an NCI-Sponsored Clinical Utility Randomized Controlled Screening Trial. Accessed January 24, 2022. <https://grants.nih.gov/grants/guide/notice-files/NOT-CA-22-033.html>

Appendix Table 1: Defining Endpoints for Evaluating MCED Screening Tests

Category	Endpoint	Definition
Clinical Performance	Clinical Sensitivity <i>*Calculated based on detection of cancer and cancer detection+TOO, both stratified by stage</i>	All-Cancer Sensitivity: The proportion of subjects with clinically confirmed cancer in whom the MCED screening test was positive. Cancer-Specific Sensitivity: The proportion of subjects with a specific clinically confirmed cancer type (e.g., breast cancer) in whom the MCED screening test accurately identified that specific cancer type.
	Clinical Specificity * <i>Calculated based on detection of cancer and cancer detection+TOO</i>	All-Cancer Specificity: The proportion of subjects without clinically confirmed cancer of any type in whom the MCED screening test was negative.
	Positive Predictive Value (PPV) * <i>Calculated based on detection of cancer and cancer detection+TOO, both stratified by stage (if applicable)</i>	All-Cancer PPV: The proportion of MCED screening test positive subjects who have any clinically confirmed cancer. Cancer-Specific PPV: The proportion of MCED screening test positive subjects with the TOO accurately identified (e.g., breast cancer) who have clinically confirmed cancer of that type.
	Negative Predictive Value (NPV)	The proportion of MCED screening test negative subjects who do not have cancer of any type.
	Cancer Detection Rate	The proportion of cancers detected by the MCED screening test out of the cancers expected in the population monitored over a defined period of time (requiring a control arm or acceptable external reference cohort).

Safety	Device Related Adverse Events (AEs) <i>Serious vs. Non-Serious Events</i>	Any untoward medical occurrence (physical or psychological) directly before, during, or directly after the MCED screening test is administered that is directly attributable to the test. Serious events include: events that may have been or were attributed to the use of the device and produce an injury or illness that is life-threatening, results in permanent impairment or damage to the body, or requires medical or surgical intervention to prevent permanent harm to the body. ²⁰
	Procedure-Related Complications <i>*Stratified by TP vs. FP, Cancer Type, and Serious vs. Non-Serious Events</i>	Any harm to screened individuals that is not directly attributable to the test itself but relates to any untoward medical occurrence (physical or psychological) after the MCED screening test is administered until definitive diagnosis (i.e., determination of cancer status).
	Adherence to SOC Screening Following Test <i>*Stratified by Test-Positive and Negative</i>	The proportion of subjects who have all their USPSTF A or B recommended cancer screening tests (e.g., mammogram, colonoscopy, low-dose chest CT, and cervical screening) completed within the recommended period.
	Frequency of Confirmation Diagnostic Tests and Time to Diagnostic Resolution	The time between receiving a positive MCED screening test result and determination of both the presence or absence of cancer, and specific cancer type.
	Number and Type of Follow-Up Procedures Performed <i>*Stratified by TP vs. FP, Cancer Type, and Invasive vs. Non-Invasive Events</i>	The number and type of medical procedures performed after the MCED screening test is administered that support the definitive diagnosis.
Clinical Outcomes	Short-Term Outcomes	
	Stage Shift <i>*Stratified by Cancer Type</i>	An increase in the proportion of cancers detected in early- versus late-stage disease with and without the MCED test, with a concurrent decrease in the proportion detected in late-stage disease.
	Late-Stage Cancer Incidence <i>*Stratified by Cancer Type</i>	The number of new cancer cases diagnosed at a late stage per 100,000 people per year.
	Proportion of Cancers Amenable to Definitive Local Intervention <i>*Stratified by Cancer Type</i>	The proportion of cancers diagnosed in the specified population where definitive, curative, local intervention is clinically feasible.

Clinical Outcomes	Progression-Free Survival <i>*Stratified by Cancer Type</i>	The length of time from diagnosis of cancer to first clinical evidence of disease progression.
	Long-Term Outcomes	
	All-Cause Mortality	The total number of deaths occurring in the population, regardless of the cause of death, in a specified time period.
	All-Cancer Mortality <i>*Stratified by Cancer Type</i>	The total number of deaths occurring in the population due to cancer in a specified time period.
	5-Year Cancer Specific Survival <i>*Stratified by Cancer Type</i>	The probability of surviving cancer in the absence of other causes of death in a 5-year period.
	5-Year Overall Survival <i>*Stratified by Cancer Type</i>	The percentage of patients alive in the population five years after their cancer diagnosis.

Exploring the Potential of External Control Arms created from Patient Level Data: A case study in non-small cell lung cancer

Xiang Yin^a, Pallavi S. Mishra-Kalyan^b, Rajeshwari Sridhara^b, Mark D. Stewart^c, Elizabeth A. Stuart^d, and Ruthanna C. Davi^a

^aIntegrated Evidence, Acorn AL, a Medidata Company, New York, NY, USA; ^bOffice of Biostatistics, U. S. Food and Drug Administration Silver Spring, MD, USA; ^cScience and Policy Friends of Cancer Research, Washington, DC, USA; ^dBloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Randomized controlled trials (RCTs) are the gold standard for evaluation of new medical products. However, RCTs may not always be ethical or feasible. In cases where the investigational product is available outside the trial (e.g., through accelerated approval), patients may fail to enroll in clinical trials or drop out early to take the investigational product. These challenges to enrolling or maintaining a concurrent control arm may compromise timely recruitment, retention, or compliance. This can threaten the study's integrity, including the validity of results. External control arms (ECAs) may be a promising augmentation to RCTs when encountered with challenges that threaten the feasibility and reliability of a randomized controlled clinical trial. Here, we propose the use of ECAs created from patient-level data from previously conducted clinical trials or real-world data in the same indication. Propensity score methods are used to balance observed disease characteristics and demographics in the previous clinical trial or real-world data with those of present-day trial participants assigned to receive the investigational product. These methods are explored in a case study in non-small cell lung cancer (NSCLC) derived from multiple previously conducted open label or blinded phase 2 and 3 multinational clinical trials initiated between 2004 and 2013. The case study indicated that when balanced for baseline characteristics, the overall survival estimates from the ECA were very similar to those of the target randomized control, based on Kaplan–Meier curves and hazard ratio and confidence interval estimates. This suggests that in the future, a randomized control may be able to be augmented by an ECA without compromising the understanding of the treatment effect, assuming sufficient knowledge, measurement, and availability of all or most of the important prognostic variables.

ARTICLE HISTORY

Received 25 February 2021
Accepted 19 November 2021

KEYWORDS

External control arm;
propensity score; oncology;
non-small cell lung cancer;
clinical trials

1. Introduction

Randomized controlled trials (RCTs) are the gold standard for measuring the effects of interventions or treatments on outcomes, but in some settings a randomized control may not be an option for reasons such as lack of equipoise, no options for a control treatment, or an inability to enroll a sufficient number of patients (FDA 0000). In addition, feasibility or cost concerns (Ali et al. 2019) may preclude the RCT and lead to the use of single arm trials to obtain preliminary evidence of efficacy in early phase trials; such trials have most often been used in oncology and rare disease.

External controls – mainly in the form of benchmarking or references to study-level summary outcomes reported in published literature – have been used over many years to assess the likely effects of investigational therapies in single arm trials where randomization is not an option (Ghadessi et al.

CONTACT Ruthanna C. Davi  rdavi@medidata.com  Acorn Ai, a Medidata Company

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

2020). However, the investigational group and the aggregated study group may not be comparable in terms of the patients' pretreatment characteristics, so direct comparison of outcomes might yield a biased estimate of the treatment effect (Guidance for Industry: E9 Statistical Principles for Clinical Trials 1998; Guidance for Industry: E10 Choice of Control Group and Related Issues in Clinical Trials 2001).

Propensity score methods are often used in observational studies to estimate the effects of treatment or exposure when the groups that are being compared might have systemic differences. The same methodologies can be applied to interventional studies, such as clinical trials when randomization is not implemented, and where external controls are being used. Such external controls utilizing patient-level data and propensity score methods for balancing baseline composition are referred to by many names, such as synthetic control arm, historical control arm, and non-experimental control. We utilize the term 'external control arm' as a general term for a control arm that is not enrolled or randomized concurrently to the investigational arm of interest and to stress that the statistical principles underlying this nonrandomized comparison are not new, although they are being applied in a novel way, in the setting of medical product development.

The propensity score is defined as a subject's probability of receiving a treatment conditional on observed baseline characteristics (Austin 2011a). Conditional on the true propensity score, subjects who are in the investigational and external control arms have similar distributions of observed (but not necessarily unobserved) baseline covariates. This allows for a fairer estimation of the treatment effects by allowing researchers to use these propensity scores in matching, weighting, covariate adjustment or subclassification to equate the treatment and control groups on the observed characteristics (Austin 2011a; Austin and Mamdani 2006; Rosenbaum 1987; Rosenbaum and Rubin 1983). However, attention should also be given to the possibility of differences in unobserved baseline covariates. Sensitivity or tipping point analyses are crucial in exploring how unmeasured or unknown key prognostic factors which cannot be balanced may impact the estimate of the treatment effect (Friends of Cancer Research White Paper 0000; Liu et al. 2013).

We now turn to a discussion of the four key ways of using propensity scores to create groups that are balanced with respect to the observed covariates.

In propensity score matching, matched sets of investigationally treated and external control subjects with similar propensity scores are constructed. If the matched sets have similar distributions of baseline covariates, then inferences can be made about treatment effects by directly comparing outcomes (Imbens 2004). A spectrum of matching algorithms has been developed, for example, k:1, variable ratio, full matching, greedy matching or optimal matching, with or without replacement, with or without caliper, etc. (Stuart 2010).

Weighting using the propensity score involves weighting the treatment and control groups to have similar covariate distributions. Inverse probability of treatment weighting is a common weighting approach, where the weight of each individual is calculated as the inverse of their probability of receiving the treatment that they actually received (Allan et al. 2020; Austin 2011a). This means that the weight of a treated subject is equal to the inverse of the subject's propensity score, and the weight of a control subject is the inverse of the subject's probability of being assigned to the control group (i.e., one minus the subject's propensity score). Alternative weighting techniques using propensity scores, such as weighting by odds of propensity score (Morgan and Todd 2008) can be used to estimate the "average treatment effect on the treated" (ATT) (Austin and Stuart 2015; Morgan and Todd 2008). A newly developed approach called the overlap weighting method can be considered when extreme propensity scores are of concern (Li et al. 2019).

Stratification (or sub-classification) on the propensity score entails ranking subjects based on their estimated propensity score, and placing them into mutually exclusive subsets based on previously defined propensity score thresholds. Often, five or more subsets are used; however, there is no consensus on the optimal number of subsets, and more are often needed for sufficient bias reduction.

Table 1. Features of previously conducted clinical trials.

Data (from multiple previously conducted clinical trials)	Design Open label or blinded, phases 2 or 3	Region Multi- national	Start/End of Trial(s) Began between 2004 and 2013. Ended between 2007 and 2016.	Indication Previously treated locally advanced or metastatic non- small cell lung cancer	Endpoints Overall survival	Number of Patients 1,399	Control regimen Docetaxel

Table 2. Eligibility criteria for patients from previously conducted clinical trials.

- (1) Inclusion in a previously conducted clinical trial accessible within this project
- (2) NSCLC stage III or IV at baseline
- (3) Received prior platinum-based chemotherapy
- (4) Men and women \geq 18 years of age
- (5) Eastern Cooperative Oncology Group (ECOG) performance status of \leq 2
- (6) Had measurable disease
- (7) Assigned to receive docetaxel as study treatment

Table 3. Clinically important baseline covariates utilized in propensity score estimation.

- (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)

Finally, covariate adjustment using the propensity score involves regressing the outcome variable on an indicator variable that denotes treatment status and the propensity score. Instead of simultaneously including many covariates, the regression models for the outcome analysis can simply include one single propensity score in place of those covariates. A linear regression model may be selected for continuous outcomes, while a logistic regression model may be chosen for dichotomous outcomes and a survival analysis model such as the Cox regression model may be used for time-to-event outcome.

In the context of simulating a randomized controlled trial, propensity score matching, weighting, and stratification have an important advantage over covariate adjustment by the propensity score. Propensity score matching, weighting and stratification can separate the analyses undertaken to balance the treatment groups (analogous to the design of a study) from analyses examining outcomes. Covariate adjustment does not have a distinct step for assessing covariate balance and directly uses the propensity score as a covariate in the outcome analysis. Separation of the balancing step from the analysis of outcomes is advantageous since it allows the usual clinical trial design feature of pre-specifying primary and secondary efficacy analyses and methods to be mirrored. This helps to avoid criticism of cherry picking the most favorable results. In addition, the covariate adjustment approach can be sensitive to the model specification, and generally does not lead to as much bias reduction as the other propensity score methods described.

For the remainder of this paper, we focus primarily on propensity score matching. This approach is appealing in its similarity to a 1:1 randomized trial in that all patients contribute approximately equally to a direct comparison of two groups. Similar case studies could be conducted using propensity score weighting or stratification.

2. Motivating example/case study

The case study is based on multiple previously conducted clinical trials in non-small cell lung cancer (NSCLC) and examines whether an external control arm (ECA) can provide similar overall survival estimates to a randomized control arm from an RCT. One of the trials, a randomized controlled completed study, was defined a priori as the target trial and patient-level data from control arms of all other trials were utilized in the creation of the ECA. Individual patients for the ECA were selected from the previously conducted clinical trials to match the control patients in the target trial in terms of key baseline characteristics and prognostic factors, and using propensity score matching. This contrasts with other external controls, such as benchmarking with wholesale study level results from historical clinical trials, where comparability of baseline composition is often not assured.

2.1. Previously conducted clinical trials

The previously conducted clinical trials data used in this case study originated from open label or blinded phase 2 and 3 multinational trials initiated between 2004 and 2013 (Table 1). Enrollment in the target trial started in February 2004, with a primary efficacy analysis time point in March 2007. All patients had previously been treated and presented at baseline with locally advanced or metastatic NSCLC. All patients had been in study arms that assigned docetaxel treatment. The docetaxel dosing regimen, 75 mg/m² docetaxel day 1, 1-hour intravenous, every three weeks, was the same for all trials included in this analysis. Overall survival was a key endpoint in all trials and is the primary focus of the case study. The objectivity of this endpoint is advantageous so that the varied open label or blinded nature of the trials likely would not adversely affect the assessment of the accuracy of the external control. The death events were clearly recorded in data collection from the clinical trials and it would be unlikely that they would be dependent on evaluator knowledge of treatment assignments. Other, more subjective outcomes may be subject to bias and could be problematic in assessing the viability of an external control in those cases. Baseline or screening measurements of the following prognostic factors were available for all patients: age, number of years from cancer diagnosis to baseline, race (white vs others), sex (female vs male), smoking (current vs former vs never), histology (squamous vs non-squamous), stage (III vs IV), ECOG (0 vs 1 vs 2), prior surgery (yes/maybe vs no), and EGFR/KRAS mutation (positive vs no/unknown). Aligning with common statistical practices, all available prognostic factors that were even possibly relevant were included in the prespecified propensity score model. The case study was created from completed NSCLC RCTs drawn from the Medidata Enterprise Data Store (MEDS) (Medidata website 0000). This platform comprises some 22,000 clinical trials, conducted by the pharmaceutical industry for drug or medical product development, with patient-level data recorded by electronic data capture. Information was also sourced from Project Data Sphere (projectdatasphere.org 0000) a platform where the research community can share patient level data from academic and industry phase 3 cancer clinical trials. However, neither Project Data Sphere, LLC nor the owners of any information from the website have contributed to, approved, or are in any way responsible for the contents of this paper.

The eligibility criteria for patients from previously conducted clinical trials to be included – shown in Table 2 – were met at baseline by the 1,399 patients included in this case study. Patient level source data, including screening and baseline measurements from previously conducted clinical trials, were used for these determinations.

Table 3 shows the baseline covariates available across all studies that were used in the propensity score matching process. Those variables were considered clinically important prognostic factors and were commonly collected in the clinical trials for the indication of locally advanced or metastatic NSCLC.

3. Methods

3.1. Matching methods: Creation of an ECA

Based on guidelines proposed by Ho, the propensity score matching was carried out using the three steps described below (Ho et al. 2007).

1: Propensity score estimates were developed, representing the probability of assignment of target trial control therapy conditional on the baseline characteristics using logistic regression:

$$p(x) = P(T = 1|X = x)$$

In this equation, T denotes the control in the target trial ($T = 1$)/historical control ($T = 0$) and X is a vector that represents the covariates to be included in the propensity score model. The predictors included in this model are all available baseline characteristics. These baseline covariates were used without further variable selection or trimming to obtain optimal balance between the matched subjects. Using a large set of covariates is recommended to help satisfy the underlying assumption that there are no unobserved confounders and to reduce bias, with some researchers recommending that the analysis should include all available baseline covariates if the sample size permits (Lim et al. 2018).

2: The ECA was created by choosing historical patients to match control patients in the target trial using estimated propensity scores. Greedy nearest-neighbor matching without replacement, with a fixed 1:1 matching ratio with a caliper of 0.25 was used; this aligns with the 1:1 randomization ratio frequently used in NSCLC trials. SAS/STAT® 15.1 PROC PSMATCH was used for propensity score matching.

3: A post-matching evaluation of covariate balance was carried out. The true propensity score should be a balancing score. This study examined how similar the distribution of measured baseline covariates was between the matched target trial control arm and the historical ECA subjects. Baseline demographic and disease characteristics were summarized with descriptive statistics for the target trial control arm and ECA – both before and after matching. Standardized differences (SD) in covariate means were computed and compared before and after matching.

For continuous covariates, the standardized difference is expressed as follows:

$$SD = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{(s_t^2 + s_c^2)/2}}$$

where

\bar{x}_t and \bar{x}_c are the sample mean of the covariate for the target trial control and historical control groups, respectively; and 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:

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

where

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

The absolute standardized differences should typically be less than 0.25 (Stuart and Rubin 2007) with a figure below 0.10 being taken to show a negligible difference between treatment groups in terms of the mean or prevalence of a covariate (Normand et al. 2001). In addition, the matching process was assessed by examining the distribution of propensity scores – as well as individual baseline characteristics, including prognostic factors – between the target trial control arm and the ECA, using graphical

Table 4. Baseline characteristics by arm before and after matching.

Baseline Characteristic	Before Matching		Matched		Unmatched Control in Target Trial (N = 93)
	Pool of Eligible Patients from Previously Conducted Clinical Trials(N = 940)	Control in Target Trial (N = 459)	ECA (N = 366)	Control in Target Trial (N = 366)	
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	267 (28%)	74 (16%)	66 (18%)	71 (19%)	3 (3%)
Former	436 (46%)	276 (60%)	211 (58%)	208 (57%)	68 (73%)
Never	237 (25%)	109 (24%)	89 (24%)	87 (24%)	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	334 (36%)	112 (24%)	85 (23%)	100 (27%)	12 (13%)
1	545 (58%)	306 (67%)	254 (69%)	233 (64%)	73 (78%)
2	61 (7%)	41 (9%)	27 (7%)	33 (9%)	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%)

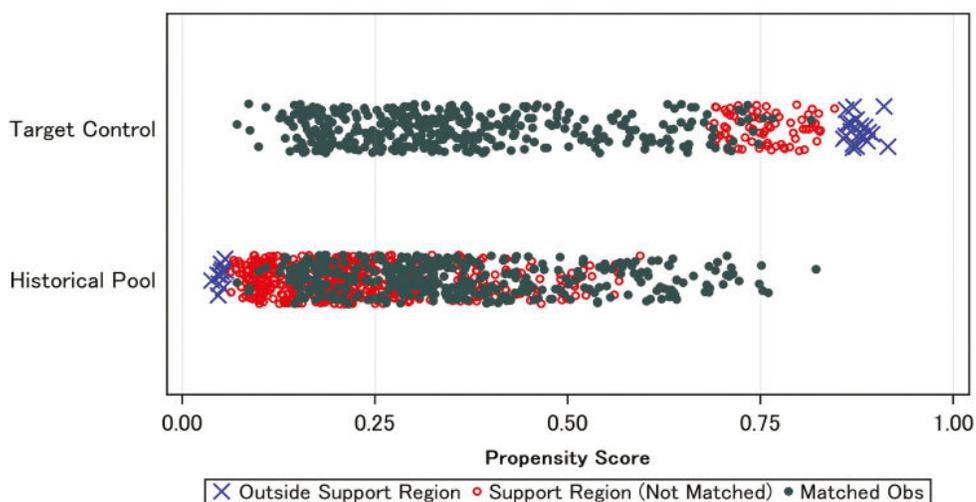


Figure 1. shows the distribution of propensity scores before and after matching. Cloud plot for illustrating distribution of propensity scores. Distributions of propensity scores indicating overlap in baseline condition of target control and historical pool patients, including matched observations (in solid dots), those within the common support region but not matched (in open dots), and those outside the support regions and not matched (in crosses). Common support region is the widest interval such that both groups have patients whose propensity scores lie within this interval. The lower endpoint of the region is the larger of the minimum propensity score for the treated group and the minimum propensity score for the control group. The upper endpoint is the smaller of the maximum propensity score for the treated group and the maximum propensity score for the control group. Only matched patients are included in the ECA and target control comparison.

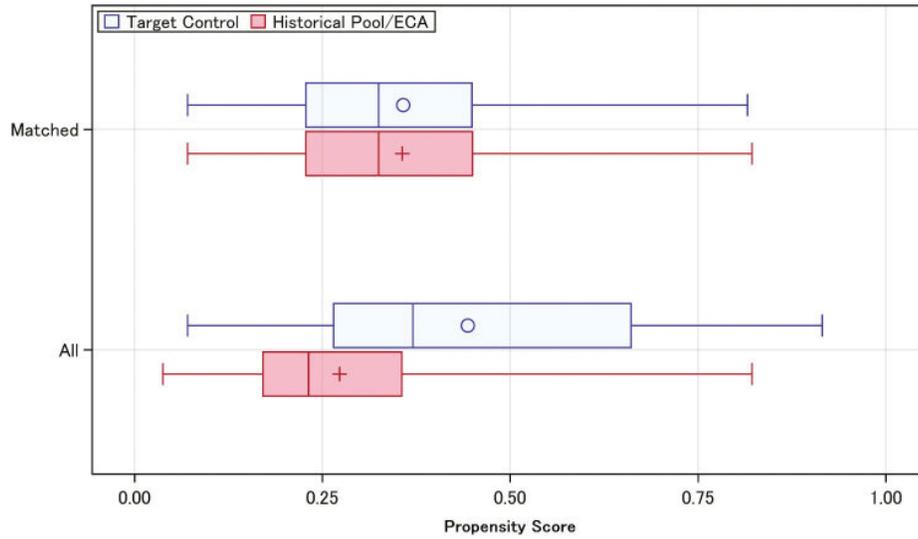


Figure 2. Distribution of propensity scores before and after matching. Distributions of propensity scores before and after matching, indicating more comparable baseline condition of target control and ECA (i.e., after matching) than between the target control and the historical pool (i.e., before matching).

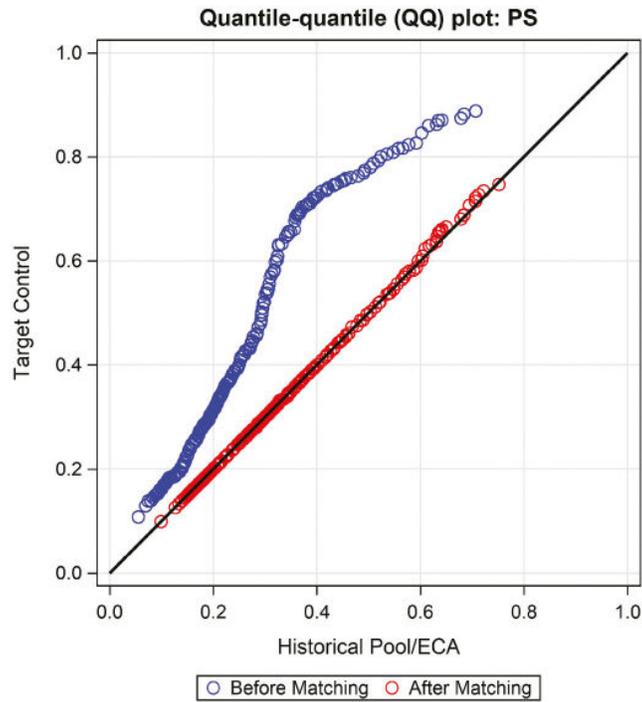


Figure 3. Propensity scores before and after matching. Quantile-quantile (Q-Q) plot compares the probability distributions of the target control and historical pool/ECA groups on the propensity score by plotting their quantiles against each other and indicates improved balance after matching relative to before.

methods such as cloud, box and quantile-quantile (Q-Q) plots. For continuous covariates, we also summarized the mean and maximum deviation between the two empirical distributions in the Q-Q plots on the scale of the variables being measured.

3.2. Outcome analysis – use of the ECA

The similarity of overall survival (OS) between the ECA and target trial was examined with the hazard ratio and associated 95% confidence interval estimated by Cox regression model and descriptive log-rank test and corresponding *p*-value for both before and after matching. The index date for overall survival was defined as the start of study treatment. This enabled an assessment of whether the overall survival in the control arm of the target trial was replicated by the ECA. Kaplan-Meier curves were presented, along with estimates of the median and other percentiles of survival times and 95% confidence intervals both before and after matching.

4. Results

4.1. Balance of baseline characteristics

As noted earlier, propensity score matching was used with the goal of selecting appropriate patients from the pool of eligible patients from previously conducted clinical trials for inclusion in the ECA, with the distribution of observed baseline characteristics well balanced between the ECA and the control arm of the target trial. This section discusses evidence suggesting that the matched groups are indeed well balanced for all observed baseline characteristics. Although not the focus of this case study, in practice, the impact of unobserved baseline characteristics on the estimation of the treatment effect should be considered.

A total of 1,399 patients were assigned to the control treatment of docetaxel in the qualified previously conducted trials (Table 4), and among those, 459 patients were included in the control arm of the target trial and the remaining 940 patients were used to create an ECA for the control arm of

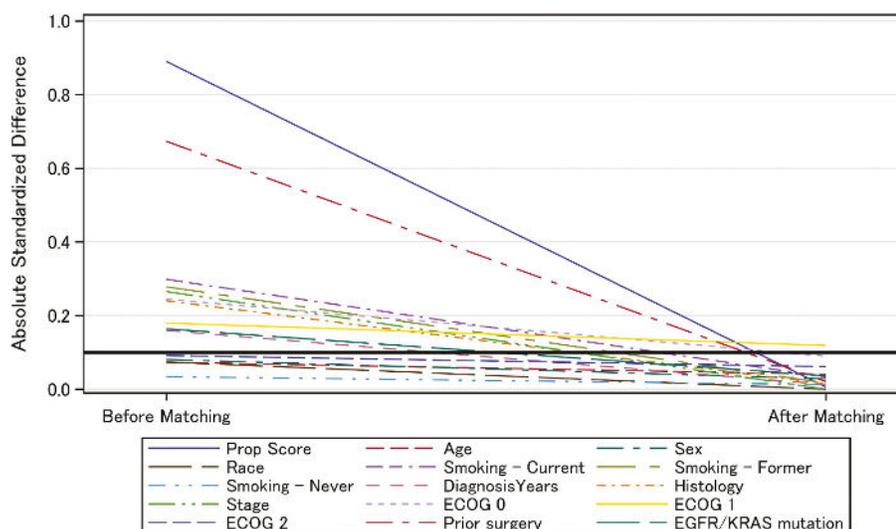


Figure 4. Plot of absolute standardized differences of important baseline covariates before and after matching. Line plots of absolute standardized differences before and after propensity score matching for the propensity score and each prognostic variable utilized in the propensity score model illustrate better balance after matching than before (lines have negative slope). All but one absolute standardized difference after matching are less than 0.01, a commonly used rule of thumb for determining adequate balance.

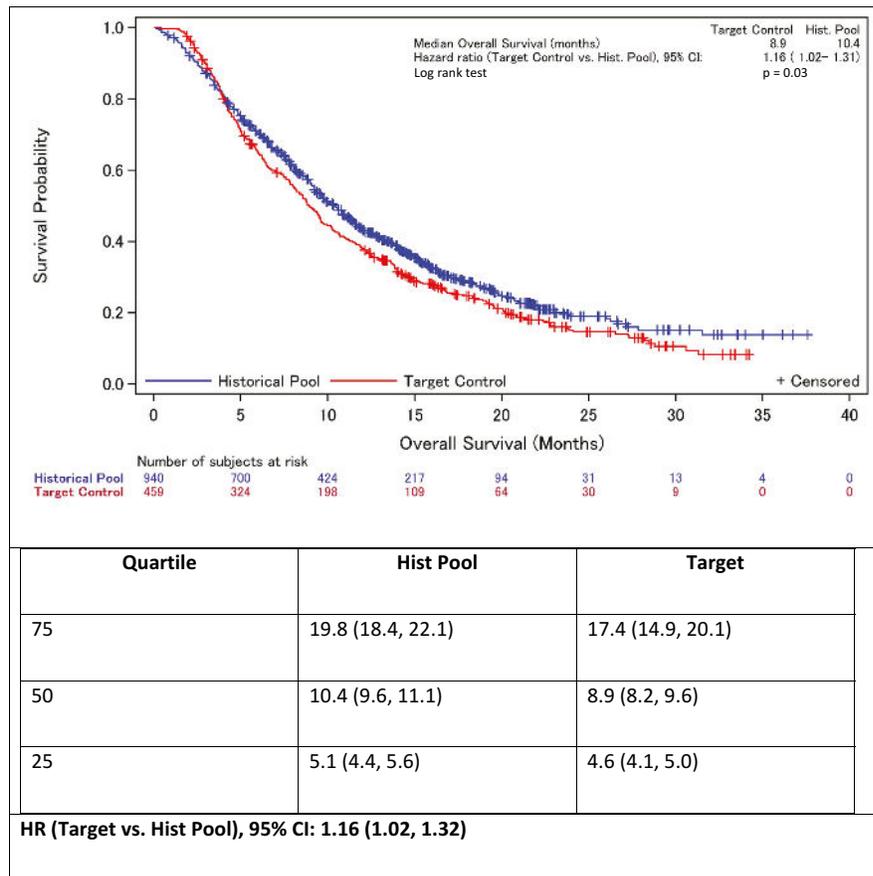


Figure 5. Comparison of overall survival in control arm of target trial versus pool of eligible patients from previously conducted clinical trials (before matching). Kaplan-Meier estimates of overall survival in the randomized control arm of the target trial versus the pool of eligible patients from previously conducted clinical trials (before matching) are different as evidenced by little overlap in the curves and a nominally statistically significant log rank test ($p = .03$).

the target trial. The cloud plot in [Figure 1](#) illustrates the high degree of overlap between the distributions of propensity scores for the control arm of the target trial and the pool of historical patients available for the ECA. The solid dots show patients who were successfully matched with a patient in the opposite group with a comparable propensity score. A common support region is defined as the largest interval that contains propensity scores for patients in both treatment groups. The region is extended by 0.25 times a pooled estimate of the common standard deviation of the logit of the propensity score (SAS/STAT® 15.1 PROC PSMATCH User Guide). The open dots (support region not matched) and crosses (outside support region) show patients where no match was available; these patients were excluded from further analysis, a common practice when using matching methods. This approach often reduces sample size and may compromise extrapolation but improves balance between groups rather than compromising it (in essence, prioritizing internal validity over external validity).

[Figure 2](#) shows a box plot of the distribution of the propensity score before and after matching; [Figure 3](#) shows a Q-Q plot of these data.

The distributions of propensity scores for the control arm of the target trial and the pool of eligible patients from previously conducted clinical trials (including all patients before matching) are shown in the lower boxplots in [Figure 2](#), while the upper boxplots show distributions after matching. Before

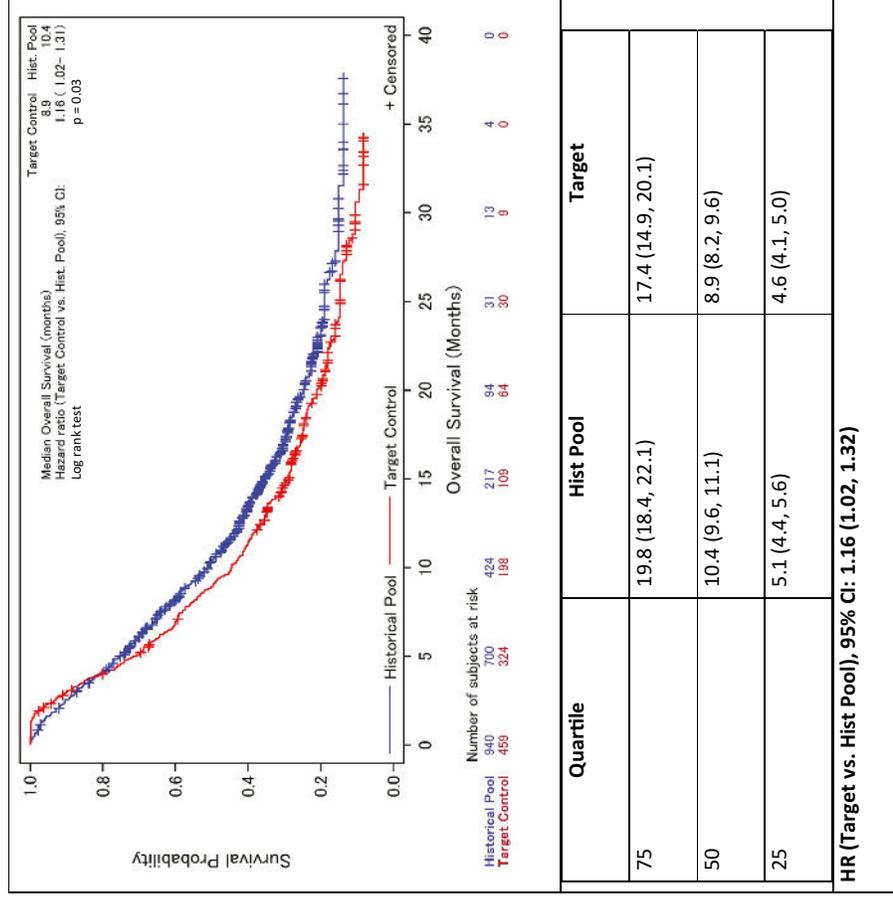


Figure 6. Comparison of overall survival in control arm of target trial versus ECA (After matching). Kaplan-Meier estimates of overall survival in the target control of the target trial versus the ECA (after matching) are similar as evidenced by overlap in the curves and a nominally insignificant log rank test ($p = .65$).

matching, there is significant discordance between the control arm of the target trial and pool of eligible patients from previously conducted clinical trials. After matching, the median and variability of the two groups (i.e., matched target control and ECA) are very similar as shown by the similar placement of the median line and width of the 'box' in the boxplots.

The Q-Q plot shown in [Figure 3](#) illustrates the similarity of propensity scores between groups before (black circles) and after matching (gray triangles). A 45-degree line, as shown after matching, confirms equal distributions. Before matching, the data points fall away from this 45-degree line, indicating that the degree of similarity in the distributions after matching is better than before matching. The mean (standard deviation) difference in propensity score between the two groups in the Q-Q plots decreased from 0.121 (0.065) before matching to 0.001 (0.003) after matching.

The standardized difference between the target and pool of eligible patients from previously conducted clinical trials (before matching)/ECA (after matching) for each key baseline characteristic is illustrated in [Figure 4](#).

In terms of baseline characteristics, most of the 459 patients included in the control arm of the target trial were white (65%), male (63%), and current smokers (16%) or former smokers (60%), as shown in [Table 1](#). Prior surgery was reported in 35% of patients; the rate of known EGFR or KRAS mutation (genetic changes that are prognostic indicators in these patients) was 7%; and the majority of patients had non-squamous type NSCLC (78%), ECOG performance status scores of 0 or 1 or 2 (24%, 67%, and 9%, respectively), and disease stage IV (87%).

As illustrated in [Table 2](#), the 940 patients from previously conducted clinical trials for potential inclusion in the ECA were similar to the target trial control arm in terms of age, years since cancer diagnosis, race, gender, ECOG score, and EGFR/KRAS mutation status. However, there were differences between the pool of eligible patients from previously conducted clinical trials and target trial controls in terms of the proportion of current smokers (28% vs. 16%), former smokers (46% vs. 60%), non-squamous disease (87% vs. 78%), disease stage IV (77% vs. 87%), and reported prior surgery (9% vs. 35%). After matching; however, all baseline characteristics were well balanced between the ECA and the target trial control arm.

[Figure 5](#) compares the overall survival for the control arm of the target trial and the pool of eligible patients from previously conducted clinical trials (before matching), while [Figure 6](#) shows overall survival for the matched target control and the ECA (after matching).

Before matching, there is limited overlap of the Kaplan–Meier curves and a visible space between the curves, indicating that the overall survival for the control arm of the target trial is shorter than for the pool of eligible patients from previously conducted clinical trials (the median survival was 8.9 months in the target group and 10.4 months in the pool of eligible patients from previously conducted clinical trials). The hazard ratio for the target relative to the pool of eligible patients from previously conducted clinical trials was 1.16 with a confidence interval that excludes 1 (95% CI 1.02, 1.32). This difference is supported by the descriptive log rank test comparing the OS curves ($p = 0.03$).

However, after matching there is significant overlap in the Kaplan–Meier curves for the target and ECA (with a median survival of 8.8 months in the target and 9.2 months in the ECA). The hazard ratio for the target relative to the ECA was 1.04 with a confidence interval that includes 1 and suggests a plausible range for the HR of 0.88 to 1.23, indicating similar overall survival for the ECA and the target trial control arm. This similarity is supported by the descriptive log rank test comparing the OS curves ($p = 0.65$).

5. Discussion and Conclusion

This NSCLC case study provides one example where an external control consisting of matched cohorts of patients from previously conducted clinical trials was successfully created using propensity scores derived from observed baseline characteristics. Further research will be needed, likely including additional case studies in additional indications, to confirm the overall acceptability, feasibility and appropriateness of the approach in a variety of clinical areas. In this case study, the

propensity score was shown to reduce the imbalance between historical patients and investigational patients. And importantly, the overall survival for the ECA was very similar to that of the randomized control from the target trial, an early step towards suggesting that an ECA could augment randomized controls in future trials involving difficult-to-study indications and populations without compromising the scientific understanding of the treatment effect. This could help mitigate many of the recruitment, retention, and crossover challenges that can occur when enrolling or maintaining a concurrent control due to rarity of the disease, or availability of the investigational agent outside the study.

This case study utilized clinical trials data from second line NSCLC patients assigned to treatment with docetaxel. The use of docetaxel as standard of care in this indication is clinically well established with few changes, if any, for many years. This may have helped to mitigate effects of temporal bias that could be more challenging for creation of external controls in indications where standard of care has been evolving. In addition, utilization of data from previously conducted clinical trials, with little missing data and consistent endpoint availability and definitions across studies, was an important advantage in this case study. Additional considerations may be necessary when constructing an external control from real-world data (e.g., data from electronic medical records, registries, etc.) (Chen et al. 2021; Levenson et al. 2021).

In the case study, we used greedy nearest neighbor matching with a caliper of 0.25, which appears to be a commonly used algorithm in the medical literature. This algorithm is straightforward and easy to understand, with the appealing feature of closely mimicking a randomized clinical trial with the matched pairs and a prespecified matching ratio. Two main types of matching algorithms using propensity scores are described in literature, the greedy matching algorithm and the optimal matching algorithm. Austin (2014) examined and compared 12 algorithms for matching on the scores, including optimal matching, greedy nearest neighbor matching without replacement, and greedy nearest neighbor matching without replacement within specified caliper widths. One of the recommendations was to use the nearest neighbor caliper without replacement in most situations.

The selection of baseline characteristics to be included in the propensity score estimation was based on the understanding of the disease population as well as data collection from all involved previously conducted clinical trials. The variables such as baseline demographics and characteristics that were usually summarized in the publications of the previously conducted clinical trials in this indication, subgroup factors, and stratification factors for randomization were considered as the candidates of important prognostic factors for balancing between the target group and external control. While we have successfully created an ECA with observed baseline characteristics that are well balanced with the control arm of the target trial, there is a need to address unobserved baseline characteristics as imbalance in these factors can confound treatment effect estimation. In addition, some target control patients were unmatched and thus excluded from the outcome analysis. While this is a common practice when utilizing matching methods and will improve baseline balance, removing patients from the analysis is undesirable especially when patients treated with the investigational drug are removed, which could restrict the matched patients to a group with more limited baseline characteristics. The option of extrapolating the analysis of this precise set and applying it to a more varied population should be considered.

Future research is needed to assess: whether the treatment effect estimated from a randomized trial can be replicated with the use of an ECA in place of randomized control since in the current case study only the control arm data was uniformly made available through data sharing programs; what effect unobserved covariates might have on treatment group comparisons; propensity score weighting rather than matching in order to remove the need to exclude unmatched investigational patients for whom significant efforts had been made to collect these data; and how this approach could be explored with additional endpoints. These may have variability in measurement (such as progression-free survival), other indications and populations, and augmentation designs (or hybrid designs) where a prospective study could be carried out with a combination of prospective randomized control patients and

external control patients providing opportunity for assessment of the accuracy of the external control through comparison of prospectively randomized and external control patients.

Acknowledgments

The authors thank Jill Dawson, Ph.D., for writing support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors reported there is no funding associated with the work featured in this article.

ORCID

Mark D. Stewart  <http://orcid.org/0000-0002-4847-0736>
Elizabeth A. Stuart  <http://orcid.org/0000-0002-9042-8611>
Ruthanna C. Davi  <http://orcid.org/0000-0003-2650-5384>

References

- Ali, M. S., D. Prieto-Alhambra, L. C. Lopes, D. Ramos, N. Bispo, M. Y. Ichihara, J. M. Pescarini, E. Williamson, R. L. Fiaccone, M. L. Barreto, et al. 2019. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Frontiers in Pharmacology* 10: 973. 10.3389/fphar.2019.00973. 10.3389/fphar.2019.00973. 2019 SEPTEMBER 18. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6760465>
- Allan, V., S. V. Ramagopalan, J. Mardekian, A. Jenkins, X. Li, X. Pan, and X. Luo. 2020. Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. *Journal of Comparative Effectiveness Research* 9 (9):603–614. Epub 2020 Mar 18. PMID: 32186922. <https://pubmed.ncbi.nlm.nih.gov/32186922>
- Austin, P. C., and E. A. Stuart. 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* 34 (28):3661–3679. doi:10.1002/sim.6607.
- Austin, P. C., and M. M. Mamdani. 2006. A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine* 25 (12):2084–2106.PubMed] [Google Scholar]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483/#R18>.
- Austin, P. C. 2011a. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46 (3):399–424. doi:10.1080/00273171.2011.568786.
- Austin, P. C. 2014. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine* 33 (6):1057–1069. doi:10.1002/sim.6004.
- Chen, J., M. Ho, K. Lee, Y. Song, Y. Fang, B. A. Goldstein, W. He, T. Irony, Q. Jiang, M. van der Laan, et al. 2021. The current landscape in biostatistics of real-world data and evidence: clinical study design and analysis. *Statistics in Biopharmaceutical Research* 1–14. doi:10.1080/19466315.2021.1883474.
- FDA. 0000. FDA clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry. <https://www.fda.gov/media/71195/download>
- Friends of Cancer Research White Paper. 0000. Characterizing the use of external controls for augmenting randomized control arms and confirming benefit. https://www.focr.org/sites/default/files/Panel-1_External_Control_Arms2019AM.pdf
- Ghadessi, M., R. Tang, J. Zhou, R. Liu, C. Wang, K. Toyoizumi, C. Mei, L. Zhang, C. Q. Deng, R. A. Beckman, et al. 2020. A roadmap to using historical controls in clinical trials – By Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). *Orphanet Journal of Rare Diseases* 15 (1):69. doi:10.1186/s13023-020-1332-x.
- Guidance for Industry: E10 Choice of Control Group and Related Issues in Clinical Trials. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). May 2001. ICH. <https://www.fda.gov/media/71349/download>

- Guidance for Industry: E9 Statistical Principles for Clinical Trials. U.S. Department of Health and Human Services Food and Drug Administration, *Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)*, September 1998. ICH. <https://www.fda.gov/media/71336/download>
- Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*. 15 (3):199–236. doi:10.1093/pan/mpl013.
- Imbens, G. W. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*. 86 (1):4–29. doi:10.1162/003465304323023651.
- Levenson, M., W. He, J. Chen, Y. Fang, F. Faries, B. A. Goldstein, M. Ho, K. Lee, P. Mishra-Kalyani, F. Rockhold, et al. 2021. Biostatistical considerations when using RWD and RWE in clinical studies for regulatory purposes: a landscape assessment, statistics in biopharmaceutical research. doi:10.1080/19466315.2021.1883473.
- Li, F., L. E. Thomas, and F. Li. 2019. Addressing extreme propensity scores via the overlap weights. *American Journal of Epidemiology* 188(1):250–257. PMID: 30189042. doi:10.1093/aje/kwy201.
- Lim, J., R. Walley, J. Yuan, J. Liu, A. Dabral, N. Best, A. Grieve, L. Hampson, J. Wolfram, P. Woodward, et al. 2018. Minimizing patient burden through the use of historical subject-level data in innovative confirmatory clinical trials: Review of methods and opportunities. *Therapeutic Innovation & Regulatory Science* 52 (5):546–559. doi:10.1177/2168479018778282.
- Liu, W., S. Kuramoto, and E. A. Stuart. 2013. An introduction to sensitivity analysis for unobserved confounding in non-experimental prevention research. *Prevention Science* 14 (6):570–580. doi:10.1007/s11121-012-0339-5.
- Medidata website. 0000 MEDS (Medidata Enterprise Data Store). <https://www.medidata.com/en/meds>
- Morgan, S. L., and J. L. Todd. 2008. A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology* 38 (1):231–281. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9531.2008.00204.x>.
- Normand, S., M. Landrum, E. Guadagnoli, J. Ayanian, T. Ryan, P. Cleary, and B. McNeil. 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 (4):387–398. doi:10.1016/S0895-4356(00)00321-8.
- Project Data Sphere website. 0000. <https://data.projectdatasphere.org/projectdatasphere/html/home>
- Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 70 (1):41–55. doi:10.1093/biomet/70.1.41.
- Rosenbaum, P. R. 1987. Model-based direct adjustment. *Journal of the American Statistical Association* . 82 (398):387–394. doi:10.1080/01621459.1987.10478441.
- Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* 25 (1):1–21. doi:10.1214/09-STS313.
- Stuart, E., and D. Rubin. 2007. Best practices in quasi-experimental designs: Matching methods for causal inference. Chapter 11 in *Best practices in quantitative social science*. In ed. J. Osborne, Vol. 11, 155–176. Thousand Oaks, CA: Sage Publications.



Innovative Drug Development: Evaluating lessons learned to inform continued progress

Abstract 1023: Evaluating the impact of the RACE act: An interim analysis

Brittany A. McKelvey¹, Mark Stewart¹, Jeff Allen,¹

¹Friends of Cancer Research, Washington, DC

Abstract

Recent approvals of novel therapeutics, including targeted small-molecules and immunotherapies, have significantly impacted cancer care. However, these advancements have not easily translated to new treatment options and approvals for pediatric cancer patients. The Research to Accelerate Cures and Equity (RACE) Act was signed into law in 2017 to accelerate the availability of drugs for pediatric cancer patients by requiring all new adult oncology therapeutics also conduct pediatric studies if the molecular target is relevant to pediatric cancer, including therapeutics with an orphan drug designation. RACE requirements were implemented on August 18th, 2020, and we evaluated the impact of the RACE Act since implementation. We evaluated all new drug applications or biologics license applications submitted and approved by the FDA from August 18, 2019-August 18, 2021. Qualifying drugs were stratified by applications approved one year before RACE implementation and applications submitted and approved one year post-RACE implementation. FDA approval letters and review documents were used to obtain information regarding pediatric study requirements and orphan drug designation. Nineteen therapeutics were identified within the study period (63.2% approved pre-RACE implementation and 36.8% approved post-RACE implementation). Only 11.8% of agents approved were indicated for pediatric use at the time of initial application, and the majority (78.9%) of approved applications received an orphan drug designation. Prior to the RACE Act, pediatric studies were waived for orphan drugs, regardless of possible applicability of the agent to pediatric cancers. However, 91.7% of the therapies approved pre-RACE implementation had a relevant mechanism of action (MoA) that may have required pediatric study if the application was submitted after RACE implementation. After RACE implementation, 71.4% of the therapies received an orphan drug designation, with 60% requiring pediatric studies due to the relevancy of the MoA. Pediatric studies were waived or exempt for all therapies during the study period prior to RACE implementation. However, following implementation of RACE, 42.9% of approved drugs require pediatric studies. The remaining therapeutics approved post-implementation had waived pediatric study requirements due to studies being impossible or highly impracticable given the pediatric prevalence in the indication. The one-year anniversary of RACE implementation provides the opportunity to begin to evaluate the effectiveness of the approach and recognize opportunities to expand its reach in the future. We found an increase in required pediatric studies after implementation, with the largest effect seen in orphan-designated therapeutics that are no longer automatically exempt. However, many pediatric study requirements are still waived for therapeutics with relevant MoA, highlighting the opportunity for future policy modifications.

Brittany A. McKelvey, Mark Stewart, Jeff Allen. Evaluating the impact of the RACE act: An interim analysis [abstract]. In: Proceedings of the American Association for Cancer Research Annual Meeting 2022; 2022 Apr 8-13. Philadelphia (PA): AACR; Cancer Res 2022;82(12_Suppl):Abstract nr 1023.

Breakthrough Therapy Designation Criteria Identify Drugs that Improve Clinical Outcomes for Patients: A Case for More Streamlined Coverage of Promising Therapies

Grace Collins, Mark Stewart, Brittany McKelvey, Hillary Stires, and Jeff Allen

ABSTRACT

The breakthrough therapy designation (BTD) process was created to expedite clinical development timelines for drugs intended to treat serious conditions and preliminary clinical evidence indicates the drug may demonstrate substantial improvement over existing therapies. This analysis demonstrates that BTD is a valuable tool for expediting approval of promising therapies in oncology. By comparing drugs indicated to treat non-small cell lung cancer (NSCLC) approved with BTD or without BTD

between January 2013 and October 2021, BTD drugs reduced the risk of death by a median of 31% and progression by a median of 48%, while drugs never receiving BTD reduced the risk of death and progression by a median of 15% and 41.9%, respectively. These findings show that BTD criteria accurately identify drugs that improve long-term outcomes for patients with cancer and warrant coordinated efforts to ensure timely coverage decisions and access for patients.

Since its inception, breakthrough therapy designation (BTD) has helped expedite clinical development timelines for drugs intended to treat a serious condition with preliminary clinical evidence indicating the drug may demonstrate substantial improvement over available therapy on a clinically significant endpoint(s). Several analyses have shown BTD facilitates earlier approval of therapies compared with therapies without BTD (1, 2). To date, the use of BTD has helped sponsors and the FDA streamline development and approval of 225 drugs, over 56% of which were oncology indications (3).

Despite faster FDA approval of these therapies, processes associated with coverage and reimbursement by insurance programs, including the Centers for Medicare and Medicaid Services (CMS), do not follow the same expedited timelines. This is particularly true for entirely novel treatments, such as first-in-class products, that involve new mechanisms of action or new technologies altogether, many of which are approved through an expedited program such as BTD. When payment processes are not finalized immediately following FDA approval, barriers to timely patient access can occur (4). While coverage of oncology drugs has not historically been an impediment to access, determinations of add-on payments or code sets can potentially delay patient access if not done in a

timely fashion. This issue was most notable with the recent introduction of chimeric antigen receptor (CAR) T-cell therapies for certain blood cancers (5).

In disease areas where recent innovations in treatment have contributed to lowered population mortality, such as in non-small cell lung cancer (NSCLC), delays between FDA approval and initiation of processes for coverage of new treatments could impede public health benefit (6). Recent discussions on CMS coverage processes for expedited approvals provide an opportunity to consider ways to align CMS and FDA procedures to ensure drugs qualifying for expedited programs, such as BTD, are covered at the time of approval. In oncology, clinical guidelines included in the National Comprehensive Cancer Network's (NCCN) Drugs and Biologics Compendium are used by insurers to inform coverage decisions. This has helped streamline reimbursement following the approval of new cancer drugs, but does not extend to other therapeutic areas, medical devices, and diagnostics, nor address timely coding processes, budgeting, or other procedures associated with payment.

As such, it is necessary to identify appropriate triggers that can help select novel products early in development to support more streamlined discussions regarding coverage. To assess whether BTD criteria identify high-priority drugs that improve outcomes for patients with cancer, and thereby evidence to support the importance of timely coverage, we compared outcomes data supporting approvals of and clinical guidelines for drugs with and without BTD indicated to treat NSCLC. NSCLC was chosen as a case study due to the high number of BTDs given to lung cancer indications and availability of long-term follow-up data. The results demonstrate that BTD drugs indicated for NSCLC improve outcomes and have more recommendations based on higher-quality data suggesting the treatments were more appropriate compared to drugs that never received a BTD. These findings support the notion that the qualifying criteria for BTD support the identification of drugs that improve outcomes for patients with NSCLC.

These findings demonstrate that BTD drugs provide improved clinical utility suggesting it would be beneficial to establish a

Friends of Cancer Research, Washington, D.C.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Mark Stewart, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC 20036. E-mail: mstewart@focr.org

Clin Cancer Res 2022;XX:XX-XX

doi: 10.1158/1078-0432.CCR-22-0983

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2022 The Authors; Published by the American Association for Cancer Research

Translational Relevance

The breakthrough therapy designation (BTD) process was created to expedite clinical development timelines for drugs intended to treat serious conditions and preliminary clinical evidence indicates the drug may demonstrate substantial improvement over existing therapies. This analysis demonstrates that BTD is a valuable tool for expediting approval of promising therapies in oncology. By comparing drugs indicated to treat non-small cell lung cancer (NSCLC) approved with BTD or without BTD between January 2013 and October 2021, BTD drugs reduced the risk of death by a median of 31% and progression by a median of 48%, while drugs never receiving BTD reduced the risk of death and progression by a median of 15% and 41.9%, respectively. These findings show that BTD criteria accurately identify drugs that improve long-term outcomes for patients with cancer and warrant coordinated efforts to ensure timely coverage decisions and access for patients.

mechanism through which a BTD would initiate processes to expedite CMS coverage. An expedited program at CMS would not include automatic coverage, but rather enable processes to ensure BTD therapies and novel technologies sufficiently meet approval and reimbursement requirements more quickly, and as appropriate, have processes in place for ensuring a product is covered at the time of approval to not delay patient access. An initial pilot of this program would provide valuable information around whether these processes are feasible, the value they bring, and identify areas needing improvement prior to wider implementation. The Parallel Review Program, which helps coordinate FDA and CMS reviews for medical devices and was initially run as a pilot for 5 years before permanent implementation, serves as a precedent for a similar process, outlined below (7).

- (i) FDA notifies CMS when a novel product (e.g., first indication) receives a BTD or is participating in another expedited regulatory pathway.
 - a. This initial notification to CMS would provide awareness of approval timelines for forthcoming products that are beneficial for patients with serious/life threatening illnesses.

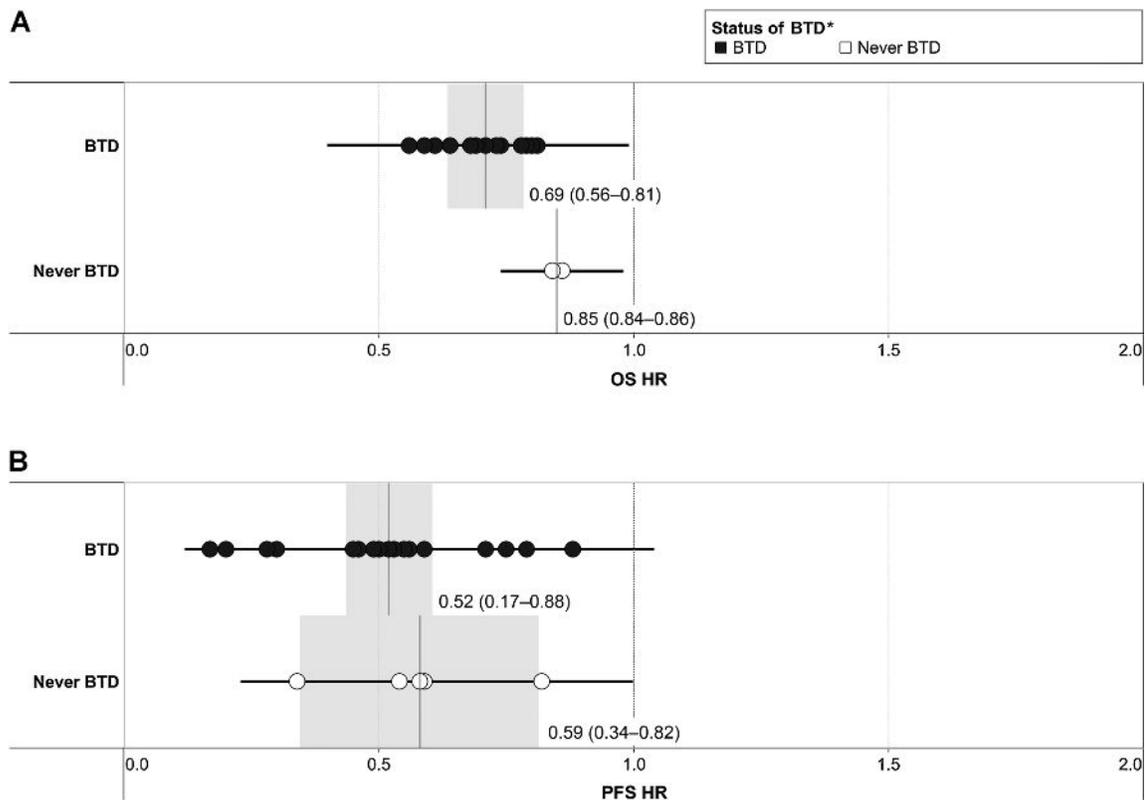


Figure 1. Outcomes supporting approvals of drugs for NSCLC. Median HR (range) for approvals supported by an RCT with the primary or coprimary endpoint of OS (A) and/or PFS (B). *"BTD", approvals for a drug or a combination of drug(s) including drugs that have ever received BTD for any indication; "Never BTD", approvals for drugs that have never received BTD for any indication.

Early notification about these products would allow additional time for CMS to coordinate resources necessary to support timely coverage decisions.

- (ii) Sponsors of products that receive BTM would have the opportunity to participate in an Expedited Coverage pilot.
 - a. The sponsor for a novel product or class of products could apply for the Expedited Coverage pilot prior to FDA approval. CMS would then evaluate whether the product or class (i) has important implications for Medicare beneficiaries and (ii) does not have a clear path to reasonable coverage (e.g., there are gaps in evidence or unique approaches to coverage may be necessary). Should CMS determine the product or product class meet the above requirements, an expedited process would begin.
 - b. This process would enable earlier discussions regarding topics such as coverage decisions, coding, eligibility for New Technology Add-On Payment (NTAP), and/or CMS budgeting implications. Sponsors would have earlier opportunities to coordinate and communicate with CMS regarding premarket data necessary to support initial

coverage at the time of FDA approval and to receive guidance from CMS on the longer-term path to coverage, including additional data that may be needed to support a national coverage decision. This would ensure clinical trials are designed to provide appropriate data supporting FDA approval and to inform coverage decisions.

Expediting development is a resource intensive process for both the FDA and sponsors, and more drugs are approved using BTM and/or other expedited pathways each year (2). For the processes proposed above to be successful, CMS will need additional resources to support their involvement. A coordinated, well-supported, and timely process for determining coverage of BTM products is necessary to ensure the value brought by BTM facilitates earlier patient access to effective treatments.

Approach

We identified 52 drug and biologics applications approved between January 1, 2013 and October 1, 2021, for an NSCLC indication and collected key clinical trial and outcomes data from

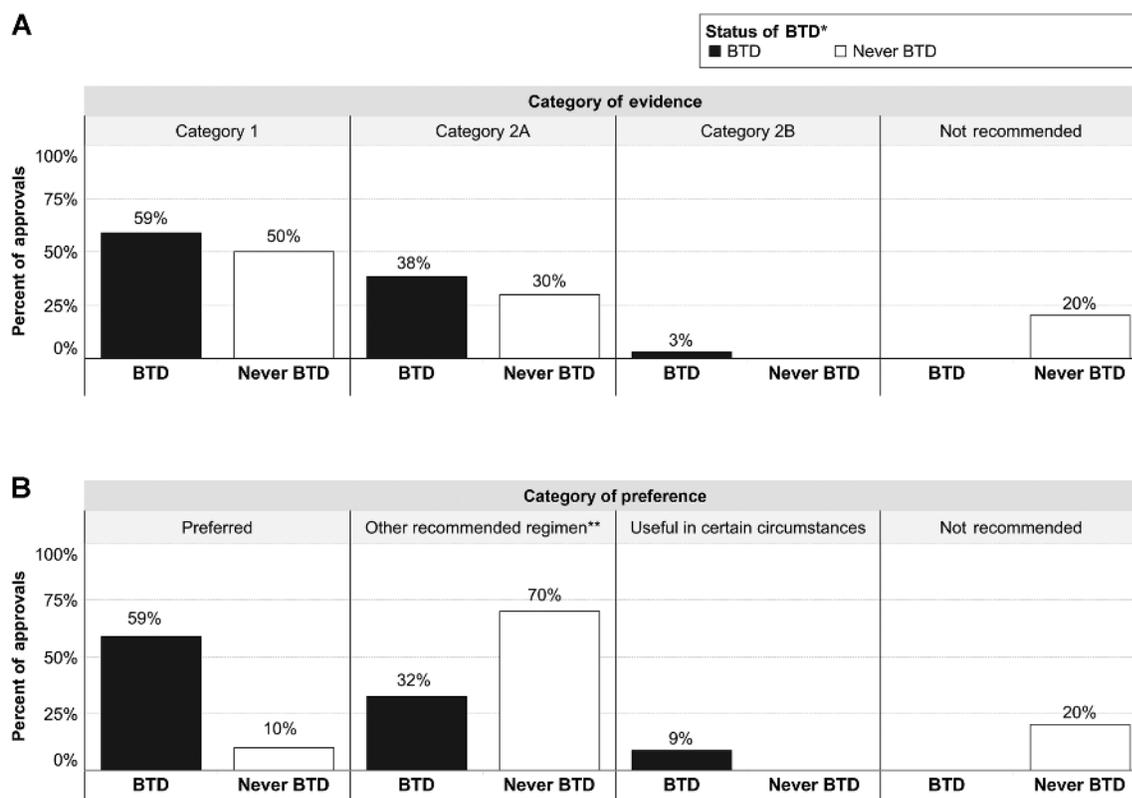


Figure 2. Characteristics of NCCN recommendations for NSCLC approvals from 2013 to 2021. **A.** Percentage of BTM approvals and percentage of Never BTM approvals by category of evidence. **B.** Percentage of BTM approvals and percent of Never BTM approvals by category of preference. **"BTM", approvals for a drug or a combination of drug(s) including drugs that have ever received BTM for any indication; "Never BTM", approvals for drugs that have never received BTM for any indication. **Other recommended regimens are uses that are more toxic, less affordable, less efficacious, and/or are based on less mature data (8).

publicly available review documents and labels published online in the Drugs@FDA database (Supplementary Table S1). We also collected recommended uses for these approvals from the NCCN Guidelines for NSCLC (version 7.2021) and noted the assigned category of preference and category of evidence.

The sample included 41 applications for drugs that had ever received BTM (BTM) and 11 applications for drugs that had never received a BTM for any indication (Never BTM). Thirty-four percent of BTM applications were also reviewed under the Accelerated Approval pathway.

Thirty-one approvals (59.6%) were supported by data from a randomized clinical trial(s; RCT) with the primary endpoint (pEP) or coprimary endpoint(s; cpEP) of progression-free survival (PFS) and/or overall survival (OS). Twenty-one approvals (40.4%) were excluded from the outcomes analysis because their labels were supported by data from nonrandomized trials and eight were excluded from the NCCN analysis to avoid double counting the same indication (Supplementary Table S1). Twenty-three approvals supported by an RCT (74.2%) included a BTM drug. The remaining eight approvals supported by an RCT were for Never BTM drugs.

Of the 16 approvals supported by trials with OS as a pEP or cpEP (14 BTM, 2 Never BTM), patients receiving BTM drugs had a 31% lower risk of death than those assigned to standard of care (SOC; median HR = 0.69; range: 0.56–0.81) whereas patients receiving a Never BTM drug had a 15% lower risk of death (median HR = 0.85; range: 0.84–0.86; **Fig. 1A**). Similarly, among approvals supported by a trial(s) with the pEP or cpEP of PFS (16 BTM, 6 Never BTM), patients receiving a BTM drug had a 48% reduced risk of progression than those assigned to receive SOC (median HR = 0.52; range: 0.17–0.88) compared with only 41.9% reduced risk of progression for patients receiving a Never BTM drug (median HR = 0.59; range: 0.34–0.82; **Fig. 1B**).

References

1. Shea M, Ostermann L, Hohman R, Roberts S, Kozak M, Dull R, et al. Regulatory Watch: Impact of breakthrough therapy designation on cancer drug development. *Nat Rev Drug Discov* 2016;15:152.
2. Collins G, Stewart M, Sigal E, Allen J. Expedited development programs at the Food and Drug Administration: insights and opportunities. *Ther Innov Regul Sci*. 2021;55:619–21.
3. US Food and Drug Administration. Breakthrough therapy approvals. 2022 Apr 13. Available from: <https://www.fda.gov/drugs/nda-and-bla-approvals/breakthrough-therapy-approvals>.
4. Roginiel AC, Dhruva SS, Ross JS. Evidence supporting FDA approval and CMS national coverage determinations for novel medical products, 2005 through 2016: A cross-sectional study. *Medicine* 2018;97:e12715.
5. Caffrey M. Delays with payers mean some patients miss window for CAR T-cell therapy, findings show. *Evidence-Based Oncology* 2021;27.
6. Howlader N, Forjaz G, Mooradian MJ, Meza R, Kong CY, Cronin KA, et al. The effect of advances in lung-cancer treatment on population mortality. *N Engl J Med* 2020;383:640–9.
7. Kux L, Slavitt A. Program for parallel review of medical devices. 81 FR 73113. *Federal Register* 2016. Available from: <https://www.federalregister.gov/d/2016-25659>.
8. National Comprehensive Cancer Network (NCCN). Development and update of guidelines. 2021. Available from: <https://www.nccn.org/guidelines/guidelines-process/development-and-update-of-guidelines>.

The NCCN Compendium serves as a source of information to support clinical decision making about the appropriate use of drugs in patients with cancer. A Category 1 recommendation is based on high-quality data and a high level of consensus among experts that the use is appropriate. Among NSCLC approvals between 2013 and 2021, including those supported by data from non-RCTs, 59% of BTM drugs, and 50% of Never BTM drugs met the bar for a Category 1 recommendation. For preference, 59% of BTM drugs are preferred interventions based on superior evidence of safety, efficacy, and, in some cases, affordability, while only 10% of Never BTM drugs were given a preferred recommendation (**Fig. 2B**; ref. 8)

This analysis is limited as the drugs included were all indicated only for treatment of NSCLC. A comprehensive analysis comparing outcomes data supporting approvals of BTM and non-BTM drugs for all oncology indications is warranted to further evaluate whether the qualifying criteria for BTM are generalizable across cancer types. In addition, the sample is necessarily limited to BTM and Never BTM drugs that reached approved status and therefore have publicly available labels.

Authors' Disclosures

No disclosures were reported.

Acknowledgments

This work was funded by Friends of Cancer Research. Friends of Cancer Research receives unrestricted grants and support from a diverse group of funders that includes industry.

Received March 28, 2022; revised May 19, 2022; accepted May 19, 2022; published first xx xx, xxxx.

Impact of the COVID-19 Pandemic Mitigation Strategies on Cancer Clinical Trials: Preliminary Findings of a *Friends-ASCO* Study

FRIENDS OF CANCER RESEARCH ANNUAL MEETING 2022

Background

Enrollment in clinical trials is key to advancing new treatments for patients with cancer. At the beginning of the COVID-19 pandemic, patient enrollment and treatment in cancer clinical trials were negatively impacted, in large part due to approaches to adapting to the COVID-19 pandemic public health emergency, including social distancing and lockdowns. Recognizing the challenges of recruiting and treating patients in clinical trials during the pandemic, researchers, regulators, and policymakers moved rapidly to support modifications to traditional clinical trial processes to enable important research and care to continue, both for ongoing trials and those initiated during the pandemic.¹

Anecdotally, many researchers have proposed that retaining these modifications in future trials could reduce inefficiencies and burdens, thereby increasing patient access to clinical trials. However, there is a knowledge gap in the published literature about how sponsors and sites adjusted clinical trial practices during the COVID-19 pandemic and what impact these changes had on the quality of trial data and patient access. To address this, Friends of Cancer Research (*Friends*) and the American Society of Clinical Oncology (ASCO) partnered to evaluate how the modifications to trial conduct adopted during the pandemic affected the conduct of clinical trials. If the impact of these changes, especially on data quality, has been sufficiently minimal, then maintaining these beneficial flexibilities could lead to increased patient access to future clinical trials and could speed the conduct of trials, thus accelerating new treatment discovery. Further, there may be an opportunity to streamline clinical trial operations by employing common reporting and documentation requirements for certain modifications, including protocol deviations (PDs) and amendments, as recommended by ASCO in its 2021 report, American Society of Clinical Oncology (ASCO) Road to Recovery Report: Learning from the COVID-19 Experience to Improve Clinical Research and Cancer Care.²

Thank You to Our Contributors

Steering Committee

Ajjai Alva, MD, Project Co-Chair, University of Michigan*

Laura Levit, JD, ASCO

Brittany McKelvey, PhD, Friends of Cancer Research

Caroline Schenkel, MSc, ASCO

Mark Stewart, PhD, Friends of Cancer Research

Hillary Stires, PhD, Friends of Cancer Research

Joseph Unger, PhD, Project Chair, Fred Hutchinson Cancer Center

*Co-chair through September 2022

Task Force

Suanna Bruinooge, MPH, ASCO

Beverly Canin, Breast Cancer Options

Nicole Connelly, PhD, IQVIA

Emily Dressler, PhD, Wake Forest School of Medicine

Peter Fredette, EQRx, formerly IQVIA

Keith Flaherty, MD, Massachusetts General Hospital Cancer Center

Elizabeth Garrett-Mayer, PhD, FSCT, ASCO

Nicole Holland, BS, ASCO

Lee Jones, MBA, Fight Colorectal Cancer

Jeff Legos, PhD, MBA, Novartis

Jennifer Lei, PhD, ASCO

Peggy McCann, DVM, PhD, Merck & Co., Inc.

Therica Miller, MBA, CCRP, Tisch Cancer Institute, Icahn School of Medicine
at Mount Sinai

Adedayo Onitilo, MD, PhD, MSCR, FACP, Marshfield Clinic

Rocio Paul, MSHS, National Cancer Institute

Fran Palmieri, MSN, Sarah Cannon Research Institute

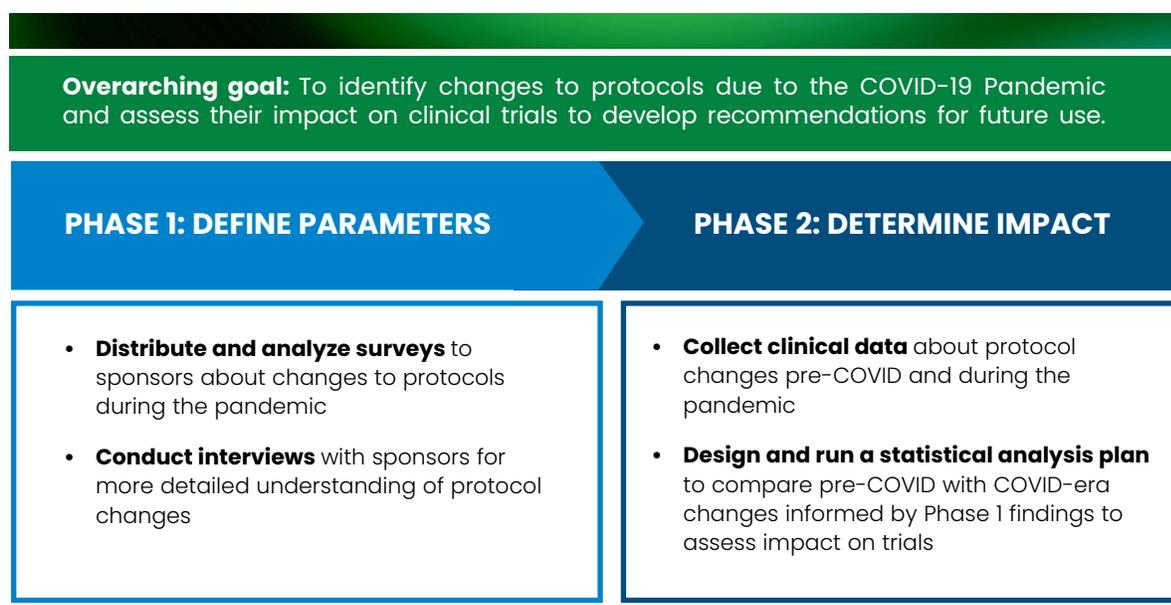
Timil Patel, MD, U.S. Food and Drug Administration

Shimere Sherwood, PhD, ASCO

Approach

ASCO and *Friends* partnered with two academic co-chairs to establish a steering committee who worked closely with a multi-stakeholder task force comprised of representatives from academic and community oncology practices (including clinical investigators and research staff), patient advocate groups, the U.S. Food and Drug Administration (FDA), the National Cancer Institute (NCI), pharmaceutical companies, a contract research organization (CRO), and ASCO and *Friends* staff. At the outset, the primary objectives of the Task Force were, 1) to assess potential changes to data quality, as reflected by changes in patterns of PDs during the COVID-19 pandemic; 2) to describe mitigation strategies that were employed to reduce PDs; and 3) to determine the broader impact of the mitigation strategies on the conduct of clinical trials. If the mitigation strategies adopted during the pandemic result in sufficiently minimal adverse consequences to data quality and trial conduct, we will formulate recommendations to retain the changes going forward.

To accomplish the research objectives, the Task Force is implementing a multi-phase approach (**Figure 1**). In Phase 1, we focused on assessing how clinical trial sponsors defined and documented PDs prior to and during the COVID-19 pandemic. We collected sponsor representatives' perceptions of the impact of the pandemic on PDs, as well as information related to trial activations and closures, mitigation strategies, and rates of adverse events. The information derived from Phase 1 will inform the design of Phase 2, in which we will conduct a meta-analysis that explicitly examines the direct impacts of the mitigation strategies on PDs, other key metrics of data quality, and patient access to clinical trials. Phase 2 will also address other pertinent research questions raised in the Phase 1 evaluation. This discussion document presents the preliminary results from Phase 1 of the project and outlines our plans for Phase 2.



Phase 1: Define Parameters

Phase 1 Aims

The aims of Phase 1 were to describe:

- Pre-COVID-19 pandemic PDs and the associated documentation requirements.
- Changes that occurred to PD descriptions, documentation requirements for PDs, and volume of PDs during the COVID-19 pandemic.
- Trial Sponsors' perception of the impact of shifting PD descriptions, documentation requirements, and volume of PDs on trial data integrity and missingness.
- Whether trial sponsors have retained or intend to retain any COVID-19 pandemic-era changes to their PD design or documentation processes moving forward.

Phase 1 Approach

ASCO and *Friends* first surveyed, and then interviewed, both industry and NCI cooperative group sponsors of anti-cancer interventional trials to understand changes to their clinical trial protocols during the COVID-19 pandemic. (A full list of survey and interview questions can be found in Appendices A-C.) Participating sponsor organizations were identified based on previous interaction with ASCO and *Friends* research activities, but all industry and NCI cooperative group sponsors who oversaw anti-cancer treatment trials (Phase 1, 2, or 3) evaluating any modality that were open in the United States between January 2015 and May 2022 were eligible to participate. Participation in the project was voluntary and at the discretion of the sponsor.

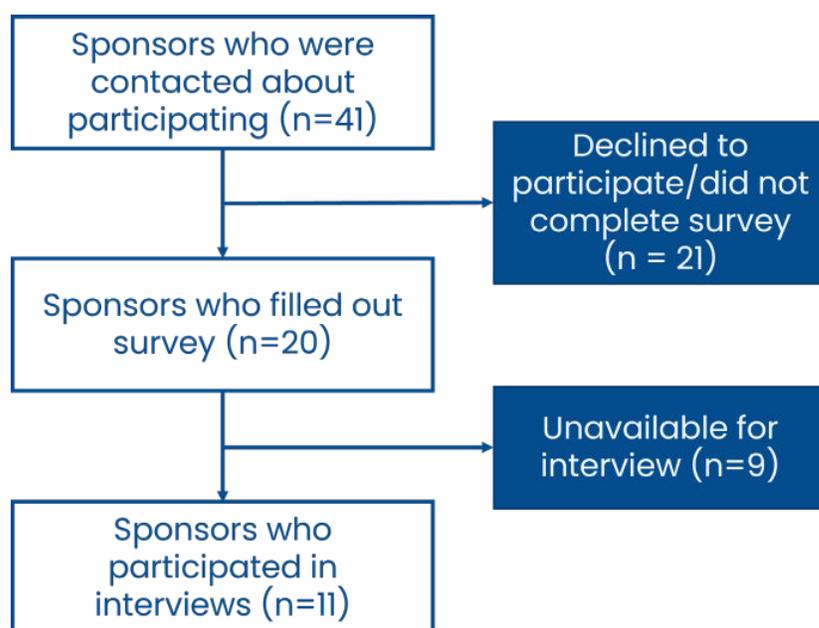
The study design was submitted for IRB review and was classified as exempt research. The survey and interview tools were created by ASCO and *Friends* staff and reviewed by the Task Force. After reviewing the study material, sponsor organizations (either industry or NCI cooperative groups) selected their own participants (henceforth, "Sponsors") to be surveyed and interviewed for the study.

The Task Force reviewed de-identified survey findings from each sponsor organization to identify areas for further exploration during semi-structured interviews. The interviews were conducted virtually over Zoom by an ASCO or *Friends* staff person with high-level oncology trial operations and data management personnel from a sample of the sponsor organizations. Sponsors received the discussion guide before the Zoom call and initial questions were the same for all participants; however, follow-up questions varied based on the discussion. During these interviews, Sponsors elaborated on their survey results and discussed their perceptions of the impact of PDs, other trial modifications, and mitigation strategies implemented during the COVID-19 pandemic.

Data collection was limited to May to July 2022 for surveys and July to October 2022 for interviews; thus, the participant sample does not include all sponsor organizations that met the eligibility criteria (**Figure 2**). Twenty sponsors (49% of those contacted) completed the survey for analysis and a subset of 11 sponsors (55% of those who completed surveys) were interviewed.

Interview findings suggested interpretation of study findings would benefit from speaking with leadership in the NCI's Cancer Therapy Evaluation Program (CTEP) about their processes for preparing and disseminating guidance during the pandemic. As such, representatives from NCI's CTEP were also interviewed using a modified version of the sponsor interview guide. Only aggregated, de-identified findings were shared with the Task Force.

Figure 2: Consort diagram of the approach for including sponsors organizations who participated in the analysis.



Findings from Phase 1

All findings reported below are based on information provided by Sponsors in surveys and interviews.

Sponsors' Perceived Impacts of the COVID-19 Pandemic

As has been previously reported, trials were most impacted early in the pandemic.^{3,4} Sponsors reported an increase in PD volume in the first wave of the COVID-19 pandemic (March–April 2020) (**Figure 3**). After the initial wave (starting in May 2020), the increase in PD volume compared to the pre-pandemic period was slightly lower (**Figure 4**). In the survey, 85% (17/20) of Sponsors reported that there was no change in how many trials closed due to low accrual since the start of the COVID-19 pandemic. While some Sponsors closed trial sites early in the pandemic for any reason related to pandemic mitigation strategies, others reported when interviewed that their sites remained open. Some interview participants specified that, during the early phases of the pandemic, very sick patients (e.g., children or patients with late-stage disease) were mostly likely to continue attending in-person appointments.

Figure 3: Change in PD volume pre-COVID-19 Pandemic to first wave. Sponsors were asked about the change in PD volume before the COVID-19 Pandemic to the first wave of the COVID-19 Pandemic (March-April 2020). Results are reported by Industry vs. Cooperative Group Sponsors. (One sponsor did not respond.)

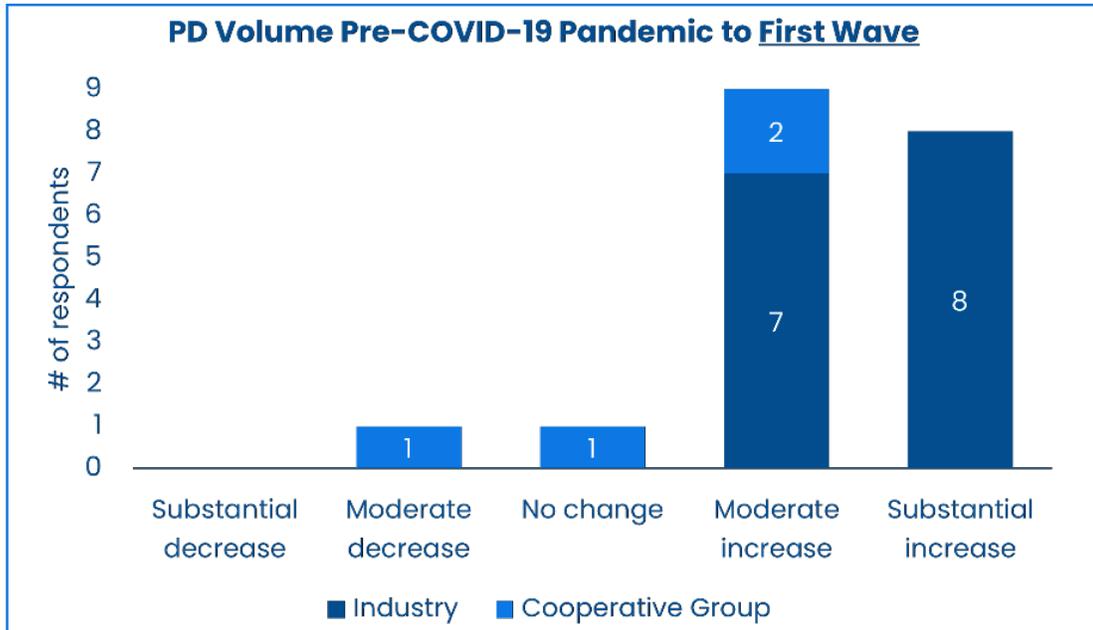
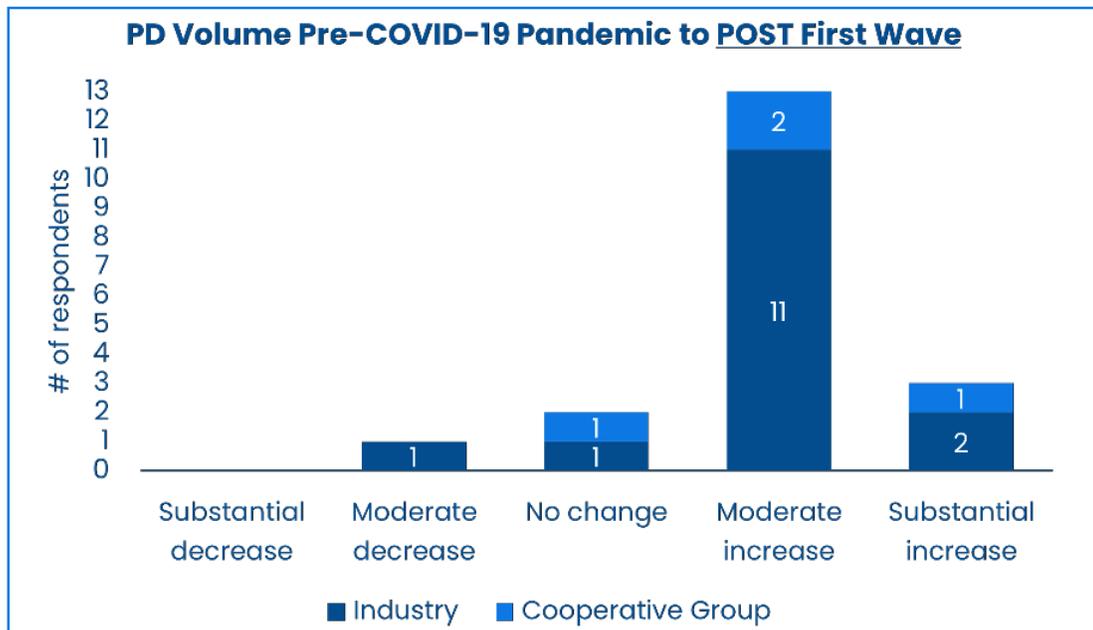
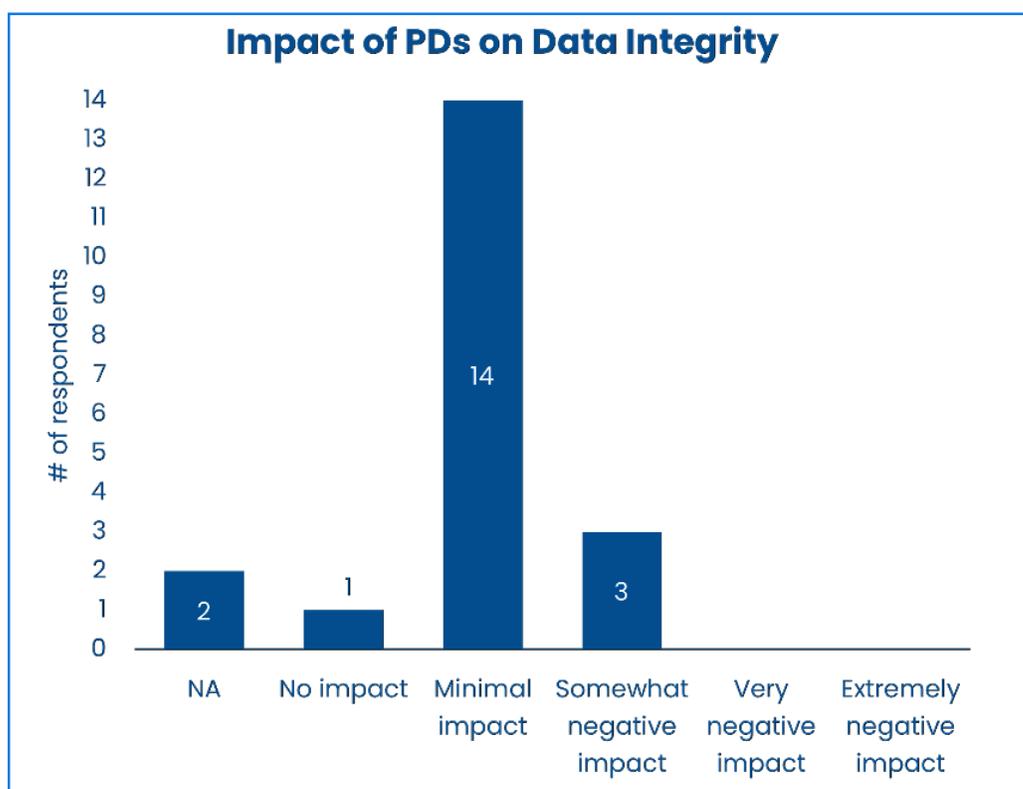


Figure 4: Change in PD volume pre-COVID-19 Pandemic to post-first wave. Sponsors were asked about the change in PD volume before the COVID-19 Pandemic to after the first wave of the COVID-19 Pandemic (May 2020 and beyond). Results are reported by Industry vs. Cooperative Group Sponsors. (One sponsor did not respond.)



In interviews, most Sponsors reported that the COVID-19 pandemic had only a minor impact on clinical trials after May 2020, which they attributed to U.S. sites pivoting quickly to allow most patients already enrolled on studies to continue with few disruptions due to flexibilities. Survey data showed that most Sponsors perceived a minimal impact of the PDs during the pandemic on data integrity (**Figure 5**). However, many Sponsors reported persistent lags in data entry related to staff shortages or turnover at trial sites. They reported that time delays were more common than data quality issues.

Figure 5: PD impact on data integrity. In the survey, sponsors were asked to rate the impact level to overall data integrity of PDs during the pandemic and provided with 5 responses ranging from “No Impact” to “Extremely Negative Impact.”



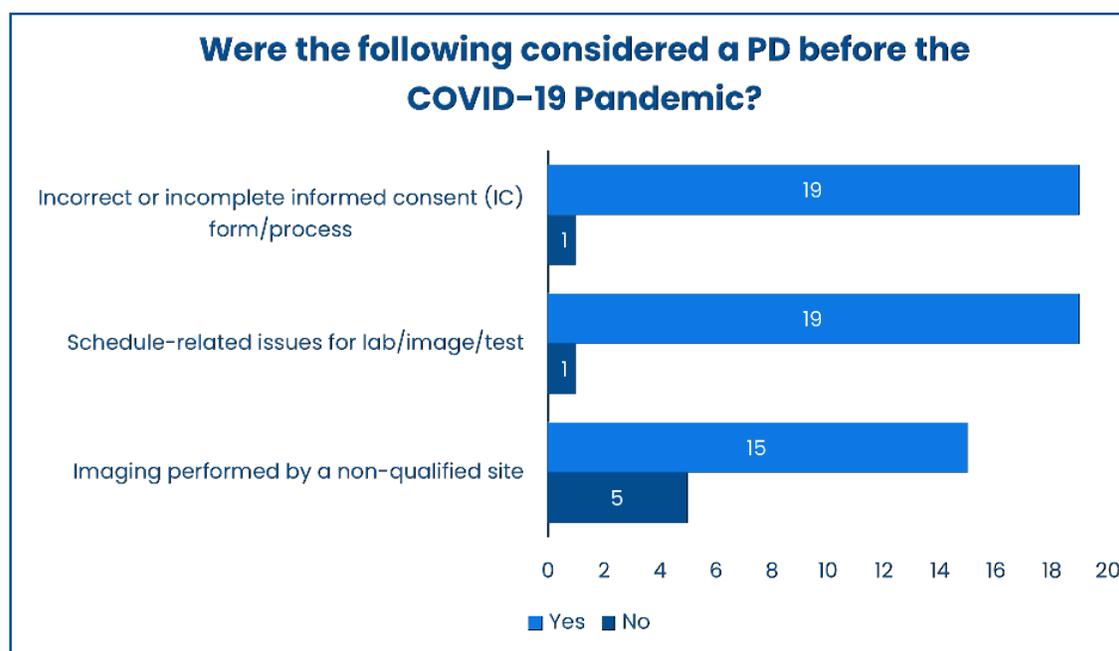
PDs During the COVID-19 Pandemic

Nearly all sponsors (95%) flagged COVID-19 pandemic-specific PDs. However, this data is often only shared with regulators when requested (i.e., during a submission). For Sponsors who analyzed the types of PDs, they reported minimal differences in the types of PDs by different trial characteristics (e.g., disease type or patient population). Sponsors observed more PDs in later phase trials, which typically have more patients and longer follow-up periods. Missed or out-of-window visits and assessments were most common early in the pandemic when patients were not traveling either due to COVID-19 pandemic restrictions or concerns about becoming ill from COVID-19, but these are no longer a prevalent challenge.

Remote Patient Monitoring

The COVID-19 pandemic accelerated the trend to make clinical trials more flexible for patients and providers through the incorporation of remote patient monitoring. Examples of frequently implemented remote patient monitoring include remote distribution of oral medication, imaging or blood draws at local facilities, remote informed consent discussion, and telemedicine visits. Many of these were classified as PDs before the pandemic (**Figure 6**). In interviews, it became apparent that part of the adaptation to the COVID-19 pandemic was to incorporate remote patient monitoring activities into trial protocols, rather than to include them as PDs.

Figure 6: Pre-COVID-19 Pandemic PD de initiions. In the survey, sponsors were asked to rate the impact level to overall data integrity of PDs during the pandemic and provided with 5 responses ranging from “No Impact” to “Extremely Negative Impact.”



Many Sponsors reported that they are considering opportunities to retain remote patient monitoring in trials moving forward, although some noted that not all flexibilities will continue. Those Sponsors who plan to retain remote patient monitoring indicated that, in the right context, it can ease patient burden while still collecting necessary data. Some Sponsors perceived that investigators may be resistant to continuing remote monitoring due to decreased oversight of their trial participants. Many highlighted that when they plan to include flexibilities for remote monitoring in their trial protocols, it would be considered optional rather than a required approach. Guidance from FDA and NCI informed modifications to trials early in the pandemic and may help shape Sponsor decisions about maintaining changes moving forward.

Guidance from Regulators and NCI

In interviews, many Sponsors indicated that guidance provided by regulators and the NCI helped facilitate the ongoing conduct of cancer clinical trials during the pandemic. Cooperative group interviewees reported that they were well-positioned to adapt to the pandemic quickly alongside NCI due to pre-existing mechanisms of communication with CTEP leadership. Industry Sponsors indicated that guidance documents from FDA (and global regulatory bodies, as relevant) were their primary reference points for clinical trial conduct. According to Sponsors, the timeliness of guidance documents from FDA and NCI was essential to mitigating the pandemic's negative effects on trials and patients, particularly early in the pandemic. At that time, FDA was permitted to bypass the usual requirements for guidance oversight and issue guidance rapidly, and NCI produced guidance documents through internal coordination. Some industry Sponsors commented that ongoing challenges outside of the U.S. – whether pandemic-related or otherwise – continue to impact regulatory guidance for global studies, and by extension, trial design and operations.

Flexibilities in the Future

Sponsors continue to evaluate which flexibilities they will retain in their interventional treatment trial protocols beyond the COVID-19 pandemic. Our findings from Phase 1 demonstrate variability among Sponsors in their approach to incorporating flexibilities; some readily adopted the strategies, while others – uncertain about whether the allowances will be permanent – have been more hesitant. Some Sponsors expressed concern about potential limitations on trial data quality with remote patient monitoring (e.g., local labs, remote auditing), while others found that these concerns diminished after flexibilities were introduced. Our hope is that findings to date and the analysis for Phase 2 will help sponsors make decisions about the appropriateness and value of bringing these flexibilities into the future.

Phase 2: Determine Impact

The Phase 1 portion of the evaluation used semi-quantitative survey data and Sponsor interviews to provide initial insights into the impact of the COVID-19 pandemic on trial conduct, and to generate hypotheses for more detailed evaluation in Phase 2. Although the focus of Phase 1 was on PDs, other domains were also evaluated including the number of active trials, trial initiations, and trial closures over time; eligibility and consent related changes; assessment, lab, and imaging changes, mitigation strategies adopted; and patterns of adverse events. Moreover, the Sponsor interviews and discussions with NCI's CTEP indicated that a more extensive and inclusive evaluation framework would be informative, which includes a detailed understanding of trial access during the pandemic and whether the pandemic affected the enrollment of diverse populations to trials. Phase 2 may include different and/or additional Sponsors from Phase 1 if the eligibility requirements are met.

Phase 2 Aims

The aim of Phase 2 is to test the hypotheses derived from Phase 1. Thus, for Phase 2, participating Sponsors will be asked to provide aggregate estimates of the key data domains highlighted by the Phase 1 evaluation as necessary to support a more determinative inference about the impact of the mitigation strategies used during the pandemic. Recognizing that participating Sponsors may have resource constraints, our aim is to request a limited, homogenous set of

data from all Sponsors to facilitate aggregate analyses and to limit the demands on Sponsor resources.

With these considerations in mind, the specific aims of Phase 2 are to characterize, over time in relation to the COVID-19 pandemic:

- The number of PDs (average per patient)
- The number of each type of PD (overall)
- Total enrollments per month to the sponsor portfolios of trials
- Grade 1 or 2 and grades 3 or 4 adverse events (average per trial)
- The number of dropouts (average per patient)
- Time delays

As suggested by the information obtained from Phase 1, an additional aim will be to examine the above outcomes from the perspective of diverse enrollment. Thus, for instance, one concern might be that the changes to trial conduct wrought by the pandemic might differentially impact sociodemographically underrepresented groups, who, for instance, might have experienced more PDs than their counterparts. Thus, for each outcome above, we will further request that Sponsors provide both overall estimates and estimates by categories of sex, age (<65 vs. 65 or older), race (Black vs. Asian vs. White vs. other) and ethnicity (Hispanic vs. not Hispanic).

Further, patterns of outcomes may differ by the nature of the trials, which could also influence the overall assessment of the impact of the pandemic on trial conduct. Thus, we will examine whether the outcomes noted above differ by study level variables (cancer type, study phase, and stage (advanced vs. adjuvant disease)).

We also plan to request the number, type, and date of implementation of mitigation strategies adopted by each sponsor during the initial pandemic wave, to determine whether the volume of strategies that were adopted is also correlated with outcomes.

Finally, as noted in Phase 1, we will represent the findings overall among all Sponsors, and also disaggregated according to sponsor type (industry vs. NCI sponsored cooperative groups). To evaluate these data, a meta-analytic approach will be used. This statistical approach requires the collection of only aggregate (deidentified) single measures for each measurement domain from each sponsor and will allow us to derive the overall average tendency (i.e., point estimates) across the trial system and to simultaneously understand accompanying variability across a diverse set of sponsors. Differences in patterns by demographic and study level variables will be examined using moderator analyses. This strategy has the distinct advantage of requiring the collection of only deidentified single measures for each measurement domain from each Sponsor but is limited by the lack of patient-level data to address within-patient patterns.

Conclusions and Next Steps

The COVID-19 pandemic impacted the conduct of cancer clinical trials and has likely accelerated a trend towards greater flexibility in trial conduct that was already emerging. The strategies implemented during the COVID-19 pandemic to provide greater flexibility in the execution of trial regulatory procedures, patient evaluation and data ascertainment can minimize clinical trial complexity, leading to reduced burden on sites and patients and improved access. Sponsors continue to include flexibilities in new protocols as they deem appropriate and engage sites and investigators in the process, while following regulatory guidance. To date, the primary aim of the Task Force has been to derive preliminary insights about the influence of the COVID-19 pandemic on trial conduct. In Phase 1 of this evaluation, sponsors reported that in their judgement, the mitigation strategies adopted in the face of the pandemic did not greatly impact data integrity. However, there is a recognition that a more detailed, quantitative, and statistical evaluation of clinical trial data integrity may provide greater and more determinative insight. We anticipate that the insights derived, and the hypotheses generated, from the Phase 1 portion of our evaluation will appropriately inform the conduct of our Phase 2 evaluation, in order to ultimately help guide the cancer clinical trial community about next steps in advancing the science of clinical trial conduct.

References

1. Boughey JC, Snyder RA, Kantor O, et al. Impact of the COVID-19 Pandemic on Cancer Clinical Trials. *Ann Surg Oncol*. 2021;28(12):7311–7316. [doi:10.1245/S10434-021-10406-2](https://doi.org/10.1245/S10434-021-10406-2)
2. Pennell NA, Dillmon M, Levit LA, et al. American Society of Clinical Oncology Road to Recovery Report: Learning From the COVID-19 Experience to Improve Clinical Research and Cancer Care. *Journal of Clinical Oncology*. 2021;39(2):155–169. [doi:10.1200/JCO.20.02953](https://doi.org/10.1200/JCO.20.02953)
3. Boughey JC, Snyder RA, Kantor O, et al. Impact of the COVID-19 Pandemic on Cancer Clinical Trials. *Ann Surg Oncol*. 28. [doi:10.1245/s10434-021-10406-2](https://doi.org/10.1245/s10434-021-10406-2)
4. Upadhaya S, Yu JX, Hodge J, Campbell J. COVID-19 impact on oncology clinical trials: a 1-year analysis. *Nat Rev Drug Discov*. 2021;20(6):415. [doi:10.1038/D41573-021-00086-8](https://doi.org/10.1038/D41573-021-00086-8)

Appendix A – Survey Instrument

Definitions

- **Protocol Deviation (PD):** Any non-compliance with Institutional Review Board (IRB)-approved protocol, including prospectively approved deviations or waivers.
- **Significant or Serious Protocol Deviation:** A protocol deviation which increases potential risk to participants or affects the integrity of study data. An isolated deviation may not be significant by itself, but significance may increase with numerous deviations of the same nature.
- **Mitigation Strategy:** Depending on the severity or frequency of one or more protocol deviations, the site may be expected to define a mitigation strategy (sometimes referred to as a Corrective and Preventive Actions (CAPA)). This strategy is broken into two parts:
 - ◊ **Corrective Action (CA),** which is the action the site takes to address the deviation. Examples of corrective actions include (but are not limited to): notifying the affected participant(s) and protocol team; re-consenting the participant(s); completing missed procedures; repeating laboratory tests; completing additional participant monitoring or management procedures; and/or destroying specimens collected in error.
 - ◊ **Preventive Action (PA),** which is the action the site takes in attempt to prevent recurrence of the product or quality problem moving forward. Examples of preventive actions include (but are not limited to): discussion of the deviation with relevant study staff, refresher training of study staff; review and/or revision of documents outlining Standard Operating Procedures (SOPs) or other study implementation materials; development of new study implementation materials; implementation of additional communication, Quality Control (QC)/ Quality Assurance (QA), or oversight/supervisory procedures; changes in day-to-day workflow; and/or changes in general participant management or laboratory procedures.
- **Remote:** Geographically separated from the research site administering the clinical trial.

Pandemic-Related Time Periods

- Pre-COVID: January 2015 through December 2019.
- Immediately Pre-COVID: January and February of 2020.
- First Wave: March and April of 2020.
- Post-First Wave: May 2020 to May 2022.

Note: All questions refer to interventional anti-cancer trials (phase 1, 2 or 3) involving any modality (e.g., systemic therapy [cytotoxic, immune, hormonal, targeted, etc.], surgery, radiation, etc.) sponsored by your organization that are/were open in the United States.

Begin Survey:

Section 1 – Cancer Treatment Trial Portfolio

Cancer trials underway immediately pre-COVID-19 pandemic (January and February 2020)

1. How many interventional Phase 1 anti-cancer trials did your organization have ongoing in January 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50

2. How many interventional Phase 2 anti-cancer trials did your organization have ongoing in January 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50
3. How many interventional Phase 3 anti-cancer trials did your organization have ongoing in January 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50

Cancer trials opened during the COVID-19 pandemic (March 2020 to May 2022)

4. How many interventional Phase 1 anti-cancer trials has your organization opened since March 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50
5. How many interventional Phase 2 anti-cancer trials has your organization opened since March 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50
6. How many interventional Phase 3 anti-cancer trials has your organization opened since March 2020?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 11-20
 - 20-50
 - More than 50

Cancer trials underway during the first wave of the COVID-19 pandemic (March and April of 2020)

7. How would you characterize the impact of trial holds at sites during the first wave of the pandemic (March 2020–May 2020) on those trials?
 - None/few (0–25%) of our trials were delayed or otherwise impacted by holds
 - Some (26–50%) of our trials were delayed or otherwise impacted by holds
 - Most (51–75%) of our trials were delayed or otherwise impacted by holds
 - Nearly all/all (>76%) of our trials were delayed or otherwise impacted by holds
8. What was the approximate average hold time at sites during the March 2020–May 2020 period? ___ (weeks)
9. How would you characterize the impact of trial closures at sites during the first wave of the pandemic (March 2020–May 2020) on those trials?
 - None/few (0–25%) of our trials were negatively impacted by closures
 - Some (26–50%) of our trials were negatively impacted by closures
 - Most (51–75%) of our trials were negatively impacted by closures
 - Nearly all/all (>76%) of our trials were negatively impacted by closures
10. Do you have any additional comments regarding trial holds and closures during the first wave of the pandemic?

Cancer trials underway post–first wave of the COVID-19 pandemic (May 2020 to May 2022)

11. Compared to the March 2020–May 2020 period (your answer to question 7), how would you characterize the impact of trial holds at sites on your organization’s interventional anti-cancer trials after May 2020 and up to the current date?
 - The percentage of trials delayed or otherwise impacted by holds was much lower
 - The percentage of trials delayed or otherwise impacted by holds was somewhat lower
 - The percentage of trials delayed or otherwise impacted by holds was the same
 - The percentage of trials delayed or otherwise impacted by holds was somewhat higher
 - The percentage of trials delayed or otherwise impacted by holds was much higher
12. Compared to the March 2020–May 2020 period (your answer to question 9), how would you characterize the impact of trial closures at sites on your organization’s interventional anti-cancer trials after May 2020 and up to the current date?
 - The percentage of trials negatively impacted by closures was much lower
 - The percentage of trials negatively impacted by closures was somewhat lower
 - The percentage of trials negatively impacted by closures was the same
 - The percentage of trials negatively impacted by closures was somewhat higher
 - The percentage of trials negatively impacted by closures was much higher
13. Do you have any additional comments regarding trial holds and closures from May 2020 to May 2022?

Section 2 – Organizational Definitions

14. Please provide your organization’s definition of a “major PD” (sometimes called a “serious PD”): ___
15. What types of major, significant, or serious PDs have been the most common during the COVID-19 pandemic? ___

16. Please provide your organization's definition of a "minor PD": ___

17. What types of minor PDs have been most common during the COVID-19 pandemic? ___

Section 3 – Pre-COVID-19 PDs

18. During the pre-COVID-19 period (January 2015 through December 2019), did your organization typically categorize eligibility and consent issues as PDs? E.g., participant did not meet eligibility criteria, incorrect or incomplete informed consent form/process, or re-consent not obtained as required.

- Yes
- No

19. During the pre-COVID-19 period (January 2015–December 2019), which of the following eligibility or consent-related changes to a patient's protocol-specified treatment plan were typically defined as a PD? [select all that apply]

- Participant did not meet eligibility criteria
- Incorrect or incomplete informed consent (IC) form/process, including:
 - ◊ Consent form document not signed/dated by study participant or parent/legally authorized representative (if applicable); signed incorrect IRB-approved version of IC form; IC form does not contain all required signatures; IC form signed after registration/enrollment; signed IC form version that was not protocol specific; patient/study participant signed IC form containing changes not approved by the CIRB/IRB; non-English speaker signed untranslated version of IC form; or did not document IC process
- Re-consent not obtained as required

20. During the pre-COVID-19 period (January 2015–December 2019), which of the following treatment-related changes to a patient's protocol-specified treatment plan were typically defined as a PD? [select all that apply]

- Failure to follow trial randomization
- Failure to discontinue treatment
- Administration of non-protocol defined therapy to treat subject's disease or concomitant medication used was not permitted per protocol
- SAE reported out of window
- Dosing issues, including agent-related issues and:
 - ◊ Study agent administered to wrong patient/study participant; Study-supplied agent substituted with non-study-supplied agent, including commercial agent; Study agent stored incorrectly; Study agent prepared incorrectly; Study agent prescribed by unauthorized prescriber
- Device-related issues, including:
 - ◊ Study device administered to incorrect subject; Study device malfunction; or Study device not returned

21. During the pre-COVID-19 period (January 2015–December 2019), which of the following assessment, lab, or imaging-related changes to a patient's protocol-specified treatment plan were typically defined as a PD? [select all that apply]

- Schedule-related issues, including:
 - ◊ Baseline assessments are out of window; Delayed image submission; Timing of Lab/Image/Test/Procedure not per protocol
- Physical assessment deviation
- Patient does not have a safety follow-up as required
- Lab/Imaging/Test/Procedure after withdrawal of consent

- Lab, imaging, or other test/procedure not done
- Imaging performed by a non-qualified site
- Other imaging-related issues, including:
 - ◊ Incorrect imaging agent administered; incorrect imaging agent dose administered; Incorrect injection to scan time; Incorrect imaging modality; Incomplete anatomical coverage; Imaging parameters not per protocol; Images lost/unavailable/corrupt; Images not submitted; Equipment not credentialed prior to imaging

Section 4 – Volume of PDs during COVID-19 Pandemic

22. How did the average volume of PDs collected during the first wave of the pandemic (March 2020 and April 2020) compare to the pre-pandemic (January 2015–December 2019) volume?
- Substantial increase after March 2020
 - Moderate increase after March 2020
 - No measurable change after March 2020
 - Moderate decrease after March 2020
 - Substantial decrease after March 2020
23. How did the average volume of PDs collected post-first wave (starting May 1, 2020) compare to the pre-pandemic (January 2015–December 2019) volume?
- Substantial increase post-first wave
 - Moderate increase post-first wave
 - No measurable change post-first wave
 - Moderate decrease post-first wave
 - Substantial decrease post-first wave
24. Compared to pre-pandemic (January 2015 through December 2019), in May 2022 how had the average number of significant/serious PDs changed relative to the average number of minor PDs?
- Increased
 - Remained stable
 - Decreased

Section 5 – PD Mitigation Strategies

25. Which of the following mitigation strategies had NOT been employed pre-COVID-19 pandemic (January 2015 through December 2019) and were introduced immediately prior to or during the pandemic (January 2020 to May 2022), at least in part to decrease the number of PDs? [select all that apply]
- Remote pre-screening for eligibility
 - Remote recruitment/trial education and counseling
 - E-consenting/remote informed consent
 - Remote routine lab testing
 - Remote study-specific lab testing
 - Remote study-required biopsies
 - Remote symptom monitoring for adverse events
 - Remote distribution of oral anticancer therapy
 - IV administration of investigational treatment outside of investigational site
 - Remote collection of patient-reported outcomes
 - Remote imaging (study-required baseline or follow-up)
 - Remote monitoring of long-term outcomes
 - Other (please describe)
 - None of the above
 - Please describe other mitigation strategies that were introduced: ___

26. Which of the following Corrective and Preventative Actions (CAPA) had NOT been employed pre-COVID-19 pandemic (January 2015 through December 2019) and were introduced immediately prior to or during the pandemic (January 2020 to May 2022), at least in part to decrease the number of PDs? [select all that apply]
- Notifying the affected participant(s) and protocol team of the deviation
 - Re-consenting the participants
 - Completing missed procedures
 - Repeating laboratory tests
 - Completing additional participant monitoring or management procedures
 - Destroying specimens collected in error
 - Discussion of deviations with relevant staff
 - Refresher training of study staff
 - Review and/or revision of Standard Operating Procedures (SOPs) or other study implementation materials
 - Implementation of additional communication
 - Quality Control (QC)/Quality Assurance (QA) or oversight/supervisory procedures
 - Changes in day-to-day workflow
 - Changes in general participant management or laboratory procedures
 - None of the above
 - Unknown

Section 6 – Impacts on Patients and Data Collection

27. Does your organization collect/flag PDs that are attributable specifically to the COVID-19 pandemic?
- Yes
 - No
28. Do you have data on PDs that were requested by sites but not approved (by the IRB, DSMB, your organization, or other) during the COVID-19 Pandemic?
- Yes
 - No
29. PDs can be attributable to the study staff (e.g., missing a lab to be ordered) or the participant (i.e., skipping a scheduled visit). Do you have data on the proportion of PDs that are attributable to staff versus participant decision-making?
- Yes
 - No
30. Approximately what percentage of PDs are attributable to study staff (as opposed to participant decision-making)? ___
31. Have your organization's cancer treatment trial drop-out rates changed since the start of the COVID-19 Pandemic (March 2020)?
- Yes, drop-out rates increased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, drop-out rates increased during the pandemic but have returned to pre-pandemic levels (or decreased further).
 - Yes, drop-out rates decreased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, drop-out rates decreased during the pandemic but have returned to pre-pandemic levels (or increased further).
 - No change in drop-out rates was observed during the pandemic.

32. Has there been a change in how many of your organization's cancer treatment trials closed due to low accrual since the start of the COVID-19 Pandemic (March 2020)?
- Yes, there was an increase in how many trials closed due to low accrual during the pandemic and these closures have not returned to pre-pandemic levels.
 - Yes, there was an increase in how many trials closed due to low accrual during the pandemic but these closures have returned to pre-pandemic levels (or decreased further).
 - Yes, there was a decrease in how many trials closed due to low accrual during the pandemic and these closures have not returned to pre-pandemic levels.
 - Yes, there was a decrease in how many trials closed due to low accrual during the pandemic but these closures have returned to pre-pandemic levels (or increased further).
 - No change was observed during the pandemic.
33. Did rates of reported grade 1-2 adverse events (AEs) change during the pandemic?
- Yes, rates of reported grade 1-2 AEs increased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, rates of reported grade 1-2 AEs increased during the pandemic but have returned to pre-pandemic levels (or decreased further).
 - Yes, rates of reported grade 1-2 AEs decreased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, rates of reported grade 1-2 AEs decreased during the pandemic but have returned to pre-pandemic levels (or increased further).
 - No change in rates of reported grade 1-2 AEs was observed during the pandemic.
34. Did rates of reported grade 3-4 AEs change during the pandemic?
- Yes, rates of reported grade 3-4 AEs increased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, rates of reported grade 3-4 AEs increased during the pandemic but have returned to pre-pandemic levels (or decreased further).
 - Yes, rates of reported grade 3-4 AEs decreased during the pandemic and have not returned to pre-pandemic levels.
 - Yes, rates of reported grade 3-4 AEs decreased during the pandemic but have returned to pre-pandemic levels (or increased further).
 - No change in rates of reported grade 3-4 AEs was observed during the pandemic.
35. How would you rate the impact level to overall data integrity of PDs during the pandemic?
- Extremely negative impact
 - Very negative impact
 - Somewhat negative impact
 - Minimal impact
 - No impact

Expedited Development of Diagnostics for Therapies Targeting Rare Biomarkers or Indications

Introduction

Drug and diagnostic co-development has traditionally occurred in a manner by which one drug is accompanied by one diagnostic test to sufficiently characterize the safety and efficacy of the drug, while contemporaneously demonstrating the analytical and clinical validity of the diagnostic test assessing the biomarker status and of the responding patients in a clinical trial. For rare biomarkers or indications, this approach may not sufficiently leverage opportunities to expedite development for therapies and balance the need for efficient development of a companion diagnostic (CDx). The field of oncology has progressed substantially with an improved understanding of the biology of cancer, which has coincided with the availability of next generation sequencing (NGS) technologies that can query many biomarkers in one test. In cancers where NGS can be employed to assess biomarker status, these advances make the traditional one drug-one test approach to development of targeted therapies less ideal and poorly aligned with clinical and laboratory practice and patient needs.

New drug development follows the typical investigational new drug (IND) processes for clinical development, and Study Risk Determination (SRD) is typically conducted to determine whether FDA investigational device exemption (IDE) approval is required for the use of an unapproved diagnostic test in the clinical study. Although local testing (e.g., tests performed at a lab affiliated with the patient's treatment facility using a laboratory

Authors

IMEIN BOUSNINA
Genentech, A member of
the Roche Group

MEGAN DOYLE
Amgen

MAGGIE HUSTON
Tempus

RASIKA KALAMEGHAM
Genentech, A member of
the Roche Group

ELAINE KATRIVANOS
Tempus

LAURA KOONTZ
Foundation Medicine

YASSINE LABIAD
Loxo Oncology at Lilly

JOCHEN LENNERZ
Massachusetts General Hospital

ELIZABETH MANSFIELD
Foundation Medicine

CARLY MCWILLIAMS
Roche Diagnostics

JENNIFER ROMANS
Foundation Medicine

LAUREN SILVIS
Tempus

ANTHONY N. SIRECI
Loxo Oncology at Lilly

developed test (LDT) or commercial test kit/platform if one exists) may be used to identify patients for studies of drug activity and biomarker assessment, one central lab test prototype is generally used for enrolling patients into the pivotal study.

Challenges with the traditional development and regulatory review of drugs and CDx range from concerns about homogenous clinical trial populations, delayed patient access to clinical trials, and pre-/post-market requirements. Drug sponsors seek to balance enrollment speed with trial integrity, i.e., ensuring that the trial is enrolling a well-defined patient population that reflects the intent-to-treat (ITT) population. There may be delayed access to clinical trials because patients may be first screened using local lab testing and then are only enrolled in the trial following confirmation that patients meet trial eligibility criteria with central lab testing. This process presents challenges to drug and diagnostic co-development and can result in undue patient burden and potential medical harm since it may entail re-biopsy. It can also lead to delayed enrollment and accrual, increased wait time for patients to be in study, and delayed development (resulting in delayed post-market access) of the diagnostic and therapeutic. Where biomarker positive samples are very rare and regulatory requirements are not adjusted to account for this rarity, pre-/post-marketing requirements for diagnostic developers may disincentivize or slow the development of an approvable CDx for rare diseases or rare variants.

The type and extent of information required by FDA to support approval of diagnostic tests may need to vary based on the benefit/risk balance for the individual device and its intended use. FDA has generally not applied differing requirements for levels of evidence or certainty when a CDx addresses rare biomarkers. This is likely because of a lack of guidance on what flexibility can be applied, a lack of well-developed alternate methodologies, and a lack of designated pathways where flexibilities can be applied. However, the Humanitarian Device Exemption seeks to address issues that exist in developing CDx for rare diseases including limited availability of positive samples, limited information about potential alterations that could be treatable, and requirements to screen large numbers of samples to find a reasonable number of useful samples. To address these issues, flexibility in development expectations would benefit both patients and product developers to overcome some of the challenges to bringing therapies and CDx to market for rare diseases/biomarkers.

A more balanced approach to patient selection and diagnostics development in oncology clinical trials is needed, particularly for patients with rare diseases. Our goal is to propose a framework that would facilitate enrollment of patients in an efficient manner while maintaining clinical trial integrity and approval of a CDx based on requirements that consider the benefit-risk profile and feasibility of obtaining samples. Furthermore, to ensure timely availability of a diagnostic at or near the time of drug approval, we propose refining validation requirements for CDx approval. This document explores recommendations to 1) improve patient access to clinical trials for rare disease/biomarker therapies via expanded use of local tests, and 2) de-risk and streamline the development of a CDx for rare cancers to align with drug development.

Improve Patient Access

Ensure that policy does not inadvertently create roadblocks or reduce patient access

As development of targeted therapies directed at rare biomarkers and rare variants of more prevalent biomarkers becomes more common, reliable testing and screening capabilities to recruit patients for studies will be increasingly important. Using a single diagnostic assay intended to support assessment of clinical trial eligibility can slow patient accrual when patients are initially screened by a local, non-FDA approved testing platform. On the other hand, enrollment based on multiple tests with potentially variable performance and varying design (e.g., DNA vs. RNA) may not optimally select patients for enrollment and may complicate later efforts to obtain CDx approval. Establishing minimum performance standards could help address and alleviate these challenges.

Recommendations

Detection of rare biomarkers and rare variants to support the development of targeted therapies poses unique challenges, particularly as it pertains to the analytical comparability of the test(s) used to enroll patients into pivotal clinical trials. However, the benefits of identifying and accruing patients using multiple local tests, particularly when identifying rare variants, may outweigh the risk of variability in the clinical trial population. To support this paradigm, alignment of minimum performance standards and variants/variant classes can help standardize biomarker measurement which in turn could reduce barriers to patient enrollment and ensure homogeneity in the trial population. Per current FDA guidance,¹ enrollment using multiple local tests is allowed (including for pivotal trials), and FDA recommends that the sponsor evaluate comparability of test results among potential sites prior to initiating trial testing at those sites. Clinical trial sponsors (drug developers) could articulate, prior to patient enrollment, the minimum performance standards needed to accrue patients based on the particular study needs.² Local labs with individual tests could then provide evidence of minimum performance data if they intend to enroll patients into trials. In keeping with FDA guidance, these data would include information regarding accuracy, precision, analytical sensitivity, and analytical specificity, which the sponsor could share with FDA and enroll patients in the pivotal study using the local test results (as already occurs) but more efficiently and potentially with less variability. While exploratory, the NCI-MATCH Designated Laboratory Network³ approach could be used as a model to qualify labs and find alignment between central/local testing through prerequisite validation standards.

¹ Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product - Draft Guidance for Industry and Food and Drug Administration Staff (fda.gov)

² Recommendations for proposed minimum performance standards available in Friends of Cancer Research White Paper—Blueprint for Breakthrough Forum: Research and Reimbursement in the Age of Precision Medicine. https://www.focr.org/sites/default/files/pdf/Friends%20Alexandria%20Blueprint%20White%20Paper_October.pdf

³ James V. Tricoli, et al. Design and development of the molecular analysis for Therapy Choice (NCI-MATCH) Designated Laboratory Network. JCO 2019 37:15_suppl, 3016-3016

Diagnostic tests have varying underlying designs and methodologies, and laboratories use different analyses, which can lead to discordance across tests. To account for potential variance, patient samples that are positive for the rare biomarker are typically used to standardize test performance and support test validation. While accepted practice, it is nonetheless a poor use of precious biomarker positive clinical samples that is costly and time consuming. FDA could consider issuing guidance recommending the use of a combination of contrived samples, representative variant validation, variant class-based validation for certain variant types, and, where available, prior data that demonstrate analytical validation of the assay (e.g., previous FDA approval of an NGS-based test) to ascertain test performance while expediting test development and patient accrual (**Table 1**). In instances where clinical samples are particularly hard to obtain, whether for a local test or a CDx in development, FDA could consider allowing substitution with similar tumor types (e.g., perform analytical validation on non-small cell lung cancer (NCSLC) samples where small cell lung cancer samples are unavailable) or a “DNA is DNA” approach allowing use of any sample with the biomarker in question, regardless of its tissue of origin. Use of a representative approach for simple genomic alterations such as single nucleotide variants (SNVs) should be considered as appropriate surrogates. The extent that these alternative approaches could be used will depend on the complexity and prevalence of the biomarker being detected.

Table 1: Proposed Minimum Requirements to Support Use of Tests Detecting Rare Variants

Performance Characteristic	Minimum Requirement*
Concordance (Sensitivity, specificity, accuracy)	30 biomarker negative samples A range up to 30 biomarker positive [#] samples If possible 6 known positives (confirmed using an orthogonal method)
Limit of Detection	1 known positive* sample in a serial dilution series with at least 3 replicates at each dilution step
Precision	Repeatability across operators, reagent lots, days, instruments using 2 positive samples per variant type, with one at 1.5x LOD and one at 2x LOD
Limit of Blank	5-10 replicates across 2-3 healthy donor samples using the same sample type
*Requirements and number of samples should be guided by the complexity and prevalence of the biomarker being detected	
[#] Can be a contrived sample	

De-risk and Streamline the Development and Review of CDx to Align with Drug Development

Review drugs for rare indications and companion diagnostics in tandem via benefit-risk assessment

Regulatory processes associated with the co-development of a targeted therapy and CDx should also be aligned if concurrent approval of the drug and diagnostic is required. As with the development and regulatory pathways for targeted therapies, the regulatory pathways for the associated CDx should be reflective of the unmet need for rare indications, which may require additional flexibilities by FDA review divisions. The goal would be to create a mechanism to identify diagnostic tests for a rare tumor type that would lead to an intensive, interactive, and collaborative development and review process. Similar to what is done for drugs used in rare diseases, this approach could include the use and publication of a formal benefit-risk assessment for the diagnostic and the level of pre- and post-market evidence could be calibrated relative to considerations in the benefit-risk assessment. Drug and diagnostic review divisions should make a concerted effort to align review processes such that the drug and diagnostic are given contemporaneous approvals.

Expedite development and regulatory pathways for companion diagnostics for rare biomarkers

In order to achieve more rapid availability of an approved diagnostic, it may be appropriate to rethink the application of FDA's benefit/risk framework. It is important to balance timely patient access with analytical and clinical validation, bridging studies, and potential post-market study requirements for PMA approval that are required of the CDx test developer. A risk-based approach to identify which data elements are essential prior to approval (minimum core data set) and which data elements could be shifted to the post-market space as a requirement for maintained approval could support expedited development of a CDx for a rare biomarker or variant and allow a sponsor the opportunity to de-risk CDx investment prior to full proof of concept on a therapy. This could serve as a means to expedite the development of high-risk tests and facilitate contemporaneous regulatory review. Likewise, and perhaps more applicable for rare biomarker CDx, FDA could reconsider the extent of required evidence based on the benefit-risk assessment for the diagnostic, the rarity of the biomarker, availability of tissue samples, and the unmet need. Although post-market studies as a condition of approval are appealing, the ability to access rare samples after approval is generally not improved and may be worse than in the pre-market setting.

Recommendations

Sponsors are afforded flexibilities to facilitate drug development for rare indications. In a similar vein, the Center for Device and Radiological Health (CDRH) and CDx developer could engage in dialogue earlier in the development process to explore flexibilities that could be applied to the CDx development for a rare biomarker. Further, CDRH should commit to an expedited review timeline of 75 days for CDx for rare indications to ensure contemporaneous approval of the CDx and the drug, as drugs for rare indications are typically reviewed in a compressed timeline.

A core set of validation data should be submitted pre-approval for all diagnostic tests, including validation of analytical performance characteristics such as sensitivity, specificity, accuracy, precision, reproducibility, and limit of detection; but FDA should have the flexibility to consider the necessity of other data requirements in the context of a rare variant. In determining when to apply such flexibilities in development requirements, FDA should consider:

- Prevalence of disease/cancer type (e.g., whether orphan disease or low prevalence cancer type)
- Prevalence of mutation/biomarker/variant within that cancer
- Tissue type and availability
- Test type and prior analytical validation generated in similar cancer types or sample types

To qualify for the rare disease/biomarker flexibilities, FDA should use a threshold of 10,000 patients likely to have the disease or condition (not be tested for it). Examples of rare variants and tissues, where it would be appropriate for FDA to apply development flexibilities due to these considerations, are included in **Table 2**.

Table 2: Example Considerations for Benefit-Risk Assessment for Rare Diseases and Rare Variants

Rare Disease/Variant/Required Tissue for Validation	Characteristic Qualifying as Rare
NTRK	<ul style="list-style-type: none"> • Prevalence of variant (0.32% across solid cancers)
ROS-1	<ul style="list-style-type: none"> • Prevalence of variant (1.0% of lung non-small cell lung cancer)
Triple negative breast cancer patients after progression on primary therapy that metastasizes to the bone	<ul style="list-style-type: none"> • Tissue type and availability
Fine Needle Aspirate (FNA) in NSCLC	<ul style="list-style-type: none"> • Paired biopsy and FNAs from the same patient needed for validation

In the case of a rare variant or rare disease where development flexibilities may be appropriate, FDA should consider a variety of options for aligning the development expectations with the risk/benefit of the test and the unmet need for the drug. FDA reviewers should have license in these rare biomarker and rare disease scenarios to modify the requisite number of samples for an analytical study, the sample types, or waive requirements for certain analytical studies if these studies are recapitulating existing data or merely being done to “check the box” rather than generating new and meaningful information. Flexibilities that could be applied are included in **Table 3**.

Table 3: Regulatory Flexibilities that Could be Applied for CDx for Rare Variants/Biomarkers.

Analytical Validation
<ul style="list-style-type: none"> • For biomarkers that have already been analytically validated on NGS tests that have previously received FDA approval, FDA should leverage this validation in order to expedite review and approval for a rare indication. • To demonstrate analytical validity of rare variants/biomarkers, FDA should allow sponsors to provide some combination of the following instead of requiring use of clinical samples: <ul style="list-style-type: none"> ○ contrived samples ○ similar tumor types/sample types ○ representative variant validation for certain variant types ○ if available, prior data that demonstrate adequate analytical validity for their assay • FDA should not require revalidation of variants if they are in the exact same location or within the same base pair as a previously validated variant, and the primers are the same. • Repeat validation should not be required for every mutation/biomarker on a test platform when adding a new variant or to enroll a trial using a previously approved test, when the variant of interest was included in the first release of the approved test. Even if a small number of samples was used to validate that specific mutation for the first release, the test should be considered validated or, if anything, additional validation should be minimal with a small number of additional samples. • FDA should not require that a variant be validated across all different types of cancers.
Clinical Validation
<ul style="list-style-type: none"> • FDA should rely on the clinical performance (based on clinical outcome data) of an assay, rather than requiring concordance studies to LDTs used in enrollment using rare or limited clinical samples if different from the proposed CDx. Often these LDTs are tests of varying design, that have not gone through FDA pre-market review and are of unknown performance. • If FDA requires bridging studies between a candidate clinical trial assay (CTA) and the to-be-marketed CDx, the agency should consider whether such studies could be conducted as a post-market commitment.
Other Regulatory Flexibilities
<ul style="list-style-type: none"> • Allow use of a prespecified modification plan for already approved CDx seeking additional indications: <ul style="list-style-type: none"> ○ A prespecified modification plan would allow a test developer to submit a validation plan for future modifications that FDA could approve for post-market use in lieu of reviewing additional post-market analytical validation data to support a modification. ○ For example, where a new mutation of clinical significance is in the same class (i.e., SNV) and same locus as a previously validated variant. • Waive or, if necessary, shift into post-market certain studies (e.g., interfering substances, reproducibility, bridging studies). • Allow for post-market collection of real-world evidence.

The alignment of review programs for drugs and CDx could be further facilitated by creating a risk-based pathway for a CDx for rare biomarkers. FDA could publish a benefit-risk assessment at CDx approval for both PMAs and supplemental PMAs, akin to what is included in the summary basis of approval published for drugs, to enable greater regulatory flexibility for tests for rare biomarkers.

Conclusion

Current CDx guidance aims to enable co-development of a diagnostic and targeted therapy, which in turn allows for demonstration of analytical and clinical validity of the diagnostic test.⁴ However, this will become more challenging as narrower subpopulations are identified (both in oncology and in rare disease spaces). The traditional pathway for drug and diagnostic test co-development may not represent the most efficient method for development of targeted drugs and their CDx for rare tumors. Policies should address how to speed CDx development and review while limiting disruption to the current framework, including leveraging current flexibilities available for rare indications and unmet medical need. Modifications to the development process can maximize patient access by not restricting the screening requirements to a single test, provide for rapid access to clinical trials by alleviating the need for repeat biopsy and test analysis, expedite clinical drug development by identifying additional eligible patients, and ensure consistency between different tests with the same intended use. Ultimately, identification of patients who would benefit from therapies can be performed more efficiently, and greater patient access can be achieved.

⁴ Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product - Draft Guidance for Industry and Food and Drug Administration Staff (fda.gov)



Complex Biomarkers: Generating evidence to support alignment in drug development

Homologous Recombination Deficiency: Concepts, Definitions, and Assays

Mark D. Stewart^{1,✉}, Diana Merino Vega¹, Rebecca C. Arend², Jonathan F. Baden³, Olena Barbash⁴, Nike Beaubier⁵, Grace Collins¹, Tim French⁶, Negar Ghahramani⁷, Patsy Hinson⁸, Petar Jelinic⁹, Matthew J. Marton⁹, Kimberly McGregor¹⁰, Jerod Parsons⁵, Lakshman Ramamurthy¹¹, Mark Sausen³, Ethan S. Sokol¹⁰, Albrecht Stenzinger¹², Hillary Stires^{1,✉}, Kirsten M. Timms¹³, Diana Turco¹³, Iris Wang¹⁴, J. Andrew Williams¹⁵, Elaine Wong-Ho¹⁶, Jeff Allen¹

¹Friends of Cancer Research, Washington, DC, USA

²Division of Gynecologic Oncology, University of Alabama at Birmingham, Birmingham, AL, USA

³Translational Medicine, Bristol Myers Squibb, New York, NY, USA

⁴Oncology Experimental Medicine Unit, GlaxoSmithKline, Philadelphia, PA, USA

⁵Tempus Labs, Inc., Chicago, IL, USA

⁶Global Medical Affairs, Diagnostics, AstraZeneca, Cambridge, UK

⁷Molecular Genetic Pathology Regional Laboratory, SCPMG Regional Reference Laboratories, Los Angeles, CA, USA

⁸Independent Cancer Research Patient Advocate, Charlotte, NC, USA

⁹Early Clinical Oncology, Merck & Co., Inc., Kenilworth, NJ, USA

¹⁰Cancer Genomics Research Group, Foundation Medicine, Cambridge, MA, USA

¹¹Global Regulatory Affairs, GlaxoSmithKline, Washington, DC, USA

¹²Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany

¹³Myriad Genetics, Inc., Salt Lake City, UT, USA

¹⁴Global Precision Medicine, Novartis Pharmaceuticals Corporation, New York, NY, USA

¹⁵Precision Medicine & Biosamples, AstraZeneca, Cambridge, UK

¹⁶Clinical Sequencing Division, Thermo Fisher Scientific, San Francisco, CA, USA

*Corresponding author: Mark D. Stewart, 1800 M Street NW, Suite 1050 South, Washington, DC 20036, USA; Email: mstewart@focr.org

Abstract

Background: Homologous recombination deficiency (HRD) is a phenotype that is characterized by the inability of a cell to effectively repair DNA double-strand breaks using the homologous recombination repair (HRR) pathway. Loss-of-function genes involved in this pathway can sensitize tumors to poly(adenosine diphosphate [ADP]-ribose) polymerase (PARP) inhibitors and platinum-based chemotherapy, which target the destruction of cancer cells by working in concert with HRD through synthetic lethality. However, to identify patients with these tumors, it is vital to understand how to best measure homologous repair (HR) status and to characterize the level of alignment in these measurements across different diagnostic platforms. A key current challenge is that there is no standardized method to define, measure, and report HR status using diagnostics in the clinical setting.

Methods: Friends of Cancer Research convened a consortium of project partners from key healthcare sectors to address concerns about the lack of consistency in the way HRD is defined and methods for measuring HR status.

Results: This publication provides findings from the group's discussions that identified opportunities to align the definition of HRD and the parameters that contribute to the determination of HR status. The consortium proposed recommendations and best practices to benefit the broader cancer community.

Conclusion: Overall, this publication provides additional perspectives for scientist, physician, laboratory, and patient communities to contextualize the definition of HRD and various platforms that are used to measure HRD in tumors.

Key words: homologous recombination; poly(ADP-ribose) polymerase inhibitors; *BRCA1*; *BRCA2*; biomarkers; tumor; DNA repair.

Implications for Practice

Analyzing deficiencies in the homologous recombination repair (HRR) machinery becomes increasingly important to identify patients responding to poly(adenosine diphosphate [ADP]-ribose) polymerase (PARP) inhibitors. Ovarian, breast, pancreatic, and prostate cancer are at the forefront of this development, but other cancer types will likely follow. Clinically, homologous recombination deficiency (HRD) is broadly defined, ranging from deleterious mutations in single HRR genes (*BRCA1/2* and non-*BRCA*) to complex genomic scars. As it currently stands, assays that determine HR status may not agree on status calls, which can be problematic for the utility of these assays in the clinic. Our work provides an overview of the diagnostic landscape of HRD including a conceptual framework and definitions which will support molecular tumor boards and clinical decision making.

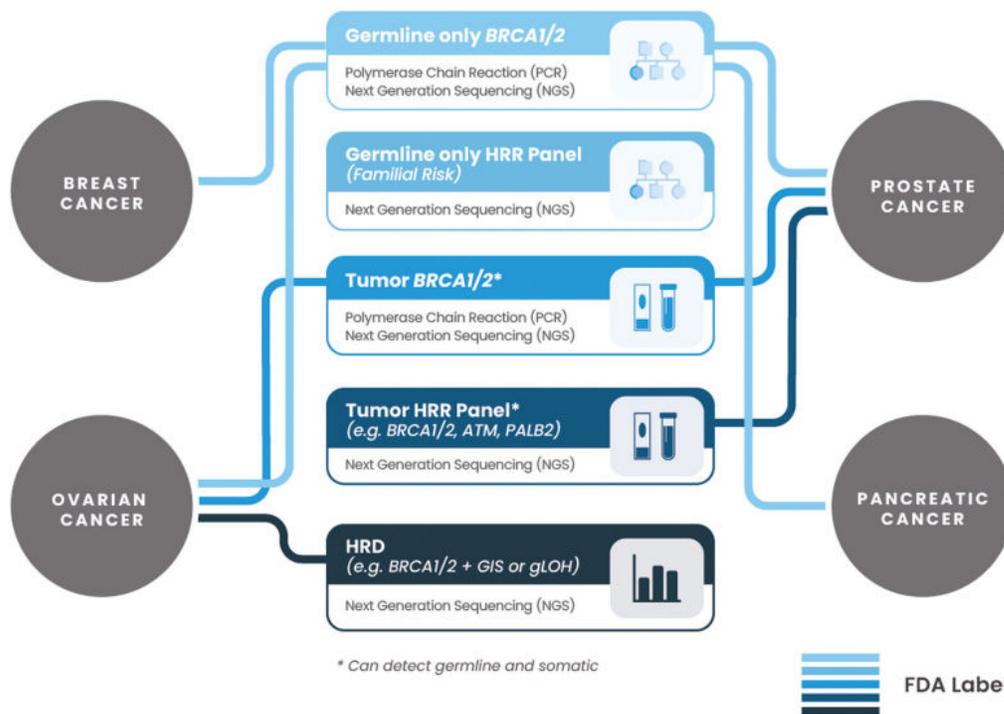
Introduction

Genomic instability is one of the most common underlying aspects of tumorigenesis, and defective DNA repair is described as a hallmark of cancer.¹ Homologous recombination repair (HRR) is a DNA repair pathway that acts on DNA double-strand breaks and interstrand cross-links (ICL).² A deficiency in the HRR pathway has been associated with several tumor types including breast, ovarian, prostate, and pancreatic cancers (Fig. 1) and has been termed homologous recombination deficiency (HRD), whereas tumors that are not HRD are termed homologous recombination proficient (HRP).^{3,4} The presence of HRD can make tumors more sensitive to ICL-inducing platinum-based therapies and poly(adenosine diphosphate [ADP]-ribose) polymerase (PARP) inhibitors (PARPi).^{5,6} Adenosine diphosphate-ribose polymerase inhibitors work via synthetic lethality; blocking base excision repair with PARPi results in an accumulation of DNA single-strand breaks and replication fork collapse resulting in DNA double-strand breaks that cannot be repaired by the HRR pathway if HRR is deficient.^{7,8} Homologous recombination deficiency is a predictive biomarker for treatment with PARPi in ovarian cancer based on patient outcomes in randomized controlled phase III trials.⁹⁻¹² Additionally, HRD is a positive prognostic marker for both progression-free survival and overall survival.¹³ Diagnostics developers have created tests to determine homologous recombination (HR) status and aid in treatment decisions; however, these assays may differ in what they measure and may lead to discordant results that can be problematic for prescribing oncologists. Patients are offered treatment at an emotionally difficult time and discordance between assays

makes the decision of diagnostic test and therapy selection more challenging due to uncertainty.

The HRD phenotype of sensitivity to platinum-based therapies and PARPi is associated with the HRD genotype defined by impairment in genes involved in the HRR pathway (“causes”) and/or genomic scarring/instability (“consequences”; Fig. 2).^{4,14} The *BRCA1* and *BRCA2* genes play prominent roles in the HRR pathway and impaired *BRCA* gene function is the most studied mechanism in tumor cells among the potential causes that results in HRD. Germline and somatic mutations, as well as epigenetic modifications in *BRCA1* and *BRCA2*, have been consistently associated with an HRD phenotype in breast, ovarian, pancreatic, and prostate cancer,¹⁵⁻¹⁷ and have been deemed archetypal in the determination of an HRD phenotype.¹⁸ Other HRR pathway genes associated with an HRD phenotype include genes such as *ATM*, *PALB2*, *RAD51*, and others.³ Either genetic or epigenetic alterations in these genes or some combination underlie the HRD phenotype in various cancer types¹⁹ including ovarian,^{16,20,21} endometrial,²² breast,²³⁻²⁵ prostate,²⁶ and pancreatic cancer.²⁷ The association between these genes and an HRD phenotype may be less consistent than *BRCA1* and *BRCA2* and may vary by the tumor’s tissue of origin. Due to the lack of understanding of the clinical implications of the mutations within HRR pathway genes, more studies to investigate the role of these genes in HRD phenotype in various cancer types are needed.²⁸⁻³⁰

Testing for the consequences of an impaired HRR pathway is performed by probing the genome for evidence of genomic abnormalities. Several studies in breast and ovarian cancer have identified genomic patterns or signatures of instability



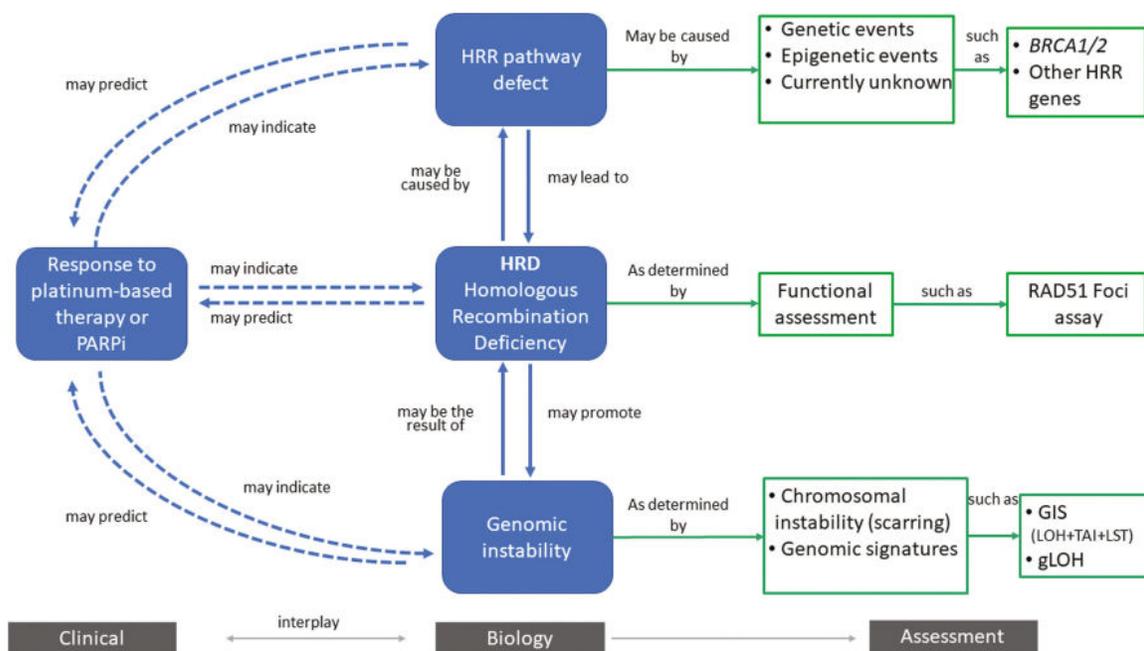


Figure 2. Overview of homologous recombination deficiency (HRD). Homologous recombination deficiency is a phenotype that is characterized by the inability of a cell to effectively repair double-strand DNA breaks using the homologous recombination repair (HRR) pathway. Alterations in these genes have been deemed “causes” of HRD (eg, genetic events and epigenetic events). This can result in an impaired HRR pathway, which can be deemed “consequences,” and assessed by probing the genome for evidence of genomic instability (eg, chromosomal instability and other genomic signatures).

associated with an HRD phenotype. These signatures of instability can include genomic patterns of loss of heterozygosity (gLOH), which are regions of intermediate size (>15 MB and < whole chromosome),³¹ number of telomeric imbalances (TAI), which are the number of regions with allelic imbalance which extend to the sub-telomere but not cross the centromere,³² and large-scale transitions (LST), which are chromosome breaks (translocations, inversions, or deletions).³³ These approaches evaluate the presence of HRD-related genomic signatures (often referred to as scars) that are thought to be a consequence of error-prone DNA repair through alternative pathways (eg, non homologous end joining [NHEJ]).

Studies have demonstrated the predictive value of assays to determine the HRD phenotype by evaluating response to platinum-based therapies and PARPi in the context of breast and ovarian cancer.^{4,14} Various multi-omic studies have investigated how a combination of the above patterns and additional genomic and transcriptomic signatures are associated with an HRD phenotype.^{5,34-36} A number of studies continue to evaluate genomic instability in breast and ovarian cancer as well as additional cancer types, which could lead to refinements in its use.²⁸ Additional approaches, such as the detection of RAD51 foci, may enable a functional assessment of HR status after a cell’s exposure to a DNA damaging agent. This approach requires multiple slides per patient which must be annotated by a trained professional. Recently, it has been shown that assessment of basal levels of RAD51 foci are possible in clinical samples and appear to show a high degree of correlation with PARPi response. The clinical validity as well as the practicality and implementation for routine clinical use is under investigation.³⁷⁻³⁹

To date, FDA has approved several companion or complementary diagnostics to facilitate tumor selection for PARPi treatments based on HR status (Table 1). Two of these (FoundationOne CDx and the Myriad myChoice CDx test) assess chromosomal instability to select patients with ovarian cancer that may benefit from an FDA-approved therapy. These assays incorporate both the causes and consequences of HRR impairment, whereas other FDA-approved assays only detect potential causes of HRR impairment without assessing consequences (eg, BRACAnalysis CDx, FoundationOne Liquid CDx, and FoundationOne CDx). It is important to note that while there are broad differences in the approach to test for HRD (causes vs consequences), there are also differences in the assays themselves that need to be considered. In addition to the FDA approved companion diagnostics, other sequencing or single-nucleotide polymorphism (SNP)-based platforms are being evaluated to measure genomic instability. Additional commercial and lab-developed assays that utilize tissue and blood to measure HR status are also available.

There is currently only partial agreement on which parameters contribute to determining the HR status of a sample and what combination of molecular measures are necessary to classify tumors as HRD.⁴⁰ Additionally, assays may use different cutoffs across tumor types, within tumor types, or for drugs within a similar class.⁴¹ For example, a clinical trial on the PARPi veliparib used a cutoff of 33 using the Myriad myChoice CDx test to define HR status in patients with high grade serous ovarian cancer¹⁰ while other trials on niraparib⁹ and olaparib¹² used a cutoff of 42. As the application of HR status is investigated in different cancer types, it is important to improve clarity regarding the way HRD is defined, measured, and reported. Addressing this lack of clarity can help

Table 1. Companion diagnostics approved for selection of PARPi.

Cancer	Assay	Sample type	Therapy	Indication	Trial
Ovarian	BRACAnalysis CDx	Blood	Olaparib	For the maintenance treatment of adult patients with deleterious or suspected deleterious germline or somatic BRCA-mutated advanced epithelial ovarian, fallopian tube or primary peritoneal cancer who are in complete or partial response to first-line platinum-based chemotherapy	SOLO1 Study
	BRACAnalysis CDx	Blood	Olaparib	For the treatment of adult patients with deleterious or suspected deleterious germline BRCA-mutated (gBRCAm) advanced ovarian cancer who have been treated with 3 or more prior lines of chemotherapy	Study 19 Study 42
	FoundationOne CDx	Tumor	Olaparib	For the maintenance treatment of adult patients with deleterious or suspected deleterious germline or somatic BRCA-mutated advanced epithelial ovarian, fallopian tube or primary peritoneal cancer who are in complete or partial response to first-line platinum-based chemotherapy	SOLO1 Study
	Myriad myChoice CDx test ^a	Tumor	Olaparib	For the maintenance treatment of adult patients with deleterious or suspected deleterious germline or somatic BRCA-mutated advanced epithelial ovarian, fallopian tube or primary peritoneal cancer who are in complete or partial response to first-line platinum-based chemotherapy	SOLO1 Study
	Myriad myChoice CDx test ^a	Tumor	Olaparib + Bevacizumab	For the maintenance treatment of adult patients with advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer who are in complete or partial response to first-line platinum-based chemotherapy and whose cancer is associated with HRD-positive status defined by either: 1) a deleterious or suspected deleterious BRCA mutation, and/or 2) genomic instability	PAOLA-1 Study
	BRACAnalysis CDx	Blood	Rucaparib	For the treatment of adult patients with a deleterious BRCA mutation (germline and/or somatic)-associated epithelial ovarian, fallopian tube, or primary peritoneal cancer who have been treated with 2 or more chemotherapies	ARIEL2 Study
	FoundationOne CDx	Tumor	Rucaparib	For the treatment of adult patients with a deleterious BRCA mutation (germline and/or somatic)-associated epithelial ovarian, fallopian tube, or primary peritoneal cancer who have been treated with 2 or more chemotherapies	ARIEL3 Study
	FoundationOne Liquid CDx	Blood	Rucaparib	For the treatment of adult patients with a deleterious BRCA mutation (germline and/or somatic)-associated epithelial ovarian, fallopian tube, or primary peritoneal cancer who have been treated with 2 or more chemotherapies	ARIEL2 Study
Prostate	Myriad myChoice CDx test ^a	Tumor	Niraparib	For the treatment of adult patients with advanced ovarian, fallopian tube, or primary peritoneal cancer who have been treated with 3 or more prior chemotherapy regimens and whose cancer is associated with HRD-positive status defined by either: (1) a deleterious or suspected deleterious BRCA mutation or (2) genomic instability and who have progressed more than 6 months after response to the last platinum-based chemotherapy	QUADRA Study
	BRACAnalysis CDx	Blood	Olaparib	For the treatment of adult patients with deleterious or suspected deleterious germline or somatic homologous recombination repair (HRR) gene-mutated metastatic castration-resistant prostate cancer (mCRPC) who have progressed following prior treatment with enzalutamide or abiraterone	PROfound Study
	FoundationOne CDx	Tumor	Olaparib	For the treatment of adult patients with deleterious or suspected deleterious germline or somatic homologous recombination repair (HRR) gene-mutated metastatic castration-resistant prostate cancer (mCRPC) who have progressed following prior treatment with enzalutamide or abiraterone	PROfound Study
	FoundationOne Liquid CDx	Plasma	Olaparib	For the treatment of adult patients with deleterious or suspected deleterious germline or somatic homologous recombination repair (HRR) gene-mutated metastatic castration-resistant prostate cancer (mCRPC) who have progressed following prior treatment with enzalutamide or abiraterone	PROfound Study

Table 1. Continued

Cancer	Assay	Sample type	Therapy	Indication	Trial
Breast	BRACAnalysis CDx	Blood	Olaparib	For the treatment of adult patients with deleterious or suspected deleterious gBRCAm, HER2-negative metastatic breast cancer who have been treated with chemotherapy in the neoadjuvant, adjuvant, or metastatic setting. Patients with hormone receptor (HR)-positive breast cancer should have been treated with a prior endocrine therapy or be considered inappropriate for endocrine therapy	OlympiAD Study
	BRACAnalysis CDx	Blood	Talazoparib	For the treatment of adult patients with deleterious or suspected deleterious germline BRCA-mutated (gBRCAm) HER2-negative locally advanced or metastatic breast cancer	EMBRACA Study
Pancreatic	BRACAnalysis CDx	Blood	Olaparib	For the maintenance treatment of adult patients with deleterious or suspected deleterious gBRCAm metastatic pancreatic adenocarcinoma whose disease has not progressed on at least 16 weeks of a first-line platinum-based chemotherapy regimen	POLO Study

^aAssesses genome-wide characteristics (potentially consequences of HRR impairment). Currently other companion diagnostic claims select patients by assessing genetic mutations (potentially causes of HRR impairment). Labels were accessed on FDA's website and are current as of June 2021.

optimize the use of this complex biomarker for the selection of patients for therapies targeting the DNA repair pathway, such as PARPi, and identify the elements used to define HR status that should be considered to best achieve consistent results.

Materials and Methods

Friends of Cancer Research convened a group of stakeholders from industry, academia, and government. We hosted bimonthly calls for 4 months with diverse stakeholders to discuss the best way to approach harmonizing HR status measurements using diagnostic assays. We set out to characterize how HRD is currently defined, measured, and used with regards to assays that measure HR status, and, ultimately, to propose common language and recommendations to improve consistency around the use of HR status as a biomarker. We reviewed literature associated with phase III trials, including those that led to FDA approvals, and current guidelines from the American Society of Clinical Oncology (ASCO) and National Comprehensive Cancer Network (NCCN) regarding the use of assays to measure HR status, investigated FDA labels of currently approved PARPi and reports on FDA-approved companion diagnostics validated to assess HR status, and discussed current laboratory and clinical practices.

Results

Assessment of current practices helped to answer what HRD is, how it is currently measured, and how assays currently assess HR status. During routine clinical use, HR status is assessed by measuring either evidence for potential causes of HRD indirectly (eg, genetic mutations) or potential consequence of deficiency in the HRR pathway (eg, genomic instability, mutational signatures; Fig. 2).

Definitions of HRD are Heterogeneous

The definition of HRD varies widely among the scientific and medical communities. Various terms have historically been used in the literature to describe HRD including BRCA-ness, BRCA-like, and genomic scarring.⁴²⁻⁴⁴ Additionally, HRD is

a complex biomarker whose definition may need refinement based on growing biological and clinical understanding. Defining HRD in terms of specific genetic mutations and/or the success of a PARPi may be a too narrow approach. Homologous recombination deficiency should not be solely defined by response to any one therapy, given that recent studies have shown that HRD has both a positive prognostic value in ovarian and other cancers and predictive value for PARPi and platinum therapy. Additionally, testing capabilities may evolve to better assess HR status with a functional assay.^{2,38}

Approaches for Assessing HR Status Vary Across Current FDA-Approved Companion Diagnostics

Given that HR status can inform treatment decisions and help predict improved outcomes for certain patients, it is essential to understand processes through which these decisions are made and identify best practices to maximize future benefit of HR status determination. To better understand how HR status is currently determined, we reviewed the labels of FDA approved companion diagnostics that test HR status for use with a PARPi. Companion diagnostics are medical devices regulated by FDA to support safe and effective use of a corresponding therapeutic. While “companion diagnostic” as a construct exists in a limited capacity in non-US markets, the clinical and analytical validation conferred by FDA review continues to inform global diagnostic use. In Europe, the European Union In Vitro Diagnostics Regulation emulates FDA's consideration of companion diagnostic regulation and is increasing review rigor by requiring validation studies.

We reviewed the FDA labels of Lynparza (olaparib), Zejula (niraparib), and Rubraca (rucaparib), PARPi that have used HRD as selection biomarkers in frontline and recurrent epithelial ovarian, fallopian tube, or primary peritoneal cancers. The language on the olaparib and niraparib labels is mostly consistent and defines HRD by either (1) a deleterious or suspected deleterious BRCA mutation and/or (2) genomic instability. Olaparib and niraparib use the FDA-approved diagnostic test the Myriad myChoice CDx test, which determines HR status by assessing the mutation status of *BRCA1/2* and/or genomic instability—measured by the evaluation of a combination of molecular measures to derive a genomic

instability (gLOH + TAI + LST; the Myriad myChoice CDx test Summary of Safety and Effectiveness Data). PARPi therapies may vary in the requirement of a companion diagnostic based on the clinical evidence for the therapeutic and depending on the indication. The rucaparib label does not mention HRD in the label but refers to patients with a *BRCA1/2* mutation (germline or somatic)-associated epithelial ovarian, fallopian tube, or primary peritoneal cancer based on an FDA-approved companion diagnostic for this therapy. Rucaparib uses the Foundation Medicine test Foundation Focus CDx ^{BRCA LOH} (now part of a broader FoundationOne CDx panel), which determines *BRCA1/2* mutation status as per the companion diagnostic claim, and determines HR status, which is defined by the mutation status of *BRCA1/2* and/or genomic instability as measured by gLOH.

In summary, the FDA-approved companion diagnostics test for somewhat different components, which may result in different HR status calls and subsequently different treatment decisions. Given that genomic instability patterns arising from HRD can look different in various tissue types, the assay used to assess HR status should be validated using samples from the intended use population being studied. Within an individual assay, cutoffs for determining HR status may differ by tissue type which requires further transparency around how thresholds are determined and whether these differences are due to chance, clinical trial approach, or biology.

Utilization of HR Status for Clinical Decision-Making Would Benefit from Addressing Uncertainties

Review of the literature, ASCO and NCCN guidelines, and expert discussions suggest uncertainty and inconsistency in how to use assays that measure HR status in the clinic, which could potentially drive low adoption for clinical decision making.⁴⁵ Because assays that measure HR status can identify cancers that are more likely to respond to PARPi therapies and predict positive outcomes, it is important to address the sources of uncertainty to enable clinicians to use these new diagnostic tools and provide their patients with optimal care.

We identified several areas which may lead to this lack of clarity including inconsistent reporting of HR status between studies and clinical trials, misaligned definitions of HR testing (eg, ‘mutation of HRR genes’ interchanged with ‘HR status’), complexity in the order of testing (eg, germline vs tumor, specific genes vs gene panel), interchangeability of tissue and plasma-based approaches, family history, and cancer type that warrant further investigation.

Discussion

Based on findings, we propose several considerations to bring better alignment in the field for use of HR assays in the clinic.

Use consistent language when describing HRD and align on a foundational definition that allows for a dynamic evolution of the term as HRD knowledge grows. We propose using the following definition: HRD is a phenotype that is characterized by the inability of a cell to effectively repair DNA double-strand breaks using the HRR pathway. Additionally, stakeholders should use the terminology “HRD” and “HRP” to define the presence or absence of an HRD phenotype, respectively.

Adopt a *minimum* set of requirements for the determination of HR status, the details on how HR status was

measured, and clearly report the type of test used should be clearly reported in publications. Greater clarity should be given to whether measures of both cause and consequence are needed to inform the determination of HR status in different contexts (Fig. 2). Evidence of consequence alone may not always be indicative of PARPi sensitivity due to the possibility of reversion mutations⁴¹; however, co-occurrence of consequence with cause can potentially support novel loss-of-function mutations.

Additionally, defining mutation status and zygosity of *BRCA1/2* (in the context of ovarian cancer) and genomic instability status to then determine HR status is complex, and we recommend transparency and standardization of the type of information used to determine *BRCA1/2* mutation status (ie, pathogenicity status of the variant), genomic instability (ie, molecular measures and parameters used to develop a score, value of continuous variable-if any), and criterion for cutoff selection. As we move beyond ovarian cancer, as well as investigate the role other HRR genes play in the cause of HRD, including different patterns of genomic instability exhibited in different cancer types, this additional information will be key to ensure consistency in results obtained across tests that determine HR status. Given that different tests may be used to determine HR status, and each uses different approaches for their determination, publications should include the name of the test used to determine HR status and specific clinical thresholds for level of deficiency (eg, HRD as measured by <assay name> and defined as <features evaluated and cutpoint(s)>).

Conduct studies to identify and assess sources of discordance among assays that assess HR status and identify sources of variability to inform optimal use of these assays for clinical decision making. This can be accomplished through a study that assesses concordance of HR status across assays. Additionally, it would be beneficial to create a clinician-targeted survey to identify major barriers to the understanding or use of HR status as a decision-making factor to determine treatment approaches.

Encourage all testing labs to report a *minimum* set of elements important for interpreting the clinical report in line with FDA reporting requirements for current FDA approved assays assessing HR status and contextualize clinical meaning to assist clinicians and patients with decision-making. Test developers should report whether the test is tumor-type dependent, what genomic findings were identified (as is being done), and genomic instability/scarring scores (with thresholds as per drug/companion diagnostic approval for that cancer type).

Conclusion

Biomarkers such as HR status play a critical role in treatment decisions for patients with cancer. It is therefore of utmost importance to build consensus on how to define HRD and the methodology for assessing HR status to promote alignment and optimal use of this biomarker to identify patients who would benefit from PARPi therapy. This publication provides findings from the group’s discussions that encourages diagnostic developers to consider the parameters that contribute to the determination of HR status. Perspectives captured in this manuscript complement notable academic efforts by professional societies, such as the European Society for Medical Oncology to assess methods for HRD testing as well as planned activities and surveys being conducted by the Association for Molecular Pathology,

Association of Community Cancer Centers, American Society for Clinical Oncology and College of American and Pathologists to assess current clinical practice and provide evidence-based subject matter expert recommendations regarding best practices and performance characteristics of clinical HRD molecular methods. We created recommendations and proposed best practices for industry stakeholders to benefit the entire cancer community that support alignment efforts and can evolve as biological and clinical advancements emerge to support robust use among oncologists and ensure assays enable the best possible care for patients.

Conflict of Interest

Rebecca C. Arend AstraZeneca, Clovis Oncology, GlaxoSmithKline, KIYATEC, Leap Therapeutics, Merck & Co, Sutro Biopharma, and VBL Therapeutics (SAB), Caris Life Sciences (C/A); Jonathan F. Baden Bristol Myers Squibb (E, OI), Johnson & Johnson (OI); Olena Barbash GlaxoSmithKline (E); Lakshman Ramamurthy GlaxoSmithKline (E); Nike Beaubier Tempus Labs, Inc. (E); Jerod Parsons Tempus Labs, Inc. (E); Tim French AstraZeneca (E, OI); Petar Jelinic Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., (E, OI); Matthew J. Marton Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., (E, OI); Kimberly McGregor Foundation Medicine (E), Roche (OI); Ethan S. Sokol Foundation Medicine (E), Roche (OI); Mark Sausen: Bristol Myers Squibb (E, OI); Albrecht Stenzinger Aignostics, AstraZeneca, AGCT, Bayer, BMS, Eli Lilly, Illumina, Incyte, Janssen, MSD, Novartis, Pfizer, Roche, Seattle Genetics, Takeda, Thermo Fisher (SAB); Bayer, BMS, Chugai, Incyte (RF); Kirsten M. Timms Myriad Genetics Inc. (E, OI); Diana Turco Myriad Genetics Inc. (E, OI); J. Andrew Williams AstraZeneca (E); Iris Wang Novartis Pharmaceuticals Corporation (E, OI); Elaine Wong-Ho Thermo Fisher Scientific (E). The other authors indicated no financial relationships.

(C/A) Consulting/advisory relationship; (RF) research funding; (E) employment; (ET) expert testimony; (H) honoraria received; (OI) ownership interests; (IP) intellectual property rights/inventor/patent holder; (SAB) scientific advisory board.

Author Contributions

Conception/design: M.D.S., D.M.V., R.C.A., J.F.B., O.B., N.B., T.F., N.G., P.H., P.J., M.J.M., K.M., J.P., L.R., M.S., E.S.S., A.S., H.S., K.M.T., D.T., I.W., J.A.W., E.W.-H., J.A.

Data analysis and interpretation: M.D.S., D.M.V., R.C.A., J.F.B., O.B., N.B., T.F., N.G., P.H., P.J., M.J.M., K.M., J.P., L.R., M.S., E.S.S., A.S., H.S., K.M.T., D.T., I.W., J.A.W., E.W.-H., J.A. Manuscript writing: M.D.S., D.M.V., R.C.A., J.F.B., O.B., N.B., T.F., N.G., P.H., P.J., M.J.M., K.M., J.P., L.R., M.S., E.S.S., A.S., H.S., K.M.T., D.T., I.W., J.A.W., E.W.-H., J.A. Final approval of manuscript: M.D.S., D.M.V., R.C.A., J.F.B., O.B., N.B., T.F., N.G., P.H., P.J., M.J.M., K.M., J.P., L.R., M.S., E.S.S., A.S., H.S., K.M.T., D.T., I.W., J.A.W., E.W.-H., J.A.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding

This work was supported by Friends of Cancer Research.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674.
2. Miller RE, Leary A, Scott CL, et al ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Ann Oncol*. 2020;31(12):1606-1622.
3. Heeke AL, Pishvaian MJ, Lynce F, et al Prevalence of Homologous recombination-related gene mutations across multiple cancer types. *JCO Precis Oncol*. 2018;1-13.
4. Marquard AM, Eklund AC, Joshi T, et al Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. *Biomark Res*. 2015;3:1-10.
5. Hoppe MM, Sundar R, Tan DSP, Jayasekharan AD. Biomarkers for homologous recombination deficiency in cancer. *J Natl Cancer Inst*. 2018;110(7):704-713.
6. Lord CJ, Ashworth A. PARP inhibitors: the first synthetic lethal targeted therapy Europe PMC Funders Group. *Science* 2017;355:1152-1158.
7. Iglehart JD, Silver DP. Synthetic lethality—a new direction in cancer-drug development. 2009;361:189-191. <https://doi.org/10.1056/NEJMe0903044>.
8. Banerjee S, Gonzalez-Martin A, Harter P, et al First-line PARP inhibitors in ovarian cancer: summary of an ESMO Open—Cancer Horizons round-table discussion. *ESMO Open*. 2020;5(6):e001110.
9. González-Martín A, Pothuri B, Vergote I, et al; PRIMA/ENGOT-OV26/GOG-3012 Investigators. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med*. 2019;381(25):2391-2402.
10. Coleman RL, Fleming GF, Brady MF, et al Veliparib with first-line chemotherapy and as maintenance therapy in ovarian cancer. *N Engl J Med*. 2019;381(25):2403-2415.
11. Moore KN, Secord AA, Geller MA, et al Niraparib monotherapy for late-line treatment of ovarian cancer (QUADRA): a multicentre, open-label, single-arm, phase 2 trial. *Lancet Oncol*. 2019;20(5):636-648.
12. Ray-Coquard I, Pautier P, Pignata S, et al; PAOLA-1 Investigators. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *N Engl J Med*. 2019;381(25):2416-2428.
13. How JA, Jazaeri AA, Fellman B, et al Modification of homologous recombination deficiency score threshold and association with long-term survival in epithelial ovarian cancer. *Cancers*. 2021;13(5):1-17.
14. Sokol ES, Pavlick D, Khiabani H, et al Pan-cancer analysis of BRCA1 and BRCA2 genomic alterations and their association with genomic instability as measured by genome-wide loss of heterozygosity. *JCO Precis Oncol*. 2020;4:442-465.
15. Vollebergh MA, Lips EH, Nederlof PM, et al Genomic patterns resembling BRCA1- and BRCA2-mutated breast cancers predict benefit of intensified carboplatin-based chemotherapy. *Breast Cancer Res*. 2014;16(3):R47.
16. Frey MK, Pothuri B. Homologous recombination deficiency (HRD) testing in ovarian cancer clinical practice: a review of the literature. *Gynecol Oncol Res Pract*. 2017;4:1-11.
17. Nguyen L, Martens J, Van Hoeck A, et al Pan-cancer landscape of homologous recombination deficiency. *Nat Commun*. 2020;11(5584):1-12.
18. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb Perspect Biol*. 2015;7(4):a016600.
19. Heeke AL, Baker T, Lynce F, et al Prevalence of homologous recombination deficiency among all tumor types. *J Clin Oncol*. 2017;35:1502.

20. da Cunha Colombo Bonadio RR, Fogace RN, Miranda VC, Diz MDPE. Homologous recombination deficiency in ovarian cancer: a review of its epidemiology and management. *Clinics (Sao Paulo)*. 2018;73(suppl 1):1-6.
21. Mirza MR, Monk BJ, Herrstedt J, et al; ENGOT-OV16/NOVA Investigators. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N Engl J Med*. 2016;375(22):2154-2164.
22. de Jonge MM, Auguste A, van Wijk LM, et al Frequent homologous recombination deficiency in high-grade endometrial carcinomas. *Clin Cancer Res*. 2019;25(3):1087-1097.
23. den Brok WD, Schrader KA, Sun S, et al Homologous recombination deficiency in breast cancer: a clinical review. *JCO Precis Oncol*. 2017;1:1-13.
24. Zhao EY, Shen Y, Pleasance E, et al Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer. *Clin Cancer Res*. 2017;23(24):7521-7530.
25. Melinda LT, Kirsten MT, Julia R, et al Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin Cancer Res*. 2016;22:3764-3773.
26. Felice D, Dove Press P. Defective DNA repair mechanisms in prostate cancer: impact of olaparib. *Drug Des Dev Ther*. 2017:547-552.
27. Waddell N, Pajic M, Patch AM, et al; Australian Pancreatic Cancer Genome Initiative. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. 2015;518(7540):495-501.
28. Lotan TL, Kaur HB, Salles DC, et al Homologous recombination deficiency (HRD) score in germline BRCA2- versus ATM-altered prostate cancer. *Mod Pathol*. 2021;34:1185-1193. <https://doi.org/10.1038/s41379-020-00731-4>.
29. Li X, Heyer WD. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res*. 2008;18(1):99-113.
30. Norquist BM, Brady MF, Harrell MI, et al Mutations in homologous recombination genes and outcomes in ovarian carcinoma patients in GOG 218: an NRG Oncology/Gynecologic Oncology Group Study. *Clin Cancer Res*. 2018;24(4):777-783.
31. Abkevich V, Timms KM, Hennessy BT, et al Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer*. 2012;107(10):1776-1782.
32. Birkbak NJ, Wang ZC, Kim JY, et al Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov*. 2012;2(4):366-375.
33. Popova T, Manié E, Rieunier G, et al Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res*. 2012;72(21):5454-5462.
34. Meijer TG, Verkaik NS, Sieuwerts AM, et al Functional ex vivo assay reveals homologous recombination deficiency in breast cancer beyond BRCA gene defects. *Clin Cancer Res*. 2018;24(24):6277-6287.
35. Sztupinski Z, Diossy M, Krzystanek M, et al Detection of molecular signatures of homologous recombination deficiency in prostate cancer with or without BRCA1/2 mutations. *Clin Cancer Res*. 2020;26(11):2673-2680.
36. Davies H, Glodzik D, Morganella S, et al HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med*. 2017;23(4):517-525.
37. Malihi PD, Graf RP, Rodriguez A, et al Single-cell circulating tumor cell analysis reveals genomic instability as a distinctive feature of aggressive prostate cancer. *Clin Cancer Res*. 2020;26(15):4143-4153.
38. Castroviejo-Bermejo M, Cruz C, Llop-Guevara A, et al A RAD 51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. *EMBO Mol Med*. 2018;10:1-16.
39. van Wijk LM, Vermeulen S, Meijers M, et al The recap test rapidly and reliably identifies homologous recombination-deficient ovarian carcinomas. *Cancers*. 2020;12:1-16.
40. Eckstein M, Bloom KJ, Riccelli P, et al HRD in ovarian cancer: defined today, evolving for the future. 2020;38:e18052. https://doi.org/10.1200/JCO.2020.38.15_suppl.e18052.
41. Takaya H, Nakai H, Takamatsu S, Mandai M, Matsumura N. Homologous recombination deficiency status-based classification of high-grade serous ovarian carcinoma. *Sci Rep*. 2020;10(1):1-8.
42. Turner N, Tutt A, Ashworth A. Hallmarks of “BRCAness” in sporadic cancers. *Nat Rev Cancer*. 2004;4:814-819.
43. Tan DS, Rothermundt C, Thomas K, et al “BRCAness” syndrome in ovarian cancer: a case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations. *J Clin Oncol*. 2008;26(34):5530-5536.
44. Konstantinopoulos PA, Spentzos D, Karlan BY, et al Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. *J Clin Oncol*. 2010;28(22):3555-3561.
45. Stover EH, Fuh K, Konstantinopoulos PA, Matulonis UA, Liu JF. Clinical assays for assessment of homologous recombination DNA repair deficiency. *Gynecol Oncol*. 2020;159(3):887-898.

Abstract STO77. Assessing Variability across HRD Assays: Findings from the Friends' HRD Harmonization Project

H. Stires¹, Z. Zhang², L. McShane², J. Bieler³, L. Chen⁴, M. Gupta⁵, A. Lazar⁶, B. McKelvey¹, S. Pabla⁷, J. Parsons⁸, O. Serang⁹, S. Shams¹⁰, E. Sokol¹¹, E. Starks¹², B. Thomas¹³, S. Yang¹⁴, J. Yen¹⁵, M. Stewart¹, J. Allen¹

¹Friends of Cancer Research, Washington, District of Columbia; ²National Cancer Institute, Bethesda, MD; ³SOPHiA GENETICS, SaintSulpice, Switzerland; ⁴Frederick National Laboratory for Cancer Research, Frederick, MD; ⁵Clinical Sequencing Division, Thermo Fisher Scientific, South San Francisco, CA; ⁶The University of Texas MD Anderson Cancer Center, Houston, TX; ⁷Omniseq, Buffalo, NY; ⁸Tempus Labs, Chicago, IL; ⁹DNAnexus, Mountain View, CA; ¹⁰Bionano Genomics, San Diego, CA; ¹¹Foundation Medicine, Cambridge, MA; ¹²Invitae, San Francisco, CA; ¹³Neogenomics, Fort Myers, FL; ¹⁴Amoy Diagnostics Co., Ltd, Xiamen, Fujian, People's Republic of China; ¹⁵Guardant Health, Palo Alto, CA.

Introduction: Homologous recombination deficiency (HRD) assays determine eligibility for treatment with poly (ADP-ribose) polymerase inhibitors and other DNA repair targeting drugs. The assays measure several factors to define homologous recombination (HR) status including causes (i.e., inactivation in HR pathway genes) and consequences (i.e., genomic instability) of HRD. Methodological variability across HRD assays has not been investigated thoroughly, and an empirical assessment of assay variability may support broader adoption of HRD and strengthen clinical interpretation of test results.

Methods: Friends of Cancer Research (Friends) initiated a unique partnership with HRD assay developers and other key stakeholders to characterize differences in assay factors and assess levels of agreement and variability across HRD assays. First, we surveyed HRD assay developers (n=20) about factors their assays measure to determine HR status. Subsequently, a subset of assay developers (n=11) measured in silico and reported HR status and the contributing factor(s) for 348 TCGA ovarian cancer samples. We performed pairwise comparisons of assay's HR status calls to determine the level of agreement and considered specific factors measured by each assay to identify potential sources of variation. Additionally, we analyzed HR status agreement for BRCA1/2 mutated versus wild-type BRCA1/2 samples.

Results: The 20 surveyed HRD assays are heterogeneous in the factors they measure. Although all assays consider BRCA1/2 mutations, assays also variously consider genomic loss of heterozygosity (gLOH; 75% of assays), additional HRR genes (55%), telomeric allelic imbalance (TAI; 45%), and large-scale state transitions (LST; 45%). For assays involved in the TCGA analysis, the range of percent positivity (% patients with HRD) was 9%-67% with a median of 49%. Rates of HRD were higher in assays that included gLOH, TAI, and/or LST. Median positive percent agreement (PPA) was 74% and median negative percent agreement was 81%. The presence of BRCA1/2 mutations was associated with an increase in PPA. The median Spearman correlation for pairwise comparisons of ranked continuous HRD scores was 0.66 and 0.70 for %gLOH.

Conclusions: Preliminary findings demonstrate variation in the factors measured and the HR status calls made across HRD assays. Some of the variation in HR status calls could be due to the nature of the TCGA dataset, and future studies will aim to understand assay agreement from freshly extracted formalin-fixed, paraffin-embedded human archival ovarian tumor samples. Understanding the agreement among assays will help to inform assay interpretation and improve consistency between HR status calls and alignment of HRD scores across HRD assays to help patients and providers make appropriate treatment decisions.

PRECISION MEDICINE

Changes in Circulating Tumor DNA Reflect Clinical Benefit Across Multiple Studies of Patients With Non–Small-Cell Lung Cancer Treated With Immune Checkpoint Inhibitors

Diana Merino Vega, PhD¹; Katherine K. Nishimura, PhD, MPH²; Névine Zariffa, PhD³; Jeffrey C. Thompson, MD⁴; Antje Hoering, PhD²; Vanessa Cilento, MPH²; Adam Rosenthal, MS²; Valsamo Anagnostou, MD, PhD⁵; Jonathan Baden, MS⁶; Julia A. Beaver, MD⁷; Aadel A. Chaudhuri, MD, PhD^{8,9,10,11}; Darya Chudova, PhD¹²; Alexander D. Fine, PhD¹³; Joseph Fiore, PharmD¹⁴; Rachel Hodge, PhD¹⁵; Darren Hodgson, PhD¹⁶; Nathan Hunkapiller, PhD^{17,18}; Daniel M. Klass, PhD¹⁹; Julie Kobie, PhD²⁰; Carol Peña, PhD²¹; Gene Pennello, PhD, MS²²; Neil Peterman, PhD²³; Reena Philip, PhD²⁴; Katie J. Quinn, PhD¹²; David Raben, MD²⁵; Gary L. Rosner, ScD⁵; Mark Sausen, PhD⁶; Ayse Tezcan, MPH, PhD²³; Qi Xia, PhD²⁶; Jing Yi, PhD²⁵; Amanda G. Young, PhD¹³; Mark D. Stewart, PhD¹; Erica L. Carpenter, MBA, PhD²⁷; Charu Aggarwal, MD, MPH²⁷; and Jeff Allen, PhD¹

abstract

PURPOSE As immune checkpoint inhibitors (ICI) become increasingly used in frontline settings, identifying early indicators of response is needed. Recent studies suggest a role for circulating tumor DNA (ctDNA) in monitoring response to ICI, but uncertainty exists in the generalizability of these studies. Here, the role of ctDNA for monitoring response to ICI is assessed through a standardized approach by assessing clinical trial data from five independent studies.

PATIENTS AND METHODS Patient-level clinical and ctDNA data were pooled and harmonized from 200 patients across five independent clinical trials investigating the treatment of patients with non–small-cell lung cancer with programmed cell death-1 (PD-1)/programmed death ligand-1 (PD-L1)–directed monotherapy or in combination with chemotherapy. CtDNA levels were measured using different ctDNA assays across the studies. Maximum variant allele frequencies were calculated using all somatic tumor-derived variants in each unique patient sample to correlate ctDNA changes with overall survival (OS) and progression-free survival (PFS).

RESULTS We observed strong associations between reductions in ctDNA levels from on-treatment liquid biopsies with improved OS (OS; hazard ratio, 2.28; 95% CI, 1.62 to 3.20; $P < .001$) and PFS (PFS; hazard ratio 1.76; 95% CI, 1.31 to 2.36; $P < .001$). Changes in the maximum variant allele frequencies ctDNA values showed strong association across different outcomes.

CONCLUSION In this pooled analysis of five independent clinical trials, consistent and robust associations between reductions in ctDNA and outcomes were found across multiple end points assessed in patients with non–small-cell lung cancer treated with an ICI. Additional tumor types, stages, and drug classes should be included in future analyses to further validate this. CtDNA may serve as an important tool in clinical development and an early indicator of treatment benefit.

JCO Precis Oncol 6:e2100372. © 2022 by American Society of Clinical Oncology

Creative Commons Attribution Non-Commercial No Derivatives 4.0 License 

INTRODUCTION

The recent approval of programmed cell death-1 (PD-1)/programmed death ligand-1 (PD-L1) inhibitors as frontline therapy for advanced non–small-cell lung cancer (NSCLC) has changed the treatment paradigm for this disease.¹⁻⁶ However, not all patients respond to immune checkpoint inhibitors (ICI), and some may experience clinically significant, and sometimes long-lived, toxicity.⁷ Disease response is currently assessed with clinical and radiographic evaluation, with the first imaging assessment usually after 8 weeks on ICIs. However, clinical assessments are subjective, difficult

to standardize, may lack the necessary sensitivity to identify very early stages of progressive disease, and may misinterpret tumor responses in the case of pseudoprogression.⁸ Hence, accurate, early, and objective predictors of response to ICI therapy are needed.

Next-generation sequencing of circulating tumor DNA (ctDNA) has been recently established as a sensitive, less invasive, and accurate means to detect therapeutically actionable mutations in patients as well as to identify the emergence of resistance mutations in patients receiving targeted therapies. However, the use of this technology to monitor response to therapy is less

CONTEXT

Key Objective

Can changes in circulating tumor DNA (ctDNA) reflect clinical benefit across multiple, independent studies of patients with non–small-cell lung cancer treated with immune checkpoint inhibitors?

Knowledge Generated

Analyses confirm an association between changes in ctDNA levels and clinical benefit for patients with non–small-cell lung cancer treated with varying lines of anti–programmed cell death-1 (PD-1)/programmed death ligand-1 (PD-L1) therapy.

Harmonization strategies were developed to help address differences in ctDNA collection timing, ctDNA assay results, and clinical variables across different clinical studies.

Relevance

Our study provides supporting evidence that ctDNA may serve as an early predictor of treatment response. Given the multitude of recent studies investigating the use of ctDNA as a minimally invasive way to measure treatment outcome, these results are timely by confirming observations seen across multiple, independent studies and by outlining harmonization strategies to support future studies and meta-analyses to validate ctDNA as an end point in drug development.

defined for ICIs.⁹⁻¹³ Serial ctDNA measurements may yield additional insights into a patient's disease, providing a more timely assessment of response to treatment than traditional clinical and radiologic assessments. If shown to correlate with treatment response, monitoring ctDNA changes during treatment may improve disease management.¹⁴⁻¹⁸

Several recent studies suggest a potential role for ctDNA in monitoring response to ICI therapy and have investigated how changes in ctDNA levels may be associated with outcomes. These studies have identified a correlation between on-therapy reductions in ctDNA and objective response rate, progression-free survival (PFS), and overall survival (OS).¹⁹⁻²³ However, uncertainty exists in the generalizability of these studies, since they often used different methods of ctDNA assessment, had variable on-treatment blood collection time points, had heterogeneity in the patient populations, and implemented a variety of methods to calculate ctDNA changes over time.

To address the need for a standardized approach to assess the role of ctDNA as a potential tool for monitoring response to ICI treatment as well as to develop a robust data set evaluating the relationship between ctDNA changes during ICI treatment and clinical outcomes, Friends of Cancer Research (Friends) launched the ctDNA for Monitoring Treatment Response (ctMoniTR) pilot project. The first step of ctMoniTR pooled and harmonized data from five independent studies focused on patients with NSCLC receiving PD-(L)1–directed monotherapy or combination with chemotherapy. The results from this multi-institutional study are presented and discussed, providing further evidence of ctDNA as a noninvasive and dynamic indicator of clinical outcome to ICI.

PATIENTS AND METHODS

Patients

Anonymized patient-level clinical and ctDNA data from five independent clinical trials were collected and included 254

patients (Data Supplement).^{20-22,24,25} Each study reviewed patients' informed consent approved by the local institutional review board to ensure their data were suitable for secondary use beyond their original intent. Patients with NSCLC who had been treated with varying lines of anti-PD-(L)1 therapy, either as monotherapy or in combination with standard chemotherapy, and who had a pretreatment ctDNA sample (no earlier than 14 days before the start of treatment) and at least one on-treatment ctDNA sample (no later than 70 days from the initiation of treatment) were included. As this was a pilot project, these time points were selected to allow inclusion of the largest number of samples. The five data sets were split into seven cohorts, with each cohort representing a unique study or trial arm. Initial criteria for patient inclusion/exclusion and strategies for minimizing bias in a combined data set were established before analysis (Data Supplement).

Clinical Outcomes and Covariates

OS and PFS were defined as the number of days between treatment initiation and death resulting from any cause, and the number of days between treatment initiation and death from any cause or progression, respectively. Tumor response was evaluated according to the RECIST, version 1.1, and confirmed by local or central review.²⁶ Durable clinical benefit was defined as maintenance of PFS at 6 months from treatment initiation (PFS6).²⁷ Patients who did not progress on study but were lost to follow-up within 6 months of treatment initiation (n = 11) were excluded from the PFS6 analysis. Additional clinical descriptors were collected and harmonized according to a common set of definitions (Data Supplement).

ctDNA Data

All studies used similar plasma collection methods (Data Supplement) that met the minimum prespecified assay standards (Data Supplement) and provided ctDNA results according to their individual protocols. Various next-generation

sequencing-based ctDNA assays (Data Supplement), including targeted panels and whole-genome sequencing, were used and, as a result, performance metrics may vary across the platforms. Variant allele frequencies (VAF), defined as the number of mutant alleles divided by the total number of mutant and wild-type alleles, were reported from four of the five studies. The fifth study assessed ctDNA changes with a whole-genome sequencing approach using copy-number alterations and local changes in ctDNA fragment length to determine a tumor fraction ratio.²¹ Variants contributing to the calculation of VAF met internal assay-specific quality standards. Germline and clonal hematopoiesis variants were removed according to each study's original protocol (Data Supplement) or, for one study, by the independent analysis center (Data Supplement).

Derived ctDNA Metrics

Mean, median, and maximum VAF values were calculated using all somatic tumor-derived variants eligible for analysis in each unique patient sample, regardless of whether they were detected at baseline. For patients with nondetectable (ND) ctDNA, the VAFs were assumed to be indeterminably low and were set to a value of 0; additional data handling details are in the Data Supplement. The percent change of the mean, median, or maximum VAF value from baseline (T0) to the first on-treatment sample collected within 70 days of treatment initiation (T1) was calculated as

$$\begin{aligned} \text{Percent Change of Mean VAF} &= (\text{mean VAF}_{T1} \\ &\quad - \text{mean VAF}_{T0}) / \text{mean VAF}_{T0} \\ \text{Percent Change of Median VAF} &= (\text{median VAF}_{T1} \\ &\quad - \text{median VAF}_{T0}) / \text{median VAF}_{T0} \\ \text{Percent Change of Maximum VAF} &= (\text{maximum VAF}_{T1} \\ &\quad - \text{maximum VAF}_{T0}) / \text{maximum VAF}_{T0} \end{aligned}$$

Then, three types of ctDNA metrics were calculated for analysis: (1) continuous percent change variable using the raw percent change value, with a cap in cases with percentage increase of 500% to mitigate the impact of outliers; (2) binary variable using a cutpoint of -50% change in VAF as the threshold, where this optimal cutpoint was determined using the running log-rank method²⁸; and (3) the three-level variable, which used cohort-specific thresholds to identify the 50% most extreme patients within each cohort exhibiting a strong decrease in ctDNA from baseline (decrease), the 50% most extreme patients exhibiting a strong increase in ctDNA (increase), and the remaining patients in a middle category with modest reductions or increases in ctDNA (intermediate; Data Supplement).

Statistical Analyses

The three-level ctDNA metric was modeled as an ordinal variable with three categories representing patients with a decrease in ctDNA from baseline, an intermediate change, or an increase. Kruskal-Wallis tests were used to compare the medians of continuous variables, and Wald chi-Square tests were used to compare proportions of categorical

variables, with Fisher's exact test used in cases where assumptions for utilization of the chi-square test were not met. Survival probabilities (OS and PFS) were estimated using the Kaplan-Meier method,²⁹ using a 70-day landmark from treatment initiation to ensure that the ctDNA metric reflected a change in ctDNA that occurred before patients were assessed for survival outcome. Overall and pairwise comparisons between strata in Kaplan-Meier analyses were calculated using log-rank tests. Univariate and multivariate Cox proportional hazards models were used to assess associations with OS and PFS, with *P* values derived from the log-likelihood test, and covariates that were measured after treatment initiation modeled as time-dependent covariates. Univariate and multivariate logistic regression models were used to assess associations with binary clinical end points (partial response [PR] or better, and PFS6). All models accounted for cohort-specific risks using cohort-stratified models, where cohort was adjusted by stratification, which allows for a different baseline risk within each cohort group. All statistical tests with *P* value < .05 were considered statistically significant. As this was an exploratory pilot project, *P* values were not adjusted for multiple tests. Analyses were done using the SAS statistical software package (SAS Institute, Cary, NC) or R (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

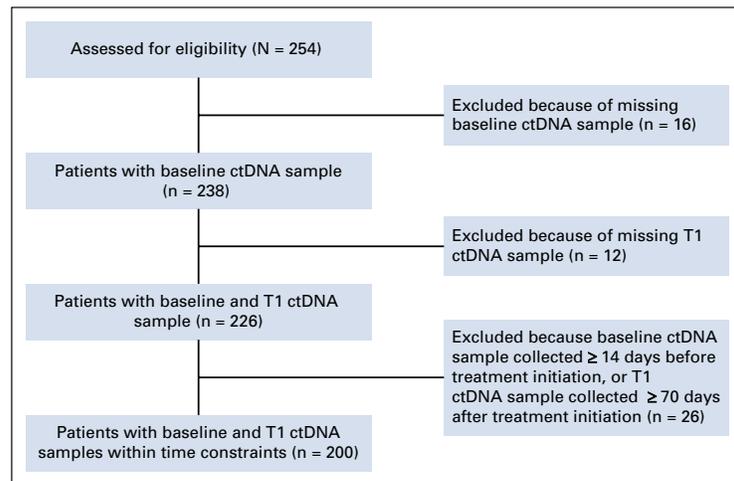
Analysis Data Set

A total of 254 patients were considered for inclusion, with 200 patients included in the final data set after excluding patients who failed to meet study criteria (Fig 1; full population demographics shown in the Data Supplement). Broad heterogeneity was observed across cohorts with noticeable differences in age, sex, stage at enrollment, histology, programmed death ligand-1 (PD-L1) expression, and number of prior lines of therapy (Table 1). Among all clinical covariates, smoking history was the only one to be univariately associated with changes in ctDNA values (Data Supplement).

ctDNA Collection Timing and ctDNA Metrics

Descriptive analyses revealed that the timing and frequency of ctDNA samples varied between cohorts because of differences in the protocols used within each study (Fig 2). There was also variability across cohorts in the number of variants detected, the magnitude of VAF values, and the range of baseline mean, median, and maximum VAF values (Data Supplement). Considering the likelihood that differences in these data could be related to the assay used, the 3-level Max VAF Percent Change Group results are shown here, since this metric accounted for differences in distributions by using cohort-specific thresholds to categorize patients. This metric also demonstrated the most consistent results for OS, PFS, and durable clinical benefit. The results for the other ctDNA metrics are available in the Data Supplement. Within the 3-level Max VAF Percent Change

FIG 1. Flow diagram. ctDNA, circulating tumor DNA.



Group metric, 63 (32%) patients had a decrease, 103 (51%) had an intermediate change, and 34 (17%) an increase in ctDNA levels from baseline while on treatment.

Changes in ctDNA Are Associated With Survival End Points

Strong and consistent associations between reductions in ctDNA levels and improved OS were observed in unadjusted Cox models (Data Supplement) and adjusted Cox models with cohort stratification and adjustment by baseline clinical covariates (Fig 3A). For example, each increase in the category of the three-level Max VAF Percent Change Group variable (from decrease, to intermediate, to increase in Max VAF) was associated with an increased risk of death (adjusted hazard ratio, of 2.28 [95% CI, 1.62 to 3.20; $P < .001$]), after adjusting for baseline clinical covariates. Baseline ctDNA values, including ND samples, were not found to be associated with OS (Data Supplement). OS Kaplan-Meier plots showed a strong separation in the different ctDNA categories, with statistically significant differences in the pairwise comparisons, and 1-year survival rates of 75%, 58%, and 32% for patients with a decrease, intermediate change, or increase in Max VAF, respectively (Fig 3B). Additional Kaplan-Meier and univariate associations for OS are available in the Data Supplement.

Similar observations occurred when examining the ctDNA associations with PFS in unadjusted Cox models (Data Supplement) and adjusted Cox models with cohort stratification and adjustment by baseline clinical covariates (Fig 4A), where the adjusted hazard ratio of 1.76 (95% CI, 1.31 to 2.36; $P < .001$) indicated that each increase in the categories of the three-level Max VAF Percent Change Group variable (from decrease, to intermediate, to increase in Max VAF) was associated with an increased risk of progression or death, after adjusting for baseline clinical covariates. Similar to the OS analysis, baseline ctDNA values, including ND samples, were not associated with PFS. The PFS Kaplan-Meier plot

revealed that patients with a decrease in the maximum VAF had better PFS compared with the other two groups, but there was no apparent separation in PFS between patients in the intermediate and increase categories (Fig 4B). Additional Kaplan-Meier and univariate associations for PFS are available in the Data Supplement. Of note, in the adjusted Cox models for both OS and PFS, smoking history was associated with improved survival outcomes. This finding is consistent with previous studies that argued that cancers resulting from the accumulation of tobacco-related mutations may have increased tumor mutational burden and respond especially favorably to immunotherapies.^{30,31} Additionally, there was a lack of association with PD-L1 positivity, which was likely because of variation in how it is measured and defined in each clinical trial.

Changes in ctDNA Are Associated With Improved Tumor Response

Reductions in ctDNA were also associated with improved tumor response, defined as achieving a RECIST classification of PR or complete response. Logistic regression models with cohort stratification and adjustment by baseline clinical covariates yielded an adjusted odds ratio of 0.19 (95% CI, 0.08 to 0.45; $P < .001$) for intermediate versus decrease and 0.11 (0.03 to 0.38) for increase versus decrease, suggesting that each increase in the strata of the three-level Max VAF Percent Change Group variable was associated with a decreased likelihood in achieving PR or better, after adjusting for baseline clinical covariates (Table 2). Baseline ctDNA values were not univariately associated with achieving PR or better (Data Supplement). Additional univariate associations and results for other ctDNA metrics are included in the Data Supplement.

Changes in ctDNA Are Associated With Durable Clinical Benefit

Logistic regression models with cohort stratification and adjustment by baseline clinical covariates found that

TABLE 1. Patient Demographics

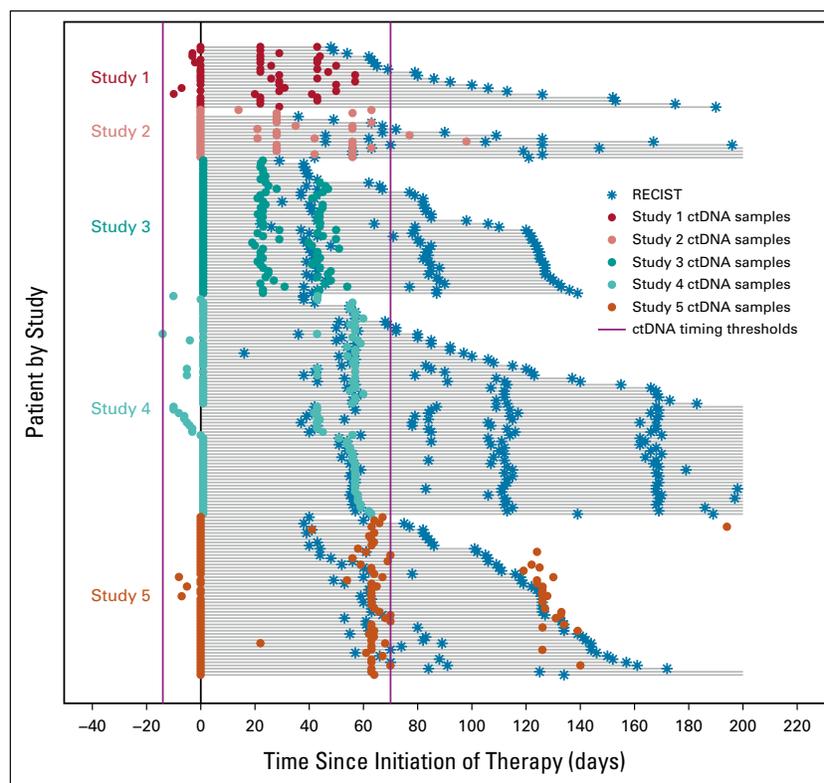
Trait	Description	Cohort								Overall, n/N (%)	P
		1	2	3a	4a	4b	5a	5b			
Age, years	Age ≥ 66 years at enrollment, No. (%)	19/20 (95)	9/16 (56)	19/43 (44)	16/55 (29)	5/15 (33)	12/24 (50)	14/27 (52)	14/27 (52)	94/200 (47)	< .001
Sex	Female, No. (%)	9/20 (45)	8/16 (50)	10/43 (23)	22/55 (40)	2/15 (13)	15/24 (63)	11/27 (41)	11/27 (41)	77/200 (39)	.013
Race	White, No. (%)	13/20 (65)	8/14 (57)	32/43 (74)	28/55 (51)	11/15 (73)	15/24 (63)	22/27 (81)	22/27 (81)	129/198 (65)	.094
Smoking status	Ever smoked, No. (%)	15/20 (75)	12/14 (86)	37/43 (86)	47/55 (85)	12/15 (80)	23/24 (96)	24/27 (89)	24/27 (89)	170/198 (85.9)	.577 ^a
ECOG	ECOG performance status ≥ 1, No. (%)	13/19 (68)	ND	26/43 (60)	35/55 (64)	12/15 (80)	11/24 (46)	18/27 (67)	18/27 (67)	115/183 (63)	.360
Stage at enrollment	Advanced stage (stage IV), No. (%)	18/20 (90)	14/16 (88)	43/43 (100)	39/55 (71)	13/15 (87)	24/24 (100)	27/27 (100)	27/27 (100)	178/200 (89)	< .001 ^a
Histology	Squamous, No. (%)	3/20 (15)	7/16 (44)	12/43 (28)	13/55 (24)	9/15 (60)	7/24 (29)	0/27 (0)	0/27 (0)	51/200 (26)	< .001
PD-L1 expression	PD-L1-positive (TPS ≥ 1%, or TC/IC PD-L1-positive), No. (%)	14/19 (74)	11/11 (100)	30/42 (71)	48/53 (91)	12/15 (80)	23/24 (96)	13/26 (50)	13/26 (50)	151/190 (79)	< .001 ^a
Prior therapy	Prior lines of systemic treatment ≥ 1, No. (%)	8/20 (40)	11/15 (73)	43/43 (100)	55/55 (100)	11/15 (73)	6/24 (25)	0/27 (0)	0/27 (0)	134/199 (67)	< .001

NOTE: Each cohort represents a unique study (cohort number) or trial arm (cohort letter) within a study. The proportion is calculated as the percent of patients with a given trait, within each cohort. If a patient was missing data on a given trait, this was reflected in the total count for the cohort; therefore, N may be < 200. Studies were blinded for analyses.

Abbreviations: ECOG, Eastern Cooperative Oncology Group; ND, no data; PD-L1, programmed death ligand-1; TC/IC, tumor cells/immune cells; TPS, tumor proportion score.

^aP value from Fisher's exact test, otherwise χ^2 .

FIG 2. Timing of plasma collection and tumor response per patient by study in the analysis data set (N = 200). Unique patients are represented as horizontal lines, with markers denoting the timing of RECIST evaluations and ctDNA samples. The x-axis is truncated at 200 days, with some patients having longer follow-up. ctDNA, circulating tumor DNA.



decreases in ctDNA were associated with achieving durable clinical benefit, defined as PFS \geq 6 months (PFS6). This analysis yielded an adjusted odds ratio of 0.13 (95% CI, 0.05 to 0.34; $P < .001$) for intermediate versus decrease, and 0.06 (95% CI, 0.02 to 0.22) for increase versus decrease, interpreted as a decreasing likelihood of achieving PFS6 with each increase in the ctDNA Max VAF metric category (Table 2). No other clinical covariates were statistically significant in the adjusted model, and the ctDNA values at baseline were also not found to be associated with PFS6. Additional univariate associations and results for other ctDNA metrics are included in the Data Supplement.

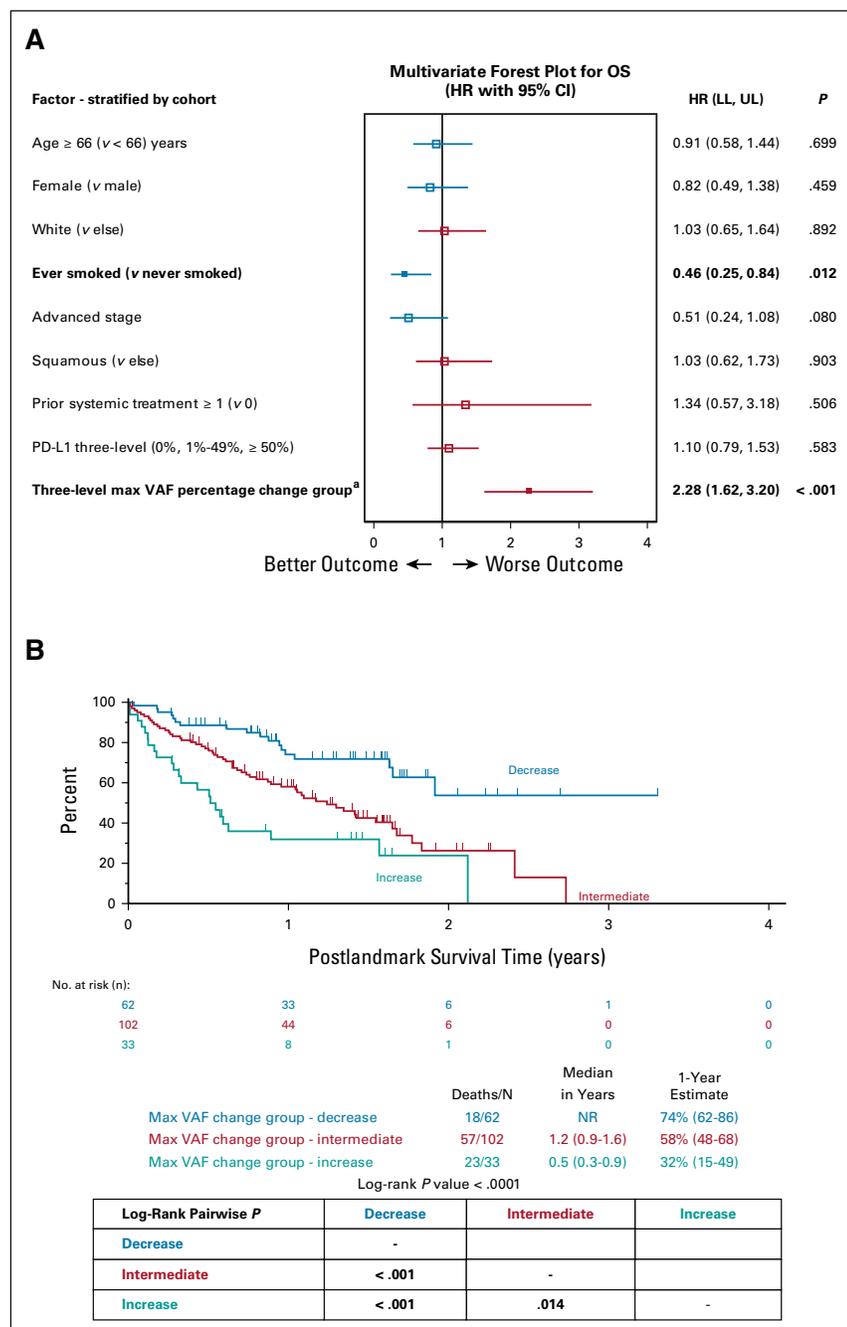
DISCUSSION

Among patients with NSCLC treated with ICI whose data were analyzed in aggregate, consistent and robust associations between reductions in ctDNA and clinical benefit were found across multiple end points. Although the results presented in this manuscript are consistent with recent reports, these individual studies have limited sample sizes, and were constrained in their generalizability, given that each study used a particular treatment and a specific ctDNA assay on a carefully selected group of patients.^{19-23,25,32} The heterogeneity of the data sets included required various harmonization strategies to address differences in ctDNA collection timing, ctDNA assay results, and clinical variables.

These strategies successfully minimized bias and confounding factors and were equally valuable in establishing useful methodologies for combining data sets collected from disparate sources. By pooling and harmonizing the results from independent studies, the results of this study show that, even when analyzed across five different clinical trials, using multiple ICIs in differing NSCLC populations, with different sample collection time points and different ctDNA assays, the on-treatment changes in ctDNA levels correlate with outcome. These correlations hold true in analyses using ctDNA as a dichotomized, trichotomized, or continuous variable, and using all outcome measures evaluated (OS, PFS, best response, and PFS6).

In the literature, there is a lack of standardization in the methods used to quantitate ctDNA changes and evaluate their association with clinical outcomes. Previous studies have generally used different metrics, such as mean, median, or maximum VAF, mutant molecules per unit volume of blood or plasma, or absolute numbers of mutations observed at one point in time.^{24,32} Moreover, different thresholds have been used to determine significant changes in ctDNA, such as one log reduction, two-fold change or statistically distinguishable changes with non-overlapping CIs, percent change in the absolute ctDNA levels, or a ratio of on-treatment VAF to baseline VAF, with a molecular response set at $> 50\%$ decrease.^{19,21,23,32,33}

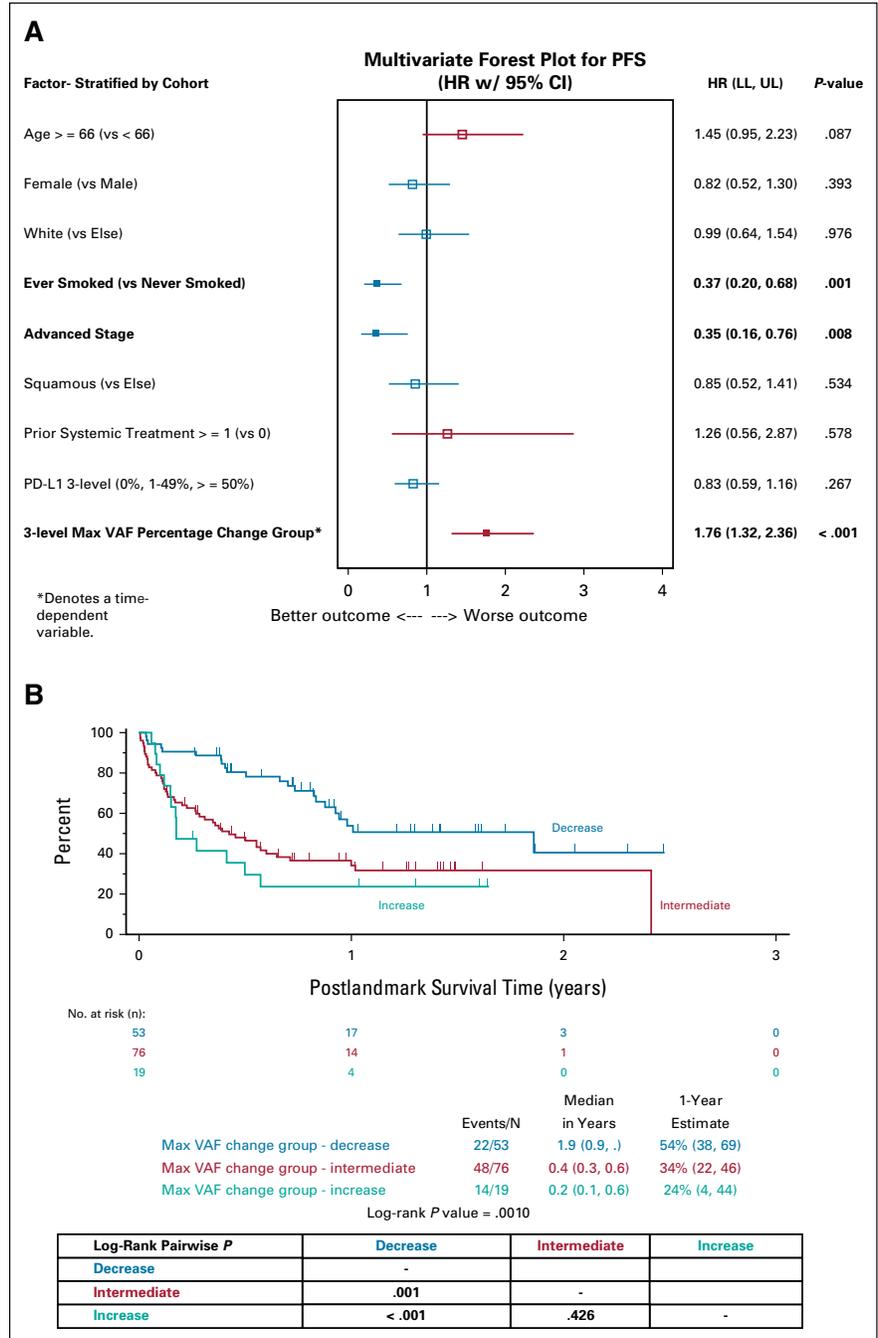
FIG 3. (A) Forest plot with Cox regression results for OS and the three-level max VAF percent change group variable, adjusted by baseline clinical covariates. Red means the HR is > 1.0 (increased risk of death) and blue means the HR is < 1.0 (decreased risk of death); unfilled box = nonsignificant *P* value, filled box = significant *P* < .05. (B) Kaplan-Meier plot for OS and the three-level max VAF percent change groups, landmarked at 70 days from treatment initiation (the sampling window for the first on-treatment ctDNA sample); patients with an event during the 70-day landmark were excluded from the analysis. ^aDenotes a time-dependent variable. ctDNA, circulating tumor DNA; HR, hazard ratio; LL, lower limit; OS, overall survival; PD-L1, programmed death ligand-1; UL, upper limit; VAF, variant allele frequency.



Thus, one aim of this study was to compare different ctDNA metrics to identify those that yielded the most consistent and robust associations across multiple technologies and clinical outcomes. The analyses presented in this manuscript were focused on metrics on the basis of VAF or tumor fraction values, since these were available for studies in this evaluation.

When comparing changes in the mean, median, or maximum VAF values, it was generally observed that the mean and maximum VAF ctDNA values showed similarly strong and consistent univariate associations with different outcomes, whereas median VAF had a weak and inconsistent signal (Data Supplement). One possibility is that median values minimize the impact of large, outlier VAF values that

FIG 4. (A) Forest plot with Cox regression results for PFS and the three-level max VAF percent change group variable, adjusted by baseline clinical covariates. Red means the HR is > 1.0 (increased risk of death/progression) and blue means the HR is < 1.0 (decreased risk of death/progression); unfilled box = nonsignificant P value, filled box = significant $P < .05$. (B) Kaplan-Meier plot for PFS and the three-level max VAF percent change groups, landmarked at 70 days from treatment initiation (the sampling window for the first on-treatment ctDNA sample); patients with an event during the 70-day landmark were excluded from the analysis. ^aDenotes a time-dependent variable. ctDNA, circulating tumor DNA; HR, hazard ratio; LL, lower limit; PD-L1, programmed death ligand-1; PFS, progression-free survival; UL, upper limit; VAF, variant allele frequency.



are clinically meaningful, suggesting that large VAF values may be the most informative when assessing treatment responses. Thus, the single highest somatic VAF value, regardless of the gene and mutation that contributed to the calculation, may be a superior proxy for disease burden, as opposed to other summary measurements that give more weight to rare variants with low VAFs. However, capturing a

single highest variant will be sensitive to the panel used, and a mean VAF may be more robust across tumor types and molecular subtypes, especially those without defined driver mutations.

When comparing the continuous, two-level, and three-level ctDNA metrics, the three-level, and to a lesser

TABLE 2. Multivariate Testing for Association With PR or Better, or PFS at 6 Months
PR or Better (N = 187) PFS at 6 Months (N = 178)

Factor	With Factor	Without Factor	OR (95% CI)	P	With Factor	Without Factor	OR (95% CI)	P
Age ≥ 66, years, No. (%)	28/88 (32)	32/99 (46)	0.85 (0.39 to 1.89)	.696	43/84 (51)	43/94 (46)	0.86 (0.38 to 2.04)	.735
Female, No. (%)	25/70 (36)	35/117 (30)	2.31 (1.00 to 5.31)	.050	35/65 (54)	51/113 (45)	1.84 (0.78 to 4.37)	.165
White, No. (%)	40/122 (33)	20/65 (31)	0.77 (0.35 to 1.71)	.521	64/119 (54)	22/59 (37)	2.12 (0.93 to 4.85)	.075
Ever smoked, No. (%)	57/160 (36)	3/27 (11)	3.23 (0.80 to 13.01)	.100	81/153 (53)	5/25 (20)	2.44 (0.76 to 7.86)	0.134
Advanced stage (stage IV), No. (%)	56/168 (33)	4/19 (21)	3.45 (0.85 to 14.04)	.097	82/160 (51)	4/18 (22)	3.52 (0.86 to 14.49)	.081
Squamous, No. (%)	18/46 (39)	42/141 (30)	2.62 (1.06 to 6.49)	.037	18/42 (43)	68/136 (50)	1.12 (0.45 to 2.81)	.812
PD-L1 3-level, No. (%)								
0%	15/39 (38)		Reference	—	21/38 (55)		Reference	—
1%-49%	14/62 (23)		0.45 (0.15 to 1.31)	.144	24/61 (39)		0.51 (0.17 to 1.55)	.236
≥ 50%	31/86 (36)		0.75 (0.24 to 2.30)	.614	41/79 (52)		0.64 (0.20 to 2.08)	.462
Prior systemic treatment ≥ 1, No. (%)	27/61 (44)	27/61 (44)	0.77 (0.19 to 3.12)	.714	44/121 (36)	42/57 (74)	0.77 (0.20 to 2.95)	.706
3-level max VAF percent change group, No. (%)								
Decrease	34/59 (58)		Reference	—	43/56 (77)		Reference	—
Intermediate	22/95 (23)		0.19 (0.08 to 0.45)	< .001	36/90 (40)		0.13 (0.05 to 0.34)	< .001
Increase	4/33 (12)		0.11 (0.03 to 0.38)	< .001	7/32 (21.8)		0.06 (0.02 to 0.22)	< .001

NOTE. The proportion is calculated as the percent of patients with a given factor, who had the outcome (PR, or better, or PFS, at 6 months). Conversely, among the 99 patients younger than 66 years, 32 patients (or 46%) achieved PR, or better.

Abbreviations: NA, not applicable; OR, odds ratio; PD-L1, programmed death ligand-1; PFS, progression-free survival; PR, partial response; P, P value from Wald χ^2 test in logistic regression; VAF, variant allele frequency.

extent, the two-level ctDNA metric (Data Supplement), consistently showed strong associations with patient outcomes. The continuous ctDNA metric on the basis of the raw percent change value had inconsistent associations with patient outcomes. Modeling the continuous variable was challenging, as the natural range of a percent change calculation (potentially ranging from -100% to +infinity) made data transformations problematic to implement and produced a distribution of the values that resulted in several outliers that could strongly bias a model that assumes a linear association. Conversely, the three-level ctDNA metric grouped extreme and moderate patients (who unquestionably had a substantial change in their ctDNA levels from baseline) and appeared to classify patients into appropriate categories despite potential differences that may exist across ctDNA platforms or clinical situations. Absolute ctDNA values, such as mutant molecules per volume of plasma, were not evaluated because these data were not available for all studies but should be examined in greater depth in future studies. Assessment of overall tumor fraction from plasma data is a field with ongoing development. Incorporating analytical characteristics of specific assays, like limit of detection and precision as well as further improvements on filtering and dynamics of variant VAFs over time, could be hypothesized to further improve predictive power of response assessment. These should continue to be integrated into assessment of molecular response, building off the standardized VAF-based approaches established within the ctMoniTR Project.

Other lines of inquiry include determining how early a change in ctDNA can accurately reflect a patient's response to

treatment, especially if it can reveal tumor responses earlier than radiographic evaluation, and whether baseline ctDNA values are associated with clinical outcomes, as this has been reported previously.^{23,32} In the current study, however, we did not observe an association between baseline ctDNA VAF and clinical outcomes, which could be related to all patients harboring advanced NSCLC or failing a prior line of systemic therapy. Still, our data suggest that ctDNA measurements may help guide treatment decisions, either independently or in conjunction with radiographic evaluation, especially in tumors that are challenging to assess.

Future work from the ctMoniTR Project will expand the scope to include additional tumor types, stages, and drug classes to further validate the association between harmonized ctDNA levels and clinical outcomes in different clinical settings. More specifically, future analyses will focus on better understanding how early changes in ctDNA could be associated with treatment outcomes, and how longitudinal ctDNA measurements can reflect ongoing changes in an actively evolving tumor. Larger data sets will also enable subgroup analyses where relevant covariates can be further investigated. Future efforts will aim to recommend common standards for ctDNA evaluation for use in pharmaceutical trials and clinical practice. Additionally, standardization of ctDNA sampling time points is recommended for future studies, and additional modeling techniques to account for left-truncated data may be considered in future analyses.³⁴ Ongoing work in the ctMoniTR Project will focus on improving measurements and comparability in ctDNA studies, facilitating acceleration in the regulatory adoption of reliable ctDNA measures of responsiveness to treatment, and investigating ctDNA as an intermediate measure of treatment success.

AFFILIATIONS

¹Friends of Cancer Research, Washington, DC

²Cancer Research And Biostatistics (CRAB), Seattle, WA

³NMD Group LLC, Bala Cynwyd, PA

⁴Division of Pulmonary, Allergy and Critical Care Medicine, Thoracic Oncology Group, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

⁵Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD

⁶Translational Medicine, Bristol Myers Squibb, Princeton, NJ

⁷Oncology Center of Excellence, Food and Drug Administration (FDA), Silver Spring, MD

⁸Department of Radiation Oncology, Washington University School of Medicine, St Louis, MO

⁹Department of Genetics, Washington University School of Medicine, St Louis, MO

¹⁰Department of Computer Science and Engineering, Washington University, St Louis, MO

¹¹Siteman Cancer Center, Washington University School of Medicine, St Louis, MO

¹²Guardant Health Inc, Redwood City, CA

¹³Research and Development, Foundation Medicine Inc, Cambridge, MA

¹⁴Oncology Development, Bristol Myers Squibb, Princeton, NJ

¹⁵Late Oncology Statistics, Oncology Biometrics, AstraZeneca, Cambridge, United Kingdom

¹⁶Translational Medicine, Oncology Research & Development, AstraZeneca, Waltham, MA

¹⁷GRAIL, Menlo Park, CA

¹⁸During the conduct of this work and development of the manuscript, N.H. was affiliated with GRAIL, Inc; however, is not affiliated with GRAIL, Inc at the time of submission.

¹⁹Assay Development, Roche Sequencing Solutions, Pleasanton, CA

²⁰Translational Oncology, Early Oncology Statistics, Merck Research Laboratories, Kenilworth, NJ

²¹Companion Diagnostics, Oncology Early Development, Merck Research Laboratories, Kenilworth, NJ

²²Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, Food and Drug Administration (FDA), Silver Spring, MD

²³Foundation Medicine Inc, San Diego, CA

²⁴Division of Molecular Genetics, Office of Health Technology 7 (In Vitro Diagnostics and Radiological Health), Food and Drug Administration (FDA), Silver Spring, MD

²⁵Product Development Oncology, Genentech Inc, South San Francisco, CA

²⁶Product Development Data Sciences, Genentech Inc, South San Francisco, CA

²⁷Division of Hematology and Oncology, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

Changes in ctDNA Reflect Clinical Benefit in NSCLC

CORRESPONDING AUTHOR

Mark D. Stewart, PhD, Friends of Cancer Research, 1800 M St NW, Suite 1050 South, Washington, DC 20036; e-mail: mstewart@focr.org.

DISCLAIMER

This publication reflects the views of the authors and should not be construed to represent FDA's views or policies.

SUPPORT

This work was funded by Friends of Cancer Research.

AUTHOR CONTRIBUTIONS

Conception and design: Diana Merino Vega, Katherine K. Nishimura, Névine Zariffa, Vanessa Cilento, Valsamo Anagnostou, Jonathan Baden, Julia A. Beaver, Darya Chudova, Joseph Fiore, Rachel Hodge, Darren Hodgson, Nathan Hunkapiller, Julie Kobie, Carol Peña, Gene Pennello, Reena Philip, David Raben, Mark Sausen, Ayse Tezcan, Qi Xia, Jing Yi, Mark D. Stewart, Erica L. Carpenter, Charu Aggarwal, Jeff Allen
Financial support: Amanda G. Young

Provision of study materials or patients: Jonathan Baden, Charu Aggarwal

Collection and assembly of data: Diana Merino Vega, Katherine K. Nishimura, Névine Zariffa, Jeffrey C. Thompson, Vanessa Cilento, Valsamo Anagnostou, Julia A. Beaver, Joseph Fiore, Rachel Hodge, Darren Hodgson, Daniel M. Klass, Neil Peterman, Katie J. Quinn, Ayse Tezcan, Qi Xia, Jing Yi, Mark D. Stewart, Charu Aggarwal

Data analysis and interpretation: Diana Merino Vega, Katherine K. Nishimura, Névine Zariffa, Antje Hoering, Vanessa Cilento, Adam Rosenthal, Valsamo Anagnostou, Jonathan Baden, Julia A. Beaver, Aadel A. Chaudhuri, Alexander D. Fine, Joseph Fiore, Rachel Hodge, Darren Hodgson, Carol Peña, Reena Philip, Katie J. Quinn, David Raben, Gary L. Rosner, Mark Sausen, Ayse Tezcan, Qi Xia, Jing Yi, Amanda G. Young, Mark D. Stewart, Erica L. Carpenter, Charu Aggarwal

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

Diana Merino Vega

Employment: AstraZeneca/MedImmune, AllStripes Research Inc

Stock and Other Ownership Interests: AstraZeneca/MedImmune

Travel, Accommodations, Expenses: Aetion

Katherine K. Nishimura

Research Funding: EUSA Pharma (Inst)

Névine Zariffa

Stock and Other Ownership Interests: AstraZeneca, Inovio Pharmaceuticals, GlaxoSmithKline, Pfizer, Moderna Therapeutics, Takeda, Vaxart, VIR Biotechnology, Janssen, Merck, Sanofi, Translate Bio, ANOVA, Intelligencia

Consulting or Advisory Role: Genentech, Bristol Myers Squibb, Cytokinetics, ANOVA, Intelligencia

Other Relationship: ZS Associates

Jeffrey C. Thompson

Consulting or Advisory Role: AstraZeneca

Antje Hoering

Research Funding: eUSA, Hookipa (Inst)

Valsamo Anagnostou

Research Funding: Bristol Myers Squibb (Inst), AstraZeneca (Inst)

Jonathan Baden

Employment: Bristol Myers Squibb, Johnson & Johnson

Stock and Other Ownership Interests: Bristol Myers Squibb, Johnson & Johnson

Aadel A. Chaudhuri

Leadership: Droplet Biosciences, LiquidCell Dx

Stock and Other Ownership Interests: Geneoscopy, Droplet Biosciences, LiquidCell Dx

Honoraria: Roche, Dava Oncology

Consulting or Advisory Role: Geneoscopy, Roche, Tempus, AstraZeneca/Daiichi Sankyo, NuProbe, Alphasights, Guidepoint Global

Research Funding: Tempus

Patents, Royalties, Other Intellectual Property: Patent filings related to cancer biomarkers, and have licensed intellectual property to Droplet Biosciences, Tempus Labs, and Biocognitive Labs

Travel, Accommodations, Expenses: Roche, Droplet Biosciences, LiquidCell Dx

Other Relationship: Roche

Darya Chudova

Employment: Guardant Health

Leadership: Guardant Health

Stock and Other Ownership Interests: Guardant Health

Patents, Royalties, Other Intellectual Property: Patents filed

Alexander D. Fine

Employment: Foundation Medicine

Stock and Other Ownership Interests: Roche

Joseph Fiore

Employment: Bristol Myers Squibb

Stock and Other Ownership Interests: Bristol Myers Squibb

Rachel Hodge

Employment: AstraZeneca

Stock and Other Ownership Interests: AstraZeneca

Darren Hodgson

Employment: AstraZeneca

Stock and Other Ownership Interests: AstraZeneca

Nathan Hunkapiller

Employment: Grailbio, Genentech (I), Pacific Biosciences (I)

Leadership: Grailbio, Pacific Biosciences (I)

Stock and Other Ownership Interests: Grailbio, Genentech (I), Pacific Biosciences (I)

Consulting or Advisory Role: Pacific Biosciences (I)

Patents, Royalties, Other Intellectual Property: Patents generated while employed at Grailbio, patents generated while at Pacific Biosciences (I)

Travel, Accommodations, Expenses: Grailbio, Pacific Biosciences (I), Genentech (I)

Daniel M. Klass

Employment: Roche Molecular Diagnostics

Stock and Other Ownership Interests: Roche

Patents, Royalties, Other Intellectual Property: Coinventor on a patent from Stanford University

Travel, Accommodations, Expenses: Roche Molecular Diagnostics

Julie Kobie

Employment: Merck

Stock and Other Ownership Interests: Merck

Carol Peña

Employment: Merck

Stock and Other Ownership Interests: Merck, Bayer

Patents, Royalties, Other Intellectual Property: Patents from work at Bayer as a Bayer employee

Neil Peterman

Employment: Foundation Medicine, Lexent Bio, Inc

Stock and Other Ownership Interests: Lexent Bio, Inc, Roche

Patents, Royalties, Other Intellectual Property: Several patents filed and pending in the space of liquid biopsy technology, submitted as part of my roles at Lexent Bio and Foundation Medicine

Katie J. Quinn

Employment: Guardant Health

Stock and Other Ownership Interests: Guardant Health

David Raben

Employment: Genentech/Roche, Amgen

Stock and Other Ownership Interests: Genentech, Amgen

Other Relationship: Savor Health, Nimbus Therapeutics

Gary L. Rosner

Stock and Other Ownership Interests: Johnson & Johnson, Walgreens

Honoraria: Novartis

Consulting or Advisory Role: Novartis, Medicago R&D, Inc

Travel, Accommodations, Expenses: Novartis

Other Relationship: Novartis

Mark Sausen

Employment: Personal Genome Diagnostics, Bristol Myers Squibb

Stock and Other Ownership Interests: Bristol Myers Squibb, Personal Genome Diagnostics

Ayse Tezcan

Employment: Foundation Medicine, Lexent Bio, N-Power Medicine

Stock and Other Ownership Interests: Lexent Bio, Foundation Medicine, Foundation Medicine (I) N-Power Medicine, N-Power Medicine (I)

Qi Xia

Employment: Genentech/Roche, AbbVie

Stock and Other Ownership Interests: Genentech/Roche, AbbVie

Jing Yi

Employment: Genentech/Roche, Genentech/Roche (I), Erasca, Inc

Stock and Other Ownership Interests: Roche/Genentech, Roche/Genentech (I), Erasca, Inc

Patents, Royalties, Other Intellectual Property: Genentech, Genentech (I)

Travel, Accommodations, Expenses: Erasca, Inc

Amanda G. Young

Employment: Foundation Medicine, Casma Therapeutics (I)

Stock and Other Ownership Interests: Roche

Erica L. Carpenter

Honoraria: AstraZeneca, Guardant Health

Consulting or Advisory Role: Bristol Myers Squibb

Research Funding: Merck (Inst), Janssen (Inst), Becton Dickinson (Inst), Merck (Inst), United Health Group

Patents, Royalties, Other Intellectual Property: Invention disclosure titled, "Methods and Compositions for Treating Neuroblastoma, Invention disclosure titled, Methods and Compositions for Identifying, Diagnosing and Treating Neuroblastoma"

Travel, Accommodations, Expenses: AstraZeneca, Foundation Medicine

Charu Aggarwal

Consulting or Advisory Role: Genentech, Lilly, Celgene, Merck, AstraZeneca, Blueprint Genetics, Shionogi, Daiichi Sankyo/Astra Zeneca, Sanofi/Regeneron, Eisai, BeiGene, Turning Point Therapeutics, Pfizer, Janssen, Boehringer Ingelheim

Speakers' Bureau: AstraZeneca (I), Roche/Genentech (I)

Research Funding: Genentech/Roche (Inst), Incyte (Inst), MacroGenics (Inst), Merck Sharp & Dohme (Inst), AstraZeneca/MedImmune (Inst)

No other potential conflicts of interest were reported.

REFERENCES

1. Socinski MA, Jotte RM, Cappuzzo F, et al: Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med* 378:2288-2301, 2018
2. Paz-Ares L, Luft A, Vicente D, et al: Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med* 379:2040-2051, 2018
3. West H, McCleod M, Hussein M, et al: Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): A multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 20:924-937, 2019
4. Ramalingam SS, Ciuleanu TE, Pluzanski A, et al: Nivolumab + ipilimumab versus platinum-doublet chemotherapy as first-line treatment for advanced non-small cell lung cancer: Three-year update from CheckMate 227 Part 1. *J Clin Oncol* 38:9500-9500, 2020
5. Reck M, Rodriguez-Abreu D, Robinson AG, et al: Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 375:1823-1833, 2016
6. Gandhi L, Rodriguez-Abreu D, Gadgeel S, et al: Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* 378:2078-2092, 2018
7. Toxicities Associated with Checkpoint Inhibitor Immunotherapy. UpToDate. <https://www.uptodate.com/contents/toxicities-associated-with-checkpoint-inhibitor-immunotherapy>
8. Chiou VL, Burotto M: Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol* 33:3541-3543, 2015
9. Herbreteau G, Vallee A, Charpentier S, et al: Circulating free tumor DNA in non-small cell lung cancer (NSCLC): Clinical application and future perspectives. *J Thorac Dis* 11:S113-S126, 2019
10. Scilla KA, Rolfo C: The role of circulating tumor DNA in lung cancer: Mutational analysis, diagnosis, and surveillance now and into the future. *Curr Treat Options Oncol* 20:61, 2019
11. Douillard JY, Ostoros G, Cobo M, et al: Gefitinib treatment in EGFR mutated caucasian NSCLC: Circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* 9:1345-1353, 2014
12. Kukita Y, Uchida J, Oba S, et al: Quantitative identification of mutant alleles derived from lung cancer in plasma cell-free DNA via anomaly detection using deep sequencing data. *PLoS One* 8:81468, 2013
13. Li X, Ren R, Ren S, et al: Peripheral blood for epidermal growth factor receptor mutation detection in non-small cell lung cancer patients. *Transl Oncol* 7:341-348, 2014
14. Merker JD, Oxnard GR, Compton C, et al: Circulating tumor DNA analysis in patients with cancer: American society of clinical oncology and college of American pathologists joint review. *J Clin Oncol* 36:1631-1641, 2018

15. Trebeschi S, Drago SG, Birkbak NJ, et al: Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 30:998-1004, 2019
 16. Siravegna G, Marsoni S, Siena S, et al: Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* 14:531-548, 2017
 17. Wan JCM, Massie C, Garcia-Corbacho J, et al: Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat Rev Cancer* 17:223-238, 2017
 18. Rossi G, Ignatiadis M: Promises and pitfalls of using liquid biopsy for precision medicine. *Cancer Res* 79:2798-2804, 2019
 19. Moding EJ, Liu Y, Nabet BY, et al: Circulating tumor DNA dynamics predict benefit from consolidation immunotherapy in locally advanced non-small-cell lung cancer. *Nat Cancer* 1:176-183, 2020
 20. Jafarnejad M, Gong C, Gabrielson E, et al: Dynamics of tumor and immune responses during immune checkpoint blockade in non-small cell lung cancer. *Cancer Res* 79:1214, 2019
 21. Davis AA, Iams WT, Chan D, et al: Early assessment of molecular progression and response by whole-genome circulating tumor DNA in advanced solid tumors. *Mol Cancer Ther* 19:1486-1496, 2020
 22. Raja R, Kuziora M, Brohawn PZ, et al: Early reduction in ctDNA predicts survival in patients with lung and bladder cancer treated with durvalumab. *Clin Cancer Res* 24:6212-6222, 2018
 23. Zhang Q, Luo J, Wu S, et al: Prognostic and predictive impact of circulating tumor DNA in patients with advanced cancers treated with immune checkpoint blockade. *Cancer Discov* 10:1842-1853, 2020
 24. Gandara DR, W Zou, J Jiang, et al: An exploratory analysis of on-treatment ctDNA measurement as a potential surrogate for overall survival for atezolizumab benefit in the OAK study. *Ann Oncol* 30:v642-v643, 2019
 25. Thompson JC, Carpenter EL, Silva BA, et al: Serial monitoring of circulating tumor DNA by next-generation gene sequencing as a biomarker of response and survival in patients with advanced NSCLC receiving pembrolizumab-based therapy. *JCO Precis Oncol* 5:510-524, 2021
 26. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 45:228-247, 2009
 27. Rizvi NA, Hellmann MD, Snyder A, et al: Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348:124-128, 2015
 28. LeBlanc M, Crowley J: Survival trees by goodness of split. *J Am Stat Assoc* 88:457-467, 1993
 29. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53:457-481, 1958
 30. Li R, Han D, Shi J, et al: Choosing tumor mutational burden wisely for immunotherapy: A hard road to explore. *Biochim Biophys Acta Rev Cancer* 1874:188420, 2020
 31. Norum J, Nieder C: Tobacco smoking and cessation and PD-L1 inhibitors in non-small cell lung cancer (NSCLC): A review of the literature. *ESMO Open* 3:e000406, 2018
 32. Bratman SV, Yang SYC, lafolla MAJ, et al: Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. *Nat Cancer* 1:873-881, 2020
 33. Goldberg SB, Narayan A, Kole AJ, et al: Early assessment of lung cancer immunotherapy response via circulating tumor DNA. *Clin Cancer Res* 24:1872-1880, 2018
 34. Hosmer DW, Lemeshow S, May S: *Applied Survival Analysis: Regression Modeling of Time to Event Data* (ed 2). Hoboken, NJ, John Wiley & Sons, Inc., 2011. [10.1002/9780470258019](https://doi.org/10.1002/9780470258019)
-

Circulating Tumor DNA in Development of Therapies For Cancer: An Evidentiary Roadmap to an Early Endpoint for Regulatory Decision Making

Approach

Friends of Cancer Research (*Friends*) assembled a multi-stakeholder working group comprised of pharmaceutical companies, diagnostic labs, government health officials, patient advocates, and academic researchers to develop an aligned strategy for generating data and evidence to support using the measurement of *circulating tumor DNA (ctDNA)*¹ levels in patients with solid tumors as a drug development tool in regulatory decision-making. This objective includes using ctDNA as an early endpoint to predict long-term outcomes in patients being treated for early-stage cancer. We thank the numerous stakeholders for their thoughtful input and expertise in the development of this evidentiary roadmap.

Background

Recent technological innovations allow for the detection of ctDNA in the blood. ctDNA is a biomarker with potentially broad clinical and regulatory applicability in oncology:²

- **ctDNA for Patient Selection based on Molecular Alteration**
 - Using molecular features identified in ctDNA to select for patients with alterations targetable by therapy, allowing for the development and evaluation of targeted therapeutic approaches that bestow the most benefit to the patient.
- **ctDNA Molecular Residual Disease for Patient Enrichment**
 - Determining the need for adjuvant therapy after definitive surgery, radiation, or chemoradiation by indicating the presence of *molecular (or minimal) residual disease (MRD)*.
 - Enabling the identification of patients with elevated risk of recurrence for enrollment in clinical studies. This may optimize clinical studies by reducing the overall

- number of trial participants needed, and in turn the time and cost of studies.
- **ctDNA as a Measure of Response**
 - o Detecting a change in the degree or extent of disease burden by serial measurements while on treatment.
 - o Supporting early response/resistance identification (signal finding) in early phase clinical trials to support decision-making in drug development.
- **ctDNA as an Early Endpoint in Clinical Trials**
 - o Detecting a change in the degree or extent of disease burden by serial measurements while on treatment.
 - o Evaluating treatment efficacy and support regulatory decision-making as an early endpoint capable of predicting long-term survival outcomes.

This document outlines considerations and evidentiary needs to support the use of changes in ctDNA levels while on treatment as an early endpoint in clinical trials that predicts long-term clinical outcomes. The term “early endpoint” signifies measuring *ctDNA changes* earlier than other longer-term endpoints (i.e., progression-free survival (PFS), event-free survival, and overall survival (OS)) rather than defining the timeframe of when the endpoint is measured (i.e., not insinuating ctDNA measurement occurs early in a clinical trial, as this may vary based on the context of different cancer types or treatment settings). This roadmap outlines evidentiary needs to support the use of this early endpoint in cancer, including an endpoint to support accelerated approval. To support accelerated approval, the *surrogate endpoint* must be reasonably likely to predict a clinical benefit. The long-term goal of using ctDNA as a validated surrogate endpoint to replace a clinical outcome is a much higher bar of evidence. As we work towards accomplishing this long-term goal, useful information will also be generated to help inform its use as an early endpoint.

The introduction of novel therapeutics, especially targeted therapies, has changed the paradigm for treating solid tumors. While the availability of these new therapies provides increased clinical benefit for patients, the concomitant increase in survival time creates a unique challenge in the development of new therapies. With these novel therapeutics, traditional clinical trial approaches using long-term clinical outcome endpoints such as PFS or OS may not allow for an efficacy determination in a timely manner. Early endpoints that are “reasonably likely to predict a clinical benefit,” are becoming increasingly important in oncology drug development. However, it is critical to obtain adequate data to fully qualify and validate ctDNA as an early endpoint for solid tumors. The use of ctDNA levels (e.g., presence, changes, or clearance) represents an emerging early endpoint that holds great promise.

Aligned methodologies are needed to support robust data generation and enable evaluation of data across trials throughout academia and industry. A recent meta-analysis by *Friends* in non-small cell lung cancer (NSCLC) suggests decreases in ctDNA levels due to therapeutic intervention are associated with improved outcomes.³ Individual trials have also demonstrated this trend. However, evaluating findings from across individual trials can be challenging due to potential impacts of differences in therapeutic modalities investigated, inconsistencies in how ctDNA is collected, measured, and reported, and the variability in the performance of the tests measuring ctDNA (Table 1). This can make it difficult to generalize findings and may not meet the necessary evidentiary threshold for use of ctDNA levels as an endpoint in regulatory decision-making.

Table 1: Sources of Variability in ctDNA Clinical Studies

Clinical Variables	Tumor type, histology, stage of disease
	Definitive therapy type (e.g., surgery, radiation, chemoradiation)
	Therapeutic setting (advanced/recurrent/metastatic, neoadjuvant, adjuvant)
	Current treatment regimens (dosing/timing) and prior regimens
	Therapeutic class (e.g., targeted, IO, cytotoxic, hormonal, etc.)
ctDNA Collection and Methodology	Sample collection timepoints
	Whole blood collection (i.e., tube type, storage, time in transport)
	Plasma sample processing (i.e., centrifugation speeds, double spins, long-term stability)
Captured Endpoints	Endpoints for clinical and radiographic associations, including methodology and definitions of endpoints (e.g., 50% decrease)
	Timing of radiographic surveillance
	Statistical plan (e.g., interim analysis timing, etc.)
Diagnostic Assay and Analysis	Performance parameters (e.g., reference range/interval, Limit of Blank (LOB), Limit of Detection (LOD), accuracy, repeatability, reproducibility, clinical cut-off for molecular residual disease, unit of measurement for ctDNA)
	Biomarker features assessed (e.g., somatic variant mutations, structural variant alterations, methylation, fragmentation, etc.)
	Tumor informed or liquid only platform
	Sequencing platform
	Algorithm design for ctDNA detection and status reporting
	Algorithm design for ctDNA quantification

Source: Adapted from Friends of Cancer Research White Paper: Assessing the Use of ctDNA as an Early Endpoint in Early-Stage Disease.⁴

Clinical Questions that Support the Use of ctDNA Measurement as an Early Endpoint

Generating evidence to support the use of changes in ctDNA levels as an endpoint requires careful consideration of several critical clinical questions.

The primary question is:

- Do changes in ctDNA levels while on treatment predict long-term outcomes (i.e., disease free survival/event free survival (DFS/EFS), overall response rate (ORR), PFS, and/or OS at the individual- and trial-level?

Secondary questions should also be explored to investigate additional nuances:

- Does the predictive value of ctDNA levels vary by:
 - o stage of disease (e.g., early stage, advanced stage)
 - o disease therapy setting (e.g., neoadjuvant, adjuvant)?
 - o therapeutic class (e.g., immunotherapy, chemotherapy, targeted therapy)?
 - o tumor type?
- How does timing of ctDNA measurement impact the predictive value, i.e., should there be set time points for measurement before and throughout treatment for all trials? How do different treatment regimens and cancer types influence timing?
- What is the optimal threshold, in terms of percent change in ctDNA levels (or clearance), that should be used to define ctDNA response? Does this threshold used to define ctDNA response depend on the disease setting (e.g., advanced disease, neoadjuvant) and tumor type?
- How does the depth and durability of ctDNA response (i.e., early response from pre-treatment to on-treatment, maintaining ctDNA response at a landmark on-treatment timepoint) correlate with long-term survival benefit?

Key Considerations for Validating the Use of ctDNA as an Early Endpoint

Understanding the application of ctDNA levels as a biomarker is important when designing studies to validate its use for ascertaining therapeutic efficacy. There are multiple technical and clinical characteristics contributing to variability that should be adequately accounted for, such as assay type, underlying disease, patient heterogeneity, therapeutic context, target of therapy, or a combination of disease parameters, when conducting validation studies.

Technology Considerations

There are a variety of assays that measure ctDNA with differing approaches, which can impact the interpretation of assay results. Assays analyze different molecules, use different technology platforms, and have different methodologies for measuring ctDNA. The variability these differences introduce to ctDNA measurement should be considered when developing evidence that supports the use of ctDNA as an endpoint.

1. *Molecular Alterations Analyzed*

Assays measure different molecules, including genetic alterations and epigenetic modifications. For genetic alterations, there is a tendency to evaluate single nucleotide variants (SNVs), while some assay developers assess insertions, deletions, and/or classic gene fusions. Most ctDNA assays do not account for large structural events or gene copy number variation (CNV), however, some may account for these alterations through low depth whole genome sequencing (WGS) or other more targeted approaches. For assays measuring epigenetic modifications, most focus on measuring methylation or fragment size distribution/DNA fragment patterns. In some cases, assays use a multimodal approach that incorporate genetic alterations and epigenetic modifications.

2. *Platforms*

Different platforms analyze different variants/signatures at varying sensitivities and specificities. Table 2 highlights platforms that measure ctDNA and the opportunities and challenges for each.

3. *Methodological Approaches*

There are two main methodological approaches for identifying ctDNA variants to monitor:

- *Tumor Informed Assays* use individualized sequencing information from a patient's tumor tissue to determine which genes should be monitored in the patient's blood. ddPCR typically focuses on a single genetic mutation that is often the target of the patient's therapeutic treatment. Monitoring a single tumor marker, however, may result in a false negative due to other subclones that can emerge while on treatment. Other assays use a proprietary algorithm to select the optimal variants from the tumor to include in the bespoke panel. Logistical challenges, including the time to develop patient-specific marker panels, require careful consideration when selecting this approach in certain disease settings (e.g., advanced disease).
- *Liquid Only or Tumor-Naïve/Tumor Agnostic Assay* does not require tumor tissue or prior knowledge of a tumor's mutation profile. This approach uses either a pre-determined gene-panel to identify ctDNA variants or a WES/WGS assay. The former approach depends on the panel of genes/methylation loci, which are sometimes selected based on the tumor of origin (e.g., a lung cancer panel). Genes can be analyzed individually or as a signature. WGS and WES can also be used in a tumor-agnostic manner to generate significant coverage across genes, increasing the assay's sensitivity.

4. *Assay Performance*

Assay performance depends on a variety of factors and alignment on metrics will be critical for harmonizing the use of ctDNA as an early endpoint. The BloodPAC Consortium developed recommendations⁵ for 11 required preanalytical attributes to support standards development and robust ctDNA assay development. There are also a series of measurements used to

Table 2: Characteristics of ctDNA Assay Platforms

Platform	Description	Opportunities	Challenges	Sensitivity	Coverage
Droplet Digital PCR (ddPCR)	<ul style="list-style-type: none"> Used for targeted somatic variants 	<ul style="list-style-type: none"> Targeted nature focuses on known variants Smaller volume necessary to measure ctDNA 	<ul style="list-style-type: none"> Not suitable for most fusions because of complexity in junctions or numbers of isoforms generated in splicing Less suited for capturing novel emerging alterations over time 	Highest	Lowest
<u>Fixed size next generation sequencing (NGS) panels</u>	<ul style="list-style-type: none"> Generally, includes targeted NGS somatic nucleotide variant panels, and “methylation only” approaches, or both May also include CNVs and fusions 	<ul style="list-style-type: none"> Can cover a broader range of genomic alterations (~30-650 genes) while controlling size and cost 	<ul style="list-style-type: none"> Sensitivity may vary based on the number of analytes included, GC-rich bias and depth of sequencing 	Moderate	Moderate
WGS/ Whole Exome Sequencing (WES)	<ul style="list-style-type: none"> Algorithms identify ctDNA from whole genome/exome 	<ul style="list-style-type: none"> Coverage advantage, whole genome/exome are covered with a single test 	<ul style="list-style-type: none"> Lower sensitivity and variable specificity for targets 	Lowest	Highest

determine assay performance that need to be harmonized:

- Sensitivity: In general, sensitivity is variable and mostly depends on shedding. There are circumstances where the gene alteration is present in the tumor but no, or very low levels of, ctDNA are detected in the blood.
- Specificity: Specificity depends on the variant targeted. Tumor sequences are unique/specific to the presence of a tumor; however clonal hematopoiesis of indeterminate potential (CHIP) can be mislabeled as a somatic variant and complicate the results if not controlled for. In addition, if germline mutations are not adequately filtered, bioinformatically, or by normal tissue sequencing, the presence of germline alterations may affect the specificity of these assays.
- Accuracy: Greater than 95% accuracy is desirable, but this may be difficult depending on the technologies, panels, target variant frequency, and availability of clinical specimens.
- Precision: Precision should be greater than 95% inter-day, intra-day and inter-instrument, inter-operator.
- Limit of Quantitation (LOQ)/Limit of Detection (LOD): LOQ/LOD is the lowest concentration of analyte that can be consistently detected 95% of the time in a defined type of specimen. LOQ/LOD will be driven by the input level, molecular conversion rate, noise reduction method (e.g., Unique Molecular Identifier), ctDNA input/blood volume, and depth of sequencing and will vary at the specific allele level. Minimum performance characteristics will differ for various platform technologies as well as various providers.

There may be opportunities to use contrived samples to align on approaches for determining assay performance across different assay platforms. However, limitations in certain factors such as number of mutations, chromosome copies, and tissue tumor fractions impact how contrived samples reflect clinical sample performance.

5. Sampling Considerations

Plasma is the default choice for nearly all liquid biopsy applications; however, a few diagnostics developers use serum. Early data shows that serum is more likely to be contaminated by leukocyte DNA and this can impact analyses. Given the need for standardization in this space, we recommend adopting a ubiquitous matrix such as plasma.

6. Standardization Needs

A few additional standardization needs include:

- Measurement outputs may vary across assays including outputs such as variant allele frequency (VAF) and mean tumor molecules per milliliter (mtm/mL).
- A common language to describe epigenetic modifications.

- Statistical considerations, including *baseline* measurement versus change from baseline. Different statistical questions require different data.

Clinical Considerations

The data that will support the use of ctDNA as an early endpoint will depend on the clinical context of use including the cancer type, disease stage, treatment setting, and treatment regimen as these may impact ctDNA kinetics. These components should be considered when developing a framework for evidence generation and designing clinical trials supporting the use of ctDNA as an endpoint.

Likely, validating the use of ctDNA levels as an early endpoint will be a stepwise process, initially validating its use in one tumor type, treatment setting, and drug class where there are strong and existing data and evidence for changes in ctDNA levels anchored to a standard measure of response to treatment in that indication. This approach was seen in FDA's pathological complete response (pCR) guidance⁶, with use of pCR as an endpoint specific to high-risk early-stage breast cancer in the neoadjuvant setting. From a single indication, there may be opportunities to use lessons learned across treatments or tumor types to support use in other settings. For ctDNA levels, it will be important to understand how and when it is feasible to extrapolate findings from one indication to other indications.

1. Current Data Availability

While the long-term goal of the evidentiary roadmap is to support the use of ctDNA levels in early-stage disease, there are lessons to be learned from data that are currently available. Trials are underway to collect data and evidence that support the use of ctDNA levels as a biomarker for treatment response and long-term outcomes in the metastatic setting in multiple tumor types including NSCLC, bladder cancer, colorectal cancer, renal cell carcinoma, and breast cancer. Additionally, trials focused on the use of ctDNA levels to determine MRD may also support an understanding of ctDNA dynamics in different cancer types, especially in earlier stages of disease. This is another active area of research focused in the colorectal and NSCLC settings.

2. Variability Based on the Disease and Therapeutic Characteristics

Identifying clinical characteristics that influence monitoring ctDNA levels in a specific disease setting may support a rationale for indications that could be categorized together and/or prioritized for evidence generation. As noted above, and through *Friends'* previous work, the value of the use of ctDNA levels to monitor treatment response has been demonstrated in the metastatic NSCLC setting, and the indication serves as a use case to apply the characteristics of the indication to other indications (Table 3).

Table 3: Characteristics of Therapeutic Indication Impacting Use of ctDNA as an Endpoint

Characteristic	Considerations	Use Case: Metastatic NSCLC
Biology of Cancer Type	The biology of the specific tumor type, histology, size, vascularity, and location may result in variable ctDNA shed rates impacting the relevance and feasibility of using ctDNA levels as an endpoint.	Shown to have high shed rates in the metastatic setting
Tumor Stage	The stage of the tumor may impact the ctDNA shed rates and early-stage tumors may shed less, impacting detection of ctDNA levels for use as an early endpoint.	
Drug Class	ctDNA kinetics in response to treatment may vary depending on the mechanism of action of the treatment, which may impact the use of ctDNA to compare outcomes across treatment arms.	Ongoing work by ctMoniTR in different drug classes (IO, TKIs)
Treatment Schedule	The treatment schedule may impact the ctDNA shed rates and ctDNA kinetics as well as the timing of ctDNA collection and measurement, therefore a fixed time to sample may not be optimal for each study.	Some drugs for NSCLC are given weekly or every 3-4 weeks in most instances
Definitive Treatment (Early-Stage Setting)	The type of definitive treatment (e.g., surgery, radiation, chemoradiation, radiofrequency ablation) and success of the resection or therapy will alter ctDNA levels and should be considered for establishing baseline ctDNA levels. Additionally, surgery and radiation may impact ctDNA levels differently with and without treatment.	Use of ctDNA is mostly in the advanced setting, where definitive treatment does not occur
Medical Necessity	The utility of other established surrogate endpoints for certain cancers or patients (e.g., patients with unmeasurable disease) may vary by indication necessitating other types of endpoints that can potentially readout sooner or be relevant in a specific patient population.	Safety/risk concerns with repeated tissue biopsies, and toxicities associated with therapies demand a need for a less invasive monitoring tool, such as liquid biopsies

Sampling Considerations

The timing of when to measure ctDNA levels, both initially (baseline) and during follow-up, is not currently standardized, and understanding ctDNA dynamics will require a level of standardization for use as an endpoint. Differences in timing decisions are driven by disease characteristics, treatment regimen and schedule, and assay technology. Additionally, the time to therapeutic response and response durations may vary by types of treatment regimens. Alignment across trials for similar drug classes and cancer types should be considered. For clinical trials that use ctDNA levels as an early endpoint to measure treatment efficacy, sponsors should:

- Ensure impacts on the patient are considered when making decisions about timing for blood draws. Focus on aligning with other clinical activities such as treatment administration and scans.
- Collect a pre-treatment (including pre-surgery) baseline sample. A surgical sample may also be valuable, especially for tumor informed assays.
- Consider timing of imaging, including collecting on-treatment ctDNA samples in parallel with CT/MRI imaging scans to facilitate the exploration of how ctDNA response correlates with accepted measures of clinical response, such as RECIST. However, additional timepoints should be considered, especially before imaging scans, since changes in ctDNA levels may be detected much earlier than disease progression assessed by an imaging scan in early-stage disease.
- Measure ctDNA levels at the end-of-treatment response assessment to fully capture the treatment effect and consider collecting ctDNA during the DFS or OS follow-up period.

Evidence is needed to understand how timepoints impact our understanding of ctDNA dynamics. Key questions to answer include:

- How frequently should samples be collected?
- Does frequency differ based on use case (e.g., tumor type, therapy)?
- How does the frequency of therapy administration impact ctDNA kinetics, and therefore collection?
- How do you define a baseline sample within a specific clinical trial and therapy setting (e.g., neoadjuvant versus adjuvant), and how soon should the sample be drawn (e.g., early-disease setting after definitive surgery)?

Knowledge Gaps and Approaches to Support Evidence Generation

Current data provide a limited understanding of the variability in ctDNA dynamics across different tumor types, tumor stages, treatment regimens, and treatment settings. Meta-analytic approaches will help evaluate and support the use of ctDNA levels as an early endpoint to monitor treatment across various disease settings and provide an opportunity for better alignment in data collection and evidence generation.

Baseline ctDNA Levels Associated with Different Cancer Types

- **Challenge:** Current data in the metastatic setting have shown variable baseline ctDNA shed rates across different cancer types.^{7,8} However, there is not a wealth of data on baseline ctDNA shed rates in the early-stage setting, and data continues to be from disparate ctDNA technologies, making pooled analyses across studies challenging. A better understanding of pre-treatment ctDNA levels across tumor types and stages, assayed with multiple ctDNA assay technologies, would be informative to begin to understand how the biology of the cancer type and stage impacts ctDNA levels.
- **Potential Solution:** Establish evidence regarding baseline ctDNA levels for cancer type and stage across assays through a collaborative effort involving multiple diagnostic developers. These data could support efforts to develop guidance on the use of ctDNA as an early endpoint that can encompass multiple cancers that have shared characteristics, such as shed rates, rather than focusing on individual cancers. Not only would baseline levels support an understanding of the range of ctDNA levels in different stages and tumor types to inform ongoing strategies, comparing different assay outputs will build a foundation and identify key questions to support the harmonization of the endpoint across ctDNA assays. This evidence could also support harmonized metrics for quantifying ctDNA levels, including in early-stage disease.

Association Between Changes in ctDNA Levels and Response to Treatment in Early- and Late-Stage Disease

- **Challenge:** The majority of available data assessing associations between ctDNA changes and patient outcomes to date is in the *late-stage disease* setting. Data is limited in early-stage disease and determining whether associations observed in the advanced setting are generalizable to early-stage disease is needed to help inform its use as an early endpoint in the early-stage setting.
- **Potential Solution:** The ctMoniTR Step 2 Project asks the question: Do changes in ctDNA reflect treatment response in metastatic disease? This project serves as a foundation for an aligned methodology to generate evidence for use of ctDNA to track treatment response in the metastatic setting, but also provides a framework to ask a similar question in early-stage disease to determine whether associations between ctDNA level and patient outcomes varies between these two settings and types of endpoints (e.g., ORR, PFS, OS).

Harmonizing Assays that Measure ctDNA

- **Challenge:** There are different methodologies, technological approaches, and metrics for quantifying and reporting ctDNA levels (e.g., VAF, mtm/mL) across assays, which also impacts the LOD of a given assay. These differences can impact the ability to conduct meta-analyses and could lead to differences in how data are interpreted.
- **Potential Solution:** Evaluating the ability of ctDNA to assess response to treatment using a meta-analytical approach across multiple clinical trials and ctDNA assays, such as in the ongoing ctMoniTR project, can help inform methodological approaches for analyzing data across different assay technologies. This data can also help identify opportunities to support greater alignment across assay quantification, determination of ctDNA “positivity”, and reporting. Comparing associations between ctDNA and other established endpoints, such as RECIST, will help evaluate differences in timing and duration of response and inform what constitutes “meaningful change.” Once established, this could serve as a performance benchmark for determining whether an assay is optimal for detecting clinically meaningful changes in ctDNA levels.

Tumor Specific ctDNA Dynamics

- **Challenge:** The extent to which changes in tumor clonality, and the tracking of specific clones, can have an impact on the interpretation of associations between measured ctDNA levels and clinical outcomes is not fully known. In studies investigating targeted therapies, assays may be tracking molecular alterations of the therapeutic target. An important consideration is whether only tracking those specific mutations impacts the understanding of tumor dynamics as new subclones may emerge during the course of treatment. Comparing results with assays that are not tumor informed may help understand the impacts of subclones.
- **Potential Solution:** For these studies, sponsors should identify the mutations known to be sensitive to the therapy and those that are known to be resistant to the therapy. If using only the target of the therapy as the ctDNA marker, it is important to understand the impact clonality changes may have on interpreting the results of the ctDNA measurements. Tumor samples from patients with late-stage/metastatic disease could be used to assess both tumor-informed and tumor-naïve approaches, where a high disease burden increases the tumor ctDNA and tissue amount for study. Alternatively, a similar analysis could be conducted in specimens collected in the adjuvant setting (i.e., following post-surgery with curative intent), with blood specimens collected pre- and post surgery.

Standards

- **Challenge:** Patient samples, especially from clinical trials, are limited. Some evidence suggests that contrived samples may be challenging to use across assays due to differences in technologies of the assays and that it may be difficult to use contrived samples for evidence development with tumor-informed methodologies.

- **Potential Solution:** A review and discussion of the various contrived samples available would most likely show that some enable more assays to be evaluated than others. There may be an opportunity to test contrived samples across multiple assays to identify the best approach to creating them, if possible. Additionally, assay developers should consider creating a shared resource of retained blinded remnants for ongoing quality control assessments for assay performance.

Regulatory Considerations for Use of ctDNA in Oncology Drug Development

Use of ctDNA as an Early Endpoint for an Accelerated Approval

Validated surrogate endpoints that predict clinical benefit can be used as the primary efficacy endpoint in some clinical trials to replace traditional clinical outcome measures like OS. In oncology drug development, this has helped spur innovation and bring life-saving therapies to patients quickly and safely. When the surrogate endpoint is intended to replace a clinical outcome for the purposes of regular approval, the validity of the surrogate endpoint to predict a clinical benefit in a specific context of use must be established using robust evidence from clinical trials and meta-analytical analyses. However, in the absence of a validated surrogate endpoint, an early endpoint supported by less extensive evidence can support an accelerated approval when the early endpoint is “reasonably likely to predict a clinical benefit.” This would be limited to drugs for serious conditions that fulfill an unmet medical need and would require confirmatory trials to verify benefit using a traditional clinical outcome measure. The use of ctDNA as an early endpoint could be a plausible approach for use in oncology drug development, as more extensive data continues to be generated, which may ultimately result in ctDNA becoming a validated surrogate endpoint to support regular approval. When using ctDNA as an endpoint in a clinical trial, it is necessary to define what is a clinically meaningful change. Use of ctDNA as an early endpoint could be defined in many ways, including as a categorical change from baseline to an on-treatment measurement (e.g., a 50% decrease from baseline), based on the absence of ctDNA (e.g., ctDNA clearance), or including an assessment of duration of the observed changes in ctDNA at a landmark timepoint. How clinical benefit is being defined based on changes in ctDNA levels and what constitutes a clinically meaningful change should be defined a priori. This will likely be dependent on the clinical setting and will also likely vary by treatment, tumor, and assay type will impact use of ctDNA as an endpoint.

Meta-Analytical Approach for Analyzing Data to Validate Use of ctDNA as an Early Endpoint

Meta-analytical approaches merge findings from independent studies to measure an overall effect and have been used to validate other novel early endpoints. Pooled analyses can also increase confidence in observed associations between ctDNA levels and patient outcomes. The terminology and definitions below provide further detail about statistical principles relevant to the validation of an early endpoint:

- *Individual-level association* is the strength of the association between the early endpoint and the true clinical endpoint.

- *Trial-level association* is the strength of the association between the effects of treatment on the early endpoint and the true clinical endpoint.

As outlined in recent FDA guidance documents,⁹ to maximize the interpretability of data aimed at supporting the use of ctDNA as an early endpoint, meta-analytic approaches should include:

- Details of trial designs, inclusion and exclusion criteria, ctDNA assessment, and disease setting as well as justification for the suitability of pooling the studies.
- Trials that include a patient population representative of the population in which the endpoint will be used.
- An adequate number of randomized trials with sufficient follow-up time. The number of trials to be included in the meta-analysis should be justified.
- An analysis based on individual patient level data to allow an assessment of individual level surrogacy.
- Prespecified criteria for concluding association based on both trial-level and individual-level association measurements, including prespecified timing and window of ctDNA assessment. Should explore sensitivity analyses based on different time windows.
- Long-term clinical endpoints, such as DFS/EFS, PFS, and OS, that have been clearly and consistently defined across studies.
- Missing ctDNA assessments and reasons for missing data.
- Sensitivity analyses to demonstrate robustness of the early endpoint and subgroup analyses.
- Statistical handling of unevaluable samples.
- Description of the rate of technical failure (e.g., no ctDNA detectable in sample, especially relevant at baseline).
- Potential confounding factors.
- Comparison of trials using different ctDNA assays and level cutoffs, capturing the assay performance metrics (sensitivity, number of alterations examined, etc.)

Evidence Needed to Support Regulatory Use of ctDNA

While the number of studies demonstrating an association between decreased levels of ctDNA and improved patient outcomes continues to grow, these studies on their own likely do not meet the necessary threshold needed to support the use of ctDNA as a validated surrogate endpoint. Evidence needed to justify use of ctDNA as an early endpoint will depend on the regulatory context, whether it is being used as a supportive or primary endpoint, and other criteria. One approach to accumulate evidence to support the use of ctDNA as an early endpoint is through meta-analytic approaches. In some instances, these meta-analyses could potentially justify the use of ctDNA as an early endpoint in the context of an accelerated approval. Meta-analyses

can be challenging given the variability across trials incorporating ctDNA, including differences in technologies used, disease and therapeutic types, and the schedule of assessment. Collaborative research partnerships, such as the consortium that supported pCR as a surrogate endpoint in breast cancer and the ctMoniTR project for use of ctDNA, are helping elucidate strategies for bringing together disparate datasets and identifying opportunities for alignment to support future meta-analyses across clinical trials.

Glossary of Key Terms

Baseline: The time before a treatment intervention for cancer, to assess ctDNA levels prior to therapy.

Clonal Hematopoiesis of Indeterminate Potential (CHIP): The alteration or mutation of bone marrow stem cells that gives rise to an outgrowth of affected cells.

Circulating tumor DNA (ctDNA): Tumor-derived fragmented DNA shed into a patient's bloodstream that is not associated with cells.¹⁰

ctDNA Changes: A variation or fluctuation in the levels of measured ctDNA over time, most likely a change from pre-treatment levels to on-treatment levels. This may also be a categorical change, such as presence or absence of ctDNA.

Early Endpoint: An endpoint that is reasonably likely to predict long-term clinical outcomes and may be used to support an accelerated approval.

Early-Stage Disease: Specific TNM staging will vary depending on the cancer type but is generally a localized cancer amenable to local intervention with curative intent.

Fixed size next generation sequencing (NGS) panels: A next generation sequencing (NGS) assay that evaluates a pre-specified number and type of alterations in ctDNA.

Late-Stage Disease: Specific TNM staging will vary depending on the cancer type, but is generally a cancer that has metastasized, and is not amenable to local intervention.

Limit of Quantitation (LOQ)/Limit of Detection (LOD): The lowest concentration of analyte that can be consistently detected 95% of the time in a defined type of specimen.

Liquid Only or Tumor-Naïve/Tumor Agnostic Assay: An assay that does not require tumor tissue for creation and use of the assay, and relies on a pre-determined gene panel or whole genome/whole exome sequencing to identify ctDNA variants.

Molecular Residual Disease (MRD): The persistence of a small number of malignant cells, which may be undetectable by conventional screening methods, that is measurable by next generation sequencing of the plasma.

Mean Tumor Molecules per milliliter (mtm/mL): The average number of tumor molecules detected in the ctDNA per milliliter of the patient's plasma.

Tumor Fraction: The proportion of cell-free DNA derived from tumor cells in a blood sample.

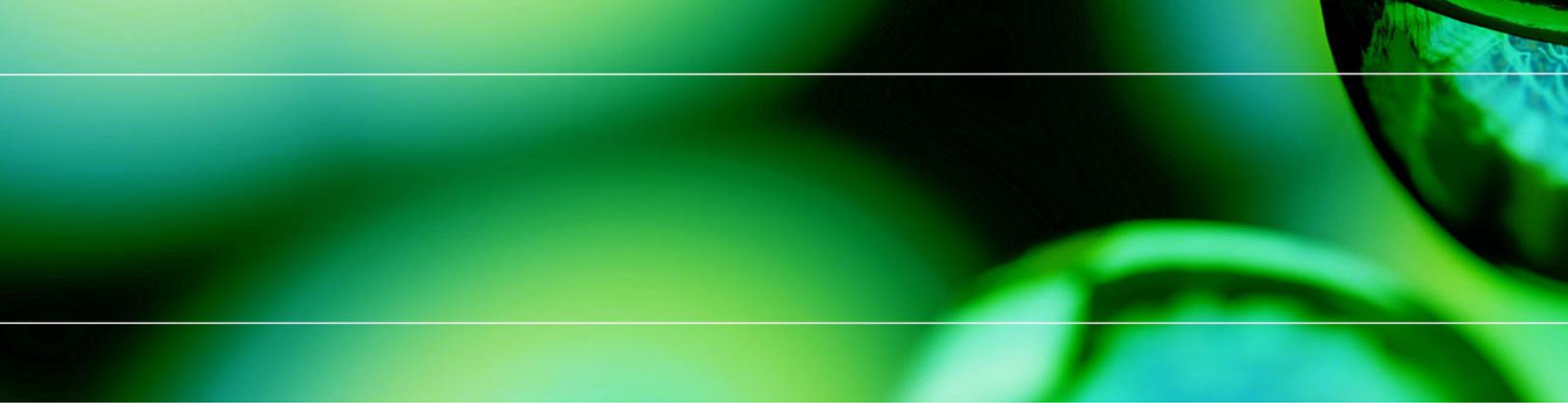
Tumor Informed Assay: An assay that uses a patient's tumor tissue to determine which genes and alterations should be monitored in the blood.

Surrogate Endpoint: An endpoint used in clinical trials that does not directly measure clinical outcomes as the clinical outcomes may take a very long time to study. The surrogate endpoint predicts, or correlates with, clinical benefit and is accepted by the FDA as evidence of benefit and can support a regular approval.

Variant Allele Fraction (VAF): The frequency at which the variant of interest at a specific locus is detected in sequencing reads from a specimen.

References

1. Throughout the document, words that are defined in the appendix are italicized and underlined at first use.
2. FDA. Use of Circulating Tumor DNA for Early Stage Solid Tumor Drug Development Guidance for Industry. May 2022. Accessed on May 20, 2022 from: <https://www.fda.gov/media/158072/download>
3. Vega et al. JCO PO 2022 [In Press].
4. Friends of Cancer Research White Paper: Assessing the Use of ctDNA as an Early Endpoint in Early-Stage Disease. https://friendsofcancerresearch.org/wp-content/uploads/Assessing_Use_of_ctDNA_Early_Endpoint_Early-Stage_Disease-1.pdf
5. Febbo et al. Minimum Technical Data Elements for Liquid Biopsy Data Submitted to Public Databases. CPT 2020.
6. FDA. Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval, Guidance for Industry. July 2020. from: <https://www.fda.gov/media/83507/download>
7. Zill OA, Banks KC, Fairclough SR, et al. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clinical Cancer Research* 2018; 24(15): 3528–3538.
8. Zhang Y, Yao Y, Xu Y, et al. Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nature Communications* 2021; 12(11): 1–14.
9. FDA. Use of Circulating Tumor DNA for Early Stage Solid Tumor Drug Development Guidance for Industry. May 2022. Accessed on May 20, 2022 from: <https://www.fda.gov/media/158072/download>
10. FDA. Use of Circulating Tumor DNA for Early Stage Solid Tumor Drug Development Guidance for Industry. May 2022. Accessed on May 20, 2022 from: <https://www.fda.gov/media/158072/download>

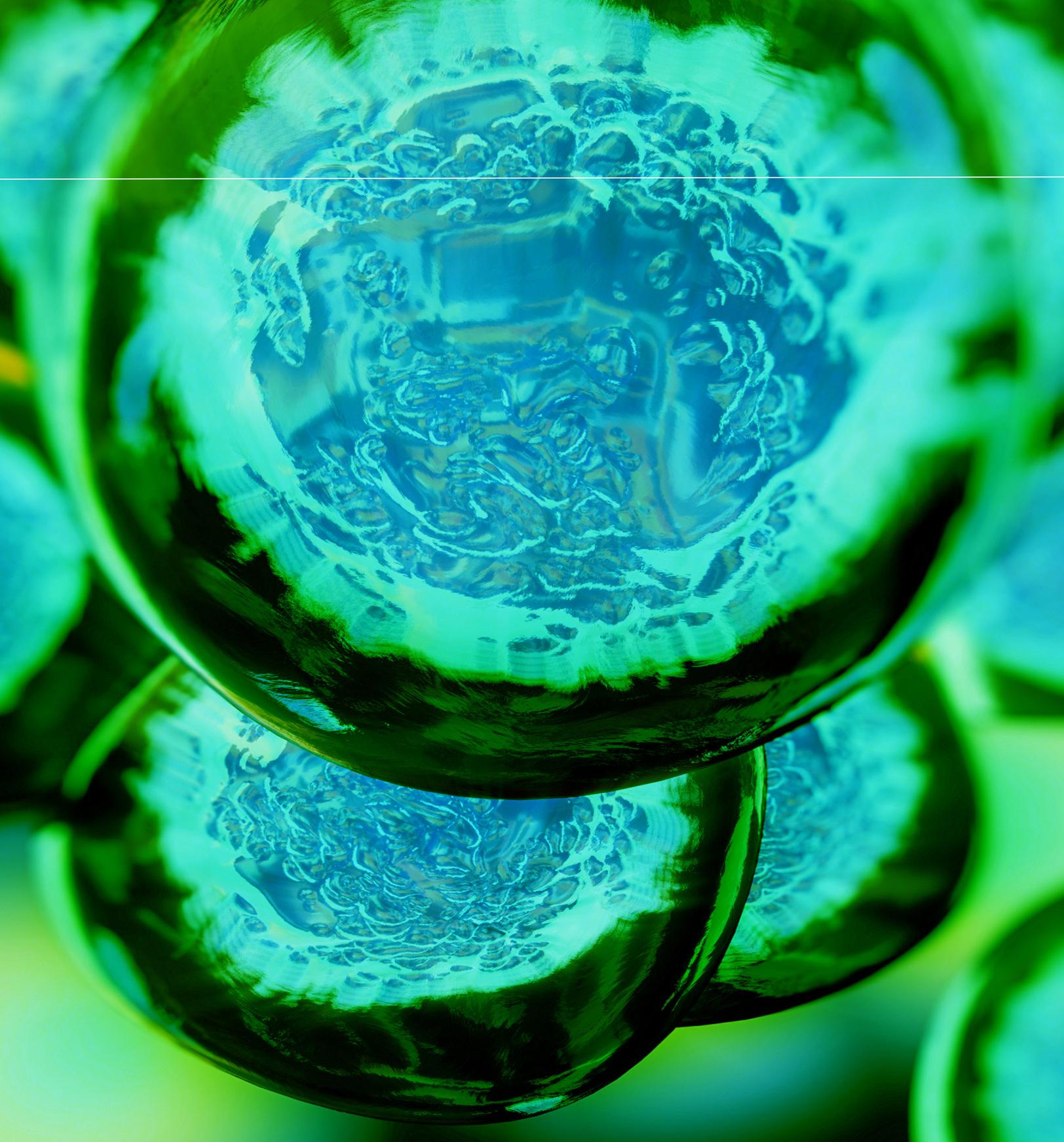




FRIENDS of CANCER RESEARCH

ABOUT FRIENDS OF CANCER RESEARCH

Friends of Cancer Research (*Friends*) works to accelerate policy change, support groundbreaking science, and deliver new therapies to patients quickly and safely. We unite scientists, industry researchers, patient advocates and policy makers with shared trust and guide them toward meaningful cooperation.



1800 M STREET NW SUITE 1050 SOUTH WASHINGTON DC 20036
TEL: 202.944.6700 FRIENDSOFCANCERRESEARCH.ORG

 @CancerResrch

FRIENDS
of **CANCER**
RESEARCH