2025 SCIENTIFIC REPORT

Regulatory Advancements for Patients

FRIENDS of CANCER RESEARCH

Introduction

Friends of Cancer Research (*Friends*) remains a leader in transforming oncology drug development, driving advancements through collaborative and innovative scientific research partnerships and policy initiatives. In 2025, *Friends* forged new partnerships between scientists, advocates, and leading experts to generate evidence-based policy solutions that address critical challenges in oncology drug development and patient care.

This year, evidence generated through our unique partnerships will help drive innovation in drug development and regulatory policy. We expanded our Early Endpoints Portfolio (see project spotlight on pages 10-11) by launching two new projects: the ai.RECIST Project, which evaluates the use of artificial intelligence (AI)-enabled imaging technologies for tumor response assessments; and the Interim OS Project, which explores approaches for interpreting early overall survival (OS) data.

These partnerships—along with the outputs from our working groups, roundtables, and policy research—generate novel insights to inform future policies and deliver scientific advances to patients. This Scientific Report reflects the significant progress achieved through these partnerships in 2025 and serves as a resource for stakeholders in drug development, regulatory policy, and advocacy by offering insights, evidence-based strategies, and collaborative solutions to continue advancing the field of oncology drug development for patients.

As we enter our 30th year, we look forward to growing these partnerships and continuing to work together toward better outcomes for patients.

The 2025 Scientific Report includes the full text of our white papers and publications, which center on four themes:

1. **Modernizing Clinical Trial Designs with an Emphasis on Patient Centricity**
2. **Advancing Diagnostic Tests and AI-Based Tools**
3. **Establishing Early Endpoints for Drug Development**
4. **Expanding Access to Cell and Gene Therapies**

2025 By the Numbers

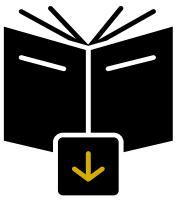
3,000+
meeting attendees

13

working groups

7

public meetings & roundtables



10,000+
publication downloads

500+
working group members

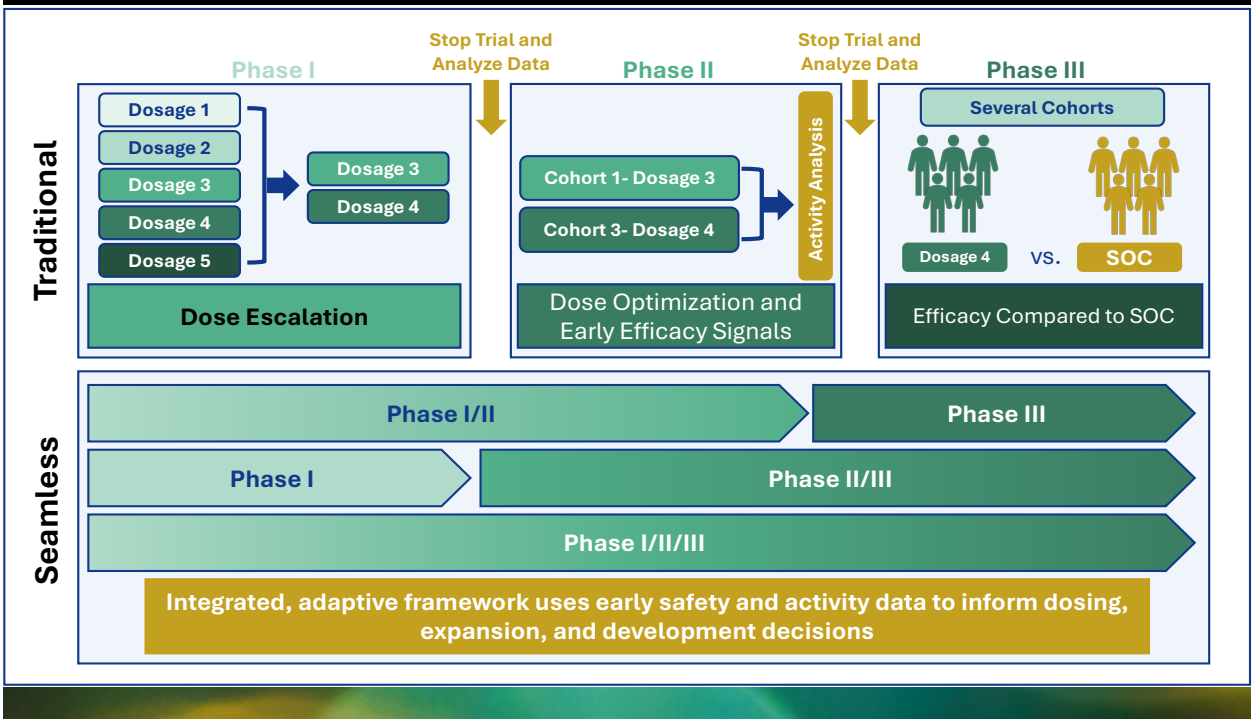
publication citations
500+

1 Modernizing Clinical Trial Designs with an Emphasis on Patient Centricity

As cancer treatments become increasingly complex, adapting clinical trial designs to meet evolving patient and scientific needs is imperative. Emerging approaches—such as the use of external control arms to minimize patients receiving the standard of care (SOC), departures from traditional factorial designs for combination therapy development, and protocol modifications that allow patients to cross-over to experimental arms—reflect a shift toward more flexible, patient-centered models of evidence generation. To ensure trials remain both scientifically rigorous and supportive of patient needs, *Friends* focused 2025 efforts on advancing modern trial design strategies that improve access to care and accelerate therapeutic development, with an emphasis on patient centricity.

Our 2025 Annual Meeting featured three white papers, each focused on adapting or modernizing traditional clinical trial approaches and regulatory frameworks to meet evolving scientific and patient needs. In the first white paper, *Friends* explored the use of seamless trial designs in rare cancer indications, which integrate traditional trial phases into a continuous process to streamline operations and reduce delays. By efficiently leveraging data from every patient across trial phases, seamless designs maximize insights while reducing development timelines, offering the potential to bring therapies to patients with rare cancers faster. In a second white paper, *Friends* addressed trial design challenges in the development of combination therapies, which evaluate multiple agents used together. Recognizing the challenges of isolating the contribution of effect of each agent within traditional trial frameworks, *Friends* examined alternative methodological approaches that can generate robust evidence without requiring prohibitively large or complex trials. These

Trajectory of Seamless Clinical Trials vs. Traditional Clinical Trials



strategies offer a more efficient pathway for the development of promising combination regimens when traditional trial designs are impractical. The final white paper examined the challenge of selecting comparator arms for multi-regional clinical trials (MRCTs), where the SOC may vary significantly across regions. The paper outlined key considerations to ensure comparator arms are both scientifically appropriate and acceptable to patients, particularly within the U.S. context. This work reinforces the importance of aligning trial designs with real-world clinical practice while maintaining global applicability.

To further advance patient-centered approaches to trial conduct, *Friends* collaborated with the American Society of Clinical Oncology (ASCO) to assess the impact of COVID-19-related flexibilities on cancer trials. The joint analysis found that adaptations such as remote monitoring and reduced visit requirements did not adversely affect data quality. These findings support broader implementation of trial designs that reduce patient burden and improve accessibility without compromising scientific standards. Through this body of work, *Friends* continues to support innovative, patient-centered trial designs that enhance access, improve trial efficiency, and accelerate the development of effective cancer therapies.

RELEVANT 2025 POLICY & REGULATORY ACTIVITIES

- **Development of Cancer Drugs for Use in Novel Combination—Determining the Contribution of the Individual Drugs' Effects,** Draft Guidance, July 17, 2025

2 Advancing Diagnostic Tests and AI-Based Tools

Advances in precision oncology are transforming cancer diagnosis and treatment. Most novel treatments approved by the U.S. Food and Drug Administration (FDA) in the past ten years require diagnostic tests to identify patients whose tumors express the relevant indicated biomarker. These tests help match patients to therapies most likely to benefit them, improving outcomes. However, test development and validation remain complex and resource intensive. In 2025, *Friends* advanced efforts to establish flexible, aligned approaches for developing and validating tests to reliably guide patient care.

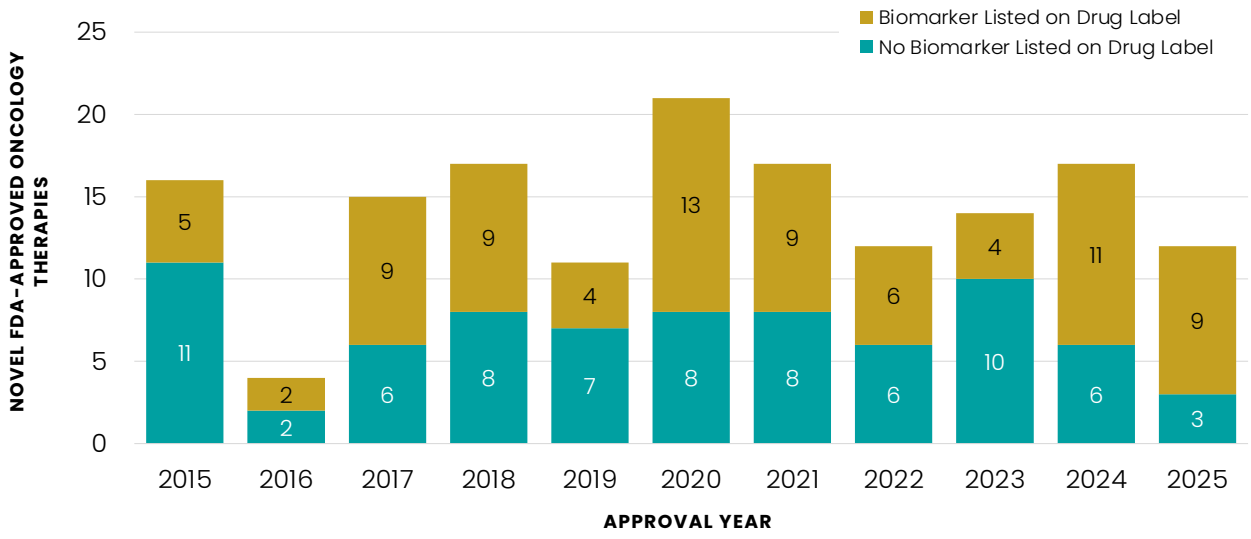
These challenges are amplified for rare biomarkers and cancers, where limited well-characterized samples and reference materials constrain validation efforts. *Friends* conducted an analysis of regulatory documents supporting companion diagnostic (CDx) tests to evaluate how alternative sample sources have been used to validate CDx tests for rare cancer biomarkers. The analysis showed that FDA has been willing to exercise regulatory flexibility to enable efficient development of CDx tests, timely approvals, and patient access—especially in areas of high unmet need. *Friends* also collaboratively developed a white paper with test developers, regulators, and sponsors that details innovative validation strategies and proposes flexible regulatory pathways to support efficient, reliable test development for rare biomarkers.

Beyond regulatory flexibility, computational pathology models powered by AI, including machine learning (ML) models, are emerging to support biomarker discovery and evaluation. These AI-enabled tools can enhance the accuracy, reproducibility, and standardization of prognostic and predictive biomarkers, accelerate diagnosis, and identify novel biomarkers. As AI-enabled tools evolve, there is a growing need for well-characterized reference datasets that can serve as a common basis for benchmarking and independent performance assessment. *Friends* also collaborated with experts from our Digital PATH working group to define considerations for developing independent reference datasets and benchmarks that support reliable and representative evaluations of AI tools. The group developed a discussion document that outlines best practices for designing and applying reference datasets in digital pathology.

RELEVANT 2025 POLICY & REGULATORY ACTIVITIES

- **Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products,** Draft Guidance, January 6, 2025
- **Development of an Artificial Intelligence Action Plan Ensuring Efficient and Innovative Regulatory Approaches for AI in Drug Development,** Request for Information, February 6, 2025
- **Benchmarks for Artificial Intelligence in Cancer Research and Care,** Request for Information, May 14, 2025

Most Novel Oncology Therapies Approved in the Past Decade Are Biomarker-Directed



Source: OncoKB. FDA-Approved Oncology Therapies. oncokb.org/oncology-therapies.

3 Establishing Early Endpoints for Drug Development

As cancer therapies continue to improve, patients are living longer, shifting expectations around clinical trial design and endpoint selection. In oncology trials, OS remains the gold standard for measuring treatment efficacy. However, the extended time required to observe OS outcomes can delay regulatory approval and limit timely patient access to promising therapies. To address this, drug developers are increasingly using surrogate or early endpoints through pathways such as Accelerated Approval. In 2025, *Friends* focused efforts on advancing the development and regulatory acceptance of early endpoints to support more efficient oncology drug development.

One promising early endpoint involves measuring on-treatment changes in circulating tumor DNA (ctDNA), fragments of DNA shed from tumors into the bloodstream. In collaboration with our multistakeholder working group, *Friends* published two papers evaluating ctDNA as a potential early endpoint in advanced non-small cell lung cancer (aNSCLC). The first study demonstrated that patients whose ctDNA levels became non-detected within 10 weeks of treatment had significantly improved OS and progression-free survival (PFS) compared to those with persistent ctDNA detection. The second *Friends* study further explored multiple ctDNA reduction thresholds (50%, 90%, and clearance), various timepoints (<7 weeks and 7 to 13 weeks), and different treatment regimens. Results showed that larger reductions in ctDNA were more strongly associated with improved OS, and that later timepoints may provide more predictive information. These findings underscore the potential of ctDNA as an early endpoint and highlight the need for further research to refine its use in trial design and to evaluate its relevance across additional cancer types.

Since its inception in 2016, **the BQP has accepted**

61 Projects

however, only 8 biomarkers have been fully qualified

In a complementary initiative, *Friends* assessed the FDA's Biomarker Qualification Program (BQP), a critical pathway for enabling the regulatory use of novel biomarkers. Since its inception in 2016, the BQP has accepted 61 projects; however, only 8 biomarkers have been fully qualified. The *Friends* analysis highlighted that none of the qualified biomarkers were developed as surrogate endpoints and called for a dedicated framework to support the development of novel response biomarkers. Through these efforts, *Friends* continues to advance innovative regulatory strategies to accelerate evidence generation and support earlier patient access to safe and effective therapies.

RELEVANT 2025 POLICY & REGULATORY ACTIVITIES

- **Approaches to Assessment of Overall Survival in Oncology Clinical Trials,** Draft Guidance, August 18, 2025

4 Expanding Access to Cell and Gene Therapies

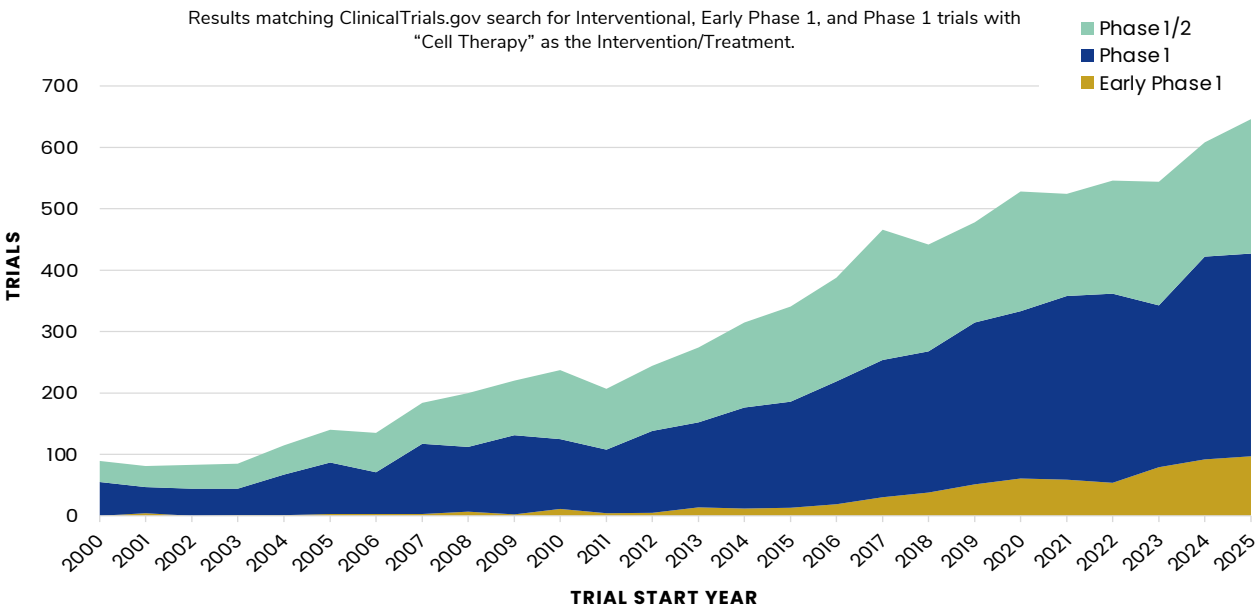
Genetically modified cell-based therapies, such as chimeric antigen receptor (CAR) T-cell and T-cell receptor (TCR)-based approaches, represent transformative scientific advances, offering some patients curative treatment options. These advances have been particularly impactful for those with rare and hard-to-treat diseases; however, sustaining and expanding access remains challenging as traditional clinical development, manufacturing, and reimbursement models have not fully evolved to support delivery of these highly personalized, complex therapies.

In 2025, *Friends* convened stakeholders to examine how collaborative, adaptable, fit-for-purpose approaches can address these barriers and ensure that advances in cell-based therapies reach patients. This work culminated in a white paper that was discussed during a May 2025 public meeting hosted in partnership with the Parker Institute for Cancer Immunotherapy (PICI). The white paper explored decentralized or point-of-care manufacturing models, risk-adjusted development strategies, and public-private cost-recovery partnerships as potential pathways to ensure that cell-based therapies with compelling early evidence are available to patients. Together, these efforts underscore the need for systemic, multi-stakeholder collaboration to create pathways that enable safe, timely, and equitable access to cell- and gene-based therapies—ensuring that scientific innovation translates into tangible benefits for patients.

RELEVANT 2025 POLICY & REGULATORY ACTIVITIES

- **Leveraging Knowledge for Facilitating the Development and Review of Cell and Gene Therapies,** Public Listening Meeting & Request for Information, July 25, 2025
- **Innovative Designs for Clinical Trials of Cellular and Gene Therapy Products in Small Populations,** Draft Guidance, September 25, 2025

An Increasing Number of Early Phase Investigational Cell Therapy Trials Are Initiated Each Year



Early Endpoints Portfolio

Goal

Friends Early Endpoints Portfolio generates evidence to support the development and use of novel endpoints and technologies that enable efficient assessments of treatment efficacy. These efforts aim to facilitate timely clinical trials, regulatory decision-making, and ultimately, patient access to life-saving therapies.

Background

Scientific innovation and therapeutic advancements over the past two decades have helped patients with cancer live longer. While this progress is positive, it also means that clinical trials may take longer to generate readouts to support FDA approval of the next generation of drugs. As a result, efficacy assessments, regulatory approvals, and patient access to new therapies may take longer while data mature, highlighting the need for reliable early endpoints to provide timely insights into treatment benefit.

Early endpoints that are reasonably likely to predict overall survival (OS) such as overall response rate and progression free survival have been incorporated in clinical trials, enabling expedited approval through the Accelerated Approval Pathway for promising therapies in areas of unmet need. Other innovative approaches to assessing treatment efficacy—including evaluating changes in circulating tumor DNA (ctDNA) and employing artificial intelligence (AI)-based technologies for tumor assessments—are emerging methods that may provide even earlier results and be more accurate predictors of benefit than current tools. These novel tools hold promise for facilitating efficient efficacy assessments, but robust evidence is needed to demonstrate their value, establish their reliability, and validate their use for regulatory decision-making.

Approach

Friends' collaborative, data-driven approach generates the evidence and frameworks necessary to support reliable use of early endpoints and novel tools for regulatory decision-making through several key projects (findings will be consolidated and presented in public meetings and peer-reviewed literature):

ai.RECIST Project – AI-driven tumor measurement tools have the potential to reduce variability, increase efficiency, and improve the accuracy of imaging-based endpoint assessments in clinical trials. The ai.RECIST Project evaluates the consistency and reliability of AI-driven tumor measurements across AI tools and compared with human readers. This project provides insights on the use of these tools for streamlining tumor response assessments in clinical trials.

ctMoniTR Project – ctDNA holds promise for measuring treatment efficacy in clinical trials. *Friends* is leading a multi-stakeholder group to develop an aligned strategy and generate the data needed to support the use of ctDNA as an early endpoint for treatment response in regulatory decision-making. Validating ctDNA as an endpoint will accelerate identification of effective therapies, enabling them to reach patients sooner.

Interim OS Project – Interim OS analyses are conducted at the time when an early endpoint is evaluated and they can help assess potential harm. However, they require careful planning to ensure accurate interpretation. *Friends* is using computational modeling to develop best practices for interpreting interim OS data in clinical trials and regulatory decision-making.

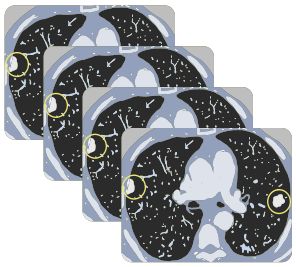
ai.RECIST Project Spotlight

Background

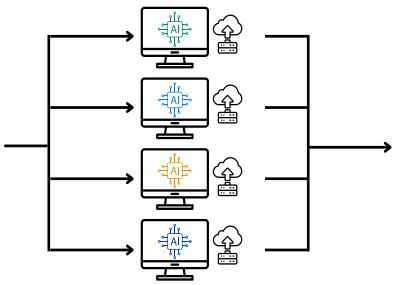
While the current approaches to tumor assessment in clinical trials rely on radiologists (i.e., human readers) applying standardized criteria (i.e., Response Evaluation Criteria in Solid Tumors [RECIST]), this process is time-intensive, subject to investigator bias, and variable across readers and sites. Additionally, associations with long-term outcomes are not always clear. Incorporating artificial intelligence (AI) to measure tumors offers the potential for more efficient, reproducible, and objective assessments of treatment response in clinical trials. To realize this potential, it is important to understand capabilities and consistency of current AI tools supporting the efforts to determine their reliability and generalizability, and paving the way for thoughtful integration into clinical trials.

Approach

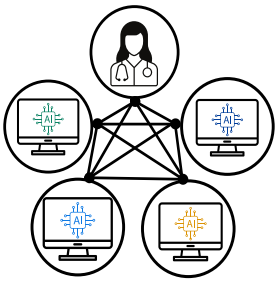
Friends has convened a multistakeholder working group to evaluate reproducibility and agreement among AI-based tumor measurements and as compared with human readers.



Identify and share a common dataset of CT scans with clinical metadata



Tool developers apply independently developed AI models to assess tumor lesions



Statisticians evaluate variability by comparing tumor assessments across AI models and with human readers

Impact

AI-driven tumor measurement tools have the potential to reduce variability, increase efficiency, and improve measurement accuracy in clinical trials, enabling effective therapies to be reliably identified. Understanding the variability of tools can help determine their generalizability and supports aligned approaches for tracking tumor growth over time. Findings from this project will inform regulatory and policy frameworks for AI use in clinical trials and contribute to the broader effort to advance early endpoint development.

Project Workflow

Identify Problem: Current approaches are resource-intensive and variable, which can extend trial timelines.

Current approaches to tumor measurement rely on human radiologists using RECIST v1.1, which has challenges with investigator bias, subjectivity in lesion selection, and variability in measurements across clinical sites and radiologists.

Identify Solutions: Can AI-based imaging tools improve tumor measurements in clinical trials?

Incorporating AI tools into tumor measurements could suppor t more efficient and effective clinical trials. **How these tools identify and report tumor measurements has not been systematically evaluated.**

Align on an approach to implement solutions

Establish a multi-stakeholder working group to establish an approach for comparing outputs from AI tools that provide tumor measurements. Align on a common statistical analysis plan and dataset for analysis.

Provide AI tools with a common set of images & assess variability

Distribute images from a clinical trial of patients with advanced non-small cell lung cancer (aNSCLC) treated with tyrosine kinase inhibitors.

Take Action on Findings

Policy Opportunities

Inform regulatory and legislative efforts focused on:

- Early endpoint development
- Frameworks for AI use in clinical trials

Data Readouts

Initial data shared during a *Friends* public meeting in 2026, subsequent outputs shared at scientific conferences and in peer-reviewed literature.

Modernizing Clinical Trial Designs with an Emphasis on Patient Centricity

FRIENDS of CANCER RESEARCH

Seamless Clinical Trial Designs in Rare Cancers: Leveraging Operational and Adaptive Strategies to Accelerate Drug Development

White Paper | 2025

Executive Summary

Rare cancers pose unique challenges for drug development. Small, heterogeneous patient populations can limit the feasibility of traditional randomized controlled trials, slowing evidence generation and resulting in delayed access to potentially life-saving therapies. Sequential evaluation of safety, dosage optimization, and efficacy in distinct phases can be slow and resource-intensive, creating inefficiencies in rare cancer development. More deliberate integration of these stages within a seamless framework can maximize learning from each patient, improve operational efficiency, and accelerate evidence generation. Additionally, development programs consisting of multiple distinct sequential clinical trials may not provide the ability to optimally leverage data from each patient, which is particularly crucial for rare cancer product development.

Seamless clinical trials build on common early-phase approaches—such as dose escalation and cohort expansion—by more thoughtfully integrating multiple development stages, including dosage optimization and efficacy evaluation, within a single framework. This approach can reduce downtime between phases, maximize learning from each patient, and allow adaptive modifications to the trial based on emerging data. Despite these advantages, seamless trial designs remain underutilized in rare cancer drug development.

To address these challenges, Friends of Cancer Research (*Friends*) convened a multi-stakeholder working group including experts from patient advocacy organizations, pharmaceutical companies, academia, National Cancer Institute (NCI) and the U.S. Food and Drug Administration (FDA) to identify critical design considerations and explore strategies for operationalizing these efficiencies offered by the seamless design framework in rare cancer development. The group explored several strategies and considerations:

- Seamless trial designs: These approaches integrate multiple development stages under a single framework, with inferential designs pooling data across stages for integrated analyses. Selecting the appropriate approach depends on objectives, patient population, and endpoints.
    - Operational considerations: Seamless designs can enable faster transitions from dose escalation to expansion, incorporate data from early-phase patients into later analyses, and can embed randomization, maximizing learning from each patient.
- Endpoints and adaptive features: Early, meaningful endpoints maximize the insights gained from each patient and can inform pre-specified adaptations, such as adjusting dose levels, expanding promising cohorts, or refining eligibility criteria. These adaptations occur according to predefined rules based on interim analyses, ensuring that decisions are guided by accumulating data while maintaining statistical rigor and trial integrity.
- Regulatory and patient engagement: Early and ongoing dialogue with regulatory authorities clarifies expectations around use of novel endpoints, adaptive features, and integrated analyses. Engaging patient advocates ensures trial designs reflect patient priorities, tolerability considerations, and operational feasibility, particularly in rare disease settings.

These insights highlight that seamless trials require careful planning, adaptive design, and close coordination with stakeholders to balance scientific rigor, operational feasibility, and patient benefit. By thoughtfully implementing these approaches, sponsors can accelerate patient access to new therapies, maximize evidence generation from limited populations, ensure safety, and provide a flexible yet rigorous framework for rare cancer drug development.

Authors

Harindra Abeysinghe, Johnson & Johnson

Kalyanee Viraswami- Appanna, EMD Serono Inc

Cong Chen, Merck & Co., Inc.

Alice Chen, National Cancer Institute (NCI)

YounJeong Choi, Genentech, Inc

David Driver, GSK

Lola Fashoyin-Aje, Parexel International

Elizabeth Garrett-Mayer, American Society of Clinical Oncology (ASCO)

Paul Haluska, Bayer AG

Colin Kielty, Patient Advocate

Jackie Kline, Amgen, Inc.

Eric Kowack, Xencor, Inc.

Misha Mehta, *Friends* Advisory Advocate

Allen S. Melemed, Chimerix, A Jazz Pharmaceuticals Company

Nassim Morsli, AstraZeneca

Bernat Navarro-Serer, Friends of Cancer Research

Wendye Robbins, Incendia Therapeutics

Zhen Su, Marengo Therapeutics

Vivek Subbiah, Sarah Cannon Research Institute

Christopher D. Turner, Nuvalent, Inc.

Chris White, *Friends* Advisory Advocate

Ying Yuan, MD Anderson Cancer Center

This white paper was developed through discussions that included these experts and other perspectives representing academia, industry, the U.S. Food and Drug Administration, and the patient advocacy community. The views expressed here represent the collective insights from working group discussions and do not necessarily reflect the official positions of any individual organization.

Table of Contents

Introduction

Challenges of Rare Cancer Drug Development

Developing new therapies for rare cancers presents complex challenges that distinguish these diseases from more common malignancies. In many cases, the small number of eligible patients globally limits the feasibility of traditional randomized controlled trials (RCTs) and slows evidence generation. This rarity, combined with heterogeneity in disease biology, limited natural history data, low-prevalence biomarker-defined populations, and inadequate or absent standards of care (SOC), constrains trial enrollment, complicates trial design, and can reduce the generalizability of results. Additionally, limited financial incentives and constrained resources for conducting trials make efficient use of available patients and data even more critical. In rare cancers, inefficient trial designs or delays can have an outsized impact—slowing patient access, limiting evidence generation, and potentially missing opportunities to identify effective therapies.

Although challenges with traditional phased development are not unique to rare diseases, they are often magnified when patient numbers are small. Generally, traditional phased development starts with dose-finding studies and may be followed by registrational trials and confirmatory trials, which are often resource-intensive and time-consuming. Pauses between phases for protocol development and site activation may introduce delays that hinder patient access, risk losing the momentum of site engagement and patient recruitment, and limit efficient use of trial data across stages. Although randomization can be challenging to implement broadly in rare cancers due to small patient populations and often, because of the limitations of SOC which precludes clinical equipoise, it can be informative in certain context, such as when multiple dosages are under evaluation or within platform trials.

The significant unmet clinical need in rare cancer populations demands more agile approaches to drug development that maintain scientific rigor while improving efficiency. Patients with rare cancers are often enrolled in early-phase trials with broad study population (i.e., all solid tumors). As a result, the data are often fragmented, and hypothesis tests are underpowered, missing opportunities to generate meaningful evidence. Additionally, foundational knowledge about the patient population, treatment patterns, and clinical outcomes should be generated alongside the clinical trial itself. These realities highlight the need for trial designs that are efficient, representative of the intended population, and adaptable—maximizing insights from every enrolled patient while maintaining the robustness and reliability of the evidence.

Regulatory and Scientific Trends

Over the past decade, the U.S. Food and Drug Administration (FDA) has issued guidance documents and demonstrated openness toward flexible strategies to generate evidence in settings of unmet medical need, including rare cancers.[1–4] Sponsors have often employed single-arm designs, intermediate clinical endpoints such as objective response, and adaptive trial designs in appropriate contexts. While early-phase oncology studies often include dose-finding and cohort expansion, more deliberate integration of additional seamless elements—such as adaptive cohort transitions or combined efficacy and safety assessments—remains limited in rare cancer development. Shared experience and best-practice frameworks for implementing these approaches are still sparse, creating uncertainty for sponsors and regulators. Broader adoption of seamless approaches in rare cancer development could enhance data continuity, maximize scientific and clinical insights, and accelerate access to promising therapies.

Scope

This white paper describes seamless clinical trials and design elements tailored to the challenges of rare cancers. These designs may integrate early-stage clinical evaluation with potential registrational intent into a single protocol, or they may be applied to specific portions of clinical development to enable more efficient evidence generation.

We describe what constitutes a seamless trial in this context, identify key considerations in trial structure and implementation, and outline specific use cases to illustrate when and how these strategies may be applied. The main objective is to provide a practical foundation to inform thoughtful, feasible, and scientifically robust trial designs that improve continuity, efficiency, and impact in rare cancer drug development.

Understanding Seamless Trials: Definitions, Benefits, and Relevance to Rare Cancers

Seamless clinical trials can accelerate and streamline drug development by integrating multiple trial stages within a single, continuous protocol. This approach can promote more efficient evidence generation, minimize delays between phases, and support real-time decision-making based on emerging data. For the purposes of this paper, we define a seamless clinical trial as:

"A clinical trial integrating multiple, sequential stages of drug development—such as dose escalation, dosage optimization, cohort expansion, and efficacy assessment—within a single framework."

By consolidating these stages, seamless trials can enhance efficiency, ensure consistency, and maximize the scientific and evidentiary value of each patient's participation. Seamless designs may incorporate pre-specified adaptive features, allowing modifications to aspects such as sample size, dosage optimization, or expansion criteria based on emerging data, provided these adaptations are pre-specified and carefully justified. Seamless designs vary in scope and complexity and can be broadly categorized as operationally or inferentially seamless, as summarized in Table 1.

Table 1. Seamless Trial Types, Key Features, and Benefits Relative to Traditional Single-Phase Trials

Approach	Description	Key Features	Purpose & Benefits
Operationally Seamless	Continuous trial conduct across multiple stages within a single protocol	Single continuous protocol; minimal enrollment gaps; early data informs later decisions; streamlined enrollment and data collection	Reduces delays, maintains trial momentum, and maximizes insights from each patient, addressing the primary challenges of rare cancers
Inferentially Seamless	Data from multiple stages are pooled and analyzed together to support unified conclusions	Combined analysis; integrated statistical plan	Enhances statistical efficiency, reduces sample size requirements, and supports cohesive decision-making across trial stages

In practice, all inferentially seamless trials are operationally seamless, but not all operationally seamless trials include inferential pooling. Operationally seamless designs may incorporate multiple expansion

cohorts, including those with potential registrational intent, and are often preferred in rare cancer settings when early-stage and later-stage endpoints differ, or when the limited scientific understanding of a novel agent warrants independent analyses—such as initiating with a dose-finding stage in a mixed tumor cohort before moving to a histology-specific expansion phase with different efficacy measures. By contrast, inferentially seamless features are most valuable when eligibility criteria, endpoints, and trial populations remain consistent across stages, allowing data to be combined—such as in a rare cancer trial where both the exploratory and confirmatory stages assess the same objective response endpoint in the same patient population.

Selecting the appropriate seamless features enables sponsors to optimize scarce patient resources, maintain statistical validity, and meet regulatory expectations while accelerating development in areas of unmet need.

Why Are Seamless Trials Important for Rare Cancers?

Seamless trials are particularly valuable for rare cancers because they help to:

- **Minimize redundancy and enrollment delays:** Avoiding separate protocols and site start-up processes helps preserve momentum, which is particularly important in rare cancers where recruitment is difficult, and patients may only be eligible for a single trial.
- **Reduce patient exposure to ineffective therapies:** By integrating early indicators of activity, seamless designs can identify ineffective interventions sooner, placing patient well-being at the center of trial efficiency.
- **Enable real-time learning and adaptation:** Seamless designs can accommodate multiple objectives from early and later stages, such as moving from early phase questions about dosage to objectives aimed at registration based on emerging signals.
- **Maximize insights per patient enrolled:** Given the limited number of eligible patients, integrating data across trial stages ensures that no clinical evidence is lost and data collected in early stages can inform later decisions and inferences.
- **Support intermediate clinical endpoints or early measures of activity:** Seamless trials allow incorporation of early signals to guide trial progress and inform dosage or cohort decisions.
- **Leverage regulatory flexibility and enhance efficient evidence:** Recent published FDA guidance reflects openness to well-justified, innovative trial designs.[1–4] Seamless trials may align well with approval pathways when thoughtfully planned and appropriately justified. For example, FDA OCE's Project FrontRunner highlights opportunities for using a seamless randomized approach to generate evidence for accelerated approval and verify clinical benefit for subsequent traditional approval in the front-line advanced/metastatic setting.[5]
- **Facilitate faster patient access to promising therapies:** By aiming to reduce pauses between phases and integrate registrational intent earlier, seamless trials can shorten timelines and provide patients with earlier access to potentially effective therapies.

Seamless design challenges relating to operational complexity, statistical considerations, and regulatory planning must be carefully managed to reduce bias impact and maintain trial integrity and interpretability. Their use in rare cancers must be grounded in both flexibility and rigor, balancing efficiency with meaningful

Challenges in Seamless Trials for Rare Cancer Drug Development

Despite the potential advantages of seamless trial designs, rare cancer development presents inherent challenges that can complicate trial approaches. These hurdles can make seamless designs more difficult to implement effectively. Table 2 summarizes the key challenges identified in both rare cancer and seamless trial contexts. Taken together, these challenges show that while seamless trials may reduce redundancy, they require deliberate planning to ensure interpretability and patient benefit.

Table 2. Key Challenges in Implementing Seamless Trials for Rare Cancer Drug Development

Scientific and Statistical Considerations	• Fully integrated seamless trial designs remain uncommon in rare cancer development. • Novel endpoints or adaptive rules require careful characterization. • Small populations amplify statistical uncertainty; traditional p-value frameworks may be underpowered, requiring alternative approaches
Operational Complexity	• Challenges coordinating early-phase developers with rare cancer disease experts. • Trials enrolling multiple rare cancer subtypes/tumors require coordination across investigators and institutions to ensure adequate representation and consistency. • Multi-regional trials face divergent regulations, SOC, and data collection requirements, which can disrupt enrollment and trial continuity.
Endpoint Selection and Evidence Generation	• Endpoints (ORR or DOR (e.g., objective response or duration of response)) for tumor activity in early phases can support go/no-go decisions; but their predictive value for long-term clinical benefit in rare cancers is not always typically established • Reliance on external or historical control data to inform go/no-go decisions can be challenging and unreliable in rapidly evolving treatment landscapes. • Traditional endpoints (e.g., OS, PFS) may require long follow-up, be underpowered and/or large sample sizes, or may be insufficient to capture other similarly meaningful aspects of patient experience; or may be impractical as patients transition through multiple therapies.
Dosage Optimization and Safety	• Limited prior clinical data and small patient cohorts complicate dosage optimization; may preclude extensive dosage optimization and safety monitoring. • Adaptive rules for dosage or cohort modifications must balance pre-specification with flexibility. • Higher risk of suboptimal dosage or missed safety signals compared with common cancers.

Abbreviations: SOC, standard of care; OS, overall survival; PFS, progression free survival.

Types of Seamless Approaches (Case Studies & Scenario Based)

Seamless trial strategies should be adapted to the unique context of each development program. Key factors such as disease rarity, heterogeneity of the patient population, robustness of understanding of the disease's natural history, prior knowledge of the therapeutic target, and regulatory goals shape both the feasibility of a seamless design and its implementation.

The following case studies illustrate a range of approaches: (1) early-phase dose-finding trials with efficacy signals; (2) seamless trials with registrational intent; and (3) seamless trials with randomized components. Together, they illustrate both the potential of seamless strategies to accelerate development in rare cancers and the operational challenges that can slow progress.

Early-Phase Dose-Finding Trials with Efficacy Signals

FIGHT-101

FIGHT-101 (NCT02393248) was a first-in-human (FiH) trial of pemigatinib, an FGFR inhibitor, in patients with advanced solid tumors (**Table 3**).[6] The trial progressed operationally seamlessly from dose-escalation into expansion cohorts that included rare cancers such as cholangiocarcinoma. Key operational considerations included balancing pharmacokinetic and broader drug development expertise alongside disease-specific insights, site selection, and protocol flexibility.

This trial exemplifies how seamless trial designs can inform subsequent registrational studies. Insights from FIGHT-101 supported the design of FIGHT-202, a dedicated study of pemigatinib monotherapy in FGFR-altered cholangiocarcinoma,[7] which ultimately supported FDA approval.[8,9]

Key Insight: Seamless early-phase trials can accelerate dose finding and provide early activity signals in rare cancers, but their success depends on aligning with experienced investigators in FiH trial conduct as well as early activity assessment, engaging trial sites equipped to manage complex protocols, and anticipating the operational trade-offs. They can generate critical pharmacology and safety data that can inform later trial decisions, and proactive regulatory engagement—through Investigational New Drug submissions, pre-New Drug Application meetings, and targeted feedback on dosing or safety questions—can help guide registrational planning. When combined with prior disease knowledge and clear regulatory engagement, seamless designs can set the stage for registrational trials and expand treatment options in areas of unmet need.

Table 3. Key Features of FIGHT-101 Trial

Key design, operational, and patient population features

Feature	Details
Population / Rarity	• Advanced, refractory malignancies with or without FGF/FGFR alterations, including NSCLC, cholangiocarcinoma, urothelial, pancreatic, head and neck, and other solid tumors ○ Alterations are uncommon (generally ~8–15% in urothelial and cholangiocarcinoma, and <5% in NSCLC, pancreatic, and head and neck cancers). • Rare molecular subtypes; later cohort enriched by FGFR status. • Data from 128 patients who received pemigatinib monotherapy (dose escalation Part 1: n=49; dose expansion Part 2: n=79), with patients receiving either intermittent (n=70) or continuous (n=58) dosing.
Seamless Features	• Dose escalation stage transitioned directly into expansion. • Multiple expansion cohorts defined by tumor type and FGFR alteration status; enrollment adapted based on emerging activity signals.
Key Design / Operational Decisions	• Standard 3+3 dose-escalation scheme to determine MTD and recommended dosage. • Safety monitoring (dose interruptions/reductions for TEAEs) • Operational flexibility to evaluate a broad spectrum of tumors and FGFR alterations within one study.
Endpoints	• Escalation: MTD, RP2D, safety, PK/PD, and biomarker correlations. • Expansion: activity (primary endpoint: ORR), DOR, PFS, OS, and safety; and exploratory assessment of predictive biomarkers and FGFR alteration–driven responses.
Regulatory Interactions / Outcomes	• Data from patients informed dosage selection for subsequent studies. ○ FIGHT-101 was conducted prior to the FDA's Project Optimus initiative. • Signals of antitumor activity and safety across multiple tumor types supported initiation of a registrational trial (FIGHT-202), with a primary endpoint that differed from FIGHT-101.

Abbreviations: NSCLC, non-small cell lung cancer; MTD, maximum tolerated dose; TEAE, treatment-emergent adverse event; RP2D, recommended phase two dose; PK/PD, pharmacokinetics/pharmacodynamics; OR, objective response, DOR, duration of response; PFS, progression-free survival; OS, overall survival.

Seamless with Registrational Intent

ARROW

ARROW (NCT03037385) evaluated pralsetinib, a RET tyrosine kinase inhibitor, in RET fusion-positive non-small cell lung cancer (NSCLC), thyroid cancer, and other RET-altered tumors using an inferentially seamless design with registrational intent (Table 4).[10] Patients who received the recommended dosage during the dose-finding portion were integrated into pivotal analyses. The design combined dose escalation and

multiple expansion cohorts, allowing early data to guide enrollment and cohort adaptations. This structure enabled data generation within a unified study framework.

The trial ultimately supported FDA approval of pralsetinib for treatment of RET fusion-positive NSCLC and thyroid cancers, with additional expansion cohorts exploring other RET-altered tumors.[11–13]

Table 4. Key Features of ARROW Trial

Key design, operational, and patient population features

Feature	Details
Population / Rarity	• Patients with advanced RET-altered solid tumors including NSCLC (~1–2% prevalence), medullary thyroid cancer (50–80% RET mutations), and other RET-fusion tumors • Data from 647 total patients (dose-escalation: 62; expansion and registrational cohorts: 585). ○ Data from 471 patients at selected dose (NSCLC: 233; thyroid cancer: 162; other RET-fusion tumors: 76)
Seamless Features	• Inferentially seamless design integrating first-in-human dose escalation, cohort expansion, and registrational intent within a single protocol • Data from patients at recommended dosage included in dataset to support approval • Multiple expansion cohorts defined by tumor type and prior therapy; cohort sizes adapted in response to emerging data.
Key Design / Operational Decisions	• BOIN dose-escalation in small cohorts, with selected dosage determined by tolerability and activity. • Adaptive cohort sizing (e.g., Group 2 NSCLC increased from ~80 → 200) was not prespecified and occurred via protocol amendments based on emerging data. • Operational flexibility achieved through protocol amendments • Safety monitoring tailored to tumor type and prior therapy
Endpoints	• Escalation: MTD, RP2D, safety, OR, PK/PD, biomarker correlations • Expansion and registrational (NSCLC cohorts): activity (primary endpoint: OR), DOR, CB, DC, PFS, OS, intracranial response, safety
Regulatory Interactions / Outcomes	• Data from patients who received the proposed recommended dosage supported U.S. approvals for RET fusion-positive NSCLC and thyroid cancers.

Abbreviations: NSCLC, non-small cell lung cancer; BOIN, Bayesian Optimal Interval; MTD, maximum tolerated dose; OR, overall response; PK/PD, pharmacokinetics/pharmacodynamics; DOR, duration of response; CB, clinical benefit; DC, disease control; PFS, progression-free survival; OS, overall survival.

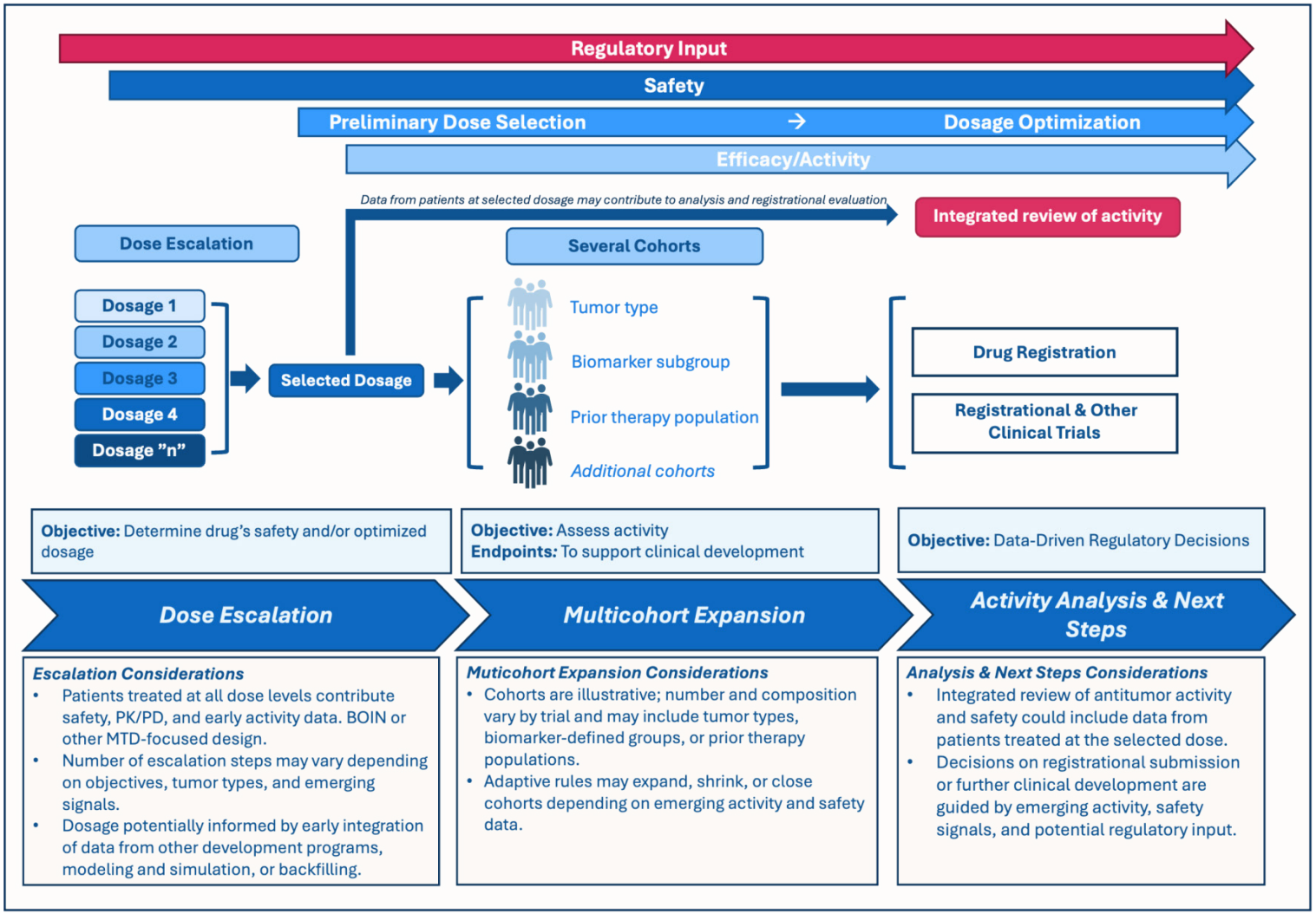
Key Insight: Seamless designs can enable faster development and support multiple approvals when expansion cohorts show consistent, high response rates. When robust efficacy is observed across tumor types, treatment lines, or specific biomarkers, efficiencies in development may be largely driven by the drug itself. Integrating patients who received the RP2D from early cohorts into analyses to support the approval can shorten timelines but requires clear regulatory alignment on dosage optimization, adaptive rules, and safety monitoring. Single studies can include multiple expansion cohorts that could support distinct

indications, though successful implementation depends on structured dosage-optimization strategies, careful endpoint selection, and balancing pre-specified elements with necessary flexibility to maintain statistical rigor and operational feasibility.

ARROW applied a seamless trial design in RET fusion-positive NSCLC, integrating adaptive dose escalation with multiple expansion cohorts and pooling patients treated at the RP2D to support registration. Other drugs used a similar approach, including LIBRETTO-001; this inferentially seamless trial was designed to evaluate selpercatinib in patients with previously treated and treatment-naive RET fusion-positive NSCLC, and incorporated data from the dose escalation portion into the primary efficacy analysis to maximize data and follow-up.[14–16] TRIDENT-1 also used an inferentially seamless design in ROS1 fusion-positive NSCLC, enrolling multiple molecularly defined cohorts—including TKI-naive and pretreated patients—and pooled data from patients enrolled in dose-finding to optimize sample size and evaluate efficacy across diverse populations.[17–19] It is important to note that these studies were conducted prior to the FDA's Project Optimus initiative final guidance, emphasizing systematic dosage optimization.

Across these trials, the seamless design strategy—combining dose-escalation, adaptive cohort expansion, and integrated analyses to support registration—demonstrates a flexible and pragmatic approach, which can incorporate elements such as pooling patients across development and tailoring dosage finding to accelerate development. This structure can help enable faster development, support multiple patient populations within a single protocol, and provide a framework for efficiently generating the evidence needed for regulatory approval. Figure 1 illustrates a potential general structure for trial designs, highlighting key elements such as cohort expansion, dose escalation, and pooling strategies, while demonstrating the adaptability of seamless designs across different disease contexts.

Figure 1. Conceptual Model of Seamless Trial Structure with Registrational Intent



Abbreviations: BOIN, Bayesian optimal interval; MTD, maximum tolerated dose; ORR, overall response rate; PK/PD, pharmacokinetics/pharmacodynamics; DOR, duration of response; PFS, progression-free survival.

Seamless with Randomized Components

RINGSIDE

RINGSIDE (NCT04871282) is an ongoing clinical trial evaluating AL102, a γ-secretase inhibitor, in patients with desmoid tumors, a rare and locally aggressive fibroblastic neoplasm (Table 5).[20] The trial was designed as an integrated Phase II/III study: an open-label Phase II dose-finding stage explored multiple dosing regimens, while a randomized, double-blind, placebo-controlled Phase III is underway to confirm efficacy and safety of the selected regimen.

Table 5. Key Features of RINGSIDE Trial

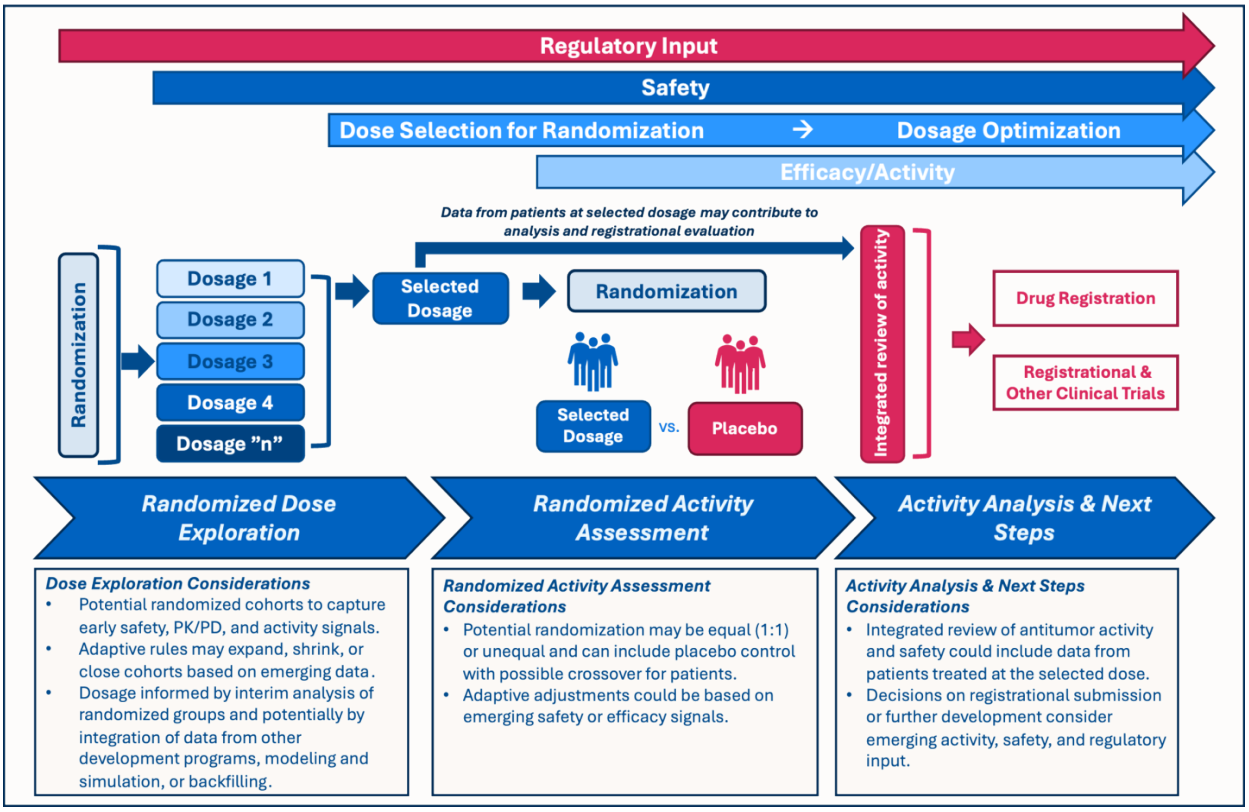
Key design, operational, and patient population features

Feature	Details
Population / Rarity	• Adults with progressing desmoid tumors (aggressive fibromatosis); 42 patients enrolled in initial activity assessment; multi-country pivotal stage ongoing. • Ultra-rare soft-tissue tumors (~2–4 per million annually)
Seamless Features	• Initial randomized evaluation of multiple doses to identify optimal regimen. • Randomization embedded at both activity/efficacy and registrational stages. • Operationally seamless design, without carryover of patients.
Key Design / Operational Decisions	• Randomized, placebo-controlled assessment to evaluate antitumor activity and safety. • Adaptive dosing and cohort selection based on emerging activity data.
Endpoints	• Early evaluation: tumor response, disease control rate, tumor volume change, T2 signal intensity; safety and tolerability. • Confirmatory evaluation: progression-free survival (primary), symptom control, quality of life, overall safety.
Regulatory Interactions / Outcomes	• Dosage selection for the ongoing Phase III study.

Key Insight: Seamless trials are not confined to single-arm expansion strategies; they can also integrate randomization at pivotal stages. By embedding randomized cohorts within a seamless framework, sponsors can capture early safety and activity signals while simultaneously generating the confirmatory evidence regulators require. Randomized seamless designs may be particularly valuable in rare cancers, where efficient use of limited patient populations must be weighed against the need for credible, comparative evidence.

Figure 2 illustrates a potential structure for seamless Phase II/III trial designs, highlighting key elements such as randomized dose exploration, dosage selection, and activity evaluation, while showing how randomization can be embedded at multiple stages to generate early safety and efficacy signals alongside confirmatory evidence.

Figure 2. Conceptual Model of Seamless Trial Structure with Randomized Components



Abbreviations: PK/PD, pharmacokinetics/pharmacodynamics

Key Design and Operational Considerations

Patient and Advocate Engagement

Engaging patient advocacy groups early in trial planning is particularly critical in rare cancers. Advocates can provide insights on patient priorities, tolerability, and feasibility, particularly in limited populations. Inclusion of patient advocates and key opinion leaders in FDA meetings allows discussion of complex trade-offs in trial design, dosing, and treatment considerations. Input from advocates helps ensure seamless trial elements are meaningful from the patient perspective, including early endpoints, adaptive features, and pragmatic design elements. **By combining iterative learning, flexible frameworks, and patient-centered input, seamless designs can accelerate drug development while generating high-quality evidence that addresses both clinical and patient priorities.**

Regulatory Engagement and Global Considerations

Early and ongoing engagement with regulators is essential for employing seamless trial designs to facilitate rare cancer drug development. In rare cancers, proactive regulatory dialogue can help clarify expectations on important elements of trial design and other aspects of development, such as inferentially designed elements (e.g., pool strategies) or dosage optimization strategies, thus reducing the risk of late-stage redesigns or delays. Formal FDA meetings, such as Type B, C, or D meetings, provide a structured forum for

clear guidance on trial design, adaptive features, dosage optimization, and regulatory expectations, and allow sponsors to align adaptations with regulatory priorities and ensure trial integrity despite mid-course adjustments.[21,22]

Beyond FDA, global regulatory coordination is increasingly important. Many rare cancer trials recruit globally to achieve sufficient enrollment, which introduces complexities related to varying SOCs, ethical frameworks, and data collection practices. Divergent expectations around acceptable endpoints or evidence thresholds can create hurdles for sponsors aiming to generate unified evidence packages. International engagement early in trial planning—through initiatives like parallel scientific advice meetings—can help facilitate international development, reduce duplication, and expand access to clinical trials. **Seamless designs, by nature of integrating multiple phases and endpoints, amplify the importance of early discussions to aid in achieving an international development program that can satisfy regulatory expectations of multiple regulatory authorities.**

Feasibility: Safety Assessments and Dosage Optimization

Dosage optimization presents unique challenges in rare cancers. Because reliable early measures of antitumor activity are often lacking, dosage selection may be primarily guided by toxicity, which can hinder dosage optimization.

Thoughtful use of validated clinical outcome assessments (COAs) can provide additional insight into tolerability and symptom burden, helping guide dosage optimization. Pre-specified dose-finding schemes may require mid-course adjustments as accumulating data refine understanding of dose- and exposure- and response relationships for safety and activity. Safety run-ins can be considered for monitoring, dosage selection, and seamless trial conduct. When a drug is already approved in other indications, extensive safety evaluations may not be needed; however, assessing potential safety issues or key pharmacokinetics (PK) interactions with new treatment regimens remain essential. Adaptive evaluations of multiple doses, supported by early PK and exposure–response analyses, can maximize learning from each patient. Dosage optimization strategies must balance scientific rigor with feasibility given small cohorts, competing SOCs, and heterogeneous trial sites.

Pediatric Considerations

For pediatric populations, potential differences in PK, between very young and older pediatric patients due to ontogeny, age-appropriate formulation considerations, and developmental-specific safety concerns for some products can affect dosage optimization and adaptive strategies. Adaptive strategies may include adjusting dosages, cohort progression, or enrollment criteria based on accumulating pediatric PK, safety, or activity data. While safety profiles are often similar in pediatric and adult patients, seamless trial designs may require modifications—such as staggered cohort enrollment or additional monitoring—to ensure integrated dose-escalation and expansion elements are safe and appropriate for younger patients.

Balancing Pre-Specification and Flexibility

Balancing pre-specification with the need for protocol amendments is a central challenge in seamless trials. Pre-specifying rules for dose expansion, dropping arms, or patient pooling across phases help maintain statistical rigor and enable confidence in the resulting data. However, in rare cancers, limited early data and evolving knowledge of patient response often make amendments inevitable. Midstream protocol amendments such as adding biomarker-defined cohorts, adjusting eligibility criteria, or incorporating new

endpoints may be necessary as insights emerge. Sponsors may consider prospectively identifying which elements can realistically be pre-specified and where amendments to the protocol and regulatory interaction may be necessary to preserve trial feasibility, efficiency, and integrity. **Addressing this challenge proactively through pre-specified interim assessment of data generated in the trial and timely regulatory engagement when needed is key to increasing adoption of seamless approaches.**

Endpoints

In rare cancer drug development, endpoints are especially important when employing seamless trial designs that integrate early- and late-phase objectives to guide dosage optimization, safety evaluation, and adaptive trial decisions. While early endpoints could provide valuable insight into drug activity and support trial adaptations, they generally lack the validation necessary for standalone regulatory approval. In settings of high unmet medical need, regulators may allow some flexibility in accepting novel early endpoints, but these must be clearly justified, interpretable, and linked to meaningful clinical outcomes.

Endpoint selection should consider feasibility, biological and clinical relevance, anticipated drug activity, interpretability, and precedent from similar rare disease settings. When traditional endpoints are not suitable or feasible, complementary endpoints can help characterize drug activity. In some rare cancers, traditional measures may show modest effect yet traditional approval can be achieved if a clinically meaningful supportive endpoint reinforces the evidence of benefit.[23] In contrast, in diseases without accepted early endpoints, such as glioblastoma, trials often rely on overall survival, which can limit opportunities for early adaptation and dosage optimization within seamless designs. Incorporating COAs can further strengthen evidence in this setting by providing longitudinal data across phases. Measures of patient experience—if collected consistently from early stages through registration—can offer early insights into tolerability and build a more comprehensive view of clinical benefit over time. Remote data collection may also reduce site burden and improve feasibility, complementing efficacy and safety endpoints as part of the totality of evidence. **Selecting endpoints that maximize learning from each patient and thoughtfully integrating both supportive and primary endpoints can help seamless trials remain efficient while generating credible evidence to inform regulatory decisions.**

Conclusions and Future Directions

Seamless clinical trials represent a paradigm shift in rare cancer drug development, addressing the limitations of conventional phased approaches by integrating multiple stages of clinical evaluation into a single protocol. When thoughtfully designed, seamless trials can accelerate timelines, reduce redundancy, and generate robust evidence from limited populations without compromising scientific integrity.

Beyond operational efficiency, seamless designs support patient-centered considerations, including early engagement with advocacy groups, meaningful endpoints, and pragmatic trial elements. As regulatory agencies increasingly embrace innovative methodologies in areas of high unmet need, seamless trials provide a practical, patient-centered framework that balances innovation with regulatory expectations.

Realizing the full potential of seamless trials requires collaboration among sponsors, investigators, regulators, and patient advocates. As experience grows, stakeholders can iteratively refine trial elements—learning from successes and challenges to improve efficiency, adaptive decision-making, and endpoint

selection. Sharing best practices and proactively seeking FDA guidance further ensures that adaptive designs remain feasible, meaningful, and aligned with regulatory priorities.

Looking forward, broader frameworks, including platform trials and other consolidated designs, offer opportunities to integrate new hypotheses, emerging patient subgroups, and novel indications within a single adaptive structure. Such approaches are particularly valuable in rare cancers, where limited populations demand careful resource allocation and coordination. By combining iterative learning, flexible frameworks, and patient-centered input, seamless trials can become increasingly efficient, informative, and aligned with both clinical and patient priorities, ultimately accelerating meaningful therapeutic advances in rare malignancies.

References

1. U.S. Food and Drug Administration. Demonstrating Substantial Evidence of Effectiveness with One Adequate and Well-Controlled Clinical Investigation and Confirmatory Evidence (Draft Guidance, September 2023). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-evidence-effectiveness-one-adequate-and-well-controlled-clinical.
2. U.S. Food and Drug Administration. Clinical Trial Considerations to Support Accelerated Approval of Oncology Therapeutics (Draft Guidance, Marh 2023). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-considerations-support-accelerated-approval-oncology-therapeutics.
3. U.S. Food and Drug Administration. Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics (Final Guidance, March 2022). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/expansion-cohorts-use-first-human-clinical-trials-expedite-development-oncology-drugs-and-biologics.
4. U.S. Food and Drug Administration. Rare Diseases: Considerations for the Development of Drugs and Biological Products (Final Guidance, December 2023). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-considerations-development-drugs-and-biological-products.
5. U.S. Food and Drug Administration. Project FrontRunner. Available from: https://www.fda.gov/about-fda/oncology-center-excellence/project-frontrunner.
6. Subbiah V, Iannotti NO, Gutierrez M et al. FIGHT-101, a First-In-Human Study of Potent and Selective FGFR 1-3 Inhibitor Pemigatinib in Pan-Cancer Patients with FGF/FGFR Alterations and Advanced Malignancies. Annals of Oncology 2022;33:522–533. DOI: 10.1016/j.annonc.2022.02.001.
7. Vogel A, Sahai V, Hollebecque A et al. An Open-Label Study of Pemigatinib in Cholangiocarcinoma: Final Results from FIGHT-202. ESMO Open 2024;9:103488. DOI: 10.1016/j.esmoop.2024.103488.
8. Patel TH, Marcus L, Horiba MN et al. FDA Approval Summary: Pemigatinib for Previously Treated, Unresectable Locally Advanced or Metastatic Cholangiocarcinoma with FGFR2 Fusion or Other Rearrangement. Clinical Cancer Research 2023;29:838–842. DOI: 0.1158/1078-0432.CCR-22-2036.
9. U.S. Food and Drug Administration. FDA Grants Accelerated Approval to Pemigatinib for Cholangiocarcinoma with an FGFR2 Rearrangement or Fusion. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-pemigatinib-cholangiocarcinoma-fgfr2-rearrangement-or-fusion.
10. Gainor JF, Curigliano G, Kim DW et al. Pralsetinib for RET Fusion-Positive Non-Small-Cell Lung Cancer (ARROW): A Multi-Cohort, Open-Label, Phase 1/2 Study. Lancet Oncol 2021;22:959–969. DOI: 10.1016/S1470-2045(21)00247-3.
11. Subbiah V, Cassier PA, Siena S et al. Pan-Cancer Efficacy of Pralsetinib in Patients with RET Fusion–Positive Solid Tumors from the Phase 1/2 ARROW Trial. Nat Med 2022;28:1640–1645. DOI: 10.1038/s41591-022-01931-y.

12. Kim J, Bradford D, Larkins E et al. FDA Approval Summary: Pralsetinib for the Treatment of Lung and Thyroid cancers with RET gene mutations or fusions. Clinical Cancer Research 2021;27:5452–5456. DOI: 10.1158/1078-0432.CCR-21-0967.
13. U.S. Food and Drug Administration. FDA Approves Pralsetinib for Non-Small Cell Lung Cancer with RET Gene Fusions. Available from: https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pralsetinib-non-small-cell-lung-cancer-ret-gene-fusions.
14. Study Details | NCT03157128 | A Study of Selpercatinib (LOXO-292) in Participants with Advanced Solid Tumors, RET Fusion-Positive Solid Tumors, and Medullary Thyroid Cancer (LIBRETTO-001) (LIBRETTO-001). Available from: https://clinicaltrials.gov/study/NCT03157128.
15. Drilon A, Oxnard GR, Tan DSW et al. Efficacy of Selpercatinib in RET Fusion–Positive Non–Small-Cell Lung Cancer. New England Journal of Medicine 2020;383:813–824. DOI: 10.1056/NEJMoa2005653.
16. U.S. Food and Drug Administration. FDA Approves Selpercatinib for Locally Advanced or Metastatic RET Fusion-Positive Non-Small Cell Lung Cancer. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-selpercatinib-locally-advanced-or-metastatic-ret-fusion-positive-non-small-cell-lung.
17. Study Details | NCT03093116 | A Study of Repotrectinib (TPX-0005) in Patients with Advanced Solid Tumors Harboring ALK, ROS1, or NTRK1-3 Rearrangements (TRIDENT-1). Available from: https://www.clinicaltrials.gov/study/NCT03093116.
18. Drilon A, Camidge DR, Lin JJ et al. Repotrectinib in ROS1 Fusion–Positive Non–Small-Cell Lung Cancer. New England Journal of Medicine 2024;390:118–131. DOI: 10.1056/NEJMoa2302299.
19. U.S. Food and Drug Administration. FDA Approves Repotrectinib for ROS1-Positive Non-Small Cell Lung Cancer. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-repotrectinib-ros1-positive-non-small-cell-lung-cancer.
20. Study Details | NCT04871282 | A Study of AL102 in Patients with Progressing Desmoid Tumors (RINGSIDE). Available from: https://clinicaltrials.gov/study/NCT04871282.
21. U.S. Food and Drug Administration. Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/formal-meetings-between-fda-and-sponsors-or-applicants-pdufa-products.
22. U.S. Food and Drug Administration. Best Practices for Communication Between IND Sponsors and FDA During Drug Development. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-communication-between-ind-sponsors-and-fda-during-drug-development.
23. U.S. Food and Drug Administration. FDA Approves Belzutifan for Pheochromocytoma or Paraganglioma. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-belzutifan-pheochromocytoma-or-paraganglioma.

FRIENDS of CANCER RESEARCH

Assessing Contribution of Effect (COE) in Oncology Combination Therapies: Lessons Learned to Inform and Optimize Future Registrational Trial Designs

White Paper | 2025

Executive Summary

The pace of oncology drug development has accelerated significantly over the past decade, with a growing emphasis on the use of combination drug therapies and their potential to offer patients earlier access to novel treatments that can significantly improve outcomes. However, introducing more drugs into a regimen may also increase the risk of toxicity, making it important to ensure that each drug contributes meaningfully to the overall therapeutic benefit. To evaluate this, researchers study the contribution of effect (COE), or the impact of each individual drug in a combination therapy. The recommended trial design to assess COE is a factorial trial design, in which each component (e.g., Drug A and Drug B) is tested individually, in combination (A+B), and against a control (C), across four arms: (A, B, A+B, and C).

As many modern combination therapies are increasingly co-developed from the outset, rather than by combining data from separately developed therapies, demonstrating each component's contribution to the overall effect can be more challenging. While understanding COE in combination therapies is essential, there are often limitations to a fully factorial study design. To address these, alternative trial designs may be considered, provided they generate sufficient COE data and are agreed with regulators before studies are initiated.

Friends of Cancer Research (*Friends*) convened a multi-stakeholder working group including experts from the U.S. Food and Drug Administration (FDA), National Cancer Institute (NCI), pharmaceutical companies, academia, and patient advocacy organizations to discuss best practices for generating COE data. The group explored several strategies and considerations:

- Alternative trial designs: Exploring options beyond traditional factorial designs for registrational trials to evaluate COE, including adaptive designs, 2-arm, and 3-arm trials, as well as trials that include descriptive statistical comparisons.
- Leveraging data from previously approved examples:
    - Using early evidence of COE in one cancer type to support streamlined development across other cancer types, especially for rare cancers or high unmet need settings.
    - Analyzing data using simple comparisons to assess COE of drugs with the same mechanism of action (MOA) as the comparator when similar outcomes are expected.
    - Considering data from early-phase trials, especially when both components demonstrate activity, to justify excluding monotherapy arms.
    - Omitting arms when there is a strong biological rationale, such as when one drug is known to lack monotherapy activity.
- Additional considerations: While we focus on contribution of effect in doublets, safety and the complexity of multi-drug regimens must be carefully evaluated when designing combination trials.

These insights emphasize the importance of integrating thoughtful trial design and practical considerations into combination therapy development in ways that maximize patient benefit while maintaining scientific rigor. By adopting more flexible approaches, clinicians can better understand the COE of each drug, paving the way for more effective treatment options in oncology.

Authors

Hillary Andrews, Friends of Cancer Research

Gideon Blumenthal, Merck & Co., Inc.

Julie Bullock, GSK

Gary Doherty, AstraZeneca

Anna Farago, Amgen, Inc.

Scott Greenfeder, Bayer AG

Yuko Ishii, *Friends* Advisory Advocate

Sreeneeranj Kasichayanula, Gilead Sciences

Michael LeBlanc, Fred Hutchinson Cancer Center

Chunze Li, Genentech Inc

PK Morrow, Takeda Pharmaceuticals Inc.

Lorraine Pelosof, National Cancer Institute

Punam Sandhu, Daiichi Sankyo

Sujay Shah, Johnson & Johnson

Elad Sharon, Dana-Farber Cancer Institute

Zhen Su, Marengo Therapeutics

Carol Vallett, *Friends* Advisory Advocate

Yolanda Villalobos, Eli Lilly and Company

This white paper was developed through discussions that included these experts and other perspectives representing academia, industry, the U.S. Food and Drug Administration, and the patient advocacy community. The views expressed here represent the collective insights from working group discussions and do not necessarily reflect the official positions of any individual organization.

Table of Contents

Introduction

The speed of oncology combination drug development has accelerated over the past decade. While this can enable earlier access to novel, highly effective treatment options for patients, it can also present challenges in demonstrating that each component meaningfully contributes to the efficacy observed in the combination, with an acceptable safety profile. Ideally, drugs combine synergistically, where their efficacies are greater than the sum of the individual effects, or additively, where the combined effect equals that of the sum of the individual components. Conversely, an unfavorable situation occurs if the combination provides no meaningful benefit relative to either of the components administered alone, highlighting the importance of ensuring contribution of effect (COE) is well understood.

Traditionally, when the pace of advancing combination therapies was less rapid, it was more common for sponsors to evaluate drugs in combination after the approval of one or more parts of the combination in that clinical setting. Increasingly, however, therapies are intentionally developed and studied as part of a combination from the outset due to scientific or clinical factors, and/or the expected safety and efficacy, which can lead to limited or unavailable monotherapy data. As such, isolating and understanding the COE of each component can be challenging, given overlapping effects, the need for methodological rigor and pressures for meeting unmet clinical needs.

Demonstrating the COE of each component can ensure that a combination therapy provides a clinically meaningful benefit over its individual components, while avoiding unnecessary toxicity and informing decisions about the adoption of treatments that may offer only marginal improvements at the cost of increased burden to patients. The U.S. Food and Drug Administration (FDA) recently released draft guidance for establishing clinical evidence to demonstrate the COE in novel combinations in oncology.[1] Building on this guidance, this white paper explores practical design considerations for registrational trials demonstrating COE, highlighting areas where additional clarity, flexibility, or alternative approaches may help to operationalize these principles effectively in clinical development programs.

Scope

The gold standard for demonstrating COE in a registrational trial is a factorial design where the individual components (A and B) are evaluated both individually and together across four arms: A, B, AB, and C (standard of care [SOC] control). While this approach may provide the most rigorous evidence, it may be infeasible due to recruitment and logistical challenges as well as the potential for added statistical complexity, particularly in rare populations or aggressive disease settings. Furthermore, the use of factorial designs may not be appropriate in settings where there is sufficient evidence that either A or B is inactive or poorly active as a single agent.

This white paper examines scenarios where alternative designs may be considered to provide sufficient evidence of COE in registrational trials. For each scenario, we provide examples of previous registrational trials and discuss 1) the data that supported deviating from a traditional factorial design, and 2) key considerations for ensuring the alternative approach remains scientifically robust, interpretable, and supportive of regulatory decision-making.

We focus on efficacy assessments rather than dosing decisions (which should ideally be established prior to registrational trials) or safety assessments (which are evaluated as part of the overall benefit-risk assessment). Discussions outlined herein are framed around patient-centricity and multi-disciplinary

perspectives, recognizing both scientific rigor and operational realities in the development of oncology combination therapies.

Factorial Design

In addition to providing treatment effect estimates for each experimental arm (A, B, and AB) relative to C, the factorial design isolates the contribution of each component by allowing either formal or exploratory comparisons of AB vs. A and B. From a statistical perspective, using a factorial trial design and making comparisons across all arms (i.e., AB vs. A, B, C and C vs. A and B) requires additional multiplicity adjustment and ultimately, more patients enrolled, prolonging the time to readout, approval, and patient access.

While statistical methods can help to minimize sample size inflation,[2] additional concerns may arise if there are sufficient data that either of the individual components is less active or have outcomes that are worse, than current SOC. Patients may be hesitant to join trials due to concerns about randomization to a less active arm, so it is critical to describe trial designs to patients as part of their consent as well as to carefully consider these aspects during study design. Clinical equipoise should be maintained, and trials should only include arms that have valid scientific questions about their effectiveness. A variety of alternative trial designs can be considered to demonstrate COE based on data availability.

Alternative Trial Designs

Many oncology combination trials do not use a factorial design. Instead, they use data to justify alternative approaches including adaptive designs, 2- or 3-arm trials, or analyses with descriptive comparisons. In some cases, more than one approach can be considered. For each, the approach should be discussed and agreed upon with regulatory agencies before initiating the trial.

Adaptive Study Design

An adaptive study design includes a pre-determined interim analysis plan to assess early signals from the experimental arms. Based on these results, the trial may undergo modifications, such as dropping one of the experimental arms that demonstrates futility, adjusting sample size, or adapting randomized probabilities to favor arms performing better. Early dropping of monotherapy arm(s) should be carefully considered to ensure that sufficient evidence has been acquired to demonstrate COE. These approaches use methods such as group-sequential testing to maintain statistical validity and control for type I error. For example, the STAMPEDE trial in advanced or metastatic prostate cancer was designed to evaluate the efficacy of adding various agents to standard androgen deprivation therapy. Arms that did not show survival benefit were discontinued and those that did were eventually incorporated into clinical SOC.[3]

3- Arm Trial Design

If one of the monotherapy arms (B) is known to be ineffective, poorly effective, or has well known activity demonstrating inferiority to the combination, a 3-arm design (AB, A, and C) could be considered where both investigational arms are compared to C. Alternatively, although less common, one of the monotherapy arms may already be the established SOC. Often, 3-arm trial designs include parallel testing (i.e., A vs. C and AB vs. C) or sequential testing (e.g., AB vs. C then, A vs. C), but do not formally test AB vs. A. Isolating the contribution of B to the combination would still require comparing AB vs. A if both AB and A are shown to be better than C.[4]

2- Arm Trial Design

A 2-arm design may be used to demonstrate the additional benefit of a second drug, particularly when it involves adding a new agent (B) to the SOC (A), where the efficacy of A is already well established and B alone would not be considered an effective treatment option. Other scenarios include when COE has already been established or where A and B are both inactive or poorly active as monotherapies. Confirming that any inactive comparator arms are indeed pharmacologically or clinically inactive to ensure the validity of the observed treatment effect would be necessary. Data from previously completed clinical trials often support an understanding of COE for comparisons that do not include all monotherapy arms.

Trials that Include Descriptive Comparisons

In some cases, the primary analysis will focus on AB vs. C while comparing AB vs. monotherapy arm(s) may be adequately guided by a less stringent exploratory rule, as long as limitations of this less stringent approach are understood by all stakeholders.[5–7] It should be appreciated that the less stringent approaches that do not involve a reproducible or quantifiable decision rule, either in favor or against the comparison, should be considered on a case-by-case basis. Furthermore, in the absence of a prespecified decision rule, the ability of the approach to address COE cannot be quantified. Decisions regarding thresholds for efficacy improvements should also be prespecified and safety profiles would need to support a positive benefit risk assessment.

Data to Support COE

Data developed before the registrational trial can support efficacy of individual components to guide decisions on choosing an alternative trial design to the factorial design. However, there is no single definition or threshold for what constitutes "sufficient" evidence to demonstrate efficacy, and the relevance of any data source depends on the molecule, mechanism of action (MOA), and the specific context of the trial. Additionally, many studies demonstrate activity, or anti-tumor effect, which does not always guarantee there is long-term clinical benefit. Below, we outline potential sources for data that support an understanding of efficacy.

Clinical Data

Clinical data, either from earlier phases of the program of interest or other trials from the sponsor, can support an understanding of COE. The strongest COE evidence may come from the early phase clinical trials that use randomized designs and report definitive clinical endpoints. Depending on the circumstance, data may be extrapolated from other settings such as different cancer types, molecules, or biomarker populations.

Early endpoints that support an understanding of potential efficacy are critical. Objective response rate (ORR), alone or in conjunction with duration of response (DOR), is an early endpoint that can be considered as a marker for activity. Other tumor markers, such as prostate specific antigen (PSA) or circulating tumor DNA (ctDNA), may be used as early signals of activity depending on the disease and drug, but they should not replace adequate demonstration of COE. Early endpoints should be relevant to the clinical setting/registrational trial. In general, assessing COE would ideally use clinical endpoints that have served for other regulatory decision-making (e.g., progression free survival (PFS), overall survival (OS), ORR) and evaluating the totality of evidence across multiple early readouts.

Data Modeling and Non-Clinical Pharmacology

Modeling can support expectations around drug activity. Pharmacokinetics (PK)/ pharmacodynamics (PD) modeling and Tumor Growth Inhibition modeling using clinical or pre-clinical data can support a mechanistic rationale to support the clinical data established in early phase studies and strengthen the evidence base. Model based meta-analysis (MBMA) uses study-level covariates to build a quantitative model of treatment-response relationships in clinical data. MBMA can be used to identify class-level COE interactions regarding ORR, PFS, and OS.[8] Other non-clinical data (e.g., *in vivo* patient-derived xenograft models, organoids, etc.) can provide supportive evidence. These approaches would not be considered adequate to establish COE on their own.

Data from Other Settings

Data including preclinical data, activity, or efficacy data from other setting(s) can support the design approach in a new setting, including different cancer types or populations (e.g., different disease stages, biomarker studies, studies completed in other countries). Additionally, real-world data and evidence from off-label use of components can support an understanding of COE and ultimately, an alternative registrational design from the factorial design when available. Why the data support the alternative trial approach should be justified with consideration for the natural history of the disease, SOC, and biomarkers, which may differ in different settings.

Scenarios

To demonstrate when each of the alternative trials could be considered and what data might justify their use, we identified scenarios for registrational trial designs of previously FDA reviewed combinations. The default assumption when designing registrational trials that assess COE is that both components of the combination are individually effective and thus a factorial design could be used. Below, we outline data that may support an alternative approach to demonstrating COE. We categorized the scenarios based on data supporting COE that were available ahead of the registrational trial and led to the use of an alternative trial design.

For the combinations where each component is anticipated to be active as a monotherapy, we considered 3 categories of data:

1) Other tumor types – Certain therapies are effective in multiple cancer types and there may be opportunities to leverage the totality of data across tumor types.
2) Drugs with similar MOA – For drugs with similar MOA that are expected to behave similarly, it may be possible to use informal comparisons (i.e., control arm vs. investigational agent with same MOA and expectation of similar activity in the disease setting).
3) Other trials in the same cancer setting – As development programs for combinations evolve, there may be opportunities to leverage data from other trials in the sponsor's portfolio that demonstrate COE for the monotherapy arms compared to the combination (e.g., earlier phase trials, other phase III trials).

Lastly, we include a section focused on scenarios where we do not anticipate activity in one or more of the arms as a monotherapy and the data that might be used to justify this.

Other Tumor Types

Anti-PD-(L)1 + anti-CTLA-4

Anti–PD-(L)1 and anti–CTLA-4 therapies are immune checkpoint inhibitors. Anti-PD-(L)1 blocks the interaction between PD-1 on T cells and PD-L1 on tumor cells, preventing immune suppression and allowing T cells to recognize and attack cancer cells. Anti-CTLA-4 targets a separate checkpoint that regulates early T-cell activation, thereby enhancing T-cell proliferation and activity. Studying these agents in combination leverages complementary mechanisms with known class effects: CTLA-4 blockade primes and expands T-cells, while PD-(L)1 inhibition sustains their activity within the tumor microenvironment, often resulting in more durable and robust anti-tumor responses. This combination strategy has been shown to improve outcomes in multiple cancer types, including melanoma, non–small cell lung cancer (NSCLC), renal cell carcinoma (RCC), and others, and has become an important approach in treatment of solid cancers. The development programs for subsequent cancer types relied on data from previous trials that support demonstration of COE. To demonstrate the stepwise data analyses across cancer types, **Table 1** outlines the various trials, outcomes, and key takeaways.

Throughout the development of anti-CTLA-4 and anti-PD-(L)1 combination therapies, data from other cancer types supported decisions to not investigate certain monotherapy arms for COE in the registrational trial. Initially, the anti-CTLA-4 alone arm was not included because evidence suggested that anti-CTLA-4 inhibition had limited efficacy as monotherapy. Eventually, the anti-PD-(L)1 arm alone became SOC for many settings or was not included due to lessons learned from other settings. Sponsors used their own internal data from earlier phase clinical trials as well as publicly available data from others to justify the decision about which arms to include.

Table 1. Anti-CTLA4 + anti-PD-(L)1

Trial {Primary Source} **Cancer type (Setting)** **Year of FDA Approval**	**Trial arms**	**Findings** *(Confidence Intervals [CI] included in parenthesis)**	**Trial Design Takeaways**
CheckMate 067[9] Unresectable or metastatic melanoma (First line) 2016	1:1:1 ratio: • Nivo + ipi (combo) • Nivo • Ipi	**mPFS** Combo = 11.5 mos (8.9-16.7) Nivo = 6.9 mos (4.3-9.5) Ipi = 2.9 mos (2.8-3.4) HR for combo vs. ipi = 0.42 (0.31-0.57); p<0.001 HR for nivo vs. ipi = 0.57 (0.43-0.76); p<0.001	• First approval of the combo – ipi and nivo were each approved separately for unresectable or metastatic melanoma • Formal planned statistics compared the combo and nivo arm with the ipi arm – when the trial started, ipi was not fully established as SOC but had shown efficacy over other options (i.e., chemo) • The 10-year follow-up[10] suggested potentially non-trivial 8% improvement in 5-year OS rates for nivo-ipi vs. nivo alone, however the study was not designed with formal comparison for the nivo-ipi vs. nivo alone
CheckMate 214[11] Advanced RCC (First line) 2018	1:1 ratio: • Nivo + ipi (combo) • Sunitinib (SOC)	**mOS** Combo = not reached Sunitinib = 26.0 mos HR = 0.63 (0.44-0.89); p<0.001	• No nivo or ipi monotherapy arm included – previous data from single-arm studies had demonstrated substantially lower ORR for each monotherapy in the first-line RCC setting compared to ORR for the combo • Supportive data of similar ORR trends for nivo or ipi monotherapy vs. the combo across other tumor types • Melanoma data supported the biological rationale that the combination may be beneficial
CheckMate 227[12] Metastatic NSCLC (First line) 2020	1:1:1 ratio (PD-L1 ≥1%): • Nivo + ipi (combo) • Nivo • Chemo (SOC)	**mOS** Combo = 17.1 mos (15.0-20.1) Nivo = 15.7 mos (13.3-18.1) SOC = 14.9 mos (12.7-16.7) HR for combo vs. SOC = 0.79 (0.65-0.96); p=0.007	• No ipi only arm • Primary comparisons were made to the SOC arm (chemo – nivo not approved at time of trial design) • Supported by data across tumor types, including data used in review of melanoma and RCC applications

Trial {Primary Source} **Cancer type (Setting)** **Year of FDA Approval**	**Trial arms**	**Findings** *(Confidence Intervals [CI] included in parenthesis)**	**Trial Design Takeaways**
CheckMate9LA[13] Metastatic NSCLC (First line) 2020	1:1 ratio: • Nivo + ipi + 2 cycles chemo (combo) • Chemo (SOC)	**mOS** Combo = 15.8 mos (13.9-19.7) SOC = 11.0 mos (9.5-12.7) HR = 0.74 (0.62-0.87); p<0.001	• Assessed whether reducing the amount of chemo with addition of nivo + ipi enhanced benefit • Trials showing no improvement in OS with addition of nivo alone to chemo already completed • Contribution of ipi was adequately established based on available data from CheckMate-227 and across other tumor types
SWOGS1616[14] Unresectable stage III or IV melanoma (After progression on anti-PD-(L)1) N/A	3:1 ratio: • Nivo + ipi (Combo) • Ipi	**OS** HR = 0.63 (0.41-0.97); p=0.036	• At this point, anti-PD-(L)1 was first-line therapy and there were ethical concerns about including an anti-PD-(L)1 alone arm after progression
CheckMate-9DW[15] Unresectable HCC (First line) 2025	1:1 ratio: • Nivo + ipi • Investigator's choice of lenvatinib or sorafenib (SOC)	**mOS** Combo = 23.7 mos (18.8-29.4) SOC = 20.6 mos (17.5-22.5) HR = 0.79 (0.65-0.96); p=0.018	• Other settings (e.g., CheckMate-067, CheckMate-214, CheckMate-227) supported not including a nivo only arm • Phase II data from CheckMate-040 further supported COE for nivo vs. combo[16,17]

*Trials differed in their definition of CI (i.e., 95-98% CI).

Nivo = nivolumab (anti-PD-1); ipi = ipilimumab (anti-CTLA-4); mPFS = median progression free survival; combo = combination; HR = Hazard ratio; mOS = median overall survival; RCC = renal cell carcinoma; NSCLC = non-small cell lung cancer; mos = months; SOC = standard of care; HCC = hepatocellular carcinoma; lenvatinib = multikinase inhibitor; sorafenib = multikinase inhibitor

BRAFi + MEKi

BRAF and MEK inhibitors (BRAFi and MEKi) are targeted therapies that disrupt signaling through the MAP kinase pathway, a key driver of tumor growth in BRAF V600–mutant cancers. BRAFi block the mutant BRAF kinase directly, halting downstream signaling and tumor proliferation, while MEKi act one step further along the cascade, overcoming resistance mechanisms that can limit the efficacy of BRAF inhibition alone. Studying these agents in combination leverages complementary mechanisms: BRAFi produce rapid tumor responses, while MEKi suppress resistance mechanisms and mitigate paradoxical MAP kinase activation. The development programs for subsequent cancer types built on the clinical proof of concept established in melanoma and demonstrated COE across multiple trials. To demonstrate the stepwise data analyses across cancer types, Table 2 outlines the various trials, outcomes, and key takeaways.

Lessons learned from other cancer types supported an understanding of COE and eventually alternative trial design approaches for BRAFi/MEKi combinations. Early clinical data, combined with biological rationale for the value of MEKi supporting BRAFi, supported initial decisions to not include the MEKi arm. In some cases, like BEACON CRC, the BRAFi arm was included, however, there were no formal comparisons between the BRAFi arm (doublet with EGFR inhibitor) and the combination arm (triplet with MEKi) for efficacy.[18] Given the similar efficacy data, the totality of data including safety and toxicity led to conclusions that there was little additional benefit from the triplet. While BREAKWATER CRC did not include a pre-specified adaptive trial design, protocol amendments allowed for the doublet arm to be dropped when it was clear there was not an efficacy benefit in that arm.[19]

When there is an unmet need, there is a particular urgency to establish effective regimens rapidly. Additionally, when the disease is rare, patient populations sizes are small, which can make it challenging to include multiple arms and randomize patients. As outlined in Table 2, trials like ROAR and CDRB436G2201 included 10-20 patients per arm due to rarity of the patient population.[20,21] Additionally, these patients often have limited treatment options, and treatments that do exist have limited efficacy, suggesting it may be appropriate to not include an SOC arm as was seen in ROAR. Because of the severity of the disease, such development approaches may rely on endpoints like ORR combined with DOR to demonstrate efficacy.

Overall Takeaways from Other Tumor Types

Early demonstrations of COE in one tumor type can serve as the foundation for expansion across multiple cancers. Initial trials developed relevant evidence, often including single-agent and SOC arms to establish COE, but over time sponsors streamlined designs by leveraging prior data and biological rationale including expected class effects. While combinations may have more durable and deeper responses than single agents, toxicity concerns and a lack of clear head-to-head analyses sometimes raise questions about risk–benefit trade-offs. Additionally, there may be differences in feedback mechanisms among cancer types. Particularly in rare or high-unmet-need populations, single-arm or small randomized trials relying on ORR and DOR can be considered. Overall, these programs show how the totality of evidence across tumor types can justify streamlined development strategies, while also highlighting the ongoing challenge of balancing efficacy, toxicity, and interpretability of COE.

Table 2. Anti-CTLA4 + anti-PD-(L)1

Trial {Primary Source} Cancer type (Setting) Year of FDA Approval	Trial arms	Findings *(Confidence Intervals [CI] included in parenthesis)**	Trial Design Takeaways
COMBI-d[22,23] BRAF V600-mutant advanced melanoma (First line) 2014	1:1 ratio: • Dabrafenib + trametinib (combo) • Dabrafenib	**mPFS** Combo = 9.3 mos Dabrafenib = 8.8 mos HR = 0.75 (0.57-0.99); p=0.03	• BRAFi were SOC before trial began • Provided a comparison of COE vs. dabrafenib to demonstrate the added value of trametinib • Study also showed a large OS benefit
COMBI-v[24] BRAF V600-mutant aNSCLC (First line) 2014	1:1 ratio: • Dabrafenib + trametinib (combo) • Vemurafenib (SOC)	**Interim OS (mOS)** Combo = 28% death (not reached) SOC = 35% death (17.2 mos) HR = 0.69 (0.53-0.89); p=0.005	• Review of COMBI-v also included data from COMBI-d as supportive data for COE assessment • Demonstrated superiority over SOC
BRF113928[25] BRAF V600-mutant aNSCLC (First line) 2014	Single arm: Dabrafenib + trametinib (combo)	**ORR** Combo = 64% (46%-79%) DOR = 10.4 mos (8.3-17.9)	• While not randomized, this study did include a cohort of patients treated with dabrafenib alone[26] • Assessment was supported by melanoma data on COE • The unmet need was a key driver that influenced the acceptance of the combination in aNSCLC in a smaller study using ORR and DOR endpoints
COLUMBUS[27] BRAF V600-mutant advanced melanoma (First line) 2018	1:1:1 ratio: • Encorafenib + binimetinib (combo) • Encorafenib • Vemurafenib (SOC)	**mPFS** Combo = 14.9 mos (11.0-18.5) Encorafenib = 9.6 mos (7.5-14.8) SOC = 7.3 mos (5.6-8.2) HR for combo vs. SOC = 0.54 (0.41-0.71); p<0.0001 HR for combo vs. encorafenib = 0.68 (0.52-0.90); p<0.001	• Historical data supported not including MEKi only arm (binimetinib) due to biological rationale and data from COMBI-d/ COMBI-v • Original findings from the study showed that combo was better than BRAFi alone (encorafenib) • Long-term OS:[28] HR for combo vs. encorafenib = 0.93 (0.73 to 1.18)
COMBI-AD[29] BRAF V600-mutant stage III melanoma (Adjuvant) 2018	1:1 ratio: • Dabrafenib + trametinib (combo) • Placebo (SOC)	**RFS (Investigator)** Combo = 38% SOC = 57% HR = 0.47 (0.39-0.58)	• Single-agent efficacy of dabrafenib in metastatic melanoma already demonstrated when the trial began • No SOC in adjuvant setting before this trial • Natural history supported the use of placebo as comparator

Trial {Primary Source} Cancer type (Setting) Year of FDA Approval	Trial arms	Findings *(Confidence Intervals [CI] included in parenthesis)**	Trial Design Takeaways
BEACON CRC[18] BRAF V600-mutant metastatic CRC (Second or third line) 2020	1:1:1 ratio: • Encorafenib, cetuximab, binimetinib (triplet) • Encorafenib + cetuximab (doublet) • Chemo (SOC)	**mOS** Triplet = 9.0 mos (8.0-11.4) Doublet = 8.4 mos (7.5-11.0) SOC = 5.4 mos (4.8-6.6) HR for triplet vs. SOC = 0.52 (0.39-0.70); p<0.001 HR for doublet vs. SOC = 0.60 (0.45-0.79); p<0.001	• The design did not have a formal comparison of triplet vs. doublet arms • The results showed that the doublet was sufficient and that the triplet did not provide a significant additional benefit but did add more adverse events • Ultimately, doublet was the approach moving forward
BREAKWATER CRC[19] BRAF V600-mutant metastatic CRC (First line) 2024	1:1:1 ratio: • Encorafenib + Cetuximab + mFOLFOX6 (doublet + SOC) • Encorafenib + Cetuximab (doublet) • Chemotherapy with or without bevacizumab (SOC)	**mPFS** Doublet + SOC = 12.8 mos (11.2-15.9) SOC = 7.1 mos (6.8-8.5) HR for doublet + SOC vs. SOC = 0.53 (0.41-0.68); p<0.001	• Protocol amendment led to closing the doublet arm early due to interim efficacy showing superiority of the doublet + SOC arm • Not an adaptive design (i.e., not prespecified) but resembled an adaptive approach
BRF117019 (ROAR)[20] BRAF V600-mutant rare cancers (mixed) 2022[30]	Single arm basket trial: Dabrafenib + trametinib (combo)	**ORR (Investigator)** ATC = 56% (38.1%-72.1%) BTC = 53% (37.7%-68.8%) ASI = 67% (9.4%-99.2%) LGG = 54% (25.1%-80.8%) HGG = 33% (20.0%-49.0%) HCL = 89% (77.8%-95.9%) MM = 50% (18.7%-81.3%)	• Rare tumors with no effective standard of care justified the use of a single arm and ORR as primary endpoint • This approach was considered acceptable due to extensive prior experience with the combo vs. monotherapy in other tumor types demonstrating COE • Demonstrated high and durable response rates

Trial {Primary Source} **Cancer type (Setting)** **Year of FDA Approval**	**Trial arms**	**Findings** *(Confidence Intervals [CI] included in parenthesis)*****	**Trial Design Takeaways**
CDRB436G2201[21] BRAF V600-mutant pediatric glioma (First line) 2023	2:1 ratio • Dabrafenib + trametinib (combo) • Chemotherapy (SOC)	**ORR (Central)** Combo = 47% (35%-59%) SOC = 11% (3%-25%) Odds ratio (CI)= 7.19 (2.30-22.40)	• Data from prior clinical trials of dabrafenib and trametinib as single agents in pediatric patients with glioma supported trial design • Body of data across other BRAF V600E mutation-positive tumor types also supported combo vs. SOC

* Trials differed in their definition of CI (i.e., 95-98% CI).

Dabrafenib = BRAFi; trametinib = MEKi; vemurafenib = BRAFi; encorafenib = BRAFi; binimetinib = MEKi; cetuximab = EGFRi; mFOLFOX6 = modified folinic acid + fluorouracil + oxaliplatin; combo = combination; SOC = standard of care; mos = months; mPFS = median progression free survival; mOS = median overall survival; RFS = recurrence free survival; ORR = objective response rate; HR = Hazard ratio; RCC = renal cell carcinoma; NSCLC = non-small cell lung cancer; CRC = colorectal cancer; ATC = anaplastic thyroid cancer; BTC = biliary tract cancer; ASI = adenocarcinoma of the small intestine; LGG = low-grade glioma; HGG = high-grade glioma; HCL = hairy cell leukemia; MM = multiple myeloma

Drugs with Similar MOA

Targeting EGFRm in aNSCLC

Tyrosine kinase inhibitors (TKIs) targeting EGFR mutations (EGFRm) were introduced in the early 2000s and iterative improvements to drug design resulted in better outcomes for patients treated with TKI monotherapy over the next two decades. The evolution of targeting EGFRm provided a multitude of studies that led to confidence in the biology of EGFRm aNSCLC and a clearer understanding of the TKIs. Subsequent combination studies that incorporated EGFRm targeting agents used data to support decisions about trial designs.

The MARIPOSA[31] trial assessed patients with untreated EGFRm aNSCLC who were randomized in a 2:2:1 fashion to first-line treatment with amivantamab + lazertinib (combo): osimertinib (SOC): lazertinib (unapproved monotherapy with similar outcomes expected as compared to SOC). The lazertinib monotherapy arm was unpowered and comparisons between the lazertinib monotherapy arm and the other arms were included as descriptive analyses with any statistical comparisons included as secondary or exploratory. The combination arm showed a mPFS (CI) of 23.7 (19.1-27.7) months compared to 16.6 (14.8-18.5) months with osimertinib (HR = 0.70 [0.58-0.85]; p<0.001). The mPFS for the lazertinib only arm was 18.5 (14.8-20.1) months.

This setting was an add on design of amivantamab to an EGFR TKI compared to osimertinib, another EGFR TKI, where outcomes were anticipated to be similar between the two EGFR TKIs. The SOC at the time was osimertinib and when designing the trial, it was appreciated that patients and providers would not forgo a third generation EGFR TKI (i.e., osimertinib) as first-line treatment to receive amivantamab alone. The lazertinib single agent arm was necessary because lazertinib was not approved in the U.S. as a single agent and had not been evaluated against osimertinib in a phase 3 trial. To show that amivantamab was necessary, the trial needed to demonstrate that the effect of lazertinib alone on efficacy endpoints was consistent with the effect osimertinib. In this scenario, the contribution of amivantamab to the efficacy observed with the combination could be projected from the demonstration of superiority of the combination compared to osimertinib, with supportive data provided by informal comparison between the combination and lazertinib alone arms. This approach was supported by data from a previous phase 3 trial comparing lazertinib to gefitinib as first-line treatment for patients with EGFR-mutated NSCLC, which demonstrated outcomes for lazertinib were similar to those reported with osimertinib.[32]

While the 2:2:1 allocation limited the size of the unapproved lazertinib arm, there were over 1000 patients in the trial, allowing for a descriptive efficacy characterization without diluting power from the main combination vs. SOC comparison. The trial was powered for the analysis comparing PFS for the combination vs. SOC and no formal alpha was spent on combination vs. lazertinib comparisons, avoiding the consequence of inflating sample size for multiple primary hypotheses. Once the combination was demonstrated to be superior to SOC, the comparison of the combination vs. lazertinib was performed by examining the magnitude of effect and confidence intervals rather than a pre-specified p-value. Given the resultant approval, the non-powered analysis was deemed to provide sufficient descriptive findings that support COE for monotherapy.

The MARIPOSA-2 trial also employed a 2:2:1 (amivantamab + lazertinib + chemotherapy [combination]: chemotherapy [SOC]: amivantamab + chemotherapy [experimental monotherapy + SOC]) design when treating patients with EGFRm aNSCLC who had progressed on osimertinib. The trial was originally designed to include the comparison of the combination to SOC for assessing COE but was later updated to include

dual hypothesis testing due to emerging Phase I data demonstrating the benefit of amivantamab + chemotherapy. The study reported PFS effects with HRs of 0.44 and 0.48, respectively, for the combination and experimental monotherapy arms compared with SOC.[33] Operational characteristic of this less stringent (i.e., unequal randomization ratio) approach should be carefully examined at the design stage.

Overall Takeaways on Drugs with Similar MOA

Therapies with similar MOA may sometimes leverage data from earlier trials suggesting similar efficacy outcomes to support COE and enable alternative registrational trial designs. Iterative improvements to therapies over time establish strong data about the biology of the MOA and associations with outcomes, providing confidence in the biology and mechanism that inform combination trials. With strong background about the MOA, designs with appropriately calibrated traditionally less stringent evidentiary thresholds may be considered to provide findings that support COE for monotherapy and help overcome ethical and practical concerns without diluting power. Across these settings, leveraging prior data can allow for efficient trial designs that preserve power for primary comparisons while providing descriptive evidence of the added value of combination therapies.

Other Trials in the Same Cancer Setting

Anti-P(D)-L1 + nectin-4 antibody drug conjugate (ADC) in urothelial carcinoma (UC)

Data from early UC clinical trials with monotherapy anti-PD-(L)1 or nectin-4 ADC arms demonstrated that each exhibited benefit as monotherapies. The biological rationale for the potential additive benefit of these two therapies together was based on high expression of nectin-4 in UC. Targeting UC with nectin-4 ADCs may increase neoantigen presentation in some patients who do not respond to anti-PD-(L)1 monotherapy, theoretically turning "cold" tumors "hot" and allowing for anti-PD-(L)1 to function in larger populations. COE data were leveraged from previously completed randomized control trials in advanced UC to build the data over time:

- *Pembrolizumab alone:* KEYNOTE-052 was a phase II trial that assessed pembrolizumab alone (ORR = 24% [19%-27%]).[34]
- *Pembrolizumab vs. chemotherapy:* KEYNOTE-045 was a phase III trial assessing pembrolizumab (mOS = 10.1 mos [8.0-12.3]) vs. chemotherapy (mOS = 7.3 months [6.1-8.1]) and a HR (CI) = 0.70 (0.57-0.85); p<0.001.[35]
- *Enfortumab vedotin alone*: EV-201 was a phase II single arm trial that assessed enfortumab vedotin monotherapy after progression on platinum chemotherapy with an anti-PD-(L)1 (confirmed ORR = 44% (CI 35.1% - 53.2%; mDOR = 7.6 months [3.6-11.3]).[36]
- *Enfortumab vedotin vs. chemotherapy*: EV-301 was a phase II trial that assessed enfortumab vedotin (mOS = 12.9 months [11.01-14.92]) vs. chemotherapy (mOS = 8.94 months [8.25-10.25]) with HR (CI) = 0.704 (0.581-0.852); p=0.00015.[36]
- *Enfortumab vedotin vs. enfortumab vedotin + pembrolizumab*: EV-103 was a phase II trial that assessed enfortumab vedotin (ORR = 45.2% [33.5-57.3]) vs. enfortumab vedotin with pembrolizumab (ORR = 64.5% [52.7%-75.1%]).[37]

In the registrational trial EV-302/KEYNOTE-A39,[38] patients with untreated advanced UC were treated with pembrolizumab + enfortumab vedotin or chemotherapy (mPFS = 12.5 months [10.4-16.6] vs. 6.3 months

[6.2-6.5]) with HR (CI) = 0.45 (0.38-0.54); p<0.001. FDA approved the combination for patients with advanced UC in 2023.[39]

Head-to-head comparisons of each component to the combination and each other were assessed in individual clinical trials rather than being analyzed all at once. These previous trials set the foundation for COE of the components and led to a 2-arm trial as the alternative registrational trial.

VEGFi + PD-(L)1 in advanced Renal Cell Carcinoma (aRCC)

aRCC is a highly vascular disease and VEGF signaling is central to disease progression. To overcome this, TKIs targeting the VEGF receptor (VEGFi) were established as SOCs in in the 2000s. aRCC is also immunogenic and responsive to immune-based therapies, so it was hypothesized that there may be synergy between VEGFi and anti-PD-(L)1 to enhance immune infiltration and T-cell activity. Early phase studies showed promising response rates and acceptable safety initially as monotherapies and later as a combination. As such, to demonstrate COE, the registrational trials used previous data to justify not including the monotherapy arms.

We include the following list to demonstrate the multitude of studies using previously completed trials to support COE in this space:

- CheckMate 9ER compared nivolumab plus cabozantinib (mPFS = 16.6 months [12.8-19.5]) vs. sunitinib (mPFS = 8.4 months [7.0-9.7]), demonstrating a HR (CI) = 0.59 (0.49-0.71).[40]
- CLEAR (KEYNOTE-581) compared lenvantinib plus pembrolizumab (mPFS = 23.9 months [20.8-27.2]) vs. sunitinib (mPFS = 9.2 months [6.0-11.0]), demonstrating a HR (CI) = 0.39 (0.32-0.49); p<0.001. Additionally, it compared lenvantinib plus everolimus (mPFS = 14.7 months [11.1-16.7]) vs. sunitinib, demonstrating a HR (CI) = 0.65 (0.53-0.80); p<0.001.[41]
- Among patients with PD-L1 positive tumors, JAVELIN Renal 101 compared avelumab plus axitinib (mPFS =13.8 months [11.1-not estimable]) vs. sunitinib (mPFS = 7.2 months [5.7-9.7]) demonstrating a HR (CI) = 0.61 (0.47-0.79); p<0.001.
- KEYNOTE-426 compared pembrolizumab plus axitinib (mPFS = 15.1 months [12.6-17.7]) vs. sunitinib (mPFS = 11.1 months (12.6-17.7)), demonstrating a HR (CI) = 0.69 (0.57-0.84); p<0.001.[42]

There were multiple sponsors with programs in this space and rather than relying on data from other programs, each used data from their own internal programs to demonstrate COE.

Overall Takeaways on Other Trials in the Same Cancer Setting

Early-phase studies may suggest that combining two agents could improve response rates beyond what either could achieve alone. In settings where an increase in response rates is known to translate into OS benefit registrational trials can subsequently focus on the comparison between the combination and SOC, omitting active monotherapy arms in part because each agent already had established single-agent activity. To demonstrate COE, the relied-on trials should ideally be randomized, include a sufficient sample size of patients treated with monotherapy compared with the combination, and demonstrate a strong indication of response rate. The absence of formal monotherapy comparators means that COE is inferred rather than directly measured, highlighting the trade-offs between trial efficiency and the ability to fully characterize the added value of each component.

One Monotherapy Arm is Inactive

KRASmuti + EGFRi in metastatic Colorectal Cancer (mCRC)

KRAS mutations are prevalent in refractory CRC, and treating with KRAS mutation inhibitors (KRASmuti) can prevent growth by blocking downstream signaling through the MAP kinase pathway. CRC tumors use the compensatory mechanism of re-activating EGFR in response to KRAS inhibition, suggesting using an EGFRi may overcome resistance. Preclinical studies showed EGFR signaling allowed cancer cells to evade KRAS inhibition, so adding EGFRi addresses a key resistance mechanism.[43] Additionally, modeling guided the rational selection of combination therapies to overcome resistance mechanisms like EGFR feedback activation.

Clinical trial data assessing cetuximab in KRAS mutant CRC demonstrated that cetuximab alone was ineffective (i.e., no difference in mOS between cetuximab vs. SOC). Additionally, mechanistic studies demonstrated that the KRASmut confers resistance to EGFRi monotherapy. As a result, the EGFRi cetuximab carries a warning that treatment could lead to increased tumor progression, increased mortality, or lack of benefit for the patient population with KRASmut CRC.[44]

As such, KRYSTAL-1 was a phase 1/2 trial that tested adagrasib (KRASmuti) and cetuximab (EGFRi) vs. adagrasib in refractory KRAS G12C mutated CRC and did not include a cetuximab monotherapy arm.[45] ORR for the combination was 34% with all partial responses and mDOR of 5.8 months, while the single arm adagrasib showed ORR of 19% and mDOR of 4.3 months. The combination was approved under accelerated approval without a phase 3 study. Subsequently, KRYSTAL-10 was a phase 3 trial that included two arms: the combination vs. chemotherapy.[45] CodeBreaK 300 also assessed sotarasib (KRASmuti at two different doses) and panitumumab (EGFRi) vs. investigator's choice (SOC) and the combination was subsequently approved.[46] Pre-clinical and clinical data demonstrated a mechanistic understanding and subsequent lack of activity with EGFRi alone, leading to a registrational trial that did not include an EGFRi monotherapy.

Anti-SLAMF7 + Immunomodulary Agents (IMiDs) + Dexamethasone in Multiple Myeloma (MM)

Anti-SLAMF7 antibodies combined with IMiDs and dexamethasone enhance complementary mechanisms of action. Anti-SLAMF7 marks MM cells for immune-mediated destruction and simultaneously activates natural killer (NK) cells. IMiDs further potentiate this effect by stimulating NK cell and T-cell activity, increasing cytokine production, and sensitizing myeloma cells to immune attack. Elotuzumab, an anti-SLAMF7 monoclonal antibody, demonstrated limited single-agent efficacy in relapsed/refractory MM as demonstrated by early-phase studies that reported ORR near 0%. This lack of efficacy was supported by clinical data demonstrating insufficient natural killer (NK) cell activation when used alone, as expected based on the MOA. Consequently, subsequent registrational trials, including ELOQUENT-2 and ELOQUENT-3, evaluated elotuzumab exclusively in combination with IMiDs (i.e., lenalidomide or pomalidomide) and dexamethasone, as a monotherapy was deemed unlikely to provide clinically meaningful benefit.[47,48] Again, pre-clinical MOA data and early clinical trial data supported not including the elotuzumab monotherapy arm for the registrational trial to demonstrate COE.

Overall Takeaways when One Monotherapy Arm is Inactive

Early pre-clinical and clinical data from clinical trials that demonstrate a single agent has very limited or no efficacy can support not investigating that agent as monotherapy in a registrational study to demonstrate COE. Preclinical data can support mechanistic understandings of single agent vs. combination activity and

limit the patients who may be exposed to poorly effective monotherapies. Clear biological rationale, supported by non-clinical and clinical data, can demonstrate that a single drug is inactive and thus support not including such a monotherapy arm as a comparator for COE in the registrational trial.

Additional Considerations

Above, we outline key considerations for determining COE, but it is critical to recognize additional considerations and settings.

Safety as Part of the Totality of Evidence

While the FDA draft guideline and this white paper do not address safety, for approvals, evidence needs may vary depending on the safety profile of the drug. Additionally, the value of a component is judged not just by its efficacy but also by its toxicity level. Patient reported outcomes can support an understanding of toxicity that includes the patient's perspective, rather than just relying on clinical safety endpoints.

Contribution of Phase

As therapies move into earlier stages of disease, they are often provided either in the neoadjuvant (i.e., before surgery) or adjuvant (i.e., after surgery) setting. Many times, therapies are provided in both, however, whether both are necessary may be unknown. Like combination therapies, understanding the contribution of efficacy of each phase is important to prevent overtreating and may necessitate a factorial design comparing treatment in adjuvant only, neoadjuvant only, both neoadjuvant, or none. This approach is further complicated when the therapy under investigation is given in combination, and a demonstration of COE is needed. This white paper provides approaches for considering alternative trial designs that could be pulled through into contribution of phase trials, specifically the need for robust evidence before adjusting the factorial approach.

Complex Combinations with Multiple Components

Demonstrating the individual contribution of each drug in a triplet or higher-order combination can be highly complex, particularly when the SOC is already a combination regimen. One common approach is to build on an existing doublet by adding a third agent, but questions often arise regarding whether all components must be evaluated individually or if some can be substituted. As combinations become more complex, it may not always be feasible or necessary to evaluate every component separately, though careful consideration in trial design is essential. Additional factors, such as market utility, also play a role: a triplet may receive approval but see limited adoption if it does not demonstrate clear superiority or practical advantage over simpler regimens. This challenge is further compounded in biomarker-selected subgroups, where a SOC may not yet exist.

Conclusions

The development of combination therapies in oncology continues to rapidly evolve, requiring a balance between methodological rigor, feasibility, and patient need. While factorial designs remain the methodological gold standard for demonstrating COE, alternative approaches supported by prior data can be considered if they are scientifically sound and operationally practical. Sponsors should engage regulatory agencies as they build strategies to demonstrate COE. The scenarios outlined in this white paper highlight the importance of carefully justifying trial design choices, prespecifying decision rules, and leveraging the totality of available evidence to support regulatory and clinical confidence in the benefit of the combination

beyond that of each monotherapy. Going forward, clearer consensus on evidentiary standards, greater use of adaptive and innovative trial designs, and thoughtful integration of early endpoints and modeling may enable more efficient development while maintaining scientific integrity. Continued dialogue between sponsors, regulators, clinicians, and patients will be essential to refine expectations for COE, ensure trial designs support the delivery of safe and effective therapies, and accelerate the availability of effective, safe, and meaningful combination treatments.

References

1. U.S. Food and Drug Administration. Development of Cancer Drugs for Use in Novel Combination - Determining the Contribution of the Individual Drugs' Effects (Draft Guidance, July 2025). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-cancer-drugs-use-novel-combination-determining-contribution-individual-drugs-effects
2. Korn EL, Freidlin B. Non-Factorial Analyses of Two-by-Two Factorial Trial Designs. Clin Trials 2016;13:651–659. DOI: 10.1177/1740774516659472.
3. Attard G, Murphy L, Clarke NW et al. Abiraterone Acetate Plus Prednisolone with or without Enzalutamide for Patients with Metastatic Prostate Cancer Starting Androgen Deprivation Therapy: Final Results from Two Randomised Phase 3 Trials of the STAMPEDE Platform Protocol. Lancet Oncol 2023;24:443. DOI: 10.1016/S1470-2045(23)00148-1
4. Korn EL, Allegra CJ, Freidlin B. Phase III Evaluation of Treatment Combinations in Three-Arm Trials. J Clin Oncol 2025;43:226–233. DOI: 10.1200/JCO-24-01476.
5. Parmar MKB, Sydes MR, Morris TP. How Do You Design Randomised Trials for Smaller Populations? A Framework. BMC Med 2016;14:1–10. DOI: 10.1186/s12916-016-0722-3.
6. FDA. Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Product (Draft Guidance, December 2019). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-evidence-effectiveness-human-drug-and-biological-products
7. Renfro LA, Ji L, Piao J et al. Trial Design Challenges and Approaches for Precision Oncology in Rare Tumors: Experiences of the Children's Oncology Group. JCO Precis Oncol 2019;3:1–13. DOI: 10.1200/PO.19.00060.
8. Turner DC, Wada R, Zhou H et al. Model-Based Meta-Analysis of Non-Small Cell Lung Cancer with Standard of Care PD-1 Inhibitors and Chemotherapy for Early Development Decision Making. CPT Pharmacometrics Syst Pharmacol 2023;12:1751–1763. DOI: 10.1002/psp4.12917.
9. Larkin J, Chiarion-Sileni V, Gonzalez R et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med 2015;373:23–34. DOI: 10.1056/NEJMoa1504030.
10. Wolchok JD, Chiarion-Sileni V, Rutkowski P et al. Final, 10-Year Outcomes with Nivolumab plus Ipilimumab in Advanced Melanoma. N Engl J Med 2025;392:11–22. DOI: 10.1056/NEJMoa2407417.
11. Motzer RJ, Tannir NM, McDermott DF et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. N Engl J Med 2018;378:1277–1290. DOI: 10.1056/NEJMoa1712126.
12. Hellmann MD, Paz-Ares L, Bernabe Caro R et al. Nivolumab plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer. N Engl J Med 2019;381:2020–2031. DOI: 10.1056/NEJMoa1910231.
13. Paz-Ares L, Ciuleanu TE, Cobo M et al. First-Line Nivolumab Plus Ipilimumab Combined with Two Cycles of Chemotherapy in Patients with Non-Small-Cell Lung Cancer (Checkmate 9LA): An International, Randomised, Open-Label, Phase 3 Trial. Lancet Oncol 2021;22:198–211. DOI: 10.1016/S1470-2045(20)30641-0.

14. VanderWalde A, Bellasea SL, Kendra KL et al. Ipilimumab with or without Nivolumab in PD1/PDL1 Blockade Refractory Metastatic Melanoma: A Randomized Phase 2 Trial. Nat Med 2023;29:2278. DOI: 10.1038/s41591-023-02498-y.
15. Yau T, Galle PR, Decaens T et al. Nivolumab Plus Ipilimumab Versus Lenvatinib or Sorafenib as First-Line Treatment for Unresectable Hepatocellular Carcinoma (Checkmate 9DW): An Open-Label, Randomised, Phase 3 Trial. The Lancet 2025;405:1851–1864. DOI: 10.1016/S0140-6736(25)00403-9.
16. Yau T, Kang YK, Kim TY et al. Efficacy and Safety of Nivolumab Plus Ipilimumab in Patients with Advanced Hepatocellular Carcinoma Previously Treated with Sorafenib: The CheckMate 040 Randomized Clinical Trial. JAMA Oncol 2020;6. DOI: 10.1001/JAMAONCOL.2020.4564.
17. Study Details | NCT01658878 | An Immuno-therapy Study to Evaluate the Effectiveness, Safety and Tolerability of Nivolumab or Nivolumab in Combination with Other Agents in Patients with Advanced Liver Cancer. Available from: https://clinicaltrials.gov/study/NCT01658878.
18. Kopetz S, Grothey A, Yaeger R et al. Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. N Engl J Med 2019;381:1632–1643. DOI: 10.1056/NEJMoa1908075.
19. Elez E, Yoshino T, Shen L et al. Encorafenib, Cetuximab, and mFOLFOX6 in BRAF-Mutated Colorectal Cancer. N Engl J Med 2025;392:2425–2437. DOI: 10.1056/NEJMoa2501912.
20. Subbiah V, Kreitman RJ, Wainberg ZA et al. Dabrafenib plus Trametinib in BRAFV600E-mutated Rare Cancers: The Phase 2 ROAR Trial. Nat Med 2023;29:1103–1112. DOI: 10.1038/s41591-023-02321-8.
21. Bouffet E, Hansford JR, Garrè ML et al. Dabrafenib plus Trametinib in Pediatric Glioma with BRAF V600 Mutations. N Engl J Med 2023;389:1108–1120. DOI: 10.1056/NEJMoa2303815.
22. Long G V., Stroyakovskiy D, Gogas H et al. Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. N Engl J Med 2014;371:1877–1888. DOI: 10.1056/NEJMoa1406037.
23. Long G V., Stroyakovskiy D, Gogas H et al. Dabrafenib and Trametinib versus Dabrafenib and Placebo for Val600 BRAF-Mutant Melanoma: A Multicentre, Double-Blind, Phase 3 Randomised Controlled Trial. The Lancet 2015;386:444–451. DOI: 10.1016/S0140-6736(15)60898-4.
24. Robert C, Karaszewska B, Schachter J et al. Improved Overall Survival in Melanoma with Combined Dabrafenib and Trametinib. N Engl J Med 2015;372:30–39. DOI: 10.1056/NEJMoa1412690.
25. Planchard D, Smit EF, Groen HJM et al. Dabrafenib plus trametinib in patients with previously untreated BRAFV600E-mutant metastatic non-small-cell lung cancer: an open-label, phase 2 trial. Lancet Oncol 2017;18:1307–1316. DOI: 10.1016/S1470-2045(17)30679-4.
26. HIGHLIGHTS OF PRESCRIBING INFORMATION for TAFINLAR (dabrafenib). Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2025/202806s038,217514s009lbl.pdf.
27. Dummer R, Ascierto PA, Gogas HJ et al. Encorafenib Plus Binimetinib versus Vemurafenib or Encorafenib in Patients with BRAF-mutant Melanoma (COLUMBUS): A Multicentre, Open-label, Randomised Phase 3 Trial. Lancet Oncol 2018;19:603–615. DOI: 10.1016/S1470-2045(18)30142-6.
28. Schadendorf D, Dummer R, Flaherty KT et al. COLUMBUS 7-Year Update: A Randomized, Open-Label, Phase III Trial of Encorafenib Plus Binimetinib Versus Vemurafenib or Encorafenib in Patients with BRAF V600E/K-Mutant Melanoma. Eur J Cancer 2024;204. DOI: 10.1016/j.ejca.2024.114073.
29. Long G V., Hauschild A, Santinami M et al. Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma. N Engl J Med 2017;377:1813–1823. DOI: 10.1056/NEJMoa1708539.

44. HIGHLIGHTS OF PRESCRIBING INFORMATION for ERBITUX (cetuximab). Available at https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/125084s279lbl.pdf.

45. Yaeger R, Weiss J, Pelster MS et al. Adagrasib with or without Cetuximab in Colorectal Cancer with Mutated KRAS G12C. New England Journal of Medicine 2023;388:44–54. DOI: 10.1056/NEJMoa2212419.

46. Pietrantonio F, Salvatore L, Esaki T et al. Overall Survival Analysis of the Phase III CodeBreaK 300 Study of Sotorasib Plus Panitumumab Versus Investigator's Choice in Chemorefractory KRAS G12C Colorectal Cancer. Journal of Clinical Oncology 2025;43:2147–2154. DOI: 10.1200/JCO-24-02026.

47. Lonial S, Dimopoulos M, Palumbo A et al. Elotuzumab Therapy for Relapsed or Refractory Multiple Myeloma. N Engl J Med 2015;373:621–631. DOI: 10.1056/NEJMoa1505654.

48. Dimopoulos MA, Dytfeld D, Grosicki S et al. Elotuzumab Plus Pomalidomide and Dexamethasone for Relapsed/Refractory Multiple Myeloma: Final Overall Survival Analysis from the Randomized Phase II ELOQUENT-3 Trial. J Clin Oncol 2023;41:568–578. DOI: 10.1200/JCO.21.02815.

FRIENDS of CANCER RESEARCH

Multi-Regional Clinical Trials: Addressing Standard of Care Variability

White Paper | 2025

Executive Summary

Global multi-regional clinical trials (MRCTs) in oncology accelerate access to new therapies, improve the diversity and generalizability of clinical data, and enable more efficient regulatory review across regions. A central challenge in MRCT design is selecting an appropriate standard of care (SOC) comparator, which anchors interpretation of a new therapy's efficacy and safety relative to existing treatments. SOC can vary widely across regions due to differences in regulatory approvals, clinical guidelines, real-world practice, access, and reimbursement. It is often not a single treatment but a range of acceptable, context-dependent options that evolves over time as new evidence emerges, presenting two key challenges for trial design: (1) SOC may shift during an ongoing trial and (2) multiple SOCs may exist simultaneously, complicating selection of a single comparator.

Beyond scientific complexity, comparator selection raises ethical and operational considerations. Patients and investigators must view the control arm as acceptable and relevant to current practice; otherwise, trials risk poor enrollment or high dropout rates. Trial sponsors must therefore take a thoughtful approach to comparator selection that balances scientific rigor, ethical integrity, and global feasibility. Friends of Cancer Research (*Friends*) convened a multi-stakeholder working group including experts from the U.S. Food and Drug Administration (FDA), pharmaceutical companies, academia, and patient advocacy to identify key considerations and potential strategies for selecting and justifying comparators in oncology MRCTs.

Key Considerations:

- Comparator arms should reflect clinically meaningful standards and not be inferior to therapies demonstrating clinical benefit.
- Strategy for comparator selection and design should evaluate possible SOC evolution during trial planning and conduct.
- Factors such as regulatory approvals, clinical guidelines, real-world use, feasibility, reimbursement, and patient and clinician preferences should all be considered.
- Balancing regional applicability, particularly in the U.S., with global feasibility is crucial, as SOC in one region may not be approved or practical in others.

Strategic Approaches:

- Predefine acceptable options (e.g., investigator's choice or regional controls) when multiple SOCs exist.
- Anticipate SOC changes as much as possible and pre-specify limited design adaptations or supplementary cohorts if needed.
- Use descriptive analyses, embedded cohorts, or real-world data (RWD) to contextualize findings when new SOCs emerge mid-trial.
- Document comparator rationale and engage early with regulators to ensure scientific and ethical acceptability.

This white paper outlines considerations to guide trial sponsors, from defining the patient population and SOC options to evaluating feasibility, ethics, timing risks, and regulatory input throughout the process. These are not intended as a strict roadmap but as flexible considerations to support alignment and transparency in MRCT design. While perfect solutions may not be attainable, a thoughtful and ongoing process can improve applicability and transparency in comparator selection, ensuring trials remain feasible, meaningful, and representative across regions.

Authors

Terrell Baptiste, Gilead Sciences

Judd Englert, Amgen

Carsten Goessl, GSK

Stacey Kalambakas, Bayer AG

Sumithra Mandrekar, Mayo Clinic

Kristin McJunkins, *Friends* Advisory Advocate

Amy McKee, AstraZeneca

PK Morrow, Takeda

Raymond Osarogiagbon, Baptist Cancer Center

Marie-Laure Papi, Bristol Myers Squibb

Kiran Patel, Johnson & Johnson

Julia Perkins Smith, Eli Lilly and Company

Susan Rouse, *Friends* Advisory Advocate

Harpreet Singh, Precision for Medicine

Mark Stewart, Friends of Cancer Research

Kerry Wentworth, Nuvation Bio

Kathleen Winson, Genentech Inc

Sunita Zalani, Merck & Co., Inc.

Victoria Zazulina, Merck KGaA

This white paper was developed through discussions that included these experts and other perspectives representing academia, industry, the U.S. Food and Drug Administration, and the patient advocacy community. The views expressed here represent the collective insights from working group discussions and do not necessarily reflect the official positions of any individual organization.

Table of Contents

Introduction

Global multi-regional clinical trials (MRCTs) in oncology drug development are commonly used to support marketing applications across multiple regulatory jurisdictions. Studying new medicines and regimens in MRCTs has the potential to improve the generalizability of results, accelerate global drug development, and support more efficient regulatory review. Both the International Council for Harmonisation (ICH) E17 guideline and the U.S. Food and Drug Administration's (FDA) guidance on clinical evidence generation from oncology MRCTs underscore the value of MRCTs in establishing efficacy and safety across diverse patient populations and geographic regions.[1,2]

Selecting an appropriate comparator arm remains one of the most challenging aspects of MRCT design. Standards of care (SOC) can vary significantly across and within regions due to differences in regulatory approvals, clinical guidelines, real-world practice, access, and reimbursement. In some settings, more than one regimen may reasonably be considered SOC, with 'gray zones' that reflect the realities of clinical decision-making for individual patients. Additionally, the therapies most used in practice may differ from those formally approved by regulatory agencies or recommended in clinical guidelines, particularly when access, reimbursement, or tolerability influence real-world uptake. In the context of clinical trial design and conduct, challenges in comparator selection generally reduce to two core issues: (1) the timing of new SOC adoption relative to an ongoing trial, and (2) the presence of multiple SOC options during trial conduct.

These challenges have far-reaching implications. Comparator selection is not only a scientific or regulatory issue, but also a patient-centered, ethical, and practical one, as trial credibility and feasibility ultimately rests on whether patients and their treating physicians view the comparator as acceptable and relevant to their treatment. A comparator must maintain clinical equipoise and be one that patients are willing to receive when randomized; otherwise, trials risk poor enrollment, high dropout rates, or ultimately becoming infeasible to enroll.[3] Comparator regimens should also allow for clear isolation of the contribution of phases and/or new products in investigational regimens that have multiple phases of treatment and/or combinations of products across one or more treatment phases (e.g., combination regimens with both neoadjuvant and adjuvant phases). Regulators have increasingly emphasized that control regimens should be therapies demonstrating substantial clinical benefit, even when access may be regionally delayed. Comparator selection directly affects the feasibility, ethical integrity, and interpretability of global oncology trials.

Without workable solutions to address SOC variability, some global MRCTs may not be pursued at all — limiting opportunities for patients to access novel therapeutics and for researchers to generate robust evidence across diverse populations. In other cases, trials may proceed but be highly impacted by skewed regional representation, which increasingly shapes submission and registration discussions. This reality underscores both the urgency of the problem and the need for predictable pathways to guide comparator selection across evolving SOC landscapes.

Scope

This white paper focuses on comparator regimen selection in MRCTs, with an emphasis on oncology trials intended to support marketing applications in the U.S. that also enable ex-U.S. regulatory submissions. A related consideration is the expectation from many health authorities for sufficient representation of

patients from their own regions. While this intersects with comparator selection, it is not the primary focus of this paper.

Building on existing regulatory guidance, this paper expands on high-level principles by: (1) outlining key considerations for defining and justifying SOC control arm selection in a global context, (2) presenting archetypes and scenarios that illustrate common timing and multiplicity challenges, (3) exploring design strategies and solutions to support more predictable and feasible trial planning, and (4) proposing a framework and future directions for advancing regulatory and operational approaches.

Understanding and Defining Standard of Care

SOC in oncology is inherently dynamic and context dependent. Rather than a single, universally accepted regimen, SOC often spans a range of options shaped by cancer types, labeled indications, clinical guidelines, real-world uptake, patient preferences, and the practical realities of access, reimbursement, and deliverability (e.g., chimeric antigen receptor T-Cell [CAR T] or radiopharmaceutical therapies). In MRCTs, these elements are magnified by the often lengthy timelines of oncology trials, during which the treatment landscape can shift in variable and sometimes dramatic ways across regions. While not all disease areas face this challenge equally, the issue is especially pronounced in fields with rapidly evolving SOCs, where new approvals or data readouts can redefine practice within the timespan of a clinical trial. Changes to a comparator arm mid-trial are operationally challenging and highly impracticable to implement, and shifts in the landscape can impact clinical equipoise, undermine enrollment, create heterogeneity, or render results less interpretable to regulators and clinicians.

Because SOC cannot always be clearly established as a single entity, especially in fields such as oncology where multiple therapies may exist or SOCs are rapidly evolving, patient experience and acceptability should remain central. What matters is not only what is regulatory approved or guideline-recommended, but whether patients and clinicians view the treatment as acceptable and relevant, in both routine care and the context of a clinical trial, given toxicity, convenience, and perceived benefit.

Appraising Comparator Options in a Changing Landscape

Sponsors must continuously track multiple factors when selecting a global control arm—a task made harder when the landscape is evolving rapidly or a new therapy may offer transformational benefit. Key factors include:

- Recent and near-term regulatory approvals, including specific labeling language for the intended population and line of therapy
- Variation in timelines for global approvals for emerging new treatments
- Clinical guideline recommendations (e.g., NCCN, ESMO) and their evidentiary strength
- Patient and clinician preferences as reflected in routine clinical practice
- Real-world uptake (if available) and typical or shifts to treatment sequencing
- Feasibility of delivery, including sourcing, site capabilities, and infrastructure requirements, especially for emerging treatment options requiring specific site expertise
- Reimbursement and access (coverage, formulary status, logistics)

It is also important to distinguish the types of limitations that may arise when evaluating comparator options. A regimen may be infeasible when it cannot realistically be delivered within the study framework due to regulatory, ethical, or logistical barriers. It may be impractical when it is technically possible but requires disproportionate operational or infrastructure investment. Such limitations may also restrict the number and types of sites that can participate, potentially narrowing patient diversity and reducing the generalizability of results to the broader population for which the drug is intended. In some circumstances, a regimen may become unacceptable when clinical equipoise no longer exists, such as when a therapy with superior benefit has entered clinical practice. Clarifying these distinctions can help align discussions of feasibility and appropriateness across regions.

Regulators have emphasized that selecting a less-active comparator, or a less-commonly-used comparator, risks undermining trial credibility if it appears designed to exaggerate the investigational therapy's benefit. When appraising the evidentiary foundation, considerations include the approval type (e.g., traditional vs. accelerated), the magnitude and consistency of benefit (including potential class effects), study design (e.g., head-to-head vs. add-on design), the associated risks (e.g., safety, tolerability, convenience of administration), and the maturity and nature of endpoint data (e.g., overall survival [OS] vs. intermediate or surrogate endpoints). However, the more practical test is whether the chosen comparator can credibly serve as SOC for the intended population such that the comparator regimen is acceptable to patients and investigators.

Considerations in Assessing Global Trial Applicability

Recent FDA deliberations have highlighted how comparator choice can directly influence regional enrollment and, ultimately, regulatory interpretation. In one recent oncology MRCT, the selected control arm may have contributed to limited U.S. enrollment and raised questions about the applicability of results to the U.S. population. Despite meeting its primary endpoint, the resulting imbalance and inconsistent effects across regional subgroups led the advisory committee to conclude that the results were not sufficiently applicable to the U.S. patient population.[4] This example illustrates how a comparator that is scientifically reasonable but misaligned with regional practice can inadvertently limit participation and undermine applicability in key regions.

U.S. law requires that new drug approvals be supported by substantial evidence of effectiveness from adequate and well-controlled investigations and sufficient evidence to establish safety under the proposed conditions of use.[5] The statute and associated regulations do not mandate that such evidence be generated exclusively in U.S. patients, but the expectation is that trial data must be applicable to the intended patient population.

In practice, FDA has increasingly emphasized the importance of U.S. applicability—both to ensure that control arms remain relevant to current U.S. practice and to provide confidence that safety signals are adequately characterized in U.S. patients. This emphasis reflects concerns about population differences, evolving SOCs, and trial credibility, but it has also created uncertainty for sponsors when global feasibility is at odds with regional expectations.

Feasibility and Applicability Tension

A central challenge in MRCT comparator selection is balancing regional applicability (particularly in the U.S.) with global feasibility. For sponsors, approval in the U.S. is often a primary objective, and FDA expects trials

to include controls aligned with U.S. practice (Table 1). Yet most new regimens are not globally approved, consistently reimbursed, or delivered quickly enough across regions to serve as immediate global control arms, creating feasibility and operational barriers when considering study designs with a single comparator.

This tension is operational as well as regulatory: supplying control arm therapies globally, variability in site capabilities, and infrastructure gaps (e.g., with administration of CAR-T or radioligands) can make otherwise appropriate SOCs impractical to implement. Moreover, every additional comparator option or adaptation layer, including expanding control arm options using investigator's choice, adds extra statistical and regulatory risk.

Comparator selection often requires balancing scientific rigor, regional applicability, and global feasibility, recognizing that no single approach may suffice across all trials.

Table 1. Considerations and Practical Challenges in Comparator Selection

Considerations	Practical Challenges
Applicability of trial results to current SOC may become a review consideration if available therapies evolve during the study	Prolonged trial timelines mean comparators may become outdated mid-study, creating uncertainty for both design and interpretability.
New SOC is not globally available	Global differences in approval, reimbursement, and access make it difficult to ensure uniform delivery of the control therapy across regions. Sponsors must balance designing a trial that reflects regional SOC while maintaining interpretability of pooled results. In some cases, new approvals are limited to specific subpopulations, creating misalignment in what constitutes SOC for the broader disease population.
Trial control arms should not be *a priori* inferior when new therapies demonstrate clinical benefit	Determining what constitutes a substantial benefit and weighing the endpoint used may rely on cross-trial comparisons and can be subjective; implementing changes when the trial is still ongoing (either accruing or awaiting primary endpoint maturity) may be impracticable due to operational complexity, disruption of enrollment, and risks to the prespecified analysis plan.
Adaptation may be expected (e.g., investigator's choice, updated control, refinement of patient subgroups, regional applicability data) depending on approval of new therapy	Careful planning at trial design and initiation to allow for adaptations, but still requires protocol amendments, additional cohorts, reconsenting, and/or new studies, which can complicate interpretation, increase time and cost, and reduce the credibility of pooled analyses.

Archetype Scenarios

Recent experience shows how rapidly oncology SOC can evolve. For example, when KRAS G12C inhibitors were approved and became available for patients with KRAS-mutated non-small cell lung cancer, many eligible patients transitioned to these targeted therapies, affecting enrollment and the feasibility of ongoing trials that used chemotherapy-based control arms.[6] Similarly, the introduction of antibody–drug conjugates in HER2-positive breast cancer reshaped expectations for control arms within only a few years. These examples highlight the need to design MRCTs that remain interpretable and feasible even as the treatment landscape shifts.

Comparator challenges often reflect three interacting dimensions:

1. The timing of SOC change relative to an ongoing trial,
2. The magnitude and scope of the new therapy's clinical benefit, including the maturity of evidence, and
3. The resulting impact on feasibility, equipoise, and interpretability.

While adaptability remains an important design consideration, substantial protocol modifications during trial conduct are impracticable. Once initiated, trials are generally intended to answer a defined scientific question using a prespecified design. Therefore, when new therapies emerge mid-study, emphasis should be placed on preserving interpretability and contextualizing findings. In these circumstances, a totality of evidence of approach leveraging complementary data sources or analyses may be used to provide additional context and reinforce confidence at the time of readout, ensuring that results remain informative in light of evolving SOC.

The following scenarios illustrate common situations sponsors may encounter. This list is not exhaustive but reflects frequently observed cases where SOC heterogeneity complicates comparator selection.

Timing and Magnitude of SOC Change

Pending Change to SOC Before Trial Initiation

A near-term transformative approval is anticipated before or during trial initiation.

Implication: Sponsors must assess whether the planned comparator will remain credible once enrollment begins and consider pre-specifying contingency strategies. Early regulatory dialogue regarding potential contingency approaches and the acceptability of the planned comparator is important if approval appears imminent.

Transformative Therapy Emerges Mid-Trial

A new therapy demonstrating substantial OS improvement, cleaner safety profile, or simpler administration (e.g., PD-(L)1 inhibitors, KRAS G12C inhibitors, next-generation antibody-drug conjugates [ADC]) becomes available and rapidly adopted in some regions or within specific patient subgroups.

Implication: Comparator relevance may erode mid-study; whether adaptation is practicable depends on the stage of enrollment, feasibility of protocol changes, and whether the new indication overlaps with the enrolled population. Ethical and clinical pressure for crossover can increase, while enrollment may slow in regions where the new therapy is accessible. Differences in uptake across regions may also introduce heterogeneity and confound OS analyses due to varying subsequent therapy use.

Late shift in SOC During On-Going Trial

A new therapy is approved close to database lock or after primary analysis.

Implication: Late shifts are typically less disruptive operationally but may affect interpretability, labeling discussions, and the perceived relevance of results in light of current practice.

Incremental Therapy Enters the Landscape

A new regimen offering modest incremental benefit (e.g., small progression-free survival [PFS] gain or an add-on to existing therapy) becomes available.

Implication: Ethical equipoise generally remains, so comparator changes are often unnecessary once enrollment has begun. However, varying regional adoption can influence accrual and introduce modest heterogeneity, particularly if the add-on becomes common in some markets but not others.

Multiplicity of SOC Options

Multiplicity challenges can arise when several regimens or combination backbones are considered standard across or within regions. In oncology, many MRCTs use an add-on design (investigational product + SOC vs. SOC alone), where the key question is which SOC backbone(s) to include globally. These differences often reflect entrenched regional practice patterns or reimbursement structures rather than recent temporal shifts. Sponsors must balance scientific rigor, feasibility, and interpretability when determining whether to use a unified global backbone or permit regional variation.

Multiple Comparators with Similar Efficacy

Situations where more than one regimen may reasonably be considered SOC because therapies provide modest and comparable benefit to each other. This occurs frequently in later-line settings but can also arise in earlier lines, particularly in diseases with multiple approved options or combinations.

Implication: Investigator's choice may be feasible if options are functionally equivalent, though heterogeneous trial results can complicate regulatory interpretation and labeling.

Regional Asymmetry in Access or Approvals

A therapy is approved in one or a limited set of regions.

Implication: Sponsors must weigh whether to exclude certain regions, supply therapy where it is not yet available but acceptable to local health authorities, or conduct parallel or bridging studies. Regulators have signaled that lack of access in a given geography is not, on its own, sufficient justification for continuing an outdated comparator.

Additional Feasibility Considerations

Infrastructure, reimbursement, or site capability differences can also affect the feasibility of implementing certain therapies as control arms. While such cases may be uncommon, complex modalities like CAR T or radioligand therapies can illustrate how practical delivery barriers may limit their inclusion as a comparator in multi-regional settings. Anticipating these constraints early and addressing them in regional planning and comparator justification can help maintain both trial feasibility and applicability.

Design and Analytical Approaches for Maintaining Trial Applicability

Approaches to selecting and designing comparator arms in MRCTs each have distinct advantages, limitations, and feasibility implications (**Table 2**). While numerous statistical and methodological approaches exist, regulators have emphasized that their acceptability depends on context and cannot be assumed. This section outlines general principles and design options that can help maintain interpretability and relevance when standards of care evolve.

General Principles

Scientific Rigor and Trial Integrity

- Retain randomized controlled comparisons as the foundation wherever feasible.

- Consider both the magnitude and maturity of benefit when assessing whether a comparator remains appropriate (e.g., whether a PFS advantage alone is sufficient or whether OS evidence is required).
- Statistical or adaptive methods (e.g., Bayesian framework to address treatment effect heterogeneity or Sequential Multiple Assignment Randomized Trial [SMART] designs for multiple response-based treatment paths) may help address heterogeneity, but they cannot substitute for comparators that are no longer aligned with current practice.

Patient and Ethical Acceptability

- Prioritize comparators that maintain clinical equipoise and patient acceptability.
- Enrollment feasibility is a practical test of SOC acceptability.

Planning for Change and Contextualizing Evidence

- Sponsors can preemptively assess the likelihood, timing, and operational implications of an SOC change, evaluating whether anticipated shifts are imminent, regionally staggered, or likely to affect enrollment and interpretability, and align these assessments with pre-specified contingency strategies (e.g., sensitivity analyses, dual primary endpoints, bridging cohorts, or exploratory analyses among patients enrolled after a new SOC emerges).
- In practice, the feasibility of completing trial enrollment may serve as an important indicator of whether the selected comparator remains acceptable. If a study is able to enroll as planned despite the introduction of a new SOC, this may signify that the trial's comparator was appropriate and that the design continues to reflect a relevant clinical context. Conversely, significant enrollment challenges may signal that the prevailing treatment landscape has shifted and should prompt re-evaluation through discussion with regulatory authorities.
- When SOC evolves during the trial, focus on augmenting the totality of evidence. Options include:
    - Supplementary clinical trial data or RWD.
    - to benchmark outcomes under the new SOC.
    - Embedded or regional cohorts that reflect updated clinical practice without undermining the primary analysis.
    - Post hoc or sensitivity analyses to test robustness of outcomes in subgroups defined by enrollment timing or geography.
    - Ensure that the overall data package, including randomized, supplemental, and contextual data, collectively supports interpretability and relevance to current clinical practice.

Early regulatory dialogue can help align on labeling expectations and contextual analyses that may be needed if SOC evolves.

Table 2. Approaches to Maintain Trial Applicability Under Evolving Standards of Care

Approach	When it May Fit	Advantages	Risks / Limitations
Design Strategies for Prospective Flexibility			
Investigator's Choice	Settings where multiple options provide modest and comparable benefit	Reflects real-world practice; endpoints remain interpretable if options are functionally equivalent; may reduce patient dropout from control arm	Unpredictable and uneven enrollment to each option; alignment on what constitutes functional equivalence can be difficult—differences in outcomes (e.g., OS vs. PFS benefit), safety profiles, or evidence maturity across the selected options may complicate labeling and regulatory interpretation
Region-specific controls	Regional asymmetry in approvals/access	Addresses local feasibility; ensures relevance to local regulators	Adds complexity for implementation if controls require different schedules and/or management, which can make blinding difficult; risks underpowered subgroups unless accrual is prespecified for each subgroup; complicates pooled analysis for overall trial effect; may not be possible due to regional implementation differences
Planned design adaptations	When a SOC change may occur during the trial and can be anticipated and pre-specified in advance	Allows prospectively defined changes (e.g., adaptation to a new comparator, adding an arm, modifying stratification) without undermining trial integrity and statistical validity	Operationally burdensome; requires extensive upfront planning, statistical adjustments, and regulatory dialogue; flexibility may be limited if changes occur earlier or later than anticipated
Contextual Evidence Additions During or Post-Trial			
Embedded cohort	When a SOC change is anticipated during the study period, but a full redesign is not feasible	Preserves the original trial design and analysis plan while allowing collection of comparator data aligned with the new SOC in selected regions; provides context without undermining the primary evidence	Operationally complex (requires protocol amendments, site-level variation, and careful delineation of how supportive vs. primary evidence will be used); may introduce heterogeneity that complicates interpretation
Supplemental bridging or U.S. specific cohorts	U.S. SOC diverges from global practice	Ensures U.S. applicability; maintains global enrollment	Added trial burden; feasibility may be limited if U.S. enrollment is already lagging; data may be viewed as less robust than fully integrated design

Approach	When it May Fit	Advantages	Risks / Limitations
Alternative or Complementary Evidence Generation			
Parallel region-specific trials	Rare populations where U.S. comparator feasibility is limited	Allows collection of U.S.-specific data while leveraging global enrollment for confirmatory endpoints	Limited regulatory precedent; risk of non-acceptance if one trial is conducted entirely ex-U.S.; potential heterogeneity in the patient cohorts, and endpoints
External Clinical Data	When comparator delivery is impractical	Provides supportive context; may reassure regulators/ clinicians	Not central to inference; limited regulatory precedent; potential bias; potentially greater heterogeneity in collection outcome data
Staggered or phased enrollment	When feasibility differs across regions or SOC is evolving in some geographies	Allows enrollment to proceed where feasible; provides flexibility to incorporate new comparators as standards evolve	Slower global enrollment; may create regional imbalances; could complicate pooled analysis and regulatory interpretation due to non-simultaneous accrual

Guided Approach to Comparator Selection

Comparator selection benefits from a structured, transparent process that progressively filters the broad landscape of potential SOCs into a justified, feasible choice. A "funnel" approach can help sponsors document rationale and demonstrate interpretability and applicability for regulators. Comparator discussions are best addressed during the pre-phase 3 meeting, when trial design and comparator decisions can still be meaningfully influenced. Once a study is underway, implementing mid-trial changes to the comparator is rarely feasible—protocol amendments can take six months to a year to operationalize across global sites. Early dialogue at this stage helps ensure that the planned comparator and contingency strategies remain acceptable, reducing the need for disruptive mid-trial modifications.

Proposed Steps:

1. Clarify the target setting: Review disease, line of therapy, key jurisdictions, and trial purpose.
2. Identify plausible SOC options: Consider regulatory approvals, guidelines, real-world practice, access, delivery feasibility, and evidentiary maturity (e.g., whether benefit is supported by PFS alone or requires OS evidence).
3. Screen for feasibility and ethical acceptability: Exclude options that are undeliverable (infeasible), impose undue logistical or operational burden (impractical), or ethically inappropriate (unacceptable).
4. Prioritize applicability: Ensure comparator(s) and planned enrollment align with key regions while considering global feasibility.
5. Assess timing and adaptation risks: Evaluate likelihood of SOC changes during the trial; pre-specify mitigation (e.g., sensitivity analyses, planned redesign, supplementary cohorts, dual primary endpoints, regional subgroup data).
6. Engage regulators at early key milestones: Seek early and structured input on comparator justification, particularly before protocol finalization or if major SOC shifts occur.
7. Reassess periodically: Review accrual patterns, regional uptake, and emerging SOC shifts during conduct to determine whether contextual or supplementary data will be needed to support interpretability.

This process provides a common framework to ensure comparator selection remains scientifically justified, ethically sound, and operationally feasible, while supporting transparency and consistent dialogue across sponsors, regulators and patients.

Future Directions

The working group acknowledged that there are no universally applicable solutions for comparator selection in MRCTs. The challenges created by evolving and heterogeneous SOCs are unlikely to be resolved by a single approach, and sponsors, regulators, and other stakeholders will need to pressure test a range of strategies to identify workable paths forward. A shift toward totality-of-evidence approaches that integrate prospective trial data with contextual external data or descriptive analyses may offer a practical way to address evolving SOC landscapes without undermining the core trial design.

A potential roadmap could include:

- Clearer articulation on how to balance regulatory requirements with practical expectations for applicability and safety would help sponsors design trials with greater predictability. Opportunities may include enhanced guidance or early, multi-agency dialogue to clarify when global evidence is sufficient and when U.S.-specific enrollment or comparators are essential.
- Use of retrospective data or prospective simulation exercises to test how different design and analysis strategies (e.g., investigator's choice, trial-within-a-trial, bridging cohorts) would perform under real-world SOC shifts.
- Convene sponsors, regulators, clinicians, and patients to assess feasibility and acceptability of different approaches, including trade-offs between scientific rigor, operational burden, and patient relevance.
- Implement statistical and design strategies to strengthen interpretability when heterogeneity cannot be fully avoided.
- Layout considerations around issues such as endpoint maturity (e.g., whether PFS alone is sufficient to redefine SOC), acceptable use of bridging data, and how much regional asymmetry can be tolerated.
- Explore models for multi-agency or joint regulatory engagement to enable earlier, more consistent feedback on comparator strategy in MRCTs (e.g., expanding components of Project Orbis to occur during clinical development phase).

Several open questions remain for the field:

- What magnitude and type of benefit should trigger reconsideration of a control regimen — is a PFS improvement sufficient, or should OS or long-term data be required?
- How can patient perspectives on acceptability and willingness to enroll be more systematically integrated into comparator selection?
- How far into trial enrollment or endpoint maturity is it reasonable to adapt a comparator strategy, and what are the implications for analysis integrity?
- How can trials balance the need for region applicability (particularly U.S. applicability) with the operational feasibility of enrolling patients in regions where new SOCs are not yet approved or reimbursed?
- How much heterogeneity can be accommodated without undermining interpretability and labeling?
- Under what circumstances can external controls or RWD provide meaningful supplemental support when SOC shifts post-initiation?

Developing answers to these questions will require structured experimentation, ongoing dialogue, and shared learning across stakeholders. While perfect solutions may not be attainable, a deliberate process to evaluate and refine strategies can bring greater predictability and transparency to comparator selection in MRCTs, ultimately ensuring that trials remain both feasible and relevant to patients.

References

1. U.S. Food and Drug Administration. Considerations for Generating Clinical Evidence from Oncology Multiregional Clinical Development Programs (Draft Guidance, September 2024). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-generating-clinical-evidence-oncology-multiregional-clinical-development-programs
2. ICH. Guideline E17: General Principles for Planning and Design of Multi-Regional Clinical Trials. Step 5 adoption (2018). Available from: https://www.ema.europa.eu/en/ich-guideline-e17-general-principles-planning-design-multi-regional-clinical-trials-scientific-guideline
3. Fallah J, et al. Equipoise Lost? Trial Conduct Challenges in an Era of Breakthrough Therapies. J Clin Oncol. 2024 Nov 10;42(32):3783-3787. DOI: 10.1200/JCO-24-01200
4. U.S. Food and Drug Administration. Oncologic Drugs Advisory Committee Meeting Briefing Document: Glofitamab for the Treatment of Relapsed or Refractory Diffuse Large B-Cell Lymphoma. Silver Spring, MD: FDA; May 16, 2025. Available from: https://www.fda.gov/media/186557/download
5. Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 355(d).
6. Fojo T, LaRose M, Bates SE. The Impact of Exuberance on Equipoise in Oncology Clinical Trials: Sotorasib as Archetype. Oncologist. 2024 Apr 4;29(4):275-277. DOI: 10.1093/oncolo/oyae042

Impact of the COVID-19 Pandemic Mitigation Strategies on Cancer Treatment Trials: A Meta-Analysis of Industry and National Cancer Institute Studies

Joseph M. Unger, PhD[1]; Hillary S. Andrews, PhD[2]; Laura A. Levit, JD[3]; Brittany A. McKelvey, PhD[2]; Mark Stewart, PhD[2]; Beverly Canin[4]; Keith Flaherty, MD[5]; Denise Kimball, RN[6]; Therica Miller, MBA[7]; Adedayo Onitilo, MD[8]; Suanna Bruinooge, MPH[3]; Elizabeth Garrett-Mayer, PhD[3]; and Caroline Schenkel, MS[9]

DOI https://doi.org/10.1200/OA-25-00021

ABSTRACT

PURPOSE The onset of the COVID-19 pandemic in early 2020 disrupted the conduct of cancer clinical trials. In response, federal agencies allowed more flexibility for trial recruitment and patient follow-up. A key question is whether the benefits of adopting these strategies outweigh the potential detriments to quality metrics.

METHODS A joint ASCO and Friends of Cancer Research task force invited industry and National Cancer Institute trial sponsors to contribute deidentified trial-level aggregate data on enrollment, major protocol deviations, dropouts, and severe adverse events (Common Terminology Criteria for Adverse Events grade 3-5). These quality metrics were examined as proportions of participants at risk during the pre–COVID-19 (January 2017-February, 2020), initial wave (March-April, 2020), initial recovery (May-December, 2020), and secondary recovery (January 2021-December 2022) periods. Multilevel beta-regression was used, adjusting for phase; study and sponsor were treated as random effects. Indicator variables were used with pre–COVID-19 as the reference.

RESULTS Ten sponsors contributed 67 analyzable trials with N = 12,000 US-based participants. Enrollment odds decreased 49% in the initial wave (odds ratio [OR], 0.51 [95% CI, 0.30 to 0.86], $P = .01$) but recovered to pre–COVID-19 levels by 2021-2022 (OR, 1.01 [95% CI, 0.56 to 1.81], $P = .97$). Major protocol deviations, dropouts, and severe toxicity all had a lower incidence in the initial wave compared with pre–COVID-19; these outcomes were also less frequent ($P < .05$) in the initial recovery period but returned to pre–COVID-19 levels by 2021-2022.

CONCLUSION In this multicollaborator evaluation, large declines in enrollment, major protocol deviations, dropouts, and severe toxicity during the acute phase of the pandemic all returned to pre–COVID-19 levels by 2021-2022. These findings highlight the impact of the temporary disruption to trial conduct during the pandemic's peak, but suggest that pandemic-related procedural flexibility did not result in long-term reduced data quality. Sponsors and regulators should consider broader adaptation of trial flexibilities moving forward.

ACCOMPANYING CONTENT

Data Supplement

Accepted April 4, 2025
Published June 13, 2025

JCO Oncology Adv 2:e2500021

INTRODUCTION

The onset of the COVID-19 pandemic in early 2020 disrupted the conduct of cancer clinical trials, with steep drops in enrollment to existing trials and reductions in the activation of new trials.[1-3] Given the challenges of recruiting patients to clinical trials during a pandemic, major federal agencies shared guidance to allow more flexibility in trial processes, supporting the continuity of clinical trials while ensuring patient safety and data integrity.[4,5] Trial sponsors rapidly adopted these measures, including strategies such as remote patient consent to participate, remote symptom monitoring, and distribution of oral anticancer agents directly to trial patients.[4-6] Enrollment in trials subsequently rebounded, especially for trials examining new cancer treatments.[7]

The strategies adopted during the pandemic to facilitate access to trials, protocol treatment, monitoring, and follow-up have been advanced and studied previously.[8,9] However, before the pandemic, their administration had not been widely adopted, largely out of concern about their impact on data quality. The COVID-19 public health emergency forced a

CONTEXT

Key Objective
Did the benefits of making cancer clinical trials easier to conduct, adopted in response to the COVID-19 pandemic in early 2020, outweigh the potential harms to data quality?

Knowledge Generated
Based on a meta-analysis of 67 trials comprising N = 12,000 US-based participants conducted by 10 sponsors, large declines in trial enrollment, major protocol deviations, dropouts, and severe toxicity were found during the acute phase of the COVID-19 pandemic in 2020. However, these metrics all returned to prepandemic levels by 2021-2022, suggesting limited or no adverse impact of trial flexibilities on data quality over the longer term.

Relevance *(P.L. Kunz)*
The COVID-19 pandemic disrupted the conduct of clinical trials and led to temporary measures to allow more flexibility, including remote consent, remote symptom monitoring, and distribution of oral anticancer agents directly to patients. A large meta-analysis showed that there were no negative effects of these changes on quality metrics. These findings suggest that we should routinely consider decentralized clinical trial principles.*

Plain Language Summary *(P.L. Kunz)*
The COVID-19 pandemic forced researchers to be more flexible with how clinical trials were conducted, including telehealth consents and visits and delivery of oral anti-cancer treatments. A research study showed that that these changes did not reduce the data quality from these trials. This suggests that we could consider adopting permanent changes to clinical trial conduct that are more patient centric.†

*Relevance section written by *JCO Oncology Advances* Editor-in-Chief Pamela L. Kunz, MD, FASCO.
†Plain Language Summary written by *JCO Oncology Advances* Editor-in-Chief Pamela L. Kunz, MD, FASCO.

rapid and systematic adoption of decentralized trial conduct procedures. A key question for researchers and policymakers is whether the benefits of adopting these strategies have outweighed the potential detriments to data quality.

To address this, in 2022, ASCO and the Friends of Cancer Research initiated an effort to systematically evaluate the impact of the COVID-19 experience for sponsors of oncology clinical trials.[10] This report is based on original data obtained from 10 industry and federal sponsors of cancer clinical trials. To our knowledge, it represents a first-of-its-kind evaluation of the nature of key data quality indicators from trials during the COVID-19 pandemic.

METHODS

Sponsor and Trial Eligibility

Study participation was open to pharmaceutical companies and National Cancer Institute (NCI) Network groups that sponsored at least one anticancer treatment trial before the onset of the COVID-19 Public Health Emergency (PHE; January 2017-December 2019) and at least one anticancer treatment trial during the PHE (January 2020-December 2022). Eligible studies included phase I-III interventional anticancer treatment trials of any modality (eg, systemic therapy, surgery, etc) open in the United States from January 2017 to December 2022.

For global trials, we requested that sponsors provide data from US patients only. To limit the risk of identification of patients and sponsors, data were aggregated. The WCG Institutional Review Board approved this study.

Conceptual Framework

The COVID-19 outbreak had both direct and indirect effects on the conduct of cancer clinical trials. Direct effects included reduced patient willingness to participate in clinical trials and decreased institutional capacity and staffing.[11-13] Indirect effects resulted from the declaration of a PHE and the accompanying mitigation strategies, such as shutdowns. Throughout this article, we generally refer to the impact of the COVID-19 pandemic itself (ie, the underlying causal mechanism of adverse consequences for trial conduct), even if, in some instances, the PHE was the more proximate cause.

Dependent Variables—Data Quality Indicators

Four data quality indicators served as the dependent variables (ie, the outcomes).

1. Enrollments were identified as the number of patients enrolled in a clinical trial. Enrollment data were chosen to reflect patients' access to clinical trials and willingness to participate in clinical trials throughout the pandemic.
2. Major protocol deviations were defined as any noncompliance with an IRB-approved protocol that presented a potential risk to participants or affected the integrity of study data. Protocol deviations were interpreted by sponsors to represent adherence to stated treatment, procedures, and data collection processes defined prospectively within trial protocols; thus, we interpreted an increase in the frequency of protocol deviations as a decrease in data integrity.[14]
3. A patient dropout was defined as a patient withdrawal from protocol therapy early (ie, before achieving the primary end point as defined in the protocol) for any reason, excluding death. Common categories of dropout include withdrawal of consent, lost to follow-up, patient noncompliance, adverse events, or progressive disease. Patient dropouts reduce the overall power of trial designs, affect the integrity of trial data, and suggest the challenges that individual patients may face in adhering to study or protocol therapy.
4. Severe adverse events (grades 3-5), coded according to the Common Terminology Criteria for Adverse Events based on the initial onset, are significant treatment-related complications often requiring hospitalization.[15] Only severe adverse events determined to be possibly, probably, or definitely related to treatment were considered. An increase in severe toxicity rates from prepandemic to during the pandemic may indicate compromised patient safety resulting from pandemic-related disruptions.

We hypothesized that enrollments declined and that major protocol deviations, patient dropouts, and severe adverse events increased beginning with the acute phase (ie, initial wave) of the pandemic.

Independent Variables

Prespecified COVID-19 pandemic–related landmark periods served as the key independent variables. The 3-year pre–COVID-19 pandemic period (pre-COVID) was defined as January 1, 2017-February 29, 2020, and was applied to enable the determination of a stable baseline period.

The initial pandemic wave was defined as beginning on March 1, 2020, commensurate with the first death because of COVID-19 observed in the United States and the announcement of a PHE.[16,17] The initial wave was 2 months, ending on April 30, 2020, when the US COVID-19 mortality rate approximately peaked.[18]

The initial recovery period was defined as May 1, 2020-December 31, 2020. The secondary recovery period was defined as January 1, 2021-December 31, 2022. Two recovery periods were used to determine whether recovery from pandemic effects occurred in stages.

Demographic variables included dichotomized age (<65 years *v* 65 or older), reflecting that those 65 years or older have access to social and medical support programs (eg, Social Security and Medicare); sex (female *v* male); race (Black *v* other race); and ethnicity (Hispanic *v* other ethnicity). Race and ethnicity were included given extensive research illustrating racial and ethnic barriers (especially for Black and Hispanic individuals) in access to care, including to COVID-19–related care.[19]

Study phase was characterized as early (I and II)- versus late (III)-phase trials. Trials with combined strategies (ie, a phase II-III trial) were coded according to the highest phase.

Statistical Methods

Patients were considered *at risk* of an event if they had ≥30 days of follow-up within the prescribed landmark periods. Furthermore, to enable consistency in the amount of time *at risk* for the study outcomes across a broad set of trials with different follow-up periods, the follow-up period was specified to end at the completion of protocol therapy or 1 year after initial enrollment, whichever came first. In this context, the validity of the analysis is predicated on the idea that patients are uniformly at risk of a given event at any time within 1 year after trial enrollment. An individual patient's follow-up time could have spanned multiple study-specified periods. Within a given period, a patient who experienced an event was coded 1; otherwise, they were coded 0 (including unknowns). Thus, in aggregate, the trial- and period-level unit of analysis was a proportion, ranging from 0 to 1.

Trial enrollments were indexed from the date of initial registration to a trial. To account for a heterogeneous mix of trials with different enrollment goals over time, enrollment totals within each trial and period level were standardized on a 0-100 scale as the proportion of maximum study-level monthly enrollment across periods.

As the units of analysis were interval-level continuous proportions bounded on a 0-1 scale, we used multilevel beta-regression analyses.[20] A Smithson transformation was used to accommodate 0 and 1 values.[20] Multivariable analyses were conducted, adjusted for trial phase (early *v* late) with study and sponsor as random effects. For evaluations by time period, pre-COVID was considered the baseline (ie, reference) period, and indicator variables were used to compare outcomes between the initial wave, the initial recovery period, and the secondary recovery period and the pre-COVID baseline period. Only studies with both pre-COVID and follow-up data were included.

Interaction analyses by age, sex, race, and ethnicity were conducted to assess whether patterns of outcomes differed over time by these factors.

All *P* values were two-sided. For marginal comparisons, alpha = .05 was considered statistically significant, with no

adjustment for multiple comparisons. For interaction tests by sociodemographic variables, we highlighted instances with alpha <.10 given more limited power for interaction analyses and for hypothesis generation.[21,22]

RESULTS

Among 41 eligible sponsors invited to participate, 10 (nine industry and one NCI) contributed data on 88 trials, among which 67 trials (76.1%) included sufficient data to analyze one or more of the specified outcomes, including enrollment (67, 100% of analyzable trials), protocol deviations (60, 89.6%), dropouts (61, 91.0%), and adverse events (61, 91.0%; Fig 1). The majority of evaluable trials (42, 62.9%) were sponsored by industry, with 25 (37.1%) sponsored by the NCI through its National Clinical Trials Network program. Fourteen trials (20.9%) were late-stage trials, and 53 (79.1%) were early-stage trials.

Overall, the 67 analyzable trials represented N = 12,000 patients. The majority of patients were younger than 65 years (58.1%), and 42.8% was female (Table 1). Black and Hispanic representation was 10.6% and 8.5%, respectively. The most common cancers among patients were prostate (12.4%), breast (12.0%), bladder (11.3%), myeloma (11.2%), lymphoma (9.9%), and lung (6.5%).

Quality Metrics Over Time

There were large reductions in reported major protocol deviations, dropouts, and severe or worse toxicity in the initial wave compared with the pre-COVID period that rebounded to near-baseline proportions by the secondary recovery period. For instance, in the pre-COVID period, the estimated proportion of patients with a reported major protocol deviation was 15.7% (Table 2). By contrast, the proportion in the initial wave was 8.2%, representing a 58% reduction in the odds (odds ratio [OR], 0.42 [95% CI, 0.30 to 0.61], $P < .001$). The average estimated major protocol deviations increased to 12.5% in the initial recovery period, a 32% reduction compared with the pre-COVID period (OR, 0.68 [95% CI, 0.49 to 0.95], $P = .02$). By the secondary recovery period, the estimated major protocol deviations were 14.9%, a reduction of 24% in the odds that was not statistically significantly different from the baseline period (OR, 0.76 [95% CI, 0.53 to 1.07], $P = .12$). Similar large reductions in enrollments and in reported dropouts and severe or worse toxicity were observed during the initial wave, which also rebounded to near baseline proportions by the second recovery period (Fig 2).

Quality Metrics Over Time by Demographic Variables

Patterns of quality metrics over time by demographic factors reflected the overall aggregate patterns with a few exceptions (Fig 3). Black patients were less likely to enroll in trials than other patients during the initial wave. In addition, during the initial wave, a larger drop in reported protocol deviations was observed for patients 65 years or older ($P = .06$) and Black patients ($P = .04$). Finally, dropouts rebounded to pre-COVID levels during the initial recovery period for Black patients, but not for other patients ($P = .007$).

By the secondary recovery period, patterns of quality metrics were not statistically different from the pre-COVID baseline period for all demographic subgroups of patients except protocol deviations among patients younger than 65 years (OR, 0.67 [95% CI, 0.45 to 0.99], $P = .04$) and dropouts among male patients (OR, 0.66 [95% CI, 0.45 to 0.97], $P = .03$; Fig 3; Data Supplement, Table S1).

DISCUSSION

In this meta-analysis of multisponsor data, we found large declines in both enrollments to cancer clinical trials and reported major protocol deviations, dropouts, and severe toxicity for patients enrolled in trials during the initial wave of the COVID-19 pandemic. All metrics rebounded to approximately pre-COVID (ie, baseline) levels by the 2021-

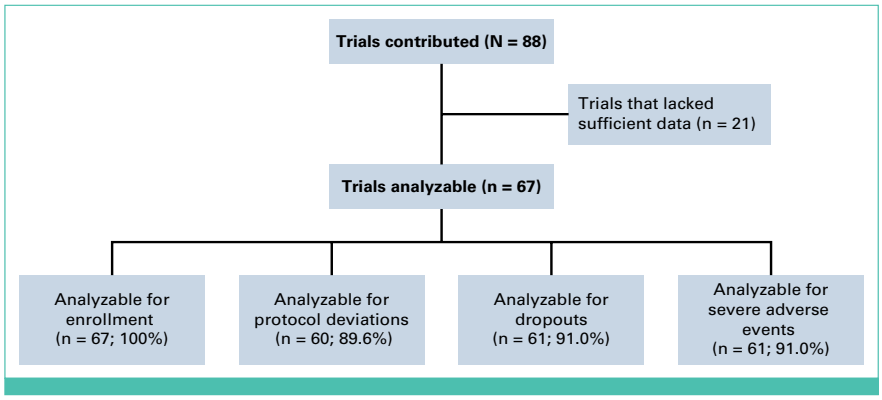


FIG 1. Flow diagram.

TABLE 1. Patient Characteristics (N = 12,000)

Characteristic	No. (%)
Age, years	
<65	6,973 (58.1)
≥65	5,027 (41.9)
Sex	
Female	5,136 (42.8)
Male	6,864 (57.2)
Race[a]	
Black	1,223 (10.6)
Other	10,303 (89.4)
Unknown	474
Ethnicity[a]	
Hispanic	991 (8.5)
Other	10,600 (91.5)
Unknown	409
Cancer type	
Biliary	495 (4.1)
Bladder	1,354 (11.3)
Breast	1,440 (12.0)
Cervical	169 (1.4)
Leukemia	177 (1.5)
Lung	779 (6.5)
Lymphoma	1,187 (9.9)
Melanoma	538 (4.5)
Myeloma	1,339 (11.2)
Prostate	1,484 (12.4)
Renal	235 (2.0)
NOS[b]	2404 (20.0)
Other[c]	399 (3.3)

Abbreviation: NOS, not otherwise specified.
[a]Percentages calculated from patients with known data.
[b]Denotes solid tumors or cancer.
[c]Includes amyloidosis (150), colorectal (52), hepatocellular (76), sarcoma (22), and urothelial (85).

2022 period. During the initial wave of the pandemic, declines in enrollment were more pronounced among Black patients compared with all other racial groups combined. In addition, we observed a greater reduction in reported major protocol deviations for both Black patients and patients 65 years or older. Among virtually all demographic subgroups of patients, levels of the specified measures returned to baseline levels by the end of the study period.

The findings of reduced enrollment during the pandemic confirm previous observations of steep reductions in accrual during the initial wave of the pandemic.[1,3,13] Data for federally sponsored trials suggested a 50% average reduction in enrollment early in the pandemic.[7,12] Industry trials demonstrated reduced trial enrollments of about 30%.[3,12] Enrollment declines were due to reductions in enrollment to active trials and reductions in the activation of new trials.[2]

Previous evaluations suggested that protocol deviations, dropouts, and severe toxicity would be higher during the acute phase of the pandemic. A survey conducted by our team revealed that 90% of study sponsors reported a moderate or substantial rise in protocol deviations.[6] However, this finding was based on sponsor perceptions rather than quantitative assessments. Bakouny et al[3] showed an increase in protocol deviations during the initial wave although the evaluation was based on only 80 patients and the observed deviations were almost entirely minor (95%) and predominantly attributable to the COVID-19 virus.

Our data on 12,000 patients demonstrate a decline in the reported protocol deviations, dropouts, and severe toxicity during the initial pandemic wave. Several factors might have contributed to these declines. Guidance issued during the pandemic introduced modifications to trial conduct, potentially leading to fewer events classified as protocol deviations (ie, missed visits). This finding aligns with our previous qualitative analysis, which suggested that trial flexibility measures had an impact on the overall occurrence of protocol deviations.[6] Logistical strategies to ease trial conduct, such as direct shipment of study drug to patients' homes, might have also reduced protocol deviations. In addition, survey results suggest that investigators and care teams adopted strategies to limit or avoid immunocompromising regimens early in the pandemic, which could have influenced the reporting of severe toxicities.[13] However, it is unlikely that actual toxicity from cancer treatment dropped so precipitously immediately after the COVID-19 outbreak. The observed decrements in quality indicators may also reflect limitations in reporting and/or patient follow-up early in the pandemic. For example, a survey found that 60% of investigators reported that the COVID-19 pandemic had a moderate or high impact on patient visits and cancer centers experienced personnel shortages because of COVID-19.[12,23–25] These factors together could have contributed to decreased data collection and reporting during the initial wave.

In May 2021, the CDC announced new guidance, stating that fully vaccinated individuals no longer needed to mask or practice social distancing.[26] In addition, by 2022, health care utilization in the United States had largely rebounded from the pandemic's beginning.[27] In this 2021-2022 timeframe, after the conclusion of the acute phase of the pandemic, we found that patterns of quality metrics had largely returned to pre-COVID (ie, baseline) levels, but importantly, did not *exceed* pre-COVID levels. This pattern emerged although trial mitigation processes and flexibilities had been widely adopted throughout the pandemic, including measures such as remote distribution of oral therapies, remote monitoring, and remote consent.[6,13]

The trial mitigation measures recommended early in the pandemic by federal agencies align with decentralized clinical trial principles.[28,29] Although decentralized clinical trial elements have been considered over decades, the

TABLE 2. Results for Quality Metrics Over Time

	Pre-COVID (January 2017-February 2020)		Initial Wave (March-April 2020)			Initial Recovery (May-December 2020)			Secondary Recovery (January 2021-December 2022)		
End Point	%	OR (95% CI)	%[a]	OR (95% CI)	*P*[b]	%[a]	OR (95% CI)	*P*[b]	%[a]	OR (95% CI)	*P*[b]
Mean monthly enrollment[c]	64.5	1.00 (reference)	47.9	0.51 (0.30 to 0.86)	.01	65.8	1.06 (0.62 to 1.83)	.66	65.0	1.01 (0.56 to 1.81)	.97
Major protocol deviations[d]	15.7	1.00 (reference)	8.7	0.42 (0.30 to 0.61)	.001	12.5	0.68 (0.49 to 0.95)	.02	14.9	0.76 (0.53 to 1.07)	.12
Dropouts[d]	37.6	1.00 (reference)	8.5	0.10 (0.06 to 0.14)	.001	24.4	0.43 (0.31 to 0.58)	.001	30.7	0.74 (0.54 to 1.02)	.07
Severe or worse toxicity[d]	35.7	1.00 (reference)	18.6	0.36 (0.26 to 0.49)	.001	29.3	0.68 (0.51 to 0.91)	.001	32.7	0.86 (0.63 to 1.17)	.34

Abbreviation: OR, odds ratio.
[a]Among trials with both pre-COVID and follow-up data.
[b]*P* values are not adjusted for multiple comparisons.
[c]To account for a heterogeneous mix of trials with different enrollment goals over time, enrollment was standardized on a 0-100 scale as the proportion of maximum study-level monthly enrollment across time periods.
[d]Percentages indicate the proportion of patients *at risk* of a given outcome who experienced ≥1 of the given outcome during the specified period.

COVID-19 pandemic enabled their rapid adoption, creating the scenario for a natural experiment in which the impact on trial conduct could be feasibly evaluated. Throughout the pandemic, decentralized clinical trial elements were implemented to ensure continuity of research while minimizing risks to safety and data integrity. Despite these procedural modifications, quality metrics remained consistent with baseline levels, demonstrating that using DCT elements did not compromise trial quality. This finding suggests that decentralized clinical trial elements can be safely adopted as permanent fixtures in the conduct of cancer clinical trials without substantial reductions in data quality that could compromise study reporting or interpretation.

The permanent adoption of these new sets of trial procedures would represent a paradigm shift in the conduct of cancer clinical trials, with the potential to improve access to trials for all patients and thereby to conduct trials more rapidly in more diverse sets of patients. The US Food and Drug Administration recently provided guidance on the conduct of decentralized trials, with a focus on digital health technologies.[29] Research and action statements by ASCO, a call to action by the American Cancer Society, and working group statements by the NCI have also highlighted the importance of further advancing decentralized clinical trial elements.[30-34] The adoption of strategies that are patient-focused and reduce the burden of trial participation is a necessary adjunct to other measures aimed at improving access to clinical trials for vulnerable patient populations.[35] Indeed, our own findings illustrated the disproportionate impact of the pandemic on some patient groups, with larger declines in enrollment among Black patients and larger reductions in protocol deviations reported for both Black patients and older patients. These differences likely

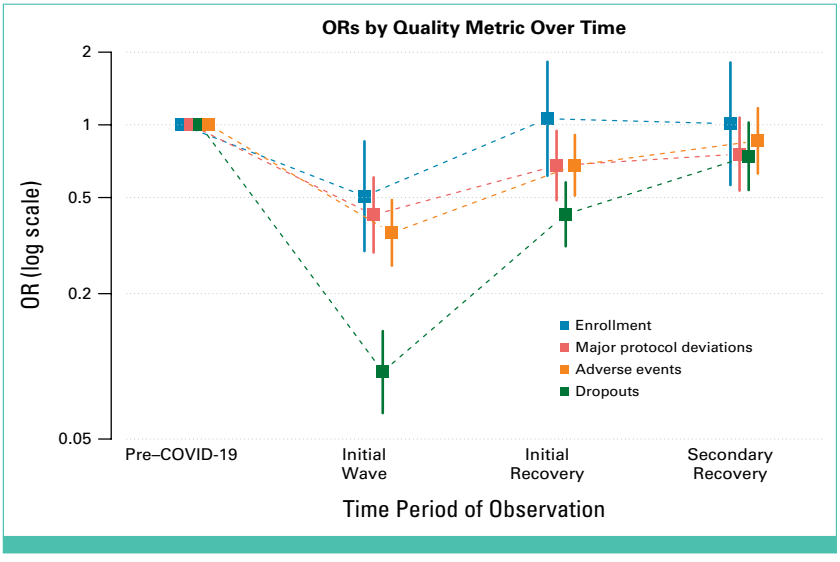


FIG 2. ORs by quality metrics over time. OR, odds ratio.

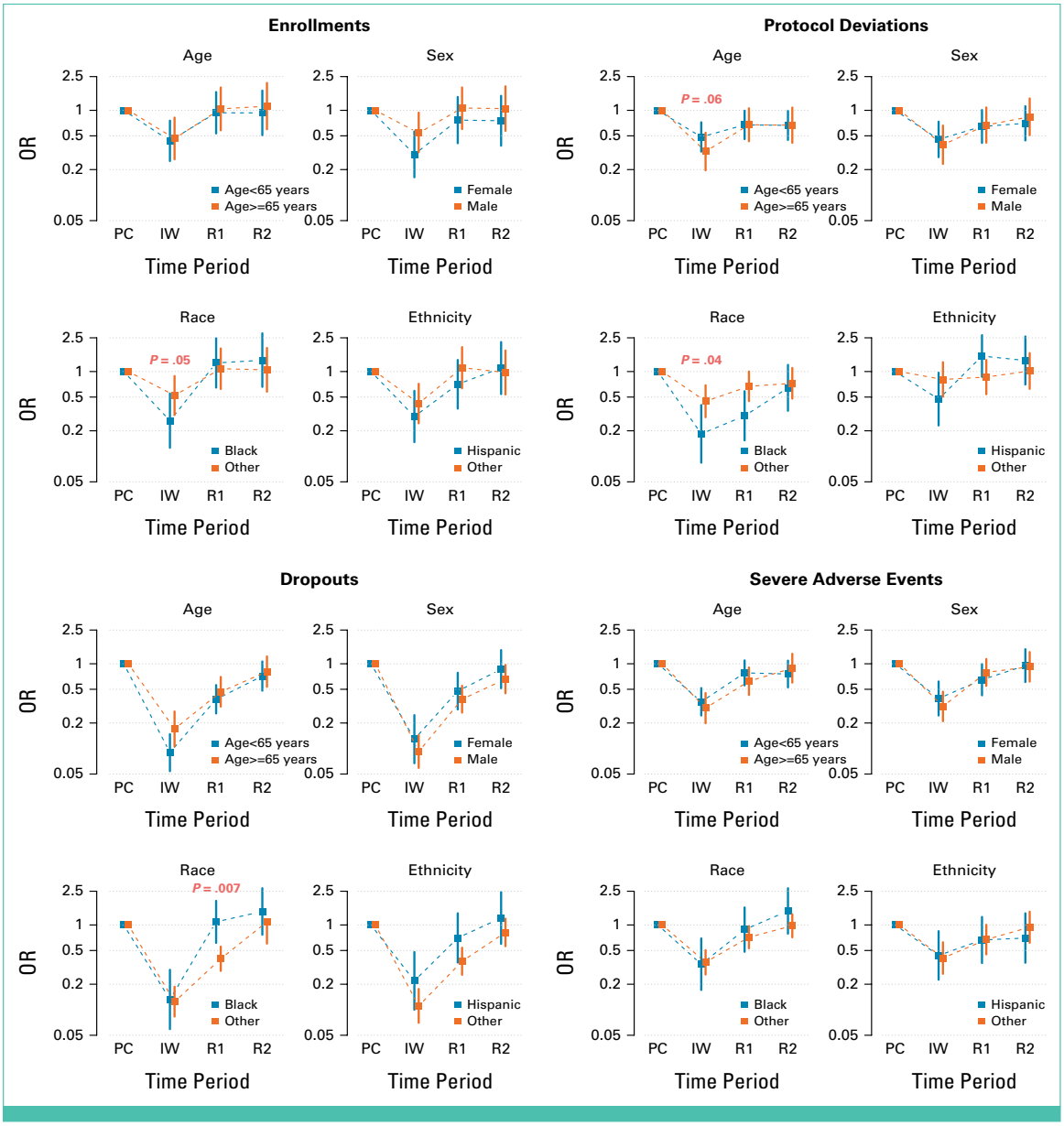


FIG 3. ORs by quality metrics by demographic variables. *P* values indicate whether patterns of outcomes differed between specified groups (based on interaction analyses) at the given time period. *P* values <.10 are highlighted in red. IW, initial wave; OR, odds ratio; PC, Pre-COVID; R1, recovery period 1; R2, recovery period 2.

reflect known differences in access to care by race and age during the pandemic, including access to clinical research studies.[19,36,37]

This study has limitations. Although the study represented 12,000 patients, each random effect was at the study level, of which only 67 trials were available, which likely limited power to detect differences between patient subgroups. We did not adjust for multiple comparisons given the observational nature of this study. However, the strength of the observed differences during the initial wave was sufficient to be statistically significant under any multiplicity adjustment. Furthermore, our evaluation was completed at the end of 2022. Data beyond 2022 would have revealed the more enduring impact of the recovery in the presence of continuous adoption of modernized trial processes, especially as they are

related to dropout rates and retention. In addition, data on other aspects of patients' backgrounds—such as rurality or additional categories of race—were not available for evaluation. Finally, data on the stage of disease were not available, and evaluations within disease type were limited by the number of studies for a given cancer.

In this comprehensive, multisponsor evaluation of quality metrics for cancer clinical trials during the COVID-19 pandemic, we found large declines in enrollment and in the occurrence of major protocol deviations, dropouts, and severe toxicity during the initial wave that rebounded to pre-COVID levels by the end of the study period. These findings highlight the temporary disruption to trial conduct during the acute phase of the COVID-19 pandemic. They also indicate that pandemic-related procedural flexibility did not lead to increased major protocol deviations, dropouts, or severe toxicity over the long term. As such, sponsors and regulators should consider broader adoption of trial flexibilities moving forward.

AFFILIATIONS

[1]Fred Hutchinson Cancer Center, Seattle, WA
[2]Friends of Cancer Research, Washington, DC
[3]American Society of Clinical Oncology, Alexandria VA
[4]Breast Cancer Options, Kingston, NY
[5]Massachusetts General Hospital, Boston, MA
[6]Johnson & Johnson, New Brunswick, NJ
[7]Tisch Cancer Center, New York, NY
[8]Marshfield Clinic Health System, Weston, WI
[9]Children's National Hospital, Washington, DC

CORRESPONDING AUTHOR

Joseph M. Unger, PhD; e-mail: junger@fredhutch.org.

SUPPORT

NIH/NCI grant awards U10CA180888, U10CA180819 (J.M.U.).

AUTHOR CONTRIBUTIONS

Conception and design: Joseph M. Unger, Hillary S. Andrews, Laura A. Levit, Brittany A. McKelvey, Mark Stewart, Beverly Canin, Keith Flaherty, Suanna Bruinooge, Elizabeth Garrett-Mayer, Caroline Schenkel
Administrative support: Laura A. Levit, Adedayo Onitilo, Caroline Schenkel
Provision of study materials or patients: Laura A. Levit, Adedayo Onitilo
Collection and assembly of data: Joseph M. Unger, Hillary S. Andrews, Laura A. Levit, Denise Kimball, Caroline Schenkel
Data analysis and interpretation: Joseph M. Unger, Laura A. Levit, Mark Stewart, Keith Flaherty, Denise Kimball, Therica Miller, Adedayo Onitilo, Suanna Bruinooge, Elizabeth Garrett-Mayer
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 https://ascopubs.org/authors.

Joseph M. Unger
Consulting or Advisory Role: Loxo/Lilly

Keith Flaherty
Leadership: Strata Oncology, Kinnate Biopharma, Scorpion Therapeutics, Clovis Oncology, Khora Therapeutics
Stock and Other Ownership Interests: Clovis Oncology, Loxo, X4 Pharma, Strata Oncology, PIC Therapeutics, Apricity Health, Oncoceutics, FOGPharma, Tvardi Therapeutics, Kinnate Biopharma, Scorpion Therapeutics, ALX Oncology, xCures, Monopteros Therapeutics, Vibliome Therapeutics, Transcode Therapeutics, Soley Therapeutics, Nextech Invest, Alterome Therapeutics, PreDICTA, Flinr Therapeutics, IntrECate, Tasca Therapeutics
Consulting or Advisory Role: Novartis, Tvardi Therapeutics, Takeda, Karkinos Healthcare

Denise Kimball
Employment: Johnson and Johnson
Leadership: Johnson and Johnson
Stock and Other Ownership Interests: Johnson and Johnson Innovative Medicine

Therica Miller
Employment: Mount Sinai Health System
Consulting or Advisory Role: Triomics, J&J, Flatiron Health, AOC Oncology
Research Funding: J&J
Travel, Accommodations, Expenses: American Society of Clinical Oncology
Other Relationship: University of Southern California, AOC Oncology, Alliance Foundation Trials, Children's Hospital Los Angeles, George Washington University, Alphasights

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank the following sponsors for contributing data to the meta-analysis: Amgen; Bayer; Eisai; EMD Serono/Merck KGaA; Genentech/Roche; Janssen/Johnson & Johnson; Merck & Co, Inc; Rahway, NJ, USA; Mirati Therapeutics (a Bristol Myers Squibb company); SWOG Cancer Research Network; and Takeda.
We also thank the following individuals for their contributions to the project: Margaret E McCann, DVM, PhD, Global Regulatory Affairs and Clinical Safety, Merck & Co, Inc, Rahway, NJ; Xiao Peng, PhD, Biostatistics and Research Decisions Sciences, Merck & Co, Inc, Rahway, NJ; Mathias Lebreton, Programming Lead, Johnson and Johnson Innovative Medicine; and Patricia Mader, MS, Kimberly Barnholt, PhD, Wei Yu, PhD, Yong Wang, MS, Bann-mo Day, PhD, Evidence Generation, US Medical Affairs, Genentech.

REFERENCES

1. Unger JM, Blanke CD, LeBlanc M, et al: Association of the Coronavirus disease 2019 (COVID-19) outbreak with enrollment in cancer clinical trials. JAMA Netw Open 3:e2010651, 2020
2. Unger JM, Xiao H: The COVID-19 pandemic and new clinical trial activations. Trials 22:260, 2021

3. Bakouny Z, Labaki C, Bhalla S, et al: Oncology clinical trial disruption during the COVID-19 pandemic: A COVID-19 and cancer outcomes study. Ann Oncol 33:836-844, 2022
4. National Cancer Institute: Coronavirus guidance, 2022. https://ctep.cancer.gov/investigatorResources/corona_virus_guidance.htm
5. US Food and Drug Administration: FDA guidance on conduct of clinical trials of medical products during COVID-19 pandemic, 2020. https://www.fda.gov/regulatory-information/search-fdaguidance-documents/fda-guidance-conduct-clinical-trials-medical-products-during-covid-19-pandemic
6. Unger JM, Stires H, Levit LA, et al: Sponsor perspectives on the impact of the COVID-19 pandemic on interventional cancer clinical trial protocols and data quality. JCO Oncol Pract 19:907-916, 2023
7. Unger JM, Xiao H, LeBlanc M, et al: Cancer clinical trial participation at the 1-year anniversary of the outbreak of the COVID-19 pandemic. JAMA Netw Open 4:e2118433, 2021
8. Khozin S, Coravos A: Decentralized trials in the age of real-world evidence and inclusivity in clinical investigations. Clin Pharmacol Ther 106:25-27, 2019
9. Sommer C, Zuccolin D, Arnera V, et al: Building clinical trials around patients: Evaluation and comparison of decentralized and conventional site models in patients with low back pain. Contemp Clin Trials Commun 11:120-126, 2018
10. Friends of Cancer Research: Imact of the COVID-19 Pandemic Mitigation Strategies on Cancer Clinical Trials. Preliminary Findings of a Friends of Cancer Research-American Society of Clinical Oncology Study. Friends of Cancer Research Annual Meeting, 2022. https://friendsofcancerresearch.org/wp-content/uploads/Impact_COVID-19_Pandemic_Strategies_Cancer_Clinical_Trials.pdf
11. Fleury ME, Farner AM, Unger JM: Association of the COVID-19 outbreak with patient willingness to enroll in cancer clinical trials. JAMA Oncol 7:131-132, 2021
12. Prindiville SA, Sarosy GA, Loose D, et al: Patterns of enrollment in cancer treatment trials during the COVID-19 pandemic at National Cancer Institute-designated cancer centers. Cancer J 28: 111-117, 2022
13. Upadhaya S, Yu JX, Oliva C, et al: Impact of COVID-19 on oncology clinical trials. Nat Rev Drug Discov 19:376-377, 2020
14. US Food and Drug Administration: Protocol Deviations for Clinical Investigations of Drugs, Biological Products, and Devices: Guidance for Industry, 2024. https://www.fda.gov/media/184745/download
15. National Cancer Institute. Division of Cancer Treatment & Diagnosis: Common Terminology Criteria for Adverse Events (CTCAE). https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm
16. Centers for Disease Control and Prevention: CDC, Washington state report first COVID-19 death, 2020. https://www.cdc.gov/media/releases/2020/s0229-COVID-19-first-death.html
17. Cucinotta D, Vanelli M: WHO declares COVID-19 a pandemic. Acta Biomed 91:157-160, 2020
18. Ahmad FB, Cisewski JA, Miniño A, et al: Provisional mortality data—United States, 2020. MMWR Morb Mortal Wkly Rep 70:519-522, 2021
19. Xiao H, Vaidya R, Liu F, et al: Sex, racial, and ethnic representation in COVID-19 clinical trials: A systematic Review and meta-analysis. JAMA Intern Med 183:50-60, 2023
20. Smithson M, Verkuilen J: A better lemon squeezer? Maximum-Likelihood regression with beta-distributed dependent variables. Psychol Methods 11:54-71, 2006
21. Cascio WF, Zedeck S: Open a new window in rational research planning: Adjust alpha to maximize statistical power. Personnel Psychol 36:517-526, 1983
22. Mathieu JE, Aguinis H, Culpepper SA, et al: Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. J Appl Psychol 97:951-966, 2012
23. Dizon DS, Szczepanek CM, Petrylak DP, et al: National impact of the COVID-19 pandemic on clinical trial staff attrition: Results of the SWOG Cancer Research Network Survey of Oncology Research Professionals. J Clin Oncol 40, 2022 (suppl 16; abstr 11049)
24. Huggins A, Husaini M, Wang F, et al: Care disruption during COVID-19: A national survey of hospital leaders. J Gen Intern Med 38:1232-1238, 2023
25. Sun G, Dizon DS, Szczepanek CM, et al: Crisis of the clinical trials staff attrition after the COVID-19 pandemic. JCO Oncol Pract 19:533-535, 2023
26. Centers for Disease Control and Prevention: Interim Public Health Recommendations for Fully Vaccinated People, 2021. https://stacks.cdc.gov/view/cdc/106177
27. Peterson-KFF Health System Tracker: How has healthcare utilization changed since the pandemic? 2023. https://www.healthsystemtracker.org/chart-collection/how-has-healthcare-utilization-changed-since-the-pandemic/
28. Daly B, Brawley OW, Gospodarowicz MK, et al: Remote monitoring and data collection for decentralized clinical trials. JAMA Netw Open 7:e246228, 2024
29. US Department of Health and Human Services, Food and Drug Administration: Conducting Clinical Trials With Decentralized Elements: Guidance for Industry, Investigators, and Other Interested Parties. https://www.fda.gov/media/167696/download
30. Harvey RD, Miller TM, Hurley PA, et al: A call to action to advance patient-focused and decentralized clinical trials. Cancer 130:1193-1203, 2024
31. Thota R, Hurley PA, Miller TM, et al: Improving access to patient-focused, decentralized clinical trials requires streamlined regulatory requirements: An ASCO research statement. J Clin Oncol 42: 3986-3995, 2024
32. American Society of Clinical Oncology: ASCO Welcomes FDA Efforts to Expand Access to Clinical Trials, Provides Agency with Additional Recommendations. ASCO, 2023. https://society.asco.org/news-initiatives/policy-news-analysis/asco-welcomes-fda-efforts-expand-access-clinical-trials
33. National Cancer Institute Clinical Trials and Translational Research Advisory Committee (CTAC) Strategic Planning Working Group: Working Group Report, 2020. https://deainfo.nci.nih.gov/advisory/ctac/1120/SPWGreport.pdf
34. Doroshow JH: Adapting NCI's Clinical Trials System to a Changed Clinical Research Environment. National Cancer Advisory Board, 2023. https://deainfo.nci.nih.gov/advisory/ncab/0223/Doroshow.pdf
35. US Food and Drug Administration: Diversity Action Plans to Improve Enrollment of Participants from Underrepresented Populations in Clinical Studies. Draft Guidance for Industry, 2024. https://www.fda.gov/media/179593/download
36. Aburto JM, Tilstra AM, Floridi G, et al: Significant impacts of the COVID-19 pandemic on race/ethnic differences in US mortality. Proc Natl Acad Sci USA 119:e2205813119, 2022
37. Eberly LA, Kallan MJ, Julien HM, et al: Patient Characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 3: e2031640, 2020

Advancing Diagnostic Tests and AI-Based Tools

Companion Diagnostic FDA Review Flexibilities: An Assessment of CDx for NSCLC to Support Aligned Approaches for Validation

Hillary S. Andrews[1] · Grace Collins[1] · Bernat Navarro-Serer[1] · Mark D. Stewart[1] · Jeff D. Allen[1]

Received: 6 February 2025 / Accepted: 7 May 2025 / Published online: 15 May 2025

Abstract

The U.S. Food and Drug Administration (FDA) recommends concurrent development of targeted therapies with an associated companion diagnostic (CDx) as the optimal approach to provide patient access to novel, safe, and effective treatments. However, CDx validation often relies on clinical samples from pivotal clinical trials for the drug, which can be challenging, particularly when there is limited sample availability. A review of Summary of Safety and Effectiveness Data (SSED) documents for CDx approved for non-small cell lung cancer (NSCLC) revealed that CDx for rare biomarkers often use alternative samples for validation. While the practice of using alternative samples for validation occurs, it is not always clear when these flexibilities are considered or how alternative samples should be used for validation. To address this, we propose the FDA establish guidance for the use of alternative sample sources for CDx validation, especially for rare biomarkers, to ensure timely and effective patient access to targeted therapies.

Background

Precision medicine is transforming cancer care by enabling the development of therapies tailored to specific biomarkers. However, the success of these therapies depends on reliable companion diagnostics (CDx) that can accurately match patients to treatments. For rare biomarkers, limited sample availability poses unique challenges to CDx validation, potentially delaying patient access to beneficial treatments.

In most instances, the U.S. Food and Drug Administration (FDA) recommends concurrent development of a CDx alongside drugs targeting a specific biomarker [1, 2]. However, this concurrent approval process does not always occur and targeted therapies have, in some cases, been approved without an accompanying CDx for patient selection, particularly when therapeutic clinical trials are expedited to address an unmet need or involve small patient populations [3]. Delayed CDx approval is not preferable, and the application of regulatory flexibility enabling the use of alternative sample sources for CDx validation is one approach that could help address these challenges. However, it requires a clear understanding of when and how such flexibilities should be applied [4].

CDx review and subsequent approval focus on analytical and clinical validation of the test, while assessing safety and effectiveness. For more prevalent biomarkers, FDA recommends clinical validation be performed using samples from the pivotal clinical trial that supports the drug's approval. However, in some cases, FDA guidance acknowledges it may be infeasible to acquire a sufficient number of clinical samples from the pivotal study for retesting, necessitating alternative approaches to validation, such as using a subset of trial samples. Should alternative validation approaches be used, FDA outlines considerations for ensuring validation studies reflect test performance in the intended use population [2].

Review of SSED Documents from CDx Based on Rarity

Understanding how diagnostic test developers address these challenges in obtaining sufficient clinical samples during the validation process is critical for informing best practices and future regulatory requirements. To explore this, we reviewed Summary of Safety and Effectiveness Data (SSED) documents from the Premarket Approval (PMA) application available on FDA's website for CDx used for

✉ Mark D. Stewart
mstewart@focr.org

[1] Friends of Cancer Research, 1800 M Street NW Suite 1050S, Washington, DC 20036, USA

Table 1 Overview of alternative samples used for assessment of clinical performance

Biomarker Group	PMAs	# PMAs Using Alternative Sample Sources with Descriptions	PMAs Using Clinical Trial Samples
Rarest *Prevalence 1–2%* ROS1, [5] BRAF V600E [6]	3	3/3 PMAs • 139 archival specimens [7] • 305 retrospective melanoma samples not obtained from clinical trial [8] • 117 negative commercially acquired samples [7]	2/3
Rare *Prevalence 3–13%* ALK, [9] KRAS G12C [10]	5	2/5 PMAs • 148 supplemental matched tissue and plasma samples from commercial vendors [11] • 303 patients from separate trial used to evaluate concordance between the two sample types [12]	5/5
Least Rare *Prevalence 24–60%* EGFR exon 19 deletions, EGFR exon 21 L858R alterations, EGFR exon 20 T790M alterations, [13] PD-L1 [14]	10	4/10 PMAs • 282 retrospective samples not obtained from a clinical trial [8] • 130(-) FFPE NSCLC archival specimens sourced from commercial vendors [7] • The NILE study provided supplemental samples to calculate NPA ($N=92$)(15) • 35 patients for whom data were previously generated on Guardant360 LDT [15]	9/10

Table 2 Bridging studies with median valid positive and negative samples included by biomarker prevalence

Biomarker Group	# of PMAs with Bridging Results	Valid Positive Samples Included in Bridging (Median (range))	Valid Negative Samples Included in Bridging (Median (range))
Rarest *Prevalence 1–2%* ROS1, BRAF V600E	3/3	67 (25–167)	119 (114–135)
Rare *Prevalence 3–13%* ALK, KRAS G12C	4/5	82 (75–179)	145 (75–754)
Least Rare *Prevalence 24–60***%** EGFR exon 19 deletions, EGFR exon 21 L858R alterations, EGFR exon 20 T790M alterations, PD-L1	9/10	182.5 (72–282)	150 (108–277)
All	**16/18**	**136 (25–282)**	142 **(75–754)**

non-small cell lung cancer (NSCLC). Findings revealed that alternative sample sources were frequently used when samples from pivotal clinical trials were limited (Table 1). For each CDx, we focused on the Primary Clinical Study section in the SSED and searched for terms related to alternative samples. These alternative sample sources included archival specimens, retrospective samples, and commercially acquired specimens. Interestingly, alternative samples were more commonly used for the rarest biomarkers (3/3 PMAs, 100% for the rarest biomarkers vs. 4/10 PMAs, 40% for the least rare biomarkers).

In some cases, pivotal study enrollment is based on one or more Clinical Trial Assays, which may include the candidate CDx test as well as local tests performed at individual trial sites. Bridging studies are then performed to evaluate the agreement between assays (e.g., the enrollment tests vs. the candidate CDx) and to link clinical data from the intended use population to the candidate CDx. This process is critical to ensure the CDx can reliably provide clinically actionable results compared to the local trial assays and supports the demonstration of its safety, effectiveness, and approval. Most CDx included in this analysis required a bridging study (16/18, 89%). We analyzed the number of samples in these bridging studies and found that those for the rarest biomarkers had fewer positive (median 67 [25–167]) or negative samples (median 119 [114–135]), while the least rare biomarkers included the greatest number of positive (median 182.5 [72–282]) or negative (median 150 [106–277]) samples (Table 2).

Regulatory Flexibilities and Consistency in Validation Strategies

Based on findings in the SSEDs, it is clear that regulatory flexibilities are often applied in situations where sample availability is limited. However, the lack of explicit regulatory guidance on when and how these samples should be used may create uncertainty for developers. The FDA could

draw on this prior experience to establish guidance clarifying situations in which regulatory flexibilities would be considered, as well as identifying the types of sample sources that could support robust validation and regulatory review.

Understanding when to use alternative sample sources could support sponsors in determining when to request flexibilities [16]. For instance, rare biomarkers, defined as a prevalence that is less than 1% of the overall population of patients with that specific cancer type, could serve as a starting point for considering flexibility. However, other factors limiting sample availability, along with considerations such as the technology or the intended use of the assay, may influence decisions to exercise flexibility. For example, blood-based biomarkers may warrant additional flexibility due to sample volume limitations.

Additionally, defining the types of data appropriate for different aspects of test validation could help sponsors select appropriate alternative data sources and prioritize limited clinical samples. For instance, cell lines such as immortalized cell lines or primary cultures, could be leveraged for analytical validation to assess interference, reagent stability, or guard banding. However, these would not be appropriate for clinical validation that requires outcomes data or other analytical studies that consider tissue complexity. For these types of studies, prioritization of clinical samples will be important. Overall, it is critical to consider which alternative sample type best represents the information that is necessary to garner from the validation analysis.

Sponsors seeking to use alternative sample sources should engage early with FDA through mechanisms such as pre-IDE meetings or Q-submissions. Early and clear communication with the FDA, supported by well-documented justifications, is critical when seeking flexibility for CDx validation to support an FDA approval. A consistent format for providing this information could streamline the sponsor discussions with FDA and the review process. For example, a document addressing validation study design, proposed samples, sample sources, and justifications for using the samples for each proposed validation study would allow reviewers to efficiently assess strategies [16].

Conclusion

Regulatory flexibilities play a critical role in CDx development when limited samples are available from the pivotal clinical trial. The FDA has considerable experience supporting the use of alternative samples for clinical and analytical validation. Establishing formal guidance on these flexibilities would facilitate more streamlined processes and support sponsors in incorporating alternative samples into their validation strategies, as appropriate. The types of alternative samples appropriate for different validation purposes should be clearly defined to ensure robust validation while maintaining regulatory standards. Additionally, a consistent, well-documented approach across development programs could facilitate efficient communication between sponsors and FDA reviewers, ultimately ensuring that biomarker-based precision therapies reach patients more effectively.

Author Contributions All authors contributed to the conception and design of the work, participated in drafting and critically revising the manuscript for important intellectual content, approved the final version to be published, and agree to be accountable for all aspects of the work to ensure its accuracy and integrity.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

1. U.S. Food and Drug Administration. In Vitro Companion Diagnostic Devices Guidance for Industry and Food and Drug Administration Staff. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/in-vitro-companion-diagnostic-devices
2. U.S. Food and Drug Administration. Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product Draft Guidance for Industry and Food and Drug Administration Staff. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-codevelopment-in-vitro-companion-diagnostic-device-therapeutic-product
3. Jørgensen JT. Missing Companion Diagnostic for US Food and Drug Administration-Approved Hematological and Oncological Drugs. JCO Precis Oncol. 2022;6:e2200100. Available from: https://pubmed.ncbi.nlm.nih.gov/35709402/
4. Bousnina I, Doyle M, Huston M, Kalamegham R, Katrivanos E, Koontz L. Expedited development of diagnostics for therapies targeting rare biomarkers or indications White Paper [Internet]. 2022. Available from: https://friendsofcancerresearch.org/wp-content/uploads/Dx_For_Rare_Biomarkers_White_Paper.pdf
5. Muminovic M, Uribe CRC, Alvarez-Pinzon A, Shan K, Raez LE. Importance of ROS1 gene fusions in non-small cell lung cancer.

Cancer Drug Resist. 2023;6:332. https://doi.org/10.20517/CDR.2022.105.
6. Planchard D, Sanborn RE, NM V, Vaishnavi A, Smit EF. BRAFV600E-mutant metastatic NSCLC: disease overview and treatment landscape. Npj Precision Oncol 2024. 2024;8:1. https://doi.org/10.1038/s41698-024-00552-7.
7. Food US, Administration D. CDRH. Summary of Safety and Effectiveness Data (SSED) for P160045 [Internet]. 2017. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160045B.pdf
8. Food US, Administration D. CDRH. Summary of Safety and Effectiveness Data (SSED) for P170019 [Internet]. 2017. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019B.pdf
9. Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG. Anaplastic lymphoma kinase Inhibition in non-small-cell lung cancer. N Engl J Med. 2010;363:1693–703. https://doi.org/10.1056/NEJMOA1006448.
10. Lim TKH, Skoulidis F, Kerr KM, Ahn MJ, Kapp S JR. F.A. KRAS G12C in advanced NSCLC: prevalence, co-mutations, and testing. Lung Cancer 2023;184. https://doi.org/10.1016/J.LUNGCAN.2023.107293
11. Food US, Administration D. CDRH. Summary of Safety and Effectiveness Data (SSED) for P210040 [Internet]. 2022. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf21/P210040B.pdf
12. Food US, Administration D. CDRH. Summary of Safety and Effectiveness Data (SSED) for P200006 [Internet]. 2020. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200006B.pdf
13. Zhang Y-L, Yuan J-Q, Wang K-F, Fu X-H, Han X-R, Threapleton D. The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. Oncotarget. 2016;7:78985–93. https://doi.org/10.18632/ONCOTARGET.12587.
14. Yu H, Boyle TA, Zhou C, Rimm DL, Hirsch FR. PD-L1 expression in lung Cancer. J Thorac Oncol. 2016;11:964. https://doi.org/10.1016/J.JTHO.2016.04.014.
15. Food US, Administration D. CDRH. Summary of Safety and Effectiveness Data (SSED) for P200010 [Internet]. 2020. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200010B.pdf
16. Andrews H, D'Apote L, Das B, Dickey J, Doyle M, Freaney J et al. Innovative validation and regulatory processes for companion diagnostic tests for rare biomarkers or indications White Paper 2025. https://friendsofcancerresearch.org/wp-content/uploads/Innovative-Validation-and-Regulatory-Processes-for-Companion-Diagnostic-Tests-for-Rare-Biomarkers-or-Indications.pdf

FRIENDS of CANCER RESEARCH

Innovative Validation and Regulatory Processes for Companion Diagnostic Tests for Rare Biomarkers or Indications

White Paper | 2025

Executive Summary

Precision therapy has become a leading approach for oncology treatment, showing continued success in improving outcomes for patients with cancer over the past 20 years. The U.S. Food and Drug Administration (FDA) reviews and approves the diagnostic tests critical for identifying patients who may benefit from precision therapy as companion diagnostics (CDx). The review process includes analytical and clinical test validation, which often requires an abundance of clinical samples. However, in situations where the biomarker or the cancer is rare, there are often limited clinical samples from the clinical trial, making it challenging to perform all necessary test validation studies. To overcome this challenge, drug sponsors and diagnostic test developers may consider using alternative sample sources for validation, such as procured human samples or contrived samples. While the use of alternative sample sources to support regulatory approval of a CDx for a rare biomarker has been in practice for some time, sponsors may lack an understanding of when this flexibility is warranted and how various alternative sample types should be considered for each validation analysis. Friends of Cancer Research convened a working group of experts to align on an approach to determine when regulatory flexibility might be considered, identify possible alternative samples, and suggest opportunities for using the samples in validation studies, including potential ways to support more streamlined discussions on validation plans and strategies between sponsors and FDA.

Authors

Hillary Andrews, Friends of Cancer Research

Lucia D'Apote, Amgen

Biswajit (Bishu) Das, Frederick National Laboratory (NIH)

Jennifer Dickey, PGDx/LabCorp

Megan Doyle, Eli Lilly and Company

Jonathan Freaney, Tempus AI

Elaine Katrivanos, Tempus AI

Laura Koontz, Foundation Medicine, Inc.

Joe Lennerz, BostonGene

Dun Liang, Eli Lilly and Company

Elizabeth Mansfield, Foundation Medicine, Inc.

Carly McWilliams, Roche Diagnostics

Gary Pestano, Biodesix

Brianna Phillips, Biodesix

Alain Silk, Tempus AI

Nino Sireci, Eli Lilly and Company

The authors thank the FDA for their participation as observers on the working group.

Table of Contents

Introduction

Rapid technological innovations and a deeper understanding of cancer biology have driven advancements in precision oncology. As treatments become increasingly tailored to the unique characteristics of each cancer, the need for diagnostic tests to identify rarer biomarkers for diagnosis, prognosis, and therapeutic decision-making has grown significantly. Especially for rare biomarkers and indications affecting a small subset of the population diagnosed with cancer (see definition of rare biomarker on page 6), there can be inherent challenges for validating assays as companion diagnostics (CDx), such as difficulty obtaining sufficient quantities of well-characterized samples and limited established reference materials. As such, it is critical to assess the current regulatory frameworks and propose strategies to facilitate continued advancements, particularly for the evaluation of diagnostic tests for rare biomarkers or indications. It is especially important to consider flexible validation approaches to ensure patients have access to validated CDx for rare biomarkers or indications in a timely manner. While the regulatory landscape continues to evolve, challenges with validating these diagnostic tests will likely remain.

One of the most frequent challenges for the validation* of a rare biomarker test is obtaining sufficient quantities of well-characterized clinical samples in a timely manner to perform analyses due to small patient numbers. In situations where the method employed to determine positivity by the test is novel and thus not regularly used in routine clinical practice, the ability to screen and identify positive samples is particularly challenging. To supplement these data, companies may need to invest considerable resources and time to acquire and screen a large number of samples to identify biomarker-positive samples. Identifying and employing alternative data or sample sources to support test validation is critical and needs to be conducted thoughtfully and collaboratively with drug sponsors, diagnostics developers, and the U.S. Food and Drug Administration (FDA).

Friends of Cancer Research (*Friends*) organized a collaborative working group of experts to propose potential approaches to facilitate oncology diagnostic test validation for rare biomarkers and indications. The considerations discussed here focused on scenarios where a diagnostic test is being validated for a rare biomarker or indication in preparation for a CDx premarket submission to FDA. These considerations may also be applicable in other regulatory contexts for biomarker testing.

The group identified three objectives for discussion:

- Identify situations where regulatory flexibilities would be appropriate and facilitate validation of diagnostics for rare biomarkers and indications in oncology.
- Develop approaches for leveraging alternative sample sources or data to support validation strategies.

* The term "validation" is used to refer to the establishment of specific performance characteristics, including (but not limited to) accuracy, precision, sensitivity, specificity, range, reference intervals, or other required performance characteristics.

- Outline a framework for capturing key information to support the proposed validation strategy, particularly when using alternative samples, to ensure clarity in premarket submissions.

Considerations for Regulatory Flexibility

Various biomarker, assay, and disease attributes inform the benefit-risk and safety and effectiveness assessment by regulators, which may influence the level of regulatory flexibility appropriate for a specific diagnostic test and validation strategies, such as the use of alternative samples or other data sources. Herein, we provide some broad categories of attributes to consider, ranked by their potential impact on decisions about flexibility. Categories for consideration should not be taken in isolation, but the sum of considerations can be used to support proposals for flexibility.

Biomarker and Indication Prevalence

A key aspect of when regulatory flexibility should be considered is the size of the population with the specific biomarker or indication, as a smaller population can make it more challenging to identify samples with the biomarker for assay validation. We propose that these flexibilities be considered for biomarker-defined subsets of cancer types with an estimated prevalence of 1% or lower (in the population of patients with that specific cancer type in the U.S.) or for rare cancer types with an estimated total prevalence of 1% or lower (in the overall population of patients with cancer in the U.S.) Determining whether a biomarker or indication qualifies as "rare" should rely on reasonable estimated prevalence with appropriate data and justification but should not be considered an exclusive criterion for applying flexibility. Biomarker or indication prevalence or incidence can be difficult to accurately identify, as it can change over time (e.g., if left undiagnosed or untreated), or can be unknown, especially among racial and ethnic minorities.[1–4] Additionally, the novelty of the biomarker may impact the degree to which testing is employed in routine clinical practice and the degree to which samples with the biomarker of interest are available at commercial biobanks.

Sample Availability

In addition to biomarker prevalence, other factors can influence the availability of clinical samples for assay validation. In certain populations, accessing adequate tissue to perform the necessary clinical and analytical validation analyses may be challenging. The location of the tumor and the risk associated with procedures required to obtain the sample may result in limited tissue availability (e.g., lung cancer). For liquid biopsies, sample volume is typically restricted, with blood providing the highest volume (albeit still limited) and other fluids, like cerebrospinal fluid and aqueous humor of the eye, yielding even less. Additionally, to represent the expected testing scenario, the sample would need to be collected in the appropriate compatible sample collection tube, which is not always the case. Some sample types, like whole blood, extracted mRNA, or frozen tissue, may degrade beyond usability faster than other sample types (e.g., FFPE), impacting the ability to do testing at later timepoints. This can be a practical issue when the time of sample collection and the time of validation extend beyond stability expectations for the analyte. Additionally, patients in biomarker driven trials are often initially screened at local sites for enrollment and, as a result, tissue for bridging studies or to support analytical validation may be more

difficult to obtain. This is especially challenging for rare biomarkers, where the test may not be part of standard practice so there are overall fewer patients who are screened at local sites. Additionally, samples may be exhausted by the time the trial initiates due to their use in supporting clinical care and management and other stakeholders (e.g., patients, providers, Institutional Review Boards) may be resistant to subjecting patients to an additional biopsy for the purposes of test validation. Ethical considerations related to biopsy for the sole or primary purpose of supporting analytical testing may preclude additional sample collection.

Unmet Needs and Expedited Review Pathways

In situations where the CDx is co-developed with a drug for a serious or unmet medical need that has Breakthrough Therapy Designation (BTD) or is ultimately approved through the Accelerated Approval pathway, development timelines may be condensed. Contemporaneous approval of the drug with a CDx ensures patients are appropriately identified and have access as soon as the product is approved.[5] However, aligning the timing of the development and approval of both the drug and the device can be challenging, particularly within the expedited timelines of FDA's drug development programs.

Approaches for Leveraging Data Sources for Validation

For rare biomarkers where accessing adequate tissue to perform all necessary assessments for assay validation may be difficult, other data sources could be considered to support the premarket submission for the CDx. Assay validation includes clinical validation, which for a CDx refers to the accuracy with which the test identifies the patients for whom the therapy is safe and effective, and analytical validation, which focuses on ensuring tests are accurate, precise, specific, and reliable.[5,6] The following sections outline various data sources, their proposed use in clinical or analytical validation, and opportunities and challenges for using each (also outlined in **Table 1**). The text and the table suggest prioritization for using samples in different types of validations. Examples are also provided for situations where the various data sources could be considered with appropriate justification.

Clinical Trial Samples

In general, clinical trial samples from the corresponding pivotal study should be prioritized for clinical validation as these samples represent the intended use population. Ideally, the candidate CDx will be used to identify all patients for inclusion in the therapeutic product's pivotal study; however, in some cases, the pivotal study enrolls patients using one or more Clinical Trial Assays (CTA) and may also include the candidate CDx test and local testing. Bridging studies assess agreement between assays (e.g., the enrollment tests vs. the candidate CDx) to bridge the intended use population clinical data from the enrollment tests to the candidate CDx to evaluate safety and effectiveness and support approval. Thus, remaining patient samples from the enrollment tests or local testing should be prioritized for conducting bridging studies. In some cases, these samples are saved as pre-processed samples such as extracted DNA or RNA from clinical trial studies and can be considered for use in CDx clinical validation. If samples are extracted using a different method/process than the one specified for the

candidate CDx, information is needed to demonstrate equivalent performance across the different method(s)/process(es).

Therapeutic clinical trial sponsors should prospectively plan for storing archival tissues or nucleic acids from the pivotal clinical trial and ensure they obtain and retain patient consent to use these tissues for test development. These archived samples may be useful for follow-on CDx development (e.g. to support the development of a liquid biopsy CDx if only a tissue-based CDx exists), or to support the need to develop multiple different CDx in various geographies. However, some archived patient specimens may be of lower-quality, and for the reasons noted above regarding ethical and practical challenges with obtaining additional biopsies, sponsors may not have sufficient samples for all activities and must therefore determine how to prioritize the use of available samples.

Provided the clinical trial samples from related clinical studies (e.g., earlier phase study in the same development program) were not used to develop the candidate CDx, these samples may be prioritized for clinical validation when the intended use population is the same as the pivotal trial and can supplement the available samples for clinical concordance studies. Biomarker-negative samples are necessary for bridging and clinical concordance studies; however, these may not be included in the target trial design due to the lack of anticipated effect in patients without the biomarker, raising ethical questions about the enrollment of biomarker-negative patients. In this scenario, it will be challenging or impossible to have sufficient clinical trial-enrolled biomarker-negative samples due to the selection criteria and limited inclusion of these patients with biomarker-negative tumors in the trial. However, this should not preclude development of a plan that includes storing biomarker negative specimens from patients that were not enrolled in the trial. In addition, well-characterized negative samples from related studies or normal healthy donors for blood-based biomarkers could be considered. Specifically, early phase trials may include biomarker negative samples that may be valuable for negative percent agreement (NPA) analyses. The value of these samples is that drug sponsors have control of the trial, the samples, and their availability, and thus similar performance could be expected. However, for analytical validation studies, the stage of the disease may not be significant in certain scenarios, such as when analyzing driver mutations in tissue samples. Stage can be highly relevant in other situations, like when dealing with resistance mutations or using ctDNA approaches, where the extent of tumor shedding can vary significantly. Therefore, whenever proposing to use samples from related trials or a different cancer stage, proper justification should be included.

Representative Clinical Approaches

Trial samples from related clinical studies, or samples from routine clinical testing of different cancer types (e.g., lung vs. colon) or specimen types (e.g., biopsy type or fixation) could be considered for analytical validation, provided such samples are applicable and relevant to the intended use of the candidate CDx. It is important to consider whether there are any differences in analytical validation due to the specimen or cancer type and to describe the rationale for using these samples. There is potential to use a more prevalent cancer type (e.g., lung cancer) for analytical validation of a CDx for a tissue agnostic indication being used in a rare tumor type (e.g., pediatric brain cancer). These samples could

also be considered when a drug is tested for a different indication where the sample type and biomarker tested are the same as the related study.

An alternative approach is to use clinical specimens that are not necessarily reflective of the intended use population to leverage representative validation approaches. Generalized conclusions about analytical validity can be based on a broad sampling of variants in the same class (i.e., substitutions, insertions, deletions, etc.) in various contexts across the queried genome. This approach may be particularly useful for assessing rare genetic variants in similar genomic contexts (e.g., GC regions, same chromosome) to other more prevalent variants. Whether there are opportunities to use a similar approach for other assay modalities beyond nucleic acid sequencing (e.g., IHC) should be explored.

Real-World Evidence (RWE)

There are opportunities to track patients in real-world settings who have been tested with a diagnostic that could be developed as a CDx and who have also received a therapy of interest. Such real-world data (RWD), when appropriately gathered and analyzed, may be proposed to support clinical validation. Leveraging RWD provides value not only for assessing clinical outcomes at single time points but also for tracking outcomes over time. However, RWD from electronic health records will likely differ from the data collected in a clinical trial, which may lead to inconsistencies in data interpretation. For example, measurements of progression in the real world often do not apply RECIST criteria and may occur with a different periodicity. These factors should be considered and addressed in proposals to use RWD in CDx regulatory filings.

RWE developed from incidence rates in developer databases can be supportive in post-market settings to demonstrate non-specific comparability with other assays measuring the same biomarker but would not be useful for demonstrating safety and effectiveness of a CDx. For example, knowing that the prevalence of ALK alterations is 3-7% in the general population, a developer might demonstrate the same rate of ALK alterations in their real-world NSCLC dataset. In any approach using RWD, use of different versions of an assay (e.g., design iteration) could confound analysis and clear explanations about the potential impact of assay versions should be described.

Procured Human Specimens

Procured human specimens that are similar to the intended use population can be purchased from a vendor, identified from data repositories or representative archival tissue, and are often useful for analytical validation, including determining the limit of detection (LOD), accuracy, precision, and other key analytical studies related to the specimen (e.g., stability). Additionally, since clinical trials often enroll only biomarker positive patients, sample procurement provides an alternative approach to identifying biomarker negative samples that could be used for analytical validation studies.

In some cases, the specimen may be from the appropriate intended use population, but the sample acquisition method may differ. Differences could include either the approach for sample collection (e.g., biopsy vs. a fine needle aspirate vs. a cytology smear) or the sample preservation approach (e.g., FFPE block vs. a frozen tissue that was secondarily fixed, or plasma collected in a K2EDTA tube and frozen vs. a Streck cfDNA BCT shipped at an ambient temperature). In each case, there may be implications for the

analytical analysis, which should be clearly described. These factors should be considered and addressed in proposals to use procured samples for validation in CDx regulatory filings.

Cell Lines

Cell lines, including immortalized cell lines with the biomarker of interest, those with CRISPR or other genetic modification to have the biomarker of interest, and primary cultures or organoids, can be considered for analytical studies. When appropriately validated, these cell lines may be beneficial for accuracy, precision, interference, reagent stability, guard banding, and other studies. Cell line identity and validity for use may vary depending on the supplier. The benefit of validated cell lines is that they can be processed to simulate tissue processing (i.e., freezing or FFPE embedding, as appropriate). However, the samples may not reflect the tumor tissue complexity of clinical samples and so may not be feasible for analyses that require tissue architecture (e.g., analytical validation of IHC) or where sample-based interfering substances are problematic. Further, prolonged culture of cell lines can lead to genetic drift, making them less representative of the original tumor. These factors should be considered and addressed in proposals to use cell line data for regulatory use in CDx regulatory filings. Such proposals should also include information to support that the performance in cell lines is not different from the performance in clinical intended use specimens.

Contrived Samples

Contrived samples such as analyte spike-in, synthesized DNA, and double-stranded DNA fragments may be useful to supplement human samples in analytical validation studies such as linearity, stability, precision, interfering substances, and dilution studies to assess limits of detection. These samples could be especially helpful in studies where large numbers of replicates are required. When appropriately validated, there is confidence that the biomarker is present. The variant type (e.g., substitutions, insertions, deletions) and level (e.g. allele frequency) can be specified and customized. Contrived samples may be especially beneficial when validating highly sensitive assays, such as liquid biopsies assessing ctDNA. In this case, distinguishing a few particles of cancerous DNA from billions of non-target molecules can be challenging. Purpose-built, patient-like contrived reference materials built using, for example, a 'plasma in plasma' approach could be used to address this challenge.[7] Ensuring "spike-in" samples are prepared using an appropriate background/matrix to mimic the intended use specimens to the extent possible is important. Strengths and weaknesses of contrived samples should be considered and addressed in CDx regulatory filings. Such proposals should include information to demonstrate the performance in contrived samples does not differ from clinical intended use specimens.

In Silico Datasets

In silico datasets can be considered for analytical validation, specifically focused on validating the bioinformatics pipeline and other informatics components. Appropriately constructed and relevant in silico datasets are stable and may be useful to re-validate an assay after a software or hardware change. It is important that the dataset used to train the algorithm is not used for validation. An in silico validation approach requires close alignment between the in silico dataset and the specific wet lab procedures,

making it challenging to establish a standardized, off-the-shelf solution using in silico reference datasets. Specific approaches to dataset construction and the ability to query important bioinformatics functions should be considered and addressed in proposals to use them for validation in CDx regulatory filings.

Table 1. Overview of Data Sources and Sample Types in Assay Validation

Data or Sample Type Category (Examples)	Use in Validation	Potential Advantages	Challenges
Clinical Trial Samples (e.g., tissue or nucleic acids from either the pivotal trial or other clinical trials with the same intended use population)	Prioritize for clinical validation (e.g., bridging studies)	• If from the therapeutic pivotal study, validation performance represents best evidence for clinical validity • Use of clinical specimens from related trials that were not used for diagnostic test development and are adequately representative of the intended use population may have well-characterized negative samples	• Pivotal trial samples are often limited and may not be available or appropriate for analysis with the CDx • Matched efficacy data may be unavailable for related trials
Representative Clinical Approaches (e.g., trial samples from related clinical study with different cancer (lung vs. colon) or specimen type (biopsy type or fixation), or clinical specimens that are not reflective of the intended use population)	Analytical validation	• Validation performance expected to be similar to validation performance in the intended use clinical specimens • For procured samples, broad sampling of variants in the same class may support generalized conclusions about analytical validity	• There may be nuanced differences between cancer types that could affect the interpretation of either analytical or clinical validation data. Therefore, any anticipated differences should be clearly described along with a biological and/or technical justification for the use of samples from different cancer types • The biomarker prevalence may be low in other cancer types as well
Real-World Evidence (e.g., tracking patients in real-world settings who have been tested with the candidate CDx and received therapy of interest)	Clinical validation	• Can include clinical outcomes and track outcomes over time • Data reflects real-world use of assays and therapies	• RWD (e.g., data from clinical practice in an EHR) will differ from data collected in a clinical trial, potentially leading to inconsistencies in data interpretation ○ Measurements of progression in the real world often do not apply RECIST and may occur with a different periodicity • Use of different versions of an assay (e.g., design iteration) could confound analysis

Data or Sample Type Category (Examples)	Use in Validation	Potential Advantages	Challenges
Procured Human Specimens (e.g., purchased from a vendor, identified from data repositories, or representative archival tissue)	Analytical validation (e.g., LOD, accuracy, precision, and other key analytical studies related to the specimen (e.g., stability))	• Commercially available • Includes the complexity/difficulty of the human specimen • Technological and/or biological justification can support representative variant detection across the genome, or within a specified genomic context • Existing data and metrics can support performance and justification	• Sometimes expensive • Commercial availability may still be a problem for some rare cancers or rare biomarkers
Cell Lines (e.g., CRISPR, immortalized cells, primary cultures/ organoids)	Analytical validation (e.g., interference, reagent stability, input of intermediate steps, guard banding, etc.)	• Materials processing can simulate tissue processing (e.g., freezing or FFPE embedding, as appropriate) • When appropriately validated, confidence that biomarker is present • Defined quality and abundant quantity • Useful when testing accuracy and reproducibility at lowest analyte levels	• May not reflect the tumor tissue complexity of clinical samples • Prolonged culture of cell lines can lead to genetic drift, making them less representative of the original tumor • Analytical validity may vary depending on the supplier • Not feasible for analyses that require tissue architecture (e.g., IHC) • Does not reflect the intended use population

Data or Sample Type Category (Examples)	Use in Validation	Potential Advantages	Challenges
Contrived Samples (e.g., analyte spike-in, synthesized DNA, double-stranded DNA fragments)	Analytical validation (e.g., linearity, stability, reproducibility, interfering substances, etc.)	• When appropriately validated, have confidence biomarker is present • Can use patient-derived materials (e.g., plasma in plasma approach) • Can perform robustness or process studies where large numbers of replicates are required • The variant type (e.g., SNPs with certain INDEL) and level (e.g. allele frequency) can be specified and customized • Can be sensitive enough to assess the most sensitive assay (e.g., ctDNA assay or highly sensitive flow cytometry)	• May not reflect the complexity or variability of actual patient samples • Does not reflect the intended use population
In Silico Datasets (e.g., sequence coverage at biomarker positions, can include synthetic approaches)	Validating bioinformatics pipeline and other informatics components	• Could be used to re-validate the assay after software or hardware changes • Does not expire	• Cannot be used to validate the wet lab portion of an assay • Requires sequencing to align with wet lab approach (i.e., challenging to establish off the shelf approach) • May not be useful across different platforms as the data processing can be mismatched to the specific analytical platform used by a test • The reference human genomes used may be different in different pipelines and can impact variant calls

Opportunities for Consistent Data Reporting and Regulatory Discussions

For discussions with FDA regarding premarket submissions of CDx for rare biomarkers, consistent descriptions of the validation strategy, including suggested samples and justifications, is important. In the **Appendix**, we provide an example snapshot to aid in sharing the validation strategy during the pre-submission and marketing application, which may be updated based on feedback throughout the development process. Using this or a similar approach would provide FDA with a clear understanding of the evidence used for validation and accompanying justification. Additionally, this snapshot could help drug sponsors and diagnostics companies align on the approach for the drug and CDx review and approval. We recommend discussing the co-development program with the FDA as early in development as possible.

Each development program will have different needs and considerations for the justification for flexibility and sample selection. However, some consistent recommendations apply. In general, clinical samples from the intended use population, particularly those with clinical outcomes data, should be prioritized for clinical validation. This is especially important for complex biomarkers, such as those incorporating sophisticated algorithmic analyses, to ensure accuracy in the clinical state or cutoff determination. For novel biomarkers, readily available reference standards and clinical samples may be limited, as testing for these biomarkers is not yet routine in clinical practice. In each case, adequate justification for the selected data source should be included.

Conclusions and Next Steps

Regulatory flexibilities can aid in demonstrating a favorable benefit-risk profile for rare biomarkers and indications, especially where there are limited clinical trial samples for validation studies. Various alternative evidence sources (e.g., samples, data, etc.) can support clinical and analytical validation for CDx biomarker tests when specimen availability is limited. Sponsors should provide an explanation for why samples would be limited and discuss plans for using alternative data or samples for validation, including a well-justified rationale for their use in early conversations with the FDA. Sponsors could consider using the proposed snapshot document in the Appendix to more effectively facilitate these discussions.

As drug development for cancers with rare biomarkers expands, consistent approaches to clinical sample storage and alternative sample selection for validation are increasingly important. To maximize the availability of trial samples, drug and device sponsors, working together, should establish proactive plans for preserving samples from all phases of clinical trials. Additionally, the field should consider aligned approaches for establishing validated reference materials and methods, for example, datasets with well-annotated samples that could support both already approved products and the rapid development of reference information for novel, rare biomarkers, which may allow for more standardized characterization of assay performance.

References

1 Aldrighetti CM, Niemierko A, Van Allen E et al. Racial and ethnic disparities among participants in precision oncology clinical studies. JAMA Netw Open 2021;4(11):e2133205. DOI: 10.1001/jamanetworkopen.2021.33205

2 Cheung ATM, Palapattu EL, Pompa IR et al. Racial and ethnic disparities in a real-world precision oncology data registry. NPJ Precis Oncol 2023;7(1):7. DOI: 10.1038/s41698-023-00351-6

3 Overman MJ, Morris V, Kee B et al. Utility of a molecular prescreening program in advanced colorectal cancer for enrollment on biomarker-selected clinical trials. Ann Oncol 2016;27(6):1068–74. DOI: 10.1093/annonc/mdw073

4 Dienstmann R, Garralda E, Aguilar S et al. Evolving landscape of molecular prescreening strategies for oncology early clinical trials. JCO Precis Oncol 2020;4:505–13. DOI: 10.1200/PO.19.00398

5 U.S. Food and Drug Administration. Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product: Draft Guidance. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-codevelopment-in-vitro-companion-diagnostic-device-therapeutic-productThe Clinical and

6 Laboratory Standards Institute (CLSI). https://clsi.org/

7 Stetson D, Labrousse P, Russell H et al. Next-generation molecular residual disease assays: do we have the tools to evaluate them properly? J Clin Oncol 2024;42(23):2736–40. DOI: 10.1200/JCO.23.02301

Appendix

Proposed Snapshot for Alternate Data Use for CDx in Rare Biomarker Validation

1. *Include a paragraph that provides justification for the biomarker of interest being considered as a rare biomarker with citations:*
    - What is the biomarker? What are the incidence and prevalence (either overall or in the specific cancer type of interest)?
    - How many clinical samples do you anticipate having access to? Explain why you believe the use of alternative evidence is necessary.
2. *Complete the table below for each proposed validation study (An example follows with proposed language in red. There should be one table for each validation study.)*

Category	Description
Validation Study Describe which study you will be using the proposed samples for	
Proposed Samples Describe the samples and include the anticipated sample size	
Sample Source Describe how the samples are procured	
Sample Justification Describe the justification behind using these samples	

Category	Description
Validation Study Describe which study you will be using the proposed samples for	Analytical validation - limit of detection
Proposed Samples Describe the samples and include the anticipated sample size	Human specimens containing the biomarker of interest from five different patients with alternative cancer types (i.e., samples representing different cancer types than the specific cancer type of interest).
Sample Source Describe how the samples are procured	All samples will be residual clinical specimens processed for routine laboratory testing and/or sourced from a biorepository. Dilutions to establish limit of detection will be prepared and analyzed by diagnostic test sponsor.
Sample Justification Describe the justification behind using these samples	Limit of detection confirmation studies require more samples than are available for [biomarker/specimen type], due to rarity of biomarker. Limit of detection is an analytical validation analysis that does not rely on clinical outcomes. As such, we are proposing to use procured samples that have [biomarker of interest] to support limit of detection confirmation analyses. The assay analyzes extracted nucleic acid. There are no unique biological characteristics of the biomarker, or biological differences between cancer types, that would make evaluation of limit of detection dependent on the cancer type from which nucleic acid is extracted. The specific variants tested for limit of detection using alternative cancer types will be representative of the specific variants relevant to the intended use population. Therefore, we believe that the limit of detection for [the biomarker of interest] can be appropriately confirmed using alternative cancer types.

FRIENDS of CANCER RESEARCH

Considerations for Developing Reference Data Sets for Digital Pathology Biomarkers

Discussion Document | 2025

Considerations for Developing Reference Data Sets for Digital Pathology Biomarkers

Discussion Document

Digital pathology enables innovative approaches for biomarker interpretation, cellular evaluation, and diagnosis. These approaches can leverage computational pathology models developed using artificial intelligence (AI) (including machine learning [ML] models) to aid in image analysis. While this holds the promise of enhancing accuracy, reproducibility, and standardization of pathology-based features to measure prognostic and predictive biomarkers, expedite diagnosis or pathological scoring, and identify novel biomarkers, there is currently a lack of robust publicly available data sets to support development and validation, and ensure consistent performance of different computational pathology models. Developing reference data sets of images and associated metadata can be challenging, requiring substantial time and money; however, adequately built data sets can support future platform development and validation and address concerns around model accuracy, reproducibility, reliability, and comparability.

Friends of Cancer Research (*Friends*) leveraged expertise from the ongoing Digital PATH Project working group, which included representatives from industry, the U.S. Food and Drug Administration (FDA), the National Cancer Institute (NCI), patient advocates, and academia, to discuss the promise of reference data sets. The Digital PATH Project evaluated the variability in HER2 assessments in a single breast cancer data set across multiple computational pathology models.[1] This document reflects a series of discussions on considerations for developing a reference data set, intended to spark ideas and facilitate further exploration of these critical topics to support future model development and validation.

Independent Reference Data Sets Provide Value

Multiple computational pathology models are often under development (by different developers) to assess the same biomarker and variability across models can cause challenges. The potential challenges are the same as when different immunohistochemistry (IHC) assays (i.e., without the application of computational pathology) are developed for the same biomarker. For example, multiple PD-L1 IHC assays were independently developed for various anti-PD-(L)1 therapies, each using different antibodies, scoring methods, and cut-offs. Analytical validation was performed on independent commercially acquired sample sets and clinical validation established using each developer's individual clinical trial data sets, resulting in inconsistency in how these IHC assays and scoring methodologies compare.[2, 3] To prevent similar challenges for future computational pathology models developed to assess the same biomarker, publicly available (non-proprietary) reference data sets can support an understanding of performance characteristics across multiple

models. Publicly available reference data sets also have the potential to enable more efficient regulatory evaluation of computational pathology models.

Many organizations have identified a need for reference data sets for assay validation,[4–6] including specifically for digital and computational pathology and AI-based models.[7–9] Unlike the development of reference data sets for assays requiring blood or tissue, data sets for digital pathology-based biomarkers are not limited by the constraints of obtaining and storing biological material. The banking of digitized slides is more feasible and provides the opportunity to develop reference data sets. Under the condition that a reference set is robustly built with relevant metadata and samples representative of the intended use of the assay, the data set can provide an objective measurement of model performance on data independent from any training data or data set unique to a specific model.

Ongoing efforts to develop digital pathology data sets largely source samples from individual academic sites.[10] These existing digital and computational pathology data sets are composed of various types of data (e.g., tissue and slide processing characteristics, and image acquisition characteristics, metadata, pathologist annotations, and clinical outcomes) with each uniquely contributing to their intended uses. However, single-source data sets may lack demographic and/or clinical characteristic representativeness of the larger patient population.

Reference data sets provide value to various groups who develop and use these models. Developers can have easy access to a rich data set to validate their model and assess performance. Drug developers can elucidate performance across multiple models to inform use in drug development and potential labeling. Clinician end-users can make informed decisions on model use as they understand the comparability of different models with a reference standard. Lastly, FDA's review process can evaluate validation with robust reference data sets as part of the body of evidence supporting regulatory decision-making.

Considerations for Developing Reference Data Sets

Possible Intended Uses of Reference Data Sets

Reference data sets may be developed for a variety of purposes and a single reference set may be leveraged to assess multiple aspects of model performance and multiple intended use populations, allowing for pre-specified analysis of data subsets in accordance with the specific use of a particular model. Alternatively, a reference data set may be designed to focus on one aspect of performance in specimens with particular characteristics. Table 1 provides an overview of possible intended uses for a reference data set. Analytical and clinical validation may require different types of data.

Each reference data set should be accompanied by a statement of its intended use(s), and the intended use(s) of any model evaluated on such a data set should be described. Any models assessed against the reference data set for validation purposes should be locked at the time of assessment, and the reference data set should be used for fully independent external validation of the model and not for training. To ensure the data set's utility is not limited to a single use, considerations for developer blinding and traceability between model versions should be explored. These approaches can help mitigate risks of overfitting to the reference data set and enable its reuse for testing modified versions of models.

Table 1. Possible Intended Uses of Reference Data Sets for Model Validation.

Intended Use	Performance Assessment	Considerations
Analytical Validation		
Accuracy	Demonstrate the extent to which the test model scores agree with the reference standard widely accepted as producing "truth"	• Accuracy refers to the assessment of the test compared to a reference standard, rather than the average of multiple values used as a proxy for the reference standard; therefore, assessment of accuracy using a consensus of multiple values is technically not a true measure of accuracy but may be necessary given the challenge to have a true "gold standard"
Precision	Demonstrate that the test model provides consistent scores when presented with similar or related inputs or under different conditions	• Reference data sets should include scenarios that capture known sources of variability, such as rescans of the same slide, scans of sequential sections, or scans of different biopsies from the same patient • The study design including number of replicates and samples, and factors considered, should be clearly defined[11]

Intended Use	Performance Assessment	Considerations
Interchangeability	Demonstrate that the test model scores are within the range of scores from multiple pathologists and/or current models (i.e., how well multiple raters agree when assessing the same sample)	• Recognizes that there may not be a singular, cost-effective, and independent reference standard, given the variability in models' and pathologists' scoring • The number and the set of readers providing the reference scores will impact the assessment of interchangeability, and careful consideration is needed to ensure the readers are appropriately selected and trained • Determine how inter-rater reliability and agreement will be assessed depending on the measurement scale • The study design should be clearly defined to ascertain whether any factors other than raters (models) are changing
Clinical Validation		
Clinical Specificity	Assess the proportion of patients who are "negative" for the clinical outcome (denominator) who are correctly identified as "negative" by the digital pathology biomarker (numerator), e.g., assessment of tumor response by biomarker status	• For non-binary clinical outcomes, such as time-to-event outcomes, other metrics of clinical performance may be needed • The definition of biomarker-positive, as well as the clinical outcome, will greatly impact the clinical specificity • Findings may be specific to the clinical setting, including disease type and stage,

Intended Use	Performance Assessment	Considerations
Clinical Sensitivity	Assess the proportion of patients who are "positive" for the clinical outcome (denominator) who are correctly identified as "positive" by the digital pathology biomarker (numerator)	and particular treatment or drug mechanism of action • For predictive biomarkers, clinical sensitivity and specificity, as defined by patients' treatment benefits, cannot generally be estimated without restrictive assumptions. Therefore, a direct assessment of treatment effect (e.g., average probability of difference in "negative" outcomes versus biomarker values) could be employed[12, 13]

Intended Use of the Models

When leveraging reference data sets to assess performance, it is important to consider the intended use of the model, as well as the purpose of the reference data set, to ensure the intentions are aligned and the reference data set has the appropriate applicable data for the model's intended use. In practice, the intended use of models may be very diverse, and therefore, it may be difficult to develop a reference data set that is relevant to all models. Considering the utility of the reference set, given the current state of the intended uses of the models being developed, is important to ensure the reference data set is as broadly usable as possible.

One important consideration for intended use of the model is whether it will be used as a standalone test or to aid or assist the pathologist in interpretation or scoring. For standalone use, a more comprehensive and detailed understanding of performance compared to the reference standard might be important, and the reference data set would need to cover a broader range of potential cases/samples. As there is no need to involve a reader end user with standalone models, assessing performance on a reference set can be completed quickly. For pathologist-aided models, the reference data set might focus on borderline cases or a larger proportion of cases where there is known to be a higher degree of variability in pathologists' scoring. These models will likely require a reader study (e.g., comparing the reader with and without the model) which takes time and may encounter feasibility challenges depending on the size of the data set. However, even if the model is intended to be used as an aid, assessment of standalone performance is usually desired.

Determining how to store, back up, and audit digitized slides will be critical, as the reference data set will likely require considerable memory storage space and cost to host the images and associated metadata. In addition to data storage, a platform to interface with model developers and allow for bidirectional data transfer will be necessary. The design of the infrastructure should align with the intended use(s) of the reference data set and the models it supports, ensuring these priorities guide subsequent decisions. The whole-slide images (WSIs) could either be transferred to the model developers, without the metadata and reference standard assessments, or the WSIs could remain sequestered on a platform within a federated framework that allows models to be executed or evaluated without requiring the WSIs to be transferred. There also needs to be a mechanism for analyzing the model output compared to the reference standard, which could be conducted by a third party. It is important to consider whether the model results remain blinded and what data would be made available to the model developer after conducting the analyses. Key governance considerations, such as contributions, quality control, accessibility, versioning, and validation, will be necessary for ensuring the data set's integrity and alignment with its intended use.

Considerations for Defining a Reference Standard

A single reference standard is necessary to establish accuracy. The current reference standard for many pathology-based biomarkers is generally considered to be the pathologist rendering an interpretation using a light microscope, which differs from reading a digital image. Given the variability in pathologists' manual biomarker readings, there may not be a single reference standard (i.e., "gold standard") for the biomarker for analytical validation. As such, a single reference standard can be based on a consensus across multiple pathologists. As biomarker development continues, including the development of novel biomarkers assessed by AI-models, pathologists' scores may not be feasible as a reference standard (e.g., HER2-ultra low may not be amenable to reproducible determination by pathologists). When considering the definition and measurement of the reference standard, there are strengths and limitations to various approaches, highlighted in Table 2. Additionally, the reference standard should align with the intended use of the data set to ensure relevance and applicability.

The target performance of a model in the assessment of accuracy compared to a given reference standard will depend on multiple factors. Performance targets for a biomarker assay will be context-dependent influenced by the level of risk due to inaccurate biomarker identification and its impact on clinical predictions and outcomes from clinical management decisions (e.g., treatment selection) guided by those predictions. Guidance on target performance goals would be helpful.

Rather than defining a single reference standard and performance goal for a model, it may be necessary to assess performance based on interchangeability with pathologists' scoring of the reference data set. This would require evaluating the level of agreement among multiple pathologists' scores on the WSIs comprising the reference data set. Following the determination of the agreement among pathologists, the level of agreement between the model and all the pathologists can be assessed to determine if the model performs within the distribution of pathologists' performance. However, this approach is only applicable when pathologist scoring is possible and does not support the development or validation of novel biomarkers or more quantitative scoring approaches that are independent of pathologists.

Considerations for Annotation of Region of Interest

Pathologists inherently have a different workflow for assessing a WSI compared to an AI-derived model, including their understanding of the overarching morphology depicted in the slide. The Digital PATH project anecdotally found that for many WSIs with discordance in HER2 scores across models, scoring variability stemmed from differences in how the models identified the area of invasive carcinoma. To promote alignment in biomarker assessment, a reference dataset could include a consensus-based reference standard for the identified invasive tumor area on a WSI, derived from a consensus of pathologist annotations. Performance could be assessed based on a model's ability to identify the area of invasive carcinoma, providing additional insight into its performance. Guidance is needed to understand how to set targets for performance specific to the task of tumor area identification. It is also important to note that similarity in identifying regions of interest may or may not support clinical validation of a computational pathology model, especially for models that identify or integrate signals not visually discernible or typically analyzed by pathologists.

Table 2. Strengths and Limitations of Reference Standards for Digital Pathology–Based Biomarkers.

Reference Standard	Considerations	Strengths	Limitations
Consensus Pathologists' Scores	• Assess the inter-rater variability of multiple raters to give context to variability observed between the model and reference standard • Capture the recruitment methods, applicable qualifications and requirements (training, board certification, specialization, etc.), inclusion/exclusion criteria for pathologists	• Established scoring guidelines and proficiency training for the biomarker (e.g., ASCO/CAP HER2) provide consistency in assessment • Current practice for ascertaining biomarker status	• Guidelines may become outdated or irrelevant to future use cases, limiting the utility of the reference data set (e.g., HER2-low/ultra-low designations) or requiring augmentation with new information • It is challenging to recruit and train experts, and the make-up of the pathologists included can impact the consensus derived and the assessment of inter-rater variability
Clinical Treatment Outcomes	• The reference standard for clinical validation will depend on clinical outcomes • Full assessment of the predictive ability of a model requires data from both biomarker-positive and biomarker-negative patients, with some in each group receiving biomarker-directed therapy versus non-biomarker-directed	• Variability in scoring might not always translate to major differences in predicted patient outcomes and variability should be viewed in the context of impact of performance on outcome prediction • Likely needed for novel biomarkers without standardized guidelines for assessment	• May be difficult to find biomarker-negative cases treated with biomarker-directed therapy when many trials use biomarker-based eligibility criteria; might only be possible to establish whether patients identified as biomarker-positive by a model benefit from a targeted therapy relative to an alternative • The reference standard will be narrowly applicable to a specific

Reference Standard	Considerations	Strengths	Limitations
			drug/mechanism of action that may not be relevant for other drugs that utilize the same biomarker, and the data set will also be biased towards any inclusion/exclusion criteria inherent to treatment selection
Other Biological Correlates	• Orthogonal assays to measure the biomarker, such as mRNA, in situ hybridization, or mass spectrometry, could provide an additional assessment of the biomarker	• More quantitative, objective measure of biomarker that is not reliant on human interpretation	• Requires additional tissue, slides to run analyses • As new technology is developed, would need additional biological material to run new assays for the reference set to stay relevant • Orthogonal methods themselves may not be standardized and introduce additional variability

Considerations for Reporting Metadata and Ensuring Representativeness in the Reference Set

Relevant clinical, sample, and patient data should be connected to the WSIs. Table 3 highlights relevant metadata to capture. A data dictionary should accompany the reference data set, including metadata definitions for demographic and clinical information, as well as how the data were identified/defined (e.g., chart review, central testing for biomarkers, consensus or single scoring for histological grade, etc.). Additionally, the data dictionary should also specify the expected format for each data field.

While de-identified data is likely to be used, patients should be properly consented for use of their samples in a publicly available reference data set. For example, certain institutions consider digital pathology images to be biospecimens, which may require additional patient consent for inclusion in a repository.

Table 3. Metadata to Include in a Reference Data Set.

Patient Characteristics	Clinical Characteristics	Tissue and Slide Processing Characteristics*	Image Acquisition Characteristics*
Age (at sample collection)	Diagnosis History (e.g., de novo or recurrence)	Glass Slide Type	Scanner Hardware and Software Versions
Sex	Histological Grade	Tissue Thickness	Scanner Software Configurable Parameters
Race	Histology	Tissue Area; Tumor area/size	Slide Viewer
Ethnicity	Clinical Stage	Tissue Artifacts	Image File Type
Geographic Location	Biomarker Status (relevant to disease of interest)	Tissue Age	Magnification
Relevant prognostic factors	Prior Treatments Received	Slide Age	Resolution

	Sample Type (e.g., core biopsy, FNA, cytology)	Antibody Used (Lot #)	
	Tumor Site/anatomic location	Staining Conditions/method of antigen retrieval	
	Associated molecular findings	Slide Storage	

*For more detail, see
Supporting the Application of Computational Pathology in Oncology.pdf

As relevant to the intended use of the reference data set, these characteristics should vary to represent the entire diagnostic spectrum of a diverse target population. The sampling strategy should be detailed in accompanying literature to the reference data set (e.g., data set includes all cases from one site within a specific time frame), as well as any relevant inclusion or exclusion criteria that impact the samples. Generally, reference data sets should include samples from multiple clinical sites to ensure diversity in patient populations and clinical practice, which allows for the potential to conduct subgroup analyses assessing the association of the model's performance with specific clinical, patient, or pre-analytical characteristics.

Several considerations are specific to ensuring representativeness in the reference data set of the biomarker. Within a biomarker category, a reference set should include a spectrum of staining positivity. For example, HER2 staining levels could include weak-to-moderate complete membrane staining (e.g., 11% or 50% of tumor cells) as well as intense membrane staining (e.g., 1% or 9% of cells). Therefore, data sets should not only consider representation across each broad biomarker category, but also within each category.

Characteristic categories (e.g., patient, clinical, etc.), including the biomarker category, may be representative of the intended use population in the reference data set as a general principle. However, if there are no indications of an association between a specific characteristic and the outcome, and no subgroup analyses are planned, strict balancing may not be necessary. This approach allows for flexibility while ensuring that the reference data set reflects the intended purpose and avoids unnecessary complexity. If risk profiles are different, subgroups should be sized for individual subgroup analyses for a more definitive understanding of performance. There is a concern about oversampling (or undersampling) as some agreement measures, such as kappa coefficients, are highly dependent on the distribution of scores in patient subgroups. It is important to understand the distribution of characteristic categories in the clinical population to ensure that subgroups can be weighted appropriately when estimating overall performance results.

Lastly, it is important to consider the sample quality included and represented in the reference data set, including artifacts, inadequate tumor cellularity, and edge cases to ensure robust and meaningful validation. Whether and how many lower-quality samples are included in the reference data set will depend on its intended use. For example, one may prioritize inclusion of more "pristine" samples to establish baseline concordance of the model, but "edge" or challenging cases should likely be included to assess robustness of the models in a data set more reflective of clinical practice. Challenging cases may include those with artifacts, complex architecture or morphology, rare or mixed histologies, etc. In biomarkers studied extensively, such as HER2 in breast cancer, defining challenging cases may be easier than in other biomarker contexts. Proactively identifying challenging cases may be difficult but could be informed by conducting interviews with pathologists to understand difficult cases.

Conclusions and Next Steps

This document provides an overview of discussions aimed at catalyzing further dialogue in the field on developing robust reference data sets. Reference data sets and models can have many intended uses, which require careful consideration during the development process. When creating a reference data set, it is important to narrow focus to a single intended use. For example, analytical validation for regulatory purposes could serve as a use case to propose specific criteria for developing a robust reference data set.

Developing a reference data set will require collaboration across multiple contributors and may emerge from community efforts, patient groups, federal initiatives, or professional societies. Recent opportunities, such as the Advanced Research Projects Agency for Health (ARPA-H) ImagiNg Data EXchange (INDEX) program, provide possible platforms to develop such reference data sets. Those interested in developing a reference data set for regulatory purposes should consult the FDA early in the planning process. An ideal opportunity for interaction with the FDA is through development of a medical device development tool (MDDT).[14] Additionally, the FDA provides various other pathways for engagement, depending on the intended use of the AI model or reference data set.[15] These options, outlined in a recent guidance document, include opportunities to discuss innovative trial designs, digital health technologies, and real-world evidence generation, among others. A voluntary pre-submission with the FDA would allow for early discussions on the scope, protocol, statistical approach, and patient population for the data set.

References

1. McKelvey B, Torres-Saavedra P, Li J, et al: Agreement Across 10 Artificial Intelligence Models in Assessing HER2 in Breast Cancer Whole Slide Images: Findings from the Friends of Cancer Research Digital PATH Project. Available from: https://friendsofcancerresearch.org/wp-content/uploads/SABCS-Poster_Digital-PATH-Project-Findings.pdf

2. September 26, 2024: Meeting of the Oncologic Drugs Advisory Committee Meeting Announcement - 09/26/2024 | FDA [cited 2025 Jan 15] Available from: https://www.fda.gov/advisory-committees/advisory-committee-calendar/september-26-2024-meeting-oncologic-drugs-advisory-committee-meeting-announcement-09262024

3. Tsao MS, Kerr KM, Kockx M, et al: PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples: Results of Blueprint Phase 2 Project. J Thorac Oncol 13:1302, 2018. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8386299/

4. Feng Z, Kagan J, Pepe M, et al: The Early Detection Research Network's Specimen Reference Sets: Paving the Way for Rapid Evaluation of Potential Biomarkers. Clin Chem 59:68–74, 2013. Available from: https://dx.doi.org/10.1373/clinchem.2012.185140

5. Khleif SN, Doroshow JH, Hait WN: AACR-FDA-NCI cancer biomarkers collaborative consensus report: Advancing the use of biomarkers in cancer drug development. Clinical Cancer Research 16:3299–3318, 2010. Available from: /clincancerres/article/16/13/3299/11147/AACR-FDA-NCI-Cancer-Biomarkers-Collaborative

6. Cancer Biomarker Measurements and Collaborations | NIST [cited 2025 Jan 15] Available from: https://www.nist.gov/programs-projects/cancer-biomarker-measurements-and-collaborations

7. Petrick N, Chen W, Delfino JG, et al: Regulatory considerations for medical imaging AI/ML devices in the United States: concepts and challenges. Journal of Medical Imaging 10:051804, 2023. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10289177/

8. Abels E, Pantanowitz L, Aeffner F, et al: Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. J Pathol 249:286, 2019. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6852275/

9. November 20-21, 2024: Digital Health Advisory Committee Meeting Announcement - 11/20/2024 | FDA [cited 2025 Jan 15] Available from: https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-20-21-2024-digital-health-advisory-committee-meeting-announcement-11202024

10. GitHub - maduc7/Histopathology-Datasets: Ressources of histopathology datasets [cited 2025 Jan 15] Available from: https://github.com/maduc7/Histopathology-Datasets?tab=readme-ov-file

11. Yee LM, Lively TG, McShane LM: Biomarkers in early-phase trials: fundamental issues. Bioanalysis 10:933, 2018. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6123886/

12. Janes H, Pepe MS, McShane LM, et al: The Fundamental Difficulty With Evaluating the Accuracy of Biomarkers for Guiding Treatment. J Natl Cancer Inst 107, 2015. Available from: https://pubmed.ncbi.nlm.nih.gov/26109106/

13. Polley MYC, Freidlin B, Korn EL, et al: Statistical and practical considerations for clinical evaluation of predictive biomarkers. J Natl Cancer Inst 105:1677–1683, 2013. Available from: https://pubmed.ncbi.nlm.nih.gov/24136891/

14. Medical Device Development Tools (MDDT) | FDA. Available from: https://www.fda.gov/medical-devices/medical-device-development-tools-mddt

15. Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products | FDA [cited 2025 Jan 15] Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological

Establishing Early Endpoints for Drug Development

ctDNA Clearance as an Early Indicator of Improved Clinical Outcomes in Advanced NSCLC Treated with TKI: Findings from an Aggregate Analysis of Eight Clinical Trials

Hillary S. Andrews[1], Nevine Zariffa[2], Katherine K. Nishimura[3], Stephanie H. Choi[4], Shibing Deng[5], Megan Eisele[3], Carin R. Espenschied[6], Emily M. Goren[3], Minakshi Guha[7], Samuel Hong[8], Dilafruz Juraeva[9], Nicole Krämer[10], Li Liu[11], Jean-Francois Martini[5], Brittany A. McKelvey[1], Geoffrey R. Oxnard[12], Gary A. Pestano[13], Lynne Poole[14], Adam Rosenthal[3], Anna M. Szpurka[12], Diana Merino Vega[15], Christine Ward[7], Sameera R. Wijayawardana[16], Antje Hoering[3], Mark D. Stewart[1], and Jeff D. Allen[1]

ABSTRACT

Purpose: Circulating tumor DNA (ctDNA) holds promise as an early endpoint to predict overall survival (OS). The creation and structured interrogation of aggregated datasets inform the hypothesis that ctDNA is reasonably likely to predict treatment benefit. Friends of Cancer Research convened a diverse working group to establish and implement an analysis plan assessing patient-level associations between changes in ctDNA levels with OS and progression-free survival (PFS).

Experimental Design: The aggregate dataset included eight clinical trials representing 940 patients with biomarker-positive advanced non–small cell lung cancer treated with tyrosine kinase inhibitors. Detection of baseline and on-treatment ctDNA was assessed for associations with OS and PFS. Additionally, combinations of ctDNA detection and RECIST measurements up to 10 weeks on treatment were considered.

Results: Patients with detected ctDNA at baseline that became nondetected on treatment ("clearance") experienced improved OS compared with patients with persistently detected ctDNA (adjusted HR = 2.12, $P < 0.001$). This pattern was also seen in the subset of patients with stable disease as measured by RECIST within 10 weeks of treatment initiation (adjusted HR = 4.15, $P < 0.001$). Results were similar for PFS.

Conclusions: In patients with oncogene-driven advanced non–small cell lung cancer treated with tyrosine kinase inhibitors, ctDNA clearance within 10 weeks of treatment initiation was associated with improved OS and PFS. These patient-level results support the growing evidence that demonstrates a change in ctDNA levels during treatment is associated with clinical benefit. Future prospective trials should include predefined thresholds of molecular response to advance the utility of ctDNA as an early endpoint.

Introduction

Clinical trial endpoints convey information about the safety and efficacy of therapies, with overall survival (OS) as the gold standard for evaluating therapeutic efficacy in oncology drug development (1). The rapidly evolving therapeutic landscape in oncology has led to much-awaited improvements in survival (2), which benefit patients but also create a need for new, earlier endpoints to support the development of emerging therapies. To reliably accelerate the assessment of new therapies, early endpoints must be associated with clinical benefit and provide accurate insights into treatment response. The U.S. Food and Drug Administration (FDA) may approve new therapies based on an early endpoint that is reasonably likely to predict clinical benefit through the Accelerated Approval pathway, which provides patients with serious conditions timely access to novel therapies that fulfill an unmet medical need (3). Early endpoints may also inform trial sponsors' internal decisions about proceeding with clinical development programs.

Currently, early clinical endpoints for solid tumors focus on radiographic response as assessed by RECIST version 1.1 to determine the best response or progression-free survival (PFS; ref. 4). These endpoints measure the effect on the tumor attributable to the drug and are generally objective and quantitative (5, 6). However, assessing radiographic response using RECIST has limitations, such as the need for central independent review, the need for repeat measures that require exposure to radiation (albeit low levels), challenges with assessing nonmeasurable disease, and limited insights into molecular responses, particularly for targeted therapies or immunotherapies (6, 7). There is an opportunity to identify and validate endpoints that may aid in evaluating treatment efficacy and provide an earlier, easier, and more comprehensive assessment of response to treatment (8).

[1]Friends of Cancer Research, Washington, District of Columbia. [2]NMD Group Inc., Bala Cynwyd, Pennsylvania. [3]Cancer Research and Biostatistics, Seattle, Washington. [4]EMD Serono, Washington, District of Columbia. [5]Oncology R&D, Pfizer, Inc., San Diego, California. [6]Guardant Health, Inc., Palo Alto, California. [7]Takeda Pharmaceuticals Inc., Cambridge, Massachusetts. [8]Agilent Technologies, Inc., Santa Clara, California. [9]The Healthcare Business of Merck KGaA, Darmstadt, Germany. [10]Boehringer Ingelheim Pharma GmbH & Co KG., Biberach, Germany. [11]Illumina, Inc., San Diego, California. [12]Loxo@Lilly, Indianapolis, Indiana. [13]Biodesix Inc., Louisville, Colorado. [14]AstraZeneca, Cambridge, United Kingdom. [15]AstraZeneca, Gaithersburg, Maryland. [16]Eli Lilly and Company, Indianapolis, Indiana.

Corresponding Author: Hillary S. Andrews, Science Policy, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC 20036. E-mail: handrews@focr.org

Clin Cancer Res 2025;31:2162–72

doi: 10.1158/1078-0432.CCR-24-3612

Translational Relevance

ctDNA holds promise as an early endpoint in solid tumor oncology trials; however, patient- and trial-level meta-analyses demonstrating associations between changes in ctDNA and long-term outcomes, such as overall survival and progression-free survival, are necessary before ctDNA can be used in regulatory decision-making. A patient-level aggregate analysis of eight previously completed clinical trials of patients with advanced non–small cell lung cancer treated with tyrosine kinase inhibitors showed that clearance of ctDNA was associated with improved overall survival and progression-free survival compared with those who had persistently detected ctDNA. Additional work is necessary to establish ctDNA as an early endpoint, but this study supports the growing body of evidence that decreased ctDNA levels are associated with improved outcomes.

Circulating tumor DNA (ctDNA) has emerged as a novel biomarker with the potential to revolutionize cancer care and accelerate the approval of new cancer medicines (9). An FDA draft guidance provides important considerations for the use of ctDNA in oncology drug development, including its potential as an early endpoint in clinical trials (10). ctDNA can be assessed through a blood draw, potentially allowing for earlier measurement of molecular response that is less invasive for patients and more frequent than radiographic response assessments.

A key step in using ctDNA change from baseline (i.e., pretreatment) to on-treatment as an early endpoint is demonstrating associations between ctDNA change and clinical outcomes, such as OS and PFS, at the trial-level and in aggregate patient-level analyses. In patient-level analyses, associations between ctDNA change and long-term clinical outcomes have been observed in a variety of cancers for certain treatments in clinical trials (11–13). However, the generalizability of these results across trials, cancer types, assays that detect or quantify ctDNA, and treatment modalities has not been comprehensively evaluated in aggregate patient-level datasets.

The Friends of Cancer Research (*Friends*) ctDNA for Monitoring Treatment Response (ctMoniTR) project is a unique collaboration to assess whether changes in ctDNA levels are associated with OS and PFS. In a first report, data from five previously completed clinical trials of patients with advanced non–small cell lung cancer (aNSCLC) treated with immune checkpoint inhibitors suggested that a greater than 50% decrease in ctDNA levels from baseline was associated with improved OS and PFS (14). This second report evaluated whether changes in ctDNA levels were associated with OS and PFS in a patient population with oncogene-driven aNSCLC treated with a tyrosine kinase inhibitor (TKI). Additional objectives assessed the relationship between ctDNA changes and early RECIST measurements with long-term outcomes. Collectively, the study aimed to provide insights into the potential role of ctDNA changes as an early endpoint for response to TKIs in oncogene-driven aNSCLC.

Materials and Methods

ctMoniTR project approach

Friends is a nonprofit advocacy organization focused on leveraging groundbreaking collaborations, generating scientific evidence, and integrating patient input to shape public policy. *Friends* coordinated a working group with representatives from pharmaceutical companies, ctDNA test developers, academia, and the FDA to discuss the analysis plan, react to emerging results, and align on the interpretation. These prespecified, retrospective analyses are considered exploratory and hypothesis-generating and are not definitive. Cancer Research And Biostatistics (CRAB) served as the data aggregator, independent statistical analysis team, and performed analyses. There were frequent meetings of an expert statistical group with representatives from the participating organizations to discuss the statistical approach and findings. The entire group also met regularly to review and interpret the findings.

Clinical trials

Clinical trial sponsors provided CRAB patient-level data from eight unique clinical trials of biomarker-positive aNSCLC treated with TKIs that were previously completed (Supplementary Table S1). The original trials were conducted with appropriate ethical oversight, including Institutional Review Board approval in accordance with the Declaration of Helsinki and written informed consent from all participants, which allowed for secondary research use of deidentified data. Each sponsor reviewed the informed consent forms to confirm the suitability of the data for secondary use. Patient-level clinical and ctDNA data were cleaned, formatted, and anonymized by the sponsors prior to submission to CRAB. Sponsors approved the final representation of their data for quality assurance purposes.

Patient inclusion criteria

A cohort was defined as the patients from a treatment arm in a randomized controlled trial or as the patients in a single-arm trial. Only patients treated with TKIs were included in the analyses (i.e., patients treated in the chemotherapy arm were not included). The working group aligned on the definition of the index date as the date of randomization for randomized controlled trials and the date of treatment initiation for single-arm trials as well as the inclusion criteria: Patients with biomarker-positive aNSCLC treated with a corresponding TKI (i.e., anti-EGFR, anti-ALK, anti-MET, or anti-RET) had assessments by RECIST version 1.1 (15), survival endpoints (OS and/or PFS), a baseline ctDNA sample (up to 14 days before index; T0), and at least one on-treatment ctDNA sample. Further inclusion criteria included refining the T1 time window to 10 weeks, the clinical covariates to be included in multivariable analyses, the outcomes of interest (i.e., OS or PFS), and minimum cohort size requirements based on the training/test approach (see Training and test approach section).

Clinical endpoints and covariates

OS and PFS were the endpoints of interest. Additional clinical covariates were standardized to a common definition across cohorts: age (<65 vs. ≥65), sex (female vs. male), race (White vs. other), smoking status (ever smoked vs. never smoked), stage (advanced stage vs. other), Eastern Cooperative Oncology Group performance status (≥1 vs. 0; ref. 16), number of prior lines of therapy (≥1 vs. 0), and histology (adenocarcinoma vs. other).

ctDNA data

Sponsors provided variant allele frequency (VAF) values for each ctDNA variant detected in plasma (not sera) cell-free DNA samples collected from clinical trial participants. The specific gene (4–523) variants assessed for each patient depended on the assay used in

each trial, which included commercially available next-generation sequencing (NGS) or Droplet Digital PCR (ddPCR) assays (**Table 1**). When relevant, assay developers established protocols to filter out nontumor-related variants, such as clonal hematopoiesis of indeterminate potential variants. For each patient at each time point, VAF values for all filtered variants were included in the derived ctDNA metrics, and the maximum VAF was calculated as previously specified (i.e., the highest VAF value within each ctDNA sample) where applicable (14). If no ctDNA was detected by the ctDNA assay used in the trial, the ctDNA measurement was reported as nondetected (ND). The limit of detection was defined by the assay used in each individual clinical trial and varied by assay (**Table 1**).

Derived ctDNA metrics

The primary ctDNA metric in this analysis was based on the change in ctDNA levels from baseline (T0) to up to 10 weeks after the index (T1). If measurements from multiple on-treatment ctDNA samples were available within 10 weeks for an individual patient, the ctDNA sample with the lowest VAF or ND was used. The percentage change in VAF was calculated as a ratio to baseline.

Training and test approach

As this was a retrospective analysis, statistical validation was predefined by the training/test strategy. The data were randomly split into 2/3 training and 1/3 test datasets using the SAS SURVEYSELECT procedure balancing on cohort, age, and tumor stage. To ensure an adequate sample size for multivariable modeling stratified by cohort, any cohort with fewer than 20 patients was dropped as it could not contribute meaningfully to multivariable models in the test dataset.

The training dataset was used to assess heterogeneity between the cohorts, review descriptive statistics to identify harmonization strategies and create derived variables, conduct multivariable association tests, and investigate different ctDNA metrics. Once the ctDNA variables were optimized in the training dataset, their performance was examined in the test dataset to confirm that overfitting was not a factor. To confirm the consistency of identified associations in the test data, Kaplan–Meier curves of the training and test data were plotted and assessed for consistency of ctDNA category ordering. C-statistics of the multivariable model in the training data were compared with the C-statistics based on the model's predictions for the test data. Once the group reviewed these comparisons, each set of analyses was aligned in both sets of data. As such, the main figures presented throughout represent all data, whereas the results using the separate training/test datasets are available in the supplementary figures.

Statistical analyses

A prespecified statistical plan outlined the key research hypotheses, validation strategies, and analytical approaches. Three research objectives were prospectively defined: (i) to evaluate associations between early ctDNA measurements and OS and PFS, (ii) to evaluate the correlation between ctDNA and the patient's response at the first RECIST measurement (early RECIST response), and (iii) to evaluate the additive value of using both ctDNA and early RECIST response (i.e., at first assessment) when assessing associations with OS and PFS (**Fig. 1**). The statistical analysis plan was finalized before any analyses that included PFS or OS except for research objective 1, in which the results of a first analysis informed the final analysis.

The primary models for interpretation were multivariable Cox proportional hazard models to estimate the association between ctDNA change and survival outcomes, with stratification by cohort to account for heterogeneity and landmarking to avoid immortal time bias (17, 18). Stratification served to account for the expected heterogeneity between cohorts, given that each cohort received different treatments and was derived from a different patient population according to a trial's recruitment strategy. All multivariable models were also adjusted for baseline demographic and clinical covariates. Cochran–Mantel–Haenszel tests stratified by cohort were used to compare proportions of categorical variables (19, 20). Survival probabilities (OS, PFS) were assessed using Kaplan–Meier plots.

All survival models used a 70-day (10-week) landmark, given that the on-treatment ctDNA sample was required to be no later than 10 weeks from the index, and thus, the change in ctDNA levels from T0 could not be determined until 10 weeks. Patients with an event (i.e., death for OS, death or progression for PFS) before 10 weeks were removed from the landmarked models (21). Landmarked OS and PFS were used to address the immortal time bias that can arise when a predictor (i.e., ctDNA) can only be obtained among patients who survive long enough to have the predictor measured. Compared with alternative methods, such as Cox models with time-varying covariates, landmarking produces easily interpretable associations between ctDNA and OS and PFS that have been corrected for this bias. The variance explained in nested versus full models was compared using likelihood ratio tests and the Akaike information criterion (22, 23). Concordance statistics, including Harrell's C-index and Uno's C-statistic, were used to assess the fit of models developed in the training dataset when applied to test datasets (24, 25). Likelihood ratio tests were also used to compare associations with ctDNA levels at baseline and outcomes to changes in ctDNA levels.

All statistical tests were two-sided, with a nominal P value <0.05 used as an indication of statistical significance for simplicity. There was no adjustment for multiplicity, and the results are exploratory or hypothesis-generating. Analyses were done using the SAS statistical software package (SAS Institute, RRID:SCR_008567) or R (R Foundation for Statistical Computing).

Data availability

Data are not publicly available due to data use agreements with pharmaceutical companies that provided data for the purposes of conducting the study and privacy and ethical considerations for patients. Upon request and subject to review by all authors, data that support the findings of the study may be provided, including code for running statistical analyses.

Results

Analysis dataset

Eight clinical trials representing 1,590 patients with biomarker-positive aNSCLC (for the trial's oncogene of interest) who were treated with a TKI were considered for inclusion (**Fig. 1**; Supplementary Table S1). The working group defined the T1 sample window as 10 weeks following the index, allowing for the maximal number of patients to be included ($n = 961$ at 10 weeks vs. $n = 774$ at 6 weeks) and to align with the previous ctMoniTR analysis, which also used a 10-week window (14). Exclusion criteria led to 940 patients for analysis (**Fig. 1**). Patient and assay characteristics are summarized in **Table 1**.

Table 1. Patient demographics and assay characteristics.

Trait	Description	Cohort												
		1	2	3	4	5	6	7	8	9	10	11	*P*	Overall, *n/N* (%)
		n = 179	*n* = 179	*n* = 170	*n* = 99	*n* = 92	*n* = 84	*n* = 68	*n* = 24	*n* = 23	*n* = 17	*n* = 5		*n* = 940
Age	Age ≥65 years at enrollment, *N* (%)	85 (47%)	84 (47%)	76 (45%)	37 (37%)	23 (25%)	18 (21%)	22 (32%)	6 (25%)	9 (39%)	3 (18%)	4 (80%)	<0.001	367 (39%)
Sex	Female, *N* (%)	111 (62%)	108 (60%)	107 (63%)	55 (56%)	59 (64%)	44 (52%)	38 (56%)	12 (50%)	19 (83%)	7 (41%)	1 (20%)	0.085	561 (60%)
Race	White, *N* (%)	54 (30%)	49 (27%)	55 (32%)	46 (46%)	46 (50%)	36 (43%)	35 (51%)	19 (79%)	9 (39%)	7 (41%)	2 (40%)	<0.001	358 (38%)
Smoking status	Ever smoked, *N* (%)	64 (36%)	71 (40%)	59 (35%)	46 (46%)	32 (35%)	32 (38%)	28 (41%)	8 (33%)	5 (22%)	8 (47%)	3 (60%)	0.484	356 (38%)
ECOG	ECOG performance status ≥1, *N* (%)	101 (56%)	112 (63%)	102 (60%)	50 (51%)	55 (60%)	55 (65%)	42 (62%)	18 (75%)	20 (87%)	8 (47%)	1 (20%)	0.029	564 (60%)
Stage[a]	Advanced stage (stage IV), *N* (%)	164 (92%)	165 (92%)	138 (81%)	98 (99%)	90 (98%)	83 (99%)	66 (97%)	24 (100%)	23 (100%)	17 (100%)	5 (100%)	<0.001	873 (93%)
Prior therapy	Prior lines of systemic therapy ≥1, *N* (%)	21 (12%)	16 (9%)	170 (100%)	11 (11%)	5 (5%)	27 (32%)	23 (34%)	22 (92%)	23 (100%)	17 (100%)	3 (60%)	<0.001	338 (36%)
Histology	Adenocarcinoma, *N* (%)	176 (98%)	176 (98%)	168 (99%)	94 (95%)	88 (96%)	76 (90%)	67 (99%)	23 (96%)	22 (96%)	17 (100%)	4 (80%)	0.008	911 (97%)
ctDNA samples	Median no. of ctDNA measurements (range)	5 (3–28)	5 (2–24)	6 (2–19)	3 (2–3)	3 (2–3)	3 (2–4)	3 (2–3)	4 (2–5)	2 (2–3)	2 (2–2)	4 (3–4)		
Assay		ddPCR	ddPCR	ddPCR	NGS	NGS	NGS	NGS	NGS	NGS	NGS	NGS		
LOD		2 copies	2 copies	2 copies	0.1%–0.3%	0.1%–0.3%	0.1%	0.1%	0.1%–0.3%	0.5%	0.1%	0.1%–0.3%		
Genes assessed	*N*	4	4	4	74	74	15	15	74	523	22	74		
ND	ND ctDNA at T0	48 (27%)	52 (29%)	47 (28%)	26 (26%)	53 (58%)	49 (58%)	42 (62%)	15 (63%)	0 (0%)	1 (6%)	0 (0%)	0.0036	257/1,015 (25.3%)
VAF ≤0.5[b]	ctDNA samples per cohort, *N* (%)	52 (29%)	47 (26%)	47 (28%)	26 (26%)	18 (20%)	26 (31%)	20 (29%)	1 (4%)	0 (0%)	3 (18%)	0 (0%)	0.020	240 (26%)

Note: *P* value from Fisher's exact test comparing cohorts.
Abbreviations: ECOG, Eastern Cooperative Oncology Group; ctDNA, circulating tumor DNA; ddPCR, Droplet Digital PCR; NGS, next-generation sequencing; LOD, limit of detection; ND, nondetected; VAF, variant allele frequency.
[a]Stage at enrollment for all except cohorts 1–3, which used stage at diagnosis.
[b]0.5 chosen because it was the max LOD across cohorts. A cohort is all patients from a treatment arm in a randomized controlled trial or all patients in a single-arm trial.

Change in ctDNA levels and associations with clinical outcomes

Initial analyses focused on the percentage change in ctDNA levels from T0 to T1. The working group predefined the percentage change categories as follows: "decrease" (>50% decrease), "increase" (>20% increase), and "intermediate" (all remaining patients) based on previous research and experience (14). After noting that the number of patients with ND ctDNA at both T0 and T1 in the OS analysis (n = 136/501), a fourth category of "never detected" was included. Patients with detected (D) ctDNA (any level) at T0 that became ND at T1 (i.e., clearance) were included in the "decrease" category (n = 260/501), and patients with ND ctDNA at T0 that became D at T1 were included in the "increase" category (n = 5/501). However, this "ctDNA percentage change" variable did not show consistent associations with outcomes (Supplementary Fig. S1A and S1B).

After noting that a substantial number of patients had ND ctDNA at T1 (n = 396/501), the working group created an analysis variable called "ctDNA detection" with four ctDNA detection categories: (i) "never detected" (ND/ND; ND at T0 and T1), (ii) "clearance" (D/ND; D at T0 and ND at T1), (iii) "emerging detection" (ND/D; ND at T0 and D at T1), and (iv) "persistent detection" (D/D; D at T0 and T1). The "emerging detection" category had few patients (n = 5/501) and was omitted from further analyses.

Never detected, clearance, and persistent detection were separated by rank ordering in the Kaplan–Meier curve and were confirmed as statistically distinct in multivariable Cox models (**Fig. 2A** and **B**). Specifically, patients with never-detected ctDNA had the best OS [adjusted HR (HR) = 2.95, $P < 0.001$ for D/ND vs. ND/ND and HR = 6.25, $P < 0.001$ for D/D vs. ND/ND] and PFS (HR = 2.11, $P < 0.001$ for D/ND vs. ND/ND and HR = 3.21, $P < 0.001$ for D/D vs. ND/ND). Among patients with D ctDNA before treatment, those with ctDNA clearance had improved OS and PFS compared with patients with persistent ctDNA detection (HR for OS = 2.12, $P < 0.001$; HR for PFS = 1.52, $P = 0.002$; **Fig. 2A** and **B**). Similar trends were observed in the training and test analyses, and C-statistics demonstrated appropriateness for combining data (Supplementary Fig. S2A–S2D).

Patients with baseline ctDNA detection (i.e., not considering T1 ctDNA levels) had a less favorable prognosis compared with those with ND baseline ctDNA (HR = 4.16, $P < 0.0001$ and

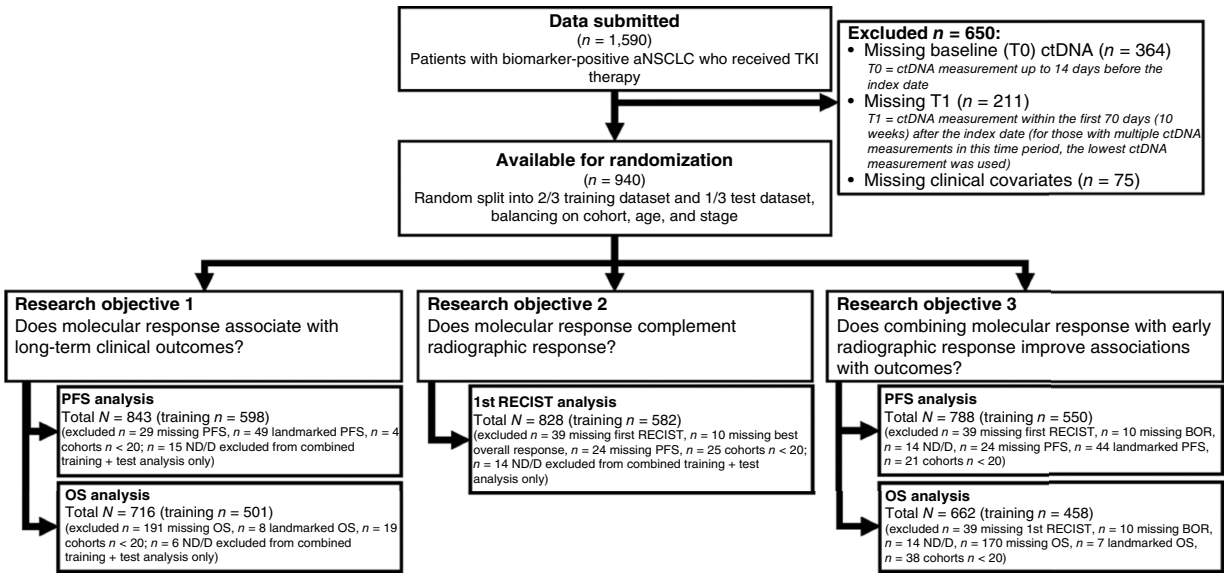


Figure 1.

CONSORT flow diagram. This figure presents the CONSORT diagram outlining patient inclusion and exclusion criteria for the analysis. The flowchart details the number of patients screened, those included in the training and test datasets, and exclusions due to missing data or other criteria. Abbreviations: aNSCLC, advanced non–small cell lung cancer; TKI, tyrosine kinase inhibitor; ctDNA, circulating tumor DNA; ND/D, nondetected/detected; PFS, progression-free survival; OS, overall survival; BOR, best overall response.

HR = 2.16, $P < 0.0001$, respectively; Supplementary Fig. S3A and S3B). However, among patients with D ctDNA at baseline, likelihood ratio tests confirmed that information from on-treatment ctDNA values further stratified patients into those with better (D/ND) or worse (D/D) OS ($P < 0.0001$) and PFS ($P = 0.0003$). Although patients with never-detected ctDNA have better outcomes, it cannot be determined whether this is simply prognostic or due to treatment. Subsequent analyses used ctDNA detection at T0 and T1 to define ctDNA categories.

Associations between ctDNA detection categories and an early RECIST measurement

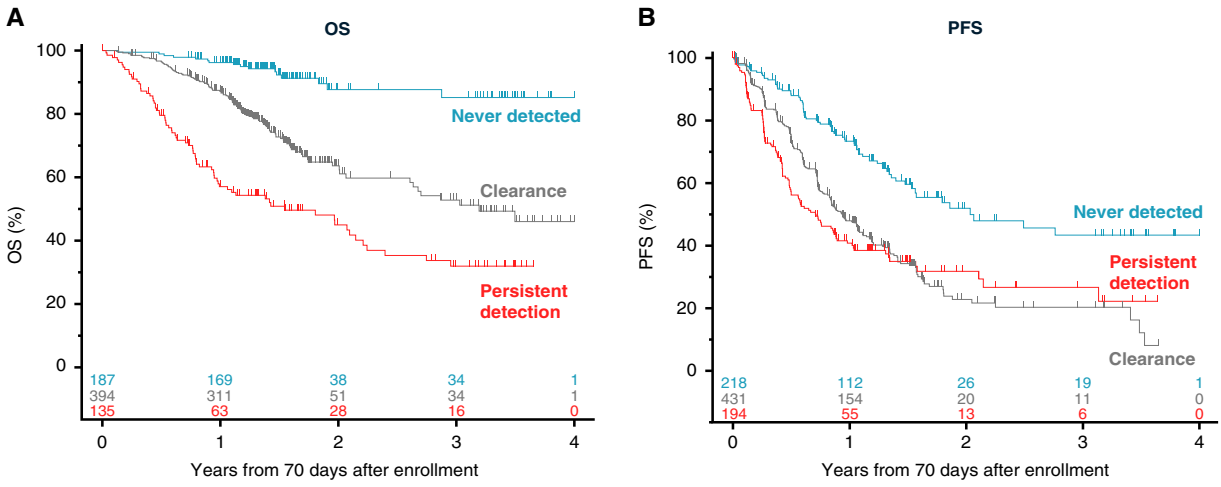
To evaluate the relationship between ctDNA detection categories and an early RECIST measurement, the working group focused on a RECIST assessment within the same time window as the T1 ctDNA (i.e., the first RECIST measure). First, the working group considered the distribution of radiographic response at only the first RECIST assessment across the ctDNA detection categories. For the early RECIST assessments, those with complete response ($n = 4$) and partial response (PR; $n = 536$) were combined into a "responder" category, and those with stable disease (SD; $n = 264$) or progressive disease ($n = 24$) were combined into a "nonresponder" category, given the small sample sizes. The goal was to determine if patients with "better" ctDNA categories (i.e., ND ctDNA on treatment) also had "better" radiographic response at first RECIST (i.e., responder). There were differences between the cohorts in the distribution of early responders and nonresponders ($P = 0.0132$); however, visually, there seemed to be a similar distribution of first RECIST response within each ctDNA category when considering all cohorts (Supplementary Fig. S4).

In contrast to the ctDNA detection category analysis, there was not an apparent association between first RECIST and OS or PFS (HR = 1.21, $P = 0.254$ and HR = 0.96, $P = 0.728$, respectively; **Fig. 3A** and **B**). Similar trends were observed in the training and test analyses assessing early RECIST response categories, and C-statistics demonstrated appropriateness for combining data (Supplementary Fig. S5A–S5D). Whether this early RECIST response was additive to the ctDNA detection categories to improve associations with long-term outcomes was then assessed. For patients with ctDNA clearance or never-detected ctDNA, the early RECIST response categorization comparing responders with nonresponders within each ctDNA category did not provide additional prognostic information for association with OS (HR = 0.95, $P = 0.917$ for ND/ND and HR = 0.91, $P = 0.675$ for D/ND; **Fig. 4A**), which was similar for the associations with PFS (HR = 0.80, $P = 0.373$ for ND/ND and HR = 0.94, $P = 0.670$ for D/ND; **Fig. 4B**). Similar trends were observed in the training and test analyses assessing early RECIST response categories, and C-statistics demonstrated appropriateness for combining data (Supplementary Fig. S6A–S6D). However, for patients with persistent ctDNA detection, nonresponders at first RECIST had an increased risk of death compared with responders at first RECIST (HR = 1.93, $P = 0.037$), a pattern not seen for progression (HR = 1.29, $P = 0.296$; **Fig. 4A** and **B**).

A likelihood ratio test for association demonstrated that radiographic response at an early RECIST assessment did not provide additional value in characterizing the association with OS beyond the contribution of the ctDNA detection categories ($P = 0.195$).

ctDNA detection categories within SD or PR at the early RECIST measurement

Finally, to determine whether the ctDNA detection categories added information about associations with outcomes to specific first RECIST classifications, the ctDNA detection categories were



	Deaths/N	Median (Years)	1-Year estimate
Never detected	16/187	NR	96% (93–99)
Clearance	114/394	3.2 (2.6–NR)	87% (84–91)
Persistent detection	72/135	1.5 (1–2.2)	57% (48–66)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	2.95 (1.58–5.48), <0.001
Persistent detection vs. never detected	6.25 (3.10–12.62), <0.001
Persistent detection vs. clearance	2.12 (1.50–3.01), <0.001

	Events/N	Median (Years)	1-Year estimate
Never detected	77/218	2 (1.6–NR)	73% (67–80)
Clearance	253/431	0.9 (0.8–1.1)	48% (43–53)
Persistent detection	114/194	0.7 (0.5–0.9)	41% (33–48)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	2.11 (1.54–2.91), <0.001
Persistent detection vs. never detected	3.21 (2.17–4.76), <0.001
Persistent detection vs. clearance	1.52 (1.17–1.98), 0.002

Figure 2.

Kaplan–Meier curves for OS and PFS by ctDNA detection categories. Kaplan–Meier curves for (**A**) OS and (**B**) PFS and ctDNA detection categories, landmarked at 70 days from the index; patients with an event during the 70-day landmark were excluded from the analysis. Data presented include the training and test datasets. Abbreviations: OS, overall survival; PFS, progression-free survival; HR, adjusted hazard ratio; CI, confidence interval.

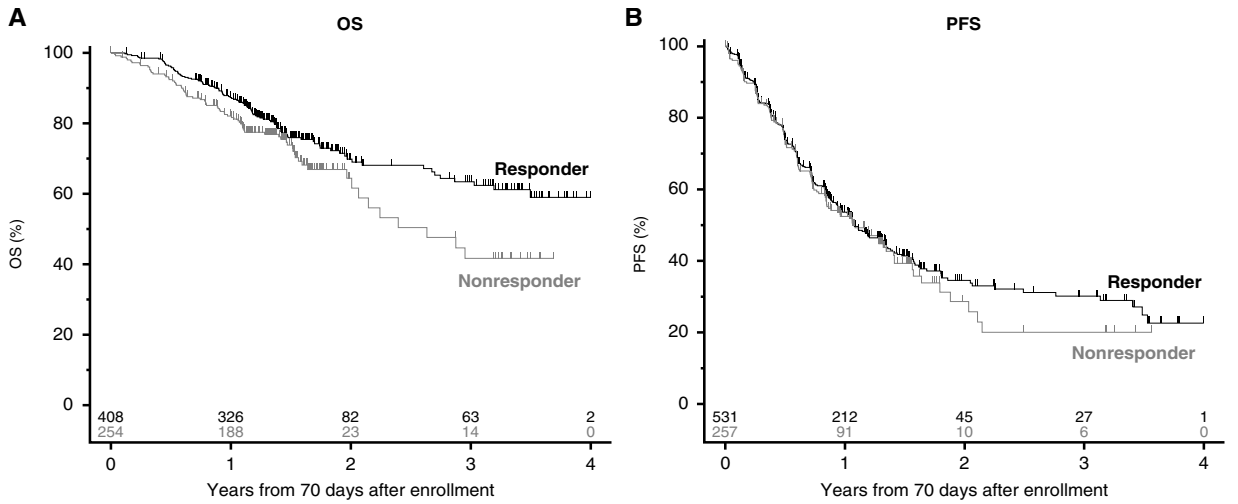
examined among patients with SD or PR. For patients classified as SD at first RECIST, the three ctDNA detection categories had differential associations with OS and PFS. Patients with persistent detection of ctDNA had an increased risk of death (HR = 4.15, $P < 0.001$) compared with those with ctDNA clearance (**Fig. 5A** and **B**). Among patients with PR at first RECIST, the predictive insight from ctDNA detection categories was less consistent yet trended toward significance across most analyses (**Fig. 5C** and **D**).

Discussion

In anonymized patient-level data from eight clinical trials, ctDNA clearance was associated with improved OS and PFS based on multivariable models that accounted for other prognostic factors. This large dataset included 940 patients with biomarker-positive aNSCLC treated with TKIs and provided clear and unifying results across a wide range of measures. The association between ctDNA detection categories and outcomes seemed largely independent of early radiographic response.

The findings build on the body of evidence from single trials of patients with *EGFR*-mutated cancer receiving anti-EGFR treatment, in which ctDNA clearance was associated with improved OS (26–28). Similarly, ctDNA clearance on anti-ALK treatment in patients with *ALK*-rearranged aNSCLC was associated with improved PFS (29). In the first report of ctMoniTR, in patients with aNSCLC treated with immune checkpoint inhibitors, reductions in ctDNA were associated with clinical outcomes including OS and PFS (14). Although the definition of a reduction in ctDNA differed (i.e., percentage change vs. detection), the previous and current aggregate analyses support the growing body of evidence demonstrating the relationship between reduced ctDNA levels and improved OS. They also highlight that different treatment modalities may require different approaches to defining clinically meaningful changes in ctDNA levels. In the future, the field would benefit from the establishment of standardized thresholds for molecular response, as feasible.

Including an early measure of radiographic response with ctDNA detection categories did not improve associations with outcomes compared with models using ctDNA detection categories alone. Additionally, this analysis and previous analyses demonstrated no association between early RECIST measures and OS (30). Objective response rate (ORR) is high for patients with aNSCLC treated with targeted TKIs in the literature (~70%) as well as across cohorts included in this study, which may make it challenging to decipher associations with outcomes because so many patients have a response (31–34). This analysis focused on an early RECIST measure



	Deaths/N	Median (Years)	1-Year estimate
Responder	103/408	NR	87% (84–91)
Nonresponder	75/254	2.6 (2.1–NR)	82% (77–87)

Multivariable associations

Comparison	HR (95% CI), *P* value
Responder vs. Nonresponder	1.21 (0.87–1.67), 0.254

	Events/N	Median (Years)	1-Year estimate
Responder	281/531	1.1 (1.0–1.3)	54% (49–58)
Nonresponder	131/257	1.1 (0.8–1.4)	52% (46–59)

Multivariable associations

Comparison	HR (95% CI), *P* value
Responder vs. Nonresponder	0.96 (0.77–1.20), 0.728

Figure 3.

Kaplan–Meier curves for OS and PFS by radiographic response at the first RECIST measurement. Kaplan–Meier curves for (**A**) OS and (**B**) PFS and radiographic response at the first RECIST measurement, landmarked at 70 days from the index; patients with an event prior to the 70-day landmark were excluded from the analysis. Data presented include the training and test datasets. Abbreviations: OS, overall survival; PFS, progression-free survival; NR, not reported; HR, adjusted hazard ratio; CI, confidence interval.

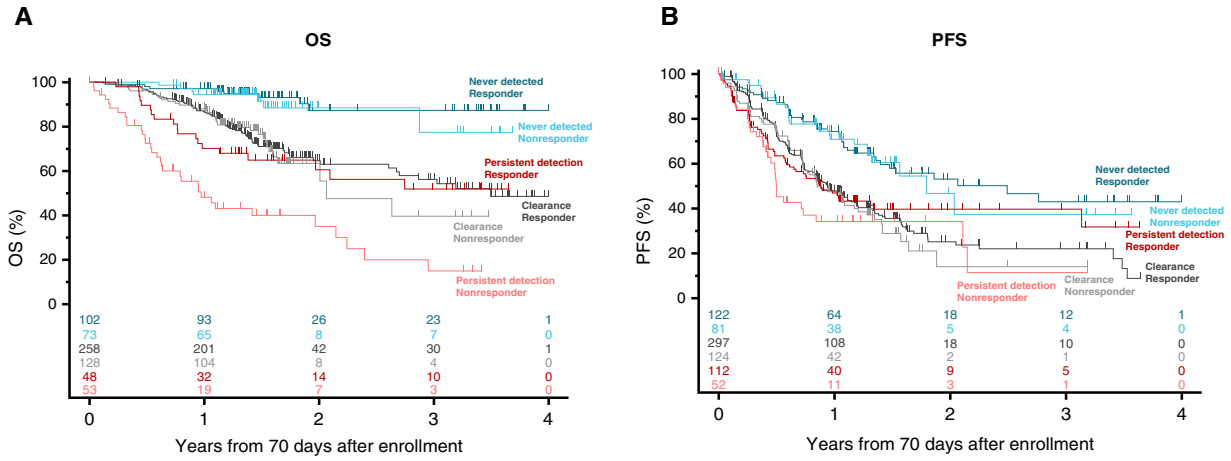
to align with the timing of ctDNA measurement, but assessing radiographic response by using a single measure is not used in clinical or regulatory decision-making. ORR would have aligned more closely with clinical practice and is an endpoint that supports some drug approvals. However, assessing ORR may take up to 6 months after the index date, which was the case for some patients in this analysis, extending the intended "early" window for response assessment and making landmarking challenging.

Landmarking was included to mitigate immortal time bias, given that patients must have survived long enough to have an on-treatment ctDNA sample collected (21). However, this also means that the landmarked analyses were only conducted among patients who were event-free prior to the landmark date, with $n = 7$ patients excluded from the OS analysis due to death and $n = 44$ patients excluded from the PFS analysis due to death or progression prior to the 70-day (10-week) landmark. The results may be biased because they exclude patients with the worst outcomes.

Focusing on early radiographic response affords an opportunity to interrogate the associations of ctDNA categories and survival outcomes in radiographic response categories with uncertainties. SD is defined by no change or a change that is not significant in the sum of tumor(s)/lesion(s) diameter and has been argued to have unclear clinical significance (35). Among patients with SD as an early radiographic response, ctDNA results meaningfully separated in terms of their associations with OS, suggesting that patients with early SD have heterogeneous survival that can be better understood with earlier ctDNA data. A similar phenomenon was seen in patients with aNSCLC treated with immunotherapy (36, 37), suggesting that the use of ctDNA to better stratify long-term outcomes among patients with SD by RECIST may apply across therapies.

Despite seeing readily interpretable results based on models stratified by cohorts, there were limitations to the analysis, which highlight current gaps in knowledge and set the stage for practical next steps to overcome challenges. Assay types (next-generation sequencing and Droplet Digital PCR), coverage (4–523 genes), performance and sensitivities, and bioinformatics pipelines (e.g., approach to filtering germline and/or clonal hematopoiesis of indeterminate potential variants) varied across the assays used in the different clinical trials. The analysis included all genes on the panels based on the intended use of the assay, rather than focusing on individual oncogenes of interest. Assessments that only consider the oncogene of interest would be interesting to explore, especially in this oncogene-driven space. However, such an approach would not be feasible for non-oncogene–driven therapies, such as immunotherapy, and is thus not favored for establishing broad recommendations for using ctDNA as an early endpoint. Regardless of the approach for assessing ctDNA, the field would benefit from a level of harmonization of assay outputs to ensure consistency and interpretability of findings.

Timing for ctDNA sample collection and the early RECIST measurements also differed across trials. To address this variability,



Category	Deaths/N	Median (Years)	1-Year estimate
Never detected Responder	8/102	NR	97% (94–100)
Never detected Nonresponder	7/73	NR	94% (89–100)
Clearance Responder	76/258	3.5 (2.7–NR)	87% (82–91)
Clearance Nonresponder	35/128	2.1 (2.0–NR)	88% (82–94)
Persistent detection Responder	19/48	NR	70% (57–83)
Persistent detection Nonresponder	33/53	1.0 (0.6–2.1)	48% (34–63)

Category	Events/N	Median (Years)	1-Year estimate
Never detected Responder	45/122	2.5 (1.4–NR)	74% (66–83)
Never detected Nonresponder	27/81	1.8 (1.3–NR)	71% (60–82)
Clearance Responder	175/297	0.9 (0.8–1.1)	48% (42–54)
Clearance Nonresponder	73/124	0.8 (0.7–1.2)	48% (38–57)
Persistent detection Responder	61/112	0.9 (0.6–1.3)	47% (37–57)
Persistent detection Nonresponder	31/52	0.5 (0.4–0.8)	34% (19–49)

Multivariable associations

Comparison	HR (95% CI), *P* value
Never detected Nonresponder vs. Responder	0.95 (0.34–2.64), 0.917
Clearance Nonresponder vs. Responder	0.91 (0.60–1.39), 0.675
Persistent detection Nonresponder vs. Responder	1.93 (1.04–3.58), 0.037

Multivariable associations

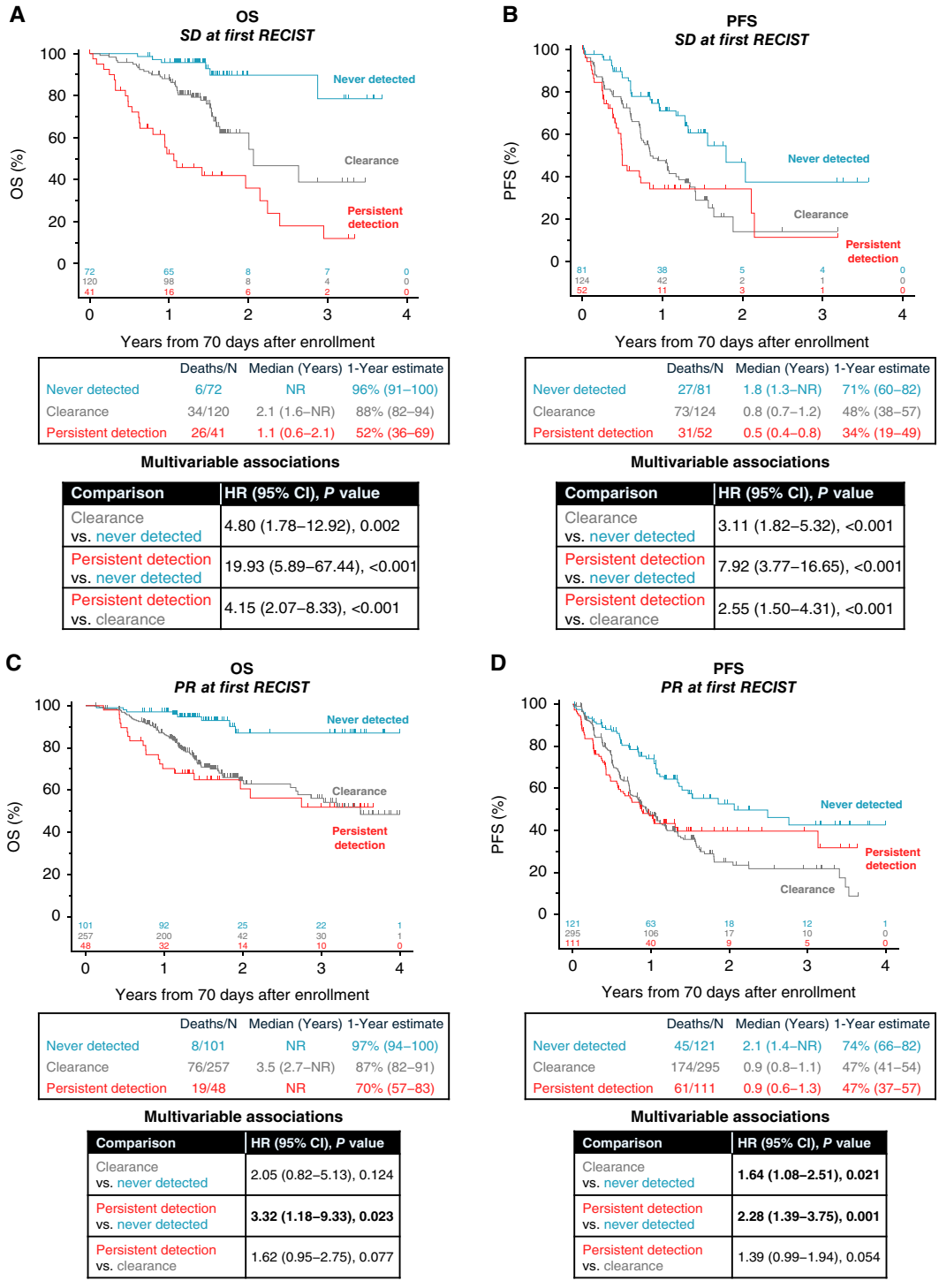
Comparison	HR (95% CI), *P* value
Never detected Nonresponder vs. Responder	0.80 (0.49–1.30), 0.373
Clearance Nonresponder vs. Responder	0.94 (0.71–1.25), 0.670
Persistent detection Nonresponder vs. Responder	1.29 (0.80–2.07), 0.296

Figure 4.

Kaplan–Meier curves for OS and PFS by combined radiographic response and ctDNA detection categories. Kaplan–Meier curves for (**A**) OS and (**B**) PFS and radiographic response at the first RECIST measurement, combined with ctDNA detection categories, landmarked at 70 days from the index; patients with an event prior to the 70-day landmark were excluded from the analysis. Data presented include the training and test datasets. Abbreviations: OS, overall survival; PFS, progression-free survival; NR, not reported; HR, adjusted hazard ratio; CI, confidence interval.

the analysis focused on the "best" ctDNA measurement, defined as the lowest VAF or ND within 10 weeks. These data support using ctDNA measurements taken within 10 weeks of treatment initiation, but the optimal timing for assessing how ctDNA is associated with OS remains unknown and will be critical to align on before ctDNA can be used as an early endpoint. Establishing specific collection schedules for ctDNA samples is an important goal for future prospective trials. Although timing may vary with disease and treatment, the field should work toward establishing best practices, including aligning on specific time points for sample collection and harmonizing assay protocols.

The FDA emphasizes the importance of patient-level meta-analyses to support the validation of ctDNA as an early endpoint for regulatory decision-making (10). Findings herein provide a patient-level aggregate analysis demonstrating an association between ctDNA clearance and improved OS. However, additional work is necessary to establish ctDNA as a reliable clinical and regulatory tool. Future studies should include prospectively defined analyses to evaluate changes in ctDNA levels and associations with long-term clinical outcomes. Focusing on disease settings that lack measurable disease by standard radiographic assessments (e.g., bone-only disease or no evidence of disease in the perioperative or postradiotherapy settings) would fulfill an unmet need in oncology. In any setting, it will be critical to predefine the clinically meaningful cutoff that establishes molecular response, which could be a change in VAF-based cutoff



	Deaths/N	Median (Years)	1-Year estimate
Never detected	6/72	NR	96% (91–100)
Clearance	34/120	2.1 (1.6–NR)	88% (82–94)
Persistent detection	26/41	1.1 (0.6–2.1)	52% (36–69)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	4.80 (1.78–12.92), 0.002
Persistent detection vs. never detected	19.93 (5.89–67.44), <0.001
Persistent detection vs. clearance	4.15 (2.07–8.33), <0.001

	Deaths/N	Median (Years)	1-Year estimate
Never detected	27/81	1.8 (1.3–NR)	71% (60–82)
Clearance	73/124	0.8 (0.7–1.2)	48% (38–57)
Persistent detection	31/52	0.5 (0.4–0.8)	34% (19–49)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	3.11 (1.82–5.32), <0.001
Persistent detection vs. never detected	7.92 (3.77–16.65), <0.001
Persistent detection vs. clearance	2.55 (1.50–4.31), <0.001

	Deaths/N	Median (Years)	1-Year estimate
Never detected	8/101	NR	97% (94–100)
Clearance	76/257	3.5 (2.7–NR)	87% (82–91)
Persistent detection	19/48	NR	70% (57–83)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	2.05 (0.82–5.13), 0.124
Persistent detection vs. never detected	**3.32 (1.18–9.33), 0.023**
Persistent detection vs. clearance	1.62 (0.95–2.75), 0.077

	Deaths/N	Median (Years)	1-Year estimate
Never detected	45/121	2.1 (1.4–NR)	74% (66–82)
Clearance	174/295	0.9 (0.8–1.1)	47% (41–54)
Persistent detection	61/111	0.9 (0.6–1.3)	47% (37–57)

Multivariable associations

Comparison	HR (95% CI), *P* value
Clearance vs. never detected	**1.64 (1.08–2.51), 0.021**
Persistent detection vs. never detected	**2.28 (1.39–3.75), 0.001**
Persistent detection vs. clearance	1.39 (0.99–1.94), 0.054

Figure 5.

Kaplan–Meier curves for OS and PFS by ctDNA detection in subgroups based on radiographic response. Kaplan–Meier curves for (**A**) OS and (**B**) PFS and ctDNA detection categories in those with SD at the first RECIST measurement and (**C**) OS and (**D**) PFS and ctDNA detection categories in those with PR at the first RECIST measurement, landmarked at 70 days from index; patients with an event during the 70-day landmark were excluded from the analysis. Data presented include the training and test datasets. Abbreviations: OS, overall survival; SD, stable disease; PFS, progression-free survival; NR, not reported; HR, adjusted hazard ratio; CI, confidence interval; PR, partial response.

or other variables such as tumor fraction or methylation-based measures. Regardless, better harmonization of assays and transparency of assay characteristics to enable aggregate analyses across trials and inform unifying definitions of response thresholds for ctDNA are needed. Overall, the data presented here support the potential for using changes in ctDNA levels as an early endpoint; however, more data are needed to validate their use. This analysis supported associations with ctDNA detection and long-term outcomes, setting the stage for future work that harmonizes the use of this unique biomarker.

Authors' Disclosures

N. Zariffa reports personal fees from Friends of Cancer Research during the conduct of the study as well as personal fees from Intelligencia AI, Beaconcure, ZS Associates, Genentech, Cytokinetics, ANOVA Enterprises Inc, AlphaLife Sciences, Bristol Myers Squibb, Pfizer, and the Bill and Melinda Gates Foundation outside the submitted work. S. Deng reports personal fees from Pfizer during the conduct of the study as well as personal fees from Pfizer outside the submitted work. C.R. Espenschied reports other support from Guardant Health outside the submitted work. M. Guha reports personal fees from Takeda Pharmaceuticals outside the submitted work. S. Hong reports employment with Agilent Technologies, Inc. D. Juraeva reports employment with Merck KGaA and ownership of Merck KGaA shares. L. Liu reports other support from Illumina outside the submitted work. J.-F. Martini reports other support from Pfizer outside the submitted work. G.R. Oxnard reports personal fees from Roche outside the submitted work. G.A. Pestano reports other support from Biodesix outside the submitted work. A.M. Szpurka reports other support from Eli Lilly and Company during the conduct of the study; in addition, A.M. Szpurka reports employment with and stock ownership in Eli Lilly and Company. D.M. Vega reports personal fees from AstraZeneca during the conduct of the study as well as personal fees from AstraZeneca outside the submitted work. In addition, D.M. Vega reports previous employment with Friends of Cancer Research, which is the organization leading these efforts; current employment with AstraZeneca; and ownership of AstraZeneca stock. C. Ward reports other support from Takeda Pharmaceutical Co. and other support from Moderna Therapeutics outside the submitted work. S.R. Wijayawardana reports personal fees from Eli Lilly and Company outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

H.S. Andrews: Conceptualization, supervision, methodology, writing–original draft, writing–review and editing. **N. Zariffa:** Conceptualization, formal analysis, supervision, validation, methodology, writing–original draft, writing–review and editing. **K.K. Nishimura:** Conceptualization, data curation, formal analysis, supervision, validation, visualization, methodology, writing–original draft, writing–review and editing. **S.H. Choi:** Resources, formal analysis, writing–review and editing. **S. Deng:** Conceptualization, resources, formal analysis, methodology, writing–review and editing. **M. Eisele:** Conceptualization, data curation, formal analysis, validation, visualization, methodology, writing–review and editing. **C.R. Espenschied:** Formal analysis, writing–review and editing. **E.M. Goren:** Data curation, formal analysis, validation, visualization, methodology, writing–review and editing. **M. Guha:** Resources, formal analysis, writing–review and editing. **S. Hong:** Formal analysis, writing–review and editing. **D. Juraeva:** Conceptualization, resources, formal analysis, methodology, writing–review and editing. **N. Krämer:** Conceptualization, formal analysis, methodology, writing–review and editing. **L. Liu:** Formal analysis, writing–review and editing. **J.-F. Martini:** Resources, formal analysis, writing–review and editing. **B.A. McKelvey:** Conceptualization, formal analysis, methodology, writing–review and editing. **G.R. Oxnard:** Conceptualization, formal analysis, writing–review and editing. **G.A. Pestano:** Formal analysis, writing–review and editing. **L. Poole:** Conceptualization, resources, formal analysis, methodology, writing–review and editing. **A. Rosenthal:** Conceptualization, data curation, formal analysis, validation, visualization, methodology, writing–review and editing. **A.M. Szpurka:** Resources, formal analysis, writing–review and editing. **D.M. Vega:** Resources, formal analysis, writing–review and editing. **C. Ward:** Resources, formal analysis, writing–review and editing. **S.R. Wijayawardana:** Conceptualization, resources, formal analysis, methodology, writing–review and editing. **A. Hoering:** Conceptualization, data curation, formal analysis, supervision, methodology, writing–review and editing. **M.D. Stewart:** Conceptualization, resources, formal analysis, supervision, methodology, writing–review and editing. **J.D. Allen:** Conceptualization, resources, formal analysis, supervision, funding acquisition, writing–review and editing.

Acknowledgments

We would like to thank the ctMoniTR Statistical Working Group members and the Broad Working Group members. We would also like to thank the companies that donated their patient-level clinical trial data and the patients for participating in the clinical trials. This work was funded by Friends of Cancer Research through unrestricted grants. Additional costs for management, aggregation, and uploading of clinical trial data were borne by each participating organization contributing data for this analysis: AstraZeneca, EMD Serono (CrossRef Funder ID: 10.13039/100004755), Eli Lilly/Loxo@Lilly, Pfizer, and Takeda Pharmaceutical Co.

Note

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

Received November 19, 2024; revised January 14, 2025; accepted March 27, 2025; posted first April 22, 2025.

References

1. Delgado A, Guddati AK. Clinical endpoints in oncology - a primer. Am J Cancer Res 2021;11:1121–31.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7–30.
3. Code of federal register 21CFR314 [cited 2025 Jan 8]. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=314&showFR=1&subpartNode=21:5.0.1.1.4.8.
4. Wilson WH, Schenkein DP, Jernigan CL, Woodcock J, Schilsky RL. Reevaluating the accelerated approval process for oncology drugs. Clin Cancer Res 2013;19:2804–9.
5. FDA. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry [cited 2025 Jan 8]. Available from: https://www.fda.gov/media/71195/download.
6. Villaruz LC, Socinski MA. The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. Clin Cancer Res 2013;19:2629–36.
7. Nishino M. Tumor response assessment for precision cancer therapy: response evaluation criteria in solid tumors and beyond. Am Soc of Clin Oncol Educ Book 2018;38:1019–29.
8. Litière S, Bogaerts J. Imaging endpoints for clinical trial use: a RECIST perspective. J Immunother Cancer 2022;10:3005092.
9. Vellanki PJ, Ghosh S, Pathak A, Fusco MJ, Bloomquist EW, Tang S, et al. Regulatory implications of ctDNA in immuno-oncology for solid tumors. J Immunother Cancer 2023;11:e005344.
10. FDA. Use of circulating tumor deoxyribonucleic acid for curative-intent solid solid tumor drug development. Final Guidance for Industry. [cited 2024 Nov 27]. Available from: https://www.fda.gov/media/183874/download.
11. Sanz-Garcia E, Genta S, Chen X, Ou Q, Araujo D V., Abdul Razak AR, et al. Tumor-Naïve circulating tumor DNA as an early response biomarker for patients treated with immunotherapy in early phase clinical trials. JCO Precis Oncol 2023;7:e2200509.
12. Raja R, Kuziora M, Brohawn PZ, Higgs BW, Gupta A, Dennis PA, et al. Early reduction in ctDNA predicts survival in patients with lung and bladder cancer treated with durvalumab. Clin Cancer Res 2018;24:6212–22.
13. Bratman S V., Yang SYC, Iafolla MAJ, Liu Z, Hansen AR, Bedard PL, et al. Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. Nat Cancer 2020;1:873–81.
14. Vega DM, Nishimura KK, Zariffa N, Thompson JC, Hoering A, Cilento V, et al. Changes in circulating tumor DNA reflect clinical benefit across multiple studies of patients with non-small-cell lung cancer treated with immune checkpoint inhibitors. JCO Precis Oncol 2022;6:e2100372.

15. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.
16. Lyman GH. Evaluation and regulation of oncology drug approval: finding the right balance. JAMA Oncol 2016;2:728–9.
17. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. J Clin Oncol 2008; 26:3913–5.
18. Cox DR. Regression models and life-tables. J R Stat Soc 1972;34:187–202.
19. Cochran WG. Some methods for strengthening the common χ^2 tests. Biometrics 1954;10:417–51.
20. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719–48.
21. Morgan CJ. Landmark analysis: a primer. J Nucl Cardiol 2019;26:391–3.
22. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki BF, editors. Proceedings of the 2nd International Symposium on Information Theory. Budapest (Hungary): Academiai Kiado; 1973. p. 267–81.
23. Wilks S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 1938;9:60–2.
24. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30:1105–17.
25. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–6.
26. Fukuhara T, Saito H, Furuya N, Watanabe K, Sugawara S, Iwasawa S, et al. Evaluation of plasma EGFR mutation as an early predictor of response of erlotinib plus bevacizumab treatment in the NEJ026 study. EBioMedicine 2020;57:102861.
27. Roldan Ruiz J, Fuentes Gago MG, Chinchilla Tabora LM, Gonzalez Morais I, Sayagues JM, Abad Hernández M, et al. The impact of liquid biopsies positive for EGFR mutations on overall survival in non-small cell lung cancer patients. Diagnostics (Basel) 2023;13:2347.
28. Mack PC, Miao J, Redman MW, Moon J, Goldberg SB, Herbst RS, et al. Circulating tumor DNA kinetics predict progression-free and overall survival in EGFR TKI–treated patients with EGFR-mutant NSCLC (SWOG S1403). Clin Cancer Res 2022;28:3752–60.
29. Soo RA, Martini JF, van der Wekken AJ, Teraoka S, Ferrara R, Shaw AT, et al. Early circulating tumor DNA dynamics and efficacy of lorlatinib in patients with treatment-naive, advanced, ALK-positive NSCLC. J Thorac Oncol 2023; 18:1568–80.
30. Wu TH, Hsiue EHC, Lee JH, Lin CC, Liao WY, Ho CC, et al. Best response according to RECIST during first-line EGFR-TKI treatment predicts survival in EGFR mutation-positive non–small-cell lung cancer patients. Clin Lung Cancer 2018;19:e361–72.
31. Jiang J, Zhao C, Zhang F, Liu Z, Zhou K, Ren X, et al. ALK inhibitors in ALK-rearranged non-small cell lung cancer with and without brain metastases: systematic review and network meta-analysis. BMJ Open 2022;12: e060782.
32. Gao G, Ren S, Li A, Xu J, Xu Q, Su C, et al. Epidermal growth factor receptor-tyrosine kinase inhibitor therapy is effective as first-line treatment of advanced non-small-cell lung cancer with mutated EGFR: a meta-analysis from six phase III randomized controlled trials. Int J Cancer 2012;131:E822–9.
33. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011;12: 735–42.
34. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non–small-cell lung cancer with mutated EGFR. N Engl J Med 2010;362:2380–8.
35. Grossi F. Management of non-small cell lung in cancer patients with stable disease. Drugs 2012;72(Suppl 1):20–7.
36. Zhang Q, Luo J, Wu S, Si H, Gao C, Xu W, et al. Prognostic and predictive impact of circulating tumor DNA in patients with advanced cancers treated with immune checkpoint blockade. Cancer Discov 2020;10:1842–53.
37. Assaf ZJF, Zou W, Fine AD, Socinski MA, Young A, Lipson D, et al. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer. Nat Med 2023;29:859–68.

Journal for ImmunoTherapy of Cancer

Molecular response cutoffs and ctDNA collection timepoints influence on interpretation of associations between early changes in ctDNA and overall survival in patients treated with anti-PD(L)1 and/or chemotherapy

Hillary S Andrews ,[1] Nevine Zariffa,[2] Katherine K Nishimura,[3] Yu Deng,[4] Megan Eisele,[3] Joe Ensor,[5] Carin Espenschied,[6] David Fabrizio,[7] Emily M Goren,[3] Vincent Haddad,[8] Minetta C Liu,[5] Dimple Modi,[9] Achim K Moesta,[9] Katie J Quinn,[6] Adam Rosenthal,[3] Diana Merino Vega ,[10] Wei Zou,[4] Antje Hoering,[3] Mark Stewart ,[1] Jeff Allen[1]

To cite: Andrews HS, Zariffa N, Nishimura KK, *et al*. Molecular response cutoffs and ctDNA collection timepoints influence on interpretation of associations between early changes in ctDNA and overall survival in patients treated with anti-PD(L)1 and/or chemotherapy. *Journal for ImmunoTherapy of Cancer* 2025;**13**:e012454. doi:10.1136/jitc-2025-012454

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jitc-2025-012454).

Accepted 27 August 2025

For numbered affiliations see end of article.

Correspondence to
Dr Hillary S Andrews; handrews@focr.org

ABSTRACT

Background Circulating tumor DNA (ctDNA) is a promising intermediate end point for oncology drug development, potentially accelerating regulatory approvals by providing early insights into treatment response. However, challenges remain in standardizing ctDNA assessment, including optimal blood collection timing and defining molecular response (MR) cutoffs. The ctDNA for Monitoring Treatment Response (ctMoniTR) project, led by Friends of Cancer Research, aggregates patient-level data from clinical trials to evaluate associations between ctDNA changes and overall survival (OS).

Methods This analysis included four randomized clinical trials of patients with advanced non-small cell lung cancer (aNSCLC) treated with either anti-programmed death (ligand) 1 (anti-PD(L)1) therapy (with or without chemotherapy) or chemotherapy alone. MR was assessed using three predefined per cent-change thresholds in ctDNA levels (≥50% decrease, ≥90% decrease, and 100% clearance). ctDNA samples were analyzed at two timepoints: an early window (T1, up to 7 weeks post-treatment initiation) and a later window (T2, 7–13 weeks post-treatment initiation). Multivariable Cox proportional hazards models and time-dependent analyses were used to evaluate associations between ctDNA changes and OS.

Results A total of 918 patients were included. In the anti-PD(L)1 group, ctDNA reductions at both T1 and T2 were significantly associated with improved OS across all MR thresholds. In the chemotherapy group, associations were weaker at T1 but became more pronounced at T2. Patients with MR at both T1 and T2 had the strongest OS associations. Overall, the results suggest that T2 had marginally stronger association with OS than T1.

Conclusions This study supports the potential of ctDNA as an intermediate end point in aNSCLC, with MR at both early (T1) and later (T2) timepoints showing significant associations with OS. Differences in ctDNA dynamics between treatment modalities highlight the importance of considering the timing of blood collection. Further research is needed to determine the optimal time window for assessing ctDNA response. Prospective trials and trial-level meta-analyses will be critical to validating ctDNA as a regulatory-grade intermediate end point for oncology drug development.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Circulating tumor DNA (ctDNA) is an emerging biomarker with potential as an intermediate end point for regulatory decision-making in oncology drug development.

⇒ Prior studies have shown that decreases in ctDNA levels are associated with improved overall survival (OS), but questions remain regarding the optimal timing of plasma collection and how to define molecular response (MR) across different treatment modalities.

INTRODUCTION

As cancer therapies improve, so does overall survival (OS), which is beneficial for patients but can make further drug development challenging. The US Food and Drug Administration (FDA) considers OS as the gold standard end point for efficacy in clinical trials of a new cancer therapy. Waiting for OS readouts can result in extensive lead times for getting novel therapies to patients. To overcome this, the Accelerated Approval pathway offers an opportunity for drug approval based on an intermediate end point that is reasonably likely to predict benefit.[1] In order to use novel intermediate end points in clinical

WHAT THIS STUDY ADDS

⇒ This study addresses a critical gap in our understanding of ctDNA dynamics, as it pertains to the association between reductions in ctDNA on treatment and improved outcomes.

⇒ We provide insight into the optimal analytical cutoffs to define MR and timepoints for plasma collection that will support harmonization across studies.

⇒ Integrating data from four randomized clinical trials advanced non-small cell lung cancer (aNSCLC) treated with anti-programmed death (ligand) 1 (anti-PD(L)1) therapy and/or chemotherapy, we evaluated MR cutoffs of 50%, 90%, and 100% reduction from baseline as well as ctDNA collection timepoints.

⇒ Results demonstrated that ctDNA reductions at both early (T1) and later (T2) timepoints were significantly associated with OS using any of the MR cutoffs in patients treated with either anti-PD(L)1 or chemotherapy.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings provide further evidence supporting ctDNA as a potential intermediate end point in oncology trials, highlighting the importance of timing considerations in assessing ctDNA response, and underscoring the need for further prospective trials and trial-level meta-analyses to validate ctDNA for use as a regulatory grade end point in drug development.

trials for regulatory decision-making, the FDA expects meta-analyses to validate the end point at both the patient and trial levels.[2]

Circulating tumor DNA (ctDNA) holds promise as an intermediate end point in oncology drug development; however, there are outstanding questions that fall into three main categories: (1) the timing and frequency of blood collection for ctDNA assessment, (2) the specific definition of molecular response (MR) (eg, what per cent ctDNA change cut-off should be used), including whether this differs across treatment modalities and/or cancer types, and (3) assay standardization—which assays are most appropriate, including which characteristics should be considered for determining appropriateness. Trials to date have focused on a variety of approaches for demonstrating associations between decreased ctDNA on treatment and improved OS,[3] but a lack of consistency in measuring and reporting ctDNA change makes establishing the necessary data to support using it as an intermediate end point challenging.

Friends of Cancer Research (*Friends*) created the ctDNA for Monitoring Treatment Response (ctMoniTR) project to aggregate patient-level data from previously completed clinical trials and assess whether changes in ctDNA levels are associated with OS in different cancer types and treatment modalities. The ctMoniTR project's goal is to identify a unifying approach for assessing associations between changes in ctDNA and OS. An initial pilot study of five clinical trials of 200 patients with advanced non-small cell lung cancer (aNSCLC) treated with an anti-programmed death (ligand) 1 (anti-PD(L)1) showed that using a >50% decrease to define MR was associated with improved OS.[4] In a second study of eight clinical trials representing >1000 patients with aNSCLC treated with a tyrosine kinase inhibitor (TKI), using clearance of ctDNA to define MR was associated with improved OS.[5] These findings suggest that there may be differences in how to define MR by treatment class, and these differences may reflect the distinct mechanisms of action of different drug classes. In each analysis, the majority of patients had a single ctDNA measurement taken at variable times up to 10 weeks after baseline, which made it challenging to compare how changes in ctDNA levels at different timepoints were associated with OS.

In the current analysis, the ctMoniTR working group analyzed four randomized clinical trials (RCTs) of patients with aNSCLC treated with either an anti-PD(L)1 with or without chemotherapy (deemed 'anti-PD(L)1 group' throughout) or chemotherapy alone. The primary goal was to explore different per cent change cutoffs to define MR in relation to OS. Once established, definitions of MR were used to analyze the similarities and differences between measurements taken during an early 'T1' window (ie, up to 7 weeks post-treatment initiation) and a late 'T2' window (ie, 7–13 weeks post-treatment initiation) in both the anti-PD(L)1 and the chemotherapy groups. Findings from this study offer key insights into ctDNA dynamics and its association with OS, which can inform future efforts to refine design and analysis parameters for prospective trials.

METHODS

Overall project approach

Friends coordinated a working group with expert representatives from pharmaceutical companies, ctDNA assay developers, academia, and the FDA to discuss the analysis plan, react to emerging analysis results, and align on the interpretation of the findings. These retrospective analyses are considered to be exploratory and hypothesis-generating. Cancer Research And Biostatistics (CRAB) served as the data aggregator and independent statistical analysis team and performed all analyses.

Clinical trials and patients

The four clinical trials enrolled patients with aNSCLC treated with anti-PD(L)1 therapy with or without chemotherapy ('anti-PD(L)1 group') or with chemotherapy alone ('chemotherapy group'; online supplemental table 1). Each Sponsor reviewed the informed consent forms approved by the local institutional review board to ensure that the data were suitable for secondary use beyond their original intent. Sponsors anonymized the data at the patient level and mapped the data to a universal data dictionary prior to submission to CRAB. Once CRAB cleaned and formatted the data, the Sponsors approved the final datasets for quality assurance purposes before all data were pooled.

ctDNA data

Sponsors provided variant allele frequency (VAF) values for each ctDNA variant detected in plasma cell-free DNA

samples collected from clinical trial participants. VAF values were provided as either numerical or non-detected (ND) as reported to them by the assay developer. The specific gene variants assessed for each patient depended on the assay used in each trial, which included commercially available next-generation sequencing (NGS) assays. Prior to data submission, results were filtered by the Sponsor to only include tumor-related variants, with clonal hematopoiesis of indeterminate potential (CHIP) and germline mutations removed using PBMC-based clearance or biopsy sequencing. The assays included in this analysis had a limit of detection (LOD) ranging from 0.1% to 0.5% VAF. For each patient at each timepoint, the maximum VAF was calculated among all VAF results reported in a given sample (ie, the highest VAF value within each ctDNA sample), where applicable.[4] To be eligible for the analysis, patients were required to have a baseline ctDNA sample (0–14 days prior to the start of therapy) and at least one on-treatment sample within 7 weeks of treatment initiation. Initial analyses focused on an 'Early' on-treatment sample (T1), defined as a ctDNA sample within 7 weeks of treatment initiation. Separate analyses investigating repeated ctDNA measurements required patients to have both a T1 and a 'Late' on-treatment sample, defined as 7–13 weeks from treatment initiation (T2). These time windows were defined based on the combined availability of samples across all four clinical trials. If more than one ctDNA sample was available within the T1 time window, then the earliest sample was used, and if more than one ctDNA sample was available within the T2 time window, then the latest sample was used.

Derived ctDNA metrics

The per cent change of the ctDNA maximum VAF from baseline was the primary measurement under investigation using the following equation (where on-treatment is either T1 or T2 depending on the analysis):

$$\text{Per cent change} = (\text{Max VAF}_{\text{On treatment}} - \text{Max VAF}_{\text{Baseline}}) / \text{Max VAF}_{\text{Baseline}}$$

Three MR thresholds were predetermined by the ctMoniTR working group based on prior experience and evidence: 50% decrease, 90% decrease, and 100% decrease (ie, clearance of ctDNA or a change from a detected to ND ctDNA value).[6] Of the 918 patients included in the analysis data set, 111 patients had ND ctDNA at baseline and were excluded from the main analyses because a per cent change cannot be calculated from an undetected baseline sample and therefore these patients cannot provide information to support the evaluation of MR rate.[7] However, there was interest in understanding how patients with ND ctDNA at baseline compared with the others, so these patients were included in a descriptive assessment aimed at demonstrating a gradient of response, with patients classified into five groups based on changes in ctDNA: (1) ND/ND (ie, ND ctDNA at baseline and on-treatment), (2) 100% decrease (ie, detected at baseline and ND on treatment or clearance), (3) ≥90%–100% decrease, (4) ≥50%–90% decrease, and (5) <50% decrease or increase. In the primary analyses, specific cutoffs to define binary MRs were prioritized (eg, a molecular responder for 50% cut-off (MR50) was a ≥50% decrease in ctDNA and non-molecular responder for 50% cut-off (nMR50) was a <50% decrease in ctDNA or an increase in ctDNA).

Statistical analyses

A statistical plan outlined the key research hypotheses and analytical approaches. Multivariable Cox proportional hazards models were used to estimate the association between the ctDNA metric and OS.[8] Given the expected heterogeneity of patient populations between the studies, all models were stratified by cohort, where each cohort was a unique trial arm, allowing for a different baseline hazard for each cohort. Multivariable models were adjusted for several clinical descriptors including age (<65 vs ≥65 years), sex, race (white vs else), smoking status (ever smoked vs never smoked), stage (stage IV vs else (stages I–III), with the majority of else (80%) being stage III), Eastern Cooperative Oncology Group Performance Status (≥1 vs 0),[9] number of prior lines of therapy (≥1 vs 0), histology (squamous vs else), PD-L1 expression (<50% vs ≥50%), and small baseline maximum VAF (>0.5 vs ≤0.5).

Given that the primary predictor category (ie, the change in ctDNA from baseline) could not be determined until the on-treatment sample was collected, two approaches were used to account for the repeated ctDNA measurements. The data were landmarked at 49 days for analyses focused on T1 ctDNA and 91 days for those that included T2 ctDNA.[10] Additionally, Cox models with time-dependent covariates allowed for the inclusion of all ctDNA measurements, accounted for sample collection timing, and did not exclude patients with events prior to the landmark time.

Model fit statistics were used to compare the predictive value of T1 and T2 ctDNA, including Akaike Information Criterion (AIC),[11] Bayesian Information Criterion (BIC),[12] Harrel's C-index,[13] and Likelihood Ratio Tests.[14] Receiver operating characteristic (ROC) curves compared the model performance of ctDNA changes association with OS at various timepoints, including 7 weeks (T1 window), 13 weeks (T2 window), and 6 months, corresponding to timeframes when both ctDNA and survival outcomes were available for all studies.

All statistical tests were two-sided, with a p value of 0.05 used for nominal significance. Analyses were performed using the SAS statistical software package (SAS Institute, Cary, North Carolina, USA) or R (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Analysis dataset and demographic and baseline characteristics

Across four RCTs, 2042 patients received anti-PD(L)1 therapy and/or chemotherapy (online supplemental

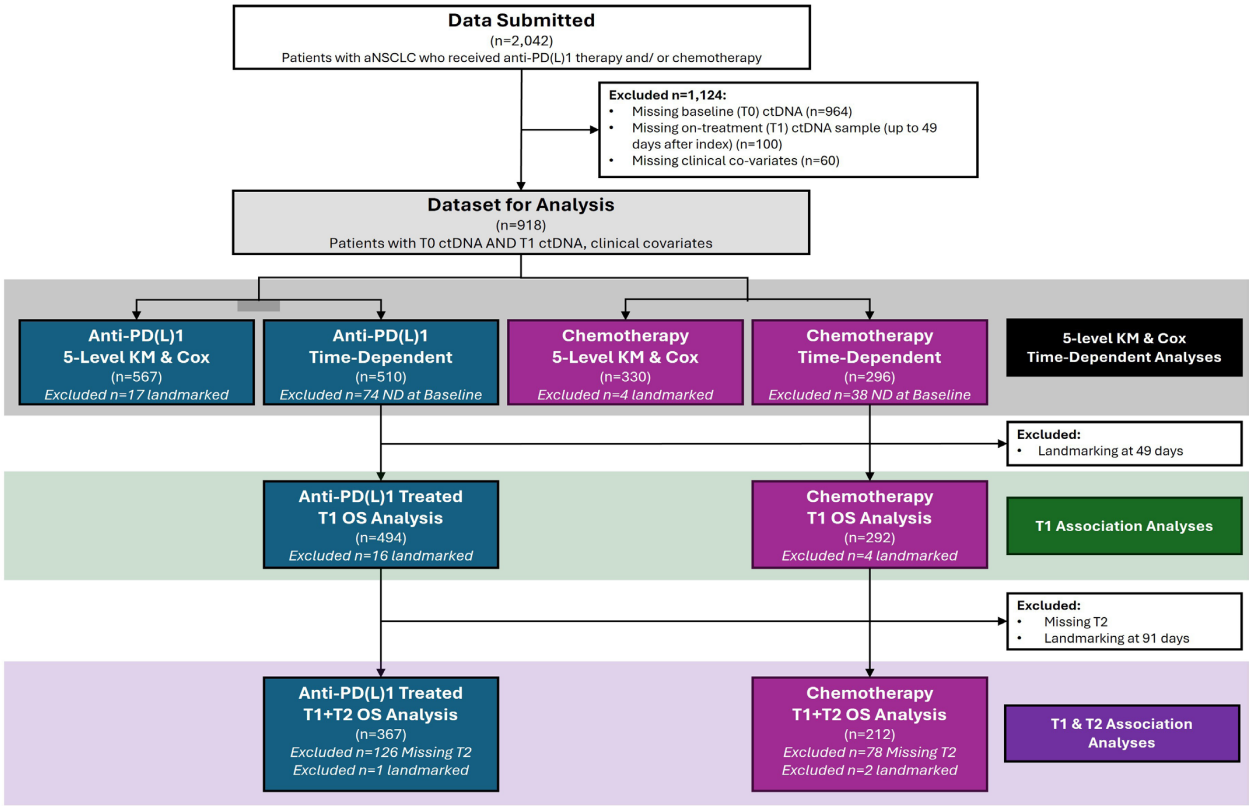


Figure 1 Consolidated Standards of Reporting Trials diagram of patient selection and analysis cohorts. Overview of the selection process for patients included in the study, beginning with the total number of submitted cases and detailing exclusions due to missing data. anti-PD(L)1, anti-programmed death (ligand) 1; aNSCLC, advanced non-small cell lung cancer; ctDNA, circulating tumor DNA; KM, Kaplan-Meier; ND, non-detected; OS, overall survival.

table 2). Of these, 1124 patients were excluded from subsequent analyses if they were missing a baseline or T1 ctDNA sample or clinical co-variates leading to an analysis dataset of 918 patients. There was further exclusion for landmarking and/or requiring a T2 ctDNA measurement depending on the analysis (figure 1).

Changes in ctDNA levels and associations with clinical outcomes in patients treated with anti-PD(L)1 and/or chemotherapy

To assess whether a gradient of response existed, patients were categorized into five groups (online supplemental figure 1). Generally, a gradient of response was observed in both anti-PD(L)1-treated and chemotherapy-treated patients. A greater per cent change in ctDNA levels was associated with improved OS, which suggests exploring each cut-off as its own MR category could be beneficial. Additionally, we analyzed detected versus ND ctDNA at baseline (without any consideration of the on-treatment ctDNA values; online supplemental figure 2) and found no significant association between ctDNA detection at baseline and OS for either the anti-PD(L)1 or the chemotherapy group. The remaining analyses excluded the patients with ND ctDNA at baseline.

When assessing the relationship between changes in ctDNA at T1 among the patients treated with anti-PD(L)1 (figure 2A–C), using 50% decrease, 90% decrease, or 100% decrease to define MR showed an association with OS (adjusted HR (aHR) for MR50=1.99, p<0.001; aHR for MR90=2.18, p<0.001; aHR for MR100=2.50, p<0.001), indicating a higher risk of death in the absence of MR. Results were similar for the chemotherapy group (figure 2D–F), but associations were weaker compared with the anti-PD(L)1 group (aHR for MR50=1.58, p=0.006; aHR for MR90=1.44, p=0.016; aHR for MR100=1.71; p=0.026).

An ongoing consideration regarding changes in ctDNA is whether small VAF values (ie, ≤1.0) influence the interpretation of findings, which led to a subgroup analysis restricted to patients who had a VAF ≤1.0 at T0 and T1 (online supplemental figure 3). The sample sizes were insufficient for Cox multivariable modeling; however, there were similar patterns of MR being associated with improved OS for each cut-off in both the anti-PD(L)1 and the chemotherapy groups.

Comparison of early versus late ctDNA associations with OS

The next set of analyses compared the strength of association between ctDNA at an early timepoint (T1) versus

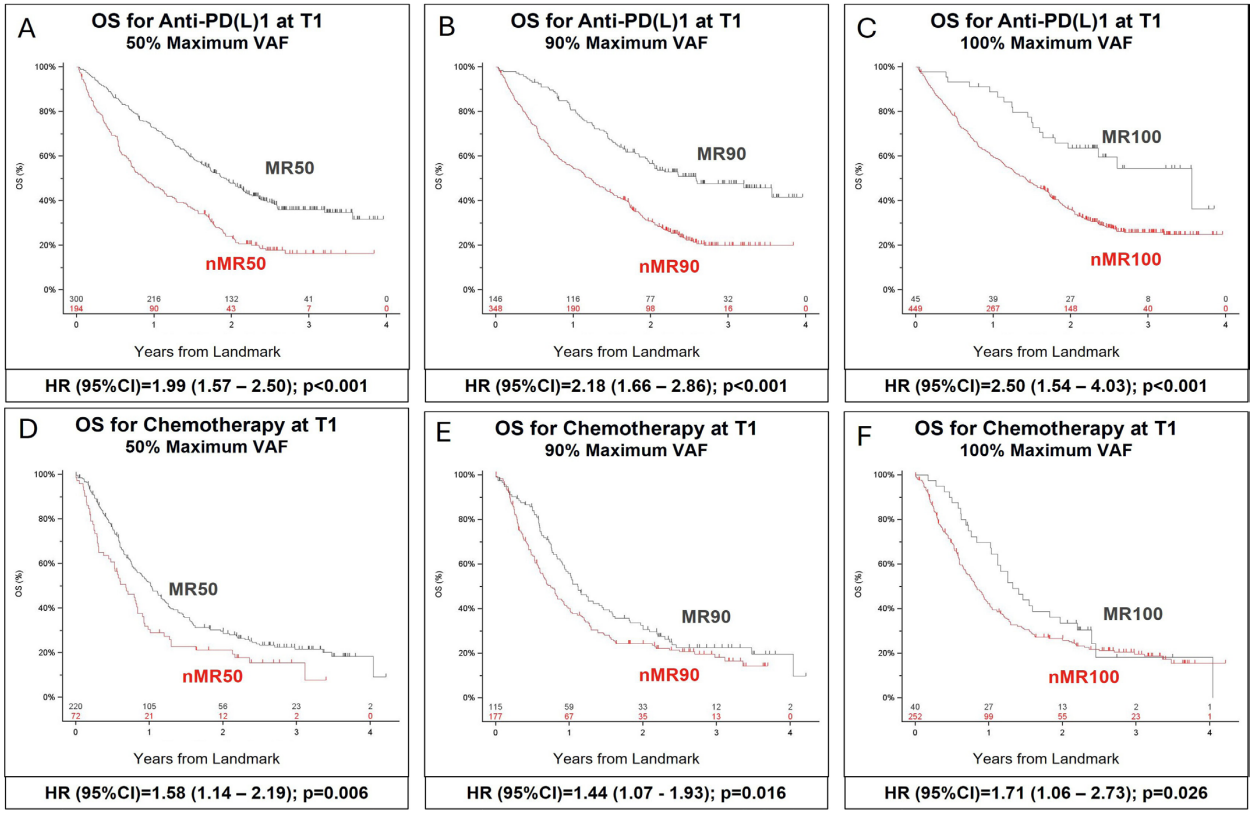


Figure 2 Multivariable association between change in ctDNA at T1 among patients treated with anti-PD(L)1 and chemotherapy. Kaplan-Meier survival curves depict the association between early (T1) ctDNA changes and OS, grouped by treatment type: anti-PD(L)1 therapy (A–C) and chemotherapy (D–F). Molecular response is categorized based on ctDNA reduction thresholds of ≥50% (A, D), ≥90% (B, E), and 100% clearance (C, F). anti-PD(L)1, anti-programmed death (ligand) 1; ctDNA, circulating tumor DNA; MR, molecular responder; nMR, non-molecular responder; OS, overall survival; VAF, variant allele frequency.

a late timepoint (T2) and OS. To ensure valid model fit comparisons, these analyses were restricted to patients who had both T1 and T2 ctDNA samples leading to a reduction in the analysis dataset from 494 to 367 patients in the anti-PD(L)1 group and from 292 to 212 patients in the chemotherapy group. OS was calculated from the day 91 landmark for each analysis. We compared the two timepoints within the same population to reduce bias due to differing data collection time schedules across cohorts and to allow for model-based comparisons. The changes in ctDNA at T1 and T2 were analyzed separately using the three MR cutoffs.

For the anti-PD(L)1 group, MR was associated with improved OS using all cutoffs and timepoints (figure 3A–F). The magnitude of the difference was larger for the T2 timepoint compared with the T1 timepoint for the 50% and 100% cutoffs (aHR for MR50 at T1=1.42, p=0.019vs at T2=2.15, p<0.001; aHR for MR100 at T1=2.08, p=0.009vs at T2=3.34, p<0.001) and similar for the 90% cut-off (aHR for MR90 at T1=1.89, p<0.001vs at T2=1.89, p<0.001). Interestingly, for the chemotherapy group (figure 4A–F), there was not a statistically significant difference in the association between each of the MR categories for the three cutoffs and OS at T1 (aHR for MR50 at T1=1.41, p=0.145; for MR90 at T1=1.32, p=0.121; for MR100 at T1=1.36, p=0.304). However, there were statistically significant associations between MR and improved OS for the chemotherapy at T2 using each of the cutoffs (aHR for MR50 at T1=1.71, p=0.031; aHR for MR90 at T1=2.16, p<0.001; aHR for MR100=2.56, p<0.001).

Model fit statistics to identify the best model given the T1 and T2 ctDNA results (online supplemental table 3) suggest that overall, T2 has a better model fit compared with T1 for both anti-PD(L)1 and chemotherapy groups. T2 alone had better or comparable fit to a model with both T1 and T2 by AIC and BIC, and likelihood ratio tests indicated T2 had a significant association above and beyond T1.

Confirmation of MR categories between T1 and T2 and associations with OS

The next analysis investigated the value of using repeated ctDNA measurements to confirm earlier results. Like Response Evaluation Criteria in Solid Tumors (RECIST) assessments,[15] there may be value in confirming MR by measuring ctDNA levels at a second timepoint. Four

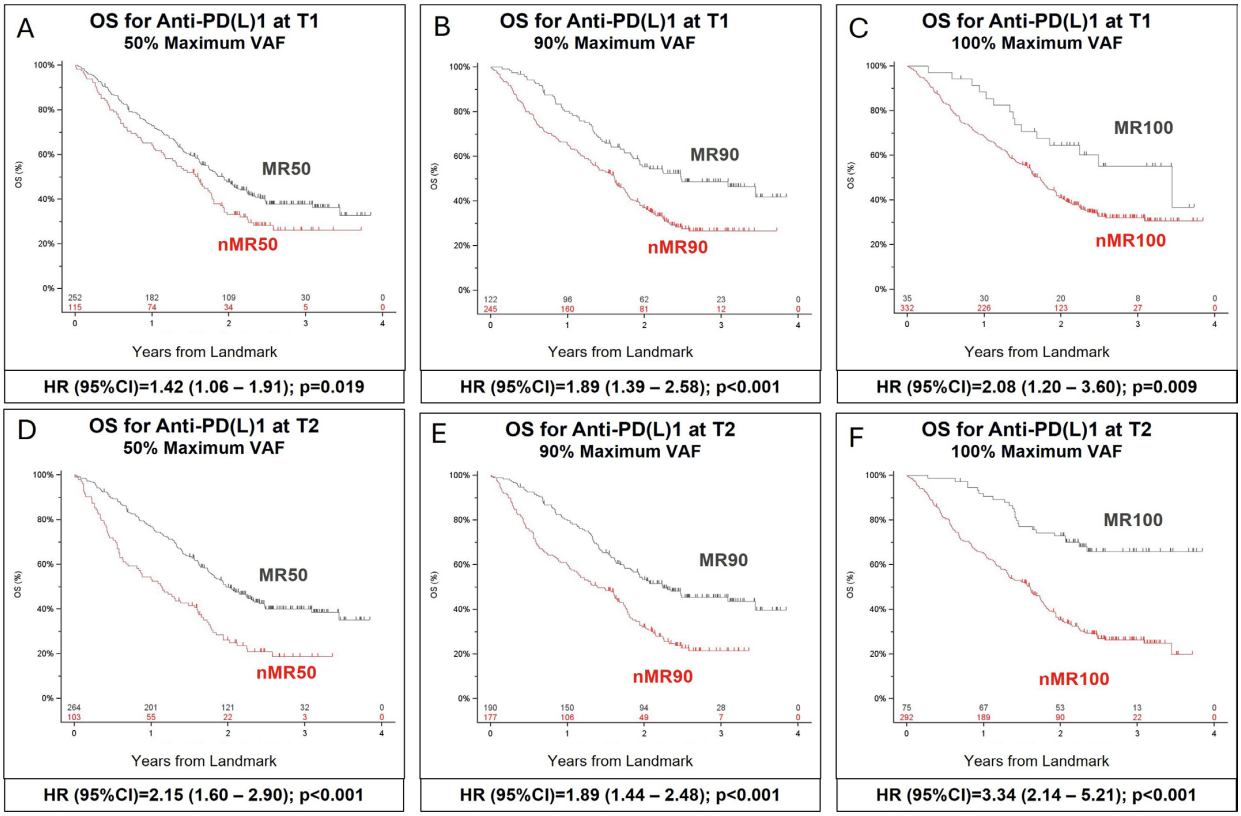


Figure 3 Multivariable association between change in ctDNA and OS at T1 and T2 among patients treated with anti-PD(L)1. Kaplan-Meier survival curves show the association between ctDNA changes and OS at early (T1; A–C) and later (T2; D–F) timepoints in patients treated with anti-PD(L)1 therapy. Molecular response is categorized based on ctDNA reduction thresholds of ≥50% (A, D), ≥90% (B, E), and 100% clearance (C, F). anti-PD(L)1, anti-programmed death (ligand) 1; ctDNA, circulating tumor DNA; MR, molecular responder; nMR, non-molecular responder; OS, overall survival; VAF, variant allele frequency.

groups were established based on categorical MR at T1 and T2: (1) confirmed MR (MR at T1, MR at T2; MR/MR), (2) confirmed nMR (nMR at T1, nMR at T2; nMR/nMR), (3) delayed MR (nMR at T1, MR at T2; nMR/MR), and (4) delayed nMR (MR at T1, nMR at T2; MR/nMR). In this dataset, most patients fell into the same response category at T1 and T2 (for the anti-PD(L)1 group, n for MR50=293/367 (80%), n for MR90=275/367 (75%), n for MR100=305/367 (83%) and for the chemotherapy group, n for MR50=184/212 (87%), n for MR90=161/212 (76%), n for MR100=190/212 (90%)).

Kaplan-Meier curves were created, and a multivariable Cox proportional hazards model was run on these four groups to understand associations with OS for the anti-PD(L)1 (figure 5A–C) and the chemotherapy group (figure 6A–C). For patients treated with anti-PD(L)1, the confirmed MR group (MR/MR) had improved OS compared with the confirmed nMR (nMR/nMR) for MR50 (aHR=2.01, p<0.001), MR90 (aHR=2.29, p<0.001), and MR100 (aHR=3.47, p=0.001). Similar results were seen for patients treated with chemotherapy using MR90 (aHR=2.05, p<0.001) and MR100 (aHR=2.55, p=0.012), but not for MR50 (aHR=1.56, p=0.100).

When the T1 and T2 MR categories did not match, the T2 MR category appeared to be a better indicator of OS outcomes than the T1 MR category (ie, a patient who was nMR at T1 and then MR at T2 appeared to have survival outcomes more consistent with a confirmed responder than a confirmed non-responder). For the anti-PD(L)1 group, this was true for confirmed MR (MR/MR) compared with delayed nMR (MR/nMR) using MR50 (aHR=3.35, p<0.001), but not the MR90 or MR100, where sample size limitations exist (n for MR/nMR=31 for MR50, 12 for MR90, and 11 for MR100). Similarly, confirmed nMR (nMR/nMR) was associated with worse OS compared with the delayed MR (nMR/MR) group for MR90 (aHR=1.55, p=0.014) and MR100 (aHR=3.36, p<0.001). There was a trend toward worse OS for the MR50 group (aHR=1.55, p=0.076).

Similar trends were seen for the chemotherapy group. When compared with those with confirmed MR (MR/MR), those with delayed nMR (MR/nMR) had associations with

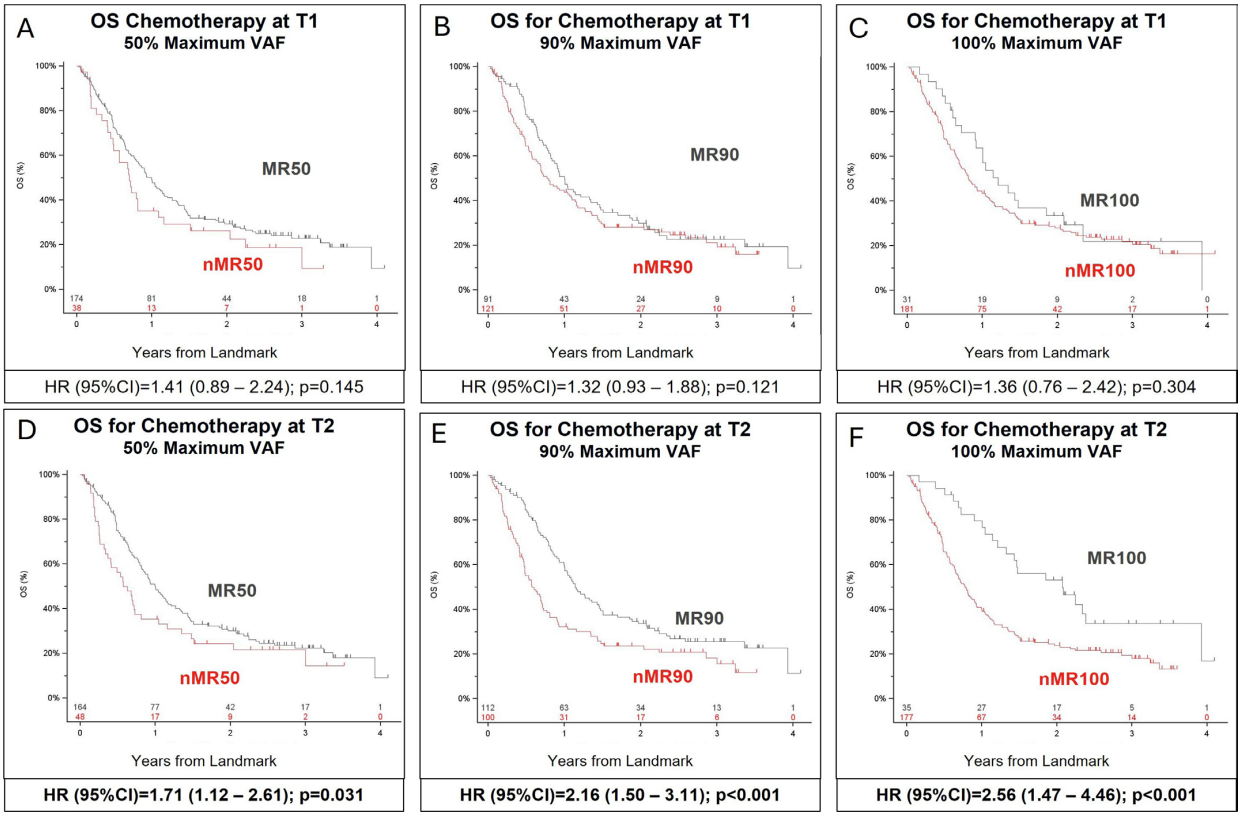


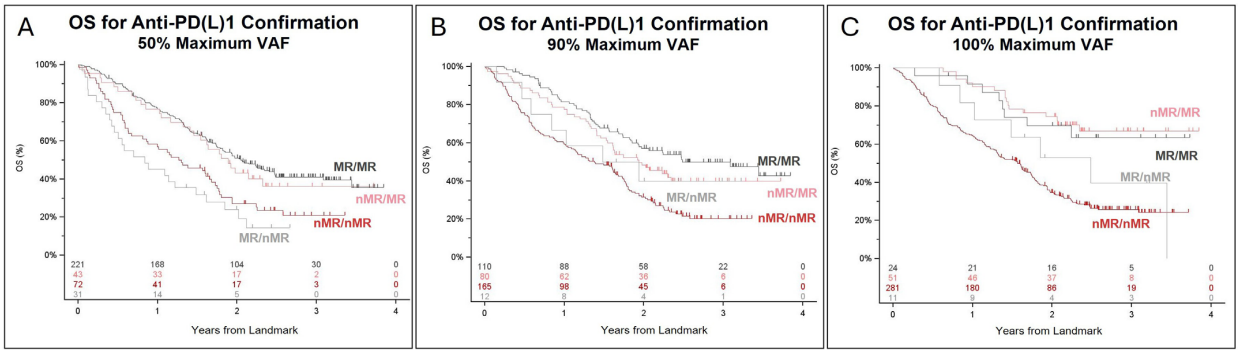
Figure 4 Multivariable association between change in ctDNA and OS at T1 and T2 among patients treated with chemotherapy. Kaplan-Meier survival curves illustrate the association between ctDNA changes and OS at early (T1; A–C) and later (T2; D–F) timepoints in patients treated with chemotherapy. Molecular response is categorized based on ctDNA reduction thresholds of ≥50% (A, D), ≥90% (B, E), and 100% clearance (C, F). anti-PD(L)1, anti-programmed death (ligand) 1; ctDNA, circulating tumor DNA; MR, molecular responder; nMR, non-molecular responder; OS, overall survival; VAF, variant allele frequency.

worse OS for MR50 (aHR=1.99, p=0.018), MR90 (aHR=3.84, p<0.001), and MR100 (aHR=4.06, p=0.004). Also, when comparing those with confirmed nMR (nMR/nMR) and those with delayed MR (nMR/MR), there were associations with worse OS for those with confirmed nMR for MR90 (aHR=1.74, p=0.035) and for MR100 (aHR=2.31, p=0.033).

To better understand the prevalence and potential influence of small VAF values (ie, ≤1.0 maximum VAF results), the distribution of VAF values over time was visualized (online supplemental figure 4). Despite concerns that per cent change values may not be appropriate when raw VAF measurements are small, there did not appear to be any visual indication that small VAF measurements were being misclassified into inappropriate confirmation categories.

Cox models with time-dependent covariates

Since the timing of the ctDNA measurements was not consistent between studies, Cox models with time-dependent covariates were conducted as an alternative to landmarked analyses. The benefit of this approach was that it allowed for the maximum inclusion of all patients and ctDNA results, using all ctDNA measurements and accounting for their true collection timing post-treatment (online supplemental table 4). Among the anti-PD(L)1 group, patients who had MR showed improved OS using all three thresholds (aHR for MR50=2.68, p<0.001; aHR for MR90=2.68, p<0.001; aHR for MR100=3.74, p<0.001). There was a similar pattern among the chemotherapy group with a lower magnitude (aHR for MR50=1.69, p<0.001; aHR for MR90=1.86, p<0.001; aHR for MR100=2.31, p<0.001). ROC curves based on the MR category after 7 weeks, 13 weeks, or 6 months of treatment were included to assess model performance. The ROC curves show minor improvements over time; however, most patients had stable MRs, resulting in ROC curves that appear very similar for all three timepoints (online supplemental figure 5). These results suggest that while there is some added benefit for collecting a later ctDNA sample, as evidenced by the modest improvements of the AUC values from T1 (7 weeks), T2 (13 weeks), and 6 months, the early ctDNA samples are often a good indication of long-term prognosis and earlier information may outweigh modest improvements in prediction.



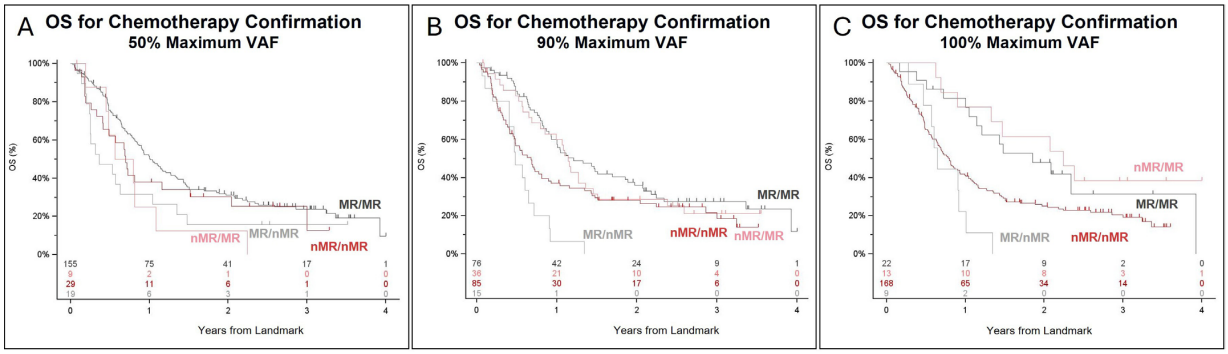
Multivariable Associations HR (95% CI); p-value		Reference								
		50% Maximum VAF			90% Maximum VAF			100% Maximum VAF		
		MR/MR	nMR/MR	nMR/nMR	MR/MR	nMR/MR	nMR/nMR	MR/MR	nMR/MR	nMR/nMR
Comparator	nMR/MR	1.29 (0.82-2.03); p=0.266	-	-	1.47 (0.97-2.23); p=0.067	-	-	1.03 (0.43-2.47); p=0.941	-	-
	nMR/nMR	**2.01 (1.42-2.83); p<0.001**	1.55 (0.95-2.53); p=0.076	-	**2.29 (1.63-3.21); p<0.001**	**1.55 (1.09-2.20); p=0.014**	-	**3.47 (1.65-7.30); p=0.001**	**3.36 (1.99-5.69); p<0.001**	-
	MR/nMR	**3.35 (2.05-5.49); p<0.001**	**2.60 (1.45-4.65); p<0.001**	1.67 (1.00-2.79); p=0.050	1.64 (0.72-3.70); p=0.237	1.11 (0.49-2.50); p=0.802	0.72 (0.33-1.56); p=0.400	2.28 (0.80-6.50); p=0.123	2.21 (0.88-5.53); p=0.092	0.66 (0.30-1.43); p=0.288

Figure 5 Confirmation and multivariable associations with OS among patients treated with anti-PD(L)1 based on categorical MR at T1 and T2. Kaplan-Meier survival curves (A–C) and multivariable HRs depict the association between confirmed molecular response MR at T1 and T2 and OS in patients treated with anti-PD(L)1 therapy. anti-PD(L)1, anti-programmed death (ligand) 1; MR, molecular responder; nMR, non-molecular responder; OS, overall survival; VAF, variant allele frequency.

DISCUSSION

We performed an aggregate analysis of four previously completed RCTs in 918 eligible patients with aNSCLC who were treated with anti-PD(L)1 and/or chemotherapy to provide data that support using ctDNA as an intermediate end point. The analysis focused on three main areas: (1) exploring optimal cutoffs to define MR, (2) the impact of the timing of the ctDNA collection timepoint, and (3) evaluating two different drug classes. Using three different cutoffs to define MR (ie, >50% decrease, >90% decrease, and 100% decrease), MR was associated with improved OS in patients with aNSCLC treated with anti-PD(L)1 and/or chemotherapy. Associations were seen using either a landmarked multivariable Cox model or a time-dependent Cox model, suggesting the associations are robust and consistent.

Defining the optimal timing for MR assessment is crucial for using ctDNA as an intermediate end point, both to provide consistency in the design of prospective clinical trials as well as interpretation of findings across future studies in future meta-analyses to validate the use of ctDNA as an intermediary end point. To date, many studies assessing associations between changes in ctDNA and OS focus on a single blood collection timepoint up to 10 weeks after treatment initiation.[3] While some analyses focus on ctDNA at multiple timepoints for tracking molecular residual disease following surgery or treatment to support clinical decision-making,[16] there are limited studies that assess multiple timepoints outside of the first 10 weeks of treatment initiation for associations with OS to support regulatory decision-making. In this study, two time windows were evaluated: 0–7 weeks (T1) and 8–13 weeks (T2) based on the time of collection in the clinical trials. In most cases, Sponsors selected ctDNA collection timepoints that aligned with the protocol-defined assessment schedule in each trial. Statistically, T2 had stronger associations with OS compared with T1. However, this effect varied by treatment group, with anti-PD(L)1 therapy showing stronger associations at T1, whereas significant associations for chemotherapy emerged primarily at T2. Practically, there are additional considerations for defining the optimal time window for assessing ctDNA, including differences in ctDNA dynamics between treatment modalities and the potential implications for early treatment decision-making. The ctDNA taken at T1 is at or before the time when the first tumor scan was analyzed for RECIST-based measurements, which highlights the value of ctDNA in tracking response earlier and across a larger set of patients. However, lead-time bias remains a concern at the second timepoint. Despite this, most patients with available data in both time windows remained in the same MR category (~80%). Multiple studies, including the results presented here, demonstrate associations between changes in ctDNA up to 10 weeks after treatment initiation and OS, suggesting this time window should be retained for consideration. Future



Multivariable Associations HR (95% CI); p-value		Reference: 50% Maximum VAF			Reference: 90% Maximum VAF			Reference: 100% Maximum VAF		
		MR/MR	nMR/MR	nMR/nMR	MR/MR	nMR/MR	nMR/nMR	MR/MR	nMR/MR	nMR/nMR
Comparator	nMR/MR	1.43 (0.63-3.22); p=0.390	-	-	1.18 (0.71-1.95); p=0.531	-	-	1.10 (0.40-3.01); p=0.850	-	-
	nMR/nMR	1.56 (0.92-2.64); p=0.100	1.09 (0.43-2.75); p=0.854	-	**2.05 (1.35-3.11); p<0.001**	**1.74 (1.04-2.91; p=0.035**	-	**2.55 (1.23-5.28); p=0.012**	**2.31 (1.07-5.00); p=0.033**	-
	MR/nMR	**1.99 (1.12-3.53); p=0.018**	1.39 (0.52-3.74); p=0.510	1.28 (0.63-2.58); p=0.494	**3.84 (2.00-7.39); p<0.001**	**3.27 (1.62-6.59); p<0.001**	1.88 (1.00-3.54); p=0.051	**4.06 (1.55-10.69); p=0.004**	**3.69 (1.33-10.25); p=0.012**	1.60 (0.75-3.38); p=0.221

Figure 6 Confirmation and multivariable associations with OS among patients treated with chemotherapy based on categorical MR at T1 and T2. Kaplan-Meier survival curves (A–C) and multivariable HRs illustrate the association between confirmed molecular response (MR) at T1 and T2 and OS in patients treated with chemotherapy. MR, molecular responder; nMR, non-molecular responder; OS, overall survival; VAF, variant allele frequency.

analyses should explore additional timepoints to better determine the optimal timing for ctDNA assessment.

An aligned approach for defining MR is essential to assess associations with outcomes.[17] In this study, all three cutoffs evaluated (ie, >50%, >90%, and 100% decrease) were associated with improved OS. While statistically, aHRs were larger for the 100% decrease group compared with the others for both the anti-PD(L)1 or chemotherapy groups, there are other factors to consider such as sample sizes within each category and whether one cut-off could serve as a unifying approach across cancer types and treatment modalities beyond those studied here. Future analyses will also aim to compare MR rates between treatment arms to identify which offers greater clinical benefit to determine trial-level association between ctDNA and OS. Additionally, if 100% (ie, clearance) is used, assay characteristics are even more important to ensure limits of detection to define clearance are aligned.[18] This study included assays that were either tumor-informed or tumor-naïve but cleared from CHIP using PBMCs, which should be considered for future studies since these approaches promote accurate classification of variants as tumor-derived.[19]

Assay sensitivities and limits of detection can influence interpretation of findings. An ongoing concern regarding using a per cent change or clearance to define MR is the role of small VAF values, as these could cause misclassification due to a large numerical value of the per cent decrease. Use of an absolute change, alone or in combination with a per cent change threshold, has been proposed to overcome this, but feasibility depends on the scale of the distribution of the ctDNA values, which can differ substantially from one assay to another. Using a per cent change is preferable when assessing multiple assays at once to facilitate an aggregate trial analysis where each trial has its own range of values for ctDNA due to the variety of assays implemented. In the current study, small VAF values (ie, ≤1.0) did not appear to influence the interpretation of the findings. Another feature to consider when aggregating results from trials with varying assays is the definition of ND, which may differ depending on the reported sensitivity of each assay. Regardless, patients with ND at baseline should be excluded from analyses focused on assessing ctDNA dynamics to interpret response to treatment due to the inability to provide information about response. Additionally, ND could be due to pre-analytical or technical assay limitations, such as low plasma volume or the assay LOD, rather than the absence of ctDNA.

While this study represents an opportunity to further understand associations between changes in ctDNA and OS, there are limitations. Importantly, this analysis was performed on previously completed RCTs; thus, it was not powered a priori and subject to limitations of

sample size in some of our analyses. Associations were seen for both the anti-PD(L)1 group, which included patients who received both anti-PD(L)1 and chemotherapy, and the chemotherapy-only group. However, there were not sufficient patients to further distinguish how patients treated with both anti-PD(L)1 and chemotherapy performed. Additionally, decisions were made about the definitions of T1 and T2 based on the timing of the collection across the various studies; however, some patients had more than one measurement within the windows. The earliest timepoint in T1 and the latest in T2 were each predefined as the values for analysis; however, other analyses might consider the 'best' or lowest ctDNA measurement within the time window.[5] To address potential limitations with either strategy, the time-dependent analysis was conducted since it includes all ctDNA results, regardless of whether they occurred in our prespecified time windows, or if multiple measurements occurred during a single time window. Our findings from the time-dependent analyses were consistent with those from the landmark analyses, supporting the conclusion that the interpretation of the results would remain unchanged regardless of the analytical approach used. Conversely, some patients lacked a ctDNA measurement in both the T1 and T2 time window. For analyses that required both T1 and T2 samples—such as the comparison of early versus late response (figure 4, online supplemental table 3) and the confirmation of MR (figures 5–6)—patients who experienced death or progression before T2 collection were necessarily excluded. This resulted in an analysis population that removed patients who had an event prior to the start of the T2 time window. While we do not have the ability to report the reason for a missing T2 sample in many instances, we acknowledge that this could result in a biased analysis population. However, to fairly compare the performance of T1 versus T2, this analysis must be conducted within identical patient populations. Also, a similar precedent exists in trials reporting a confirmed best response, which typically require a confirmatory RECIST scan approximately 4 weeks after the first assessment.[15]

These findings add to the growing body of evidence demonstrating that on-treatment reductions in ctDNA are associated with improved OS. However, important questions remain regarding the optimal timing of blood collection, the definition of MR, and how changes in ctDNA may vary across treatment modalities and cancer types.[20–23] Additionally, while ctDNA holds significant potential as an intermediate end point in advanced diseases, particularly when OS follow-up is prolonged or image-based end points are unreliable, it may be especially valuable in earlier-stage settings, where OS can take many years to mature. In these settings, ctDNA may also help inform more timely clinical decisions regarding treatment response and therapeutic cycling. Both applications require additional research and can build on the foundational evidence established here regarding ctDNA collection timepoints and cutoffs to support future work.

Despite differences in timing of ctDNA collection and assays used for patients with aNSCLC, MR was associated with OS for both the anti-PD(L)1 group and the chemotherapy group. In many cases, associations were not as strong for the chemotherapy-only group, which may be due to sample sizes or drug modality-based differences in ctDNA dynamics. As the scientific community works toward aligning on the ideal timing of blood collection for ctDNA assessment and approaches for defining MR, the potential influence of drug class and cancer type should be considered as well. It is important to note that these analyses are confined to aNSCLC, and the applicability of these MR cutoffs beyond aNSCLC, including in other disease settings such as resectable disease, will require further investigation in future translational studies.

The data outlined here provide a critical step toward using ctDNA as an intermediate end point; however, additional work is necessary before ctDNA can be used in regulatory decision-making. This analysis was a patient-level aggregate assessment of RCTs, and trial-level meta-analyses are also necessary.[18 22] To perform trial-level meta-analyses, prospectively designed trials should include a harmonized approach in key features of trial design, execution, and collection of data. In fall 2024, *Friends* convened a working group of experts to outline key considerations for what to include in trials and outlined the need for aligned timing of blood collection for ctDNA and assay considerations.[22] To maximize regulatory impact, future trials should incorporate prespecified analyses of ctDNA-based end points and adopt harmonized definitions for molecular response and timing of collection, to ensure consistency and robustness in the evidence generated. Once these prospectively designed trials are completed, a trial-level meta-analysis will be performed to further support the use of ctDNA as an intermediate end point.

Author affiliations
[1]Friends of Cancer Research, Washington, District of Columbia, USA
[2]NMD Group Inc, Bala Cynwyd, Pennsylvania, USA
[3]Cancer Research And Biostatistics, Seattle, Washington, USA
[4]Genentech Inc, South San Francisco, California, USA
[5]Natera Inc, Austin, Texas, USA
[6]Guardant Health Inc, Redwood City, California, USA
[7]Foundation Medicine Inc, Cambridge, Massachusetts, USA
[8]AstraZeneca, Cambridge, UK
[9]Regeneron Pharmaceuticals Inc, Tarrytown, New York, USA
[10]AstraZeneca, Gaithersburg, Maryland, USA

Present affiliations The present affiliation of Vincent Haddad is: Adcendo ApS, Copenhagen, Denmark.

Contributors HSA, NZ, KKN, YD, ME, JE, CE, DF, EMG, VH, MCL, DM, AKM, KJQ, AR, DMV, WZ, AH, MS, and JA discussed the approach, aligned on findings, and reviewed the manuscript. HSA and MDS wrote manuscript drafts. KKN, ME, EMG, AR, and AH performed statistical analyses. HSA is the guarantor.

Funding This work was funded by Friends of Cancer Research through unrestricted grants. Additional costs for the management, aggregation, and uploading of clinical trial data were borne by each participating organization contributing data to this analysis: AstraZeneca, Genentech, and Regeneron Pharmaceuticals.

Competing interests NZ reports personal fees from Friends of Cancer Research during the conduct of the study as well as personal fees from Intelligencia AI,

Beaconcure, ZS Associates, Genentech, Cytokinetics, ANOVA Enterprises, AlphaLife Sciences, Bristol Myers Squibb, Pfizer, and the Bill and Melinda Gates Foundation outside the submitted work. YD reports employment with Genentech. JE reports employment with Natera. CE reports other support from Guardant Health outside the submitted work. DF reports employment with Foundation Medicine. VH reports previous employment with AstraZeneca and is currently employed with Adcendo. MCL is currently employed with Natera. DM is employed with Regeneron Pharmaceuticals. AKM is employed with Regeneron Pharmaceuticals. DMV reports personal fees from AstraZeneca during the conduct of the study as well as personal fees from AstraZeneca outside the submitted work. In addition, DMV reports previous employment with Friends of Cancer Research, which is the organization leading these efforts; current employment with AstraZeneca; and ownership of AstraZeneca stock. WZ is currently employed by Genentech. No disclosures were reported by the other authors.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data are not publicly available due to data use agreements with pharmaceutical companies that provided data for the purposes of conducting the study and privacy and ethical considerations for patients. On request and subject to review by all authors, data that support the findings of the study may be provided, including code for running statistical analyses.

ORCID iDs
Hillary S Andrews http://orcid.org/0000-0002-0177-2024
Diana Merino Vega http://orcid.org/0000-0002-6173-6957
Mark Stewart http://orcid.org/0000-0002-4847-0736

REFERENCES

1 FDA Center for Drug Evaluation and Research. Guidance for Industry: Expedited Program for Serious Conditions-Accelerated Approval of Drugs and Biologics. 2014. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/expedited-programs-serious-conditions-drugs-and-biologics
2 FDA Oncology Center of Excellence. Use of Circulating Tumor DNA for Curative-Intent Solid Tumor Drug Development: Guidance for Industry. 2024. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-circulating-tumor-deoxyribonucleic-acid-early-stage-solid-tumor-drug-development-guidance
3 Murray JC, Sivapalan L, Hummelink K, *et al*. Elucidating the Heterogeneity of Immunotherapy Response and Immune-Related Toxicities by Longitudinal ctDNA and Immune Cell Compartment Tracking in Lung Cancer. *Clin Cancer Res* 2024;30:389–403.
4 Vega DM, Nishimura KK, Zariffa N, *et al*. Changes in Circulating Tumor DNA Reflect Clinical Benefit Across Multiple Studies of Patients With Non-Small-Cell Lung Cancer Treated With Immune Checkpoint Inhibitors. *JCO Precis Oncol* 2022;6:e2100372.
5 Andrews HS, Zariffa N, Nishimura KK, *et al*. ctDNA Clearance as an Early Indicator of Improved Clinical Outcomes in Advanced NSCLC Treated with TKI: Findings from an Aggregate Analysis of Eight Clinical Trials. *Clin Cancer Res* 2025;31:2162–72.
6 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* 2021;5:510–24.
7 Huang RY, Bi WL, Weller M, *et al*. Proposed Response Assessment and Endpoints for Meningioma Clinical trials: Report from the Response Assessment in Neuro-Oncology Working Group. *Neuro Oncol* 2019;21:26–36.
8 Cox DR. Regression Models and Life-Tables. *J R Stat Soc Series B* 1972;34:187–202.
9 Oken MM, Creech RH, Tormey DC, *et al*. Toxicity and Response Criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;5:649–55.
10 Morgan CJ. Landmark Analysis: A Primer. *J Nucl Cardiol* 2019;26:391–3.
11 Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. *Biogeochemistry* 1998;199–213.
12 Schwarz G. Estimating the Dimension of a Model. *Ann Statist* 1978;6:461–4.
13 Harrell FE Jr, Califf RM, Pryor DB, *et al*. Evaluating the Yield of Medical Tests. *JAMA* 1982;247:2543–6.
14 Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann Math Statist* 1938;9:60–2.
15 Eisenhauer EA, Therasse P, Bogaerts J, *et al*. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* 2009;45:228–47.
16 Anagnostou V, Forde PM, White JR, *et al*. Dynamics of Tumor and Immune Responses during Immune Checkpoint Blockade in Non–Small Cell Lung Cancer. *Cancer Res* 2019;79:1214–25.
17 Stadler J-C, Belloum Y, Deitert B, *et al*. Current and Future Clinical Applications of ctDNA in Immuno-Oncology. *Cancer Res* 2022;82:349–58.
18 Vellanki PJ, Ghosh S, Pathak A, *et al*. Regulatory Implications of ctDNA in Immuno-oncology for Solid Tumors. *J Immunother Cancer* 2023;11:e005344.
19 Abbosh C, Kuziora M, Stetson D, *et al*. Abstract 2299: Misclassification of Clonal Hematopoietic Mutations as Tumor-Derived Confound Accurate ctDNA-Response Assessment in Patients with Metastatic Non-Small-Cell Lung Cancer: Analyses from the Phase 3 MYSTIC Trial. *Cancer Res* 2023;83:2299.
20 Brenner Thomsen C, Dandanell Juul A, Lefèvre AC, *et al*. Reporting on Circulating Tumor DNA Monitoring in Metastatic Cancer-From Clinical Validity to Clinical Utility. *Cancer* 2022;128:2052–7.
21 Thompson JC, Scholes DG, Carpenter EL, *et al*. Molecular Response Assessment using Circulating Tumor DNA (ctDNA) in Advanced Solid Tumors. *Br J Cancer* 2023;129:1893–902.
22 Friends of Cancer Research. Framework for Integrating Change in ctDNA Levels in Advanced Cancer Clinical Trials to Support Meta-Analyses for Intermediate Endpoint Validation. 2024. Available: https://friendsofcancerresearch.org/wp-content/uploads/Framework-for-Integrating-Change-in-ctDNA-Levels-in-Advanced-Cancer-Clinical-Trials-to-Support-Meta-analyses-for-Intermediate-Endpoint-Validation.pdf
23 Ribba B, Roller A, Helms HJ, *et al*. Circulating Tumor DNA: Opportunities and Challenges for Pharmacometric Approaches. *Front Pharmacol* 2022;13:1058220.

Hurry Up and Wait: Timelines and Takeaways from the Biomarker Qualification Program

Grace Collins[1] · Jeff D. Allen[1] · Hillary S. Andrews[1] · Bernat Navarro-Serer[1] · Mark D. Stewart[1]

Received: 1 August 2025 / Accepted: 13 October 2025

Abstract

Background The Biomarker Qualification Program (BQP), formally established in 2016 under the *21st Century Cures Act*, is a key pathway for developing novel biomarkers for regulatory use. We evaluated eight years of BQP experience to assess whether it has facilitated the qualification of novel biomarkers.

Methods We collected characteristics and submission dates for accepted biomarker qualification projects from the FDA's Drug Development Tool Qualification Project Search database.

Results As of July 1, 2025, 61 projects were accepted into the BQP. Safety (30%), Diagnostic (21%), and PD Response (20%) biomarkers were the most common. Projects primarily used molecular (46%) and radiologic/imaging (39%) methods and were split between measures of a disease/condition or drug response/effect of exposure. Few projects included surrogate endpoint biomarkers ($n=5$). Half of the accepted projects remained at the initial Letter of Intent (LOI) stage, and only eight biomarkers were qualified through the program. LOI and Qualification Plan (QP) reviews frequently exceeded FDA targets by three months and seven months, respectively. For projects reaching the QP stage, QP development took a median of 32 months, with surrogate endpoints taking 47 months.

Conclusion The BQP supports the development of certain biomarkers but has seen limited use for biomarkers intended as surrogate endpoints. Coupled with longer timelines for their QP development, these trends demonstrate the program may not be well-suited for advancing novel response biomarkers. Given significant stakeholder interest in novel surrogate measures, a dedicated program may better support novel response biomarker development, particularly for biomarkers with applicability across multiple drug development programs.

Keywords Oncology · Biomarker · Endpoint · Qualification · FDA

Introduction

Novel biomarkers are increasingly valuable tools for accelerating evidence generation and regulatory decision-making by informing patient selection and providing earlier measures of treatment efficacy and safety. Currently, the U.S. Food and Drug Administration (FDA) offers two primary "pathways" for stakeholders seeking to develop and validate novel biomarkers for regulatory use, both of which require extensive evidence generation for a novel biomarker to be considered reliable for regulatory use. Most commonly, novel biomarker development occurs during a clinical development program, with validation through regulatory review and approval of a new drug application (NDA) or biologics license application (BLA). The second "pathway", the Biomarker Qualification Program (BQP), began in 2007 and its review framework was formally established under the *21st Century Cures Act* on December 13, 2016, in response to stakeholder calls for a more collaborative, structured, and transparent process for biomarker development and validation [1, 2]. The BQP's goals are to "support outreach to stakeholders for the identification and development of new biomarkers; provide a framework for the review of biomarkers for use in regulatory decision-making; and qualify biomarkers for specific contexts of use that address specified drug development needs"[3].

To achieve BQP goals, participants submit information and receive FDA feedback through a structured, three-phase

✉ Grace Collins
gcollins@focr.org

[1] Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC, USA

process: (1) letter of intent (LOI), (2) qualification package (QP), and (3) full qualification package (FQP) (Fig. 1).

In the final stage, the Center for Drug Evaluation and Research (CDER) qualifies the biomarker for the defined context of use (COU) in any drug development program to support regulatory decision-making. The phased and collaborative BQP process is intended to facilitate qualification of regulatory-grade biomarkers by providing an avenue for development outside the context of a single clinical development program, offering feedback at each submission, and the ability to seek advice through meetings with BQP staff.

Previous analyses have examined participation in and the impact of other drug development tool (DDT) qualification programs, such as the clinical outcome assessment (COA) qualification program, and characterized the stakeholder groups engaged in the BQP [4, 5]. Here, we assess the use of the BQP over the past eight years to evaluate whether the program facilitated the qualification of novel biomarkers as intended, including those for use as early measures of treatment efficacy (i.e., surrogate endpoints).

Materials & Methods

We collected data on biomarker qualification projects from the CDER and Center for Biologics Evaluation and Research (CBER) Drug Development Tool (DDT) Qualification Project Search database [6].

Project Characteristics and Submission Progress

For each project, we collected information on the latest submission stage (LOI, QP, FQP), latest submission status as of July 1, 2025 (accept, not accept, qualified, submitted, withdrawn/rescinded), and project characteristics (biomarker category, biomarker type, measure of [disease or condition, drug response/effect of exposure], and surrogate endpoint [yes/no]). Where available, we also recorded the project's date of acceptance into the program, dates of each submission (LOI, QP, FQP), and the date of the FDA's determination for each submission. Although not a formal stage of the BQP submission process, dates of submission and determination for Section 507 Transition Plans were also collected for projects initiated under the FDA's legacy process for drug development tool qualification.

Submission Timelines

For projects with available submission and determination dates, we calculated LOI review time (months from LOI submission to FDA determination) and QP review time (months from QP submission to FDA determination). For projects that submitted a qualification plan, we also calculated QP development time (months from FDA determination for the LOI to QP submission). For biomarker projects that reached full qualification, the database included only the date(s) of qualification; therefore, time to qualification could not be calculated. In some cases, multiple submission and determination dates were reported (e.g., an initial QP submission that received a "not accept", followed by a QP resubmission, and an FDA "accept" determination). In these instances, we used the last reported submission and determination date.

Results

As of July 1, 2025, the FDA's DDT Qualification Project Database listed 99 projects under the biomarker qualification program. Sixty-one of these projects (62%) were accepted into the program.

Characteristics of Accepted Biomarker Qualification Projects

Accepted projects represented a diverse range of biomarkers with varying methods of assessment and intended uses (Table 1).

Safety biomarkers were most common (18/61, 30%) followed by diagnostic biomarkers (13/61, 21%), pharmacodynamic (PD) response biomarkers (12/61, 20%), and prognostic biomarkers (12/16, 20%). For the method of assessment (i.e., biomarker type), molecular biomarkers (28/61, 46%) and radiologic/imaging biomarkers (24/61,

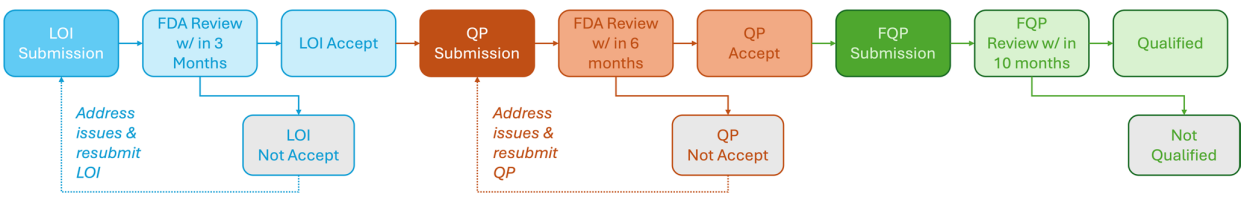


Fig. 1 Overview of the drug development tool qualification process. The *21st Century Cures Act* established a structured, three-stage process for drug development tool qualification to provide a transparent, collaborative pathway to qualification. LOI, Letter of Intent; FQP, Full Qualification Package; QP, Qualification Plan

Table 1 Characteristics and latest submission stage of accepted biomarker qualification projects. The biomarker qualification program (BQP) has had broad application with accepted projects focusing on a variety of biomarker types, categories, and measures

	Total	507 Transition Plan[1]	Stage 1: LOI	Stage 2: QP	Stage 3: FQP
Total	61 (4 wd)	5 (1 wd)	30 (1 wd)	18 (2 wd)	8
Biomarker category					
Diagnostic	13 (1 wd)		8	4 (1 wd)	1
Monitoring[2]	5		4		1
PD Response[2]	12 (1 wd)	2	4	6 (1 wd)	
Prognostic	12		8	2	2
Safety	18 (2 wd)	3 (1 wd)	5 (1 wd)	6	4
Susceptibility/Risk	2		2		
Measure of					
Disease or Condition	30 (1 wd)		19	7 (1 wd)	4
Drug Response/Effect of Exposure	30 (3 wd)	5 (1 wd)	10 (1 wd)	11 (1 wd)	4
Not Specified	1		1		
Biomarker type					
Composite	1		1		
Histologic	4	1	2	1	
Molecular	28 (1 wd)	2 (1 wd)	11	8	7
Physiologic Characteristics	4 (1 wd)	1	3 (1 wd)		
Radiologic/Imaging	24 (2 wd)	1	13	9 (2 wd)	1
Surrogate endpoint					
Yes	5	1		4	
No	56 (4 wd)	4 (1 wd)	30 (1 wd)	14 (2 wd)	8

[1]507 Transition is not a formal phase of the DDT qualification process established under Section 507 of the 21st Century Cures Act; it serves as a transition phase for projects initiated under the legacy process for biomarker qualification (pre-2017).

[2]One project is categorized as both a PD Response and a Monitoring biomarker and is included in the total projects for each category. LOI, Letter of Intent; QP, Qualification Plan; FQP, Full Qualification Package; wd, Withdrawn/Rescinded; PD, pharmacodynamic.

39%) were most frequently explored. Projects were evenly split between biomarkers intended to measure a disease or condition (n=30/61, 49%) and those intended to measure drug response/effect of exposure (n=30/61, 49%), with one project uncategorized. While about half of accepted projects included biomarkers considered to be measures of drug response/effect of exposure, only a small fraction included biomarkers intended to be used as surrogate endpoints (5/61, 8%).

Submission Stage and Status

About half of all accepted projects (30/61, 49%) have not progressed past the initial LOI stage of the BQP process, and four projects were withdrawn or rescinded after acceptance. Many projects that remain at the LOI stage and have not been withdrawn or rescinded are likely still developing a qualification plan. Only eight biomarkers were qualified through the program, and seven of these were qualified before the *21st Century Cures Act* was enacted in 2016 under the FDA's legacy biomarker qualification process (Fig. 2).

The most recent qualification was granted in 2018. Among the biomarkers that achieved qualification, most were safety biomarkers (4/8, 50%) and predominantly used molecular methods for assessment (7/8, 88%). Importantly, no biomarker projects that included a surrogate endpoint have reached qualification; however, 4/5 of these projects submitted a qualification plan, of which 3 were accepted by the FDA.

Submission Timelines

Median LOI and QP review times exceeded the time frames specified in the November 2020 Final FDA guidance document titled "Qualification Process for Drug Development Tools"[1] [7]. Among the 43 projects with LOI submission and determination dates reported, LOI reviews took a median of 6 months—twice as long as the 3-month target time frame outlined in the guidance. Notably, most (31/43, 72.1%) of these projects were accepted into the BQP pre-final guidance. However, among the 12 projects submitted since the finalization of this guidance, LOI reviews have continued to exceed expected timelines, taking a median of

[1] The FDA guidance document specifies reviews are, "The time taken to review a submission once FDA has deemed it reviewable and a memorandum notifying the requestor of receipt of a reviewable submission has been sent to the requestor. For LOI, QP, and FQP submissions, the time frames are targeted to be completed within 3, 6, and 10 months, respectively, from the date on the reviewable memorandum".

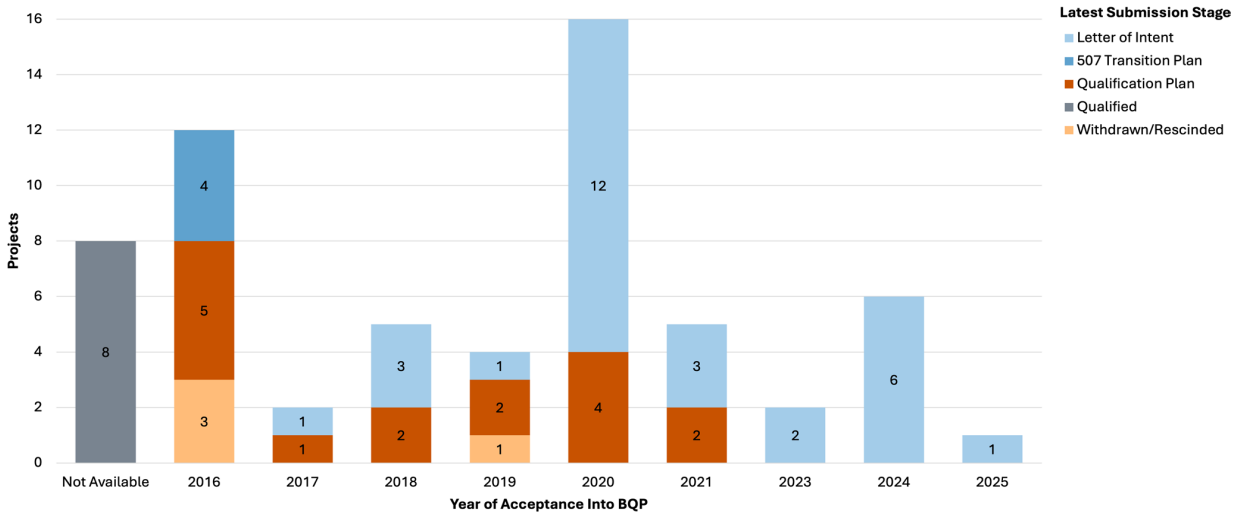


Fig. 2 Latest submission stage by year of acceptance into the biomarker qualification program. Note: "Not Available" includes qualified biomarkers for which acceptance and submission dates were not reported in the database. BQP, Biomarker Qualification Program

13.4 months (compared to 5.1 months for the 31 projects submitted pre-guidance). For the 13 projects with available QP submission and determination dates, 10 were submitted post-final guidance. Overall, QP reviews took a median of 14 months—7 months longer than the guidance-specified time frame. For post-guidance submissions, QP reviews took a median of 11.9 months.

QP development time was also calculated for the 16 projects with LOI determination and QP submission dates available. For these projects, it took a median of 32 months (2.7 years) to develop and submit a qualification plan. QP development timelines varied depending on the project characteristics. PD response biomarkers ($n=6$) and biomarkers assessing drug response/effect of exposure ($n=11$) had longer QP development timelines, with a median of 38 months (3.2 years) from LOI acceptance to QP submission, compared to other biomarker categories and measures (Fig. 3).

Projects across the various biomarker types (molecular, radiologic/imaging, and histologic) had similar QP development times. Qualification plans for surrogate endpoints ($n=4$) took the longest to develop overall, with a median of 47 months (3.9 years), reflecting the extensive requirements to validate a novel surrogate endpoint.

Discussion

The formalization of the DDT qualification process, including the process for qualifying biomarkers under the BQP, created a transparent, structured process for collaboration and information exchange to optimize the development and validation of novel biomarkers. However, in the eight years since the *21st Century Cures Act* formalized this framework, the program's overall impact appears limited. Only eight biomarker projects have reached full qualification, and all were reviewed under the legacy qualification process before the Section 507 process was established by the *21st Century Cures Act*. For ongoing projects, reviews often exceed expected timelines, and about half have not progressed beyond initial LOI acceptance. These trends have continued for the 12 projects submitted in the five years since the November 2020 final guidance on the qualification framework; however, additional time may be needed to assess whether the guidance has an impact on QP development times, or if these trends continue.

The BQP appears to be more impactful for certain biomarker categories, particularly safety biomarkers, which account for roughly one-third of accepted projects and four of the eight qualified biomarkers. In contrast, despite stakeholder interest in developing novel biomarkers to measure treatment efficacy, the program has seen very limited use for biomarkers intended as surrogate endpoints. It is important to understand what drives the lack of engagement and whether the current framework is ill-suited for qualifying these more complex biomarkers. Surrogate endpoint projects advance more slowly, with QP development requiring nearly 4 years–16 months longer than other biomarkers. The BQP lacks a funding mechanism, and there is limited guidance on opportunities for iterative FDA interaction throughout the process. Many factors can lead to delays or a failure to reach qualification, and targeted engagement and collaboration can help to efficiently plan, navigate challenges, and advance surrogate endpoint qualification. An optimized

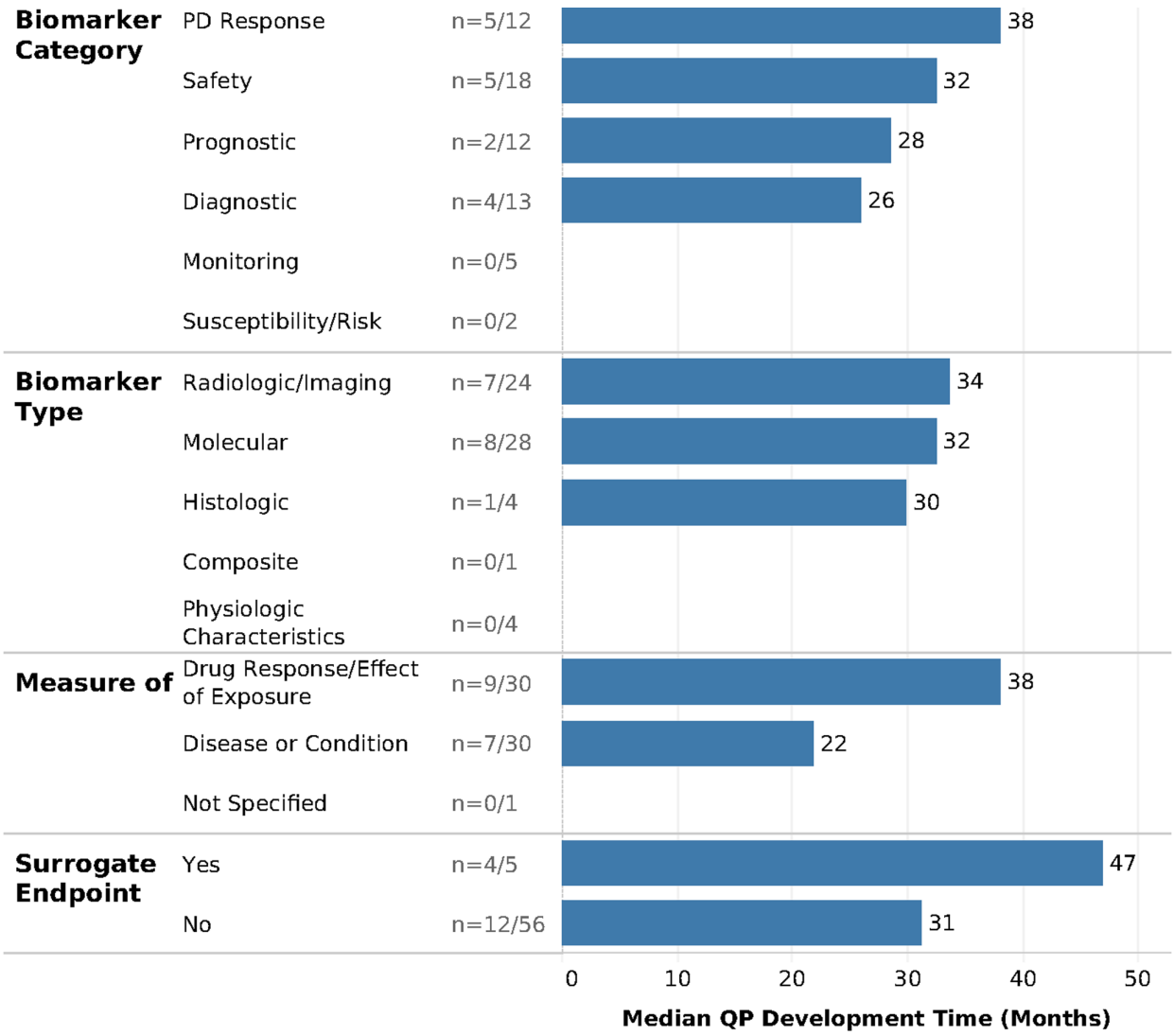


Fig. 3 Qualification plan development time by characteristics of biomarker qualification projects. Qualification plan development timelines varied across projects depending on the biomarker's characteristics, with surrogate endpoint biomarkers requiring 16 months longer for QP development than other biomarker projects. PD, pharmacodynamic; QP, Qualification Plan

program could provide timely advice to help researchers identify and address potential challenges and, if they are unable to be overcome, facilitate more efficient withdrawals from the qualification program.

Given the promise of surrogate endpoints to expedite treatment evaluations—and the extensive evidence required for them to be deemed reliable for regulatory decision-making—targeted enhancements and dedicated funding for the BQP, potentially supported by allocating PDUFA resources to divisions participating in the review of a given qualification plan, could enable more timely, structured engagement with specific FDA therapeutic area experts. Additional resources to support the divisions overseeing biomarker qualification should be accompanied by expected timelines for interactions such as meeting requests and conduct, feedback deadlines, and review periods. Such an approach would help embed BQP review into the established workflow associated with traditional user-fee-supported application review.

The Rare Disease Endpoint Advancement (RDEA) Pilot program, initiated under PDUFA VII, demonstrates a model for supporting the development of novel endpoints. However, RDEA is limited to endpoints that are associated with an active investigational new drug application (IND) [8].

Unlike the BQP, participating in RDEA may not enable qualification of the endpoint for broader use in any development program within a specific COU. While this approach may be sufficient for enabling endpoint development for rare diseases, it is less suited for endpoints with broader applicability, such as those relevant across multiple oncology development programs (e.g., circulating tumor DNA [ctDNA], minimal residual disease [MRD], pathologic complete response [pCR]), which typically require pooled data from multiple studies and sponsors [9–15]. In these instances where the biomarker could be applied to numerous future development programs, encouraging development and validation across multiple stakeholders and sponsors would be important to not limit use of the biomarker to a single program or sponsor's portfolio. While the BQP has been suitable for developing certain types of biomarkers (e.g., safety biomarkers), for novel endpoints and response biomarkers, particularly surrogate endpoints, which generally face a higher evidentiary bar for regulatory acceptance, funding and additional opportunities for engagement could provide the targeted and collaborative framework necessary to efficiently advance their development.

Conclusion

Enhancements to the existing BQP to support the development of novel biomarkers and endpoints for assessing treatment response not tied to a single IND or development program could help better prioritize and advance collaborative efforts to validate these tools. A more flexible framework that enables collaborative, pre-competitive evidence generation and sustained regulatory engagement would be needed for such biomarkers to achieve validation. Building on the phased structure of the BQP, the framework could incorporate iterative submissions and feedback touchpoints on data and validation plans, while offering enhanced opportunities for engagement between FDA and stakeholders to prioritize high-impact measures and guide their advancement. Creating this type of pathway could unlock the potential of novel response biomarkers to expedite drug development and improve patient access to effective therapies.

Author's contribution G.C. conception and design, data acquisition, analysis and interpretation of data, drafting and revisions, final approval. J.D.A. conception and design, drafting and revisions, final approval. H.S.A. conception and design, drafting and revisions, final approval. B.N. conception and design, drafting and revisions, final approval. M.D.S. conception and design, drafting and revisions, final approval.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker qualification: toward a multiple stakeholder framework for biomarker development, regulatory acceptance, and utilization. Clin Pharmacol Ther. 2015;98(1):34–46. https://doi.org/10.1002/cpt.136.
2. Lavezzari G, Womack A. Industry perspectives on biomarker qualification. Clin Pharmacol Ther. 2016;99(2):208–13. https://doi.org/10.1002/cpt.264.
3. U.S. Food and Drug Administration. Drug Development Tool (DDT) Qualification Programs: Biomarker Qualification Program. Accessed June 30, 2025. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program
4. Shahzad M, Stern AD. Participants in the FDA's biomarker qualification program. Clin Pharmacol Ther. 2025;118(1):80–4. https://doi.org/10.1002/cpt.3661.
5. Yang F, Bousnina I, Madej A, Kalamegham R. Has FDA's drug development tools qualification program improved drug development? Ther Innov Regul Sci. 2025;59(4):871–81. https://doi.org/10.1007/s43441-025-00790-2.
6. U.S. Food and Drug Administration. CDER & CBER Drug Development Tool Qualification Project Search. Accessed July 1, 2025. https://force-dsc.my.site.com/ddt/s/
7. U.S. Food and Drug Administration. Guidance for Industry and FDA Staff: Qualification Process for Drug Development Tools. Published online November 2020. Accessed June 1, 2025. https://www.fda.gov/media/133511/download
8. U.S. Food and Drug Administration. Rare Disease Endpoint Advancement (RDEA) Pilot Program. January 3, 2024. Accessed July 15, 2025. https://www.fda.gov/drugs/development-resources/rare-disease-endpoint-advancement-pilot-program#Goals%20of%20the%20RDEA%20Pilot%20Program
9. U.S. Food and Drug Administration. Guidance for Industry: Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment. Published online January 2020. Accessed July 31, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hematologic-malignancies-regulatory-considerations-use-minimal-residual-disease-development-drug-and
10. U.S. Food and Drug Administration. Guidance for Industry: Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support

Accelerated Approval. Published online July 2020. Accessed July 31, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pathological-complete-response-neoadjuvant-treatment-high-risk-early-stage-breast-cancer-use
11. U.S. Food and Drug Administration. Guidance for Industry: Use of Circulating Tumor Deoxyribonucleic Acid for Early-Stage Solid Tumor Drug Development. Published online November 2024. Accessed July 31, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-circulating-tumor-deoxyribonucleic-acid-early-stage-solid-tumor-drug-development-guidance
12. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384(9938):164–72. https://doi.org/10.1016/S0140-6736(13)62422-8.
13. Circulating Tumor DNA in Development of Therapies for Cancer: An Evidentiary Roadmap to an Early Endpoint for Regulatory Decision-Making. Published online July 2022. Accessed July 24, 2025. https://friendsofcancerresearch.org/wp-content/uploads/Circulating_Tumor_DNA_in_Development_of_Therapies_for_Cancer-Evidentiary_Roadmap.pdf
14. Landgren O, Prior TJ, Masterson T, et al. EVIDENCE meta-analysis: evaluating minimal residual disease as an intermediate clinical end point for multiple myeloma. Blood. 2024;144(4):359–67. https://doi.org/10.1182/blood.2024024371.
15. Exploring the Use of Circulating Tumor DNA As A Monitoring Tool for Drug Development. Published online November 2018. Accessed July 24, 2025. https://friendsofcancerresearch.org/wp-content/uploads/Circulating-Tumor-DNA.pdf

Expanding Access to Cell and Gene Therapies

FRIENDS of CANCER RESEARCH

PARKER INSTITUTE FOR CANCER IMMUNOTHERAPY

Regulatory and Manufacturing Pathways to Expand Access to Genetically Modified Cell-Based Therapies

White Paper | 2025

Table of Contents

Contributors

Mehrshid Alai-Safar – Gilead Sciences

Erin Bassett - BioCanRx

Jacqueline Bussolari – Johnson & Johnson Innovative Medicine

Ben Beneski – Allogene Therapeutics

Julio Delgado – University of Barcelona

Amanda Dinofia – Children's Hospital of Philadelphia

Boro Dropulic – Caring Cross

Steven Feldman – Stanford University

Holly Fernandez Lynch – Perelman School of Medicine at the University of Pennsylvania

Judy Fitzgerald – Patient Advocate

Lee Fleisher – Rubrum Advising

Patrick Hanley – Children's National Hospital

Kristen Hege –Independent Board of Directors Member

Natasha Kekre – The Ottawa Hospital

Seraphin Kuate – Bristol Myers Squibb

Bruce Levine - Perelman School of Medicine at the University of Pennsylvania

Deepu Madduri – Johnson & Johnson

Crystal Mackall – Stanford University

David Mitchell – CAR T Patient and Patients For Affordable Drugs

Dattesh Suthar – Novartis

Chris White – Patient Advocate

Friends of Cancer Research and Parker Institute for Cancer Immunotherapy

Jeff Allen – Friends of Cancer Research

Chris Cabanski – Parker Institute for Cancer Immunotherapy

Grace Collins – Friends of Cancer Research

John Connolly – Parker Institute for Cancer Immunotherapy

Ute Dugan – Parker Institute for Cancer Immunotherapy

Bernat Navarro-Serer – Friends of Cancer Research

Mark Stewart – Friends of Cancer Research

Enjun Yang – Parker Institute for Cancer Immunotherapy

The views expressed in this white paper do not necessarily reflect the views of all individuals or organizations listed. Participation in the development process does not imply endorsement of specific recommendations or conclusions.

Executive Summary

Genetically modified cell-based therapies, including chimeric antigen receptor (CAR) T-cell and T-cell receptor (TCR)-based approaches, are reshaping possibilities in the treatment of cancers and other complex diseases. Despite their potential, these therapies sometimes struggle to advance beyond early clinical trials—particularly for rare diseases or small patient populations, where traditional models for development, manufacturing, and reimbursement may not be conducive. Further compounding these challenges, fewer than one-quarter of relapsed or refractory hematologic malignancy patients eligible for these therapies receive them, often due to the real or perceived complexity of their use.

This white paper outlines potential regulatory, manufacturing, and cost recovery strategies to address the barriers that prevent promising therapies from reaching patients. The proposals are intended to inform future policy discussions, highlight areas for regulatory clarity, and identify operational solutions to support sustained therapy development. While exploratory in nature, the concepts aim at synergies between scientific rigor, operational feasibility, and patient need.

Key Focus Areas

1. Regulatory Engagement and Flexibility: The paper proposes clarifying and structuring the application of existing regulatory flexibilities—particularly in small populations, where traditional evidentiary expectations from large numbers of patients may not be practical. This includes aligning Chemistry, Manufacturing, and Controls (CMC) expectations with phase-appropriate standards and using early regulatory engagement to support risk-based development. These flexibilities could build on existing programs like the Regenerative Medicine Advanced Therapy (RMAT) and Breakthrough Therapy Designation (BTD) without reducing FDA's statutory approval standards.
2. Manufacturing Adaptability: Scalable access to genetically modified cell-based therapies will depend on flexible manufacturing ecosystems. The paper explores frameworks to support comparability, quality oversight, and site certification, along with mechanisms for implementing iterative process improvements without triggering full regulatory reassessment.
3. Sustainable Pre-Market Access: For therapies that show early clinical promise but face financial and commercial barriers due to the exceptionally small size of the relevant patient population, structured cost recovery and pre-approval access mechanisms could support continued development. Potential strategies include public-private partnerships, supplier collaboration, and grant-based funding. These concepts are not intended to replace traditional reimbursement pathways but may serve as transitional tools in select high-need settings.

The goal of this white paper is to identify actionable solutions that can help promising genetically modified cell-based therapies move from early development to clinical use—especially in areas of high unmet need. Continued dialogue with regulators, payors,

developers, and patient advocates will be essential to refining these ideas and ensuring that future pathways remain responsive, responsible, and focused on improving patient outcomes.

Introduction

Genetically modified cell-based therapies, including chimeric antigen receptor (CAR) T-cell and T-cell receptor (TCR)-based approaches, are beginning to transform treatment paradigms for complex diseases such as cancer. These therapies hold great potential for personalized medicine by targeting underlying genetic or cellular causes of disease. However, despite their promise, several barriers remain that prevent some of these therapies from transitioning from clinical trials to sustained patient access. These challenges are particularly acute for rare cancers and other small populations, where uncertainties in regulatory flexibilities and high costs associated with current manufacturing requirements create hurdles to sustainable development, often limiting commercial interest. At the same time, the specialized infrastructure required for production may be difficult to scale efficiently under traditional manufacturing models, underscoring the need for more adaptable solutions.

As a result, many promising therapies stall after early clinical development, leaving patients with few or no treatment options. Without clear regulatory pathways and cost-effective manufacturing solutions, these therapies may remain in limbo, lacking a viable pathway for continued development and access. This challenge is becoming more common as advances in cancer biology and therapeutic technology make it increasingly feasible to develop highly targeted cell-based therapies for narrowly defined patient populations—including rare adult and pediatric cancers—where traditional development and commercialization models may not be viable.

To address these challenges, a structured approach is needed that balances regulatory oversight and development of evidence to demonstrate safety and effectiveness with operational feasibility and sustainable reimbursement and/or cost recovery. This white paper explores solutions addressing several barriers that hinder genetically modified cell-based therapies from advancing beyond early-phase development, focusing on three critical areas:

1. Regulatory Uncertainty: While existing regulatory pathways offer some flexibility in demonstrating safety and efficacy, particularly for small patient populations with high unmet medical needs, there is no structured framework that defines when and how these flexibilities should—and should not—be applied to Chemistry, Manufacturing, and Controls (CMC) and manufacturing requirements. Tailored evidentiary requirements, including stage- and context-specific (e.g., fit-for-purpose) Good Manufacturing Practice (GMP) requirements, may be accepted on a case-by-case basis depending on the development program. Without transparency around how flexibilities have been applied in past scenarios, developers face uncertainty when

trying to align their development plans with regulatory expectations.[1] Establishing a more predictable framework for development-stage appropriate regulatory flexibilities, without compromising demonstrated product safety, efficacy, and quality, could enhance clarity, reduce inefficiencies, and foster greater alignment between developers and regulators.

2. Manufacturing Feasibility: Current regulatory requirements for product quality, safety, and site GMP compliance are often designed for large-scale commercial manufacturing, which can pose challenges for low-throughput production models. These challenges are especially true when therapies are developed or produced outside of traditional commercial settings. Iterative updates to manufacturing processes, such as adopting new technologies or refining production platforms, can also introduce regulatory complexity, increasing uncertainty for developers. These challenges are particularly relevant for decentralized manufacturing models, where maintaining product consistency and regulatory compliance across multiple sites adds another layer of complexity. While existing tools like pre-approved comparability protocols can help facilitate process changes, further clarity or guidelines on the application of this framework to distributed manufacturing for genetically modified cell-based therapies would be valuable. Without clearer pathways to support implementation of different manufacturing models and manufacturing improvements, developers may struggle to enhance turnaround times, reduce manufacturing costs, and expand patient access.
3. Reimbursement Barriers: A predictable reimbursement pathway is essential to ensuring long-term patient access to approved engineered cell-based therapies. However, even before approval, investigational genetically modified cell-based therapies, particularly those targeting rare diseases with limited commercial viability, often face financial barriers during development, as they typically fall outside the scope of traditional payor coverage. While the FDA approval process enables entry into standard reimbursement systems, clearer pathways to support development-stage access may be needed. Exploring pre-approval cost recovery strategies, such as limited, regulated mechanisms or public-private support models, may help sustain access in select, high-need cases while additional evidence is generated.

This white paper explores policy solutions across these three interdependent areas to support broader access to genetically modified cell-based therapies. While particularly relevant for manufacturing CAR T-cell treatments for rare cancers and small patient populations, many of the proposed manufacturing strategies could have broader applications across the landscape of genetically modified cell-based therapies, such as TCR-based approaches and cell and gene therapies for rare, non-malignant diseases. These solutions also aim to strengthen

national biomanufacturing infrastructure by promoting more resilient and distributed models that enhance domestic and local preparedness.

By aligning regulatory flexibility, adaptable manufacturing approaches, and predictable reimbursement models, the goal is to support access to therapies that might otherwise remain out of reach. This effort seeks to balance scientific rigor with operational feasibility, ensuring timely access to innovative, safe, and effective treatments while maintaining appropriate regulatory oversight. Importantly, this white paper does not propose lowering regulatory standards. Rather, it emphasizes applying existing standards in a structured, risk-based, and context-appropriate way. Patient safety, product quality, and regulatory integrity remain central to all proposals outlined herein. These proposals would apply exclusively to genetically modified cell-based therapies regulated under Section 351 of the *Public Health Service Act* and developed under active Investigational New Drug (IND) applications. The aim is to thoughtfully apply existing regulatory tools to improve access in high-need settings without compromising safety or efficacy.

Scope of Application: Illustrative Development Scenarios

Representative scenarios can help illustrate where the proposed solutions might have the greatest impact. While the overarching goal is to expand patient access to promising genetically modified cell-based therapies for rare diseases, the path to achieving this can vary widely depending on the nature of the sponsor, maturity of the clinical program, and anticipated commercial potential.

This white paper focuses on scenarios in which a therapy has demonstrated preliminary evidence of both safety and efficacy in early-phase studies but faces obstacles to initiating or completing a registrational trial due to limited commercial incentives, insufficient funding, or regulatory uncertainty. The proposals are intended to enable continued development by establishing regulatory and financial frameworks that support pivotal trial execution and approval.

While many of these challenges are particularly acute for programs led by research institutions or public-sector developers, the intent is not to create a framework limited to any one type of organization, but rather to address scenarios in which therapies with demonstrated potential face barriers due to scale, feasibility, or financial constraints. By shifting the economic and regulatory calculus, these models could create viable opportunities to advance these therapies. Several illustrative development scenarios are outlined below:

- Therapy with early-phase data in a rare disease: A genetically modified cell-based therapy developed and tested in a single-center, Phase 1 study in a rare disease population with no existing therapies. The therapy has demonstrated acceptable safety and preliminary efficacy, has a selected dose or dose range, and has received

a designation such as Breakthrough Therapy (BTD) or Regenerative Medicine Advanced Therapy (RMAT), acknowledging its potential clinical value. However, the pathway to a multi-center, registrational trial is unclear due to limited commercial interest and financial or regulatory constraints.

- Development of a therapy for a niche indication: A developer identifies a therapeutic opportunity in a small patient population that may not be commercially viable under current models. Regulatory flexibility and cost-sharing mechanisms—such as limited pre-approval coverage or shared public-private funding—could make development more feasible and support long-term access.
- Optimization of manufacturing for an approved therapy: An approved therapy could benefit from more efficient manufacturing processes. Regulatory processes that allow streamlined comparability between manufacturing processes without requiring a full new clinical development program could enable greater scalability and cost-effectiveness, ultimately improving patient access.

These scenarios are not exhaustive but are intended to reflect the range of programs that could benefit from targeted regulatory and financial innovation. The recommendations that follow aim to be broadly applicable across these settings while remaining grounded in operational feasibility and regulatory rigor.

Regulatory Pathways to FDA Approval for Genetically Modified Cell-Based Therapies in Small Patient Populations

Regulatory frameworks currently allow for flexibility in the development and approval of genetically modified cell-based therapies, particularly when supported by strong biological rationale and early clinical evidence. These flexibilities, such as use of surrogate endpoints, acceptance of single-arm trial data, and tailored post-approval commitments, are available through existing mechanisms like accelerated approval, INTERACT meetings, the CMC Development and Readiness Pilot (CDRP) Program, and the RMAT or Breakthrough designation.[2] They are especially relevant when traditional development models are infeasible due to factors such as small patient populations, disease severity, or lack of alternative therapies.

However, the how and when regulatory flexibility could extend to CMC requirements remains less well defined. While FDA has tools to support modified manufacturing approaches, such as risk-based GMP implementation or comparability protocols, these are often applied on a case-by-case basis, with limited transparency or precedent.[3–6] This lack of clarity can hinder planning, particularly for therapies developed in low-throughput, decentralized, or academic settings. A more structured and predictable fit-for-purpose approach to CMC flexibility could

reduce inefficiencies, support risk-based oversight, and ultimately improve patient access without compromising quality or safety.

To ensure such approaches remain appropriately scoped, it is important to outline circumstances where flexibility may be warranted. The following illustrative factors, when considered in combination, could help define appropriate use of regulatory flexibilities:

- A rare disease or narrowly defined patient subset, potentially affecting a very small number of patients annually.
- Lack of existing approved therapies and a serious or life-threatening condition.
- Preliminary clinical evidence suggesting meaningful clinical benefit or potential to address an unmet medical need.
- A therapy that has received a designation such as RMAT or BTD, reflecting compelling biological rationale and early data.

Likewise, clear boundaries should be defined for when flexibility would not be appropriate. Providing examples of acceptable evidence and fit-for-purpose manufacturing strategies would enable developers and regulators to align on a fit-for-purpose, risk-based framework that maintains rigorous standards while accounting for practical constraints.

Under this framework, core quality and safety principles would remain intact. Developers and regulators could collaboratively define fit-for-purpose GMP expectations tailored to low-throughput or site-specific manufacturing models. To support this approach, FDA could consider issuing guidance to clarify fit-for-purpose, adaptable CMC requirements that may be acceptable for genetically modified cell-based therapies in rare or underserved populations. This would build on existing programs such as RMAT and BTD, while specifically addressing manufacturing and feasibility constraints that may prevent promising therapies from advancing.

For example, similar to how the accelerated approval framework accepts surrogate endpoints, a complementary approach could define when fit-for-purpose manufacturing standards may be used. This might include cases where therapies are developed in autologous or low-volume settings, or where delays in production or distribution prevent timely access for patients.

By providing clearer expectations, such a framework could improve predictability for developers and payors while maintaining rigorous oversight. It would not be a prerequisite for regulatory flexibility but could serve as a tool to streamline engagement, align stakeholders, and support development and access in high-need settings.

Together, these strategies can support a more predictable, risk-based regulatory pathway for genetically modified cell-based therapies in small patient populations, helping to bridge the gap between early clinical promise and sustained patient access while allowing CMC

requirements to be more appropriately tailored to benefit-risk considerations that support timely availability. .

Manufacturing Models to Support Scalable and Sustainable Genetically Modified Cell-Based Therapy Production

A major barrier to sustained patient access is the absence of a flexible manufacturing ecosystem that can support a range of production models—particularly those tailored for small patient populations. Making genetically modified cell-based therapies available for patients requires manufacturing models that balance regulatory oversight and quality standards with operational feasibility.[7,8] Traditional large-scale commercial manufacturing requirements present challenges for autologous cell-based therapies, especially when production may occur at low-throughput or in decentralized, and point-of-care (POC) settings.[9–13] A more structured framework that supports risk-based, fit-for-purpose manufacturing approaches could help ensure product consistency, compliance with regulatory expectations, and scalability while allowing for process efficiencies. For example, fewer GMP requirements may be appropriate in early-stage development taking place in very limited populations than when a product advances further in clinical development toward more widespread use and full licensure, at which time somewhat more rigorous GMP might be required.[4,14] In addition to the above, strengthening manufacturing capacity for genetically modified cell-based therapies may also support national health security and align with broader efforts to bolster the national and local biomanufacturing infrastructure.

This section explores strategies for comparability and quality oversight, decentralized and mobile manufacturing solutions, and regulatory flexibility that could enable adaptive, scalable manufacturing.

Comparability and Quality Oversight

An operational consideration for decentralized, POC, and academic-based manufacturing is ensuring product consistency across multiple sites. To address this, standardized definitions and frameworks for comparability and quality oversight could be established, tailored to the therapeutic context and specific stage of therapy development.[15] This approach could help maintain product quality and consistency while allowing the flexibility necessary for feasibility, particularly in early-phase development and low-throughput production settings.

Early-Phase Development (e.g., Phase 1 multi-center academic trials):

- Comparability assessments may focus on foundational analytical measures (e.g., cell viability, sterility, and potency assays) to ensure product consistency across sites while providing predictability for developers and maintaining feasibility for small-scale production.

- A risk-based approach to identifying critical quality attributes (CQAs) could guide validation strategies, minimizing unnecessary data generation while still supporting regulatory expectations for investigational studies.
- Flexibility in demonstrating comparability—for example, allowing smaller, fit-for-purpose datasets in lieu of extensive at-scale comparability runs—would maintain quality standards while ensuring early-phase development remains feasible.

Late-Phase Considerations (e.g., submission package for rare disease genetically modified cell-based therapy):

- As therapies advance toward regulatory submission, the CMC package would need to evolve beyond early-phase expectations to include more structured data demonstrating batch-to-batch and site-to-site consistency.
- Validated analytical assays (e.g., flow cytometry for purity and identity, PCR for vector copy number) could serve as the basis for comparability assessments aligned with regulatory expectations.
- While split-batch comparability studies are a well-established standard, particularly for technology transfer, regulators could consider allowing alternative data sources in specific contexts. For example, small-scale representative runs or non-donor-matched material may be acceptable to support comparability, provided they are scientifically justified, validated, and supported by a risk-based assessment.

Post-Approval Modifications (e.g., process improvements that do not trigger classification as a new product):

- Regulatory flexibility could enable iterative manufacturing refinements without requiring extensive new clinical data, when supported by a risk-based assessment. This could include updates to automation, manufacturing platforms, or site-specific optimizations, provided quality parameters remain within pre-specified and validated bounds.
- A centralized Pharmaceutical Quality System (PQS) could serve as a mechanism for remotely governing multiple decentralized manufacturing sites, ensuring adherence to GMP while allowing for site-specific adjustments.

Decentralized Manufacturing Models for Genetically Modified Cell-based Therapies

A hub-and-spoke manufacturing model offers a structured approach to decentralization, enabling multiple sites ("spokes") to operate under the oversight of a lead-site ("hub"). This approach can help promote consistency, regulatory alignment, and quality control (QC) across multiple locations. Key components may include:

- New sites could undergo gap assessments, regulatory audits, and compliance agreements (e.g., MOUs or contractual frameworks) to ensure alignment with lead-site standards.
- A comprehensive tech transfer program could help ensure that standard operating procedures (SOP), batch records, personnel training, and equipment align with standardized expectations.
- Product release and QC testing could be centralized at the hub or designated testing facilities to promote consistency in release criteria, support regulatory compliance, and reduce variability across manufacturing locations.
- Virtual and in-person site reviews and third-party quality audits could support new site onboarding, compliance verification, and troubleshooting of manufacturing challenges.
- A comprehensive CMC package, potentially incorporating split-batch comparability studies, could help demonstrate consistency and support regulatory submissions.

This model has the potential to enhance scalability and regulatory predictability while supporting a more distributed domestic manufacturing infrastructure. However, it places significant operational responsibility on the hub, particularly for sustaining training and oversight at sites with intermittent production, which can be resource-intensive in low-volume settings.

Mobile Point-of-Care (POC) Manufacturing as an Emerging Solution

In addition to fixed decentralized sites, mobile point-of-care (POC) manufacturing units offer a promising solution for flexible, localized production of genetically modified cell-based therapies.[16] To be viable, these units would require clear regulatory pathways and alignment with GMP expectations. Key considerations include:

- Predefined GMP compliance standards, including sterility, product consistency, and quality control.
- Integration within an existing quality oversight framework, ensuring that mobile POC units align with lead-site regulatory governance.
- Defined regulatory expectations for including mobile POC units as part of the product license, ensuring they meet the same quality and safety standards as fixed GMP sites.

Mobile manufacturing or cell collection may be especially valuable in geographically dispersed regions or in settings requiring immediate cell collection and on-site processing. As these models continue to evolve, regulatory clarity and operational feasibility will be essential for broadening patient access safely.

Pre-Certification and Accreditation Models for Manufacturing Scalability

A potential mechanism to support decentralized manufacturing scalability is the pre-certification of manufacturing sites through an accreditation-based model. Pre-certification could:

- Establish clear regulatory expectations for non-commercial GMP facilities. For example, additional clarity on how phase-appropriate CGMPs apply in low-throughput or resource-constrained settings—such as appropriate documentation, environmental monitoring, or quality oversight expectations—could support more consistent implementation and reduce uncertainty.
- Enable pre-certified sites to function under centralized regulatory oversight within a hub-and-spoke manufacturing structure.
- Leverage existing accreditation frameworks, such as those from Foundation for the Accreditation of Cellular Therapies (FACT) or Association for the Advancement of Blood & Biotherapies (AABB), to ensure minimum infrastructure standards, validated analytical assays, and appropriate personnel training. These accreditation frameworks can help support elements of infrastructure readiness and could inform context-appropriate GMP expectations.

This approach could help build a distributed, domestically anchored manufacturing ecosystem, enhancing both scalability and national manufacturing readiness. By addressing gaps before product onboarding, pre-certified sites may be better positioned to support multi-site manufacturing efforts efficiently and compliantly while ensuring product quality and regulatory alignment.

Regulatory Flexibility for Manufacturing Process Evolution for Approved Products

To support scalable and sustainable manufacturing of genetically modified cell-based therapies, a regulatory approach could enable certain pre-defined, risk-based process modifications without requiring extensive additional regulatory reassessment. Through pre-defined process modifications plans, updates such as changes to automation technologies, site-specific optimizations, or adoption of new production platforms could be pre-defined, provided that CQAs and other relevant process controls remain within scientifically justified and pre-established parameters. Flexibility in process evolution should be accompanied by careful risk assessment and, when appropriate, additional supporting data or staged clinical evaluation to ensure that modifications do not introduce unintended variability or impact clinical outcomes.

Uncertainty around regulatory expectations and inconsistency of those expectations can delay or prevent critical refinements, such as optimizing production efficiency or reducing vein-to-vein time. Addressing these challenges through risk-based manufacturing flexibility could lower costs and broaden access without compromising product quality or safety.

Such a model could facilitate:

- Regulatory recognition of iterative improvements, allowing agreed-upon modifications across sites and product versions without triggering new clinical studies for each change.
- Structured comparability assessments, leveraging prior product knowledge to refine validation strategies for process changes, particularly in decentralized settings.
- Flexibility in oversight, enabling decentralized manufacturing sites to implement process refinements while maintaining product consistency and regulatory compliance.

Integrating predefined modification plans with comparability assessment strategies could allow developers to refine processes in real time. This approach aligns with broader risk-based strategies used in other regulatory contexts to streamline data requirements.[17,18]

Exploring Cost Recovery and Pre-Market Access Strategies for Genetically Modified Cell-Based Therapies

Enabling continued development of promising investigational genetically modified cell-based therapies—particularly for rare diseases with limited commercial viability—remains a critical challenge. While regulatory approval typically enables traditional reimbursement mechanisms, therapies in early stages often face financial barriers that limit evidence generation. In some cases, structured funding approaches may be needed to support participation in pre-approval studies where commercial investment or trial infrastructure is lacking. These strategies are not intended to replace the clinical trial process, but rather to supplement it in settings where resource limitations may otherwise halt development.

This section explores potential approaches to support financial sustainability during the investigational phases of development, particularly in cases where promising therapies for rare or underserved populations might otherwise stall due to limited commercial incentives. These proposals are intended to enable continued development and evidence generation in select, high-need cases. Structured access mechanisms, such as cost recovery, may offer a bridge where traditional funding models fall short. These ideas are exploratory and would require engagement with regulators, payors, patients, and other stakeholders to evaluate feasibility and ethical implementation.

Considerations for Pre-Market Cost Recovery and Access

In select cases where a therapy demonstrates strong early evidence of safety and potential clinical benefit—but lacks a clear commercial path—cost recovery approaches could help support continued development and patient access.

Options for exploration may include:

- Structured cost recovery mechanisms, consistent with existing FDA regulations, that allow limited reimbursement to offset manufacturing and delivery costs under defined conditions, such as through expanded access protocols with FDA authorization.
- Public-private partnerships or grant-based funding models to sustain access and continued evidence generation, especially for ultra-rare conditions or small populations with no alternative options.
- Supplier collaboration models, such as at-cost provision of critical materials, or academic-CMO partnerships or cost-sharing agreements to reduce the financial burden of continued production and delivery.

Any such mechanisms would need to be limited in scope, carefully defined, and clearly distinguished from traditional reimbursement for approved products. Congressional action would likely be needed to enable such mechanisms under fee for service Medicare and could be explored in the context of a targeted pilot program.

Transitioning to Traditional Coverage Pathways

Once a therapy receives regulatory approval, it qualifies traditional coverage frameworks under Centers for Medicare and Medicaid Services (CMS) and other insurers. Existing mechanisms—such as coverage determinations or clinical guidelines—would govern reimbursement and inform access decisions.

If a therapy is made available during the investigational phase through a structured cost recovery model, the developer could be expected to generate ongoing evidence to support regulatory approval and future coverage decisions. The evidence collected during this period could be critical in informing long-term coverage policies and ensuring a smooth transition to traditional reimbursement pathways following regulatory approval.

Opportunities for Future Dialogue

Meaningful discussion with CMS, private payors, and regulatory agencies will be critical to exploring these concepts further. Key questions include:

- Under what conditions—if any—might early access models be appropriate for therapies with limited commercial viability but high potential impact?

- How could such models be structured to ensure ethical safeguards, scientific rigor, and fiscal accountability?
- What mechanisms could support a transition from investigational access to traditional reimbursement without disrupting patient care?

Conclusion

Advancing the development and availability of genetically modified cell-based therapies, particularly for rare diseases, requires a coordinated approach that integrates regulatory flexibility, adaptable manufacturing models, and mechanisms to support evidence generation and associated access prior to approval. The proposals outlined in this white paper aim to address persistent barriers to development, helping ensure that promising therapies for patients do not stall.

Key considerations include:

1. Regulatory Engagement and Flexibility: Establishing a more structured regulatory pathway, within existing statutory approval standards, that aligns fit-for-purpose evidentiary requirements with the distinct challenges of developing therapies for small patient populations. This may include leveraging accelerated approval frameworks not only for clinical evidence, but also for fit-for-purpose manufacturing and CMC requirements, supported by early engagement with regulators.
2. Manufacturing Adaptability: Supporting decentralized and scalable production through comparability frameworks, pre-certification of GMP sites, and clearly defined mechanisms for implementing manufacturing improvements. Strengthening domestic and local infrastructure, particularly through distributed manufacturing and POC models, could also enhance national readiness.
3. Sustainable Pre-Market Access: Exploring structured cost recovery and early access strategies to support investigational therapies with compelling clinical promise but limited commercial viability. While exploratory in nature, these approaches may help facilitate continued development and evidence generation in high-need areas and inform future policy dialogue.

By aligning innovations across development, manufacturing, and access, the proposals in this white paper aim to create viable processes for delivering transformative genetically modified cell-based therapies to patients with limited treatment options. Continued dialogue with regulators, developers, payors, and patient advocates will be essential to refining these proposals and ensuring they remain grounded in both scientific rigor and patient need.

References

1. Janiaud P, Irony T, Russek-Cohen E, Goodman SN. U.S. Food and Drug Administration Reasoning in Approval Decisions When Efficacy Evidence Is Borderline, 2013–2018. *Ann Intern Med*. 2021;174(11):1603-1611. doi:10.7326/M21-2918
2. U.S. Food and Drug Administration. Chemistry, Manufacturing, and Controls Development and Readiness Pilot (CDRP) Program. U.S. FDA. November 14, 2024. Accessed April 20, 2025. https://www.fda.gov/drugs/pharmaceutical-quality-resources/chemistry-manufacturing-and-controls-development-and-readiness-pilot-cdrp-program
3. U.S. Food and Drug Administration. Considerations for the Development of Chimeric Antigen Receptor (CAR) T Cell Products. January 2024. Accessed April 28, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-development-chimeric-antigen-receptor-car-t-cell-products
4. U.S. Food and Drug Administration. Current Good Manufacturing Practice for Phase 1 Investigational Drugs. July 2008. Accessed April 28, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/current-good-manufacturing-practice-phase-1-investigational-drugs
5. U.S. Food and Drug Administration. Comparability Protocols for Postapproval Changes to the Chemistry, Manufacturing, and Controls Information in an NDA, ANDA, or BLA. October 2022. Accessed April 28, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/comparability-protocols-postapproval-changes-chemistry-manufacturing-and-controls-information-nda
6. U.S. Food and Drug Administration. Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs). Published online July 2020. Accessed April 28, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/chemistry-manufacturing-and-control-cmc-information-human-gene-therapy-investigational-new-drug
7. O'Connor DJ, Moss P, Wood M, et al. The Rare Therapies Launchpad: a pilot program for individualized medicines in the UK. *Nat Med*. Published online March 10, 2025. doi:10.1038/s41591-025-03547-4

8. U.S. Food and Drug Administration. FDA Rare Disease Innovation Hub. January 2025. Accessed April 28, 2025. https://www.fda.gov/industry/medical-products-rare-diseases-and-conditions/fda-rare-disease-innovation-hub

9. Elsallab M, Maus M V. Expanding access to CAR T cell therapies through local manufacturing. *Nat Biotechnol*. 2023;41(12):1698-1708. doi:10.1038/s41587-023-01981-8

10. Elsallab M, Maus M V. Charting the course for CAR T-cell manufacturing. *Blood Adv*. 2024;8(23):6131-6132. doi:10.1182/bloodadvances.2024014653

11. Harrison RP, Rafiq QA, Medcalf N. Centralised versus decentralised manufacturing and the delivery of healthcare products: A United Kingdom exemplar. *Cytotherapy*. 2018;20(6):873-890. doi:10.1016/j.jcyt.2018.05.003

12. Wang K, Tseng CY, Li Z, et al. A simulation-based comparison of centralized and point-of-care supply chain strategies for autologous cell therapy. *Cytotherapy*. 2023;25(12):1370-1379. doi:10.1016/j.jcyt.2023.08.007

13. The Human Medicines (Amendment) (Modular Manufacture and Point of Care) Regulations 2025.

14. U.S. Food and Drug Administration. INDs for Phase 2 and Phase 3 Studies Chemistry, Manufacturing, and Controls Information. July 2008. Accessed April 29, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/current-good-manufacturing-practice-phase-1-investigational-drugs

15. FDA Office of Pharmaceutical Quality. *Distributed Manufacturing: Considerations for Cell and Gene Therapies*.; 2022. Accessed April 20, 2025. https://www.fda.gov/media/173449/download

16. Harrison RP, Ruck S, Medcalf N, Rafiq QA. Decentralized manufacturing of cell and gene therapies: Overcoming challenges and identifying opportunities. *Cytotherapy*. 2017;19(10):1140-1151. doi:10.1016/j.jcyt.2017.07.005

17. Stewart MD, Keane A, Butterfield LH, et al. Accelerating the development of innovative cellular therapy products for the treatment of cancer. *Cytotherapy*. 2020;22(5):239-246. doi:10.1016/j.jcyt.2020.01.014

18. Stewart MD, Kalos M, Coutinho V, et al. Accelerating the development of genetically engineered cellular therapies: a framework for extrapolating data across related products. *Cytotherapy*. 2024;26(7):778-784. doi:10.1016/j.jcyt.2024.03.009

Journal for ImmunoTherapy of Cancer

Intentional heterogeneity in autologous cell-based gene therapies: strategic considerations for first-in-human trials

Christopher R Cabanski ,[1] EnJun Yang ,[1] Mark D Stewart ,[2] Jeff D Allen,[2] John E Connolly,[1] Ute Dugan,[1] Philip D Greenberg,[1,3,4,5] Crystal L Mackall,[6,7,8,9] Carl H June,[10,11,12] Alexander Marson,[13,14,15,16] Marcela V Maus ,[17,18] Antoni Ribas [19,20]

To cite: Cabanski CR, Yang EJ, Stewart MD, *et al*. Intentional heterogeneity in autologous cell-based gene therapies: strategic considerations for first-in-human trials. *Journal for ImmunoTherapy of Cancer* 2025;**13**:e011301. doi:10.1136/jitc-2024-011301

MVM and AR contributed equally.

Accepted 14 May 2025

For numbered affiliations see end of article.

Correspondence to
Dr Antoni Ribas;
aribas@mednet.ucla.edu

Dr Marcela V Maus;
mvmaus@mgh.harvard.edu

ABSTRACT

Cell-based gene therapies, including chimeric antigen receptor-T, T-cell receptor-T, and tumor-infiltrating lymphocyte therapies, have transformed the treatment landscape for certain cancers, yet their efficacy in solid tumors remains limited. Next-generation therapies aim to overcome biological barriers, enhance potency and safety, and streamline development timelines through innovative approaches. Recent advances in genome editing technologies have identified hundreds of gene edits that improve T-cell functionality in preclinical models. However, the limited direct translatability of these findings and the impracticality of testing each of the individual edits in a traditional clinical trial highlight the need for more efficient strategies.

This article provides an overview of genome-wide screens that identify gene knockouts and knock-ins to enhance T-cell function and the limitations with translating these results to human trials. Next, we propose a novel clinical trial design for testing multiple gene modifications simultaneously within a single T-cell infusion product. This approach would enable head-to-head evaluation of edits in an internally controlled setting, accelerating the identification of promising candidate edits. Key considerations for Chemistry, Manufacturing, and Controls, non-clinical evaluation, and clinical protocols are discussed, with an emphasis on patient safety and ethical transparency.

This framework is informed by insights shared at the "Unlocking Complex Cell-based Gene Therapies" workshop, held on May 6, 2024. Co-hosted by Friends of Cancer Research and the Parker Institute for Cancer Immunotherapy, the event brought together participants from academia, the US Food and Drug Administration, and patient advocacy groups. By fostering collaboration among these stakeholders, this innovative approach aims to accelerate the development of effective cell-based therapies for complex diseases.

INTRODUCTION

The adoptive cell transfer of autologous cellular therapies, including genetically engineered T-cell receptor (TCR-T) and chimeric antigen receptor (CAR-T) cells, as well as tumor-infiltrating lymphocyte (TIL) therapies, has emerged as a new form of cancer treatment with proven benefits in specific indications.[1–7] As of November 2024, the US Food and Drug Administration (FDA) has approved seven CAR-T therapies for hematologic malignancies, one TCR-T therapy for HLA-A2.1 positive synovial sarcoma, and one TIL therapy for unresectable or metastatic melanoma. With multiple ongoing trials evaluating TCR-T, CAR-T and TIL therapies, there is promise for broader applications, particularly against solid tumors. To achieve this, next-generation therapies must integrate innovative scientific strategies to enhance safety and efficacy, overcome biological constraints, and address operational challenges such as manufacturing and clinical development costs and timelines.

This article provides a narrative of the concepts discussed at a workshop titled "Unlocking Complex Cell-based Gene Therapies", co-hosted by Friends of Cancer Research and the Parker Institute for Cancer Immunotherapy on May 6, 2024, in Washington, DC, USA. At this event, academic investigators and clinicians, as well as FDA leaders and patient advocates, discussed frameworks and proposals for developing the next generation of cell-based gene therapies. Discussion panels addressed operational and biological challenges to advancing cell and gene therapies, novel clinical trial designs, approaches to increase the efficacy of cell therapies for solid tumors, and included patient insights and experiences with these complex therapies.

In this article, we provide an overview of in vitro and in vivo genome-wide screens for identifying candidate gene targets to enhance T-cell function and describe a candidate approach for evaluating multiple gene modifications within a single human

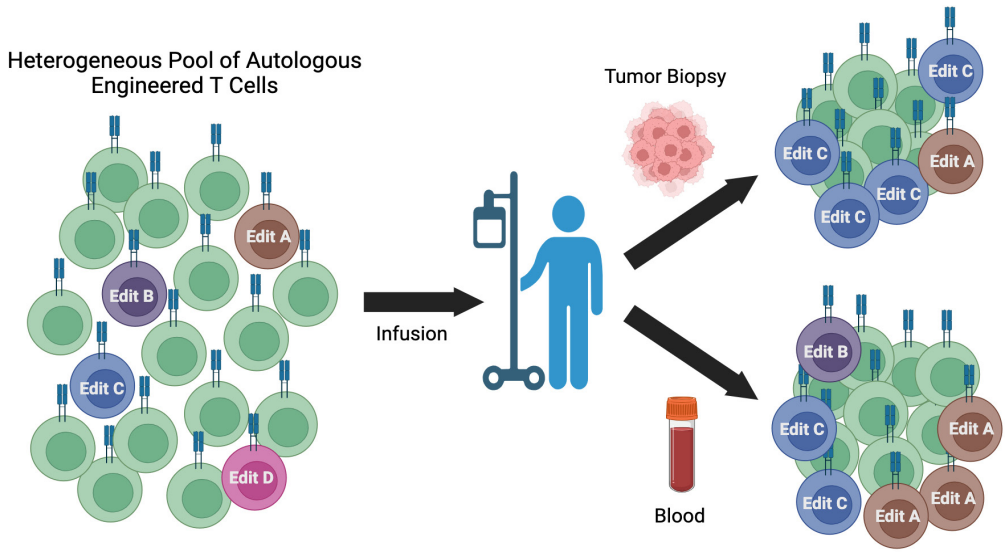


Figure 1 Testing a heterogeneous pool of genetically modified T cells. On the left, the infusion product is depicted, consisting of T cells uniformly modified to target one specific tumor-associated antigen. Distinct subpopulations within this infusion product would carry unique secondary gene edits (denoted as Edits A through D), each introduced to potentially enhance antitumor efficacy. Following infusion, analyses of tumor biopsies and blood samples would examine the persistence, expansion, function, and trafficking patterns of each genetically distinct subpopulation, as illustrated on the right. This would enable the assessment of how individual gene edits may influence the cells' in vivo behavior, such as Edits A and C appearing in both tumor and blood, Edit B localized exclusively in blood, and Edit D not prevailing in either. Patient safety and efficacy would be continuously monitored throughout the study. Created in BioRender. Yang (2025) https://BioRender.com/j69j443.

clinical trial. Subsequently, we discuss the Chemistry, Manufacturing, and Controls (CMC), non-clinical, and clinical considerations essential for designing such trials, with a focus on prioritizing patient safety.

TRANSLATING GENOME-WIDE SCREEN INSIGHTS INTO THERAPEUTIC ADVANCES

Any genetically modified autologous T-cell investigational product for therapy-resistant malignancies, including solid tumors, faces a myriad of biological challenges, such as a low peripheral antigenic stimulus, the need to traffic to and function in the immunosuppressive tumor microenvironment, and heterogeneous tumor composition.[5 7 8] Prevailing evidence from preclinical models demonstrates that the integration of potency enhancements into current platforms will be needed to deliver meaningful clinical benefits in patients with solid cancers. Fortunately, recent scientific and technological advances in cellular genome editing and manufacturing allow for the design of more advanced therapeutic products, including those that contain genetic modifications that offer the potential for increased efficacy. For instance, several in vitro and in vivo CRISPR screens have identified hundreds of genes that improve T-cell function when knocked out or overexpressed in preclinical models.[9–17] Other review articles have summarized the candidate gene modifications discovered in these studies.[18 19] Despite these advances, transitioning from preclinical models to human clinical trials remains a significant challenge.

Current genome-wide screens in preclinical models face several limitations. The mouse models employed use artificial systems that are unable to fully capture the complexities of human T-cell biology.[9 11 14 15] In immunocompetent murine models, T cells differ from human counterparts in their receptor repertoires, regulatory networks, and tumor microenvironment interactions, raising concerns about the direct translatability of identified targets.[20 21] Conversely, screens using human T cells in immunodeficient mice with human tumors lack key immunological structures, such as properly developed secondary lymphoid organs, that are critical for normal T-cell maturation, trafficking, and long-term persistence.[20–22] Furthermore, human tumor antigens and their expression patterns in these models may not fully recapitulate clinical settings, potentially skewing the assessment of T-cell efficacy and specificity.[20 23] These issues, combined with the broader shortcomings of preclinical cancer models in predicting clinical outcomes, lead to uncertainty regarding which genetic modifications are most likely to improve T-cell therapy and yield clinical benefits. Consequently, we currently lack a robust model for prioritizing candidate gene edits from these genome-wide screens for single-edit human clinical trials.

LEVERAGING INTENTIONAL CELL HETEROGENEITY IN CLINICAL TRIALS

There is an established regulatory framework for developing cell-based gene therapies. The FDA has released several guidance documents that contain recommendations and considerations related to product characterization, release testing criteria, and patient safety for cell-based gene therapies.[24,25] These regulatory guidance documents have enabled the clinical testing of several cell-based gene therapies in oncology, many of which incorporate a single genetic modification in addition to the antigen targeting receptor. In addition, a growing number of clinical trials are testing products that contain multiple gene edits.[26,27]

Despite preclinical screens identifying hundreds of candidate gene edits with the potential to enhance T-cell function, the combination of a lack of directly translatable preclinical models and the scarcity of candidate edits tested in human trials to date means that there is a significant barrier to clinical progress in the field. The resources required, particularly patient numbers and trial duration, present practical limitations to testing all potentially beneficial gene edits within a feasible time frame using current standard clinical frameworks. Furthermore, even if individual edits were evaluated in separate trials, interpreting the true impact of each modification across studies would be challenging.

Recognizing these limitations, future clinical trials could incorporate intentional heterogeneity, which we define as the introduction of variability through genetic modifications, by testing a diverse pool of engineered T cells. Testing multiple gene edits simultaneously within a single human clinical trial would enable us to observe whether any modifications yield consistent comparative benefits across a small cohort of patients. By enabling head-to-head comparison both within each patient and across patients, this design has the potential to expedite the identification of promising therapeutic enhancements and address the current challenge of limited resources for testing an expanding list of gene edit candidates. However, since no clinical trial of this kind has been conducted yet, patient safety and ethical considerations must remain priorities in its design and execution. This approach is visually represented in figure 1 and detailed below.

Uniform antigen specificity

The infusion product would consist of genetically modified autologous engineered T cells (eg, CAR-T or TCR-T) targeting a common antigen, with selection processes employed as needed to ensure that all infused cells retain this specificity. The CAR or TCR should already have established safety and initial clinical activity from prior clinical testing. By testing a CAR or TCR with established safety and baseline clinical activity, the infusion product would provide the potential for therapeutic benefit.

Subpopulations with secondary gene edits

All T cells would target a common antigen, while small subpopulations would contain a secondary, traceable gene edit that may include both loss-of-function and gain-of-function modifications. These secondary edits would differ across cell subpopulations, with each subpopulation containing a unique modification. Each individual gene edit should be supported by preclinical data, including molecular characterization to evaluate off-target effects and provide reasonable assurance of safety. Additionally, the edit should be hypothesized to enhance T-cell functionality, potentially improving antitumor activity after adoptive cell transfer. Although no single cell is intended to have multiple secondary edits, and statistically this would be highly unlikely, pooled gene engineering methods could inadvertently introduce multiple edits in an individual cell. A major challenge with current methods is that analyzing off-target effects for multiple gene edits is significantly more complex than for single gene edits. This complexity may require the development of specific lot-release criteria for the edited T-cell subpopulations. The remaining T cells, without additional functional modifications, would serve as the control, providing a baseline to measure the effectiveness of the edited cells.

The number of distinct secondary gene edits introduced into the infusion product should be contingent on the specific objectives of the study and constrained by the therapeutic target, patient population, manufacturing feasibility, and ability to appropriately characterize each edit for safety. Each of these secondary edits should be exclusive; that is, efforts should be taken to minimize single cells in the product containing more than one of these additional gene edits. This is a controlled way to keep as many variables fixed as possible and better understand the impact of each individual gene edit. In addition, these gene edits should be traceable, allowing for the determination of which cells contain which, if any, secondary gene edit.

Decisions on the composition of the cell subpopulations, specifically the prevalence of cells containing each secondary gene edit, should be guided by several considerations, including the number of candidate gene edits and available clinical and non-clinical data on both the gene edits and the CAR-T or TCR-T cell infusion product. Previous studies have shown that beneficial gene edits can be identified from initially rare cells in the CAR-T cell product,[28,29] especially if they become selectively enriched, suggesting that introducing secondary-edited cells at low prevalence is practical and sufficient for assessing dynamic changes in the composition of the T-cell pool within the patients' blood and tumor over time. While achieving precise ratios of each gene edit within the infusion product is not feasible with many of the current manufacturing approaches, the final product should undergo characterization to ascertain the proportional distribution of each edit at therapy initiation.

Manufacturing and guide RNA (gRNA) diversity

For products testing gene knockouts, the infusion product could be manufactured in bulk using a library of gRNAs in a single manufacturing run. The size of the gRNA library will depend on the intended number of secondary gene edits being compared. Transducing T cells at a low multiplicity of infection (MOI) would ensure that most cells receive at most one variable gRNA, enabling subsequent analysis of how individual gene edits affect T-cell performance post-infusion. This approach would result in three cell populations following transduction: (1) no gRNA, (2) exactly one gRNA, and (3) more than one gRNA. The distribution of these populations can be estimated using a Poisson model, and adjusting the MOI can increase the proportion of cells containing exactly one gRNA while attempting to minimize the proportion of cells with more than one gRNA.

Clinical trial design

The clinical trial could enroll a limited cohort of patients with advanced cancers that have progressed after completing treatments with curative intent. The primary objective should be assessing the safety and feasibility of infusing an intentionally heterogeneous pool of genetically modified autologous T cells. Secondary objectives could include a description of the dynamic changes in the composition of the T-cell pool within the patients' blood and tumor over time. The "beneficial" gene edits could be identified based on the proliferation and trafficking capabilities of their respective cell subpopulations in the blood and at the tumor site, as well as evidence of sustained function. Insights from this clinical trial would guide the selection of specific gene edits for further evaluation, paving the way for their testing and evaluation via traditional drug development pathways.

Well-established analytical approaches exist to track the persistence, expansion, and localization of individual subpopulations post-infusion. Flow cytometry is a reliable option for monitoring edits that result in altered surface marker expression. For gene edits that do not lead to readily detectable surface marker changes, alternative molecular approaches could be considered. Next-generation sequencing, quantitative PCR (qPCR), and other molecular assays can accurately identify and quantify genetically distinct T-cell subpopulations from peripheral blood and tumor biopsies. Longitudinal single-cell sequencing methods could also provide insights into the dynamics, persistence, and functional characteristics of each edited subpopulation over time, complementing flow cytometric assessments and ensuring comprehensive traceability across all introduced gene edits.

SAFETY CONSIDERATIONS FOR INTENTIONALLY HETEROGENEOUS CELL PRODUCTS

This innovative clinical trial design aims to prioritize safety. Several FDA guidance documents contain recommendations and considerations related to product characterization, release testing criteria, and patient safety for cell and gene therapies,[24 25] and most of these recommendations remain pertinent in this setting. However, the additional complexity introduced by a gene-edited, heterogeneous pool of T cells requires a critical reevaluation of appropriate safety measures. Specifically, infusion products that contain intentional heterogeneity may potentially increase the risk of the following safety events: (1) the possibility that an edit enhancing T-cell function could increase the frequency or severity of acute events such as cytokine release syndrome, and (2) the potential for edits promoting T-cell proliferation to initiate cellular transformation or a secondary T cell-derived malignancy.[30 31]

As sponsors work towards operationalizing intentionally heterogeneous designs, several important factors should be given careful thought to prioritize patient safety. These safety considerations fall into four principal categories: product design, CMC, non-clinical, and clinical. Each clinical trial may necessitate additional or fewer precautions, tailored to its specific context. A summary of these considerations is presented in table 1.

Table 1 Safety considerations to guide product and clinical trial designs that test an intentionally heterogeneous pool of autologous T cells

Category	Examples of safety considerations
General considerations for product design	► Selection of validated antigen target ► Endogenous TCR knockout ► Use of Immunogenic CAR or TCR constructs ► Incorporation of a safety switch
Considerations for Chemistry, Manufacturing, and Controls	► Adherence to standard analytical testing criteria ► Confirm the frequency of edited cells ► Transformation assay testing
Non-clinical considerations	► Characterization of gene edits ► Characterization of gRNA libraries ► Additional non-clinical testing considerations: – Evaluation of uncontrolled proliferation – Characterization of activation profile – Phenotypic characterization
Clinical considerations	► Study population ► Dose and dose schedules ► Staggered enrollment ► Dose-limiting toxicities and stopping rules ► Drug Safety Monitoring Board ► Assessment of product-related adverse events ► Biospecimen collection ► Long-term follow-up

CAR, chimeric antigen receptor; gRNA, guide RNA; TCR, T-cell receptor.

GENERAL CONSIDERATIONS FOR PRODUCT DESIGN

The design of the product is fundamental to ensuring the safety of an intentionally heterogeneous T-cell investigational product. Here, we elaborate on general considerations related to product design, which serve as examples to be considered when designing clinical trials with cell-based gene therapies.

Selection of validated antigen target

The product should incorporate a single targeting construct, either a CAR or TCR, that has previously demonstrated an acceptable safety profile and potentially an initial indication of benefit in clinical trials. This strategy leverages historical data, allowing the analysis and attribution of novel safety signals to focus on the secondary gene edits.

Endogenous TCR knockout

The product design could optionally include a fixed gRNA guide targeting the TCR-alpha constant chain gene within the vector in addition to the "variable" gRNA guide. This strategy may provide dual benefits: it could mitigate the risk of enhancing an auto-reactive TCR from the peripheral repertoire and allow for post-manufacturing selection of successfully gene-edited cells via magnetic separation of CD3-negative T cells. Knocking out the endogenous TCR does entail an additional genetic modification per cell, which could increase the risk of gene translocations and must be empirically tested and factored into the overall risk-benefit assessment. However, there is reproducible evidence that anti-self toxicities can be induced when infusing large quantities of TILs, including ex vivo activated T cells with self-antigen recognition or autologous T cells with transgenic TCR-T or CAR-T cells that lack endogenous TCR knockout,[1–3] and this risk would be eliminated by knockout of the endogenous TCR.

Use of immunogenic CAR or TCR constructs

Implementing a CAR or TCR containing murine sequences introduces a form of immunogenicity that may result in rejection of the therapy after several months. This built-in temporal limitation could act as a secondary safety mechanism, potentially allowing for the patient's immune system to reject the infusion product, particularly in solid tumor contexts. The downside is a possible limitation on the duration of engraftment, which could affect the analysis of gene edit effects, which more importantly may limit the benefit of long-term antitumor activity of the gene-engineered T-cell preparation. Therefore, the inclusion of this strategy should depend on the nature of the antigen and tumor type(s) being targeted.

Incorporation of a safety switch

Safety switches offer a means of inducing apoptosis in transduced cells if clinically necessary and could be considered in this context.[32] While various safety switches have been incorporated into investigational products, it is important to recognize that their efficiency might not be absolute and the clinical evidence supporting their utility remains limited. Additionally, in a scenario where a single gene edit leads to severe toxicity, activating a safety switch would eliminate transduced cells indiscriminately rather than selectively removing only the problematic subpopulation. Although this broad elimination is not ideal, patient safety considerations necessitate rapid action without waiting for the identification of specific edits associated with toxicity. Including a safety switch should therefore depend on careful evaluation of the antigen target, tumor type, and anticipated risk-benefit balance of eliminating the entire transduced cell population.

CONSIDERATIONS FOR CHEMISTRY, MANUFACTURING, AND CONTROLS

Adhering to established benchmarks for manufacturing and release testing is a fundamental aspect of ensuring the safety and efficacy of cell therapy products. These considerations encompass both standard and specialized protocols and assays designed to manufacture and validate the final product.

Adherence to standard analytical testing criteria

Sponsors should follow established CMC guidelines for evaluating the sterility, identity, purity, and activity of the investigational product. This may include performing flow cytometry, vector copy number, editing efficiency, identity, and potency assays as part of the product release testing criteria.

Confirm the frequency of edited cells

Rigorous testing is essential to confirm the frequency of edited cells in the final infusion product. This quantification is crucial for assessing the therapy's potential efficacy, ensuring batch-to-batch consistency, and understanding the precise composition of the product for patient safety and for evaluating changes post-administration. Although ideally the proportions of cells carrying each individual genetic modification would be similar, achieving precise or equal proportions across a large library of gRNAs may be practically infeasible for product release. Therefore, release criteria might instead focus on verifying the total number of edited cells rather than the proportion of each specific edit. Quantitative assessments such as flow cytometry, qPCR, or sequencing-based assays (eg, GUIDE-seq[33] or RhAmpSeq[34]) can accurately measure the presence and abundance of edited cells. Additionally, these methods can support targeted evaluations of potential off-target effects, ensuring specificity and safety within acceptable limits.

Transformation assay testing

The transformation assay is considered the gold standard for assessing the safety of CRISPR and base-edited T cells, as well as many CAR-T and TCR-T products that exhibit heterogeneity from non-targeted lentiviral integrations. This assay detects the potential for cytokine or antigen-independent proliferation among edited T cells.

For infusion products that contain intentional heterogeneity, the strategy for this assay may include testing at multiple MOIs, including exceeding the levels used in the final product. By validating the cell product against transformation at these higher MOIs, the risk of unexpected oncogenic transformation can be mitigated, particularly in the rare event that a T-cell acquires more than one variable gRNA. For illustrative purposes, the MOI might be set to transduce a minor fraction (eg, 1%) of the CAR-T or TCR-T cells with any single gRNA, thereby reducing the chance that any individual T cell would acquire multiple secondary gene edits.

NON-CLINICAL CONSIDERATIONS

A thorough characterization of the infusion product, including all gene edits, is essential for understanding the safety profile of the therapy. Here we detail considerations for characterizing the product.

Characterization of CAR or TCR construct

Previous clinical experience with the CAR or TCR construct should be described in the study protocol and Investigational New Drug (IND) application.

Characterization of gene edits

A comprehensive rationale for each gene edit, including empirical data supporting its safety profile, should be provided in the study protocol and IND application. This rationale should draw on relevant scientific literature as well as available clinical and non-clinical results to substantiate the hypothesized safety risks and potential benefits of each genetic modification.

Characterization of gRNA libraries

The selection and usage of gRNAs warrant meticulous scrutiny. Each gRNA should undergo rigorous analysis, using multiple methods such as in silico predictions, biochemical assays, and cellular-based assays, all of which should include genome-wide analyses to evaluate the likelihood of off-target effects.[25] This comprehensive approach helps reduce bias in identifying potential off-target sites and enables the preferential selection of gRNAs with minimal risks. Historically, systematic experimental validation has been feasible for cell-based gene therapies employing a limited number of genetic modifications. However, this comprehensive approach may be impractical when dealing with extensive libraries of gRNAs, as proposed in this intentionally heterogeneous design. To address this challenge, computational prediction methods could first be applied to identify potential off-target effects, followed by targeted biochemical and cellular assays (eg, GUIDE-seq[33] or RhAmpSeq[34]) to experimentally validate prioritized guides. Additionally, quantitative assays may be developed to measure residual levels of gene-editing reagents, such as Cas9.

Additional non-clinical testing considerations

Additional non-clinical testing may be needed for novel accessory molecules and gene modifications to evaluate the functionality of specific elements and the safety of the infusion product. This includes evaluating the risk of uncontrolled proliferation, which can be assessed through cytokine-independent growth assays and in vitro proliferation studies. Characterizing the activation profile of T cells is essential and can be done through assessments of antigen-dependent activation markers. Phenotypic characterization could include flow cytometry or mass cytometry to analyze surface and intracellular markers that define the T cells' activation status, exhaustion markers, memory phenotype, and potential for persistence or exhaustion. T-cell persistence may also be evaluated using in vivo models designed to mimic the clinical environment. When novel suicide genes are incorporated, non-clinical studies should be conducted to demonstrate their functionality and establish the appropriate dosing of any additional drug or biological product critical to inducing depletion of CAR-T or TCR-T cells in the event of adverse reactions.

CLINICAL CONSIDERATIONS

Protocol design is critical to the safety of clinical trials, especially when testing innovative therapies such as intentionally heterogeneous T-cell populations. Safety measures that could be incorporated into the study protocol include:

Study population

Eligibility criteria should be clearly defined in the protocol. Sponsors may consider restricting enrollment to patients whose cancers have progressed despite having received the best current treatments for that indication. However, sponsors should keep in mind that patients with advanced cancers are often heavily pretreated, resulting in baseline T cells that likely exhibit diminished proliferation potential and may carry multiple mutations, possibly affecting the expansion, efficacy, and safety of the infused cell product. If patients are opting to forgo approved therapies, this information should be clearly stated in the Informed Consent Form. As an additional preventative measure, sponsors may consider excluding patients harboring mutations associated with clonal hematopoiesis of indeterminate potential, which have been associated with increased risk of blood cancers, including T-cell malignancies.[35] However, excluding such patients could introduce bias and potentially confound the interpretation of results.

Dose and dose schedules

The overall number of infused T cells should be consistent with prior experiences with the CAR or TCR construct. The number of T cells carrying secondary gene edits could initially make up a small proportion of the infusion product, potentially increasing in prevalence for subsequent cohorts as initial safety data are collected and analyzed. This escalation approach would allow for a

cautious assessment of the safety profile associated with the gene edits.

Staggered enrollment

Treating several patients simultaneously may represent an unreasonable risk. Sponsors could consider staggered enrollment to limit the number of patients who might be exposed to an unanticipated risk at the same time.

Dose-limiting toxicities (DLTs) and stopping rules

Robust safety stopping rules are essential to safeguard patient well-being and should be clearly defined in the study protocol. These stopping rules are activated in response to DLTs, such as serious adverse events, unanticipated side effects, or events of special interest. Potential DLTs include severe cytokine release syndrome, uncontrolled T-cell proliferation, and aberrant cellular and chromosomal changes. Crossing predefined safety thresholds should necessitate a temporary or permanent pause in the trial to thoroughly evaluate the risks and make informed decisions about the continuation of the study.

Drug Safety Monitoring Board (DSMB)

A dedicated DSMB could actively oversee safety by conducting regular and ad hoc reviews to swiftly address any emerging safety concerns. The presence of an independent DSMB would ensure unbiased monitoring of patient safety throughout the trial.

Assessment of product-related adverse events

During the trial, any significant safety event could be evaluated for associations with specific gene modifications by examining the post-infusion behavior of individual T-cell subpopulations (eg, expansion, localization, and function) using sequencing and other analytical methods.

Biospecimen collection

Longitudinal blood samples and pre-infusion and post-infusion tumor biopsies are critical for measuring the persistence, expansion, function, and trafficking patterns of each distinct T-cell subpopulation. A pre-infusion tumor biopsy should be mandatory for study enrollment, while a post-infusion biopsy should be performed whenever medically feasible. To ensure consistency, the post-infusion biopsy should ideally be obtained from the same lesion as the pre-infusion biopsy.

Long-term follow-up

A long-term follow-up protocol could be established to monitor and record adverse events, facilitating a better understanding of the long-term safety profile of the product.

Collectively, these safety considerations provide a multilayered safety net for trial participants, balancing the innovative aspects of the infusion product with rigorous safety oversight. They are reflective of the commitment to patient safety and the ethical conduct of research, while also providing a framework for collecting vital safety data that informs both current and future trials.

ETHICAL CONSIDERATIONS AND BROADER CHALLENGES

While the previous sections address the safety and technical considerations essential for implementing this novel approach, it is equally important to reflect on the potential ethical, societal, and broader risks associated with its application.

Informed consent in the context of experimental complexity

The inclusion of multiple gene edits in a single trial introduces unprecedented complexity, both in the investigational product and in the potential outcomes. Patients must be fully informed of the heightened uncertainty regarding both safety and efficacy. Particular care must be taken to communicate these complexities in a clear and accessible manner, ensuring that patients understand not only the risks but also the rationale and potential benefits of the trial design.

Risk-benefit balance

Balancing the potential benefits of accelerating therapeutic discovery against the risks of testing multiple gene edits simultaneously is critical. Unanticipated interactions between edits, off-target effects, or emergent safety events highlight the importance of robust safety oversight. Trial designs must weigh these risks carefully, adopting the safeguards previously described to minimize harm.

Patient-specific variability and uncertainty

The response to gene-edited T cells may vary significantly between patients due to differences in tumor biology, immune system function, and genetic background. This variability creates challenges in defining success, as a positive outcome in one patient may reflect unique, patient-specific factors rather than a broadly applicable benefit. Transparency in communicating these uncertainties to patients is essential to maintain trust.

Patient resources and engagement

Patients often report difficulty finding accessible information, such as websites, videos, or discussion boards, to help them understand the investigational product and the trial's objectives. Although this lack of information may not always deter patients from enrolling, it can create anxiety and undermine their confidence in the decision-making process. Patients have expressed a strong desire to feel engaged and to understand what is going into their bodies. To address this need, sponsors should adopt patient-centered approaches to improve accessibility and engagement, such as creating educational materials tailored to diverse literacy levels, offering virtual consultations with trial coordinators or clinical investigators, or establishing interactive platforms to foster a sense of community and support for trial participants.

Access and representation

Barriers to participation, such as geographic, financial, or cultural factors, should be proactively addressed to ensure diverse patient populations have equitable access to trials. Additionally, data from diverse populations are essential

to ensure that findings are generalizable and that benefits from novel therapies are equitably distributed.

DISCUSSION

Two major strengths of cell-based gene therapies include the ability to rapidly and iteratively engineer cells, and the ability to track and characterize the evolution of the infused cells over time. Genome-wide screens have yielded valuable insights into the potential of gene knockouts and knock-ins to enhance T-cell function. Challenges such as impaired T-cell trafficking, reduced function on repeated antigen exposure, limited expansion capabilities, and short persistence can potentially be addressed by knocking out or overexpressing the right gene or set of genes. However, the breadth of potential solutions to these obstacles in preclinical studies also highlights the urgent need for more efficient methods to translate these findings from the laboratory into clinical settings.

Infusing a heterogeneous pool of genetically modified T cells, in which each subpopulation contains a different secondary gene edit, could potentially allow researchers to identify the most effective gene modifications more quickly. While this framework has primarily been considered in the context of engineered CAR-T and TCR-T cells, recent advances suggest that gene editing may also enhance the therapeutic potential of TILs.[4 6] This approach might significantly reduce both the time and costs of development. Rapid testing cycles could quickly eliminate ineffective gene edits, enabling researchers to focus on the most promising candidates. Such a streamlined process should expedite the delivery of groundbreaking therapies to patients.

Before initiating trials testing intentionally heterogeneous T-cell populations, several key details must be addressed. Many aspects are intentionally presented at a high level in this article to encourage creativity and allow for iterative refinement, rather than dictating exact approaches that must be followed. Open questions include the number of secondary gene edits to evaluate, appropriate testing strategies for potential off-target effects, product release criteria, study design, and patient inclusion–exclusion criteria. Collaboration among all stakeholders will be essential to reach consensus on the finer details of this approach, ensuring scientific progress is balanced with a steadfast commitment to patient safety, ethical considerations, and the goal of enhancing therapeutic outcomes.

An additional challenge inherent to testing heterogeneous T-cell populations is the management of toxicity attributable to individual gene edits. If severe toxicity arises from one specific subpopulation, current approaches such as safety switches would potentially eliminate the entire transduced population, including subpopulations thought to be beneficial. Therefore, future work should focus on developing advanced strategies to more precisely balance safety and efficacy in these complex therapeutic designs, potentially through engineering selective safety mechanisms or using molecular analytics to rapidly associate specific gene edits with adverse events. Such innovations would be crucial to fully realizing the promise of intentionally heterogeneous cell-based gene therapies.

Achieving the full potential of cell-based gene therapies requires ongoing cooperation among diverse stakeholders, including regulatory agencies, academic investigators, biopharmaceutical companies, nonprofit organizations, and, importantly, patients. This collaborative approach promises to usher in a new era of more effective and widely applicable cell-based gene therapies for complex diseases.

Author affiliations
[1]Parker Institute for Cancer Immunotherapy, San Francisco, California, USA
[2]Friends of Cancer Research, Washington, District of Columbia, USA
[3]Division of Translational Science and Therapeutics, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
[4]Department of Immunology, University of Washington, Seattle, Washington, USA
[5]Department of Medicine, Division of Hematology and Oncology, University of Washington, Seattle, Washington, USA
[6]Center for Cancer Cell Therapy, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
[7]Division of Pediatric Hematology/Oncology/Stem Cell Transplant and Regenerative Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA
[8]Division of Bone Marrow Transplant-Cell Therapy, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA
[9]Parker Institute for Cancer Immunotherapy at Stanford Medicine, Stanford Medicine, Stanford, California, USA
[10]Center for Cellular Immunotherapies, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA
[11]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[12]Parker Institute for Cancer Immunotherapy at University of Pennsylvania, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[13]Gladstone-UCSF Institute of Genomic Immunology, Gladstone Institutes, San Francisco, California, USA
[14]Department of Medicine, University of California San Francisco, San Francisco, California, USA
[15]University of California San Francisco Helen Diller Family Comprehensive Cancer Center, San Francisco, California, USA
[16]Parker Institute for Cancer Immunotherapy at Gladstone Institutes, Gladstone Institutes, San Francisco, California, USA
[17]Cellular Immunotherapy Program, Massachusetts General Hospital, Boston, Massachusetts, USA
[18]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[19]Department of Medicine, Division of Hematology/Oncology, University of California Los Angeles, Los Angeles, California, USA
[20]Parker Institute for Cancer Immunotherapy at UCLA, University of California Los Angeles, Los Angeles, California, USA

Acknowledgements We extend our gratitude to the FDA Center for Biologics Evaluation and Research (CBER) staff, including Asha Das, Anna Kwilas, Nicole Verdun, Iwen Wu, and Peter Marks, for their valuable feedback and participation as panelists in the workshop. We also thank patient advocates Rick Bangs, Desirée Walker, Chris White, and Tom Whitehead for sharing their perspectives and insights, which were instrumental in shaping the discussion.

Contributors CRC, EY, MDS, JDA, JEC, UD, PDG, CLM, CHJ, AM, MVM, and AR contributed to manuscript drafting, revising, generation of figures and generation of tables. MVM and AR contributed equally and are the guarantors.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests JEC is a cofounder of Twain Therapeutics, director at Dispatch Therapeutics and 3T Biosciences, and receives research funding from Janssen. PDG is a founder and scientific advisory board (SAB) member of, has equity in, and is receiving research support from Affini-T Therapeutics; is an SAB member of and has equity in Immunoscape, RAPT Therapeutics, Earli, Metagenomi, Nextech, and Catalio. CLM is an inventor on multiple patents related to CAR-T cell therapy and has received royalties for licenses to Juno Therapeutics and CARGO Therapeutics; is a co-founder of Lyell Immunopharma, CARGO Therapeutics and Link Cell Therapies, which are developing CAR-T cells and holds equity in these companies; is a consultant for CARGO, Link, Immatics, Ensoma, Adaptimmune and received research funding from Lyell and Tune Therapeutics. CHJ is an inventor on patents and/or patent applications licensed to Novartis and receives license revenue from such licenses; is a scientific founder of Tmunity Therapeutics, Dispatch Biotherapeutics, Bluewhale Bio, Marble Therapeutics and Capstan Therapeutics; is a member of the scientific advisory boards of AC Immune, BluesphereBio, Cabaletta, Carisma, Cartography, Cellares, Cellcarta, Celldex, Danaher, Decheng, ImmuneSensor, Kite Gilead, Poseida, Replay Bio, Treadwell Therapeutics, Verismo, Viracta, Vittoria and WIRB-Copernicus; and receives sponsored research support from Kite Gilead. AM is a co-founder of Site Tx, Arsenal Biosciences, Spotlight Therapeutics, and Survey Genomics; serves on the boards of directors of Site Tx, Spotlight Therapeutics, and Survey Genomics; is a member of the scientific advisory boards of Site Tx, Arsenal Biosciences, Cellanome, Spotlight Therapeutics, Survey Genomics, NewLimit, Amgen, and Tenaya; owns stock in Arsenal Biosciences, Site Tx, Cellanome, Spotlight Therapeutics, NewLimit, Survey Genomics, Tenaya, and Lightcast; has received fees from Site Tx, Arsenal Biosciences, Cellanome, Spotlight Therapeutics, NewLimit, Gilead, Pfizer, 23andMe, PACT Pharma, Juno Therapeutics, Tenaya, Lightcast, Trizell, Vertex, Merck, Amgen, Genentech, GLG, ClearView Healthcare, AlphaSights, Rupert Case Management, Bernstein, and ALDA; is an investor in and informal advisor to Offline Ventures and a client of EPIQ; has received research support from the Parker Institute for Cancer Immunotherapy, the Emerson Collective, Arc Institute, Juno Therapeutics, Epinomics, Sanofi, GlaxoSmithKline, Gilead, and Anthem and reagents from Genscript and Illumina. MVM is an inventor on patents related to adoptive cell therapies, held by Massachusetts General Hospital (some licensed to Promab) and the University of Pennsylvania (some licensed to Novartis); reports grant or research support from Kite Pharma and Moderna; reports consulting fees from several companies involved in cell therapies; has equity in 2Seventy Bio, A2Bio, Cargo, Century Therapeutics, Neximmune, Oncternal and TCR2; and has served on the board of directors for 2Seventy Bio. AR reports personal/consulting fees from Amgen, Arcus, Bioncotech, Bristol Myers Squibb Company, Chugai, Compugen, Cytomex, Five Prime, FLX-Bio, Genentech/Roche, MapKure, Merck, Merus, Rgenix, PACT Pharma, Synthekine, and Tango Therapeutics; support for other professional activities from Arsenal Bio; is a fiduciary officer at Apricity Health, Arcus Bio, and Lutris Pharma; holds a patent with Arsenal Bio; and owns stock or stock options in Appia, Apricity Health, Arcus Bio, Bioncotech, Compugen, Cytomex, Five Prime, FLX-Bio, Immpact Bio, Highlight Therapeutics, Inspirina, Lutris Pharma MapKure, Merus, Rgenix, PACT Pharma, Synthekine, and Tango Therapeutics all outside the submitted work. CRC, EY, MDS, JDA, and UD report no competing interests.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs

Christopher R Cabanski http://orcid.org/0009-0007-9938-1331
EnJun Yang http://orcid.org/0000-0003-1511-7511
Mark D Stewart http://orcid.org/0000-0002-4847-0736
Marcela V Maus http://orcid.org/0000-0002-7578-0393
Antoni Ribas http://orcid.org/0000-0003-3669-8458

REFERENCES

1 Boyiadzis MM, Dhodapkar MV, Brentjens RJ, *et al*. Chimeric antigen receptor (CAR) T therapies for the treatment of hematologic malignancies: clinical perspective and significance. *J Immunother Cancer* 2018;6:137.
2 Hinrichs CS, Rosenberg SA. Exploiting the curative potential of adoptive T-cell therapy for cancer. *Immunol Rev* 2014;257:56–71.
3 Klebanoff CA, Chandran SS, Baker BM, *et al*. T cell receptor therapeutics: immunological targeting of the intracellular cancer proteome. *Nat Rev Drug Discov* 2023;22:996–1017.
4 Lou E, Choudhry MS, Starr TK, *et al*. Targeting the intracellular immune checkpoint CISH with CRISPR-Cas9-edited T cells in patients with metastatic colorectal cancer: a first-in-human, single-centre, phase 1 trial. *Lancet Oncol* 2025;26:559–70.
5 Martinov T, Greenberg PD. Targeting Driver Oncogenes and Other Public Neoantigens Using T Cell Receptor-Based Cellular Therapy. *Annu Rev Cancer Biol* 2023;7:331–51.
6 Met Ö, Svane IM. Unlocking tumor-infiltrating lymphocyte potential with CRISPR-Cas9. *Lancet Oncol* 2025;26:530–1.
7 Ribas A. T Cells as the Future of Cancer Therapy. *Cancer Discov* 2021;11:798–800.
8 Bailey SR, Berger TR, Graham C, *et al*. Four challenges to CAR T cells breaking the glass ceiling. *Eur J Immunol* 2023;53.
9 Belk JA, Yao W, Ly N, *et al*. Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors that limit T cell persistence. *Cancer Cell* 2022;40:768–86.
10 Carnevale J, Shifrut E, Kale N, *et al*. RASA2 ablation in T cells boosts antigen sensitivity and long-term function. *Nature* 2022;609:174–82.
11 Chen Z, Arai E, Khan O, *et al*. In vivo CD8$^+$ T cell CRISPR screening reveals control by Fli1 in infection and cancer. *Cell* 2021;184:1262–80.
12 Dong MB, Wang G, Chow RD, *et al*. Systematic Immunotherapy Target Discovery Using Genome-Scale In Vivo CRISPR Screens in CD8 T Cells. *Cell* 2019;178:1189–204.
13 Freitas KA, Belk JA, Sotillo E, *et al*. Enhanced T cell effector activity by targeting the Mediator kinase module. *Science* 2022;378:eabn5647.
14 Guo A, Huang H, Zhu Z, *et al*. cBAF complex components and MYC cooperate early in CD8+ T cell fate. *Nature* 2022;607:135–41.
15 Kumar S, Zeng Z, Bagati A, *et al*. CARM1 Inhibition Enables Immunotherapy of Resistant Tumors by Dual Action on Tumor Cells and T Cells. *Cancer Discov* 2021;11:2050–71.
16 Li W, Xu H, Xiao T, *et al*. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 2014;15:554.
17 Shifrut E, Carnevale J, Tobin V, *et al*. Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* 2018;175:1958–71.
18 Labanieh L, Mackall CL. CAR immune cells: design principles, resistance and the next generation [published correction appears in Nature. *Nature* 2023;614:E26–648.
19 Xiang M, Li H, Zhan Y, *et al*. Functional CRISPR screens in T cells reveal new opportunities for cancer immunotherapies. *Mol Cancer* 2024;23:73.
20 Ahmed EN, Cutmore LC, Marshall JF. Syngeneic Mouse Models for Pre-Clinical Evaluation of CAR T Cells. *Cancers (Basel)* 2024;16:3186.
21 Sanmamed MF, Chester C, Melero I, *et al*. Defining the optimal murine models to investigate immune checkpoint blockers and their combination with other immunotherapies. *Ann Oncol* 2016;27:1190–8.
22 Duncan BB, Dunbar CE, Ishii K. Applying a clinical lens to animal models of CAR-T cell therapies. *Mol Ther Methods Clin Dev* 2022;27:17–31.
23 Cassidy JW, Caldas C, Bruna A. Maintaining Tumor Heterogeneity in Patient-Derived Tumor Xenografts. *Cancer Res* 2015;75:2963–8.
24 Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs): Guidance for Industry, 2020. Available: https://www.fda.gov/media/113760/download
25 Human Gene Therapy Products Incorporating Human Genome Editing: Guidance for Industry, 2024. Available: https://www.fda.gov/media/156894/download
26 Chiesa R, Georgiadis C, Syed F, *et al*. Base-Edited CAR7 T Cells for Relapsed T-Cell Acute Lymphoblastic Leukemia. *N Engl J Med* 2023;389:899–910.
27 Stadtmauer EA, Fraietta JA, Davis MM, *et al*. CRISPR-engineered T cells in patients with refractory cancer. *Science* 2020;367:eaba7365.
28 Fraietta JA, Nobles CL, Sammons MA, *et al*. Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature* 2018;558:307–12.

29 Shah NN, Qin H, Yates B, *et al*. Clonal expansion of CAR T cells harboring lentivector integration in the CBL gene following anti-CD22 CAR T-cell therapy. *Blood Adv* 2019;3:2317–22.
30 Levine BL, Pasquini MC, Connolly JE, *et al*. Unanswered questions following reports of secondary malignancies after CAR-T cell therapy. *Nat Med* 2024;30:338–41.
31 Verdun N, Marks P. Secondary Cancers after Chimeric Antigen Receptor T-Cell Therapy. *N Engl J Med* 2024;390:584–6.
32 Sahillioglu AC, Schumacher TN. Safety switches for adoptive cell therapy. *Curr Opin Immunol* 2022;74:190–8.
33 Tsai SQ, Zheng Z, Nguyen NT, *et al*. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 2015;33:187–97.
34 Dobosy JR, Rose SD, Beltz KR, *et al*. RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnol* 2011;11:80.
35 Niroula A, Sekar A, Murakami MA, *et al*. Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat Med* 2021;27:1921–7.

Journal for ImmunoTherapy of Cancer

Enabling access to genetically modified cell therapies through flexible approaches to manufacturing and cost recovery

Mark D Stewart ,[1] Christopher R Cabanski ,[2] Jeff D Allen,[1] John E Connolly,[2] Ben M Beneski,[3] Boro Dropulić,[4] Steven A Feldman,[5] Lee A Fleisher,[6,7] Patrick J Hanley,[8,9] Kristen Hege,[10] Natasha Kekre ,[11] Holly Fernandez Lynch,[12] Crystal L Mackall [13,14,15]

To cite: Stewart MD, Cabanski CR, Allen JD, *et al*. Enabling access to genetically modified cell therapies through flexible approaches to manufacturing and cost recovery. *Journal for ImmunoTherapy of Cancer* 2025;**13**:e013518. doi:10.1136/jitc-2025-013518

Accepted 17 November 2025

For numbered affiliations see end of article.

Correspondence to
Dr Mark D Stewart;
mstewart@focr.org

ABSTRACT

Genetically modified cell-based therapies hold transformative potential, particularly for patients with rare cancers and ultra-rare diseases. However, progress toward regulatory approval, reimbursement, and broad patient access is often constrained by misaligned regulatory, manufacturing, and financial frameworks that do not reflect the realities of treating small populations and low-throughput production models. Drawing on a collaborative white paper and public meeting convened by Friends of Cancer Research and the Parker Institute for Cancer Immunotherapy in May 2025, this commentary outlines three strategies to streamline regulatory pathways and enable timely, sustainable access: (1) flexible approaches to Chemistry, Manufacturing, and Controls requirements in small populations, (2) adaptable regulatory frameworks to support diverse manufacturing models, and (3) limited cost recovery mechanisms to bridge early access and development gaps. Recent regulatory and policy discussions have echoed these priorities, signaling an opportunity to align oversight with operational realities to advance innovation and access for patients in high-need settings.

INTRODUCTION

In recent years, scientific progress in genetically modified cell-based therapies has accelerated dramatically. Innovative products such as chimeric antigen receptor T-cell and T-cell receptor therapies have demonstrated remarkable clinical potential, including in patients with cancer who previously had few or no effective treatment options. In many cases, we are no longer waiting for the science—robust biological rationale and encouraging clinical data exist. Instead, the limiting factor is misalignment across regulatory, manufacturing, and financial systems, which each presenting distinct challenges that prevent promising therapies from achieving regulatory approval and reaching patients. As a result, promising therapies may be shelved despite strong clinical signals and urgent unmet need, particularly in settings that fall outside traditional biopharmaceutical development models, such as autologous therapies for low-incidence tumor types, including pediatric cancers.[1] While some genetically modified cell-based therapies may begin in small, defined populations and later expand to broader indications, the challenges discussed here are most acute for therapies inherently limited to rare or ultra-rare diseases. These products often lack the opportunity to achieve scale and therefore require tailored, proportional approaches to regulatory flexibility and manufacturing feasibility.

To address these barriers, Friends of Cancer Research and the Parker Institute for Cancer Immunotherapy convened experts across the research, regulatory, academic, and patient communities to identify actionable strategies for greater regulatory flexibility, adaptable manufacturing approaches, and limited cost recovery to sustain development. These discussions informed a collaborative white paper reflecting broad alignment on the need for systems-level change—not in the rigor of scientific standards nor the regulatory standards governing safety and effectiveness for product approval, but in how they are operationalized.[2] Recent regulatory conversations have echoed many of the same themes, including the recently described "plausible mechanism" pathway, reinforcing the urgency to modernize the development and access ecosystem around cell and gene therapies (CGT).[3 4]

CMC FLEXIBILITY IN SMALL POPULATIONS

Current regulatory expectations for Chemistry, Manufacturing, and Controls (CMC) are

based on models suited to large-scale, commercial manufacturing. However, for therapies aimed at small, genetically defined, or ultra-rare populations, expectations such as extensive stability data, full process validation, and batch-based comparability studies can be difficult to meet due to limited starting material, low manufacturing throughput, and constrained timelines. The traditional "one-size-fits-all" model may impose significant burdens on developers operating with low-throughput processes or constrained resources. At the same time, the US Food and Drug Administration (FDA) has demonstrated flexibility through risk-based approaches and case-by-case considerations, particularly for rare disease products.

For therapies targeting small populations, safe and reproducible manufacturing may be achievable through abbreviated stability protocols at the time of product release, especially when supported by robust in-process controls and scientific rationale. In these contexts, alternative evidence, such as representative engineering runs, process simulations, or non-donor-matched comparability material, may sufficiently demonstrate process control when traditional batch-based comparability is infeasible. Advanced submission of select CMC components and iterative data updates—rather than requiring a complete CMC package upfront—can further improve feasibility, particularly where material availability is limited.

Importantly, this is not about lowering standards but rather applying them in a way that is proportional to the risk, context, and scale of the therapy (ie, proportional regulation). Predictable application of existing flexibilities, discussed early in development between the developer and the regulator, can help prevent delays or product abandonment. Establishing clearer guardrails for when and how flexibility may be applied could improve confidence for developers while maintaining patient safety and benefit and product quality. To support timely and efficient adaptation to evolving regulatory requirements, regulators should routinely disseminate aggregated, non-identifiable data and examples related to flexible approaches that regulators have accepted in related contexts. Such data would enable developers to align strategies with emerging regulatory expectations in a more proactive and resource-efficient manner. Flexibilities should be considered in settings where traditional development pathways may not be viable, as proportional application of regulatory standards can maintain quality and safety while remaining feasible.

ADAPTABLE MANUFACTURING MODELS

Enabling access to cell-based therapies also requires a regulatory framework that can accommodate a range of manufacturing approaches. Centralized production may be appropriate in some cases, for example in allogenic therapies, but for therapies with logistical constraints or narrow windows for administration, decentralized or point-of-care (POC) manufacturing models may also play an important role due to cost and timing advantages of products manufactured locally.[5,6] Each approach presents trade-offs, but diverse manufacturing models can introduce beneficial innovation and redundancy, enhancing the system's agility in developing CGT.

Academic centers and other translational research hubs, such as hospital-based cell therapy programs or centers of excellence, can be well positioned to support POC or decentralized manufacturing approaches. Emerging models, including mobile POC platforms and modular manufacturing systems, could enable treatment delivery closer to patients while preserving product quality. To support this shift, regulatory oversight can be tailored through site-level accreditation, real-time monitoring tools, and chain-of-identity and custody safeguards.[7] For adaptable manufacturing models, the transition from early-phase to commercial Good Manufacturing Practices (GMP) requirements often requires substantial upgrades to facilities, processes, and compliance—a leap that can be prohibitive for rare disease therapies. While rigorous Quality Assurance (QA) and Quality Control (QC) standards must remain in place, more flexible, risk-based approaches to facility design or environmental controls could ensure product quality and safety while improving feasibility and access. Performance data from existing distributed platforms can serve as an evidence base for shaping flexible, fit-for-purpose oversight models. Mechanisms such as modular comparability protocols, automated monitoring, and shared learning networks may help ensure reproducibility across sites.

Rather than prioritizing a single model, regulatory systems should support the approach best suited to a therapy's attributes and patient population. The goal is not to replace centralized manufacturing but to enable a more flexible ecosystem capable of supporting a diversity of therapeutic contexts.[8] Regulators have also begun developing frameworks to enable such diversity that include FDA's FRAME (Framework for Regulatory Advanced Manufacturing Evaluation) initiative, which prioritizes distributed and POC production, and the UK The Medicines and Healthcare products Regulatory Agency's 2025 Modular Manufacture and Point-of-Care Regulations, which establish an operational pathway for decentralized production of advanced therapies.[9,10] These developments demonstrate growing recognition of the need for adaptable oversight frameworks.

EXPLORING LIMITED COST RECOVERY TO SUPPORT ACCESS

Even with flexible CMC and operational pathways in place to support CGT manufacturing, many therapies targeting small populations may remain inaccessible due to lack of commercial viability. These therapies face a predicament: traditional reimbursement is unavailable due to lack of marketing approval, and continued development is financially unsustainable without external support.

Structured, time-limited mechanisms may provide interim support in narrowly defined settings, thereby facilitating early patient access and providing the funding

needed to progress toward approval. Emphasis should be placed on public–private partnerships, National Institutes of Health-sponsored protocols, or targeted grant support as preferred options to sustain development and enable preapproval access in rare diseases. In exceptional circumstances, regulated preapproval access pathways (eg, FDA's Expanded Access program), when paired with defined cost recovery guardrails, may also help bridge short-term gaps. Such approaches should have appropriate safeguards to preserve incentives for full approval, while ensuring patient safety, product quality, and appropriate oversight. These mechanisms are envisioned as temporary, bridging measures to sustain development and enable access until full regulatory approval and traditional reimbursement can be achieved.

Current statutes limit Centers for Medicare & Medicaid Services (CMS) from covering unapproved products, which is an important protection to ensure public resources are not spent on therapies lacking evidence of safety and effectiveness. However, publicly supported mechanisms, such as those previously mentioned, could serve as transitional channels for patient access and evidence generation leading up to FDA approval. Any framework for interim support should be transparent, have clearly defined eligibility criteria, and promote equitable access. Importantly, the structured approaches described here differ from other early access pathways, such as Right to Try, which lack components to prioritize product development and avoid exploitation of patients with unmet needs.[11]

CONCLUSION AND PATH FORWARD

Achievable regulatory standards and sustainable patient access to genetically modified cell-based therapies that face barriers to traditional commercial development will require coordinated action across regulatory, manufacturing, and financial dimensions. Applying a framework of proportional regulation—where standards remain rigorous but are tailored to the scale, risk, benefit, and

Table 1 Illustrative barriers and proposed targeted solutions across the cell therapy development lifecycle

Stage of development	Key barriers	Proposed targeted solutions
IND preparation	► Limited starting material availability. ► Uncertainty around CMC data expectations.	► Modular CMC submissions. ► Use of representative engineering runs. ► Early engagement with FDA to align expectations in low-throughput, high-need therapies.
CMC development	► Challenges meeting stability, release, and comparability expectations for low-throughput processes.	► Abbreviated stability protocols. ► Use of in-process controls and scientific justification. ► Use of non-donor-matched comparability material. ► Allowance for iterative data updates.
Early clinical development	► Logistical challenges of patient enrollment and treatment delivery for geographically dispersed rare disease populations. ► Timelines were prolonged due to scale-up and facility constraints.	► Facilitate earlier access through academic or decentralized GMP manufacturing with stage-appropriate risk-based expectations. ► Site-level accreditation models to support reliability and oversight.
Late-stage/pivotal trials	► Operational burden of scaling manufacturing for broader clinical trial use. ► Barriers to qualifying multiple sites.	► Scalable, modular oversight frameworks. ► Real-time control and monitoring systems. ► Structured site qualification protocols to facilitate broader site activation.
Post-trial, preapproval phase	► Maintaining GMP compliance and manufacturing operations between trial completion and approval is often infeasible in low-throughput settings. ► Lack of financial pathways to support continued access.	► Structured cost recovery pathways. ► Expanded access with defined reimbursement guardrails. ► NIH/public infrastructure-supported treatment protocols as bridging channels.
Postapproval phase	► Scaling and sustaining manufacturing and distribution for rare indications.	► Continued application of proportional oversight. ► Use of shared infrastructure or public–private partnerships to maintain product availability in low-volume settings.

CMC, Chemistry, Manufacturing, and Controls; FDA, Food and Drug Administration; GMP, Good Manufacturing Practices; IND, Investigational New Drug; NIH, National Institutes of Health.

context of the therapy and underlying disease—can ensure that scientific and safety expectations are met for regulatory approval while enabling operational feasibility (table 1). Such alignment may also lower perceived development barriers and incentivize investment in therapies that currently fall outside traditional commercial models. Progress will also depend on workforce training and continued innovation to improve manufacturing efficiency and reduce costs, ensuring that new therapies can be delivered sustainably at scale.

To move from concept to implementation, several targeted opportunities should be pursued:

- Clarify and right-size regulatory expectations for CMC development in small-population and low-throughput settings through targeted guidance. This should include examples of representative engineering runs, modular submissions, and scientifically justified alternatives to traditional stability and comparability protocols, enabling developers to demonstrate product quality and process control in a feasible manner.
- Support adaptable manufacturing models through pilot approaches that enable decentralized or academic manufacturing of rare disease CGT products. These mechanisms could include site-level accreditation, use of real-world performance data, and risk-based GMP considerations tailored to the stage of development and intended use.
- Establish structured cost recovery mechanisms in collaboration with FDA, CMS, private payors, and public–private partnerships to support preapproval access and continued development where traditional development is not viable, particularly for rare diseases.

Ongoing initiatives and recent policy dialogs have signaled a shared recognition that regulatory and operational innovation is necessary. These efforts offer an actionable blueprint for ensuring that innovation in cell-based therapies progresses through regulatory pathways to reach all patients, regardless of market size or geography.

Author affiliations
[1]Friends of Cancer Research, Washington, District of Columbia, USA
[2]Parker Institute for Cancer Immunotherapy, San Francisco, California, USA
[3]Allogene Therapeutics, South San Francisco, California, USA
[4]Caring Cross, Gaithersburg, Maryland, USA
[5]Laboratory for Cell and Gene Medicine, Stanford University School of Medicine, Stanford, California, USA
[6]Rubrum Advising, Philadelphia, Pennsylvania, USA
[7]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[8]Center for Cancer and Immunology Research, Children's National Hospital, Washington, District of Columbia, USA
[9]Department of Pediatrics, The George Washington University, Washington, District of Columbia, USA
[10]Independent Advisor, Burlingame, California, USA
[11]Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada
[12]Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[13]Center for Cancer Cell Therapy, Stanford University School of Medicine, Stanford, California, USA
[14]Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA
[15]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

Acknowledgements The authors thank the working group members who contributed to the development of the proposals outlined in this commentary through their participation in planning discussions and review of the accompanying meeting white paper. Their insights and perspectives helped shape the recommendations and identify practical solutions to support the development and access of genetically modified cell-based therapies. Working group members included Mehrshid Alai-Safar, Gilead Sciences; Erin Bassett, BioCanRx; Jacqueline Bussolari, Johnson & Johnson Innovative Medicine; Ben Beneski, Allogene Therapeutics; Julio Delgado, University of Barcelona; Amanda Dinofia, Children's Hospital of Philadelphia; Boro Dropulic, Caring Cross; Steven Feldman, Stanford University; Holly Fernandez Lynch, Perelman School of Medicine at the University of Pennsylvania; Judy Fitzgerald, Patient Advocate; Lee Fleisher, Rubrum Advising; Patrick Hanley, Children's National Hospital; Kristen Hege, Independent Advisor; Natasha Kekre, The Ottawa Hospital; Seraphin Kuate, Bristol Myers Squibb; Bruce Levine, Perelman School of Medicine at the University of Pennsylvania; Deepu Madduri, Johnson & Johnson Innovative Medicine; Crystal Mackall, Stanford University; David Mitchell, CAR T Patient and Patients For Affordable Drugs; Dattesh Suthar, Novartis; Chris White, Patient Advocate.

Contributors MS and CRC drafted the initial manuscript. All authors contributed to critical review, provided edits and feedback, and approved the final version for submission. MS is the guarantor of this work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests PH is an advisor to or on the advisory board of: Autolomous, CARTx, Cellenkos, Cipla, March Biosciences, MFX, and Y Innovations. KH is a member of the board of directors for: KSQ Therapeutics, EvolveImmune Therapeutics, Kelonia Therapeutics, Adaptimmune, and Mersana Therapeutics. KH also is an advisor to multiple biotech companies. NK receives honoraria from BMS, Novartis and Gilead. BMB is employed by Allogene Therapeutics. CRC receives consulting fees from Georgiamune. JEC is a co-founder of Twain Therapeutics, director at Dispatch Therapeutics and 3T Biosciences, and receives research funding from Janssen. HFL receives research funding from Arnold Ventures, the Greenwall Foundation, and NIH. She received honorarium and travel support for a 2024 bioethics lecture to Eli Lilly on FDA approval standards. SF is on the advisory board of Autolomous, MFX, Achieve Clinics, Oslo University Hospital, Division of Cancer Medicine and Institute for Cancer Research; board member Biotech Partners. CM is a Founder and Board Member of Link Cell Therapies, holds equity in Link Cell Therapies and Ensoma and consults for Link Cell Therapies, Red Tree Venture Capital, Immatics, Ensoma, Grace Science, Nektar, Kite Pharma, and AstraZeneca. MS, JA, LAF, and BD report no COIs.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Mark D Stewart https://orcid.org/0000-0002-4847-0736
Christopher R Cabanski https://orcid.org/0009-0007-9938-1331
Natasha Kekre https://orcid.org/0000-0001-8394-0855
Crystal L Mackall https://orcid.org/0000-0003-0359-9023

REFERENCES

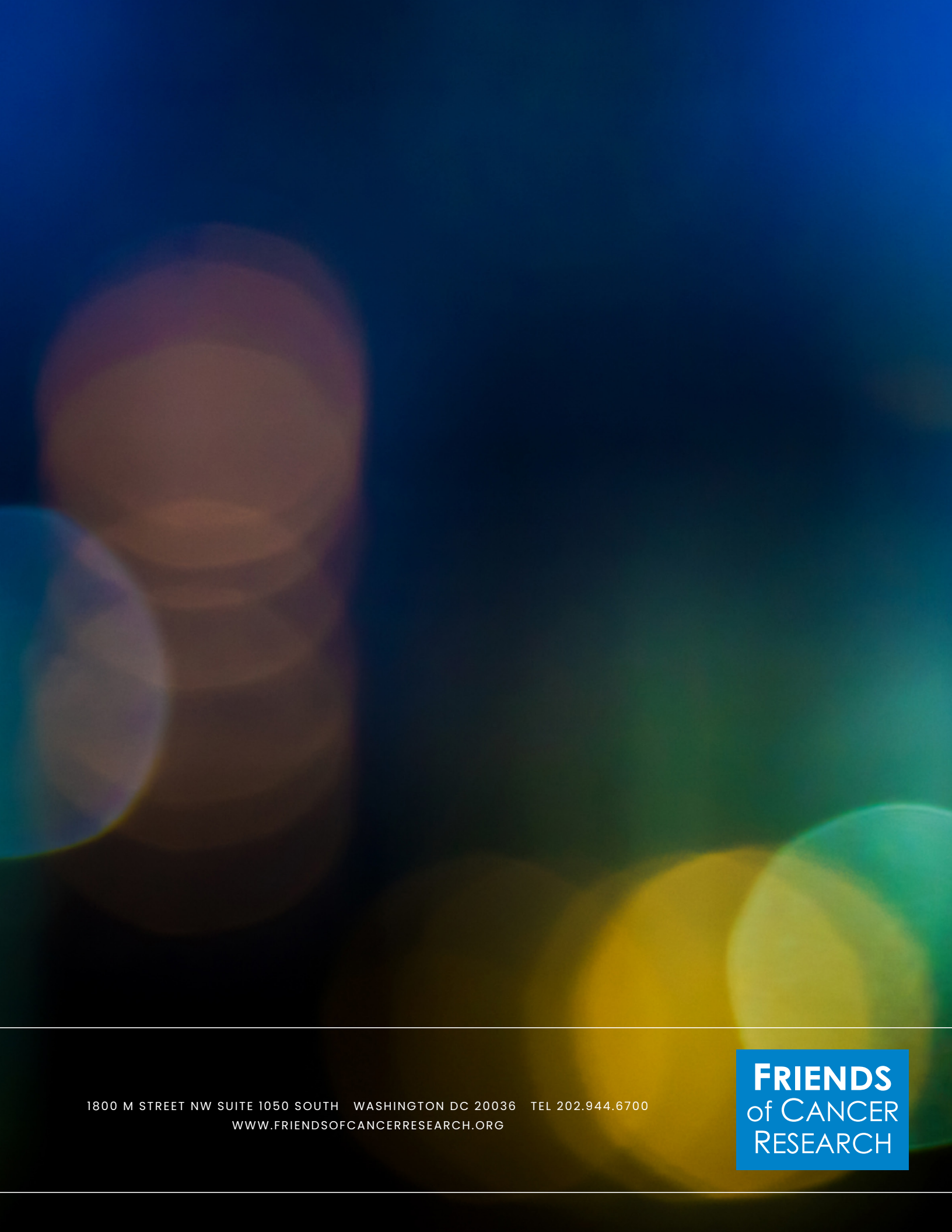
1 Mackall CL, Bollard CM, Goodman N, *et al*. Enhancing pediatric access to cell and gene therapies. *Nat Med* 2024;30:1836–46.
2 Friends of Cancer Research and Parker Institute for Cancer Immunotherapy. Regulatory and manufacturing pathways to expand access to genetically modified cell-based therapies [white paper].

2025 Available: https://friendsofcancerresearch.org/wp-content/uploads/Regulatory-and-Manufacturing-Pathways-to-Expand-Access-to-Genetically-Modified-Cell-Based-Therapies.pdf
3 FDA. Cell and gene therapy roundtable. 2025. Available: https://www.fda.gov/vaccines-blood-biologics/workshops-meetings-conferences-biologics/cell-and-gene-therapy-roundtable-06052025 [Accessed 6 Aug 2025].
4 Prasad V, Makary MA. FDA's New Plausible Mechanism Pathway. *N Engl J Med* 2025.
5 Forum on Regenerative Medicine, Board on Health Sciences Policy, Health and Medicine Division, *et al*. *Emerging technologies and innovation in manufacturing regenerative medicine therapies*. Proceedings of a Workshop in Brief. Emerging Technologies and Innovation in Manufacturing Regenerative Medicine Therapies: Proceedings of a Workshop in Brief 2024:1–15. Washington, D.C, Available: https://www.nap.edu/catalog/27483
6 Johnson B. Reducing the costs of blockbuster gene and cell therapies in the Global South. *Nat Biotechnol* 2024;42:8–12.
7 8th edition of fact-jacie standards. EBMT. Available: https://www.ebmt.org/8th-edition-fact-jacie-standards [Accessed 6 Aug 2025].
8 Urbano-Ispizua Á, Shah NN, Kekre N. Point-of-care CAR T manufacturing solutions: can 1 model fit all? *Blood Adv* 2024;8:6133–6.
9 Federal Register. Discussion paper: distributed manufacturing and point-of-care manufacturing of drugs; request for information and comments. Available: https://www.federalregister.gov/documents/2022/10/14/2022-22386/discussion-paper-distributed-manufacturing-and-point-of-care-manufacturing-of-drugs-request-for [Accessed 6 Aug 2025].
10 MHRA. Decentralised manufacture hub. Available: https://www.gov.uk/government/collections/decentralised-manufacture-hub [Accessed 6 Aug 2025].
11 Right to try. FDA. Available: https://www.fda.gov/patients/learn-about-expanded-access-and-other-treatment-options/right-try [Accessed 6 Aug 2025].

FRIENDS of CANCER RESEARCH

ABOUT FRIENDS OF CANCER RESEARCH

Friends of Cancer Research (*Friends*) powers advances in science and policy that speed life-saving treatments to patients. *Friends* aims to accelerate cutting edge cancer care that both extends and improves quality of life for patients. To accomplish this, we leverage groundbreaking collaborations, generate scientific evidence, and integrate patient input to shape public policy.



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

FRIENDS of CANCER RESEARCH