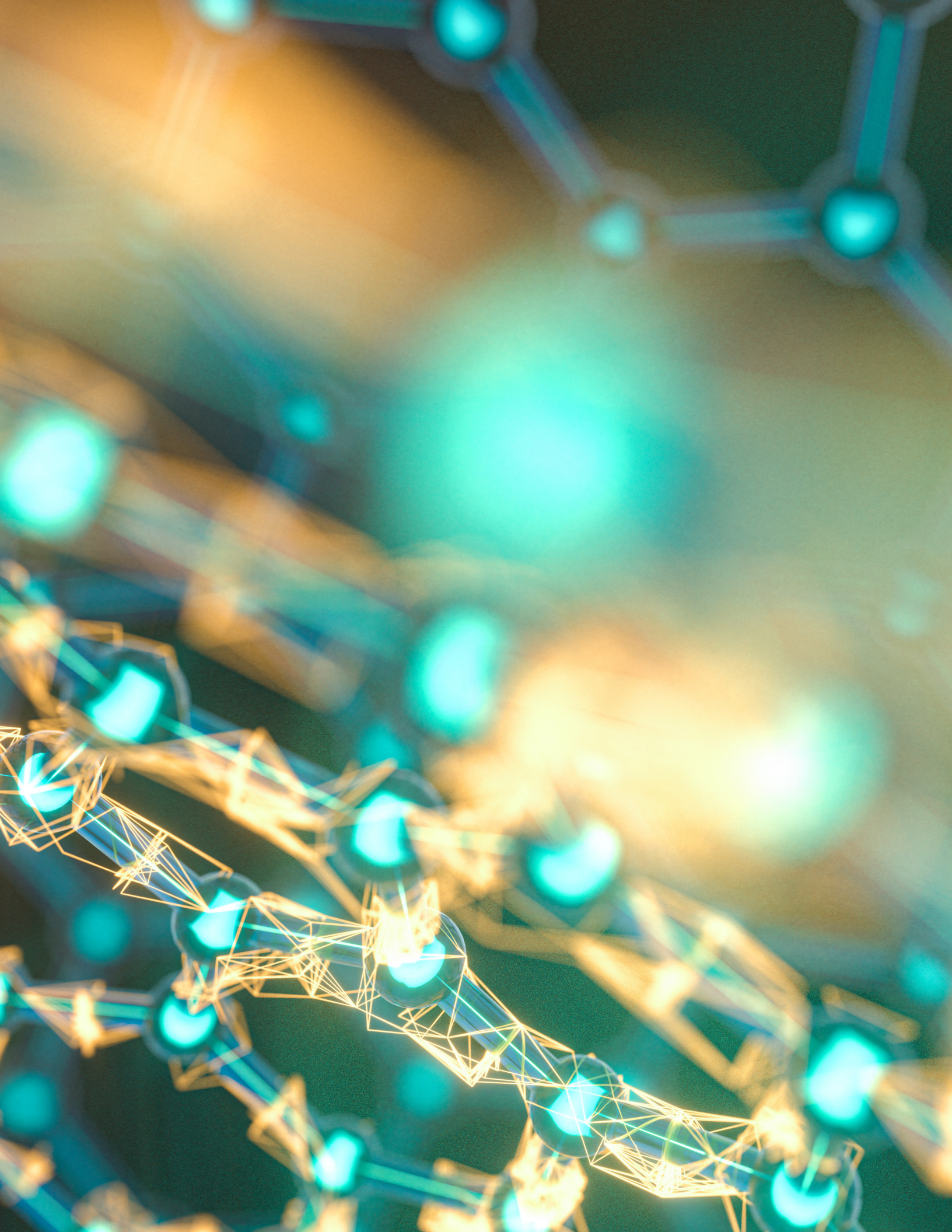
2024 SCIENTIFIC REPORT

# Regulatory Advancements for Patients

FRIENDS of CANCER RESEARCH



FRIENDS of CANCER RESEARCH


4 **INTRODUCTION**

15 **PATIENT-FOCUSED DRUG DEVELOPMENT: ENHANCING REPRESENTATIVENESS AND EQUITY IN CLINICAL TRIALS**

16 Impact of the COVID-19 Pandemic Mitigation Strategies on Cancer Treatment Trials: A Meta-analysis of Industry and NCI Studies
17 An Evaluation of Novel Oncology Approvals with a PMR/C for Assessing Data in Racial and Ethnic Populations Underrepresented in Premarket Clinical Trials
24 Improving Equity in Oncology Clinical Trials: Challenges and Strategies for Setting Diversity Enrollment Goals

45 **REAL-WORLD EVIDENCE: LEVERAGING DATA FROM ROUTINE CLINICAL PRACTICE IN ONCOLOGY**

46 Evaluation of Real-World Tumor Response Derived From Electronic Health Record Data Sources: A Feasibility Analysis in Patients With Metastatic Non–Small Cell Lung Cancer Treated With Chemotherapy*
55 Bridging Research and Practice: Enhancing Regulatory Decisions with Pragmatic Clinical Trials in Oncology
59 Pragmatism in Postmarket Trials

81 **INNOVATIVE DRUG DEVELOPMENT: ADVANCING EARLY ENDPOINTS AND NOVEL EVIDENCE PATHWAYS**

82 Accelerating the Development of Genetically Engineered Cellular Therapies: A Framework for Extrapolating Data Across Related Products
89 Enhancing Study Designs and Interpretation of Interim Overall Survival Data in Oncology Trials
109 The Impact of Treatment Modalities on Use of CTDNA as an Early Endpoint in ANSCLC Trials
110 Framework for Integrating Change in ctDNA Levels in Advanced Cancer Clinical Trials to Support Meta-analyses for Intermediate Endpoint Validation
127 Appendix: Proposed Meta-Analysis Strategy for Validation of ctDNA as an Intermediate Endpoint in aNSCLC Treated with Immunotherapy and Summary of Additional Substudies to Consider
131 Advancing Evidence Generation for Circulating Tumor DNA: Lessons Learned from A Multi-Assay Study of Baseline Circulating Tumor DNA Levels across Cancer Types and Stages

141 **COMPLEX BIOMARKERS: HARMONIZING MEASUREMENT AND AI APPLICATIONS**

142 Analysis of 20 Independently Performed Assays to Measure Homologous Recombination Deficiency (HRD) in Ovarian Cancer: Findings From the Friends' HRD Harmonization Project*
154 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


**The journal's top 5 downloaded publications of 2024*

# Introduction

Friends of Cancer Research (*Friends*) is a leader in transforming oncology drug development and regulatory policy, driving advancements in treatment through collaborative and innovative initiatives. In 2024, *Friends* continued to foster partnerships between scientists, advocates, and other experts to generate evidence-based solutions that tackle critical challenges in oncology drug development and patient care.

*Friends* made strides in 2024, particularly in advancing our diagnostic harmonization portfolio. Notably, we presented final results from our Homologous Recombination Deficiency (HRD) Harmonization Project and launched a new project, the Digital and Computational Pathology Tool Harmonization (Digital PATH) Project, which aims to evaluate alignment across computational pathology models that assess HER2 status in breast cancer. Our diagnostic harmonization projects contribute to our broader goal of developing harmonized approaches for biomarker and test performance (see more on these projects in the Project Spotlight on page 12). *Friends*' commitment to generating novel data to support regulatory policy is exemplified through these efforts and our other research partnerships and policy projects including our Real-world Evidence (RWE) Portfolio and our ctDNA for Monitoring Treatment Response (ctMoniTR) Project.

The data generated from these partnerships, along with the outputs of our working groups, roundtables, and policy research, constitute the core content of this Scientific Report, contribute novel insights, and support ongoing policy discussions. This report aims to serve as a resource for stakeholders in drug development, regulatory policy, and advocacy, by offering insights, evidence-based strategies, and collaborative solutions that advance the field of oncology drug development for patients.

The 2024 Scientific Report includes the full text of our white papers and publications, which center on four themes:

1. **PATIENT-FOCUSED DRUG DEVELOPMENT**
   Enhancing Representativeness and Equity in Clinical Trials
2. **REAL-WORLD EVIDENCE**
   Leveraging Data from Routine Clinical Practice in Oncology
3. **INNOVATIVE DRUG DEVELOPMENT**
   Advancing Early Endpoints and Novel Evidence Pathways
4. **COMPLEX BIOMARKERS**
   Harmonizing Measurement and AI Applications

# 2024 By the Numbers

**5,000+**
meeting attendees

**12**
working groups

**10**
roundtables & public meetings

**15,000+**
publication downloads

**400+**
working group participants

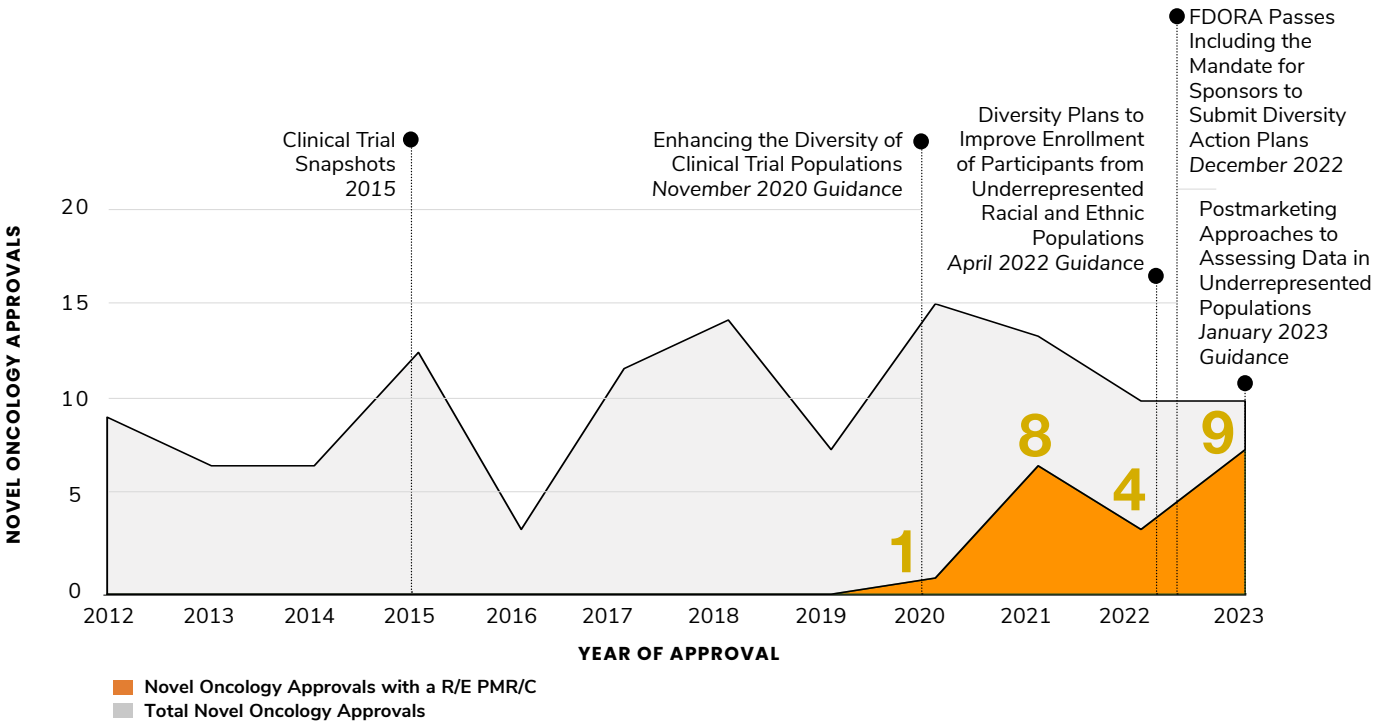
**13**
white papers, abstracts, posters, & publications

## Patient-Focused Drug Development: Enhancing Representativeness and Equity in Clinical Trials

Representation in clinical trials based on characteristics such as race, ethnicity, gender, age, and performance status is critical for generating generalizable data on the safety and efficacy of oncology therapies. In 2024, *Friends* assessed the current state of diversity in clinical trials and postmarketing studies and identified recommendations to support development of diversity plans and trial designs for future oncology clinical trials.

Specifically, *Friends* conducted an analysis of novel oncology drugs approved between 2012-2023 and observed an increasing number of postmarketing requirements or commitments (PMR/C) to conduct additional studies in populations that reflect the racial and ethnic (R/E) diversity of the U.S. population. This increase in R/E PMR/Cs aligns with recent policies, such as the diversity action plan (DAP) mandate that requires trial sponsors to consider diversity and representation when planning, designing, and conducting clinical trials intended to support regulatory decisions. To help implement the DAP mandate, *Friends* conducted a survey to assess sponsors' current approaches to establishing enrollment goals, strategies for recruiting, enrolling, and retaining diverse patients, and future needs for ensuring representative trials. The findings from the survey informed a discussion document that was shared during our public meeting in February.

**The recent increase in novel oncology approvals with a R/E PMR/C aligned with the timing of key guidance and policy on enhancing diversity in clinical trials.**




**Source:** Friends of Cancer Research. Drug Development Dashboard: Postmarketing Requirements (PMR) and Commitments (PMC) for Novel Oncology Therapies.

Ensuring diverse participation also requires addressing broader barriers to trial accessibility. Flexibilities implemented during the COVID-19 pandemic, such as remote monitoring and telemedicine, have demonstrated potential to reduce patient burden and improve accessibility. *Friends*' joint analysis with the American Society of Clinical Oncology (ASCO) of 83 trials found that these flexibilities had no major impacts on reported protocol deviations and other trial quality metrics, underscoring their feasibility for broader, longer-term implementation. By integrating these lessons into trial designs, sponsors can enhance both diversity and inclusivity in clinical research.

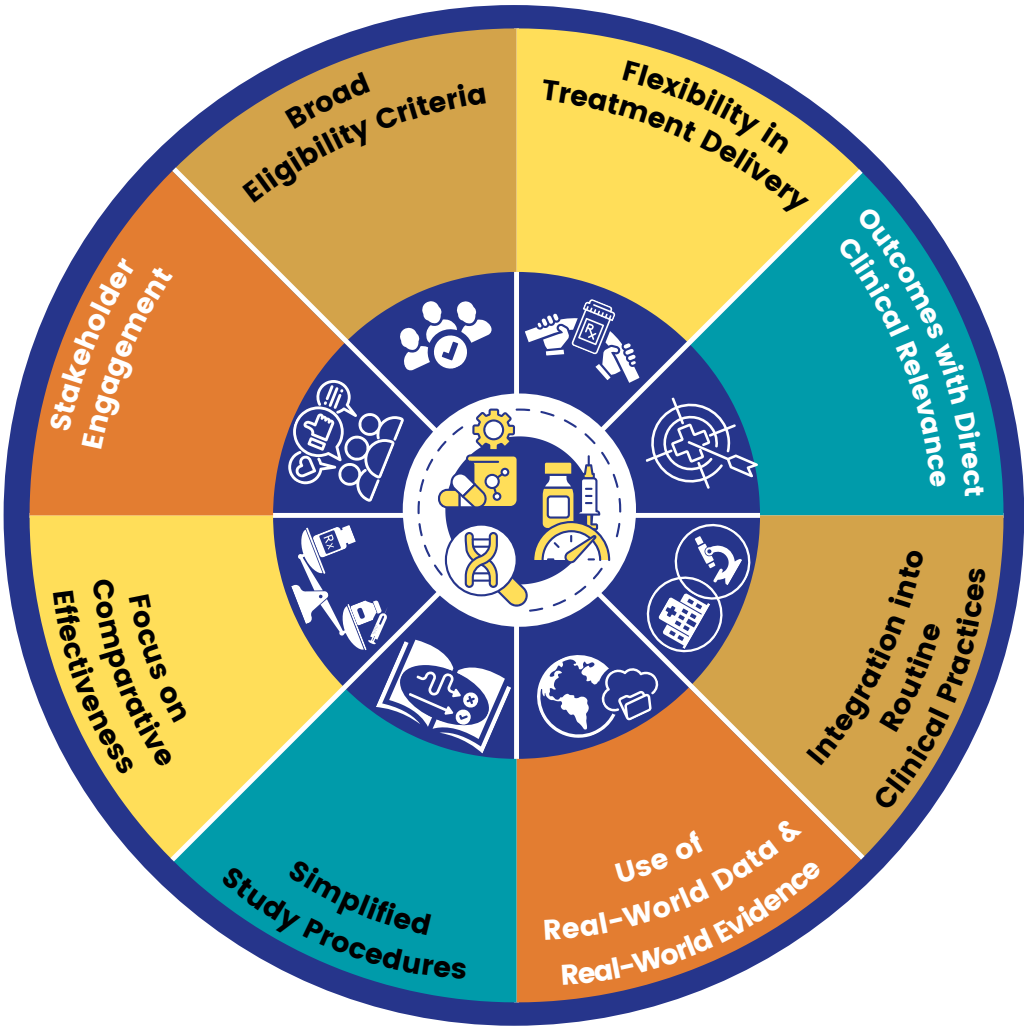
**GUIDANCE DOCUMENTS**

- **Cancer Clinical Trial Eligibility Criteria: Washout Periods and Concomitant Medications,** Draft Guidance, April 25, 2024
- **Cancer Clinical Trial Eligibility Criteria: Performance Status,** Draft Guidance**,** April 25, 2024
- **Cancer Clinical Trial Eligibility Criteria: Laboratory Values,** Draft Guidance, April 25, 2024
- **Diversity Action Plans to Improve Enrollment of Participants from Underrepresented Populations in Clinical Studies,** Draft Guidance, June 26, 2024
- **Considerations for Generating Clinical Evidence from Oncology Multiregional Clinical Development Programs,** Draft Guidance, September 17, 2024
- **Patient-Focused Drug Development: Core Patient-Reported Outcomes in Cancer Clinical Trials,** Final Guidance, October 18, 2024

## Real-World Evidence: Leveraging RWD for Insights on Real-World Response

Traditional clinical trials provide critical evidence to support regulatory decision-making, but they can be resource intensive, restrictive, and not always reflective of real-world clinical practice. Clinical trial flexibilities— such as the use of alternative data sources like real-world data (RWD) from electronic health records (EHRs) and claims data—and the incorporation of pragmatic elements, including streamlined safety data collection and telemedicine, offer patient-centered approaches to help bridge the gap between clinical trials and real-world clinical practice. In 2024, *Friends* identified additional opportunities to use these approaches to support oncology drug development. Building on our RWE Portfolio, *Friends* published findings from our rw-Response Project, which demonstrated that information on tumor response to therapy is consistently captured in clinician notes across various RWD sources.

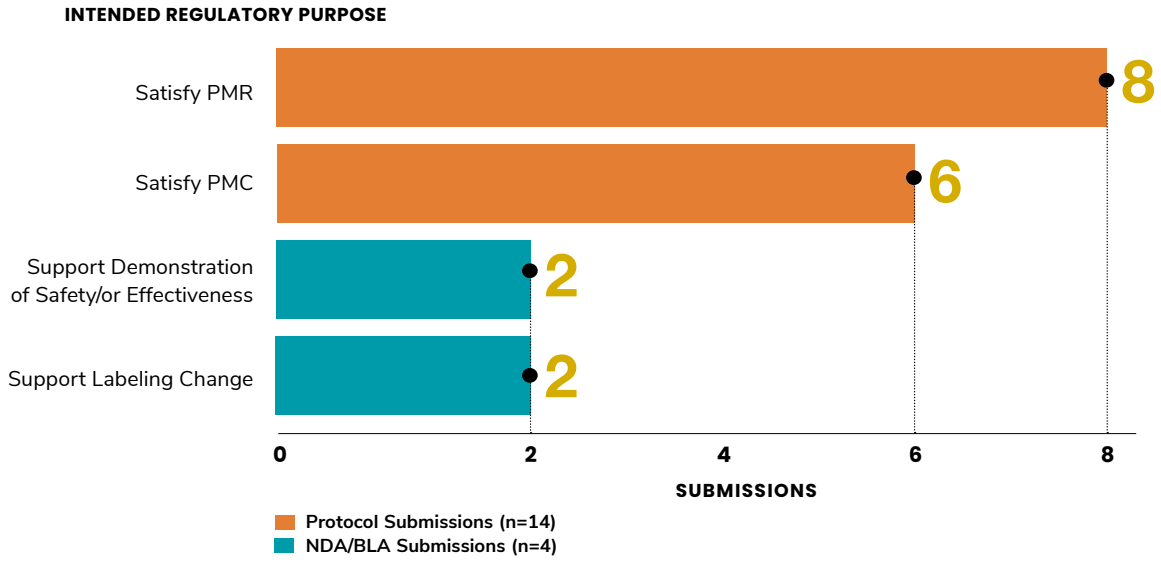
These findings show how RWD can be used to assess response to treatment, providing insights into how RWE may be used to supplement clinical trial data to support regulatory decision-making. This work aligns with recent final FDA guidance on Assessing EHRs and Medical Claims Data to Support Regulatory Decision-Making.




**Hallmarks of pragmatic clinical trials for application in hybrid designs.** These features are indicative of a pragmatic approach, although not all need to be present for a trial to be classified as pragmatic. The inclusion of these elements can vary, reflecting a spectrum rather than an all-or-nothing requirement.

**Source:** Stewart MD, et al. Bridging research and practice: enhancing regulatory decisions with pragmatic clinical trials in oncology. See page 55.

**Of the submissions containing RWD received by CBER & CDER in 2023, most were intended to satisfy a PMR/C.**




**Source:** U.S. FDA. Real-World Evidence Submissions to the Center for Biologics Evaluation and Research & the Center for Drug Evaluation and Research, 2023.

Incorporating pragmatic approaches, such as using insights from RWD, will depend on the research question and trial setting, which may not be appropriate for every trial objective. *Friends* 2024 Annual Meeting discussions and accompanying white paper outlined considerations for incorporating pragmatic elements in clinical trials and presented several use cases where pragmatic approaches could be implemented in postmarket clinical trials. Pragmatic elements facilitate patient-centered, streamlined, and timely clinical trials, offering potential to enhance accessibility and efficiency. Future efforts should be directed towards refining best practices for incorporating pragmatic elements and clarifying regulatory expectations around their acceptability.

**GUIDANCE DOCUMENTS**

- **Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products,** Draft Guidance, March 18, 2024
- **Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products,** Final Guidance, July 25, 2024
- **Integrating Randomized Controlled Trials for Drug and Biological Products Into Routine Clinical Practice,** Draft Guidance, September 17, 2024

## Innovative Drug Development: Advancing Early Endpoints and Novel Evidence Pathways

**In 2024, FDA Approved**

**3** **New Cell and Gene Therapies for Cancer**

**An increase** compared to previous years.

Novel therapeutic classes continue to improve patient survival, challenging current trial designs and drug development paradigms. As our understanding of cancer and new therapeutic classes progress, new regulatory approaches and evidence-generation strategies will be needed to facilitate timely approvals and access to therapies. Cell and gene therapies exemplify an area poised for such a regulatory paradigm shift. In 2024, the FDA approved three novel cell and gene therapies for the treatment of cancer, including the first ever cell therapy indicated for treatment of solid tumors. These therapies are uniquely complex and require resource intensive development and manufacturing. In certain instances, it may be appropriate to consider how data can be safely extrapolated across products or product versions to support efficient development of next generation products. In 2024, *Friends* partnered with the Parker Institute for Cancer Immunotherapy to host a public meeting focused on considerations for reimagining cell therapy trials and treatments for patients. The meeting identified scientific and operational challenges to current cell and gene therapy development approaches, and key opinion leaders discussed innovative approaches to reducing both the cost and duration of manufacturing and clinical development.

The use of early endpoints is critical to support early access to lifesaving therapies. However, benefits observed using intermediate endpoints can conflict with interim data on overall survival (OS) used to provide insights into efficacy and safety, making it challenging to interpret the interim OS data alongside intermediate endpoints, such as progression-free survival (PFS) that are frequently used to support expedited approvals. Interim OS data are often immature and may be unreliable due to small event counts and limited follow-up. *Friends* 2024 Annual Meeting included discussions around strategies for improving the analysis and interpretation of interim OS data in oncology clinical trials, to ensure these data can continue to provide information to support expedited approvals.

To further support timely and innovative oncology drug development, *Friends* continued advancing research and frameworks that support use of ctDNA as an early endpoint in oncology clinical trials, including through our multi-year research partnership, the ctMoniTR Project.

**GUIDANCE DOCUMENTS**

- **Considerations for the Development of CAR-T Cell Products,** Final Guidance, January 30, 2024
- **Use of ctDNA for Early-Stage Solid Tumor Drug Development,** Final Guidance, November 27, 2024
- **Accelerated Approval – Expedited Program for Serious Conditions**, Draft Guidance, December 6, 2024

## Complex Biomarkers: Harmonizing Measurement and AI Applications

In oncology, biomarker assessments provide critical information to clinicians and patients that guide treatment decisions, monitor disease progression, and evaluate patient prognosis. Diagnostic tests used to detect and quantify biomarkers, present in either tumors or blood samples, are increasingly sophisticated and complex, incorporating capabilities such as artificial intelligence and machine learning (AI/ML) to enhance biomarker assessment.

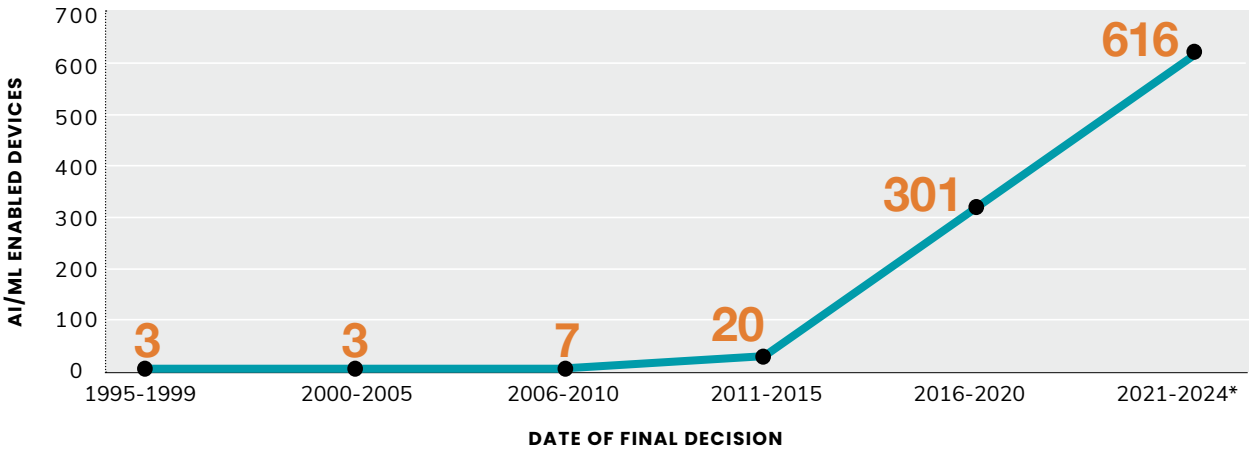
*Friends* has led several initiatives to assess variability and identify areas for alignment across assays for several biomarkers used in oncology research and care (see project spotlight on the next page). In 2024, *Friends* presented and published results from the final phase of our HRD Harmonization Project. The observed variability among HRD assays reiterated the opportunity for improved alignment on defining and measuring HRD, as inconsistent results can impact treatment decisions for patients and providers.

In 2024, *Friends* launched our Digital PATH Project to assess variability across computational pathology models evaluating HER2 biomarker status in breast cancer samples. Initial results shared in 2024 demonstrated variability in HER2 scoring across AI-models, with greater variability across algorithms when scoring HER2 low tumors. As diagnostic testing technologies continue to evolve, such as through integration of AI-enabled approaches, the development of unique validation approaches will be needed to ensure that assays remain accurate, reliable, and suitable for clinical use.

**GUIDANCE DOCUMENTS & POLICIES**

- **Laboratory Developed Tests,** Final Rule, May 6, 2024

**The number of AI/ML-enabled devices has rapidly increased in recent years.**




*Data up to 8/07/24.
**Source:** U.S. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices.

# Diagnostics Harmonization Portfolio

## Goal

Friends of Cancer Research (*Friends*) Diagnostics Harmonization Portfolio aims to assess variability across different diagnostic tests, inform approaches to support harmonized test performance, and support policy frameworks to facilitate the development of reliable and consistent tests for patients.

## Background

Diagnostic tests play a crucial role in cancer research and care, including identifying new drug targets, informing treatment decisions, and monitoring treatment efficacy and patient outcomes. An improved understanding of human biology has enabled the development of more effective and safer targeted treatments for patients with cancer. Diagnostic tests identify the presence or absence of biomarkers, which are measurable characteristics of tumors, to select which patients may benefit from specific targeted therapies.

Differences in regulatory pathways and methodological approaches for measuring and reporting biomarker status can lead to inconsistent results among tests assessing the same biomarker. This lack of harmonization can make it challenging for patients and providers to navigate test options and interpret test results.

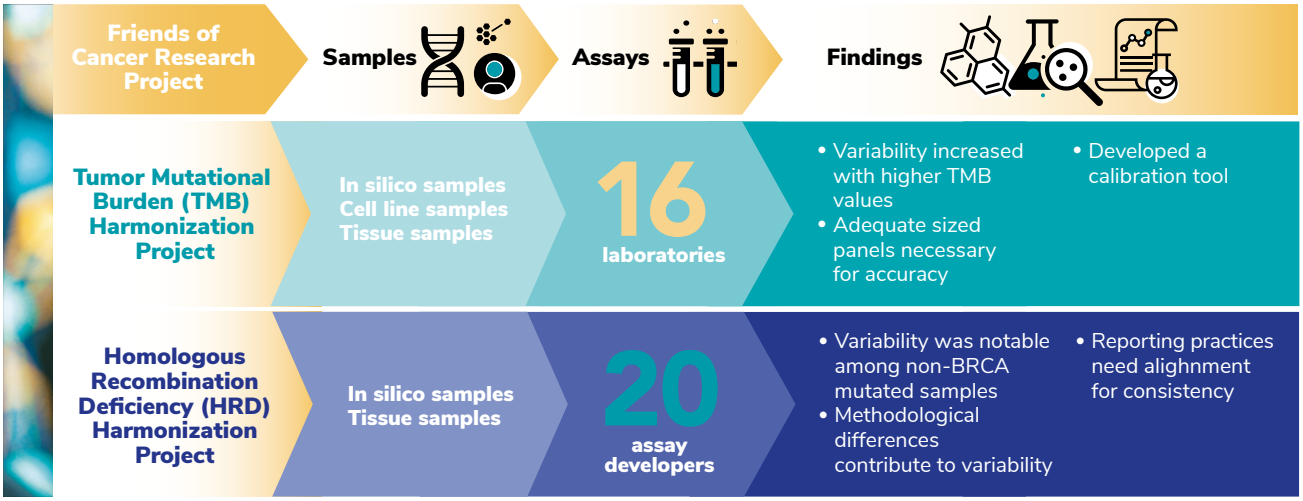
## Solutions

In a series of collaborations with diagnostic developers, patient advocates, government officials, pathologists, clinical researchers, and drug developers, we assess the comparability of biomarker measurements across different tests analyzing a common dataset, compare results, and identify opportunities to improve alignment. The goal is to assess variability and concordance of biomarker measurements across different diagnostic platforms. This can help ensure that patients have accurate results no matter which test they receive.

## Policy Implications

In recent years, both the U.S. Food and Drug Administration (FDA) and Congress have worked to modernize the regulatory approach to improve diagnostic test oversight.

*Friends*' Diagnostics Harmonization Portfolio provides important scientific evidence to demonstrate the need for improved oversight to ensure consistency and inform future policy considerations, including opportunities for promoting transparency in test performance. *Friends* remains deeply engaged in these policy discussions and leverages results from our research partnerships to develop innovative approaches for assuring test accuracy.



# Digital PATH Project

## Digital and Computational Pathology Tool Harmonization Project



### Background

Traditionally, pathologists examine tissue obtained through a biopsy to diagnose cancer, determine the type and stage, and identify biomarkers that may indicate a tumor's response to certain treatments and other clinical insights. Digital pathology enables innovative approaches to these assessments through the scanning and digitization of tissue slides for storage, viewing, and analysis. Tissue analysis of these digitized slides can include the use of computational pathology platforms with artificial intelligence (AI)/ machine learning (ML) algorithms.

### Approach

*Friends* convened a multi-stakeholder working group to evaluate HER2, an important biomarker for identifying patients for HER2 targeting agents, in >1,100 whole slide images of breast cancer tissue across ten computational pathology platforms to understand the level of variability in biomarker assessment across platforms and factors impacting variability.

### Impact

Digital and computational pathology platforms have the potential to provide greater accuracy, reproducibility, and standardization of pathology features, expedite diagnosis or pathological scoring, establish new biomarkers, and identify and select the appropriate patients for treatments—all of which can contribute to improving patient outcomes. Supporting the robust development of these platforms and identifying potential sources of variability will help to inform future use and advancements in technology to deliver more precise patient care. Without *Friends*' coordination and support from collaborative sponsors, alignment across assays in evaluation and reporting of biomarkers critical to patient care may not occur, resulting in continued challenges for patients and providers in interpreting test results.

# Patient-Focused Drug Development:

## Enhancing Representativeness and Equity in Clinical Trials

# Impact of the COVID-19 pandemic mitigation strategies on cancer treatment trials: A meta-analysis of industry and NCI studies.

Joseph M. Unger, Hillary Andrews, Laura A. Levit, Brittany Avin McKelvey, Mark Stewart, Emily Van Meter Dressler, Keith T. Flaherty, Peter Fredette, Lee Jones, Therica Miller, Adedayo A. Onitilo, Timil Patel, Suanna Steeby Bruinooge, Elizabeth Garrett-Mayer, Caroline Schenkel; SWOG Statistics and Data Management Center/Fred Hutchinson Cancer Research Center, Seattle, WA; Friends of Cancer Research, Washington, DC; ASCO, Alexandria, VA; Wake Forest University School of Medicine, Winston-Salem, NC; Massachusetts General Hospital, Boston, MA; EQRx, Cambridge, MA; Fight Colorectal Cancer, Arlington, VA; Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY; Marshfield Clinic-Weston Center, Marshfield, WI; U.S. Food and Drug Administration, Silver Spring, MD; Children's National Hospital, Washington, DC

**Background:** ASCO and Friends of Cancer Research established a task force to evaluate trial mitigation strategies allowed by US regulators during the COVID-19 pandemic, including the use of telemedicine and remote monitoring. We report the results of a meta-analysis quantifying the impact of these strategies on quality metrics and the recovery time to pre-COVID levels. **Methods:** We invited 41 sponsors with active US cancer treatment trials from January 2015-May 2022 to contribute deidentified trial-level aggregate data on major protocol deviations (PDs), dropouts, severe or worse toxicity (CTCAE Grade 3-5), and enrollment. We examined outcomes as proportions of participants at-risk during the pre-COVID, initial wave (IW), initial recovery (IR), and secondary recovery (SR) assessment times (Table). Multi-level beta-regression analyses were adjusted for trial phase ("early", phases I, II, or I/II, vs. "late", phase III) with study and sponsor as random effects. Indicator variables were used for post-COVID time periods with pre-COVID as the reference. **Results:** Ten sponsors (9 industry and 1 NCI Cooperative Group) contributed 82 evaluable studies: 63 early and 19 late phase trials. Among the 15,679 participants, enrollment odds decreased 64% in the IW and 45% in the IR but recovered to approximately pre-COVID levels by the SR (Table). Major PDs, dropouts, and severe or worse toxicity all had lower incidence in the IW compared to pre-COVID; these outcomes were also less frequent in IR ($p<.05$ for each), but not in the SR ($p>.05$ for each) compared to pre-COVID. **Conclusions:** Large declines in enrollment rates during the IW rebounded to pre-COVID levels by 2021-2022. We found steep reductions in the rates of reported occurrence of major PDs, dropouts, and severe or worse toxicity during the initial outbreak, which also recovered to pre-COVID levels by 2021-2022. Findings suggest pandemic-related procedural flexibility did not lead to increased reporting of PDs or dropouts and highlight how use of mitigation strategies likely corresponded with the temporary disruption to trial conduct during the pandemic's peak. Sponsors could consider broader adaptation of trial flexibilities moving forward. Research Sponsor: None.

| | Pre-COVID (Jan 2017-Feb 2020) | Initial Wave (IW) (Mar-Apr 2020) | | Initial Recovery (IR) (May-Dec 2020) | | Secondary Recovery (SR) (Jan 2021-Dec 2022) | |
|---|---|---|---|---|---|---|---|
| **Endpoint** | % | %[1] | OR (95% CI) | %[1] | OR (95% CI) | %[1] | OR (95% CI) |
| **Mean monthly enrollment[2]** | 69.0 | 48.2 | 0.36 (0.21-0.63) | 59.1 | 0.55 (0.32-0.96) | 64.5 | 0.90 (0.52-1.62) |
| **Major PDs** | 14.8 | 8.2 | 0.37 (0.26-0.52) | 11.5 | 0.65 (0.47-0.90) | 12.7 | 0.72 (0.52-1.00) |
| **Dropouts** | 37.8 | 8.3 | 0.09 (0.06-0.13) | 24.7 | 0.44 (0.32-0.59) | 31.2 | 0.80 (0.58-1.10) |
| **Severe or worse toxicity** | 35.2 | 18.4 | 0.35 (0.26-0.48) | 28.0 | 0.65 (0.49-0.87) | 31.3 | 0.83 (0.61-1.13) |

[1]Among trials with both pre-COVID and follow-up data;
[2]Standardized 0-100 as proportion of maximum study-level monthly enrollment across time periods.

# An Evaluation of Novel Oncology Approvals with a PMR/C for Assessing Data in Racial and Ethnic Populations Underrepresented in Premarket Clinical Trials

Grace Collins[1], Hillary S. Andrews[1], Brittany McKelvey[1], Carrigan Rice[2], Jeff D. Allen[1], and Mark D. Stewart[1]

## ABSTRACT

Clinical trials supporting oncology drug approvals frequently underrepresent diverse racial and ethnic populations. Recent policies have focused on ensuring premarket clinical trials are more inclusive and representative of racial and ethnic diversity in the general U.S. population or intended patient population; however, recent U.S. Food and Drug Administration (FDA) guidance on postmarketing approaches to collecting data in underrepresented populations demonstrates that, in certain circumstances, postmarketing requirements and/or commitments (PMR/Cs) may be issued to conduct more representative studies if there are remaining questions about safety or efficacy. This analysis demonstrates that prior to 2020, no drugs had PMR/Cs to further characterize use in a more representative population, and in the last 3 years, more than half of novel oncology approvals have had such a PMR/C (21/40, 53%). In addition, this analysis helps to identify characteristics, such as single-arm pivotal trial design, U.S. enrollment, and results of safety subgroup analyses based on race and ethnicity, that may contribute to decisions to issue a PMR/C to conduct a study that is more representative of the racial and ethnic diversity of the U.S. or intended patient population. These results can inform efforts to improve premarket clinical trials to ensure they are representative and able to characterize use in any patient who may need the drug.

## Introduction

Ensuring clinical trials include patients who represent the demographics of the intended treatment population is necessary to support a comprehensive assessment of a therapy's benefits and risks and inform optimal use. Recognizing this imperative, the federal government, including Congress and the U.S. Food and Drug Administration (FDA), has implemented policies to ensure clinical trials are more inclusive and representative of the general U.S. population and intended patient population based on characteristics such as race, ethnicity, sex, gender, socioeconomic status, and age. As part of these efforts, the FDA released several guidance documents detailing standards for collecting race and ethnicity data, launched clinical trial snapshots to improve transparency in reporting demographic variables for pivotal clinical trials, and, with their new authority provided under the Food and Drug Omnibus Reform Act (FDORA) of 2022, established the requirement for sponsors to submit diversity action plans which must include representative enrollment targets for registrational trials supporting new drug applications and biologics license applications (1–4).

The focus of these efforts is primarily on improving representation in premarket clinical trials; however, FDA may exercise regulatory flexibility and issue postmarketing requirements and/or commitments (PMR/Cs) to further evaluate a drug in situations where additional information is needed to ensure the safety, effectiveness, and quality of the drug after approval. Recent FDA guidance has clarified that in certain circumstances, PMR/Cs may be used to further characterize safety or efficacy in populations underrepresented in premarket clinical trials (5). Patients from certain racial and ethnic populations are disproportionately underrepresented in clinical research, including premarket pivotal trials supporting oncology approvals. To address this gap, in some cases, approvals were accompanied by PMR/Cs requesting additional data in underrepresented racial and ethnic populations (R/E PMR/C; refs. 6–8). This analysis evaluates the trends in oncology PMR/C studies from 2012 to 2023 focused on greater representation of race and ethnicity, assesses characteristics of pivotal trials that may prompt these studies to provide insights into FDA expectations, and informs the development of more effective diversity plans and premarket study designs.

## Methodology

Using publicly available information on Drugs@FDA and the Center for Biologics Evaluation and Research's (CBER) page of Licensed Biological Products with Supporting Documents, a list of all PMR/Cs issued in the original approval letters of novel oncology drugs and biologics (referred to herein as "drugs") approved by the FDA between 2012 and 2023 was compiled (9, 10). PMR/C descriptions, statutes under which the PMR/Cs were issued, and final report due dates were also collected from the original approval letters. To identify PMR/Cs that emphasized the need for data in a representative population, a key word search of PMR/C descriptions was conducted using the terms "represent," "racial," "race,"

[1]Friends of Cancer Research, Washington, District of Columbia. [2]Georgetown University, Washington, District of Columbia.

**Corresponding Author:** Grace Collins, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC 20009. E-mail: gcollins@focr.org

Clin Cancer Res 2024;XX:XX–XX

**doi:** 10.1158/1078-0432.CCR-24-0852

## Translational Relevance

The federal government, including Congress and the U.S. Food and Drug Administration (FDA), has implemented policies to ensure clinical trials are more inclusive and representative of the general U.S. population and intended patient population based on characteristics such as race, ethnicity, sex, gender, socioeconomic status, and age. In some cases, additional studies may be necessary to further characterize safety or efficacy and FDA may exercise regulatory flexibility by issuing postmarketing requirements and/or commitments (PMR/Cs) specifying study should be conducted in a representative population. This analysis helps to identify characteristics that may contribute to decisions to issue a PMR/C to conduct a study that is more representative of the racial and ethnic diversity of the U.S. or intended patient population and can inform efforts to improve premarket clinical trials to ensure they are representative and able to characterize use in any patient who may need the drug.

"ethnic," and "ethnicity," and matching PMR/C descriptions were reviewed to identify those that specifically addressed representation based on race and ethnicity (R/E PMR/Cs).

Publicly available review documents on Drugs@FDA and CBER's web page of Licensed Biological Products with Supporting Documents were used to collect pivotal trial characteristics including trial size, design (single-arm vs. randomized), patient demographics by geographic region, race, and ethnicity and to assess results from efficacy, safety, dosing, and pharmacokinetics (PK) subgroup analyses based on race and ethnicity. Review documents were also used to identify instances where FDA commented whether the trial(s) supporting approval were representative of the racial and ethnic makeup of the intended patient population and/or U.S. population.

## Results

Between January 1, 2012, and December 31, 2023, the FDA approved 144 novel oncology drugs and issued PMR/Cs to 98% (141/144), 22 of which had a PMR/C specifying the need for additional data in a population representative of the racial and ethnic diversity of the U.S. and/or intended patient population (R/E PMR/C). Most drugs with a R/E PMR/C (21/22, 96%) were approved from 2021 to 2023. Prior to 2021 (2012–2020), only 1 of the 104 drugs approved had a R/E PMR/C (**Fig. 1**). Further analyses focused on the 40 novel oncology drug approvals between 2021 and 2023, of which 53% (21/40) had a R/E PMR/C (**Fig. 1**).

A total of 25 R/E PMR/Cs were issued across 21 drugs (three drugs had two R/E PMR/Cs). Most were PMCs subject to annual reporting requirements under section 506B of the Federal Food, Drug, and Cosmetic Act (13/25, 52%). The remaining 12 were Accelerated Approval (AA) requirements (8/25, 32%) or PMRs under 505(o) of the Federal Food, Drug, and Cosmetic Act (4/25, 16%) that assess serious risks or safety signals related to the use of the drug.

### Approval characteristics

More than half of the drugs approved from 2021 to 2023 were AAs (21/40, 53%). Of these, the majority had a R/E PMR/C (16/21, 76%). In contrast, only 26% (5/19) of drugs approved through the traditional approval pathway had a R/E PMR/C (**Table 1**). Many approvals in this time frame received Priority Review (37/40, 93%), Orphan Drug Designation (29/40, 72%), Breakthrough Therapy Designation (22/40, 55%), and/or were first-in-class approvals (13/40, 33%)—this was true for both drugs with and without a R/E PMR/C.

Approvals supported by a single-arm pivotal trial ($n = 31$) were more likely to have a R/E PMR/C than approvals supported by a randomized controlled trial (RCT; $n = 11$). Two approvals were supported by one randomized trial and one single-arm trial and are included in the counts for both trial design categories. While 58% (18/31) of approvals supported by a single-arm trial had a R/E PMR/C, only 36% (4/11) of approvals supported by a RCT had a R/E PMR/C. Likely because of these differences in trial design (i.e., single-arm trials are on average smaller than RCTs), trials supporting approvals with a R/E PMR/C were smaller on average than trials supporting approvals without a R/E PMR/C (201 vs. 267). Notably, all AA drugs were supported by single-arm trials.

When considering cancer type, R/E PMR/Cs were assigned to drugs indicated for non–Hodgkin lymphoma (NHL; 6/8, 75% of drugs approved for the indication received R/E PMR/Cs), non–small cell lung cancer (NSCLC; 3/6, 50%), multiple myeloma (5/6, 83%), gynecologic cancers (1/3, 33%), gastrointestinal cancers (3/3, 100%), breast cancer (2/2, 100%), and nasopharyngeal cancer (1/1, 100%; **Table 1**).

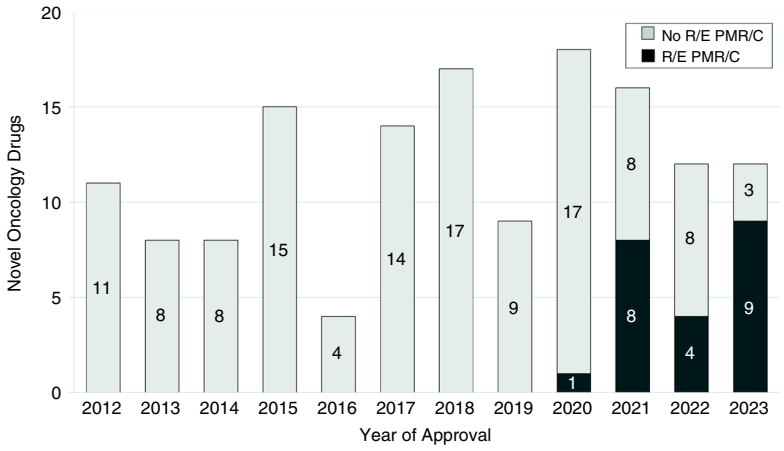
### Pivotal trial demographics

The 40 drugs approved in 2021 to 2023 were supported by data from 43 pivotal clinical trials. For each pivotal trial, patient demographics (race, ethnicity, and U.S. enrollment) of the primary efficacy population used to support approval were compiled. Pivotal trial demographics for indications that had both drugs with and without a R/E PMR/C (i.e., NHL, NSCLC, multiple myeloma, and gynecologic cancers) were further evaluated to identify any differences in pivotal trial patient demographics for drugs with a R/E PMR/C ($n = 15$) versus those without a R/E PMR/C ($n = 8$; **Table 2**).

Most pivotal trials reported race using the categories White (23/23, 100%), Black or African American (21/23, 91%), and Asian (23/23, 100%). On average, pivotal trials supporting approvals with a R/E PMR/C enrolled 75.3% (36.8%–96%) White, 3.7% (0%–7.3%) Black or African American, and 12.5% (0.6%–59.6%) Asian patients compared to 74.4% (34%–95%) White, 6.6% (1%–18%) Black or African American, and 12.9% (1%–59%) Asian patients for drugs without a R/E PMR/C (**Table 2**). Other categories included for reporting of race were American Indian or Alaska Native (AI/AN; reported for 10/23 trials, 43%), Native Hawaiian or Other Pacific Islander (NHPI; 8/23, 35%), Multiple (5/23, 22%), and Other (13/23, 57%). Patients who identify as AI/AN or NHPI were on average less represented in pivotal trials for drugs with a R/E PMR/C [0.9% (0%–1.7%) and 0.5% (0.35%–1%)] than in pivotal trials for drugs without a R/E PMR/C [1.6% (0.7%–4.2%) and 1.3% (1%–1.6%)]. In addition to the reported categories, 78% (18/23) of pivotal trials reported some degree of missing data for race (Race Not Reported/Unknown/Missing). On average, pivotal trials for drugs with a R/E PMR/C had more missing data for race reporting compared to drugs without a R/E PMR/C (7.6% vs. 3.3%; **Table 2**).

Reporting of ethnicity (i.e., Hispanic/Latinx, Non-Hispanic/Latinx, Other Ethnicity, and Not Reported/Missing/Unknown) was also variable across pivotal trials (**Table 2**). Patients identifying as Hispanic/Latinx were less represented in pivotal trials supporting drugs

**Figure 1.**
Novel oncology approvals with and without a R/E PMR/C over time. Most drugs with a R/E PMR/C were approved from 2021 to 2023. Prior to 2021, only 1 of the 104 drugs approved had a R/E PMR/C.




with a R/E PMR/C [4.8% (0%–9.7%)] compared to drugs without a R/E PMR/C [10.5% (2.7%–35%)] (**Table 2**). Most pivotal trials had some degree of missing ethnicity data (18/23, 78%) and pivotal trials for drugs with a R/E PMR/C on average had a greater percentage of patients without ethnicity reported compared to drugs without a R/E PMR/C [14% (0%–21.9%) vs. 3.5% (0%–6.2%)] (**Table 2**).

**Table 1.** Characteristics of novel oncology drugs approved by the FDA between 2021 and 2023.

| | Number of approvals *n* (%) | | |
|---|---|---|---|
| | **R/E PMR/C** | **No R/E PMR/C** | **Total** |
| **Total** | 21 (53) | 19 (47) | 40 |
| **Year of approval** | | | |
| 2021 | 8 (50) | 8 (50) | 16 |
| 2022 | 4 (33) | 8 (67) | 12 |
| 2023 | 9 (75) | 3 (25) | 12 |
| **Approval pathway** | | | |
| Accelerated approval | 16 (76) | 5 (24) | 21 |
| Traditional approval | 5 (26) | 14 (74) | 19 |
| **Drug designation** | | | |
| Priority review | 21 (57) | 16 (43) | 37 |
| Orphan drug | 15 (52) | 14 (48) | 29 |
| Breakthrough therapy | 11 (50) | 11 (50) | 22 |
| First in class | 7 (54) | 6 (46) | 13 |
| **Trial characteristics** | | | |
| Avg. primary efficacy population in pivotal trial (range) | 201 (69–708) | 267 (45–831) | 231 (45–831) |
| Single-arm[a] | 18 (58) | 13 (42) | 31 |
| Randomized[a] | 4 (36) | 7 (64) | 11 |
| **Indication** | | | |
| Non-Hodgkin lymphoma[b] | 6 (75) | 2 (25) | 8 |
| Non-small cell lung cancer | 3 (50) | 3 (50) | 6 |
| Multiple myeloma | 5 (83) | 1 (17) | 6 |
| Leukemias[b] | — | 4 (100) | 4 |
| Skin cancer | — | 3 (100) | 3 |
| Genitourinary cancers[c] | — | 3 (100) | 3 |
| Gynecologic cancers[d] | 1 (33) | 2 (67) | 3 |
| Gastrointestinal cancers[e] | 3 (100) | — | 3 |
| Breast cancer | 2 (100) | — | 2 |
| Liver cancer | — | 1 (100) | 1 |
| VHL-related cancers | — | 1 (100) | 1 |
| Nasopharyngeal cancer | 1 (100) | — | 1 |

[a]Two approvals were supported by one randomized trial and one single-arm trial and are included in the counts for both trial design categories.
[b]One drug was approved for leukemia and non-Hodgkin lymphoma and is counted for both indications under No R/E PMR/C.
[c]Bladder, prostate, and renal cell carcinoma.
[d]Cervical, uterine, and ovarian (ovarian had R/E PMR).
[e]Colorectal and cholangiocarcinoma.

**Table 2.** Consistency in reporting of demographic groups in pivotal trials for select indications of novel drug approvals between 2021 and 2023.[a]

| Demographic category | | n Drugs (%) | | | Average % of trial population (Range) | |
|---|---|---|---|---|---|---|
| | | Total drugs (*n* = 23) | R/E PMR/C (*n* = 15) | No R/E PMR/C (*n* = 8) | R/E PMR/C | No R/E PMR/C |
| **Race** | White | 23 | 15 (100) | 8 (100) | 75.3% (36.8–96) | 74.4% (34–95) |
| | Black or African American | 21 | 14 (93) | 7 (88) | 3.7% (0–7.3) | 6.6% (1–18) |
| | Asian | 23 | 15 (100) | 8 (100) | 12.5% (0.6–59.6) | 12.9% (1–59) |
| | American Indian or Alaska Native | 10 | 5 (33) | 5 (63) | 0.9% (0–1.7) | 1.6% (0.7–4.2) |
| | Native Hawaiian or Other Pacific Islander | 8 | 6 (40) | 2 (25) | 0.5% (0.35–1) | 1.3% (1–1.6) |
| | Multiple Races | 5 | 2 (13) | 3 (38) | 0.5% (0.35–0.7) | 0.5% (0–1) |
| | Other Race | 13 | 9 (60) | 4 (50) | 3.1% (0–7) | 1.6% (0–3.6) |
| | Race Not Reported/Unknown/Missing | 18 | 12 (80) | 6 (75) | 7.6% (0.6–19.5) | 3.3% (0–9) |
| **Ethnicity** | Hispanic/Latinx | 19 | 13 (87) | 6 (75) | 4.8% (0–9.7) | 10.5% (2.7–35) |
| | Non-Hispanic/Latinx | 20 | 13 (87) | 7 (88) | 81.8% (15–99) | 88% (61–97) |
| | Ethnicity Not Reported/Unknown/Missing | 18 | 12 (80) | 6 (75) | 14% (0–21.9) | 3.5%(0–6.2) |
| | Other Ethnicity | 1 | 1 (7) | 0 (0) | 3.9% | — |
| **Geographic Region** | U.S. Population[b] | 20 | 13 (87) | 7 (88) | 40.9% (12–73) | 66.2% (14.9–100) |

[a]Indications for which there was at least one approval with a R/E PMR/C and at least one approval without a R/E PMR/C (i.e., NSCLC, NHL, multiple myeloma, gynecologic cancers).
[b]Two trials (one supporting a drug with a R/E PMR/C and one supporting a drug without a R/E PMR/C) reported enrollment for North America, which included patients enrolled in the U.S. and Canada.

For reporting by geographic region, on average, pivotal trials for drugs with a R/E PMR/C, reported fewer U.S. patients than pivotal trials for drugs without a R/E PMR/C. Pivotal trials supporting approvals without a R/E PMR/C were made up on average of 66% (14.9%–100%) U.S. patients, compared to an average of 41% (12%–73%) U.S. patients for approvals with a R/E PMR/C (**Table 2**).

### Subgroup analyses

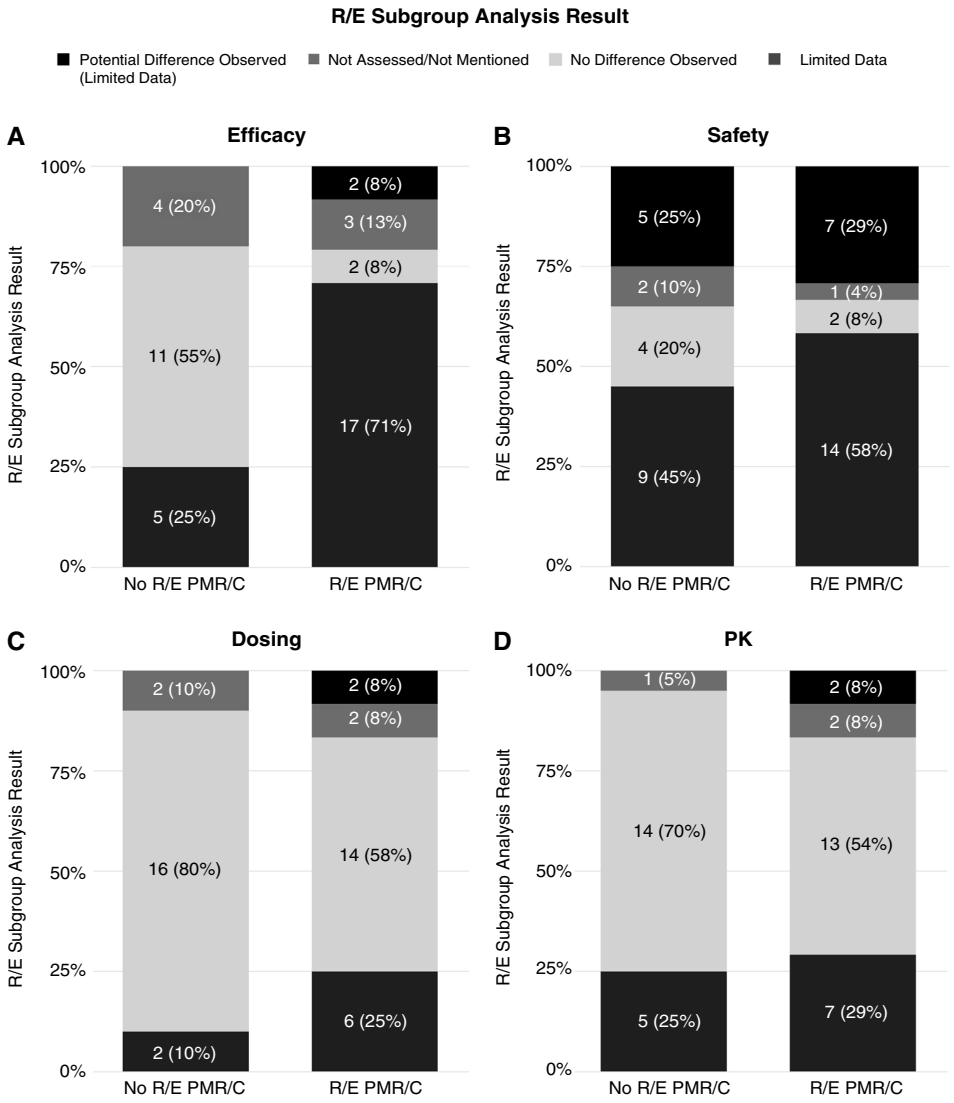
Efficacy, safety, PK, and dosing subgroup analyses based on race and ethnicity were assessed for all 40 drugs approved from 2021 to 2023. Three review documents for drugs with a R/E PMR/C and one review for drugs without a R/E PMR/C included subgroup analyses for multiple cohorts or clinical trials. Each subgroup analysis was assessed leading to differences in the total subgroup analyses for drugs with a R/E PMR/C ($n = 24$) and without a R/E PMR/C ($n = 20$) compared with the total number of drugs in each group ($n = 21$ and $n = 19$, respectively). Review documents frequently noted insufficient data or small samples sizes which limited the ability to draw conclusions from subgroup analyses. More reviews for drugs with a R/E PMR/C noted there were limited data for efficacy (17/24, 71%), safety (14/24, 58%), PK (7/24, 29%), and dosing (6/24, 25%) subgroup analyses based on race/ethnicity compared reviews for drugs without a R/E PMR/C (**Fig. 2A–D**). A greater number of subgroup analyses for drugs with a R/E PMR/C indicated there was a potential difference across racial and/or ethnic subgroups than for drugs without a R/E PMR/C, although these differences may not have been clinically meaningful (**Fig. 2A–D**). Differences were observed most in safety subgroup analyses for 29% (7/24) of drugs with a R/E PMR/C and 25% (5/20) of drugs without a R/E PMR/C (**Fig. 2B**).

## Discussion

Clinical trials supporting oncology drug approvals frequently underrepresent diverse racial and ethnic populations. Prior to 2020, the FDA had not explicitly requested postmarket studies for novel oncology drugs be conducted in more representative populations to address this gap. However, since 2020, there has been a notable increase in the number of novel oncology approvals with a PMR/C specifying the study should be conducted in a population more representative of the racial and ethnic diversity of the U.S. population or intended use population. In the last 3 years (2021–2023), more than half of novel oncology approvals (21/40, 53%) were issued such a PMR/C. The timing of this increase in R/E PMR/Cs aligned with the release of FDA guidance and legislation aimed at enhancing representation in clinical trials. This analysis assessed differences between approval characteristics for drugs with a R/E PMR/C and those without a R/E PMR/C to identify factors that may contribute to a drug being issued a R/E PMR/C.

Most notably, drug approvals supported by a single-arm trial and/or that were accelerated approvals were more likely to receive a R/E PMR/C. All accelerated approvals during this period were based on a single-arm trial. Additionally, on average, trials for drugs with a R/E PMR/C reported more missing data for both race and ethnicity compared to trials without a R/E PMR/C. For indications in which there were drugs with a R/E PMR/C and without a R/E PMR/C (i.e., NHL, NSCLC, multiple myeloma, and gynecologic cancers) pivotal trial demographics were consistent with previous analyses of pivotal trial demographics that show a lack of representation of certain racial and ethnic groups, in particular, Black or African American and Hispanic/Latinx populations (7).

**Figure 2.**
Subgroup analyses based on race and ethnicity for all novel drugs approved between 2021 and 2023. Results of subgroup analyses based on race/ethnicity for (**A**) efficacy, (**B**) safety, (**C**) dosing, and (**D**) pharmacokinetics (PK) were compared for drugs without a R/E PMR/C.




Despite efforts to standardize and improve transparency in reporting of demographic data for novel drugs (e.g., Clinical Trial Snapshots) there continues to be variable reporting of race and ethnicity data. For reporting of race, FDA recommends sponsors include, at a minimum, options to select AI/AN, Asian, Black or African American, NHPI and White, as well as directions clarifying that one or more of these may be selected (2). This analysis showed some reviews did not report enrollment for all recommended race and ethnicity categories, which may have been due to the timing of data collection for the clinical trials or because there were no patients who identified as the missing races or ethnicities; however, in some cases zero was reported if this was true. White and Asian were the only two demographic categories assessed that were reported in all review documents. Most trials included in this analysis were conducted globally, so the missing data may be related to global restrictions related to protected characteristics, such as race and ethnicity, that prevent collection and reporting of these data (11, 12). Among the 20 trials that included information on U.S. enrollment, trials for drugs with a R/E PMR/C had an average U.S. enrollment of 40.9% (12%–73%). In contrast, trials for drugs without a R/E PMR/C reported a higher average U.S. enrollment rate of 66.2% (14.9%–100%). Notably, there were four drugs supported by trials that enrolled only U.S. patients and these drugs did not receive a R/E PMR/C.

Another factor influencing whether FDA issues a R/E PMR/C is the presence of potential safety signals observed in subgroup analyses. Review documents posted in support of drug approvals often include subgroup analyses that assess whether there are differences in efficacy, safety, PK, and dosing based on intrinsic factors such as race and ethnicity. Review documents for drugs with a R/E PMR/C and those without a R/E PMR/C indicated subgroup analyses are often limited by small sample sizes and/or are incorporated as secondary/exploratory analyses, limiting the ability to draw conclusions about whether meaningful differences in efficacy, safety,

dosing, or PK exist across racial and ethnic subgroups. Despite these limitations, subgroup analyses based on race and ethnicity occasionally identify potential signals of a difference. Safety subgroup analyses most often indicated a potential signal of a difference ($n = 12$). Of note, there were five instances where a safety subgroup analysis by race and/or ethnicity indicated a potential difference and no R/E PMR/C was issued.

The FDA's commitment to ensure oncology trials reflect the diversity of the U.S. population should be considered with the evolving nature of oncology clinical trials, including enhanced research capabilities outside the U.S. Global clinical trial practices and regulatory expectations influence diversity efforts and reciprocally, U.S. diversity initiatives affect clinical trials worldwide. It will be important to know how many patients should be enrolled in the U.S. to be considered representative of the U.S. population, as well as how FDA considers patients enrolled outside of the U.S. when determining whether a trial is adequately representative. One PMC provided a benchmark for the number of patients to be enrolled in the U.S. requesting the sponsor, "conduct a clinical trial enrolling a total sample size of 100 patients in the U.S. and Canada, that includes a sufficient representation of patients in racial and ethnic minority subgroups and is reflective of the U.S. population of patients with nasopharyngeal carcinoma (NPC)." Another PMC specified a benchmark for what FDA considers appropriate representation of Black or African American patients for a trial in multiple myeloma stating, "Ensure that the representation of the African American subpopulation in the studies is reflective of the Black population in the geographical location/country. Therefore, approximately 15% of the population that is enrolled from the US should comprise of African Americans." Additionally, there is a need for more consistent and transparent reporting of race and ethnicity. Varying global definitions and restrictions on reporting for protected characteristics, in particular race and ethnicity, can impact the ability to assess whether a trial is adequately representative. Efforts to have more uniform and complete reporting across clinical trials will help to accurately assess the extent of underrepresentation, inform effective strategies for enhancing diversity and inclusion in trials, and help with assessing progress toward equitable clinical trials. Approaches for addressing these gaps in data and global coordination around efforts to enhance clinical trial diversity can be useful for ensuring data necessary to assess differences across subgroups are available.

Finally, studies should be designed to assess effectiveness and safety in different populations. In several cases, reviews noted the subgroup analyses were limited due to small sample sizes. In certain instances, such as when there are known differences between populations or disparities in the disease burden, oversampling patients from underrepresented subgroups may be warranted to improve the ability to assess whether meaningful differences across subgroups exist. As such, it will be important to evaluate when to power studies adequately to assess whether meaningful differences in safety and efficacy exist across racial and ethnic or other underrepresented subgroups. When enrolling patients is challenging due to the rarity of the disease, regulatory flexibility can be applied to answer these questions following approval through PMR/Cs to avoid delaying patient access to promising therapies.

### Looking forward

Greater representation in clinical trials can improve the generalizability of results to real-world patients who may need the drug, inform optimal use, and ensure equitable access and benefit from novel therapies. Our findings demonstrate that PMR/Cs are being used to ensure stakeholders conduct representative studies following approval; however, it is too early to assess whether this results in timely studies in more representative populations or if the results of these studies have an impact on labeling. While these types of PMR/Cs may be appropriate in certain instances, additional work is necessary to ensure premarket oncology clinical trials are representative of all patients. To achieve this goal, it will be essential to implement strategies that address barriers to participation and promote inclusivity. Although this analysis focuses on representation of racial and ethnic groups, the mandates in FDORA, such as the Diversity Action Plan requirement, also aim to address underrepresentation of populations based on characteristics such as sex, gender, age, geographic region, and other social determinants of health. As part of the effort to fulfill these mandates, trial sponsors have described strategies for setting representative enrollment goals, enrolling patients to meet these goals, and retaining these patients on clinical trials. For example, sponsors now incorporate diversity considerations into their trial planning processes by using epidemiologic data sources to assess disease burden and inform site selection. Trial sponsors are also working to lower barriers to participation when designing trials by incorporating decentralized elements and removing overly restrictive eligibility criteria. To address the financial burdens of participating, sponsors are offering reimbursement for travel, time, meals, and other costs incurred during trial participation. In addition to trial-specific efforts, sponsors are working to cultivate sustained partnerships with diverse communities and providers to build trust, better understand and address patient and provider needs, and more effectively communicate about trial opportunities (13). To continue this progress, it is critical that the future clinical trials, especially clinical trials conducted in the premarket setting used to support drug approvals, prioritize and enhance representation of diverse patient populations.

## Authors' Disclosures

No disclosures were reported.


## Acknowledgments

This work was supported by Friends of Cancer Research.

Received March 19, 2024; revised May 21, 2024; accepted June 12, 2024; published first XXX XX, 2024.


## References


1. U.S. Food and Drug Administration. Draft guidance for industry: diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials. [cited 2023 Dec 8]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations.
2. U.S. Food and Drug Administration. Guidance for industry: collection of race and ethnicity data in clinical trials. [cited 2024 Mar 15]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/collection-race-and-ethnicity-data-clinical-trials-and-clinical-studies-fda-regulated-medical.

3. Center for Drug Evaluation & Research (CDER). Drug trials snapshots. [cited 2023 Dec 11]. Available from: https://www.fda.gov/drugs/drug-approvals-and-databases/drug-trials-snapshots.
4. Congress.gov. Text - H.R.2617—117th Congress. (2021–2022): consolidated appropriations act, 2023. (public law No: 117–328). [cited 2023 Dec 11]. Available from: https://www.congress.gov/bill/117th-congress/house-bill/2617/text.
5. U.S. Food and Drug Administration. Draft guidance for industry: post-marketing approaches to obtain data on populations underrepresented in clinical trials for drugs and biological products. [cited 2023 Dec 8]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarketing-approaches-obtain-data-populations-underrepresented-clinical-trials-drugs-and.
6. Aldrighetti CM, Niemierko A, Van Allen E, Willers H, Kamran SC. Racial and ethnic disparities among participants in precision oncology clinical studies. JAMA Netw Open 2021;4:e2133205.
7. Loree JM, Anand S, Dasari A, Unger JM, Gothwal A, Ellis LM, et al. Disparity of race reporting and representation in clinical trials leading to cancer drug approvals from 2008 to 2018. JAMA Oncol 2019;5:e191870.
8. Kim J, Kester R, Blumenthal G. Clinical trial diversity in oncology: FDA takes action with post–marketing requirements or commitments. The Oncologist 2022;27:993–7.
9. U.S. Food and Drug Administration. Drugs@FDA. [cited 2023 Dec 8]. Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm.
10. U.S. Food and Drug Administration. Licensed biological products with supporting documents. [cited 2023 Dec 8]. Available from: https://www.fda.gov/vaccines-blood-biologics/licensed-biological-products-supporting-documents.
11. European Commission. Directorate general for justice and consumers, Farkas L. analysis and comparative review of equality data collection practices in the European Union: data collection in the field of ethnicity. [cited 2023 Dec 15]. Available from: https://data.europa.eu/doi/10.2838/447194.
12. Institut National de la Statistique et des Études Économiques. Ethnic-based statistics. [cited 2023 Dec 8]. Available from: https://www.insee.fr/en/information/2388586.
13. Friends of Cancer Research. Improving equity in oncology clinical trials: challenges and strategies for setting diversity enrollment goals. [cited 2024 Mar 15]. Available from: https://friendsofcancerresearch.org/wp-content/uploads/Improving_Equity_in_Oncology_Clinical_Trials.pdf.

FRIENDS of CANCER RESEARCH

# Improving Equity in Oncology Clinical Trials: Challenges and Strategies for Setting Diversity Enrollment Goals

2024 Discussion Document

## Table of Contents

Acknowledgements ..... 2

Executive Summary ..... 3

Introduction & Background ..... 4

Applying FDA Guidance ..... 5

Data Challenges ..... 10

Measures to Achieve Enrollment Goals ..... 13

Conclusion ..... 16

References ..... 18

## Acknowledgements

We thank the following organizations for contributing insights and expertise to help inform the development of this discussion document.

Amgen
Astellas
AstraZeneca
BeiGene
Bristol Myers Squibb
Certara
Daiichi Sankyo
Eisai
EMD Serono
Flatiron Health
Genentech
Genmab
Gilead Sciences
Gritstone Bio
GSK
Johnson & Johnson Innovative Medicine
Loxo@Lilly
Merck & Co., Inc.
Mirati, A Bristol Myers Squibb Company
Novartis AG
Parexel
Pfizer Inc.
Pharmacyclics, An AbbVie Company
Takeda
Tempus AI, Inc.
Sanofi
Xencor

## Executive Summary

Roughly 8% of adult patients with cancer participate in clinical trials, and among these participants, there has historically been a lack of diversity.[1] This underrepresentation impacts the generalizability of trial results and perpetuates health inequities. Recognizing this, the U.S. Food and Drug Administration (FDA) announced guidance documents and initiatives, including Project Equity, to encourage efforts to improve representativeness in oncology drug development. The recent Food and Drug Omnibus Reform Act (FDORA) further solidified this effort by requiring Diversity Action Plans for Phase III clinical trials, which must consider race, ethnicity, age, and sex/gender.

A survey by Friends of Cancer Research evaluated how 23 drug sponsors are implementing FDA guidance and FDORA mandates. Findings show that key steps include characterizing the population of patients with a particular disease, identifying and analyzing diverse data sources, and setting enrollment goals. This discussion document details two proposals to address challenges in data availability and integration:

1. Central Repository for Biomarker Data in U.S./Canada: Create a centralized, nationally representative repository for cancer biomarker data, inclusive of race and ethnicity data.
2. Collaborative Data Consolidation Efforts: Consolidate and harmonize data sources to bridge gaps in data coverage and establish standards for collecting and reporting race and ethnicity variables.

In addition to establishing enrollment goals, diversity plans must incorporate patient-directed measures, community engagement, workforce-directed measures, and trial design considerations to achieve these goals. Measures include:

- Building trust and partnerships in diverse communities.
- Lowering barriers to participation by addressing financial burdens and removing restrictive eligibility criteria.
- Intentional site selection focusing on health centers serving diverse populations.

Sponsors should implement mechanisms to track progress towards achieving enrollment goals, enabling them to reassess and adapt strategies, as necessary. This discussion document emphasizes that to achieve the shared goal of more inclusive and representative patient populations in clinical trials, a multifaceted approach involving robust data analysis, strategic planning, community engagement, and inclusive trial practices is required.

## Introduction & Background

It is estimated that around 8% of adult patients with cancer participate in clinical trials in the United States (U.S.).[1–3] Further, of those patients participating in clinical trials, there is often a lack of diversity and representativeness of the overall patient population with the disease.[4] Patients from certain racial and ethnic populations are frequently underrepresented in oncology clinical trials, and clinical research more broadly, despite these patients experiencing a disproportionate burden of disease for several cancer types, such as breast, prostate, and multiple myeloma.[5,6] This lack of inclusion and representativeness in current clinical trials may hinder the generalizability of results to the intended patient population, contribute to existing health inequities, and limit the potential to personalize treatment to meet the unique needs of various patient populations. Actions to improve inclusion of patients from underrepresented racial and ethnic groups in clinical trials are necessary to achieve the broader goals of providing equitable healthcare and reducing health disparities.[7] The U.S. Food and Drug Administration (FDA) recognizes the need for improved representativeness in clinical trials as evidenced by the release of guidance documents, policies, public meetings, and other initiatives such as Project Equity. These efforts provide recommended standards for race and ethnicity data collection and reporting in clinical trials, provide considerations for broadening eligibility criteria to be more inclusive, and describe measures that can lower barriers to participation.[8–14]

In April 2022, the FDA released a new draft guidance document titled "Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials," recommending trial sponsors develop Race and Ethnicity Diversity Plans for most investigational medical products.[14] The guidance states these Diversity Plans should include representative enrollment goals for historically underrepresented racial and ethnic populations in the U.S., including Black or African American, Hispanic/Latinx, Indigenous and Native American, Asian, Native Hawaiian and Other Pacific Islander populations, and strategies for enrolling and retaining these patients on clinical trials.

In December 2022, Congress passed the Food and Drug Omnibus Reform Act (FDORA), which includes several provisions to enhance diversity and representativeness in clinical trials.[15] Among these, the law codified components of the April 2022 guidance and expanded requirements to consider age and sex/gender in Diversity Action Plans for Phase III or other pivotal clinical trials for investigational medical products, which will be represented in an updated guidance document from the FDA. As outlined in the law, drug sponsors must submit Diversity Action Plans to the FDA by the time they submit the study protocol for any Phase III or other pivotal drug study, excluding bioavailability or bioequivalence studies and include enrollment goals, rationale supporting these goals, and a strategy for achieving these goals.

Considering these recommendations and requirements, drug sponsors have mobilized their teams to implement measures that support the development, submission, and implementation of diversity planning as part of the clinical development process.[16] Friends of Cancer Research (*Friends*)

surveyed 27 drug sponsors, as well as data aggregators, to assess specific approaches used to implement the recommendations and requirements outlined in the April 2022 draft guidance and FDORA and identify strategies for enhancing adoption of FDA recommendations. The following questions were posed:

1. How are sponsors applying FDA guidance and recent FDORA mandates to set diversity enrollment goals for oncology clinical studies? (e.g., U.S. enrollees and/or international)?
2. What key factors do sponsors consider when identifying data sources (e.g., the Surveillance, Epidemiology, and End Results [SEER] data, EHRs, past clinical trials, registries, etc.) for establishing benchmarks for population diversity (i.e., by race, ethnicity, sex, age group)? What are known strengths and limitations associated with different data sources?
3. What types of data are difficult or not feasible to obtain from data sources? What approaches are used to access information/data that may not be readily accessible/available (e.g., information on biomarker-defined subgroups)? What are the limitations of this information and what approaches can be taken to overcome them?

In addition to responses to these questions, the goal was to better understand measures to achieve enrollment goals.

## Applying FDA Guidance

Since the release of the April 2022 draft guidance (and prior to its release in some instances), and in anticipation of the FDORA Diversity Action Plan requirement coming into effect, sponsors have been proactively implementing steps to achieve greater diversity in trials and voluntarily submitting diversity plans to the FDA. Between April 2022 and April 2023, 42 sponsors submitted 76 diversity plans across 40 oncologic indications to the Center for Drug Evaluation and Research's (CDER) oncology divisions.[17] Although the currently available guidance focuses on diversity plans for enrolling underrepresented racial and ethnic groups, sponsors indicated they are also incorporating considerations such as age and sex/gender, and social determinants of health (SDoH) to ensure enrollment goals represent the disease burden across patient populations. As the community works toward implementing concepts in the guidance document and law, it is important to align on the goals and intentions of these requirements, which can include 1) ensuring a sufficient number of patients enroll and are retained from underrepresented racial and ethnic groups to determine whether demographic factors impact safety and efficacy; 2) having global studies that represent disease epidemiology and are generalizable to the intended use population in the U.S.; and 3) enrolling as many underrepresented U.S. patients into clinical trials as possible to provide equitable opportunities to participate in oncology clinical research and thereby reduce disparities in oncology health outcomes across diverse, U.S. patient groups with cancer. The specific intention of including more diverse patients in a clinical trial will have implications on the trial design, enrollment goal setting, and statistical analysis plan.

## Data Analysis and Goal Setting

One of the key steps towards achieving more representative enrollment in clinical trials is characterizing the population affected by a particular disease, including who it affects, where these patients live, and understanding treatment and testing patterns. However, there is no standardized source for these data or aligned methodology for capturing them, and therefore, goal setting can be a complicated task because it may require synthesis of data from disparate sources. Various data sources that include information on U.S. population-level demographic variables and disease incidence and prevalence need to be identified and analyzed. Using these data, sponsors set enrollment goals for U.S. enrollment in global studies and provide the rationale for these goals.

Setting enrollment goals for achieving diversity is part of broader U.S. initiatives to have more diverse patients represented in clinical trials and clarify expectations around the proportion of patients who should be enrolled from the U.S. This includes understanding what constitutes a clinical trial population that is representative of the epidemiology and demographics of U.S. patients for whom a therapy is intended to be used. Many sponsors run global development programs and conduct clinical trials spanning multiple countries including the U.S. Therefore, sponsors may monitor enrollment outside of the U.S. and identify ways to tailor enrollment from these countries to supplement U.S. enrollment goals.

However, there is often a lack of robust, decentralized data sources to obtain similar information about diversity outside of the U.S., which is largely due to incomplete collection, varying definitions of race and ethnicity, and laws that prevent collecting this information in some countries.[18] Additionally, lived experiences among similar racial and ethnic groups often vary from one country to another. As a result, it is not clear whether or how enrollment of diverse patients from outside the U.S. would be considered when determining whether diversity requirements are fulfilled, and importantly, it also does not address the issue of unequal access to or participation in clinical trials within the U.S.

## Data Sources Used for Enrollment Goal Setting

Sponsors use a variety of data sources to inform clinical trial enrollment goals. Data sources are selected based on several key factors, including the availability, completeness, and granularity of variables in the data source; the timing of data collection; the representativeness of the data source; accessibility of the data; and the expected reliability and acceptability of the data source by the FDA.

Data of interest include clinical factors such as histology, stage, co-morbidities, and relevant biomarkers, demographic and non-demographic variables such as age, sex assigned at birth, race, ethnicity, and SDoH such as income, education level, healthcare utilization, and insurance status. Table 1 outlines a range of examples for select data sources used by sponsors to set enrollment goals, which generally fall into four categories:

1. Epidemiological Data Sources are publicly accessible and useful for understanding disease incidence, prevalence, survival, mortality, and other clinical information stratified by variables such as age, race and ethnicity, and geographic area. However, sources like these lack

granularity about clinical variables such as biomarker status and prior therapies. In addition, disease progression data can be lacking and there can be time lags in data reporting of one to several years for certain data elements leading to potential misalignment with other data sources. *Examples include the SEER Database and Centers for Disease Control and Prevention (CDC) Databases.*

2. Past Clinical Trial Data & Literature provide helpful estimates for benchmarking based on prior clinical trial enrollment or evidence from retrospective database studies, prospective observational studies, and multicenter studies. There may also be patient-level data on clinical outcomes and clinical variables of interest. However, these data sources may not represent the current standard of care and historically lack representation of patients from diverse racial and ethnic groups. Additionally, race, ethnicity, and other socio-demographic data tend to be poorly and inconsistently documented across published clinical trials. *Examples include sponsor-specific data/records from past clinical trials, literature reviews, and meta-analyses of past clinical trials.*
3. Real-world Data (RWD) Sources contain patient-level data and capture a range of treatment information and other clinical data. RWD sources also have a variety of ways in which to capture and define race and ethnicity. These data sources often lack SDoH information and have variability in available demographic information, though some efforts have been made to leverage other data points to establish SDoH variables.[19] Additionally, these data sources may not always represent the general population. There can also be inconsistency in the quality and completeness of data across patients and RWD sources, and thus, the quality and robustness of the data source will need to be evaluated. *Examples include Electronic Health Records (EHRs), healthcare medical claims data, and disease-specific registries.*
4. Genomic Databases/Repositories are the most readily available source of biomarker data, but they have inconsistent categorization of race and ethnicity data and include largely patients served by large academic medical centers and patients of European descent. *Examples include The Cancer Genome Atlas (TCGA) Program, American Association for Cancer Research (AACR) Project GENIE, and other clinical-genomic databases.*

Table 1. Select Examples of Data Sources Used to Inform Enrollment Goals.

| Type | Specific Examples | Usages | Strengths | Limitations |
|---|---|---|---|---|
| Publicly Available Epidemiological Data | SEER Data, Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR) | Understanding disease prevalence, incidence, and demographic estimates | Representative of the civilian U.S. population, low missing data on demographics, publicly available/easily accessible | Limited data on rare indications, some data may not be up-to-date given timing/cadence of data collection and publishing of results, incomplete race and ethnicity data, and if included are limited to existing race and ethnicity categories, and these sources may also not include specific biomarker data and treatment history |
| Real-World Data (RWD) | Healthcare medical claims data (e.g., CMS, private insurers), Electronic Health Records (EHRs) data (e.g., Flatiron Health, Tempus, Cerner Health), other clinical provider databases | Identifying geographies and areas for high-incidence diseases, evaluating treatment effectiveness and safety | Patient-level data, captures treatment and genomic information, demographic data, near real-time | Capture of insured populations with healthcare access, missing data on race and ethnicity in some cases, variable data quality/completeness, variable reporting on biomarker data |
| Literature Searches | PubMed, clinicaltrials.gov | Informing historical disease characteristics, benchmarking goals | Provides insights into historical disease characteristics | May not provide current and representative data, selection bias in clinical trial data, biomarker data variability |
| Government Sources and Surveys | U.S. Census Data, National Health Information Survey (NHIS), National Health and Nutrition Examination Survey (NHANES), Medical Expenditure Panel Survey (MEPS) | Estimating the total population at risk, assessing disease burden in health-specific surveys, demographics, and treatment patterns | Large sample size, relevant for benchmarking population-level diversity, low missingness for race/ethnicity data | Some data may not be up-to-date given timing/cadence of data collection and publishing of results, no specific cancer staging, tumor size, etc. |

| Type | Specific Examples | Usages | Strengths | Limitations |
|---|---|---|---|---|
| Registries | Disease-specific advocacy group registries | Aggregating data from a targeted patient population (disease-specific), offering network, and disease-specific demographics | Can provide specific demographic data, helpful for certain rare hereditary cancers/diseases | Availability may vary (i.e., disease and/or population specific), not universally accessible, potential for bias in who participates, may not include all demographic information, small size |
| Past Clinical Trial Data | Clinical trial data, Meta-analyses combining data from past clinical trials including non-interventional (or observational) studies | Estimating placebo rates, understanding historical enrollment rates by demographic group, benchmarking, estimating biomarker prevalence prior to treatment | Comprehensive data from multiple trials, robust estimations, outcomes data | Limited to data from previous trials that may lack representation of diverse racial and ethnic groups, may not reflect current disease landscape, selection bias in clinical trial data, race/ethnicity reporting can vary |
| Genomic Databases | AACR Project GENIE, TCGA, other clinical-genomic databases (e.g., Flatiron Health-FMI database, Tempus) | Understanding biomarker prevalence | Most readily available source of genomic information | Limited institutions contributing (e.g., academic medical centers), potential for bias in patients included, inconsistent categorization of race and ethnicity. Historic lack of testing in minority populations |

## Data Challenges

Sponsors must leverage multiple heterogeneous data sources to set enrollment goals, which can be resource intensive and complex. As described in **Table 1,** different data sources have different uses, strengths, and weaknesses. Combining multiple sources can help to collect all necessary data; however, when using this approach to inform representative enrollment goals and develop strategies to provide more equitable opportunities for participation in clinical trials to meet these goals, it can be difficult to synthesize data across sources, particularly where data may be overlapping or are inconsistent. In addition to the resources required and methodology needed for aggregating data across sources, several gaps were identified in the existing data, including several variables of interest that are challenging to obtain even when combining data:

- **Availability of clinical variables across data sources** – With the increasing number of approvals for targeted therapies that rely on biomarker testing to select eligible patients, there is a need to improve approaches and sources for assessing biomarker frequency stratified by race and ethnicity.[20] In the absence of sufficient biomarker data by demographic group, especially for novel biomarkers, one approach is to assume that the frequency of the biomarker is equal across racial and ethnic groups thereby setting enrollment goals based on the overall prevalence of the cancer, irrespective of biomarker status. Assumptions like this may be difficult to test or validate with a high degree of confidence, within a particular clinical context. These assumptions can also lead to underestimating disease burden in underrepresented patients, and in turn, underestimating enrollment targets. Thus, it is difficult to project whether a group may be underrepresented in a trial due to gaps in data for certain populations. Other clinical variables that are difficult to obtain in some data sources include the stage of cancer, tumor histology, line of therapy, and prior therapies.
- **Availability of non-clinical or non-medical variables** – SDoH variables such as income, education level, built environments, and social and community contexts are often not routinely collected or reported likely due to a lack of standards for how this information should be collected.[21] Some national data or U.S. Census data may have information related to SDoH, but these data are not specific to cancers of interest. However, these data can provide essential information for assessing barriers to, and facilitators of, patients' participation in a clinical trial and how lived experiences influence health outcomes. In addition, a lack of standards and reporting limit availability of data on the inclusion of sexual orientation and gender identity and people with disabilities in clinical trials.[22]
- **Variable definitions for race and ethnicity data** – The lack of appropriate and consistent definitions for race and ethnicity impacts data collection, analysis, and reporting. The granularity in which race and ethnicity data are collected also can vary. More granular reporting of Asian populations (e.g., Korean, Japanese, and Chinese), and Hispanic and Latinx populations (e.g., Spanish vs. Central/South American, Mexican, Argentinian, etc.) may be necessary in some instances, and proposals are in place to implement a separate Middle

Eastern or North African (MENA) race category to better distinguish individuals of MENA descent who are frequently reported within the White race category. Currently, the Office of Management and Budget (OMB) is reviewing proposals to update existing race and ethnicity categories.[23] These efforts are important because broad categories such as White, Asian, Black, Hispanic, and non-Hispanic are frequently used, and there may be instances where individuals may not identify with any of these broadly characterized groups or some individuals may be multiracial. This in turn can result in inaccurate data, thereby skewing the ability to establish and measure enrollment goals.

- Ex-U.S. data – Obtaining robust data from outside the U.S. presents another challenge. Definitions for race and ethnicity not only vary in the U.S. but also vary globally, and there can be legal restrictions in reporting and sharing this type of patient-level data in certain countries. This poses challenges when clinical trials conducted with the intent to support U.S. submissions include ex-U.S. sites that lack race and ethnicity data. The lack of unified race and ethnicity data outside the U.S. makes it difficult to set enrollment goals for ex-U.S. populations and to estimate the number of patients from underrepresented racial and ethnic populations that could be enrolled outside the U.S. to help meet enrollment goals outlined by sponsors in their diversity action plans. Though, even with more unified race and ethnicity data availability outside the U.S., how these data would apply to achieving enrollment goals in diversity plans in support of U.S. regulatory submissions is unclear. Additionally, while sponsors set current enrollment goals with a U.S. focus, there is also a need to enroll clinical trial populations representative of the entire population who will benefit from use of the drug, particularly targeting patients in countries outside of the U.S. where there is an intent to apply for approval or market the drug. Sponsors will also need to consider variations in lived experiences among racial and ethnic groups in different countries if leveraging ex-U.S. populations to meet U.S. enrollment goals.

## Addressing Data Challenges

More work is needed to address these noted data challenges and several forward-leaning proposals have been identified to address different aspects of data integration. Specifically, statistical considerations will also need be considered for combining data sources to strengthen and minimize limitations of any one data source.[24] Additionally, clarity around the level of acceptable uncertainty in estimating the characteristics of the intended patient population with respect to setting enrollment goals and how the relevance/reliability of the data used to set enrollment goals will be considered.

### Proposal 1: Central Repository for Biomarker Data in U.S./Canada

One approach to addressing the availability of clinical variables, particularly for biomarker data, is to create a centralized repository that is nationally representative for multiple cancer types, includes race and ethnicity data, and is broadly accessible. The BROAD Institute's Repository for prostate cancer serves as one example.[25] These efforts aim to identify sources of variability across race and ethnicity groups, improve reporting standards, and promote alignment on definitions for race and ethnicity. This initiative may also highlight inequities in

biomarker testing, and thus, highlight the need for resources and strategies to close the gap in biomarker testing across race and ethnicity groups.[20]

Proposal 2: Collaborative Data Consolidation Efforts

To address the challenge of needing to combine multiple data sources, efforts are needed to consolidate and harmonize curated data sources. Collaborative data consolidation bridges gaps in data coverage, providing a more comprehensive and accessible dataset for informed enrollment goal decisions. To assist with consolidating multiple data sources, standards will be necessary.

Government agencies are currently seeking proposals to establish standards for collecting and reporting race and ethnicity variables to enhance primary data collection.[23] The SEER program recently implemented changes to race and Hispanic ethnicity towards five mutually exclusive categories: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian/Pacific/Islander, Non-Hispanic American Indian/Alaska Native, and Hispanic.[26] Additionally, legislative and policy efforts may be necessary to enhance how race is assigned by the U.S. Census and reduce the misclassification of race in cancer data. By working collectively, stakeholders can share the responsibility of data collection and integration, making it a more efficient and cost-effective endeavor.

Additionally, broad initiatives to improve reporting standards and promote alignment of definitions for race and ethnicity are needed. This can include using Clinical Data Interchange Standards Consortium (CDISC) standards as a framework for the structured exchange of clinical and non-clinical research data to ensure that race and ethnicity data are collected and reported in a consistent manner across different studies and data sources. Efforts to create and pilot updated eCRFs can help to ensure that race and ethnicity data are consistent and comparable across different countries and regions. This not only helps in achieving uniformity but also facilitates setting more precise enrollment goals and ensures that the representation of diverse racial and ethnic groups is accurate.

Sponsors recognize the need for efficient data integration to inform enrollment goals and have responded by investing in data integration solutions, establishing partnerships with data providers, and developing standardized data collection protocols. Several strategies may help alleviate data challenges. The use of standard electronic case report forms (eCRF) within the U.S. to capture patient demographic information consistently across all clinical trials can help ensure a more holistic view of representativeness across a sponsor's clinical development programs as well as across different sponsors. Additionally, providing definitions and guidance on race and ethnicity categories in eCRF instructions can help improve the accuracy of data. Epidemiologists should be part of diversity planning strategic discussions and several important questions are noted to consider:

- Is the occurrence of disease higher/lower in specific underrepresented racial and ethnic populations?
- How does the age distribution of disease vary across racial and ethnic groups?

- Do disease characteristics including biomarkers differ across racial and ethnic groups such that we need to show efficacy in each?
- What is the burden of disease across underrepresented racial and ethnic populations, including access to biomarker testing and treatment and morbidity/mortality?
- Do trial inclusion/exclusion criteria disproportionately impact enrollment of certain racial and ethnic populations and/or geographic locations?

Given the increasing number of precision medicine trials, biomarker data is becoming increasingly important. Summarizing published and/or other available evidence per geographical region, number of patients screened for the biomarker, type of biomarker tests, and other parameters can allow for more accurate estimates of the target trial population. Such a comprehensive review of data helps in determining how closely the study conditions mirror real-world settings (external validity) and the degree to which the study findings are free from biases (internal validity).

## Measures to Achieve Enrollment Goals

In addition to setting enrollment goals, diversity plans will need to outline measures for achieving these goals. FDA's assessment of experience with diversity plans in the first-year after the April 2022 guidance identified strategies sponsors currently employ to achieve enrollment goals, including patient-directed measures (84% of plans), community engagement (82%), clinical research workforce-directed measures, and trial design considerations such as use of decentralized elements (21%), and eligibility criteria considerations (21%).[17] Survey responses highlight measures being taken and outline some of the approaches that should be leveraged to recruit, enroll, and retain diverse patient populations:

### Building Trust and Partnerships in Diverse Communities

Sponsors should actively and continually work to cultivate new partnerships and sustain relationships within diverse communities by partnering with community health centers serving diverse populations, diverse providers, and other community organizations and patient advocacy groups. These relationships can help build patient and provider trust in clinical trials, promote participation, and gather valuable patient and provider feedback crucial for informing clinical development programs. Engagement includes partnering with sites experienced in recruiting diverse patients to understand successful approaches and leveraging these learnings to train and support other clinical trial sites on the importance of including patients from underrepresented groups in clinical trials. Partnerships with diverse sites and providers can also help to facilitate dialogue regarding the specific needs of site staff to support effective recruitment and retention of patients. Depending on the needs identified by site staff, participating sites can be supported with tailored plans and resources including accessible patient-facing materials in various languages, transportation services for trial participants, and trainings on communicating clinical trial opportunities and processes.

Sponsors should also consider how to effectively communicate clinical trial conduct and outcomes with patients. Regular and accessible updates on trial processes, progress, and results at the conclusion of the study (e.g., lay summaries of data) can help to build trust by enhancing transparency, and help to empower patients by providing information to support self-advocacy. Additionally, Sponsors should seek the input of health care providers and patient navigators from underrepresented populations in all aspects of trial conduct and planning including collaborative development resources and educational materials and trial design.

Engagement with diverse communities outside of the healthcare setting is also necessary to build trust. Active participation in community events addressing SDoH and collaborative efforts with community- and faith-based organizations on relevant public policy endeavors are critical components of forming these sustained partnerships. Collaboration with community outreach organizations and patient advocacy groups focused on narrowing health equity gaps is also important. In addition, efforts should be made to develop tailored media and advertising, provide translation services and multilingual materials to bridge language barriers to ensure there is accessible information being disseminated about available clinical trials. It is critical that all patients are provided the necessary information and asked to participate in clinical trials.

A deeper understanding of local dynamics within a community, as well as the power dynamics between the community and research/healthcare system, can help to clarify how these factors influence healthcare utilization and clinical trial enrollment and retention. A clearer understanding of these dynamics can inform strategies to address these factors head on to enhance inclusion and participation and facilitate a sustained engagement and commitment to diverse communities.

## Lowering Barriers to Participation

To enhance enrollment and retention, sponsors should actively assess and address barriers that hinder patient recruitment in clinical trials. Understanding these obstacles can facilitate access for participants interested in clinical trials. For instance, sponsors should consider the financial burden on patients enrolled onto trials, offering pre-loaded reimbursements for transportation, accommodations, meals, and potential compensations for loss of earnings incurred due to trial participation. In addition, financial burden (beyond travel expenses and other out of pocket costs) continues to be a hurdle for many clinical trial participants, and can disproportionately affect some therapeutic areas, such as those requiring very frequent, lengthy, or complex assessments, indications that require extended research timelines, and/or treatment areas where even the standard of care is not adequately covered for patients who have insurance or are participants in government healthcare programs, such as Medicaid.[27] FDA should work with HHS and other agencies to ensure that these roadblocks are addressed in a way that allows sponsors to provide the support needed to help ensure that clinical research is a realistic option across different communities.

Sponsors should also evaluate protocols to identify areas for lowering barriers to enrollment, such as removing overly restrictive eligibility criteria, when scientifically justified.[28,29] Additionally,

decentralizing aspects of a trial through the use of mobile units, telemedicine, and/or distributing medicine through the mail can enhance accessibility. Other trial design aspects should be considered to streamline protocols and reduce operational burden for both patients and investigators. This process should include patient advocates and advocacy groups to regularly evaluate protocol complexity and pinpoint areas where reducing the burden could encourage greater participation. Industry should share best practices, and in particular, strategies that have a positive impact on diversity in enrollment to learn from one another.

## Intentional Site Selection

In addition to setting enrollment goals, sponsors should be intentional in their site selection by identifying health centers and providers in community settings that serve catchment areas with diverse patient populations and have diverse representativeness in trial personnel. Intentional site selection is critical to ensure diverse communities have access to clinical trials, which can lead to enrollment and retention of representative patient populations. Traditional site selection has focused on historical site performance metrics (e.g., GCP/protocol compliance, data quality, ability to efficiently recruit, enroll, and retain patients). However, as part of efforts to enroll more representative populations, it is important to incorporate diversity considerations in site selection processes. For example, site surveys and questionnaires, such as the Diversity Site Assessment Tool (DSAT) developed by the Society for Clinical Research Sites, can be used to evaluate site readiness in recruiting, enrolling, and retaining patients from underrepresented populations.[30] These assessments should encompass evaluating whether care incorporates cultural humility/safety, availability of language services, site staff diversity, and patient-centric services. Given that practices caring for underrepresented populations may be less likely to participate in clinical trials, dedicated training programs should be offered to onboard and enhance the capabilities of sites without previous experience engaging with clinical research, ensuring readiness to effectively participate in clinical studies. These programs to bolster site readiness are necessary to achieve the longer-term goal of cultivating a network of sites equipped to engage diverse patient populations effectively.

## Real-time Tracking of Enrollment Progress

Implementing real-time tracking mechanisms to monitor enrollment progress can help assess progress toward the achievement of enrollment goals and identify potential areas for improvement. This approach allows sponsors to proactively understand actual versus projected enrollment status, especially in enrolling individuals from historically or currently underrepresented racial and ethnic groups, enabling them to reassess and adapt strategies, as necessary. Implementing a comprehensive dashboard integrating site performance data, local diversity metrics, incidence data, and risk factors could be one approach for providing a holistic view of trial progress. Analysis of screen failure reasons offers insights into the effectiveness of tactics employed and facilitates potential or appropriate adaptations. Overall, frequent evaluation of diversity plan progress can allow for adjustments as needed.

## Conclusion

Improving the representativeness of diverse racial and ethnic groups in clinical trials while also considering other diversity dimensions such age, sex/gender, and SDoH is necessary to address health disparities and ensure equitable healthcare access. The lack of inclusivity in current clinical trials can impact the generalizability of findings and enable continued disparities in health outcomes. Efforts by the U.S. FDA underscored by draft guidance documents and the passage of key provisions in FDORA signal a substantial commitment to enhancing diversity and representativeness in clinical trials.

However, as sponsors navigate the implementation of these recommendations, systemic challenges, particularly regarding availability of comprehensive data sources, need to be addressed and best practices established for achieving enrollment goals. Between April 2022 and April 2023, 82% of diversity plans submitted to CDER included enrollment goals and many included various measures for achieving these goals. FDA provides feedback to sponsors who submit plans to support effective implementation, which indicates the need for additional guidance in several areas to support diversity planning: 90% of feedback focused on enrollment goals, 29% of feedback was on strategies for enhancing accrual to meet the goals, and 29% on trial enrollment monitoring, with some feedback focusing on multiple topics.[17] To achieve more inclusive trials, a multifaceted approach is needed that encompasses robust data analysis, strategic planning, community engagement, clinical trial designs, and thoughtful site selection (Figure 1). While this effort will require significant investment and resources, by addressing data challenges, partnering with communities, and implementing inclusive trial practices, the community will realize a more equitable and representative clinical trials system.

Table 1. Select Examples of Data Sources Used to Inform Enrollment Goals.

| Type | Specific Examples | Usages | Strengths | Limitations |
|---|---|---|---|---|
| Publicly Available Epidemiological Data | SEER Data, Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR) | Understanding disease prevalence, incidence, and demographic estimates | Representative of the civilian U.S. population, low missing data on demographics, publicly available/easily accessible | Limited data on rare indications, some data may not be up-to-date given timing/cadence of data collection and publishing of results, incomplete race and ethnicity data, and if included are limited to existing race and ethnicity categories, and these sources may also not include specific biomarker data and treatment history |
| Real-World Data (RWD) | Healthcare medical claims data (e.g., CMS, private insurers), Electronic Health Records (EHRs) data (e.g., Flatiron Health, Tempus, Cerner Health), other clinical provider databases | Identifying geographies and areas for high-incidence diseases, evaluating treatment effectiveness and safety | Patient-level data, captures treatment and genomic information, demographic data, near real-time | Capture of insured populations with healthcare access, missing data on race and ethnicity in some cases, variable data quality/completeness, variable reporting on biomarker data |
| Literature Searches | PubMed, clinicaltrials.gov | Informing historical disease characteristics, benchmarking goals | Provides insights into historical disease characteristics | May not provide current and representative data, selection bias in clinical trial data, biomarker data variability |
| Government Sources and Surveys | U.S. Census Data, National Health Information Survey (NHIS), National Health and Nutrition Examination Survey (NHANES), Medical Expenditure Panel Survey (MEPS) | Estimating the total population at risk, assessing disease burden in health-specific surveys, demographics, and treatment patterns | Large sample size, relevant for benchmarking population-level diversity, low missingness for race/ethnicity data | Some data may not be up-to-date given timing/cadence of data collection and publishing of results, no specific cancer staging, tumor size, etc. |

## References


1 Unger JM, Vaidya R, Hershman DL et al. Systematic Review and Meta-Analysis of the Magnitude of Structural, Clinical, and Physician and Patient Barriers to Cancer Clinical Trial Participation. JNCI J Natl Cancer Inst 2019;111:245–255.

2 Murthy VH, Krumholz HM, Gross CP. Participation in Cancer Clinical Trials: Race-, Sex-, and Age-Based Disparities. JAMA 2004;291:2720.

3 Tejeda HA, Green SB, Trimble EL et al. Representation of African-Americans, Hispanics, and Whites in National Cancer Institute Cancer Treatment Trials. JNCI J Natl Cancer Inst 1996;88:812–816.

4 Woods-Burnham L, Johnson JR, Hooker SE et al. The Role of Diverse Populations in US Clinical Trials. Med 2021;2:21–24.

5 Zavala VA, Bracci PM, Carethers JM et al. Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer 2021;124:315–332.

6 Fashoyin-Aje L, Beaver JA, Pazdur R. Promoting Inclusion of Members of Racial and Ethnic Minority Groups in Cancer Drug Development. JAMA Oncol 2021;7:1445.

7 Scharff DP, Mathews KJ, Jackson P et al. More than Tuskegee: Understanding Mistrust about Research Participation. J Health Care Poor Underserved 2010;21:879–897.

8 U.S. Food and Drug Administration. Inclusion of Older Adults in Cancer Clinical Trials. 2022. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/inclusion-older-adults-cancer-clinical-trials. Accessed December 8, 2023.

9 U.S. Food and Drug Administration. Cancer Clinical Trial Eligibility Criteria: Patients with Organ Dysfunction or Prior or Concurrent Malignancies. 2020. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-patients-organ-dysfunction-or-prior-or-concurrent. Accessed December 8, 2023.

10 U.S. Food and Drug Administration. Cancer Clinical Trial Eligibility Criteria: Brain Metastases. 2020. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-brain-metastases. Accessed December 8, 2023.

11 U.S. Food and Drug Administration. Cancer Clinical Trial Eligibility Criteria: Minimum Age Considerations for Inclusion of Pediatric Patients. 2020.Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-minimum-age-considerations-inclusion-pediatric-patients. Accessed December 8, 2023.

12 U.S. Food and Drug Administration. Cancer Clinical Trial Eligibility Criteria: Patients with HIV, Hepatitis B Virus, or Hepatitis C Virus Infections. 2020. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-patients-hiv-hepatitis-b-virus-or-hepatitis-c-virus. Accessed December 8, 2023.

13 U.S. Food and Drug Administration. Draft Guidance for Industry: Postmarketing Approaches to Obtain Data on Populations Underrepresented in Clinical Trials for Drugs and Biological Products. 2023. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarketing-approaches-obtain-data-populations-underrepresented-clinical-trials-drugs-and. Accessed December 8, 2023.

14 U.S. Food and Drug Administration. Draft Guidance for Industry: Diversity Plans to Improve Enrollment of Participants From Underrepresented Racial and Ethnic Populations in Clinical Trials. 2022. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations. Accessed December 8, 2023.

15 The Food and Drug Omnibus Reform Act of 2022 (FDORA); Section 3602-3604. 2022.

16 Fashoyin-Aje LA, Tendler C, Lavery B et al. Driving Diversity and Inclusion in Cancer Drug Development - Industry and Regulatory Perspectives, Current Practices, Opportunities, and Challenges. Clin Cancer Res Off J Am Assoc Cancer Res 2023;29:3566–3572.

17 Fashoyin-Aje LA, Boehmer J, Antony R et al. Voluntary Submission of Diversity Plans to Oncology Review Divisions in the Center for Drug Evaluation and Research (CDER) Following FDA Draft Guidance: Year one experience: April 2022 – April 2023. 2023. Available at https://www.fda.gov/media/174160/download?attachment. Accessed December 14, 2023.

18 Institut national de la statistique et des études économiques. Ethnic-based statistics. Inst. Natl. Stat. Études Économiques. Available at https://www.insee.fr/en/information/2388586. Accessed December 8, 2023.

19 Guadamuz J, Calip GS, Altomare I et al. Mediators of racial/ethnic inequities in clinical trial participation among US patients with cancer, 2011-2022. 2023. Available at https://cdmcd.co/ZKnAvw. Accessed December 15, 2023.

20 Gregory A. Vidal, Jain N, Fisher A et al. Practice- and provider-level inequities in next-generation sequencing (NGS) testing by race/ethnicity for patients (pts) with advanced non-small cell lung cancer (aNSCLC) treated in the community setting. J Clin Oncol 2023. doi:10.1200/JCO.2023.41.16_suppl.6508.

21 White House Office of Science and Technology Policy. The U.S. Playbook To Address Social Determinants of Health. 2023. Available at https://www.whitehouse.gov/wp-content/uploads/2023/11/SDOH-Playbook-3.pdf.

22 The White House Office of Science and Technology Policy. FACT SHEET: Biden-Harris Administration Releases First-Ever Federal Evidence Agenda on LGBTQI+ Equity | OSTP. White House. 2023. Available at https://www.whitehouse.gov/ostp/news-updates/2023/01/24/fact-sheet-biden-harris-administration-releases-first-ever-federal-evidence-agenda-on-lgbtqi-equity/. Accessed December 8, 2023.

23 Office of Management and Budget (OMB). Notice and Request for Comment: Initial Proposals For Updating OMB's Race and Ethnicity Statistical Standards. Fed. Regist. 2023. Available at https://www.federalregister.gov/documents/2023/01/27/2023-01635/initial-proposals-for-updating-ombs-race-and-ethnicity-statistical-standards.

24 Jacobs F, D'Amico S, Benvenuti C et al. Opportunities and Challenges of Synthetic Data Generation in Oncology. JCO Clin Cancer Inform 2023;7:e2300045.

25 Crowdis J, Balch S, Sterlin L et al. A patient-driven clinicogenomic partnership for metastatic prostate cancer. Cell Genomics 2022;2:100169.

26 Surveillance, Epidemiology, and End Results Program (SEER). Race and Hispanic Ethnicity Changes - SEER Documentation. SEER. Available at https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/index.html. Accessed December 8, 2023.

27 Chino F, Zafar SY. Financial Toxicity and Equitable Access to Clinical Trials. Am Soc Clin Oncol Educ Book 2019;:11–18.

28 Kim ES, Bruinooge SS, Roberts S et al. Broadening Eligibility Criteria to Make Clinical Trials More Representative: American Society of Clinical Oncology and Friends of Cancer Research Joint Research Statement. J Clin Oncol 2017;35:3737–3744.

29 Kim ES, Uldrick TS, Schenkel C et al. Continuing to Broaden Eligibility Criteria to Make Clinical Trials More Representative and Inclusive: ASCO–Friends of Cancer Research Joint Research Statement. Clin Cancer Res 2021;27:2394–2399.

30 Society for Clinical Research Sites. Diversity Site Assessment Tool. Available at https://myscrs.org/dsat/. Accessed January 25, 2024.

# Real-World Evidence

**Leveraging Data from Routine Clinical Practice in Oncology**

# Evaluation of Real-World Tumor Response Derived From Electronic Health Record Data Sources: A Feasibility Analysis in Patients With Metastatic Non–Small Cell Lung Cancer Treated With Chemotherapy

Brittany A. McKelvey, PhD[1]; Elizabeth Garrett-Mayer, PhD, FSCT[2]; Donna R. Rivera, PharmD, MSc[3]; Amy Alabaster, MS, MPH[4]; Hillary S. Andrews, PhD[1]; Elizabeth G. Bond, MPH[5]; Thomas D. Brown, MD, MBA[6]; Amanda Bruno, PhD, MPH[7]; Lauren Damato, ACNP-BC[8]; Janet L. Espirito, PharmD[9]; Laura L. Fernandes, PhD[10]; Eric Hansen, MD[10]; Paul Kluetz, MD[3]; Xinran Ma, MS[8]; Andrea McCracken, MPH[5]; Pallavi S. Mishra-Kalyani, PhD[11]; Yanina Natanzon, PhD, MS[4]; Danielle Potter, PhD, MPH[12]; Nicholas J. Robert, MD[9]; Lawrence Schwartz, MD[13]; Regina Schwind, MS, MHI[14]; Connor Sweetnam, BS[6]; Joseph Wagner, BS[12]; Mark D. Stewart, PhD[1]; and Jeff D. Allen, PhD[1]

DOI https://doi.org/10.1200/CCI.24.00091

## ABSTRACT

**PURPOSE** Real-world data (RWD) holds promise for ascribing a real-world (rw) outcome to a drug intervention; however, ascertaining rw-response to treatment from RWD can be challenging. Friends of Cancer Research formed a collaboration to assess available data attributes related to rw-response across RWD sources to inform methods for capturing, defining, and evaluating rw-response.

**MATERIALS AND METHODS** This retrospective noninterventional (observational) study included seven electronic health record data companies (data providers) providing summary-level deidentified data from 200 patients diagnosed with metastatic non–small cell lung cancer (mNSCLC) and treated with first-line platinum doublet chemotherapy following a common protocol. Data providers reviewed the availability and frequency of data components to assess rw-response (ie, images, radiology imaging reports, and clinician response assessments). A common protocol was used to assess and report rw-response end points, including rw-response rate (rwRR), rw-duration of response (rwDOR), and the association of rw-response with rw-overall survival (rwOS), rw-time to treatment discontinuation (rwTTD), and rw-time to next treatment (rwTTNT).

**RESULTS** The availability and timing of clinician assessments was relatively consistent across data sets in contrast to images and image reports. Real-world response was analyzed using clinician response assessments (median proportion of patients evaluable, 77.5%), which had the highest consistency in the timing of assessments. Relative consistency was observed across data sets for rwRR (median 46.5%), as well as the median and directionality of rwOS, rwTTD, and rwTTNT. There was variability in rwDOR across data sets.

**CONCLUSION** This collaborative effort demonstrated the feasibility of aligning disparate data sources to evaluate rw-response end points using clinician-documented responses in patients with mNSCLC. Heterogeneity exists in the availability of data components to assess response and related rw-end points, and further work is needed to inform drug effectiveness evaluation within RWD sources.

ACCOMPANYING CONTENT

Data Supplement

Accepted June 25, 2024
Published August 15, 2024

JCO Clin Cancer Inform
8:e2400091




## INTRODUCTION

Despite the rigor of clinical trials, further understanding of a therapy's effectiveness may still be needed. The use of real-world data (RWD) to generate real-world evidence (RWE) may fill these gaps and support evaluation of therapeutic effectiveness. RWD may more readily capture the heterogeneity of the intended use population, provide information on long-term safety and effectiveness, and identify off-label use.[1] Recent efforts to increase research on and support use of RWE include the 21st Century Cures Act,[2] Prescription Drug User Fee Act VI[3]–VII,[4] the Food and Drug Omnibus Reform Act of 2022,[5] and President Biden's Cancer Moonshot.[6] To support drug development and regulatory decision making, there is a need to align on and further evaluate the use of RWD, including standardizing data

## CONTEXT

**Key Objective**
To develop an aligned methodology for assessing real-world response to treatment across disparate data sources.

**Knowledge Generated**
This methodological exercise supports the ability to align disparate data sources to evaluate rw-response in an aligned patient population. Real-world response end points using clinician-documented response show relative consistency across data sources.

**Relevance**
Using real-world data (RWD) in clinical practice can greatly enhance the understanding of treatment effectiveness, inform personalized care plans, and identify emerging trends in patient populations, ultimately improving health care quality and outcomes. This study evaluated patients with metastatic non–small cell lung cancer (mNSCLC) who were treated with first-line platinum doublet chemotherapy. It focused on the consistency and availability of data components in RWD sources, such as clinician assessments and radiology reports. The objective was to develop a methodology for determining real-world response (rw-response) and to explore its potential application in oncology research. The study demonstrated the feasibility of integrating diverse data sources to evaluate rw-response end points using clinician-documented responses in patients with mNSCLC. It highlighted the relative consistency of real-world response, underscoring the potential of RWD to support oncology research and inform clinical decision making.

elements, aligning definitions, and reproducing methodology across real-world (rw) data sets.

Friends of Cancer Research (*Friends*) previously convened key stakeholders to participate in collaborative pilots[7-9] to define rw-end points, including rw-overall survival (rwOS), rw-time to treatment discontinuation (rwTTD), and rw-time to next treatment (rwTTNT), and align these definitions across multiple RWD sources to enhance generation of RWE on patient outcomes. These pilots highlighted areas of concordance in the direction and magnitude of treatment effect measured through rw-end points across data sources when using a common research protocol. However, the projects found the common limitation that progression events were not consistently captured, requiring an additional concerted effort to evaluate approaches for capturing end points assessing change in tumor burden, such as objective response rate (ORR) and progression-free survival (PFS).

ORR is an informative regulatory measure that can be used as an end point in single-arm trials, as causality is reasonably inferred (ie, tumors do not typically shrink spontaneously). Response rate is also evaluated earlier in the treatment course and may be reasonably likely to predict clinical benefit (ie, PFS and OS).[10] The duration and magnitude of response is important to understand the treatment-response trajectory and to ascribe clinical meaningfulness. Within clinical trials, RECIST 1.1 outlines a standardized approach (ie, consistent and objective mode of evaluation and cadence of assessment) to capture the response of solid tumors to an oncology treatment. However, there are challenges with characterizing rw-response in solid tumors, as the components necessary to measure RECIST-based response are not often accessible or available in the electronic health record (EHR) or assessed in a standardized manner outside of a protocol-driven study. Recognizing the increased heterogeneity of routine clinical practice, when compared with clinical trials, this pilot project sought to (1) understand the availability and feasibility of using specific RWD elements to assess rw-response, (2) evaluate the potential to ascertain rw-response using available data elements from the EHR, and (3) evaluate the consistency of these measures across data sources.

## MATERIALS AND METHODS

### Standardization of Methods

A collaborative partnership of RWD providers, pharmaceutical companies, academics, and government agencies jointly developed the common protocol and statistical analysis plan, including definitions on patient selection criteria, data elements, and outcomes (Data Supplement, Tables S1-S8). Each RWD provider (cohort) assessed their deidentified, patient-level EHR data to report uniform summary results (Data Supplement). Contributing data providers included ConcertAI, COTA Inc, Flatiron Health, Guardian Research Network and IQVIA, Ontada, Syapse, and Tempus AI.

### RWD Cohort Development

Each cohort identified adult patients (age 18 years or older at metastatic diagnosis) with histologically confirmed metastatic non–small cell lung cancer (mNSCLC) by structured or abstracted data, diagnosed between January 1, 2015, and March 31, 2018 (inclusive) in their databases, a time frame reflective of cohorts selected for previous pilots.[7-9] All

cohorts received institutional review board approval or exemption. Patients received first-line (1L) treatment with platinum doublet chemotherapy (PDC) regimens with or without vascular endothelial growth factor (VEGF) receptor antagonists (Data Supplement, Fig 1). Eligible patients were documented as physically present at a practice or having an encounter in the database on at least two separate occasions, and patients were excluded if there was incomplete treatment data (Data Supplement). Of the eligible patients, each data provider performed random sampling to achieve a cohort size of 200 patients. After sampling, an additional 20 patients were excluded from cohort G for not meeting eligibility criteria. This sample size was chosen to ensure uniformity across cohorts and for feasibility reasons, because of the level of data curation necessary. Clinical and demographic characteristics were summarized using descriptive statistics.

### Assessment of Availability of Response Data Components

Cohorts assessed the availability of core data components during the assessment period. Components included images (magnetic resonance imaging [MRI], positron emission tomography-computed tomography [PET-CT], CT, and other), image reports (MRI, PET-CT, CT, and other), and clinician assessment of response (as stated in notes, where response evidence was referenced from imaging, symptoms, laboratory results, physical examination, pathology reports, other sources, or was not specified). The data component assessment was divided into two periods, baseline (time from the metastatic diagnosis date to the day before the start of 1L therapy, defined as the index date) and postbaseline (time from the index date up to the earliest of the start of new [second-line] treatment, 30 days after the last administration of 1L treatment, death, or data cutoff), to identify both baseline and postbaseline images or image reports for response assessment. Evaluation of clinician assessment of response was only conducted in the postbaseline period. Results were summarized for the proportion of patients in each cohort with each data component available within the assessment period. Medians and IQRs were reported for the number and timing of data components per patient. The component source (image modality and indication for image reports, or source for clinician response assessment in the record) was treated as a categorical variable and reported as a proportion of the total number of available data components. Additional statistical considerations are described in more detail in the Data Supplement.

### Methodology for rw-Response End Points and Parameter Estimation

Clinician assessment of response was used to determine rw-response for all patients using the categories rw-complete response (rwCR), rw-partial response (rwPR), rw-stable disease (rwSD), rw-progressive disease (rwPD), rw-mixed response (rwMR), and not evaluable (NE; Data Supplement, Table S3). The rw-best overall response (rwBOR) was defined as the patient's best response, where rwCR was the most

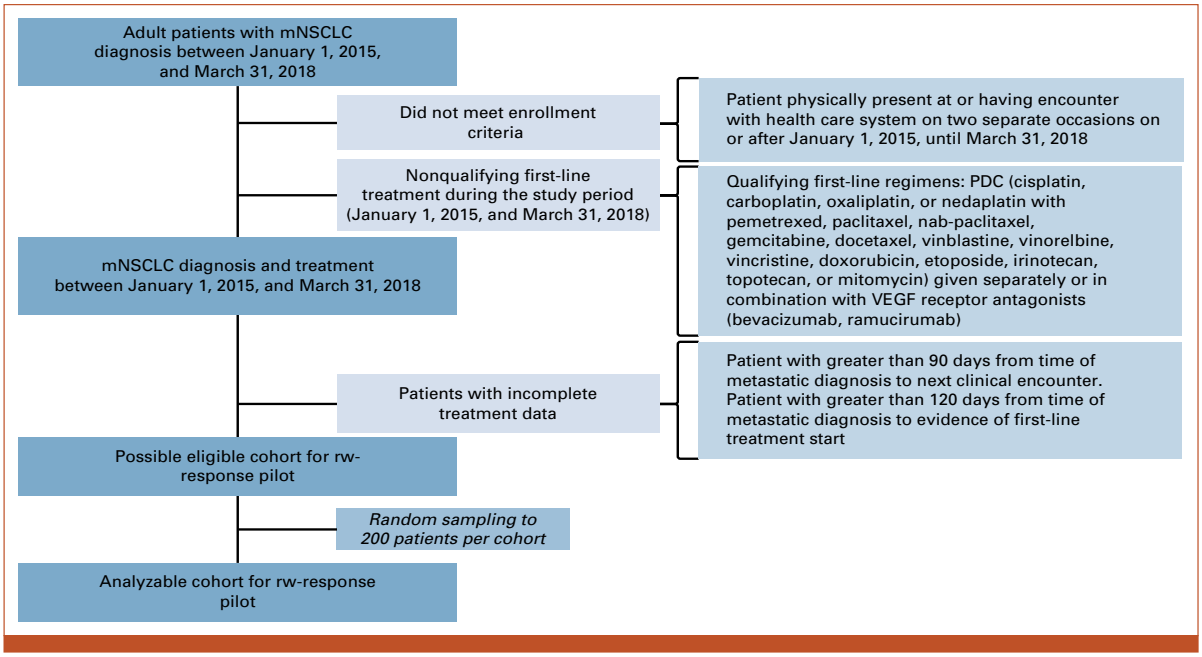



**FIG 1.** Flow diagram. mNSCLC, metastatic non–small cell lung cancer; PDC, platinum doublet chemotherapy; VEGF, vascular endothelial growth factor.

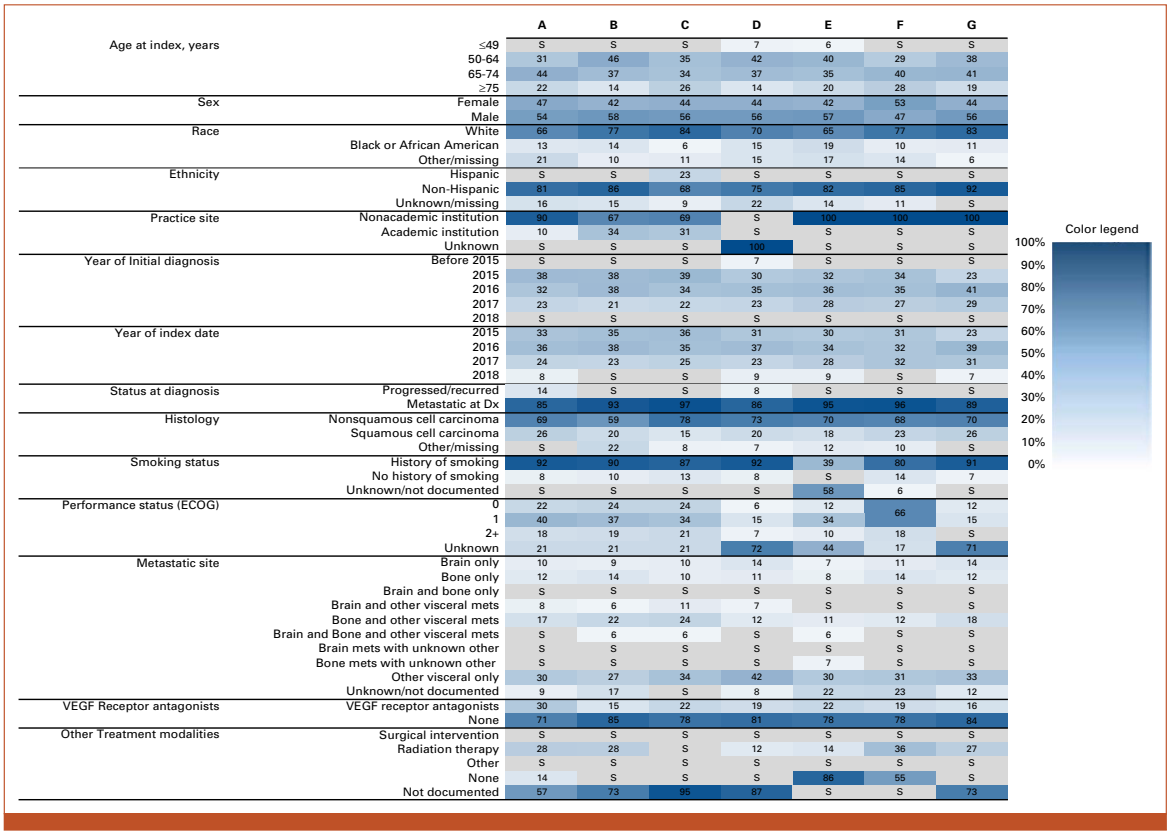



**FIG 2.** Demographic and clinical characteristics of cohorts. Numbers represent the proportion of patients in each category. Shading denotes the proportion of patients from white (0%) to dark blue (100%) to aid in visual comparison across cohorts. Data are suppressed (S, in gray) if ≤5%. Dx, diagnosis; ECOG, Eastern Cooperative Oncology Group; VEGF, vascular endothelial growth factor.

across cohorts (range, 32%-61%) had more than one clinician assessment (Data Supplement, Table S13). The timing of clinician assessments was relatively consistent across cohorts, with a median of 7.9 weeks between both the index date to first assessment and first to second assessment (Table 1, Data Supplement, Table S13). Across all cohorts, imaging was the most frequently cited source of evidence for clinician assessments of response (Data Supplement, Fig S3), followed by symptoms.

### rw-Response Estimates and End Points

There was relative consistency in rwRR (median, 46.5%, range, 38%-53%) using clinician-documented response across cohorts (Fig 4). A median of 22.5% (range, 11.7%-26.0%) of patients did not have a response assessment during the assessment period, and these patients had the shortest follow-up time compared with responders and nonresponders (Data Supplement, Fig S4). There was variability in rwDOR across data sets (Fig 5, Data Supplement, Fig S5), and accounting for interval censoring substantially increased the estimated variance.

The results of the sensitivity analyses were relatively consistent with the primary analyses (Data Supplement, Table S14).

The relationships between rw-response and rwOS, rwTTD, and rwTTNT were analyzed. Relative consistency was observed in the median estimates and directionality of the time-to-event end points (rwOS, rwTTD, and rwTTNT) across cohorts for responders compared with nonresponders (Fig 6, Data Supplement, Fig S6). Like the short follow-up time seen for patients with no response assessment, rwTTD, rwTTNT, and rwOS were consistently shorter for those with no response assessment than for both nonresponders and responders (Data Supplement, Fig S6).

## DISCUSSION

Overall, this collaborative effort assessed the availability of data components to measure rw-response and evaluated the consistency of the measure across RWD sources. The pilot demonstrates the feasibility of aggregating data from various rw-data sets to generate RWE. Findings highlight

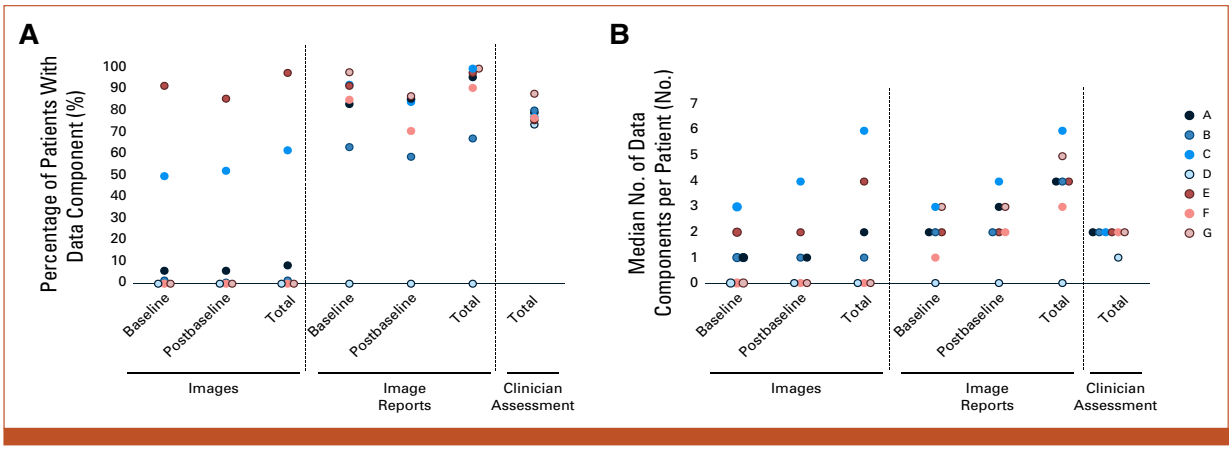



**FIG 3.** Availability of rw-response assessment data components. Dot plots depict (A) the percentage of patients in each cohort with each data component and (B) median number of data components per patient. The median number of data components is calculated only for patients with at least one data component in the record (patients with 0 assessments are not included). rw-response, real-world response.

reasonable consistency in rw-response across disparate data sources in an aligned patient population using clinician-documented response.

The pilot used a common protocol, with all data providers following an a priori agreed upon eligibility criteria, statistical analysis plans, and standardized definitions. For this methodological exercise, the patient population reflected previous *Friends*' pilots during a time frame when chemotherapy was frequently used, focusing on PDC to remove potential confounding of pseudoprogression with immunotherapy treatment.

Using RWD for causal inference can be challenging for many reasons, including the need to ascertain relevant and reliably detailed, longitudinal clinical characteristics. Data generation currently requires significant manual abstraction and curation, which limited the sample size, highlighting the challenges with evaluating rw-response and the need for standardized structured RWD. RWD can be generated from multiple sources, including EHR-derived and administrative claims data; however, EHR data were necessary to ascertain rw-response. Although many areas showed relative consistency across EHR-derived RWD cohorts, areas such as specific clinical characteristics (eg, other treatment modalities) and availability of imaging were more variable or limited for some cohorts. To support causal inference, other variables must be appropriately controlled to demonstrate that tumor response is due to the treatment, not factors such as concomitant therapies, additional modalities, or other confounding factors.

The availability and extractability of images was limited and varied significantly across cohorts. Privacy, contractual, and/

**TABLE 1.** Medians Across Cohorts Calculated From Summary-Level Statistics of Each Cohort

| Component | Baseline to Index | Baseline to Postbaseline | First to Second Postbaseline |
|---|---|---|---|
| Images | | | |
| Proportion with data, median (range) | 28% (1.5%-92%) | 22% (0.5%-79.5%) | 29% (0.5%-86%) |
| Time between in weeks, median (range) | 2.95 (2.4-5) | 13.2 (7.3-18) | 6 (3.29-7) |
| Image reports | | | |
| Proportion with data, median (range) | 88.80% (63.5%-98.3%) | 75% (55%-85.6%) | 85% (59%-87.2%) |
| Time between in weeks, median (range) | 3.63 (2.3-4) | 9.62 (7.5-18) | 5 (3.7-6.3) |

| | Index to Assessment | First to Second Assessment |
|---|---|---|
| Clinician assessment | | |
| Proportion with data, median (range) | 77.50% (74%-88.3%) | 44.50% (32%-61%) |
| Time between in weeks, median (range) | 7.9 (6.9-8) | 7.9 (6-9) |

NOTE. The median time between data components is calculated only for patients with at least one data component in the record (patients with 0 assessments are not included).

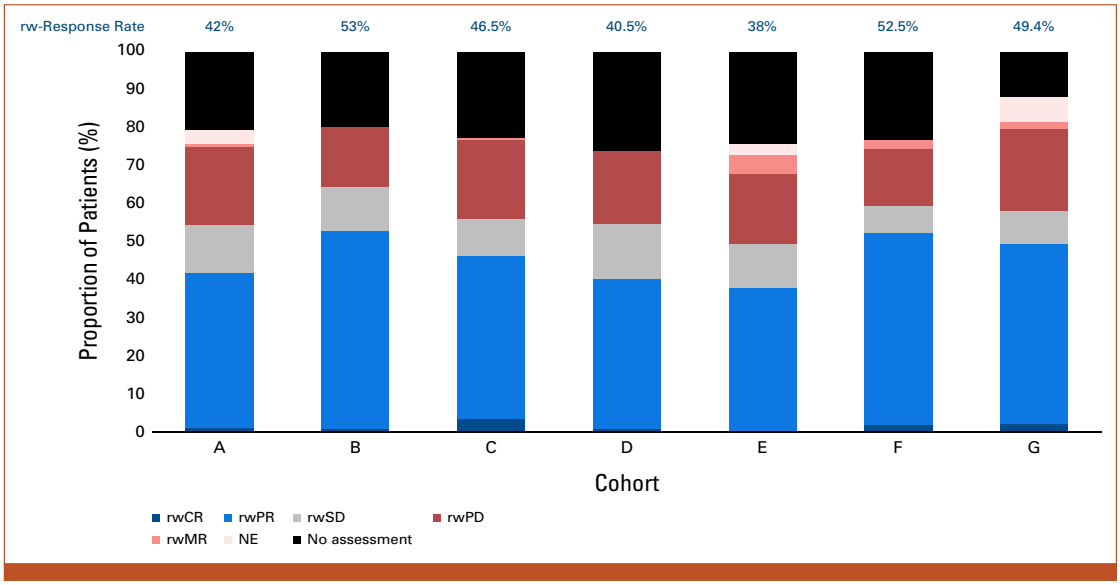



**FIG 4.** rw-best overall response and response rate across cohorts. The proportion of patients in each cohort with a given rw-best overall response by clinician assessment of response. Response rate (above bars) is derived from patients with rwPR and rwCR, out of total patients. Cohorts A-F, n = 200 patients; cohort G, n = 180 patients. NE, not evaluable; rw, real-world; rw-response, real-world response; rwCR, rw-complete response; rwMR, rw-mixed response; rwPD, rw-progressive disease; rwPR, rw-partial response; rwSD, rw-stable disease.

or compliance issues were stated as barriers to obtaining and sharing images. Additionally, linking images to the EHR requires a high level of interoperability, data management (privacy and deidentification considerations), and storage that may not be feasible for all institutions. This remains a technological and infrastructural challenge to using rw-end points.

Ascertaining rw-response from currently available EHR data will likely need to rely on clinician assessments. Response evaluated by the clinician's assessment of a patient's change in disease burden was available for most patients across all cohorts. Multiple imaging modalities were used, which may make applying a RECIST-like assessment of response difficult. The clinician assessment considers a variety of inputs (eg, radiology, physical examination, biomarkers, pathology, and patient-reported symptoms), which introduces heterogeneity and subjectivity, although findings reported herein demonstrate the source of evidence for most assessments was imaging and image reports. The timing of clinician assessments was relatively consistent across cohorts and reflects the timing prescribed in PDC clinical trials where patients are assessed every 6-8 weeks after random assignment,

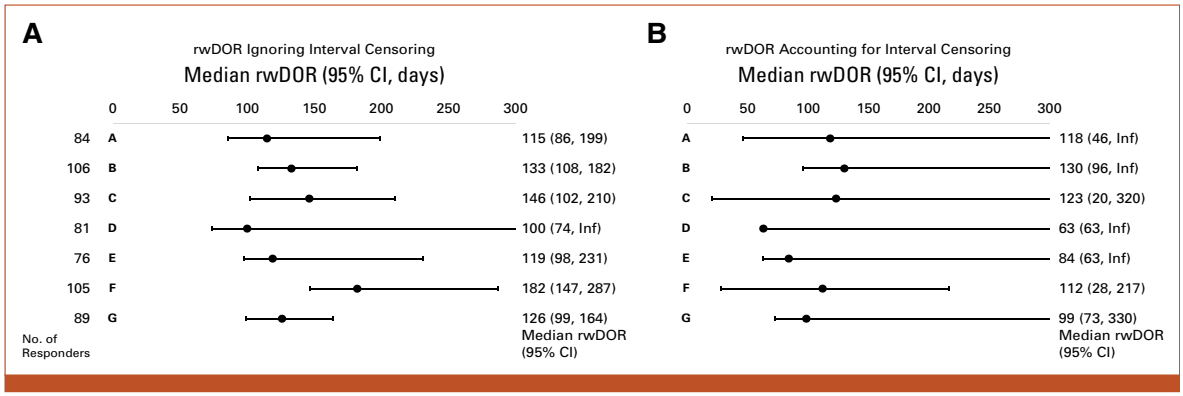



**FIG 5.** rwDOR across cohorts. rwDOR (A) ignoring interval censoring and (B) accounting for interval censoring for patients with complete or partial response (responders) across cohorts. Graphs show the median rwDOR with 95% CIs. rwDOR, real-world duration of response.

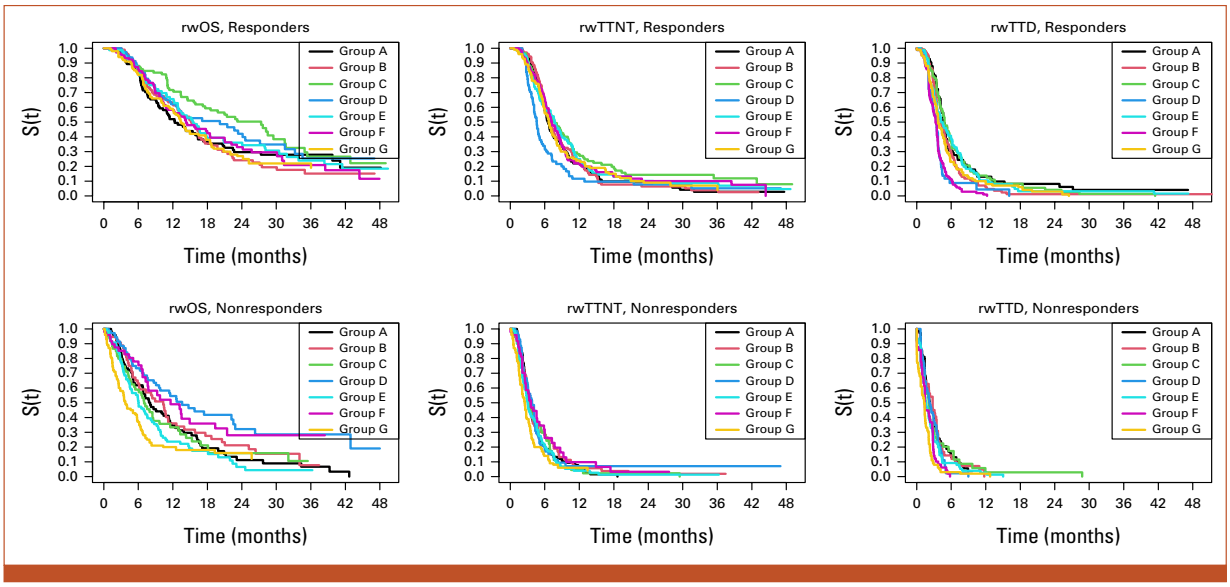



**FIG 6.** rw-time to event end points by rw-response to treatment. Kaplan-Meier curves for responders and nonresponders for rwOS, rwTTNT, and rwTTD, across cohorts. rw, real-world; rwOS, rw-overall survival; rwTTD, rw-time to treatment discontinuation; rwTTNT, rw-time to next treatment.

indicating that patients treated outside clinical trials are likely under similar active assessment or surveillance at regular intervals. However, a proportion of patients did not have a response assessment, possibly due to being lost to follow-up, rapid decline, transfer of patient care, discontinuation of treatment because of toxicity, or patient choice.

Using clinician assessment to evaluate rwRR was relatively consistent across all RWD sources, albeit notably higher than values observed in mNSCLC trials for patients treated with PDC (rwRR median 46.5% compared with ORRs of 19.4%[11] and 38.4%[12]). Given the lack of application of standardized RECIST assessment criteria outside of clinical trials, a rwPR can include any reduction of the tumor burden, not the minimum of 30% reduction required by RECIST 1.1. Likewise, the results showed a median of 11.5% of patients classified as rwBOR of rwSD, while the trials referenced above had 51% and 37% of patients classified as having stable disease, respectively. Therefore, patients with small decreases in tumor burden in routine clinical practice may be categorized as partial responders, while these same patients would likely be categorized as stable disease based on RECIST 1.1 criteria. Durability of response can provide additional insight into therapeutic efficacy, and rwDOR varied across cohorts in the study, possibly because of the variability in timing of patient assessments, variability in reporting of data, or other unmeasured or residual factors.

This study has several limitations. Data were aggregated from various data providers, such that duplication of patients may have occurred, and therefore data in the different cohorts may contain some of the same patients. Furthermore, interval censoring may have made interpretation challenging. The study also did not require patients to have measurable disease, as would be required in clinical trials using RECIST. Finally, although each data provider used patient-level data, aggregate analyses across cohorts were limited to interpretations from summary-level data.

The demonstrated feasibility of data providers' adherence to a common data model with relative consistency in rw-response end points on the basis of clinician assessment suggests rw-response warrants further exploration to inform drug effectiveness evaluation. There is a degree of uncertainty in the relationship between RECIST-based assessment and clinician assessment, which requires additional methodological development. Therefore, rw-response end points are not directly comparable with RECIST-based clinical trial response assessments and may best be leveraged for evaluation of response within RWD. Use of rw-response may support evaluation of a treatment effect in a specific population in the rw-setting or in subpopulations that were underrepresented in clinical trials. The measure may also be valuable for signal seeking to aid in identifying populations in which to explore efficacy in future clinical trials or for evidence to support label expansion of an already approved therapy. Aligning methodologies for aggregating and analyzing RWD will support use of RWD as a reliable and consistent source of RWE to support oncology drug development and regulatory decision making.

## AFFILIATIONS

[1]Friends of Cancer Research, Washington, DC
[2]American Society of Clinical Oncology, Alexandria, VA
[3]Oncology Center of Excellence, US FDA, Silver Spring, MD
[4]ConcertAI, Cambridge, MA
[5]Guardian Research Network, Spartanburg, SC
[6]Syapse, San Francisco, CA
[7]Syneos Health, Morrisville, NC
[8]Flatiron Health, New York, NY
[9]Ontada, Boston, MA
[10]COTA, Inc, New York, NY
[11]US FDA, Silver Spring, MD
[12]IQVIA, Durham, NC
[13]Memorial Sloan Kettering Cancer Center, New York, NY
[14]Tempus AI, Inc, Chicago, IL

## CORRESPONDING AUTHOR

Brittany A. McKelvey; e-mail: bmckelvey@focr.org.



## PRIOR PRESENTATION

Presented in part at ASCO Annual Meeting, Chicago, IL, June 2-6, 2023.

## SUPPORT

Supported by Friends of Cancer Research.


## AUTHOR CONTRIBUTIONS

**Conception and design:** Brittany A. McKelvey, Elizabeth Garrett-Mayer, Donna R. Rivera, Hillary S. Andrews, Thomas D. Brown, Amanda Bruno, Lauren Damato, Janet L. Espirito, Laura L. Fernandes, Eric Hansen, Xinran Ma, Pallavi S. Mishra-Kalyani, Yanina Natanzon, Nicholas J. Robert, Lawrence Schwartz, Regina Schwind, Connor Sweetnam, Mark D. Stewart, Jeff D. Allen
**Provision of study materials or patients:** Thomas D. Brown, Janet L. Espirito, Andrea McCracken, Nicholas J. Robert
**Collection and assembly of data:** Brittany A. McKelvey, Amy Alabaster, Elizabeth G. Bond, Thomas D. Brown, Lauren Damato, Andrea McCracken, Yanina Natanzon, Nicholas J. Robert, Regina Schwind, Connor Sweetnam, Joseph Wagner, Mark D. Stewart
**Data analysis and interpretation:** Brittany A. McKelvey, Elizabeth Garrett-Mayer, Donna R. Rivera, Amy Alabaster, Hillary S. Andrews, Elizabeth G. Bond, Thomas D. Brown, Amanda Bruno, Lauren Damato, Janet L. Espirito, Eric Hansen, Paul Kluetz, Xinran Ma, Pallavi S. Mishra-Kalyani, Yanina Natanzon, Danielle Potter, Nicholas J. Robert, Lawrence Schwartz, Connor Sweetnam, Mark D. Stewart, Jeff D. Allen
**Manuscript writing:** All authors
**Final approval of manuscript:** All authors
**Accountable for all aspects of the work:** All authors

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

**Amy Alabaster**
**Employment:** ConcertAI, Kaiser Permanente
**Research Funding:** ConcertAI (Inst)

**Hillary S. Andrews**
**Consulting or Advisory Role:** Avalere Health

**Thomas D. Brown**
**Employment:** Syapse
**Leadership:** Syapse
**Stock and Other Ownership Interests:** Syapse, Sygnomics
**Consulting or Advisory Role:** Sygnomics
**Research Funding:** Syapse
**Travel, Accommodations, Expenses:** Syapse

**Amanda Bruno**
**Employment:** Bayer, Syneos Health
**Stock and Other Ownership Interests:** Bristol Myers Squibb, Syneos Health

**Lauren Damato**
**Stock and Other Ownership Interests:** Roche

**Janet L. Espirito**
**Employment:** McKesson
**Stock and Other Ownership Interests:** McKesson

**Laura L. Fernandes**
**Employment:** Cota Healthcare
**Stock and Other Ownership Interests:** Cota Healthcare

**Eric Hansen**
**Employment:** COTA Healthcare
**Stock and Other Ownership Interests:** AbbVie, Johnson & Johnson, UnitedHealthcare, COTA Healthcare

**Xinran Ma**
**Employment:** Flatiron Health
**Stock and Other Ownership Interests:** Roche
**Research Funding:** Flatiron Health
**Travel, Accommodations, Expenses:** Flatiron Health

**Yanina Natanzon**
**Employment:** Syapse, ConcertAI
**Stock and Other Ownership Interests:** Syapse, ConcertAI
**Travel, Accommodations, Expenses:** Syapse, ConcertAI

**Danielle Potter**
**Employment:** IQVIA
**Stock and Other Ownership Interests:** AstraZeneca, IQVIA

**Nicholas J. Robert**
This author is a member of the *Journal of Clinical Oncology Clinical Cancer Informatics* Editorial Board. Journal policy recused the author from having any role in the peer review of this manuscript.
**Employment:** McKesson
**Leadership:** McKesson
**Stock and Other Ownership Interests:** McKesson, Oncolytics, Moderna Therapeutics

**Lawrence Schwartz**
**Consulting or Advisory Role:** Novartis, Regeneron, Bristol Myers Squibb/Celgene, Merck
**Research Funding:** Merck Sharp & Dohme (Inst), Boehringer Ingelheim (Inst)
**Patents, Royalties, Other Intellectual Property:** Varian Medical Systems

**Regina Schwind**
**Employment:** Tempus AI
**Stock and Other Ownership Interests:** Tempus AI

**Connor Sweetnam**
**Employment:** Syapse
**Stock and Other Ownership Interests:** Syapse

**Joseph Wagner**
**Stock and Other Ownership Interests:** Pfizer, Merck, Viatris

No other potential conflicts of interest were reported.

## REFERENCES


1. Rudrapatna VA, Butte AJ: Opportunities and challenges in using real-world data for health care. J Clin Invest 130:565-574, 2020
2. US Congress: 21st Century Cures Act, 2016. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf
3. PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2018 Through 2022. https://www.fda.gov/media/99140/download
4. PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2023 Through 2027. https://www.fda.gov/media/151712/download?attachment
5. US Congress: Consolidated Appropriations Act, 2023. https://www.congress.gov/117/bills/hr2617/BILLS-117hr2617enr.pdf
6. The White House: FACT SHEET: On One Year Anniversary of Reignited Cancer Moonshot, Biden-Harris Administration Announces New Actions to End Cancer as We Know It. https://www.whitehouse.gov/briefing-room/statements-releases/2023/02/02/fact-sheet-on-one-year-anniversary-of-reignited-cancer-moonshot-biden-harris-administration-announces-new-actions-to-end-cancer-as-we-know-it/
7. Stewart M, Norden AD, Dreyer N, et al: An exploratory analysis of real-world end points for assessing outcomes among immunotherapy-treated patients with advanced non-small-cell lung cancer. JCO Clin Cancer Inform 3:1-15, 2019https://ascopubs.org/doi/10.1200/CCI.18.00155
8. Rivera DR, Henk HJ, Garrett-Mayer E, et al: The Friends of cancer research real-world data collaboration pilot 2.0: Methodological recommendations from oncology case studies. Clin Pharmacol Ther 111:283-292, 2022https://onlinelibrary.wiley.com/doi/full/10.1002/cpt.2453
9. Lasiter L, Tymejczyk O, Garrett-Mayer E, et al: Real-world overall survival using oncology electronic health record data: Friends of Cancer Research pilot. Clin Pharmacol Ther 111:444-454, 2022
10. US FDA: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. https://www.fda.gov/media/71195/download
11. Gadgeel S, Rodríguez-Abreu D, Speranza G, et al: Updated analysis from KEYNOTE-189: Pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer. J Clin Oncol 38:1505-1517, 2020
12. Paz-Ares L, Luft A, Vicente D, et al: Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. N Engl J Med 379:2040-2051, 2018

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO REAL WORLD DATA AND DIGITAL ONCOLOGY

COMMENTARY

# Bridging research and practice: enhancing regulatory decisions with pragmatic clinical trials in oncology

## INTRODUCTION

Pragmatic clinical trials (PCTs) evaluate the effectiveness of interventions in settings that more closely resemble real-world settings, aiming to produce evidence directly applicable to clinical practice. There is growing interest in using PCTs as alternatives to explanatory clinical trials to support regulatory decision making. Explanatory clinical trials represent the conventional approach, driven by familiarity with the methodologies and acceptance by regulatory authorities.[1] Advocating for a shift away from the conventional trials toward PCTs highlights the need for evolving clinical trial designs to enhance research impact. This shift reflects growing recognition of the challenges with conventional trials, such as increasing design complexity and highly selected patient populations.

To aid in adopting PCTs, drug sponsors may consider a hybrid approach, integrating pragmatic elements into traditional randomized controlled trials to streamline research, enhance data relevance, ease patient burden, and expand access to diverse patient populations. A 'hybrid PCT' balances real-world applicability with rigorous scientific methodologies, addressing challenges with conventional trial approaches while leveraging the benefits of pragmatic approaches. In oncology drug development, incorporating pragmatic elements can accelerate the availability of new therapies and ensure the adaptability of research findings to clinical practice while meeting regulatory standards.

## DESIGN PRINCIPLES OF CLINICAL TRIALS WITH PRAGMATIC ELEMENTS

Hybrid PCTs blend conventional methodologies with pragmatic elements to meet specific research goals.[2] This integration aims to reflect real-world conditions and align with regulatory frameworks. Explanatory trials often operate under controlled conditions with stringent eligibility criteria, appropriate for new molecular entities with limited safety data early in development. Such trials require rigorous tumor measurement and patient follow-up to ensure a comprehensive assessment of response and safety. These trials often enroll homogenous populations to minimize confounding factors and isolate drug effects. This control can simplify measuring treatment effects but can result in complex protocols, limiting eligible sites and patients.



In comparison, trials with pragmatic elements can enhance the trial result applicability across broader patient populations by reflecting everyday healthcare settings. This is achieved through hallmarks such as broadened eligibility criteria, accommodating patients with comorbidities, poorer performance status, and older patient populations (Figure 1).[2-5] Other pragmatic features are flexible treatment delivery, and streamlined data collection, relevant for therapies with known mechanisms of action or those not first-in-class. The suitability of therapies for trials with pragmatic elements depends on having an established safety profile, a wider therapeutic index, and the feasibility of administration in nonacademic settings.

Clinical trials with pragmatic elements can leverage components to enhance the relevance and applicability of their findings, such as using real-world data (RWD) and real-world evidence from electronic health records, registries, and patient-reported outcomes. These sources ensure the trial setting mirrors real-world environments. Many PCTs focus on comparative effectiveness, providing direct evidence of the relative benefits and risks between treatment options. These trials prioritize outcomes such as overall survival (OS), patient experience, and quality of life, aligning with patient and clinician priorities. Tumor-based endpoints, such as progression-free survival, may not routinely be included in a fully pragmatic study; however, they may still be deployed in hybrid PCTs with other trial elements more pragmatic, such as selective safety data collection or broadened eligibility criteria. In addition, more pragmatic intermediate disease endpoints such as time-to-discontinuation and time-to-next-treatment can be considered indicators of greater treatment effectiveness.[6]

## PRAGMATICA-LUNG: A CASE STUDY IN PRAGMATIC CLINICAL TRIAL DESIGN

Pragmatica-Lung (NCT05633602) is a case study for integrating pragmatic elements into an oncology trial.[7] The trial evaluates a novel combination regimen within a real-world context, aiming for regulatory submission based on clinically meaningful outcomes.[8,9] The design reflects a fully pragmatic approach: broadened eligibility criteria to encompass diverse patient profiles and streamlined data collection prioritizing critical safety and efficacy endpoints. This helps mitigate participant and site burden, enhances enrollment rates, ensures the relevancy of findings to a broader population, and expedites the drug development timeline. By reducing administrative and financial burdens, Pragmatica-

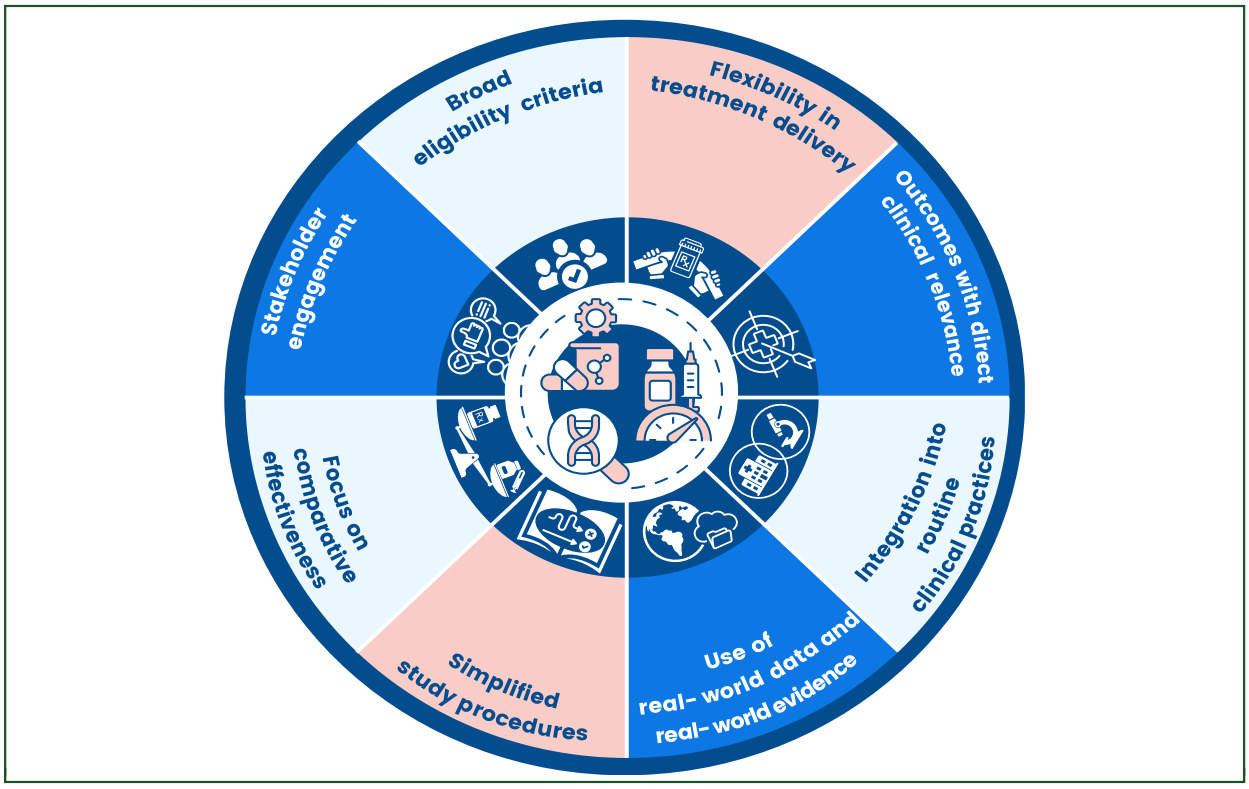



**Figure 1. Hallmarks of pragmatic clinical trials for application in hybrid designs.** These features are indicative of a pragmatic approach, although not all need to be present for a trial to be classified as pragmatic. The inclusion of these elements can vary, reflecting a spectrum rather than an all-or-nothing requirement.

Lung can make the trial more cost-effective and attractive to sites, accelerate the time to activate a trial, and be less disruptive to patients.

In Pragmatica-Lung, the endpoints align with the primary research question: does ramucirumab plus pembrolizumab extend OS compared with standard of care? Both therapies in this novel combination are Food and Drug Administration (FDA) approved and have extensive safety and efficacy data as monotherapies for non-small-cell lung cancer. Rather than measuring tumor size reduction or disease progression, the trial measures OS as the primary outcome and incidence of severe adverse events as the secondary outcome. This minimizes radiographic scans and additional visits, prioritizing survival measures and key safety signals. This regulatory-focused approach implements an efficacy endpoint that is simple to measure and acceptable to regulators. Pragmatica-Lung was rapidly implemented within 7 months and is available at >500 sites across the National Clinical Trials Network. It is on pace to complete enrollment in half the estimated timeframe across a more representative set of patients compared with historical rates, demonstrating the efficiency and ability of PCTs to reach more patients.[10]

Other examples of PCTs are the Targeted Agent and Profiling Utilization Registry (TAPUR) study, exploring the effectiveness and safety of approved cancer therapies used for genomic indications not in the FDA-approved label.[11] Although the number of PCTs in oncology is difficult to quantify, their use for regulatory decision making remains limited.[12]

## REGULATORY CONSIDERATIONS FOR CLINICAL TRIALS WITH PRAGMATIC ELEMENTS

Pragmatica-Lung extensively incorporated pragmatic elements due to both products being approved for the same population and supported by substantial efficacy and safety data. Not all trials need to be as pragmatic; integration of pragmatic elements can be tailored based on known drug characteristics and research questions. The use of pragmatic elements does not necessarily need to be limited to situations where a product is late in its development lifecycle. Early regulatory engagement is critical to align on pragmatic elements' acceptability and required for drug assessment and approval. These interactions ensure trial designs adhere to regulatory expectations while leveraging flexibilities associated with pragmatic trial elements. Given differing familiarity and acceptance among health authorities, it is essential to align with health authorities where submission has been prioritized. Programs such as FDA's Project Pragmatica and Project 5 in 5 exemplify the growing endorsement of PCTs to support regulatory decision making while reducing complexity and improving the generalizability of data.[13,14] In addition, the FDA's C3TI program has initiatives

focused on Bayesian supplementary analysis, selective safety data collection, and streamlined trials embedded in clinical practice, highlighting further commitment to enhancing trial efficiency and relevance through innovative trial designs.[15]

## CHALLENGES AND FUTURE DIRECTIONS

Conducting oncology research is challenging due to high costs and logistical complexities. Simplified study protocols and data collection processes can lessen the burden on participants and providers, improving recruitment, retention, and compliance. Aligning trials with clinical workflows minimizes disruptions and ensures settings mirror real-world environments. This can also enable additional sites to implement the study, enabling access to more patients.[16,17]

While offering advantages, trials with pragmatic elements can present challenges that must be navigated carefully, especially when intended for regulatory use. Early engagement with regulatory authorities is necessary to align on the acceptability of pragmatic elements and data adequacy for benefit—risk assessments. Operational complexities may arise when integrating research into routine care, requiring investments in infrastructure and training. Balancing pragmatism with scientific rigor remains critical. The degree of pragmatism will depend heavily on the phase of the trial and the safety profile of the treatment under study.

Variability in data quality and consistency can vary in less controlled settings depending on the types of pragmatic elements incorporated into a study design, which can complicate interpretation. For example, broad inclusion criteria, while beneficial for generalizability, can introduce variability. Use of electronic health records, digital health technologies, and other RWD sources should be evaluated to ensure they are suitable for answering the research question. Integrating RWD into trials allows researchers to observe the interaction of new therapies with standard treatments and understand the practicalities of their use in typical healthcare environments. It also aids in identifying patient subgroups that benefit most from certain treatments, a crucial aspect of personalized medicine in oncology. However, RWD quality, completeness, and consistency are concerns given that it is collected for various purposes beyond research. Lack of data interoperability between different healthcare systems and the absence of standardized collection methods also complicate aggregation and analysis of RWD. Trials with pragmatic elements should have robust methods for data verification and validation to meet regulatory standards.

To successfully incorporate pragmatic elements into future trials, several key strategies are necessary. First, there is a need to continue broadening patient inclusion criteria to ensure that trial populations accurately reflect the diversity seen in clinical practice. Second, a greater emphasis should be placed on patient-relevant outcomes. Third, leveraging technological advancements and data science to harness RWD effectively will be important. Fourth, prospective agreement among stakeholders on core data elements, processes for collection, and analyses to be carried out will help ensure successful implementation and maximize the utility of the results.

## CONCLUSION

Integrating pragmatic elements into oncology trials to create hybrid PCTs offers a promising avenue for efficient, relevant, and patient-centered drug development. Collaboration with patient advocacy groups, providers, regulators, pharmaceutical companies, and other stakeholders ensures trial designs are patient-centric and reflective of real-world practice. Sharing experiences, challenges, and successes in designing and implementing clinical trials with pragmatic elements will build a knowledge base to guide future trials.

By addressing challenges and capitalizing on opportunities, the drug development community can make significant strides in advancing cancer care. The future of oncology trials is poised to embrace pragmatic elements, aiming to bridge the gap between research and practice, ultimately improving patient care and outcomes.

M. D. Stewart[1]*, A. E. McKee[2], R. S. Herbst[3], H. S. Andrews[1], B. A. McKelvey[1], E. V. Sigal[1] & J. D. Allen[1]

[1]*Friends of Cancer Research, Washington;*
[2]*Parexel International, Washington;*
[3]*Yale Cancer Center, Yale School of Medicine, New Haven, USA*
(*E-mail: mstewart@focr.org).

Available online xxx

https://doi.org/10.1016/j.esmorw.2024.100065

## FUNDING

None declared.

## DISCLOSURE

RSH serves on the board of directors for Immunocore and Junshi Pharmaceuticals; serves as a consultant for AbbVie Pharmaceuticals, AstraZeneca, Bolt Biotherapeutics, Bristol-Myers Squibb, Candel Therapeutics, Checkpoint Therapeutics, Cybrexa Therapeutics, DynamiCure Biotechnology, eFFECTOR Therapeutics, Eli Lilly and Company, EMD Serono, Genentech, Gilead, HiberCell, I-Mab Biopharma, ImmuneOnc Therapeutics, Immunocore, Janssen, Johnson and Johnson, Loxo Oncology, Merck and Company, Mirati Therapeutics, NextCure, Normunity, Novartis, Ocean Biomedical, Oncocyte Corp, Oncternal Therapeutics, Pfizer, Regeneron Pharmaceuticals, Revelar Biotherapeutics, Ribbon Therapeutics, Roche, Sanofi, Seattle Genetics, and Xencor; reports research support from AstraZeneca, Eli Lilly and Company, Genentech/Roche, and Merck and Company; and leadership role in American Association for Cancer Research, International Association for the Study of Lung Cancer, Society for

Immunotherapy of Cancer, and Southwest Oncology Group. JDA serves as an advisory board member with n-Power (unpaid) and on the boards for Friends of Cancer Research and the Alliance for a Stronger FDA. EVS reports receiving consulting fees as an advisor for Grail, Inc. and EQRx, Inc.; honorarium from AstraZeneca; leadership and/or advisory roles at Friends of Cancer Research, MD Anderson Cancer Center, PCORI, Reagan-Udall Foundation, and Stand Up 2 Cancer. All other authors have declared no conflicts of interest.

## REFERENCES


1. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20(8):637-648.
2. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147.
3. Friends of Cancer Research. Incorporating pragmatic elements in study designs to enhance oncology randomized clinical trials. Available at: https://friendsofcancerresearch.org/wp-content/uploads/Incorporating_Pragmatic_Trial_Elements_Study_Designs_Oncology.pdf. Accessed May 12, 2024.
4. Kim ES, Bruinooge SS, Roberts S, et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research Joint Research Statement. *J Clin Oncol*. 2017;35(33):3737-3744.
5. Kim ES, Uldrick TS, Schenkel C, et al. Continuing to broaden eligibility criteria to make clinical trials more representative and inclusive: ASCO-Friends of Cancer Research Joint Research Statement. *Clin Cancer Res*. 2021;27(9):2394-2399.
6. Rivera DR, Henk HJ, Garrett-Mayer E, et al. The friends of cancer research real-world data collaboration pilot 2.0: methodological recommendations from oncology case studies. *Clin Pharmacol Ther*. 2022;111(1):283-292.
7. The ASCO Post. *Pragmatica-lung trial begins enrolling patients with NSCLC who did not respond to previous therapy*. Available at: https://ascopost.com/news/april-2023/pragmatica-lung-trial-begins-enrolling-patients-with-nsclc-who-did-not-respond-to-previous-therapy/. Accessed May 12, 2024.
8. Reckamp KL, Redman MW, Dragnev KH, et al. Phase II randomized study of ramucirumab and pembrolizumab versus standard of care in advanced non-small-cell lung cancer previously treated with immunotherapy-Lung-MAP S1800A. *J Clin Oncol*. 2022;40(21):2295-2306.
9. Herbst RS, Blanke CD, Sigal EV. Novel approach to accelerate lung cancer research: Lung-MAP and the potential of public-private partnerships. *Clin Cancer Res*. 2024;30(1):29-32.
10. Reckamp KL, Redman MW, Dragnev KH, et al. SWOG S2302, PRAGMATICA-LUNG: a prospective randomized study of ramucirumab plus pembrolizumab (PR) versus standard of care (SOC) for participants previously treated with immunotherapy for stage IV or recurrent non-small cell lung cancer. *J Clin Oncol*. 2024;42:TPS8657.
11. Mangat PK, Halabi S, Bruinooge SS, et al. Rationale and design of the Targeted Agent and Profiling Utilization Registry (TAPUR) study. *JCO Precis Oncol*. 2018;2018(2):1-14.
12. Sankar K, Redman MW, Dragnev KH, et al. Pragmaticism in cancer clinical trials. *Am Soc Clin Oncol Educ Book*. 2024;44(3).
13. U.S. Food and Drug Administration. Project pragmatica. Available at: https://www.fda.gov/about-fda/oncology-center-excellence/project-pragmatica. Accessed May 12, 2024.
14. U.S. Food and Drug Administration. *Project 5 in 5*. Available at: https://www.fda.gov/about-fda/oncology-center-excellence/project-5-5. Accessed May 12, 2024.
15. U.S. Food and Drug Administration. C3TI demonstration program. Available at: https://www.fda.gov/about-fda/cder-center-clinical-trial-innovation-c3ti/c3ti-demonstration-program. Accessed July 16, 2024.
16. Mahoney MR, Sargent DJ, O'Connell MJ, Goldberg RM, Schaefer P, Buckner JC. Dealing with a deluge of data: an assessment of adverse event data on North Central Cancer Treatment Group trials. *J Clin Oncol*. 2005;23(36):9275-9281.
17. Kaiser LD, Melemed AS, Preston AJ, et al. Optimizing collection of adverse event data in cancer clinical trials supporting supplemental indications. *J Clin Oncol*. 2010;28(34):5046-5053.

FRIENDS of CANCER RESEARCH

# Pragmatism in Postmarket Trials

FRIENDS OF CANCER RESEARCH WHITE PAPER | 2024

## Executive Summary

Incorporating pragmatic clinical trial elements (i.e., pragmatic elements) into trial designs provides an opportunity to reduce patient, site, and investigator burden, while increasing the generalizability and applicability of trial results to the intended use population by more closely reflecting routine clinical practice. Considerations for incorporation of pragmatic elements include the specific research question, trial objectives and clinical setting, available safety and efficacy data on the treatment of interest, and intended use of the trial results, including whether the data will be submitted for regulatory review. These factors will influence the appropriateness and operationalization of incorporating selected pragmatic elements and the level of risk assumed regarding trial integrity, data quality and missing data, and the rigor of endpoint assessment.

Friends of Cancer Research assembled a working group of experts, including members from the U.S. Food and Drug Administration (FDA) and National Cancer Institute (NCI), drug developers, patient advocates, health technology data experts, and academic clinicians, to identify specific trial objectives in the postmarketing setting to frame a discussion on the benefits and risks of incorporating pragmatic elements into future trials. Introduction of pragmatic elements may be most feasible initially in the postmarketing setting, where more is known about the safety of the product, and additional questions remain about its optimal use in practice. Objectives in the postmarketing setting include postmarketing requirements or commitments issued by the FDA following initial approval, or new interventional studies initiated by sponsors seeking expansion of a product's indication to additional patient populations.

The working group evaluated the following scenarios as example research objectives to guide discussion of incorporating pragmatic elements in postmarket clinical trials. For each, we provide specific considerations for increasing pragmatism:

- Conduct a clinical trial that enrolls racially and ethnically underrepresented patients in proportion to their representation in the U.S. population of patients within the disease indication, in sufficient numbers to characterize the safety and efficacy of the approved drug in the patient population.
- Conduct a clinical trial to further characterize the risk of a cumulative toxicity and potential mitigation measures in patients receiving the drug.
- Conduct a clinical trial to characterize the safety and efficacy of the drug in a biomarker-selected population expanded from the biomarker cutoff used for the initial indication.

As is true for any trial objective, for each of these three scenarios, not all pragmatic elements may be appropriate. The scenarios illustrate opportunities to introduce pragmatism into a clinical trial and provide considerations applicable to additional trial objectives. While incorporating pragmatic elements may decrease burden, there may be an increase in risk for the data to be used for regulatory decision-making. Therefore, thoughtful consideration should be given to the potential benefits and risks and early interactions with FDA on trial design will be essential.

## Authors

**Ashita S. Batavia**, Johnson and Johnson Innovative Medicine
**Trixia Camacho**, Bristol Myers Squibb
**Gil Carrigan,** Amgen
**Allen Chen**, AstraZeneca
**Li Chen**, Amgen
**Scot Ebbinghaus**, Merck
**Roy Herbst**, Yale Cancer Center
**Camille Jackson**, Flatiron Health
**Qi Jiang**, Pfizer
**Laura Lasiter**, AstraZeneca
**Eva A. May**, Patient Advocate
**Kristin McJunkins**, Patient Advocate
**Brittany McKelvey,** Friends of Cancer Research
**Jodi McKenzie**, Eisai
**Neal J. Meropol**, Flatiron Health
**Phuong Khanh Morrow**, Takeda Pharmaceuticals
**Peter J. O'Dwyer**, ECOG-ACRIN Cancer Research Group; University of Pennsylvania
**Sheila Prindiville**, NCI
**Kelly Johnson Reid**, Johnson and Johnson Innovative Medicine
**Richard L. Schilsky**, University of Chicago

## Acknowledgements

**Timil Patel,** U.S. FDA
**Donna Rivera**, U.S. FDA

## Introduction

Traditional randomized controlled trials (RCTs) have generally included standardized patient selection, specific assessment and monitoring intervals, and substantial follow-up to generate robust data to inform regulatory decision-making. However, clinic visits and data collection requirements that are required beyond routine clinical care can be burdensome to trial participants, investigators, and trial sites, and can limit the participation of some patients and trial sites.[1] Overly strict eligibility criteria can further reduce both participation and the generalizability of clinical trial results to the intended use population.[2] Furthermore, unnecessary data collection and frequent monitoring can be resource intensive (e.g., time and cost) for sponsors.

Incorporating pragmatic clinical trial elements (henceforth pragmatic elements) into trial designs, where appropriate, can introduce operational efficiencies in a less burdensome framework, and generate data that are more reflective of intended use populations.[3,4] The U.S. Food and Drug Administration (FDA) has signaled interest in incorporating pragmatic elements into clinical trials through the Center for Drug Evaluation and Research (CDER) Center for Clinical Trial Innovation (C3TI) Streamlined Trials Embedded in clinical Practice (STEP) demonstration project[5], launch of the Oncology Center of Excellence (OCE) Project Pragmatica[6], and more recently Project 5 in 5[7], focusing on pragmatic clinical trials in oncology. Friends of Cancer Research (*Friends*) assembled a collaborative working group in 2023 to draft a white paper, "Incorporating Pragmatic Elements in Study Designs to Enhance Oncology Randomized Clinical Trials[8]," which laid out considerations to inform the appropriateness of incorporating pragmatic elements into RCTs for evidence generation across the lifecycle of a drug.

Considerations for incorporating pragmatic elements into a clinical trial include the specific research question, trial objectives and clinical setting, the available safety and efficacy data on the treatment of interest, and the intended use of the trial results, specifically whether or not the data will be submitted for regulatory review. These factors will influence the appropriateness and operationalization of incorporating pragmatic elements as well as the level of risk assumed by trial sponsors regarding trial integrity, data quality and missing data, and the rigor of endpoint assessment. *Friends* assembled a new working group of experts, including members from the FDA and National Cancer Institute (NCI), drug developers, patient advocates, health technology data experts, and academic clinicians, to build on the foundation and operationalize concepts from the 2023 white paper. To better discuss the opportunities to incorporate pragmatic elements into future clinical trials, the group focused on the postmarket setting, a specific phase of drug development with high potential value for incorporating pragmatism.

## Defining Opportunity in Postmarket Clinical Trials

Incorporating pragmatic elements into prospective studies offers the opportunity to support evidence generation across the life cycle of a drug. The introduction of pragmatic elements may be most feasible in the postmarket setting, where more is known about the safety of the product, but

additional questions remain about its optimal use in practice. Such questions might include a better understanding of the safety and/or efficacy of an agent in populations underrepresented in the registrational trial(s) or information about potential new uses of the treatment. Additional research questions may be driven by the interests of the drug sponsor, regulatory authorities (i.e., through postmarketing requirements or commitments), or clinical investigators, and evidence generated may be used to support regulatory decision-making, such as updating a label or approving a new indication. Importantly, results from pragmatic studies can also inform decisions outside of regulatory agencies. Examples include informing clinical practice, supporting updates to clinical practice guidelines, or providing evidence for coverage decisions by payers. Given the level of safety and efficacy data already available from the pivotal trial(s), introducing pragmatic elements in the postmarket setting may be viewed as a lower risk for sponsors regarding trial integrity than in the premarket setting.

## Data Considerations to Inform Pragmatic Trial Designs

Data regarding safety and efficacy from prior, completed registrational trials should inform the appropriateness of implementing specific pragmatic elements in a postmarket trial. Sponsors could consider which data elements from the pivotal trial were or were not critical for determining safety and efficacy. Through formal discussions with regulatory agencies, trial protocols may be revised to improve efficiency. This could involve reducing the collection of unnecessary data elements or allowing for greater heterogeneity in data collection, when appropriate. For trials intended for regulatory approval, early engagement with the relevant FDA review division is essential to discuss currently available data and clarify the evidentiary requirements for demonstrating safety and/or efficacy needed to support a new regulatory submission.

For instance, available safety data from a pivotal trial may demonstrate no discernible difference in toxicity in patients with mild versus moderate renal dysfunction, suggesting that broadening eligibility criteria to include patients with higher levels of renal dysfunction may be appropriate if also supported by non-clinical data and knowledge of the drug's pharmacokinetics. Alternatively, existing safety data may show that an adverse event occurs commonly in relation to the administration of a therapeutic agent, suggesting that additional trials should continue to include frequent assessment and mitigation strategies for the event.

Phase II trials, often investigator-initiated or led by NCI cooperative groups, or real-world data (RWD), may suggest areas of additional efficacy or effectiveness, respectively, and/or novel safety findings, which could be used to support and identify potential patient populations for further study in a prospective clinical trial and inform the degree and type of pragmatism to incorporate into the design. The use of RWD can improve understanding of the potential impact of broadening eligibility criteria on representativeness and on outcomes[9], as well as inform flexibility in follow-up approaches and frequency of assessments. For example, a recent study found that heterogeneity in real-world visit frequency for patients with newly diagnosed multiple myeloma contributed to surveillance bias but that bias could be quantified in evaluating endpoint measurements.[10] Information such as this

example and others[11] could inform a trial design with pragmatic elements where flexibility could be introduced in the assessment interval for patients, and the study could be more tolerant of shifts in visit schedule. This could allow for reduced patient burden without substantively compromising efficacy insights.

For trials incorporating multiple pragmatic elements, the cumulative impact on the sensitivity to detect treatment effect must be carefully considered. For example, pragmatic elements such as introduction of broader eligibility criteria or allowing flexibility in assessment intervals, may increase variability and decrease statistical sensitivity to detect small treatment effects, thus requiring a larger sample size. Larger clinical trial populations typically result in trial delays and additional costs, but this concern could be mitigated if the cumulative effect of all pragmatic elements incorporated ultimately result in more rapid accrual and/or reduced attrition. Products or treatment sequences that are expected to have a large effect size may be more appropriate for higher degrees of pragmatism. Conversely, a highly pragmatic design may not be appropriate for a non-inferiority trial design.

## Introducing Pragmatic Elements into Postmarket Trial Designs

The pragmatic-explanatory continuum indicator summary (PRECIS)-2[12] is a conceptual framework that provides nine domains to consider for determining the degree of pragmatism in a given trial design, including eligibility, recruitment, setting, organization, delivery, adherence, follow-up, primary outcome, and primary analysis. The level of pragmatism is graded on a scale within each domain, and within each domain the amount of pragmatism that is appropriate or necessary may vary depending on the context in which the study is conducted. Use of pragmatic elements should aim to create the highest generalizability and reduction in burden while maintaining appropriate rigor to answer the prespecified research objectives in the population of interest. Many applications of pragmatic approaches and their considerations are relevant across research questions in the postmarket setting. Table 1 outlines these considerations by the PRECIS-2 domains. Below are further insights into how pragmatic elements may be incorporated across research questions.

### Operational Efficiencies through Technology

To facilitate research participation in routine care settings, digital health and data technologies can enhance operational efficiencies. These tools and technologies include the use of telemedicine, electronic health record (EHR) to electronic data capture (EDC) data transfer software, and automated patient clinical trial matching based on EHR documentation. Telemedicine can support remote consenting, clinical assessments, monitoring, and follow-up, and data collection. EHR-to-EDC software leverages routine clinical workflows, automating transfer of clinical data quickly and accurately to the research database, helping to avoid time-consuming, error-prone, and duplicative data entry tasks. Patient trial matching software can aid in recruitment by helping sites evaluate the suitability of a particular study by identifying trial-eligible patients at the point of care. This approach can reduce site burden and also mitigate potential unconscious biases associated with patient

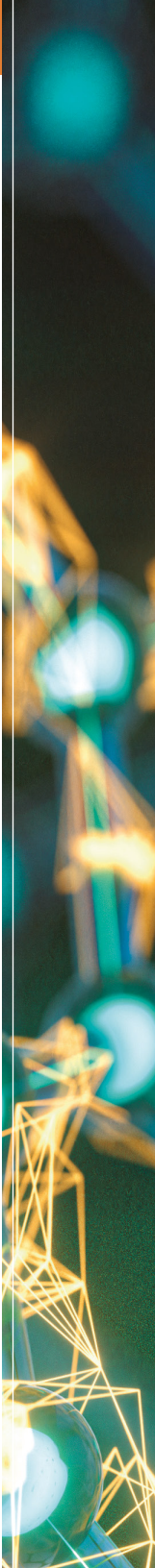
ascertainment, ultimately supporting more equitable and representative study participation. These technologies can enable operational efficiencies that allow sites to identify, recruit, enroll, and evaluate trial participants more effectively, introducing pragmatic elements across PRECIS-2 domains.

## Streamlined Safety Data Collection

If the data suggest a similar adverse event profile for a drug in the new trial population of interest compared with the patient population included in the registrational trial, selective safety data collection may be appropriate. The International Council for Harmonization (ICH) draft guidelines for Optimization of Safety Data Collection- E19[13] note data collection may be limited or stopped for non-serious adverse events, routine laboratory tests, concomitant medications, or physical examinations, as appropriate. In these scenarios, capturing serious adverse events and grade 3 or higher adverse events and reducing collection of low-grade events may be appropriate. These recommendations are similar to those recently proposed by the NCI Streamlining Clinical Trials Working Group[14]. However, collection of only high-grade events may diminish the ability to assess treatment tolerance and chronicity of low-grade adverse events. Therefore, strong existing evidence to support the safety profile and a rationale for why the expanded patient population will likely have a similar safety profile should be provided.

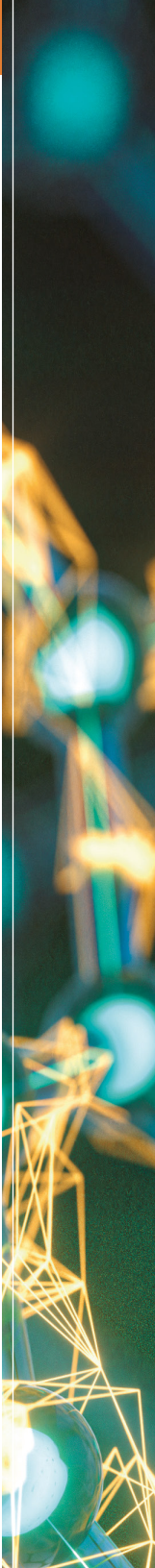
Table 1. Select Considerations for Incorporating Pragmatic Elements into Postmarket Trials by PRECIS-2 Domains.

| PRECIS-2 Domain | Pragmatic Element to Introduce | General Considerations | Potential Impacts on Patients, Sites, and Sponsors |
|---|---|---|---|
| Eligibility | • Less restrictive eligibility criteria (e.g., expanding lab values, organ function)<br>• Reduced number of eligibility criteria (e.g., not requiring certain lab values) or requirements for extra tests to confirm eligibility | • Eligibility criteria should be based on the known preclinical and early clinical safety data. A rationale for exclusion criteria focused on patients' safety should be provided | • Patient- Potentially lower burden, increased accessibility; fewer screening procedures<br>• Site- Reduced screening simplifies workflow, less pre-study documentation needed<br>• Sponsor- Faster accrual; more diverse population |
| Recruitment | • Tech/AI enabled trial matching for screening patients using existing EHR data<br>• Integration of research fields into the EHR<br>• Simplify informed consent document<br>• Permit electronic consent | • Enables rapid identification of potentially eligible patients<br>• Infrastructure needed to support enabling technology<br>• Training of site staff to engage potentially eligible patients and support informed consent<br>• Improved patient understanding of study designs, risks, benefits and alternatives<br>• Facilitates consenting process | • Patient- Less burden and reduced complexity in informed consent<br>• Site- Reduced site burden with technology enabled features but may require more upfront resources and infrastructure investments. Facilitates better communication between physicians and patients<br>• Sponsor- Faster accrual |
| Setting | • Multi-site trial conducted in community setting, including community-based clinical practices | • Meeting patients where they receive care increases the likelihood of accrual and retention<br>• Community sites serve a more representative patient population | • Patient- Reduces travel/cost, maintains provider relationship<br>• Site- Allows site to retain patients, provide continuity of care<br>• Sponsor- Increased complexity of trial management* |



| PRECIS-2 Domain | Pragmatic Element to Introduce | General Considerations | Potential Impacts on Patients, Sites, and Sponsors |
|---|---|---|---|
| Organization | • Care given by community and local providers<br>• Technology-enabled trial management | • Leverage technology for operational efficiency- i.e., EHR to EDC integration to reduce duplicative data entry<br>• Invest in robust technology infrastructure | • Patient- Less burden by allowing local care<br>• Site- Burden associated with maintenance of site research infrastructure, training, but could be alleviated with EHR to EDC<br>• Sponsor- Increased monitoring burden (e.g., CRO may be needed) with decentralization or decreased with EHR to EDC; Increased training cost for less experienced research sites* |
| Delivery | • Embedded in routine care, with treatment and assessment cadence aligned with routine care<br>• Site-based treatment visits would be scheduled per approved dosing schedule, but other visits may be remote and flexible on timing, potentially using technology-enabled delivery<br>• Shipping of oral medications as appropriate | • Invest in robust technology infrastructure<br>• Develop and implement effective patient engagement strategies<br>• Potentially widens the pool of participating sites as trials embedded in practice are not as resource intensive<br>• Adequate training and support to enable remote and flexible visits | • Patient- Reduced burden with fewer visits/ assessments<br>• Site- Does not require significant additional data capture<br>• Sponsor- May increase burden to provide more resources/ infrastructure but technology has the potential of reduced cost and increased efficiency |
| Adherence | • Patient-centered adherence strategies that prioritize patient autonomy, engagement and self-management | • Treatment at local site or at home promotes adherence, treatment and assessment compliance, as well as retention | • Patient- Reduced burden<br>• Site- Does not require additional work |

| PRECIS-2 Domain | Pragmatic Element to Introduce | General Considerations | Potential Impacts on Patients, Sites, and Sponsors |
|---|---|---|---|
| | • Technology-enabled adherence support to enhance patient adherence and engagement | | • Sponsor- May require more resources and infrastructure investment* |
| Follow-up | • Follow-up visits reflective of standard of care, in routine care location<br>• Flexibility in monitoring cadence, and intervals, including remote patient monitoring<br>• Selective safety monitoring- as per NCI Streamlining Clinical Trials Working Group (and ICH E6 and E19) recommendations<br>• Use of DHTs to prompt patients, inclusion of PROs | • Flexibility in monitoring cadence may be dependent on the drug's mechanism of action, PK, and the natural history of disease.<br>• Patients may not be amenable/adherent to rigorous schedule of assessments<br>• The appropriateness of PROs inclusion will depend on the research question. This may introduce additional patient burden, but may offer insights into low grade AEs, Quality of Life. | • Patient- Reduce burden of follow-up (frequency, duration) visits<br>• Site- Reduce burden of follow-up visits<br>• Sponsor- May require more resources on trial management and logistics, and infrastructure investment* |
| Primary Outcome | • Outcome should not require specialized or central review (rw-RECIST, OS)<br>• Efficacy endpoints based on data routinely collected in clinical practice (e.g., rwPFS, time to subsequent therapy, OS)<br>• AEs: Might collect G3-5 or those that result in treatment change<br>• Flexible outcome assessment schedules | • Need statistical analyses to understand what magnitude of effect would be acceptable because of heterogeneity in patient population and/or assessments at trial design stage. | • Patient- Minimizes extra visits/tests beyond routine care<br>• Site- Minimizes data collection as is in routine clinical care, reduce workload<br>• Sponsor- May require more resources and infrastructure investments to ensure relevant clinical data are being collected* |



| PRECIS-2 Domain | Pragmatic Element to Introduce | General Considerations | Potential Impacts on Patients, Sites, and Sponsors |
|---|---|---|---|
| | • Technology enabled outcome monitoring | | |
| Primary Analysis | • Intention to treat principle<br>• Clinically relevant primary outcome<br>• Minimize secondary objectives | • Define statistical endpoints, and a statistical model to guide interpretation of routinely collected data | • Patient- Patient-centric<br>• Site- Reduces burden of collection<br>• Sponsor- Likely no impact on burden |

* It is acknowledged by the working group that the initial burden in logistics, infrastructure and trial management that may be associated with pragmatic or decentralized elements are likely to improve over time and with experience.

## Postmarket Trial Objectives

To frame the working group's discussion on incorporating pragmatic elements in postmarket clinical trials, the following trial objectives were selected by the working group to explore as examples of where pragmatism would be feasible and impactful. For each, we provide specific considerations for increasing pragmatism in the PRECIS-2 domains, specifically focusing on unique considerations related to eligibility, setting, delivery, follow-up, and primary outcome.

### Trial Objective 1: Conduct a clinical trial in a specific patient population to further characterize the safety and efficacy of the treatment.

Evaluating the safety and/or efficacy of a drug in a specific population underrepresented in the trial is a common research question in many postmarketing studies and may be appropriate for incorporating a more pragmatic approach to evidence generation for a variety of reasons.[15] There may be an initial signal in a registrational trial that demonstrated differential safety or efficacy in a subgroup of the patient population, or there may have been too few patients in this subgroup to make robust conclusions. Additionally, some specific patient populations may have been excluded from the registrational trial due to strict eligibility criteria, prompting interest to characterize product safety and/or efficacy in this population. In such cases, a postmarketing trial further studying the population may lead to important FDA label updates. Evidence from the registrational trial(s) will influence the extent to which pragmatic elements are appropriate to incorporate into a postmarket study. For example, differential safety identified in subgroups within the registrational trial may impact the types and frequency of safety data collected in the postmarket trial, making safety assessment not amenable to a highly pragmatic approach.

#### Examples of specific patient populations for Trial Objective 1:

- Underrepresented Racial and Ethnic Group
    - Study including a racial and/or ethnic population underrepresented in the registrational trial.
- Underrepresented Age Group
    - Study including older adult populations underrepresented in the registrational trial.
- Patients with Organ Dysfunction
    - Study including patients excluded from the registrational trial due to organ dysfunction.
- Patients from a Specific Geographic Location
    - Study including patients underrepresented in the registrational trial from a specific geographic location.

*Case Study #1: Conduct a clinical trial that enrolls racially and ethnically underrepresented patients in proportion to their representation in the U.S.*

*population of patients within the disease indication, in sufficient numbers to characterize the safety and efficacy of the approved drug in this patient population.*

A common objective of postmarketing requirements or commitments is postmarket investigation with sufficient numbers of patients in an underrepresented racial or ethnic group[16]. Considerations for incorporating pragmatic elements by PRECIS-2 domains, specific to the case study:

**Eligibility–** Understanding the factors associated with underrepresentation of the patient population of interest will be informative. Patients may not be eligible as restrictive inclusion/exclusion criteria may disproportionately exclude underrepresented populations. Less restrictive eligibility criteria (e.g., expanding laboratory value requirements, comorbidities, performance status) could increase eligibility. This approach may come with potential risks, not unique to studying minority populations, but due to the broadening of eligibility criteria resulting in trial participants with different risk/benefit profiles compared to the initial trial. For patients, there may potentially be differential outcomes (both adverse events and clinical outcomes) than in the registrational clinical trial that may be attributable to other factors (e.g., organ dysfunction) given the more heterogeneous patient population. For sponsors, the increased heterogeneity of the trial population may obscure modest clinical benefits, raising the risk of trial failure and possibly making results interpretation more challenging.

**Setting–** Another factor contributing to the underrepresentation of the patient population of interest may be the trial sites selected for patient enrollment. Patient populations historically underrepresented in oncology clinical trials, including racial and ethnic marginalized groups, are more likely to be treated at community sites with limited access to clinical trials or that are relatively inactive (e.g., sites that do not have clinical trial programs or are have programs with low enrollment).[17] Expanding access to clinical trials at these community sites by designing studies better suited to routine care settings could increase the ability to recruit more representative patient populations. Meeting patients where they receive routine care in the community also increases the likelihood of accrual and retention, reducing costs and burden for patients while maintaining the patient-provider relationship and continuity of care.[18] However, some community sites may lack the infrastructure to effectively conduct clinical trials, and there may be increased complexities of trial management for sponsors. The diversity of sites may lead to increased regulatory risks such as non-compliance or trial failure due to difficulties in maintaining protocol adherence.

**Follow-Up and Primary Outcome–** Design of a trial better suited to routine care settings will be driven by the degree of protocol specified safety assessment and primary outcome measures. If the existing safety data and mechanism of action do not suggest a differential safety in the patient population of interest, selective safety data monitoring, as per the NCI Streamlining Clinical Trials Working Group[14] and ICH E19[13] guidelines, may be appropriate, such as assessing only grade 3 or higher adverse events or those that result in a treatment change. The inclusion and frequency of collection of patient-reported outcomes (PROs) could also be streamlined, if appropriate. The efficacy endpoints may also be more pragmatic, assessing outcomes that do not require specialized

or central review, such as overall survival, or real-world (rw) assessment of tumor response that employ RECIST criteria based on tumor measurements, but permit more flexibility than standard RECIST criteria (e.g. scan cadence as per routine practice rather than prespecified)[19]. Assessment of rw-response based on the clinician assessment of response may also be used to gauge efficacy in place of RECIST measurements. This measure could be further supported by a retrospective review of imaging data where available, acknowledging that imaging type and frequency would not be prespecified. This can minimize the extra visits, paperwork, and tests for patients beyond what is expected in routine care, also minimizing the data collection and workload for sites.

It is acknowledged that there may be areas of potential variability associated with reduced data collection. For instance, there may be delayed identification of imaging progression and treatment change, due to non-standardized assessment schedules. For sponsors, there may be a risk that outcomes are not directly comparable to registration-directed clinical trials given the potential increase in heterogeneity. Reduced safety data collection may diminish the ability to assess treatment tolerance and chronicity of low-grade adverse events, although expected symptomatic toxicities may be characterized with electronic PRO data. For this reason, reduced safety collection may be best suited for mature products (e.g., later in lifecycle management) with a well-characterized safety profile. An a priori statistical analysis plan with strong clinical rationale will be important to understand what magnitude of effect would be acceptable because of the potentially less fit population and heterogeneity in assessments.

## Trial Objective 2: Conduct a clinical trial to further characterize a specific adverse event/toxicity and its management.

Conducting additional studies focused on a specific toxicity or adverse event seen in the registrational trial to better characterize its frequency and management is also a common postmarketing study objective. Given that the impetus for the study often comes from a signal from the registrational trial, leveraging the existing data on the temporality (frequency, onset, reversibility, chronicity) and mitigation strategies of the toxicity can inform the appropriate pragmatic elements to include in a study design. This study may be in a specific patient subpopulation found to have differential toxicity, such as those with organ dysfunction, or be more broadly studied in the intended use population. Evidence generation may result in a label modification for management of the adverse event and/or could inform clinical management and/or practice guidelines. The type of adverse event under study will dictate the ability to incorporate flexibility in trial design.

### Examples of specific adverse event categories for Trial Objective 2:

- Long-term Toxicities
    - Specific adverse events that may be late or cumulative.
- Short-term Toxicities

- Specific adverse events that occur while on treatment (acute) within a fairly reproducible timeframe but were rarely seen or incompletely characterized in the registrational trial(s).

*Case Study #2: Conduct a clinical trial to further characterize the risk of a cumulative toxicity and potential mitigation measures in patients receiving the drug.*

This case study focuses on long-term, significant chronic toxicities, and may be applicable to toxicities or adverse events that require long-term follow-up, such as neurological toxicities. Considerations for incorporating pragmatic elements by PRECIS-2 domains, specific to the case study:

**Eligibility–** There would likely be minimal expansion of eligibility, as the risk of chronic toxicity needs to be better understood in the patient population studied in the registrational trial. The significant expansion of eligibility may run the risk of coming to an erroneous conclusion about the presence, absence, or quantitative parameters (e.g., frequency, severity) of a safety risk. There may be an opportunity to broaden eligibility to allow for the assessment of the relationship of risk to the severity and chronicity of the toxicity and to better understand potential confounding factors. If patients with an increased risk were allowed to enroll, this may require more careful and frequent monitoring to better assess the nature and severity of the toxicity and predefined design and statistical plans. This may also increase the risk to these patients, as high-risk patients may experience worse or more prolonged toxicity. By including high-risk patients, sponsors may also risk higher toxicity findings in the product label. However, this could be offset by comfort in the prescribing community to expand treatment outside of the strict eligibility criteria of the trial if safety is felt to be similar or marginally higher. Nonclinical pharmacology and toxicology data will inform the rationale for a more narrow or broad eligibility criteria.

**Setting and Delivery–** Robust data from the registrational trial(s) on the toxicity, including the time to onset, management, mitigation strategies, and outcomes, will dictate the level of flexibility and pragmatism appropriate for the postmarket trial design. A prospectively designed highly pragmatic trial may approach the simplicity of a disease registry, with prespecified evaluations capturing the relevant safety data while reducing the level of burden associated with an explanatory trial. However, more specialized testing may be required to assess causation or functional impact, especially when there is a desire to characterize the frequency of the event in a representative population. Specialized testing may also be required to adequately assess the severity and potential cause of an individual toxicity (e.g., for neurological toxicity, referrals to the neurologist, nerve conduction velocity studies, nerve biopsies, EEG, circulating neurotoxin levels) and therefore certain community settings with lesser access to specialists may not be appropriate.

**Follow-up and Primary Outcome–** If the toxicity onset window is well characterized with a fairly standard cadence across patients and easily captured through standard of care assessments, it may be appropriate to conduct follow-up visits focusing on more rigid assessment windows within the predicted onset window and less stringent assessments outside of onset based on the biology and

pharmacology of the medical product. This approach will be more easily implemented if the drug label characterizes the toxicity and its management, which will likely lead to a more standardized approach to assessment in routine care as clinicians use the label as guidance. The use of digital health technologies (DHTs) can aid in prompting patients to provide PROs and other assessments of treatment-related symptom and functional outcomes to capture low grade adverse events and their impacts that persist. Overall, this will reduce the patient and site burden of follow-up by reducing the frequency or duration of in-person follow-up visits. However, if the toxicity onset is variable and not well captured in standard of care assessments, assessment windows will likely need to be prespecified throughout, thus necessitating less pragmatism. As data generation is focused on safety, efficacy data capture can be reduced, further minimizing data collection and trial complexity. Capturing toxicities in routine practice settings allows for a more generalizable understanding of the safety of a therapeutic agent, and the opportunity to characterize exacerbating and mitigating factors. However, variability in routine practice and local assessment could impact the interpretability of the study.

## Trial Objective 3: Conduct a clinical trial intended to expand the indication to characterize the safety and efficacy of the treatment in a similar disease setting.

Another common objective for post-approval clinical trials is to generate safety and efficacy data to provide evidence supporting an approved drug in a new patient population. This trial objective facilitates identification of patients that are responsive to the drug beyond the label indication, meaning that more patients that could benefit from a safe and effective therapy are identified. Expanding a drug indication requires strong scientific and clinical justification with an adequate and well controlled investigation(s) that provide substantial evidence of drug efficacy with an acceptable safety profile to provide meaningful clinical benefit. The specific populations of interest may be identified in RWD or sponsor-supported expanded access programs, where retrospective analysis of efficacy and safety data may be feasible. A trial design to support this objective will likely be a randomized, prospective study. The appropriate level of pragmatism for such a trial would depend on the primary efficacy endpoint, as well as what is already known about the adverse event profile of the drug(s) and how or whether it would be expected to differ in the new population of interest.

### Examples of new uses for Trial Objective 3:

- Changing the Biomarker Cut Point for a Biomarker-Selected Population
    - Study medical product in a biomarker-selected population outside of the biomarker cutoff for the initial indication or defined by a new biomarker.
- New Therapeutic Combination
    - Study two medical products already approved in the indication of interest in a novel combination.

- New Drug Formulations
    - Study medical product already approved in the indication of interest, with a new formulation (e.g., intravenous to subcutaneous).

*Case Study #3: Conduct a clinical trial to characterize the safety and efficacy of the drug in a biomarker-selected population expanded from the biomarker cutoff used for the initial indication.*

To conduct a trial expanding the biomarker cutoff of the initial indication to a larger biomarker-selected population, there must be strong scientific and clinical rationale to support the new cutoff. This objective requires precision around both the biomarker status of the patients and intermediate tumor-based endpoints, if used (typically RECIST based ORR and/or PFS), to evaluate smaller but important differences in efficacy between the new biomarker subgroup and the approved biomarker-selected population. Considerations for incorporating pragmatic elements by PRECIS-2 domains, specific to the case study:

**Eligibility–** Select eligibility could be expanded from lessons learned in the accumulated clinical experience, but would likely be kept more similar to the registrational trial, except for the expansion of the biomarker selected population. There is a risk to patients in the new biomarker population, that they do not achieve adequate efficacy to overcome the known toxicity of the treatment. As such, the subgroup of patients that would be expanded by the new cut point would need to be analyzed separately to assure that the overall efficacy is not predominately attributed to the previously approved population that used a higher threshold. While local testing may lower patient and site burden with fewer screening procedures, the precision of the biomarker is critical for this research objective and tests with variable performance could negatively impact the reliability of trial results. If the study has regulatory intent, early discussion with FDA would be important to obtain advice on companion diagnostic development.

**Follow-up and Primary Outcome–** Safety data collection may be streamlined if the expanded biomarker selected population is expected to have similar safety signals. In this case, grade 3 or higher adverse events, serious adverse events, and those leading to dose changes or discontinuation should be collected. If the trial is intended to support a label update, the extent of safety data collection should be discussed with regulators prior to the start of the study. Efficacy outcomes will likely necessitate a more explanatory approach given the importance of the precision around the efficacy outcome to assess the risks and benefits in the new biomarker-selected population.

## Balancing Risk with Opportunity

The workgroup discussion highlighted the context-dependent nature of integrating pragmatic elements into prospective clinical trials. Pragmatic elements can help to reduce burden for patients and sites while answering critical research questions, but may come with uncertainty and potential risks. When determining the appropriateness of incorporating pragmatic elements, balancing the

potential risks of increased data variability with the benefits in reduced complexity and burden is important. Uncertainties inherent in new approaches to trial conduct naturally create perceived risks to incorporating pragmatic elements, however these risks may not be founded or supported by data. It is expected that perceived risks and uncertainties as well as operational complexity will be reduced with experience as more trials integrate pragmatic and decentralized elements.

A commonly stated perceived risk for sponsors is conducting trials outside of specialized centers in community-based clinical practices that may not be well versed in clinical trial conduct. Concerns include protocol deviations due to a site's inexperience with clinical trials, regulatory non-compliance or trial failure due to difficulties in maintaining protocol adherence, or data quality and integrity concerns. These risks are not inherent to conducting a trial at a community site, but rather whether the site has established infrastructure and appropriate resources to conduct the trial. Importantly, highly pragmatic designs require less protocol-directed conduct which can facilitate community site participation that is closer to routine clinical care. While there may initially be a cost to the sponsor to stand up the required infrastructure at a community site, introducing operational efficiencies and technology enablement can reduce costs over time and provide long-term benefit to support enrollment and retention at these sites. Sponsors should support inclusion of community sites and balance perceived risks with the opportunity to enhance accrual and enrollment of more diverse patient populations.

## Conclusions and Future Directions

Prospective trial designs that incorporate pragmatic elements provide the opportunity to reduce patient, site, and investigator burden and increase the generalizability of trial results by more closely reflecting routine clinical practice. By aligning research more with routine clinical care, pragmatic study designs hold promise to reduce complexity and burden of trial conduct and participation and expand access in community settings where most patients receive their care. However, not all pragmatic elements will be appropriate for every clinical trial context and design selection depends on the research questions, available data, and intended use of the trial results. While incorporating pragmatic elements may decrease burden, there may be an increase in potential risk and uncertainty regarding consistency and quality of data collected and interpretability of trial results. Uncertainty and sponsor burden may decrease as more experience is gained conducting trials with pragmatic and decentralized elements. In the near-term, consideration should be given to the potential benefits and risks of introducing pragmatic elements, and discussion with regulatory agencies regarding trial design is essential.

Trials conducted in the postmarket setting to answer additional questions are likely to be most amenable to the initial introduction of more pragmatic elements, as the safety and efficacy of the product have been established. The postmarket research questions and case studies provided herein are not exhaustive or representative of all scenarios in which introduction of pragmatic elements may be considered. The case studies described illustrate factors to consider when introducing pragmatism into a clinical trial and will likely apply to additional postmarket scenarios. Additional

statistical aspects should be considered, including sample size and power calculations that may mitigate some of the uncertainty around potential variability in outcomes that may be instilled by more pragmatic approaches.

We focused our discussion on post-marketing settings, but lessons learned from pragmatic approaches to post-marketing trials can inform premarketing trial designs conducted prior to regulatory approval. While limited knowledge of safety and efficacy data in the premarket setting may make certain pragmatic elements inappropriate, opportunities to decentralize trial conduct or expand eligibility criteria can be considered in most contexts and may lead to more rapid accrual and more representative patient populations. The working group also discussed the opportunity to conduct a more pragmatic premarket trial in parallel to an explanatory registrational trial to provide complementary data on a broader patient population. Data from such a parallel pragmatic study could obviate the need to conduct some postmarket studies if acceptable data on underrepresented populations can be generated.

Recent FDA guidance documents, including Conducting Clinical Trials with Decentralized Elements[20] and Integrating RCTs for Drug and Biological Products Into Routine Clinical Practice[18], provide helpful guidance that can be applied to many of the considerations discussed. As these trials move into the community setting, there should be a focus on infrastructure to allow such sites to participate in the trials more feasibly, as there are significant constraints on staffing and resources. Investment in site education and infrastructure are steps toward accomplishing the objective of embedding research into routine care.

As more trials incorporate pragmatic elements, evidence-based insights on which elements have the greatest impact on reducing burden and complexity can lead to prioritizing best practices for introducing pragmatism. Trials with pragmatic and decentralized elements led by the European Organisation for Research and Treatment of Cancer (EORTC)[21,22] and the Alliance and NCTN[4,23] will provide additional lessons learned. Uncertainties and regulatory risks highlighted by sponsors are acknowledged, and continued discussions with sponsors and FDA on acceptability of trial designs is encouraged.

## Citations


1. Ford I, Norrie J. Pragmatic Trials. New England Journal of Medicine. 2016;375(5):454-463. https://pubmed.ncbi.nlm.nih.gov/27518663/
2. Denicoff AM, Ivy SP, Tamashiro TT, et al. Implementing Modernized Eligibility Criteria in US National Cancer Institute Clinical Trials. JNCI: Journal of the National Cancer Institute. 2022;114(11):1437-1440. https://pubmed.ncbi.nlm.nih.gov/36047830/
3. Cannon TL, Rothe M, Mangat PK, et al. Pertuzumab Plus Trastuzumab in Patients With Biliary Tract Cancer With ERBB2/3 Alterations: Results From the Targeted Agent and Profiling Utilization Registry Study. J Clin Oncol. 2024;42(27). https://pubmed.ncbi.nlm.nih.gov/38748939/
4. Reckamp KL, Redman MW, Dragnev KH, et al. SWOG S2302, PRAGMATICA-LUNG: A prospective randomized study of ramucirumab plus pembrolizumab (PR) versus standard of care (SOC) for participants previously treated with immunotherapy for stage IV or recurrent non-small cell lung cancer. Journal of Clinical Oncology. 2024;42(16_suppl):TPS8657-TPS8657. https://ascopubs.org/doi/10.1200/JCO.2024.42.16_suppl.3008
5. Streamlined Trials Embedded in clinical Practice (STEP) Demonstration Project | FDA. Accessed October 29, 2024. https://www.fda.gov/about-fda/cder-center-clinical-trial-innovation-c3ti/streamlined-trials-embedded-clinical-practice-step-demonstration-project
6. Project Pragmatica | FDA. Accessed October 29, 2024. https://www.fda.gov/about-fda/oncology-center-excellence/project-pragmatica
7. Project 5 in 5 | FDA. Accessed October 29, 2024. https://www.fda.gov/about-fda/oncology-center-excellence/project-5-5
8. Friends of Cancer Research. Incorporating Pragmatic Elements in Study Designs to Enhance Oncology Randomized Clinical Trials; 2023. Incorporating_Pragmatic_Trial_Elements_Study_Designs_Oncology.pdf
9. Kaur M, Frahm F, Lu Y, et al. Broadening Eligibility Criteria and Diversity among Patients for Cancer Clinical Trials. NEJM evidence. 2024;3(4). https://pubmed.ncbi.nlm.nih.gov/38771994/
10. Ackerman B, Gan RW, Meyer CS, et al. Measurement error and bias in real-world oncology endpoints when constructing external control arms. Frontiers in Drug Safety and Regulation. 2024;4:1423493. https://www.frontiersin.org/journals/drug-safety-and-regulation/articles/10.3389/fdsfr.2024.1423493/full
11. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB. Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. Pharmacoepidemiol Drug Saf. 2022;31(1):46-54. https://pubmed.ncbi.nlm.nih.gov/34227170/
12. Loudon K, Zwarenstein M, Sullivan F, Donnan P, Treweek S. Making clinical trials more relevant: improving and validating the PRECIS tool for matching trial design decisions to trial purpose. Trials. 2013;14. https://pubmed.ncbi.nlm.nih.gov/23782862/

13. E19 A Selective Approach to Safety Data Collection in Specific Late-Stage Pre-Approval or Post-Approval Clinical Trials | FDA. Accessed October 29, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e19-selective-approach-safety-data-collection-specific-late-stage-pre-approval-or-post-approval
14. NIH. NATIONAL CANCER INSTITUTE CLINICAL TRIALS AND TRANSLATIONAL RESEARCH ADVISORY COMMITTEE (CTAC) STREAMLINING CLINICAL TRIALS WORKING GROUP WORKING GROUP REPORT. Published online 2024. https://deainfo.nci.nih.gov/advisory/ctac/0324/Mandrekar2.pdf
15. PMR/Cs for Novel Oncology Therapies - Friends of Cancer Research. Accessed October 29, 2024. https://friendsofcancerresearch.org/pmr-cs-for-novel-oncology-therapies/
16. Collins G, Andrews HS, McKelvey B, Rice C, Allen JD, Stewart MD. An Evaluation of Novel Oncology Approvals with a PMR/C for Assessing Data in Racial and Ethnic Populations Underrepresented in Premarket Clinical Trials. Clin Cancer Res. 2024;30(16):3388-3394. https://pubmed.ncbi.nlm.nih.gov/38884580/
17. Altomare I, Wang X, Kaur M, et al. Are community oncology practices with or without clinical research programs different? A comparison of patient and practice characteristics. JNCI Cancer Spectr. 2024;8(4):pkae060. https://pubmed.ncbi.nlm.nih.gov/39041606/
18. Integrating Randomized Controlled Trials for Drug and Biological Products Into Routine Clinical Practice | FDA. Accessed October 29, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/integrating-randomized-controlled-trials-drug-and-biological-products-routine-clinical-practice
19. Izano MA, Tran N, Fu A, et al. Implementing Real-World RECIST-based Tumor Response Assessment in Patients With Metastatic Non-small Cell Lung Cancer. Clin Lung Cancer. 2022;23(3):191-194. https://pubmed.ncbi.nlm.nih.gov/35283071/
20. Conducting Clinical Trials With Decentralized Elements | FDA. Accessed October 29, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/conducting-clinical-trials-decentralized-elements
21. Preusser M, Kazda T, Le Rhun E, et al. Lomustine with or without reirradiation for first progression of glioblastoma, LEGATO, EORTC-2227-BTG: study protocol for a randomized phase III study. Trials. 2024;25(1). https://pubmed.ncbi.nlm.nih.gov/38849943/
22. Grisay G, Turco F, Patrikidou A, et al. EORTC GUCG 2238 De-escalate: A pragmatic trial to revisit intermittent androgen-deprivation therapy in metastatic hormone-naïve prostate cancer in the era of new AR pathway inhibitors. Journal of Clinical Oncology. 2024;42(4_suppl):TPS232-TPS232. https://pubmed.ncbi.nlm.nih.gov/38779087/
23. Miller ED, Hitchcock K, Shi Q, et al. Alliance A022101/NRG-GI009: A pragmatic randomized phase III trial evaluating total ablative therapy for patients with limited metastatic colorectal cancer—Evaluating radiation, ablation, and surgery (ERASur). Journal of Clinical Oncology. 2024;42(16_suppl):TPS3633-TPS3633. https://pubmed.ncbi.nlm.nih.gov/38350888/

# Innovative Drug Development

## Advancing Early Endpoints and Novel Evidence Pathways

CYTOTHERAPY
journal homepage: www.isct-cytotherapy.org

FULL-LENGTH ARTICLE
Regulatory Practice

# Accelerating the development of genetically engineered cellular therapies: a framework for extrapolating data across related products

Mark D. Stewart[1,*], Michael Kalos[2], Vicki Coutinho[3], Marc Better[4], Jonathan Jazayeri[5], Jennifer Yohrling[6], Julie Jadlowsky[7], Miriam Fuchs[8], Shalini Gidwani[9], Carsten Goessl[10], Patrick J. Hanley[11], Jane Healy[12], Wen Liu[13], Brittany A. McKelvey[1], Laura Pearce[14], Shari Pilon-Thomas[15], Hillary S. Andrews[1], Monica Veldman[16], Judy Vong[17], Susan P. Weinbach[18], Jeff D. Allen[1]

[1] *Friends of Cancer Research, Washington, DC, USA*
[2] *Next Pillar Consulting, Philadelphia, PA, USA*
[3] *Geni Consulting Ltd., London, UK*
[4] *Pharmefex, Bellevue, WA, USA*
[5] *Kite, A Gilead Company, Collegeville, PA, USA*
[6] *Johnson & Johnson Innovative Medicine, Raritan, NJ, USA*
[7] *Center for Cellular Immunotherapies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
[8] *Novartis, Basel, CH*
[9] *Allogene Therapeutics, San Francisco, CA, USA*
[10] *GSK, Waltham, MA*
[11] *Center for Cancer and Immunology Research, Children's National Hospital, The George Washington University, Washington, DC, USA*
[12] *Merck & Co., Inc., Rahway, NJ, USA*
[13] *Lyell Immunopharma, South San Francisco, CA, USA*
[14] *Canadian Cancer Trials Group, Queens University, Kingston, Ontario, Canada*
[15] *Department of Immunology, Moffitt Cancer Center, Tampa, Florida, USA*
[16] *Genentech Inc., Washington, DC, USA*
[17] *A2Bio, Los Angeles, CA, USA*
[18] *Bristol Myers Squibb, San Diego, CA, USA*

ARTICLE INFO

*Article History:*
Received 23 October 2023
Accepted 11 March 2024

*Key Words:*
CAR-T cell therapy
drug development
regulatory policy

ABSTRACT

*Background:* Significant advancements have been made in the field of cellular therapy as anti-cancer treatments, with the approval of chimeric antigen receptor (CAR)-T cell therapies and the development of other genetically engineered cellular therapies. CAR-T cell therapies have demonstrated remarkable clinical outcomes in various hematological malignancies, establishing their potential to change the current cancer treatment paradigm. Due to the increasing importance of genetically engineered cellular therapies in the oncology treatment landscape, implementing strategies to expedite development and evidence generation for the next generation of cellular therapy products can have a positive impact on patients.
*Methods:* We outline a risk-based methodology and assessment aid for the data extrapolation approach across related genetically engineered cellular therapy products. This systematic data extrapolation approach has applicability beyond CAR-T cells and can influence clinical development strategies for a variety of immune therapies such as T cell receptor (TCR) or genetically engineered and other cell-based therapies (e.g., tumor infiltrating lymphocytes, natural killer cells and macrophages).
*Results:* By analyzing commonalities in manufacturing processes, clinical trial designs, and regulatory considerations, key learnings were identified. These insights support optimization of the development and regulatory approval of novel cellular therapies.
*Conclusions:* The field of cellular therapy holds immense promise in safely and effectively treating cancer. The ability to extrapolate data across related products presents opportunities to streamline the development process and accelerate the delivery of novel therapies to patients.


* Correspondence: Mark Stewart PhD, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC, 20036, USA.
*E-mail address:* mstewart@focr.org (M.D. Stewart).

## Introduction

Genetically engineered cellular therapies have emerged as a new treatment pillar and are poised to change the therapy landscape for patients with serious or life-threatening malignancies. To date, the U. S. Food and Drug Administration (FDA) has approved six autologous chimeric antigen receptor (CAR)-T cell-based immunotherapies, showing remarkable activity in certain hematologic malignancies. However, considerable scientific and operational obstacles must be overcome to enable broader application of this therapeutic modality in additional cancers, including solid tumors, and advance emerging technologies such as allogeneic and *in vivo* engineered cell therapies. Data extrapolation approaches that build on current products may reduce manufacturing costs and the time to develop next generation genetically engineered cellular therapies.

During the development of genetically engineered cellular therapies, sponsors investigating an autologous CAR-T cell product may also test different versions of the primary product (e.g., an altered CAR protein domain to enhance CAR-T cell activity, additional functional enhancements or co-stimulatory domains, a CAR-T cell derived from an alternative starting material, a more purified cell subtype) in parallel or in tandem [1]. As such, leveraging data from related product versions combined with prior platform technology knowledge are reasonably likely to make the drug development, manufacturing process and the regulatory review more efficient across related product versions. This concept is not exclusive to CAR-T cell products and the principles may apply to a variety of immune therapies such as T cell receptor (TCR) or other genetically engineered cell-based therapies (e.g. tumor infiltrating lymphocytes, natural killer cells and macrophages). Accordingly, adaptations of clinical development models and regulatory frameworks are needed to support more flexible development strategies and allow for product improvements based on empirical learnings. The Food and Drug Omnibus Reform Act of 2022 includes a provision for FDA to create a designation program for "platform technologies" that can be used with more than one drug and may be eligible for certain expedited development or review actions [2]. Within the platform technology program, sponsors may "reference or rely upon data and information" from a previous drug/biologics licensing application incorporating the same platform manufacturing technology. Data extrapolation strategies should consider the totality of evidence collected from preclinical research, clinical trials, and characterization of the manufactured product as well as any available published literature or post-marketing surveillance from related products to inform the safety and biological activity of iterative product versions. Ultimately, leveraging the data from the initial product can optimize the development of genetically engineered cellular therapies and may accelerate access to patients.

The FDA continues to refine guidance to increase efficiencies and facilitate development of genetically engineered cellular therapies and released several guidance documents focused on informing development and streamlining regulatory processes for novel cellular and gene therapies [3–5]. Agency expectations around the types of data and necessary comparability studies required to enable process changes (e.g., changing serum-containing media to serum-free media, changing from adherent to suspension cell culture, or adding a new manufacturing site) by sponsors during the lifecycle of a cellular therapy product are becoming clearer [6–8]. However, agency expectations regarding product changes that sponsors may introduce (e.g., refining the cell source, modifying a CAR transgene, adding a second transgene) to enhance product safety and/or efficacy attributes are beginning to be explored. Specifically, FDA outlines an innovative trial design to investigate different versions of a cellular or gene therapy in a single "umbrella" trial using a single trial infrastructure, design, and master protocol during early clinical evaluation, rather than the traditional design of initiating individual trials for each product version. FDA provides several examples of changes that result in different versions, which would require separate investigational new drug applications (INDs) [5]. Within these different versions, one version would be the primary version with the "Primary IND" containing the clinical protocol, the chemistry, manufacturing, and controls (CMC), and pharmacology/toxicology information. Each of the "Secondary INDs" would cross-reference the clinical information in the Primary IND and contain additional CMC and pharmacology/toxicology information specific to each of the secondary versions (Figure 1).

As our experience with genetically engineered cellular therapies continues to improve and FDA's expectations for the types of data necessary to support product changes are clarified, Friends of Cancer Research convened an expert group of stakeholders and hosted a meeting on May 22, 2023 to develop specific strategies for leveraging data from product versions across the stages of development. Extending the concept of cross-referencing information from one product to a related product version could enable informed trial designs and refined data collection to improve operational activities, developmental efficiencies and streamline regulatory data packages. A risk-based data extrapolation approach is proposed to evaluate when, to what extent, and how data from one product can support development of another related product version. A conceptual, risk-based data extrapolation approach is described to leverage the totality of evidence e.g.—available manufacturing, product quality, analytical

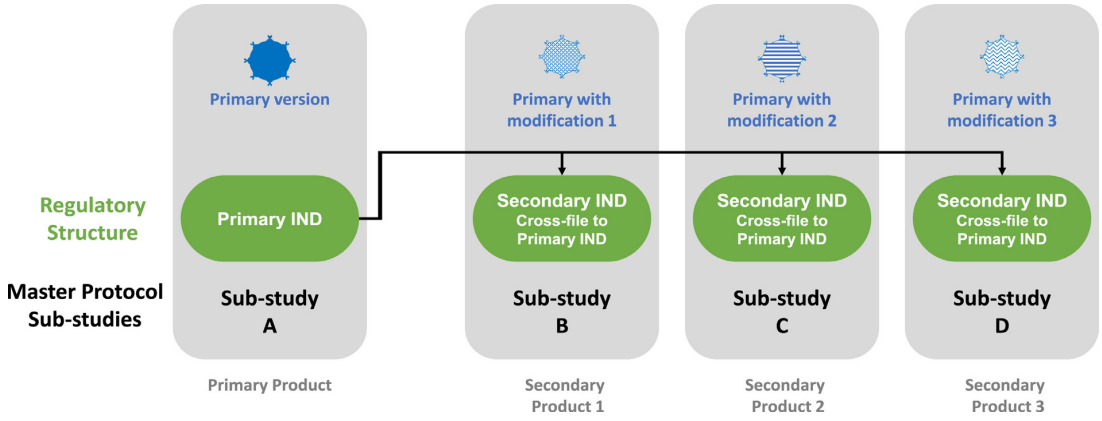



**Fig. 1.** Umbrella trial design for primary and secondary products. The proposed umbrella trial can simultaneously evaluate multiple product versions for a specific disease or condition using a single-trial infrastructure, design, and master protocol, allowing for more efficient product development. (Color version of figure is available online.)

characterization, non-clinical and clinical knowledge, to support development of multiple related product versions. This strategy minimizes redundant data collection, and optimize and accelerate the development of next generation genetically engineered cellular therapies. The data extrapolation concepts discussed draw upon drug development and regulatory processes in the United States, but the principles are congruent in other regions.

## Leveraging Data Across Product Versions to Support Clinical Development

Data extrapolation to advance new versions of investigational products has occurred for several decades across therapeutic classes due to an understanding of the biology, mechanism of action, and manufacturing processes (Supplementary Table S1). Lessons learned from leveraging the totality of evidence in other therapeutic classes to support inferences for new product versions or indications provide a basis for data extrapolation for genetically engineered cellular therapies.

The extent to which data can be meaningfully extrapolated from a primary product to related genetically engineered cellular therapy product(s) depends on the type of modification (including prior knowledge of its impact on related constructs) and phase of development of the primary and secondary products, as well as how "similar" the two versions are to each other. Notably, a case-by-case assessment should be done to determine if a version may be considered the "same" therapeutic [9]. The appropriateness of data extrapolation between two product versions may vary throughout the product lifecycle (e.g., first-in-human studies, early phase, late phase, and postmarket) and across product versions.

Axicabtagene ciloleucel and brexucabtagene autoleucel provide an example of extrapolation in genetically engineered cellular therapy products. The secondary product, brexucabtagene autoleucel, shares the same anti-CD19 CAR construct, vector used in the manufacturing, drug product composition, and similar safety profiles of cytokine release syndrome (CRS) and neurological toxicities as axicabtagene ciloleucel, the primary product. However, brexucabtagene autoleucel has a modified manufacturing process, which includes a white blood cell enrichment process. Nonclinical, clinical, and certain CMC data were extrapolated from axicabtagene ciloleucel to support development and approval of brexucabtagene autoleucel (Table 1). Further, data extrapolation strategies using letetresgene autoleucel (autologous T cell receptor [TCR] T cell therapy targeting NY-ESO-1 and/or LAGE-1a) have been deployed to clinically evaluate next generation versions in a master protocol [10]. The concept of leveraging prior data and the totality of evidence can be extended to other genetically engineered cellular therapy products.

**Table 1**
Use of data extrapolation between axicabtagene ciloleucel and brexucabtagene autoleucel CAR-T cell therapies targeting CD19. Publicly available FDA review documents include examples where data extrapolation has been used in the development and approval of CAR-T cell therapies [11–13].

| Data type extrapolated | Data extrapolation noted in FDA review documents |
|---|---|
| Non-Clinical Data | • Due to several identical features between axicabtagene ciloleucel and brexucabtagene autoleucel, –further safety pharmacology, pharmacokinetic, toxicology, tumorigenicity, and genotoxicity studies were not required for brexucabtagene autoleucel. |
| Clinical Data | • Starting dose in the clinical study to assess the safety and efficacy of brexucabtagene autoleucel in subjects with relapsed/refractory (r/r) mantle cell lymphoma (MCL) was selected on the prior dose of axicabtagene ciloleucel in subjects with r/r MCL in the same clinical study. The typical dose escalation cohorts, inter-patient intervals and stopping rules were minimized.<br>• Due to several identical features existing across the two product versions and similar safety profiles of cytokine release syndrome (CRS) and neurological toxicities, the FDA supported a combined risk evaluation and mitigation strategies (REMS) program for axicabtagene ciloleucel and brexucabtagene autoleucel. |
| CMC Data | • Due to several similarities in the manufacture (vector construct, vector manufacturing process, product manufacturing process, controls, formulation, container closure system validation, storage, equipment, and same manufacturing sites), several sections of CMC data were not generated for brexucabtagene autoleucel, but information resubmitted in the brexucabtagene autoleucel biologics license application (BLA).<br>• Certain facility inspections were waived due to axicabtagene ciloleucel and brexucabtagene autoleucel sharing the same licensed manufacturing site. |

## Developing a Risk-Based Approach to Support Data Extrapolation Between Product Versions

Extrapolating data across genetically engineered cellular therapy product versions necessitates a fundamental understanding of the primary product and its functional and biophysical properties (Table 2), which in turn requires sufficient non-clinical, CMC, and clinical data, and adequate scientific justification for extrapolation. A framework for evaluating risk in pharmaceutical development is well established in the International Council for Harmonization (ICH) Q9 (R1) and Q8(R2) guidelines on Quality Risk Management and Product Development [14,15]. Extensive knowledge of critical process parameters, product quality attributes, and well-established, robust analytical methods are essential to allow for data comparability and justify extrapolation to support development of subsequent product versions [16–18].

To support this, qualified and fit-for-purpose analytical methods that characterize quality attributes are necessary for a variety of critical parameters (e.g., safety, purity, potency, and identity) to define risk categories. Based on the magnitude of difference in assay outputs relative to the original product version and other data governing the modification that may exist, a risk assessment can demonstrate the probability and severity of risk to patients due to a product modification. Of note, especially for products with highly variable incoming starting material, variability between final products can be expected, especially early in development, making extrapolations potentially more challenging. Furthermore, the sensitivity and degree of qualification of the assays utilized for in-process controls and final product release must be considered. Consequently, evaluating the totality of the manufacturing, characterization, and release data as well as clinical data are critical when extrapolating between product versions.

The type and amount of required additional data for extrapolation will vary and depend on whether a change has a minor or major impact on product quality, efficacy, or safety. A modification that results in a low-risk impact may allow for data extrapolation across products with targeted data collection to address data gaps and support regulatory requirements, whereas a modification that results in a high-risk impact may require more extensive studies. For example, a low-risk impact that has a minor bearing only on product quality may require an analytical comparability assessment, while a moderate-risk impact that involves patient safety/efficacy may require a clinical bridging study, and a high-risk impact may require a larger clinical trial to confirm safety and efficacy in accordance with the degree of expected similarities. The patient population and magnitude of unmet need should also be considered and may lead to a shift in risk tolerance for a particular development program. An assessment aid-like tool (Table 3) could support a systematic approach for

**Table 2**
Proposed best practices in process and product development to support data extrapolation.

| |
|---|
| **1. Generate comprehensive product knowledge**<br>Gather appropriate non-clinical, clinical, and CMC knowledge based on the stage of drug development. |
| **2. Evaluate the relationship between product attributes**<br>While initial assessments can be performed based on non-clinical and clinical data, as the product advances through clinical development, more robust information on the product efficacy and safety profile will enable a more meaningful determination of how a potential change can impact critical quality attributes (CQAs) or product safety and efficacy. A stepwise approach is necessary to:<br>1) Assess the relationship between manufacturing process parameters and CQAs (e.g., identity, purity, potency, and safety).<br>2) Assess the impact of each CQA on product safety and efficacy (i.e., clinical activity). |
| **3. Develop parameters to define risk and perform risk assessment of secondary products**<br>Based on the defined relationships between any changes in quality attributes and safety and efficacy profiles between the primary and secondary product, define:<br>1) The relative risk of a change on product safety and efficacy<br>2) Appropriate action(s) to be taken based on the assigned risk. |
| **4. Develop data packages based on identified risk and actions to mitigate risk in regulatory submissions**<br>Determine the appropriate actions based on the totality of evidence from the primary and secondary products and assigned level of risk of the change(s) on safety and efficacy of the secondary product. Such actions could include:<br>• Extrapolation of data from the primary product<br>• Generation of additional or new data<br>• Develop clinical risk mitigation strategies to facilitate clinical development.<br>There should be frequent and early discussions with FDA particularly when there are uncertainties regarding regulatory and clinical pathways (i.e., will the data extrapolation package be acceptable, will safety run in data or additional data necessary to support the secondary products etc.). |

determining the appropriateness of data extrapolation within clinical development programs of secondary products and serve as a summary for FDA submissions.

Classifying the risk impact of modifications may not be easily determined at the outset of development of the related product. The extent to which prior data can be extrapolated will depend on several factors, including the intended development plan of the new product version and risk determination for the impact of the changes on safety and efficacy. In a risk evaluation, it is important to assess the robustness and types of existing data available from the primary product such as information from analytical and *in vitro* studies, non-clinical *in vivo* studies, clinical pharmacokinetic/dynamic (PK/PD) studies (i.e., biomarker correlates, product correlates of response), and clinical efficacy and safety studies (Supplementary Table S2). The analytical methods deployed will vary based on the type of genetically engineered cellular therapy product (e.g., autologous, allogeneic, CAR, TCR, etc.) as well as the types and extent of modifications introduced. Methods to analyze risk should be defined early in development and an adequate level of sensitivity to identify expected differences between two product versions and support a risk-based extrapolation plan.

## Leveraging the Totality of Evidence to Support Product Development at Specific Stages of Clinical Development

As products progress through development, the amount of data available to determine risk and extrapolate across versions increases (e.g., extrapolating data from a primary product in early phase, a primary product in late phase, or an already approved product). Table 4 provides examples of how, when justified, data extrapolation can streamline evidence generation, assist in a more seamless transition from one phase of development to another (i.e., academic to industry, early- to mid-phase, and late-phase to post-market), minimize repetitive data collection, and potentially shorten clinical development timelines. The transition from early to later phase clinical development often aligns with a transition from the academic to biopharmaceutical setting and a pivotal step where the product manufacturing process might be modified to support commercialization [19]. Assessment of the impact for such process modifications is captured under more mature FDA guidance; [6–8] however, it is possible that modifications may impact product attributes and thus be informed by the herein proposals. Some example scenarios that might support an accelerated transition of a secondary product through various stages of clinical development are presented below.

### *Early clinical development*

Early phase safety and efficacy data from the primary product could support an understanding of the preliminary safety and efficacy profile, to establish the dosing and schedule, and an approach to data collection in later-phase studies for the secondary product. For example, if appropriately justified, sponsors could propose a similar starting dose for a secondary product as the recommended phase 2 dose for the primary product and/or use the primary product profile to inform more targeted dose limiting toxicity (DLT) criteria to advance a secondary product through early phase studies more efficiently. In early and late phase trials, prior product knowledge could help prepare for expected toxicities and/or inform monitoring strategies to reduce or mitigate symptomatic adverse events.

### *Late phase clinical development*

In instances where a primary product is in late phase development or approved, the totality of data from the primary product may allow a secondary version to move straight into a Phase 2/3 clinical trial. Additionally, data extrapolation may be appropriate and generation of a reduced clinical dataset for the secondary product may be justified based on the similarities with the primary product. For instance, a Phase 3 randomized controlled trial (RCT) readout of the primary product paired with a single-arm clinical bridging study of the secondary product in the same indication may be used to support registration of the secondary product, which could dramatically accelerate patient access to improved product variations.

### *Post-market phase*

Prior product knowledge and the totality of evidence could aid in identification of potential longer-term treatment effects, inform safety surveillance activities, and support patient management in clinical practice for a secondary product. Additionally, post-market data from a related product may justify a shorter duration of patient safety follow-up and reduce the 15-year long-term follow-up period for a secondary product in development or postmarket to decrease costs, resources, and patient burden [20].

## Mechanisms for Exploring Data Extrapolation Opportunities and Engaging with FDA

Considerable progress is being made in the development and use of genetically engineered cellular therapies and the field is still evolving. The conceptual framework herein outlined, intends to accelerate investigation and development of the next generation of genetically engineered cellular therapy products and may act as a guide when expanding to other indications and patient populations. As data extrapolation across product versions becomes more common in development programs for genetically engineered cellular therapies,

**Table 3**
Data extrapolation assessment aid prototype. This document could be submitted as part of an initial IND and/or subsequent IND amendments for a secondary product or as justification to support amendments to a protocol based on learnings from a related product version for FDA meetings. Part A and Part B describe supportive information and data to justify and evaluate data extrapolation in the clinical development of secondary products.

| Supportive data | Key information | Guidance for providing information |
|---|---|---|
| | **Part A- Background/Overview** | |
| Overview of the Primary Product | • What is the stage of development of the primary product?<br>• Summary of product characteristics (e.g., type of genetically engineered cellular therapy, mechanism of action, target, CMC overview)<br>• Summary of data related to safety, efficacy and pharmacologic properties (e.g., safety summary, efficacy summary, dosing, dose/response relationships, any correlations or association between CQAs and clinical data, PK characteristics, clinical studies) | Articulate key non-clinical, CMC, preclinical and clinical safety, and efficacy data set. |
| Overview of the Secondary Product | • What is the stage of development of the secondary product?<br>• Summary of shared characteristics and differences between product versions<br>• Summary of data from secondary product [*if applicable*]<br>• Summary of known information gaps | Articulate similarities and differences between product versions with a focus on patient safety and pharmacologic properties. |
| Summary of Development Plan for Primary and Secondary Product | • Summary of development strategy (i.e., will both products be developed in parallel, or will the secondary product replace the primary product?)<br>• Timeline of development strategy | Describe development strategy for product versions.<br>Outline anticipated timelines for data readouts and how this informs development decisions for the secondary product. |
| | **Part B- Extrapolation strategy** | |
| Data Extrapolation Details | • What data are being extrapolated?<br>• How will the extrapolated data from the primary product be used in the development of the secondary product? | Information collected in this section could be presented in a tabulated format:<br>• Data being extrapolated |
| Justification for Data Extrapolation | • What is the rationale and justification for data extrapolation (i.e., risk assessment)? | • Sponsor assessment of associated risk |
| Risk Mitigation | • How will known information gaps and risks be mitigated? | • Mitigation strategy |

optimal methods to analyze, interpret, and present data in a rigorous and standardized manner will be critical. As product and process knowledge increases within individual development programs and within the field, adaptive regulatory processes that adjust based on the potential risks associated with the modification or stage of development should be in place and support data extrapolation in development of iterative product versions.

Sponsors should consider engaging the FDA early in the clinical development lifecycle when they are interested in justifying the use of prior product knowledge and data extrapolation to inform a specific program and establish pre-specified parameters for risk tolerance. Sponsors should have adequate product quality data or published data to demonstrate that distinct product versions are "similar" in a manner that mitigates concerns about product safety and efficacy when engaging with the FDA and can use the data extrapolation assessment aid prototype (Table 3). Since much of the data to support these assessments will not be publicly available, these assessments will be considered individually by each sponsor. However, public information available could be leveraged by sponsors as has been observed with industry coalescing around published data supporting starting doses for CAR-T cell therapies.

If the relationship between product attributes and patient safety and/or efficacy is not yet fully established (e.g., if the development of both primary and secondary products are in early stages), it is important to identify the uncertainties and knowledge gaps and have a plan for continued assessment of the relationship (e.g., setting milestones after a predetermined number of patients are treated or at the end-of-phase 1 or end-of-phase 2 studies). Pre-defined opportunities for meetings between sponsors and the FDA can be used to address issues relating to product development and to propose mechanisms for data extrapolation to align the core components of such a data package. FDA guidance is available that describes the various FDA meetings, meeting formats, how to submit a request, meeting package requirements, and the different timings for such meetings [22,23]. Ultimately, meetings can help ensure aspects of manufacturing, data capture, and trial designs are sufficient to support a data package for new INDs and BLAs for the next generation versions. Several regulatory opportunities exist that may be particularly advantageous to present the data extrapolation plan and propose the study design for clinical development:

- **Type B Meetings:** Pre-IND, end-of-phase 1, end-of-phase 2, pre-phase 3 meetings, or pre-biologics license application (BLA) can introduce the data extrapolation plan, available data and risk assessment, and how data extrapolation will support the development of a secondary product.
- **Type D Meetings:** Meeting to discuss a narrow set of issues (i.e., not more than 2 focused topics) and should not require input from more than 3 disciplines or Divisions, which may also consider discussion on data extrapolation. Type D meetings may also be available without having an IND.
- **Regenerative Medicine Advanced Therapy (RMAT)/Breakthrough Therapy Designation (BTD) products:** Products that receive these designations signal an organizational commitment by the FDA that involves senior managers. Additionally, products that leverage expedited development programs have shorter clinical development timelines [24]. Designated products are eligible for further FDA meetings that can include data extrapolation for new product version(s).
- **CMC Development and Readiness Pilot (CDRP):** Under the pilot, FDA will provide product-specific CMC advice during product development for products with RMAT/BTD designation, including two additional CMC-focused Type B meetings, as well as a limited number of additional CMC-focused discussions. The pilot will enable additional interactions with FDA during product development and, if applicable, warrant the use of science- and risk-based regulatory frameworks allowing streamlining of CMC development activities to provide earlier clinical access to patients.
- **Designation Program for Platform Technologies:** This is a designation program for platform technologies that have the potential to increase efficiencies in drug development. Applications for drugs or biologics that use or incorporate platform technologies may be eligible for certain expedited development or review actions. The intent of this designation program is to bring significant efficiencies to the drug development or manufacturing

**Table 4**
Potential opportunities for data extrapolation from a primary product.

| Data | Opportunities |
|---|---|
| CMC | • Extrapolate viral vector/gene editing tools/cell engineering product information, and product/process characterization data<br>• Extrapolate drug product presentation information including container and closure systems, fill volumes and cell concentration<br>• Use stability data from primary product to support initial stability for secondary product<br>• Reduced stability programs leveraging prior programs<br>• Include representative engineering batches in the initial IND of a secondary product and commit to provide certificate of analysis from good manufacturing practice (GMP) batch prior to initiating patient dosing<br>• Reuse gene editing safety data (i.e., translocation information, on and off target editing data) if same edits are used with different CAR<br>• Risk-based microbiology control strategy based on primary product to minimize redundant safety testing requirements<br>• Same analytical methods including potency assays<br>• Orthogonal assays to support similar characteristics of potency<br>• Extrapolate residual control strategy as applicable, and apply to new product<br>• Leverage specifications of primary product |
| Pre-clinical | • Same relevant animal model and, if not available, justify not conducting toxicity studies<br>• Potential to reduce/waive *in vivo* studies and use *in vitro* studies for proof of concept by referencing primary product data<br>• Use comparative potency data to support *in vivo* study design for secondary product (i.e., dose) |
| Clinical safety | • Inform starting dose using primary product data<br>• Extrapolate safety data from primary product to optimize, reduce testing (i.e., replication competent lentivirus [RCL]/ replication competent retrovirus [RCR]), and timepoints for long-term safety<br>• Extrapolate potency data to determine potential support for or differentiation of the safety profile for the secondary product<br>• Extrapolate safety data from the primary product<br>• Modified or combined REMS programs for products and use operational efficiencies as proposed by the American Society for Transplantation and Cellular Therapy (ASTCT) 80/20 Task Force [21] |
| Clinical efficacy | • Support the starting dose and minimize the number of dose levels needed to be tested in early clinical studies, where appropriate<br>• Extrapolate certain clinical data from one indication to support other indications with the secondary product<br>• Potential for fewer clinical trial patients to be treated subject to clinical comparability<br>• Potential short follow up time for the patients treated with the new version, as appropriate<br>• Extrapolate biomarkers/assays for measuring clinical efficacy based on product similarity or support clinical cutoff for patient selection |

process as well as to the review process for products across the platform. Many of the concepts and areas for data extrapolation outlined above may be within scope of cell therapy platforms and thus leveraged in subsequent platform products.

In addition to the meeting types and mechanisms noted above, the Initial Targeted Engagement for Regulatory Advice on CBER/CDER Products (INTERACT) and CBER Advanced Technology Team (CATT) may be appropriate to discuss data extrapolation plans or use of new technology/methods to enable data extrapolation.

## Moving Forward

Given the uniqueness of genetically engineered cellular therapies, opportunities for continued dialogue beyond the post-approval setting with the FDA, including the Office of Therapeutic Products (OTP), will be important to encourage continued innovation. Additional data and evidence generation, as well as learnings from leveraging safety data across different versions of products, should inform risk-based approaches to defining the optimal safety follow-up period as the field of genetically engineered cellular therapies continues to grow and evolve. FDA workshops could help inform updated guidance on, for example, generating long-term follow-up data for genetically engineered cellular therapy products and clarifying opportunities to streamline data or compress development timelines based on known or expected safety events. Additionally, workshops and other mechanisms should be explored to capture and disseminate best practices and case studies of data extrapolation in clinical development as well as learning from pilot projects like CDRP, which will help educate sponsors in exploring adequate development pathways. A question-and-answer resource could provide timely answers to questions that are commonly asked and applicable across development programs. The concepts and proposals put forward hold promise in streamlining data requirements, while still adequately and robustly assessing products, and ultimately accelerate timelines for patients to access these transformative therapies.

As the field progresses, developers are investigating genetically engineered cellular therapies to not only expand into new disease areas (e.g., CD19-CAR-T cell therapy trials in autoimmune diseases, gene-modified stem cells for genetic disorders) and lines of therapy, but also to improve upon available genetically engineered cellular therapies. For innovation to reach patients in a meaningful timeframe, leveraging available data and extrapolation from s related product version is one mechanism to accelerate development. Additional strategies for accelerating the development of the next generation of genetically engineered cellular therapy products should be explored. In addition to data extrapolation, trial design considerations, alternative and adaptive study designs, real-world data sources, novel endpoints, and use of bioinformatics may accelerate development and require thoughtful discussion among key stakeholders, including regulators, investigators, patient advocacy groups and sponsors.

## Declaration of competing interest

M.K. holds IP related to cell therapy, assigned to the University of Pennsylvania and licensed to Novartis, board memberships with IMVinc, Nanocell therapeutics, and is on scientific advisory boards for AdicetBio, Annoca AG, cTRL-therapeutics, Cue Biopharma, Lykan Bioscience, Senti Biosciences, Vittoria Therapeutics. J.J is an employee of Kite, A Gilead Company. J.Y. is an employee of Janssen R&D, LLC. M.F is an employee of Novartis. S.G. is an employee of Allogene Therapeutics. C.G. is an employee of GSK. P.J.H. is a Scientific Advisor or Advisory Board for Cellenkos, Cellevolve, Discovery Life Sciences, Microfluidx, Autolomous, Capsida and is co-founder, board of directors of Mana Therapeutics. J.H. is an employee of Merck & Co. Inc. W. L. is an employee of Lyell Immunopharma. L.P. is an employee of the Canadian Cancer Trials Group and past employee of GSK, and holds shares in GSK. S.P.T reports that Moffitt Cancer Center has licensed Intellectual Property (IP) related to the proliferation and expansion of tumor infiltrating lymphocytes (TILs) to Iovance Biotherapeutics. Moffitt has also licensed IP to Tuhura Biopharma. S.P.T. is an inventor on such Intellectual Property. SPT has received consulting fees from Seagen Inc., Morphogenesis Inc., and KSQ Therapeutics. M.V. is an employee of Genentech, Inc. J.V. is an employee of A2 Biotherapeutics, Inc. S.W. is an employee of Bristol Myers Squibb. All other authors report no potential conflicts.

## Acknowledgments

The authors thank Grace Collins (Friends of Cancer Research) for assistance with data collection and analysis. The authors also thank

representatives of the FDA for their review of the manuscript. This manuscript was developed based on working group discussions and public discussions at the meeting titled "The Next Generation of Cellular Therapies: A Blueprint to Accelerate Development" as well as a roundtable discussion in Washington, DC, supported by Friends of Cancer Research to have public discourse of a preliminary white paper [25]. The public meeting agenda is provided in Supplementary Table 3. The authors would like to thank the meeting participants and the panelists who facilitated the discussions and whose views contributed to the information provided in this manuscript.


### Funding

This work was funded by Friends of Cancer Research.


### Author Contributions

Conception and design of the study: MDS, BAM, HSA, JDA, MK, VC, MB, JJ. Analysis and interpretation of data: MDS, MK, VC, MB, JJ, JY, JJ, MF, SG, CG, PH, JH, WL, BAM, LP, SPT, HSA, MV, JV, SPW, JDA. Drafting or revising the manuscript: MDS, MK, VC, MB, JJ, JY, JJ, MF, SG, CG, PH, JH, WL, BAM, LP, SPT, HSA, MV, JV, SPW, JDA. All authors have approved the final article.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jcyt.2024.03.009.

### References


[1] Tomasik J, Jasiński M, Basak GW. Next generations of CAR-T cells - new therapeutic opportunities in hematology? Front Immunol 2022;13:1–13. https://doi.org/10.3389/fimmu.2022.1034707.

[2] H.R.2617 - 117th Congress (2021-2022): Consolidated Appropriations Act, 2023. (2022). Retrieved from https://www.congress.gov/bill/117th-congress/house-bill/2617. Accessed May 14, 2023.

[3] FDA. Considerations for the development of chimeric antigen receptor (CAR) T cell products; draft guidance for industry. Retrieved from http://www.regulations.gov. Accessed May 14, 2023.

[4] FDA. Human gene therapy products incorporating human genome editing; draft guidance for industry. Retrieved from http://www.regulations.gov. Accessed May 14, 2023.

[5] FDA. Studying multiple versions of a cellular or gene therapy product in an early-phase clinical trial; guidance for industry. Retrieved from https://www.fda.gov/media/106369/download. (2022). Accessed May 14, 2023.

[6] FDA. Q9(R1) quality risk management guidance for industry ICH-quality. Retrieved from https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidance. Accessed May 14, 2023.

[7] FDA. Chemistry, manufacturing, and controls changes to an approved application: certain biological products; guidance for industry. Retrieved from https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information. (2021). Accessed May 14, 2023.

[8] FDA. Manufacturing changes and comparability for human cellular and gene therapy products; draft guidance for industry. Retrieved from https://www.fda.gov/media/170198/download. Accessed July 28, 2023.

[9] FDA. Interpreting sameness of gene therapy products under the orphan drug regulations; guidance for industry. https://www.fda.gov/media/113760/download. (2021). Accessed on May 14, 2023.

[10] Schoenfeld AJ, Altan M, Owonikoko TK, D'Angelo S, Ladle BH, Noujaim J, et al. Abstract CT219: master protocol to assess the safety and recommended Phase 2 dose of next generation NY-ESO-1-specific TCR T-cells in HLA-A*02 patients with synovial sarcoma and non-small cell lung cancer. Cancer Res 2021;81:CT219.

[11] YESCARTA (axicabtagene ciloleucel). https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/yescarta-axicabtagene-ciloleucel (2022). Accessed on May 14, 2023.

[12] TECARTUS (brexucabtagene autoleucel). https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/tecartus-brexucabtagene-autoleucel (2022). Accessed on May 14, 2023.

[13] FDA. Approved Risk Evaluation and Mitigation Strategies (REMS). https://www.accessdata.fda.gov/scripts/cder/rems/index.cfm?event=IndvRemsDetails.page&REMS=375 (2022). Accessed on Moay 14, 2023.

[14] International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use Pharmaceutical Development Q8 (R2). (2009).https://www.ema.europa.eu/en/ich-q8-r2-pharmaceutical-development-scientific-guideline. Accessed on May 14, 2023.

[15] International Council For Harmonisation Of Technical Requirements For Pharmaceuticals For Human Use Ich Harmonised Guideline Quality Risk Management Q9 (R1). https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_-Step4_2022_1219.pdf. Accessed on May 13, 2023.

[16] Kalos M. An integrative paradigm to impart quality to correlative science. J Transl Med 2010;8:26. https://doi.org/10.1186/1479-5876-8-26.

[17] Sarikonda G, Mathieu M, Natalia M, Pahuja A, Xue Q, Pierog PL, et al. Best practices for the development, analytical validation and clinical implementation of flow cytometric methods for chimeric antigen receptor T cell analyses. Cytometry B Clin Cytom 2021;100:79–91.

[18] FDA. Comparability Protocols for human drugs and biologics: chemistry, manufacturing, and controls information guidance for industry. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm. Accessed on May 14, 2023.

[19] Lu TL, Pugach O, Somerville R, Rosenberg SA, Kochenderfer JN, Better M, et al. A rapid cell expansion process for production of engineered autologous CAR-T cell therapies. Hum Gene Ther Methods 2016;27:209–18.

[20] FDA. Long term follow-up after administration of human gene therapy products; guidance for industry 2020. https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-. Accessed on May 14, 2023.

[21] Nikiforow S, et al. Paving the road for chimeric antigen receptor T cells: American Society for Transplantation and Cellular Therapy 80/20 task force consensus on challenges and solutions to improving efficiency of clinical center certification and maintenance of operations for commercially approved immune effector cell therapies. Transplant Cell Ther 2023;29:228–39.

[22] FDA. Formal meetings between the FDA and sponsors or applicants of PDUFA products guidance for industry DRAFT GUIDANCE. https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-. Accessed on May 13, 2023.

[23] FDA. Interactions with office of therapeutic products. https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/interactions-office-therapeutic-products (2023). Accessed on July 17, 2023.

[24] Collins G, Stewart M, Sigal E, Allen J. Expedited Development programs at the food and drug administration: insights and opportunities. Ther Innov Regulat Sci 2021;55:619–21. https://doi.org/10.1007/s43441-021-00258-z.

[25] Friends of Cancer Research. The next generation of cellular therapies: a blueprint to accelerate development. https://friendsofcancerresearch.org/event/the-next-generation-of-cellular-therapies-a-blueprint-to-accelerate-development/ (2023). Accessed on August 12, 2023.

FRIENDS of CANCER RESEARCH

# Enhancing Study Designs and Interpretation of Interim Overall Survival Data in Oncology Trials

FRIENDS OF CANCER RESEARCH WHITE PAPER | 2024

## Executive Summary

In oncology drug development, early endpoints such as progression-free survival (PFS) and objective response rate (ORR) are commonly used to support expedited development of therapies by facilitating earlier efficacy readouts and regulatory review. This can help provide timely access to potentially life-saving treatments. Challenges arise when there are limited overall survival (OS) data available at the time of this early assessment, leaving uncertainty about the true benefit-risk profile of a drug. In these settings, interim OS data may be evaluated as a safety endpoint to assess potential harm. However, the interpretation of interim OS data can be challenging due to small event numbers, limited duration of follow up, and trial dynamics such as patient crossover.

A collaborative, multidisciplinary working group outlined key considerations for improving the analysis and interpretation of interim OS data in oncology clinical trials. These include a proposal for a multi-step approach to be incorporated into trial designs to guide both qualitative and quantitative evaluations, ensuring a more complete understanding of the data. Taken together, an improved design and more structured interpretation of interim OS data will lead to better informed decision-making during drug development.

### Key Insights and Recommendations

- Early assessments of OS can provide unreliable results due to small event counts, limited follow-up, and overall immature data. We examined case studies, which highlight how patient subgroups, treatment crossover, and other trial design factors complicate interim OS data interpretation.
- A carefully considered study design can enhance the reliability of interim OS data interpretations. This includes pre-specifying criteria for patient crossover, planning for sufficient follow-up duration, and simulating potential scenarios to inform analysis timing and threshold setting. Design elements, paired with a structured analysis approach, can ensure more accurate and timely decision-making during drug development.
- A structured, multi-step approach to interpreting interim OS data is proposed:
    - Perform a qualitative descriptive analysis, including a review of event counts, patient comorbidities, the timing of adverse events, rates of dose interruptions or reductions, and subsequent therapies. This provides essential clinical context for early signals of harm or efficacy.
    - Apply a streamlined/comprehensive quantitative framework that balances the risk of mistakenly concluding that a treatment is harmful (false positive) or missing a true safety issue (false negative) when interpreting interim OS data. This includes calculating hazard ratios (HRs) and their confidence intervals, setting thresholds for identifying potential harm, and using predictive models taking into account the data maturity to assess whether the final OS outcome is likely to show benefit, harm, or no difference.

These insights emphasize the importance of integrating careful design and comprehensive analysis of interim OS data to help ensure that oncology trials can better balance early efficacy signals with expected long-term survival outcomes.

## Authors

**Arunava Chakravartty**, Novartis

**Ray Lin**, Genentech

**Andy Chi**, Takeda

**Andrew Lithio**, Eli Lilly and Company

**Tai-Tsang X. Chen**, GSK

**Qiang Liu**, Amgen

**Ruthanna Davi**, Medidata Solutions

**Briggs Morrison**, Crossbow Therapeutics

**Maura Dickler**, Genentech

**Pabak Mukhopadhyay**, AstraZeneca

**Alex Fleishman**, Amgen

**Abderrahim Oukessou**, Bristol Myers Squibb

**Sebastiano Gattoni-Celli**, Daiichi Sankyo, Inc.

**Antony Sabin**, GSK

**Thomas Gwise**, Medidata Solutions

**Sabeen Shah**, Johnson and Johnson Innovative Medicine

**Lisa Hampson**, Novartis

**Michael Sweeting**, AstraZeneca

**Philip He**, Daiichi Sankyo, Inc.

**Yevgen Tymofiyev**, Johnson and Johnson Innovative Medicine

**Laura Huggins**, EMD Serono

**Chris White**, Patient Advocate

**Qi Jiang**, Pfizer, Inc.

**Nevine Zariffa**, NMD Group

**Michael Leblanc**, Fred Hutchinson Cancer Center

**Helen Zhou**, GSK

**Jianchang Lin**, Takeda

## Background

In oncology drug development, early endpoints such as progression-free survival (PFS) and objective response rate (ORR) are commonly used to provide early indications of a drug's efficacy. These endpoints help to expedite drug development, addressing unmet medical needs by enabling timely regulatory approvals and patient access to potentially beneficial therapies through Accelerated Approval, or traditional approval in certain circumstances. While overall survival (OS), the gold standard for assessing clinical benefit of cancer therapies, is traditionally evaluated as an efficacy endpoint at the end of a trial interim looks at OS data at the time of ORR or PFS assessment can serve as a safety endpoint, providing additional context for assessing the benefit-risk profile of new cancer drugs. However, interim OS data can be challenging to interpret due to the immaturity of the data at the time of an early endpoint readout.

In some cases, a statistically significant effect observed in PFS or ORR efficacy results may be overshadowed by a potential risk of harm based on interim OS data (e.g., an observed hazard ratio above 1.0). This scenario poses a conundrum due to potential conflicting data on the benefit-risk assessment of a drug. At early looks, like any endpoint, statistical estimates of endpoint readouts can be highly variable due to small sample size, limited follow up, low information fraction, and potentially delayed treatment effects. As the data mature, these fluctuations can stabilize to provide a clearer picture of the true treatment effect.[1] If interim OS analyses lead to the erroneous conclusion that a drug is harming patients, its approval may be unduly delayed, depriving patients of potential benefits based on a false conclusion of harm. However, if interim OS data correctly identify a potential safety issue early, a potentially harmful drug is kept off the market, thus protecting patients from adverse outcomes. Differentiating between these two scenarios requires careful planning and robust data, raising important questions about how best to minimize the risk of drawing false conclusions from interim OS data. To navigate these challenges, a robust, standardized, and context-specific framework can help guide the analysis and interpretation of interim OS data, ensuring reliable evaluation and good regulatory decision-making.

Friends of Cancer Research established a multidisciplinary working group to address these challenges and to develop best practices for assessing interim OS data in oncology trials. This white paper outlines key design and analysis considerations when interim OS data are evaluated in a trial and proposes a strategy for simulation studies that could provide data driven insights, with a goal of improving the understanding and application of interim OS data.

## Learnings from Recent Clinical Trials on Interim OS Data

Recent clinical trials provide insights into the challenge of interpreting interim OS data, particularly in relation to early endpoints like PFS and ORR. These trials demonstrate how factors such as treatment crossover, the immaturity of interim OS data, and patient subgroups can affect the interpretation of OS results. By examining these factors, we can gain insights into how to optimize the design and interpretation of endpoints in future trials, particularly in terms of balancing early efficacy signals with long-term survival outcomes.

The following case studies provide further context, exploring how these insights apply to individual trials and offering lessons for future study designs (**Appendix 1** summarizes study details and outcome data at interim analyses, when available):

## 1. PSMAFore Trial (177Lu-PSMA-617 in Metastatic Castration-Resistant Prostate Cancer)

The PSMAFore trial explored the use of radioligand therapy in 468 patients with metastatic castration-resistant prostate cancer (mCRPC).[2] Patients who progressed on standard therapies in the control arm were allowed to crossover to the experimental arm. The radiographic PFS (rPFS) results were highly favorable at the primary analysis (cutoff: ~7 months median time from randomization until cutoff; Hazard Ratio; HR = 0.41, 95% Confidence Interval; CI: 0.29–0.56), showing a strong treatment effect. The interpretation of interim OS analyses was complicated by high rates of patient crossover, making it difficult to accurately assess the long-term survival benefits of the therapy. By the time of the second interim OS analysis, over half of all patients randomized to the control arm (123 of 234 patients) had crossed over to the experimental arm, and the unadjusted OS HR at this analysis (targeting a treatment policy estimand, not adjusting for crossover) was 1.16 (95% CI: 0.83-1.64). At the time of the third interim OS analysis (cutoff: ~24 months median time from randomization until cutoff), 134 of 234 patients randomized to control had crossed over. This analysis showed an unadjusted OS HR of 0.98 (95% CI: 0.75–1.28). The final OS analysis is pending.

**Key Insight:** High crossover rates complicate OS interpretation and it is often necessary to evaluate the OS data under a variety of sensitivity and supplementary analyses to investigate the robustness of the results. These can include statistical estimation approaches, such as those based on the rank preserving structural failure time (RPSFT) model used in this study, though interpretation of these analyses can still be challenging. This trial highlights a common issue across oncology studies, where patient crossover allowed under the protocol design can make it harder to see the true effect of the new treatment relative to the standard treatment than if patients had not been permitted to crossover.

## 2. monarchE Trial (Abemaciclib + ET in Early HR+/HER2- Breast Cancer)

The monarchE trial examined adjuvant abemaciclib, a CDK4/6 inhibitor, in combination with endocrine therapy (ET) for patients with high-risk of recurrence early-stage HR+/HER2- breast cancer.[3] The trial included a large population of 5,637 patients. Though the trial previously demonstrated a significant benefit in invasive disease-free survival (IDFS), the immaturity of the OS data was still evident at the first interim analysis of OS (cutoff: 36 months from study start). While the IDFS readout was statistically significant (HR = 0.696, 95% CI: 0.588–0.823), the initial interim OS analysis showed an OS HR of 1.091 (95% CI: 0.818–1.455), favoring the control arm. The initial FDA approval was limited to patients at high risk of recurrence and high Ki-67 expression.[4] This was based on careful consideration of prespecified subgroups and additional analyses including a gated hierarchical testing strategy that included the additional endpoint of IDFS in patients with a KI-67 score ≥20% which also demonstrated a statistically significant IDFS (HR=0.626, 95% CI: 0.488, 0.803) and an interim OS analysis showed an OS HR of 0.767 (95%CI: 0.511,

1.152) favoring the abemaciclib arm. A subsequent interim analysis provided additional information on the OS effect. At the time of the interim analysis (cutoff: 51 months from study start), the IDFS in the intent-to-treat (ITT) population remained statistically significant and the observed OS HR was 0.929 (95% CI: 0.748–1.153). Upon review of updated data at this second interim OS analysis, FDA broadened the approved population by removing the requirement for high Ki-67 expression. Although still pending final analysis, these results indicate a more favorable OS trend with further follow-up.

**Key Insight:** Interim OS data, especially when based on a small proportion of events relative to a large trial population, may not provide sufficient insight into clinical benefit. Benefit was initially observed in patients with high Ki-67 expression. As the data matured, including the observed OS HR dropping below 1, FDA determined that "… although OS remains immature and not statistically significant, a potential detriment in survival was no longer observed for the ITT population."[5] The indication was subsequently expanded to remove the requirement of a Ki-67 score of ≥20%. The original indication in the Ki-67 ≥20% population was only granted because the population was prespecified in the statistical hierarchy and had an OS HR <1, highlighting the importance of prespecifying subgroups in the statistical analysis plan. The indication was expanded after further follow up, highlighting the importance of ensuring long-term data collection to support broader treatment decisions.

## 3. MONALEESA-2 Trial (Ribociclib + Letrozole in HR+/HER2- Metastatic Breast Cancer)

In the MONALEESA-2 trial, ribociclib, a CDK4/6 inhibitor, was tested in combination with letrozole in 668 patients with HR+/HER2- metastatic breast cancer.[6–8] The PFS data indicated a statistically significant outcome (HR = 0.556, 95% CI: 0.429–0.720). However, the initial interim OS analysis (cutoff: 24 months from study start) revealed a hazard ratio greater than 1 (HR = 1.128, 95% CI: 0.619–2.055). At that time, only 43 OS events had been observed, with an information fraction of 11%. A pre-planned OS analysis one year later (cutoff: 36 months from study start) showed improved OS results with an HR of 0.746 (95% CI: 0.517–1.078). These updated data were submitted to regulators and considered in the initial approval. The final OS analysis (cutoff: 78 months from study start) also showed a statistically significant survival benefit (HR = 0.76, 95% CI: 0.63–0.93).

**Key Insight:** The timing of interim OS analyses is critical, as early assessments may not reflect the true treatment benefits. This underscores a broader challenge in drug development, where a small number of events and/or low information fractions can lead to misleading conclusions. It is crucial to consider the HR in combination with its confidence interval (and width), which can characterize the uncertainty that is present at interim analyses.

## 4. Bellini Trial (Venetoclax + Bortezomib in Relapsed/Refractory Multiple Myeloma)

The Bellini trial investigated the combination of venetoclax, a targeted BCL-2 inhibitor, with bortezomib in patients with relapsed or refractory multiple myeloma.[9] The trial included 291 patients. The PFS analysis

showed a statistically significant outcome (HR = 0.63, 95% CI: 0.44–0.90). However, OS at the first interim analysis (cutoff: 18 months from study start) showed a hazard ratio greater than 1 (HR = 2.03, 95% CI: 1.04–3.95), suggesting detriment when analyzed in the overall ITT population. This risk of increased mortality was added to the FDA label for venetoclax under the Warnings and Precautions section, and a partial clinical hold was placed on clinical trials of venetoclax in patients with multiple myeloma. The subgroup of patients with the t(11;14) translocation did not show the same level of OS detriment, however, meaningful conclusions were not possible given the small size of the population (n=35). The final analysis of OS (33 months from study start) showed an OS HR of 1.19 (95% CI: 0.80-1.77), suggesting a lack of benefit and risk of increased mortality in the ITT population, but with a wide confidence interval. A Phase 3 trial (CANOVA) of venetoclax plus dexamethasone compared to pomalidomide plus dexamethasone was subsequently conducted in patients with t(11;14)-positive multiple myeloma; however, the trial failed to demonstrate statistical significance on the primary endpoint of PFS superiority.

**Key Insight:** Significant improvement in an early endpoint was observed, but with an OS detriment in the ITT population, which was later confirmed with more mature data. OS data indicated a possible benefit in a subgroup of patients, but the assessment was limited by the small sample size in this subgroup. This highlights the need for careful preplanning of interim analyses to assess harm, as well as appropriate powering of subgroup analyses to identify both potential benefits and risks within distinct patient populations. This approach ensures that meaningful effects are not overlooked and that potential detriment in other subgroups is properly addressed.

## 5. PI3K Inhibitors in Hematological Malignancies

Phosphatidylinositol 3-kinase (PI3K) inhibitors have been explored for their therapeutic potential in hematological malignancies. Four PI3K inhibitors—idelalisib, copanlisib, duvelisib, and umbralisib—received FDA approval for indications involving relapsed or refractory indolent non-Hodgkin lymphoma (NHL) or chronic lymphocytic leukemia (CLL). Despite showing promising results in terms of durable ORR or PFS, significant concerns emerged regarding their OS outcomes and tolerability.[10]

These drugs demonstrated substantial toxicities, including severe immune-mediated side effects such as hepatotoxicity, pneumonitis, colitis, and increased risk of infections. For example, idelalisib had halted trials in untreated CLL and indolent NHL due to increased deaths and severe adverse events. The UNITY-CLL trial of umbralisib showed an interim OS HR of 1.23, suggesting a potential increase in mortality compared to control.[11] Similar trends were observed across trials with other PI3K inhibitors, leading to safety concerns and voluntary withdrawals of certain indications.

**Key Insight:** The class-wide issues with PI3K inhibitors highlight the importance of evaluating both efficacy and safety comprehensively, particularly when using early endpoints like ORR to support initial approval. While these drugs improved ORR, the interpretation of their impact on OS is complicated by substantial toxicities that may negate the benefits. This underscores the need for careful assessment of the benefit-risk balance and ongoing OS monitoring, especially in cases where early endpoints show benefit but OS data suggest harm.

## Frameworks and Strategies for Interpreting Interim OS Data

Interim OS data can provide early insights into both the potential benefits and risks associated with a treatment, but as noted, they are often challenging to interpret due to data maturity, leading to variable HRs and wide CIs. Over-reliance on point estimates, which do not reflect underlying uncertainties about harms or benefits, may lead to misinterpretation. More robust approaches for designing and interpreting interim OS data are needed to ensure reliability and accuracy.

Current study design considerations may not adequately account for the complexities of interim OS analyses. Recent FDA-sponsored discussions and subsequent external publications emphasize the need for more thoughtful trial designs and comprehensive planning to consider these complexities when assessing potential harm using interim OS data.[12, 13] To address these challenges effectively, we consider two quantitative frameworks for interpreting interim OS data. Additional work to refine each and establish standards for trial sponsors and regulators regarding their practical implementation is desirable. These quantitative frameworks are briefly summarized below:

A **streamlined quantitative framework** can provide a more straightforward, predefined approach focusing on a standardized set of criteria for interim OS interpretation. This framework may be optimal in trials where patient crossover is not permitted and an assumption of proportional hazards is plausible, meaning that the underlying treatment effect is expected to remain consistent over time. In such cases the focus is on quantifying the uncertainty around the potential for unacceptable harm. Pre-specified thresholds, such as a minimum number of events and information fraction, and an upper limit for the HR CI, guide the interpretation and are particularly efficient when limited variability is expected in interim OS outcomes. However, it is important to tailor these thresholds to reflect the clinical considerations specific to each trial, including factors such as the disease setting, expected survival on the control therapy, and unmet medical need.

Alternatively, a **comprehensive quantitative framework** may be necessary for trials with more complexity, such as those involving non-proportional hazards or patient crossover.[4] This approach would incorporate a broader set of tools to enable deeper analysis when interim OS results are expected to be less conclusive or when complex trial dynamics may make it harder to get a reliable estimate of the treatment's true effect. This may include probabilistic assessments to quantify the likelihood of harm or benefit and the integration of qualitative factors such as patient comorbidities and subsequent therapies. This framework ensures that early signals of potential harm or benefit are not overlooked due to the complexity of the trial design or the mechanism of action of the novel drug.

### A Multi-Step Approach for Study Design and Interpretation of Interim OS Data

These two proposed quantitative frameworks can be integrated into a multi-step approach for interpreting interim OS data. This multi-step approach incorporates a descriptive evaluation with the quantitative evaluation to provide a structured methodology for comprehensively understanding the data and ensuring decisions are evidence-based and aligned with trial objectives. This approach can also be used for

interpretation of results and to prospectively align on the trial design features that can make the interpretation more reliable. The proposed multistep approach is summarized in Figure 1.

| | Design Phase | Analysis Phase | | Interpretation Phase |
|---|---|---|---|---|
| Objectives | **Step 1: Trial Design & Analysis Planning Phase**<br>• Pre-specify interim OS analysis framework<br>• Pre-specify criteria and assumptions for trial | **Step 2: Qualitative Analysis**<br>• Review event counts per arm<br>• Conduct patient-level safety review (adverse events, comorbidities)<br>• Perform aggregate analysis of deaths and related covariates (e.g., age, treatment tolerance) | **Step 3: Quantitative Analysis**<br>• Follow streamlined or comprehensive quantitative framework<br>• Calculate OS Hazard Ratio (HR) and confidence interval<br>• Compare HR to predefined harm threshold<br>• Apply Bayesian or predictive probabilities (if applicable). | **Step 4: Interpretation & Decision-Making**<br>• Synthesize qualitative and quantitative analyses<br>• Perform benefit-risk evaluation incorporating multiple endpoints (e.g., PFS, ORR, quality of life) |
| Output | Set criteria and assumptions for trial initiation | Qualitative summary and insights into potential early signals of harm or benefit | Statistical assessment of potential OS detriment or benefit | Description of decisions based on outputs |

Figure 1. Proposed Multi-Step Approach for Evaluating Interim OS Data in Oncology Trials.

While randomization allows for unbiased comparison across treatment arms, immature data and other factors previously noted make interpreting interim OS data difficult. In this context, it may be useful to consider alternative summary measures as supplementary analyses for the comparison of the interim OS data, rather than just the hazard ratio. Adequately powered, randomized comparisons are still the best approach for generating reliable effect estimates, but early insights can be gained by supplementing these comparisons with contextual analyses. Thus, the first step of interim OS data interpretation involves qualitative assessment, based on a structured descriptive summary of the available data. Sponsors could provide a review of the number of events, establishing a rate per person per year in each arm, a case-by-case examination of each death including precursor safety findings. A more in-depth patient-level assessment could explore whether adverse events (AEs) or lab abnormalities were related to, or led to, death. An analysis of comorbidities and dose considerations could be conducted at the patient and at the aggregate population level, which would involve determining if dose interruptions or reductions occurred in response to these AEs and whether they resolved after the changes. Evaluating patient baseline characteristics, such as age, existing comorbidities, and other risk factors, can provide additional insights into how these factors may have influenced outcomes, as comorbidities could exacerbate AEs or affect treatment tolerance. Further, examining pharmacokinetic (PK) exposure data may help identify whether unusually high drug exposure contributed to toxicity or death. While a case-by-case patient-level assessment might be necessary in some scenarios, a more practical approach for larger Phase 3 trials, may involve conducting an aggregate-level analysis that compares patients with different outcomes, such as those who survived and those who did not, to identify any meaningful differences. However, additional

assumptions or alternative estimators may be needed to establish whether these observed differences reflect a harmful or beneficial causal effect of the novel drug.[14]

These descriptive insights can help provide necessary clinical context and identify any evidence of excess mortality or early signals of harm and confounding factors. Such descriptive analyses can help determine whether the interim OS results warrant deeper quantitative exploration, if there are sufficient data to do so.

Once a descriptive understanding is established, the prespecified streamlined or comprehensive quantitative framework can then be applied to further evaluate the data. Evaluations may include calculating the HR point estimate and an associated CI. One can envisage setting a threshold for harm and assessing the upper end of the CI to quantify the degree of uncertainty around potential for harm, as is done in classical statistical frameworks. As the data mature, the evidentiary threshold required to rule out harm may become more stringent, and it may help to consider two-sided confidence intervals less than 95% at interim analyses for assessment of harm.[12] Assessment of the risk of erroneous conclusion with regard to unacceptable OS detriment (False Positive and False Negative), under different assumptions, should be provided to determine the reliability of the results and facilitate a more transparent trade-off of risks associated with any decision making based on such interim OS data. Conditional probabilities or Bayesian predictive probabilities, based on current data and external evidence, may help predict whether the final OS outcome is likely to be neutral, beneficial, or detrimental. This quantitative step may support assessments of how early OS detriment can be established with high certainty, if it exists, and how frequently the wrong conclusions may be drawn.

The final step may involve synthesizing the descriptive and quantitative findings into a broader benefit-risk evaluation, considering multiple endpoints. This can include a totality of evidence approach, incorporating not only OS but also other endpoints such as PFS and ORR as well as safety, tolerability, and quality of life endpoints beyond OS.

## Design Stage Considerations

Effective study design is crucial to ensuring reliable interim OS data interpretations. This section outlines key considerations to manage factors such as patient crossover or non-proportional hazards in the design stage. To allow for appropriate data capture and analysis, the decision to use a streamlined or comprehensive quantitative framework should be pre-specified during the design stage of the trial. This decision should consider factors such as risks for non-proportional hazards (e.g. the potential for delayed treatment effect or subgroups with heterogeneous treatment effects), early patient crossover, data maturity, duration of follow-up, and overall study power. Robust trial designs plan for adequate follow-up duration to ensure sufficient data collection and maturity at interim analyses as well as pre-specify criteria for patient crossover and minimize/manage missing data.

If heterogeneous treatment effects are expected in subgroups, the study would need to be sized appropriately to enable a thorough benefit-risk assessment in each of the subgroups. Leveraging historical data from similar therapies and/or patient populations is one strategy that can help estimate relationships between early endpoints and OS, as well as predict HRs and whether hazards are proportional throughout

allowing for more informed predictions of potential outcomes. Evaluating how control groups perform on key endpoints can also help set expectations and provide context for interpreting OS findings. However, gaps remain in standardizing methods to determine the impact of early safety events on OS and in using historical data to establish specific thresholds for defining harm.

Including simulation of expected survival curves and determining the operating characteristics over a range of plausible assumptions and aligning these with the planned OS assessment criteria are important in the trial design stage. Such simulations can help set the timings for analyses and optimize the study's power, especially given that the timing of this OS interim analysis is often driven by PFS or ORR analysis milestones. Furthermore, it is important to plan for the collection of OS data even after final PFS analyses are completed, and to pre-specify OS interim analysis milestones and approaches for handling patient crossover. Many analysis techniques that adjust for crossover make an assumption of no unmeasured confounding. To support this assumption, trials would need to capture detailed information on potential (fixed or time-varying) confounders—specifically, patient-level covariates linked to both prognosis and the likelihood of crossover.[15] Collecting data on factors such as comorbidities, baseline characteristics, and treatment-related considerations can help mitigate confounding, enhancing the reliability of analytical assumptions. The impact of non-proportional hazards can also be anticipated and planned for at the design stage through simulations of various patterns such as delayed separation or early small excess harm followed by benefit. When non-proportional hazards are expected, additional metrics beyond the traditional OS HR, such as restricted mean survival time (RMST) or milestone survival rates (i.e., KM estimate at 1, 2, or 3 years), or piecewise hazard ratios (e.g., HRs from 0-6 months and after 6 months) may be considered and prespecified as supplementary analyses.[16] If these supplementary analyses are being considered to address non-proportional hazards, the design stage is the appropriate time to set the analysis interval cutoffs.

Timing is another critical consideration in interim OS analyses. The timing should balance the need for early decision-making with the risk of making incorrect decisions based on incomplete data. Conducting an analysis too early may lead to uncertain conclusions if there are not enough events to provide reliable information. This can be partially avoided through the pre-specification and agreement of a harm threshold for the interpretation of early OS data, and the level of evidence required at each analysis time point to rule out harm.

It is not only the timing of interim analyses that matters, but also the overall event accumulation rate for OS. In some scenarios, low event rates and the associated power for OS may mean that even waiting longer may not lead to significantly improved probability of detecting an OS treatment effect. The design stage can also be used to assess false positive and false negative rates based on the harm threshold and alternative (OS benefit) threshold, either through simulations or in some simple settings through modified power calculations.

## Interpretation Stage Considerations

With a robustly established design which includes OS as a safety endpoint, a comprehensive and methodical interpretation of OS can begin at the primary endpoint assessment (e.g., the early endpoint). The interpretation stage can benefit from both descriptive and quantitative evaluations described above. Accurately interpreting interim OS data requires a range of analytical methods and metrics.

When efficacy trends in a subgroup differ from the overall study population or other subgroups, it is important to determine whether the observed OS detriment is likely due to chance or is plausible from a scientific, biological, or clinical perspective. For instance, does the difference align with the treatment's mechanism of action? Do patients in the subgroup have distinct clinical or biological characteristics that predispose them to a higher risk of adverse effects? Could variations be attributed to differences in clinical practices across sites or regions? It's also essential to examine the totality of the data in subgroups, including other safety and efficacy endpoints. If the detrimental OS is associated with higher incidences of serious or high-grade adverse events or lab abnormalities, this suggests a potential concern rather than a chance finding—this holds true whether observed in specific subgroups or across the overall trial population. Similarly, if the detriment is observed consistently across multiple endpoints, such as ORR, PFS, and OS, this further raises safety concerns.

Interpreting trial data alongside external evidence—such as literature, prior trials, or real-world data—may provide additional valuable insights, particularly when the trial sample size is limited or the data are immature. If the observed OS trend aligns with findings from prior trials of the same agent or others with a similar mechanism of action, this trend is less likely to be due to random chance.

When interpreting immature OS data at the time of an early endpoint analysis, revisiting design assumptions based on accrued information can offer valuable insights. Viewing this data in a Bayesian framework, where assumptions range from implausible to more likely scenarios, can help reviewers better visualize uncertainty and refine their expectations for future events. Additionally, other trial monitoring methods may be adaptable for evaluating OS as an early safety indicator. Characterizing error rates (e.g., false positives and negatives) and using tipping point analyses or other methods (e.g., Bayesian) to evaluate the robustness of interim OS results can help account for potential future variations.

Table 1 provides a summary of evolving strategies, outlining both current approaches and emerging best practices for improving the design and interpretation of interim OS data. It serves as a starting point to help navigate the complexities of using interim OS as a safety endpoint, managing trial design considerations, and handling data immaturity at interim analyses.

**Table 1.** Evolving Strategies for the Design and Interpretation of Interim OS Data.

| Category | Current Approaches and Emerging Best Practices |
|---|---|
| **Clarification of OS as a Safety Endpoint -** OS is frequently used as a safety endpoint when evaluating early endpoint data. | **Current Approach:**<br>- Typically, specify OS as a co-primary or secondary efficacy endpoint when feasible and clinically relevant.<br>**Emerging Best Practices:**<br>- Pre-specify OS analysis plans for safety, including clear definitions of OS detriment and thresholds for defining harm informed by discussions with regulatory authorities. |
| **Trial Design Considerations -** Factors like non-proportional hazards, patient crossover, data maturity and completeness, duration of follow-up, evolving standard of care, and study power can complicate interim OS interpretation. | **Current Approach:**<br>- Some consideration of evaluation of design factors such as non-proportional hazards, crossover, data maturity and completeness aiming to minimize bias of OS assessment at OS planned interim analyses.<br>**Emerging Best Practices:**<br>- Emphasis on OS data collection beyond approval milestones.<br>- In cases where cross-over is unavoidable, additional data collection on key baseline and time-varying covariates associated with patient prognosis and likelihood to crossover.<br>- Systematic application of a quantitative framework for the transparent trade-off of risk of false negative vs false positive assessment of potential OS detriment<br>- Monitor design assumptions closely and avoid deviations when possible. |
| **Handling OS Data Immaturity at Interim Analysis -** OS data is often immature at interim analyses, leading to variability and potential misinterpretation. | **Current Approach:**<br>- Focus on OS driven interim analysis frameworks, which are most often group-sequential in nature.<br>**Emerging Best Practices:**<br>- A thorough assessment is performed to quantify the degree of uncertainty around potential for unacceptable detriment in OS.<br>- Incorporate both qualitative and quantitative assessments to provide context for immature data. Engage patients through patient preference studies to define acceptable margins for potential OS detriment in specific settings.<br>- Focus on standard key analyses (to be defined) such as comparing HR and CI to predefined harm thresholds and conducting qualitative patient-level assessments.<br>- Use simulations and/or Bayesian models to refine predictions of final OS results, conditional on existing data and/or using external data if appropriate.<br>- Use of tipping point or Bayesian framework to assess the robustness of interim OS data with respect to any future potential risk of detriment. |

## Future Considerations for Tool Development and Best Practice Alignment

This section lays out potential strategies for tool development and further best practice development that may support stakeholders in designing and interpreting interim analyses. These tools would be intended to streamline trial processes, improve decision-making, and enhance the robustness of OS data interpretation.

### Use of Simulations for Enhanced Decision-Making

Simulations (see Appendix 2 for a concept proposal) can provide a powerful means to predict and explore the various outcomes at the interim stages of a trial, quantifying the operating characteristics of clinically driven decision-making thresholds, and ensuring robustness in trial design. Simulations can be particularly useful in oncology trials where data immaturity and crossover effects can obscure true treatment effects, providing a means to model outcomes under different conditions. Specifically:

- Simulations may help in understanding the possible trajectories of survival outcomes under different scenarios, such as varying treatment effects, patient heterogeneity, crossover, and information fractions.
- Simulations could be used to define the thresholds for potential harm or benefit and evaluate their operating characteristics, assess the impact of treatment crossover and non-proportional hazards, and inform the timing of interim analyses. They may also be valuable for identifying scenarios where low power for OS could result in less reliable conclusions even with extended follow-up.
- Simulations could be used to predict future outcomes based on current study data in various endpoints and integration of relevant external data if appropriate.

### Development of Tools for Design and Interpretation

In addition to simulations, practical tools could be developed to guide sponsors and researchers in designing trials and interpreting interim OS results more effectively. These may include:

- A structured assessment aid could be developed to assist sponsors during the trial design stage. This tool may consist of a uniform set of questions to help guide thinking around key aspects such as patient crossover, pre-specifying OS analysis milestones, determining adequate follow-up durations, data collection on important patient covariates associated with prognosis and key intercurrent events, and patient heterogeneity.
- Providing standardized methodology for key decisions that impact quality and completeness of data (OS and other data) collection could help ensure consistency across trials and provide clear guidance on how to mitigate bias and enhance the reliability of interim OS data.
- Bayesian modeling approaches could be incorporated as a complementary tool to simulations and standard conditional probabilities. These models may provide probabilistic statements regarding the magnitude of the final OS treatment effect (e.g., HR) based on the observed interim data. Priors may be informed by historical relationships between early endpoints (e.g., PFS, ORR) and OS,

allowing for a more nuanced and evidence-based interpretation. Using Bayesian frameworks could allow for the integration of new information as it becomes available as well as data outside of the study, thus improving the precision of interim OS estimates and supporting better-informed decisions.

### Best Practices Alignment

To facilitate consistent application of best practices in designing and interpreting interim OS data, aligned best practices could be developed and uniformly adopted by stakeholders, reducing variability in approaches to interpreting interim OS data.

## Conclusion

The interpretation of interim OS data in oncology trials poses unique challenges. Early endpoints, such as ORR and PFS, are often the basis for accelerated approvals, allowing timely access to potentially beneficial therapies. However, instances in which immature OS data conflicts with efficacy signals detected with early endpoints can lead to uncertainty around the treatment's benefit-risk profile. This white paper highlights the importance of carefully evaluating and considering interim OS data, so it provides meaningful data to support evaluation of new drugs.

The case studies provided illustrate how factors such as immature OS data, patient subgroups, early patient crossover, and information fractions impact the interpretation of results. These examples reinforce the need for a framework that integrates descriptive and quantitative analyses, supported by thorough preplanning, to ensure accurate conclusions.

## Future Directions

To address these challenges and optimize the use of interim OS data in oncology drug development, several next steps should be considered:

1. Adopting a structured, multi-step approach for interpreting interim OS data, starting with descriptive assessments and followed by quantitative analyses tailored to the therapy and clinical context.
2. Prioritizing robust trial designs that pre-specify OS milestones, harm thresholds, and strategies for incomplete data handling. Simulations can be used to predict outcomes and optimize designs.
3. Fostering continued collaboration among regulators, sponsors, and statisticians to harmonize methods for evaluating interim OS data. Future efforts should focus on refining simulation methods, threshold setting, and predictive modeling, tailored to oncology trials.

By addressing these key areas, the interpretation of interim OS data can be improved, leading to more accurate, timely decisions that benefit patients.

## References


1. Zhang JJ, Blumenthal GM, He K, et al: Overestimation of the Effect Size in Group Sequential Trials. Clin Cancer Res 18:4872–4876, 2012. Available from: https://pubmed.ncbi.nlm.nih.gov/22753584/

2. Morris MJ, Castellano D, Herrmann K, et al: 177Lu-PSMA-617 Versus a Change of Androgen Receptor Pathway Inhibitor Therapy for Taxane-Naive Patients with Progressive Metastatic Castration-Resistant Prostate Cancer (PSMAfore): A Phase 3, Randomised, Controlled Trial. Lancet 404:1227–1239, 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39293462/

3. Johnston SRD, Toi M, O'Shaughnessy J, et al: Abemaciclib Plus Endocrine Therapy for Hormone Receptor-Positive, HER2-Negative, Node-Positive, High-Risk Early Breast Cancer (Monarche): Results From a Preplanned Interim Analysis of a Randomised, Open-Label, Phase 3 Trial. Lancet Oncol 24:77–90, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36493792/

4. Royce M, Osgood C, Mulkey F, et al: FDA Approval Summary: Abemaciclib with Endocrine Therapy for High-Risk Early Breast Cancer. J Clin Oncol 40:1155–1162, 2022. Available from: https://pubmed.ncbi.nlm.nih.gov/35084948/

5. Royce M, Mulkey F, Osgood C, et al: US Food and Drug Administration Expanded Adjuvant Indication of Abemaciclib in High-Risk Early Breast Cancer. J Clin Oncol 41:3456–3457, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/37104738/

6. Hortobagyi GN, Stemmer SM, Burris HA, et al: Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. N Engl J Med 375:1738–1748, 2016. Available from: https://pubmed.ncbi.nlm.nih.gov/27717303/

7. Hortobagyi GN, Stemmer SM, Burris HA, et al: Updated Results from MONALEESA-2, A Phase III Trial of First-Line Ribociclib Plus Letrozole Versus Placebo Plus Letrozole in Hormone Receptor-Positive, HER2-Negative Advanced Breast Cancer. Ann Oncol 29:1541–1547, 2018. Available from: https://pubmed.ncbi.nlm.nih.gov/29718092/

8. Hortobagyi GN, Stemmer SM, Burris HA, et al: Overall Survival with Ribociclib plus Letrozole in Advanced Breast Cancer. N Engl J Med 386:942–950, 2022. Available from: https://pubmed.ncbi.nlm.nih.gov/35263519/

9. Kumar SK, Harrison SJ, Cavo M, et al: Venetoclax or Placebo in Combination with Bortezomib and Dexamethasone in Patients with Relapsed or Refractory Multiple Myeloma (BELLINI): A Randomised, Double-Blind, Multicentre, Phase 3 Trial. Lancet Oncol 21:1630–1642, 2020. Available from: https://pubmed.ncbi.nlm.nih.gov/33129376/

10. Meeting of the Oncologic Drugs Advisory Committee Meeting Announcement - 04/21/2022 | FDA. Available from: https://www.fda.gov/advisory-committees/advisory-committee-calendar/updated-information-april-21-22-2022-meeting-oncologic-drugs-advisory-committee-meeting-announcement

11. TG Therapeutics Announces Voluntary Withdrawal of the BLA/sNDA for U2 to Treat Patients with CLL and SLL | TG Therapeutics, Inc. Available from: https://ir.tgtherapeutics.com/news-releases/news-release-details/tg-therapeutics-announces-voluntary-withdrawal-blasnda-u2-treat

12. Fleming TR, Hampson L V, Bharani-Dharan B, et al: Monitoring Overall Survival in Pivotal Trials in Indolent Cancers. 2023. Available from: https://arxiv.org/abs/2310.20658v3

13. Rodriguez LR, Gormley NJ, Lu R, et al: Improving Collection and Analysis of Overall Survival Data. Clin Cancer Res 30:OF1–OF9, 2024 Available from: https://pubmed.ncbi.nlm.nih.gov/39037364/

14. Bas BB, Groenwold RHH: Identification of Causal Effects in Case-Control Studies. BMC Med Res Methodol 22, 2022. Available from: https://pubmed.ncbi.nlm.nih.gov/34996386/

15. Watkins C, Huang X, Latimer N, et al: Adjusting Overall Survival for Treatment Switches: Commonly Used Methods and Practical Application. Pharm Stat 12:348–357, 2013. Available from: https://pubmed.ncbi.nlm.nih.gov/24136868/

16. Lin RS, Lin J, Roychoudhury S, et al: Alternative Analysis Methods for Time to Event Endpoints under Non-proportional Hazards: A Comparative Analysis. Stat Biopharm Res 12:187–198, 2019. Available from: https://www.tandfonline.com/doi/full/10.1080/19466315.2019.1697738

# Appendices

**Appendix 1.** Summary of Select Oncology Clinical Trials: Early Endpoint and Interim Overall Survival Readouts.

| Study | Patient Population | N | Early Endpoint (HR, 95% CI) | Early Endpoint Events | OS Readout | OS Events | OS HR (95% CI) | Cut-Off Date |
|---|---|---|---|---|---|---|---|---|
| **PSMAFore Trial** (177Lu-PSMA-617) | Metastatic Castration-Resistant Prostate Cancer (mCRPC) | 468 | rPFS HR = 0.41 (95% CI: 0.29–0.56) | 166 | Interim OS #2 | 134 | 1.16 (0.83, 1.64) | 21-Jun-2023 |
| | | | | | Interim OS #3 | 216 | 0.98 (0.75, 1.28) | 27-Feb-2024 |
| | | | | | Final OS | Pending | Pending | Pending |
| **MonarchE Trial** (Abemaciclib + ET) | Early HR+/HER2- Breast Cancer | 5637 | IDFS HR = 0.696 (0.588–0.823) | 565 | Interim OS #1 | 186 | 1.09 (0.82, 1.46) | 1-Apr-2021 |
| | | | | | Interim OS #2 | 330 | 0.93 (0.75, 1.15) | 1-Jul-2022 |
| | | | | | Interim OS #3 | 442 | 0.90 (0.75, 1.09) | 3-Jul-2023 |
| | | | | | Final OS | Pending | Pending | Pending |
| **MONALEESA-2** (Ribociclib + Letrozole) | HR+/HER2- Metastatic Breast Cancer | 668 | PFS HR = 0.556 (0.429–0.720) | 243 | OS Interim #1 | 43 | 1.13 (0.62, 2.06) | 29-Jan-2016 |
| | | | | | OS Update | 116 | 0.75 (0.52, 1.08) | 2-Jan-2017 |
| | | | | | Final OS | 400 | 0.76 (0.63, 0.93) | 10-Jan-2021 |
| **Bellini Trial** (Venetoclax + Bortezomib) | Relapsed/ Refractory Multiple Myeloma | 291 | PFS HR = 0.63 (0.44–0.90) | 129 | OS Interim #1 | 52 | 2.03 (1.04, 3.95) | 26-Nov-2018 |
| | | | | | Final OS | 114 | 1.19 (0.80, 1.77) | 15-Mar-2021 |

## Appendix 2. Concept Plan and Future Directions for Interim OS Data Interpretation.

The simulation workplan is divided into distinct components, as described below. This initial work focuses solely on OS, largely independent of PFS. Future work can incorporate PFS directly into joint models or indirectly as part of a scenario regarding the totality of evidence across multiple endpoints.

1. Establish and evaluate various thresholds in a 'streamlined' criteria for harm based solely on the observed events available at the interim. This is done under the simplest assumptions to triangulate the initial set of criteria to be included in the evaluation. For example, we may find that a stringent criterion such as the upper bound of the confidence interval of the hazard ratio of 1.3 is almost equivalent to a test of efficacy, making it irrelevant to the intent of ruling out harm. Likewise, a value of 1.8 may be found to be too lenient, allowing obviously concerning scenarios to occur in an undesirably large proportion of simulation trials.
    a. It may be possible to determine a reasonable range of target operating characteristics equivalent to Type I and Type II error from the work above.
2. Evaluate existing frameworks and/or devise a new mathematical representation of the more complex scenarios that have occurred in practice, including considerations such as non-proportional hazards (e.g., early overlap of Kaplan-Meier curves followed by later separation, or early harm followed by separation), patients crossing over to the treatment arm at disease progression, dropout rates, information fraction, and the number of events available at the interim. This can be done by digitizing real examples, such as those described above, into piecewise hazard functions or by generating hypothetical scenarios. In either case, we can then evaluate the operating characteristics of the various frameworks, as outlined below.
3. Provisional Scenarios:
    a. Neutral effect on OS, proportional hazard of 1.0 throughout the trial.
    b. Separation of OS Kaplan-Meier curves, proportional hazard of modest scale (e.g., HR 0.9).
    c. Separation of OS Kaplan-Meier curves, proportional hazard of significant scale (e.g., HR 0.6).
    d. Delayed and modest separation of OS Kaplan-Meier curves after the interim (non-proportional hazard: 1.0 prior, 0.9 thereafter).
    e. Delayed and significant separation of OS Kaplan-Meier curves after the interim (non-proportional hazard: 1.0 prior, 0.6 thereafter).
    f. Small excess harm prior to the interim and small separation thereafter (non-proportional hazard: 1.15 prior, 1.0 for a period, 0.9 thereafter).
    g. Small excess harm prior to the interim and wide separation thereafter (non-proportional hazard: 1.15 prior, 1.0 for a period, 0.6 thereafter).

h. Early modest benefit with later modest harm, with initial separation of OS Kaplan-Meier curves showing a proportional hazard of 0.9, followed by a reversal to show modest harm (HR 1.15 thereafter).

i. Sustained modest harm throughout, reflected by a consistent proportional hazard of 1.15 maintained throughout the trial.

j. Increasing harm over time, with initial modest harm (HR 1.15) that intensifies to a more significant level (HR 1.3 thereafter).

4. For each scenario of interest, use a variety of approaches to evaluate the following:
    a. How often do we incorrectly conclude harm when there isn't any?
    b. If there is harm, how often can we conclude correctly based on early data?
5. For both questions, establish whether there is a minimum amount of data (e.g., number of events) that is optimal for reasonably reliable decision making.
6. The concepts above are mainly applied to the interpretation of interim OS data; however, the true value lies in translating them to the design stage. We propose a few examples that will identify a recommended process flow for design considerations and the approach at each step.

## 35 THE IMPACT OF TREATMENT MODALITIES ON USE OF CTDNA AS AN EARLY ENDPOINT IN ANSCLC TRIALS

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

1 Friends of Cancer Research, Washington, DC, USA; 2 NMD Group Inc., Bala Cynwyd, PA, USA; 3 Cancer Research And Biostatistics, Seattle, WA, USA; 4 Genentech, Inc, South San Francisco, CA, USA; 5 Natera Inc., Austin, TX, USA; 6 Foundation Medicine, Inc., Cambridge, MA, USA; 7 Guardant Health, Inc., Palo Alto, CA, USA; 8 AstraZeneca, Gaithersburg, MD, USA; 9 Regeneron Pharmaceuticals, Tarrytown, NY, USA; 10 Guardant Health, Redwood City, CA, USA

**Background** Circulating tumor DNA (ctDNA) holds promise as an early endpoint in oncology drug development, particularly in advanced non-small cell lung cancer (aNSCLC) treated with immunotherapy. Friends of Cancer Research established the ctMoniTR Project to aggregate and analyze patient-level data from clinical trials and generate evidence that characterizes the association between change in ctDNA levels on-treatment and associations with overall survival (OS). Using well characterized data from 4 randomized control trials (RCTs), we com-pared change in ctDNA levels among patients treated with an anti-PD(L)1 and/or chemotherapy.

**Methods** Patients received treatment with either anti-PD(L)1 with or without chemotherapy (IO; n=537) or chemotherapy alone (n=291). Each patient had a baseline ctDNA measure-ment (T0) and on-treatment ctDNA measurement within two windows: 2-6 weeks (T1) and 7-13 weeks (T2) after treatment initiation. We evaluated change in ctDNA levels by applying cutoffs tailored for immunotherapy of >50% and >90%decrease in ctDNA (molecular response; MR50 and MR90, respectively). A third group of clearance (non-detected ctDNA on treatment) was included. We used multivariable Cox mod-els to assess associations with OS and compared results from T1 to T2.

**Results** Patients treated with IO with either MR50 or MR90 at T1 or T2 showed improved OS compared to patients without a MR (for MR50 aHR=0.70 [0.54-0.91] p=0.008 at T1 and aHR=0.53 [0.40-0.69] p<0.001 at T2; for MR90 aHR=0.51 [0.35-0.72] p<0.001 at T1 and aHR=0.66 [0.49-0.88] p=0.006 at T2). For IO treated patients, ctDNA at T2 was a significant predictor of OS beyond T1 (Likelihood Ratio Test for MR50 p<0.001; for MR90 p=0.006). However, in patients treated with chemotherapy, ctDNA clearance was asso-ciated with improved OS at T2 (vs. MR50 HR=0.55 [0.34-0.89] p=0.015, vs. MR90 HR=0.60 [0.35-0.99] p=0.048) but achieving MR50 or MR90 did not differentiate survival outcomes from non-MR at either T1 or T2.

**Conclusions** Change in ctDNA levels is strongly associated with OS in patients with aNSCLC treated with IO, which sets the stage for using ctDNA in future prospective trials assessing immunotherapy in patients with aNSCLC. However, more work is needed for assessing change in ctDNA levels in che-motherapy including assessing different times and understand-ing the impact of assays. These findings highlight the potential impact of treatment modality for assessing change in ctDNA levels and further investigation is warranted to refine the tim-ing and clinical thresholds of universally applicable ctDNA metrics.

http://dx.doi.org/10.1136/jitc-2024-SITC2024.0035

FRIENDS of CANCER RESEARCH

# Framework for Integrating Change in ctDNA Levels in Advanced Cancer Clinical Trials to Support Meta-analyses for Intermediate Endpoint Validation

FRIENDS OF CANCER RESEARCH WHITE PAPER | 2024

## Executive Summary

In oncology clinical trials, using intermediate endpoints that are reasonably likely to predict clinical benefit accelerates access to therapies. For these endpoints to support regulatory decision-making, the U.S. Food and Drug Administration (FDA) expects meta-analyses at the patient- and trial-level that demonstrate associations between the intermediate endpoint and long-term clinical outcomes, such as overall survival (OS; i.e., clinical benefit). Circulating tumor DNA (ctDNA), a biomarker found in the blood, can serve as an indicator of tumor burden, and has shown promise as an intermediate endpoint. Initial findings from multiple clinical trials, including the Friends of Cancer Research (*Friends)* ctMoniTR Project that aggregates patient- level data from several clinical trials, demonstrate that decreases in ctDNA levels while on treatment associate with improved OS. However, evidence is lacking for trial-level meta-analyses due to inconsistencies in approaches across trials including study design, data collection, and ctDNA measurement methods, making it difficult to combine results.

To address this gap, *Friends* assembled a working group of experts, including representatives from the FDA, pharmaceutical companies, diagnostics developers, patient advocate organizations, and academia, to align on key considerations for prospectively designed clinical trials that collect ctDNA in a standardized manner. The considerations focus on advanced non-small cell lung cancer (aNSCLC) treated with immunotherapy (IO) due to the robust data established to date and ongoing drug development in this space. With a standardized approach, these trials may be more appropriate to combine with regards to data quality and coherence into a trial-level meta-analysis to support the use of ctDNA as an intermediate endpoint in oncology drug development.

The working group prioritized several critical recommendations for alignment in terms of study design and data collection; however, additional considerations are also outlined. The most critical recommendations for alignment of study design and data collection are:

- Collect a baseline ctDNA measurement before treatment initiation, preferably on C1D1 before infusion.
- Collect four on treatment ctDNA measurements: three during subsequent treatment cycles (i.e., C2D1, C3D1, and C4D1) and one at 6-months post-treatment initiation; align with RECIST measurements as is feasible.
- Use an assay that is sensitive enough to detect ctDNA in at least 70% of patients at baseline.
- Report data related to ctDNA analysis and measurement in an aligned approach (specific recommendations are included in Table 1 of the white paper).

These recommendations aim to align ctDNA collection and analysis in future clinical trials, supporting validation efforts for using ctDNA as an intermediate endpoint in regulatory decision- making, and ultimately accelerating the delivery of treatments for serious and life-threatening diseases to patients.

## Authors

**Charu Aggarwal**, Abramson Cancer Center, University of Pennsylvania
**Valsamo Anagnostou**, Johns Hopkins University
**Hillary Andrews**, Friends of Cancer Research
**Zoe June Assaf**, Genentech, A Member of the Roche Group
**Biswajit Das**, Molecular Characterization Laboratory, Frederick National Laboratory
**Yu Deng**, Genentech, A Member of the Roche Group
**Carin Espenschied**, Guardant Health
**David Fabrizio**, Foundation Medicine, Inc.
**Keith Flaherty**, Mass General Hospital
**Leon Freytor**, Johnson & Johnson
**Minakshi Guha**, Takeda
**Glenn J. Hanna**, Dana-Farber Cancer Institute
**Janet Jin**, Amgen
**John Lee**, Daiichi Sankyo
**Anne-Marie Martin**, GSK
**Dimple Modi**, Regeneron
**Achim Moesta**, Regeneron
**Kiran Patel**, Johnson and Johnson
**Reena Philip**, U.S. FDA
**Kai Qu**, Amgen
**Rajiv Raja**, GSK
**Matthew Silver**, Takeda
**Mark Stern**, BMS
**Carol Vallett**, *Friends*' Advisory Advocate
**Diana Merino Vega**, AstraZeneca
**Paz Vellanki**, U.S. FDA
**Jian Wang**, AstraZeneca
**Kathie Winson**, Genentech, A Member of the Roche Group
**Alexander W. Wyatt**, Vancouver Prostate Centre, University of British Columbia and Clinical Cancer Genomics Program and Michael Smith Genome Sciences Centre, BC Cancer
**Nevine Zariffa**, NMD Group

## Background

Advancements in oncology drug development have significantly improved outcomes for many patients with solid tumor cancers. Given these successes, it can be lengthy and resource-intensive to conduct studies for newer therapies due to the time required for mature survival endpoint readouts, especially for overall survival (OS), which remains the gold standard for evaluating clinical benefit. There is an opportunity to enhance the availability of more treatment options for patients, and thus, a need for additional, novel intermediate endpoints that are reasonably likely to predict clinical benefit, enabling earlier evaluation of efficacy and regulatory decision-making. The Accelerated Approval Pathway can be leveraged for therapies that treat a serious condition and fill an unmet medical need, allowing for U.S. Food and Drug Administration (FDA) approvals based on an intermediate endpoint that is reasonably likely to predict clinical benefit.

While radiographic-based intermediate endpoints exist, such as objective response rate (ORR) and progression-free survival (PFS), there are some challenges with these approaches. In some settings, an objective radiographic baseline measurement, which is a requirement for these approaches, cannot be made (e.g., patients with large pleural effusions or with bone-only metastases). Some cancer types (e.g., metastatic head and neck cancer) are challenging to measure by radiographic measurements and some therapies (e.g., novel treatments with immune-mediated efficacy) may lead to what appears to be progression on imaging but is in fact pseudo-progression. Additionally, guidance from the FDA[1] and recent discussions at an FDA Oncologic Drugs Advisory Committee meeting[2] suggest there is a need for earlier endpoints in the perioperative setting and highlight the challenges with radiographic based endpoints in early-stage disease as surgery often removes any measurable lesions.

In early- and late-stage settings, an objective and standardized intermediate endpoint that can predict long-term clinical benefit is needed to overcome these limitations of radiographic imaging and support efficacy evaluation in a timely manner. On-treatment change in circulating tumor DNA (ctDNA) levels from baseline can capture response at a molecular level and could potentially be used as an intermediate endpoint. Many sponsors recognize the value of ctDNA and leverage early change in on-treatment ctDNA to predict clinical benefit and inform internal Go/No-Go decisions. A more coordinated effort to have a consistent and unified approach to define molecular response based on ctDNA and for the analysis of such endpoints could support the development of ctDNA endpoints in regulatory decision-making.

To qualify novel intermediate endpoints, FDA guidance recommends meta-analyses of randomized controlled trials at both the individual patient- and trial-levels.[1, 3] The Friends of Cancer Research (*Friends*) ctMoniTR Project combines data from multiple clinical trials in the metastatic setting to assess associations between change in ctDNA levels and OS and PFS at the patient-level. These retrospective patient-level analyses have demonstrated that a decrease in ctDNA is associated with improved PFS and OS. The focus herein will be on patients with advanced solid tumors, as these

reflect the bulk of aggregate data analyses conducted to date. Where there are parallels in late-stage that are relevant for early-stage disease, the same approaches could be considered or adapted, as appropriate.

## Scope and Approach to Assessing Change in ctDNA to Date

*Friends* coordinated a working group with representatives from pharmaceutical companies, diagnostics developers, FDA, academia, and patient advocacy groups to align on recommendations for standardized, harmonized, and robust data collection to include in prospectively designed trials that can support meta-analyses. The primary focus is to set the stage for collaborative evidence collection that assesses change in ctDNA levels and associations with OS, supporting the use of change in ctDNA levels as an intermediate endpoint in regulatory decision-making. (While we recognize that the evidence developed to support using ctDNA as an intermediate endpoint could support approaches for using ctDNA to inform clinical practice, the proposed scope of work is not intended to evaluate the use of ctDNA to guide treatment decisions for individual patients.)

Current approaches to assessing change in ctDNA levels and associations with OS set the stage for which data need to be collected. To date, ctDNA is measured early in clinical trials, with many trials including a baseline blood collection before treatment starts and an on-treatment measurement usually taken 3-12 weeks after treatment initiation.[4, 5] There are a variety of assays to measure ctDNA, including next generation sequencing (NGS) and digital droplet PCR (ddPCR) assays. Often, the variant allele frequency (VAF) for all variants included at each timepoint is determined (e.g., mean or maximum VAF) and used to calculate a percent change in ctDNA from baseline to on-treatment. In some cases, a single variant is tracked (e.g., in oncogene driven cancers), and increasingly, various measures of tumor fraction are used to measure ctDNA.[6–9] Some studies have shown that results are similar regardless of whether multiple genes from a panel test are considered or just the gene of interest.[10]

The following sections provide recommendations for incorporating ctDNA into prospectively designed clinical trials. As a use case, we developed initial recommendations based on observations from advanced non-small cell lung cancer (aNSCLC) treated with immunotherapy (IO), due to the robust data established to date and ongoing drug development in this space. The most critical recommendations for alignment in terms of study design and data collection are prioritized, however, additional considerations are also outlined. Robustly designed trials that have incorporated these recommendations may support meta-analyses for validating the use of change in ctDNA levels as an intermediate endpoint. We also outline initial thoughts for how to approach a meta-analysis to support ctDNA as an intermediate endpoint using these prospectively collected data in the Appendix.

## Criteria for a Molecular Response Measurement

A key aspect of each of the following sections is considering which data should be collected and reported for inclusion in a meta-analysis. In addition to outlining these data throughout the following sections, Table 1 provides recommendations for which datapoints should be reported.

**Table 1.** Suggested data to collect and report when running prospective trials that incorporate ctDNA.

| Reportable | Category | Description | Required or Recommended |
|---|---|---|---|
| Plasma volume | Per sample | The volume (mL) of plasma input into assay workflow | Required |
| cfDNA extracted | Per sample | The amount of total cfDNA (nanograms) extracted from plasma | Required |
| cfDNA used | Per sample | The amount of cfDNA (nanograms) input into assay workflow | Required |
| ctDNA detected | Per sample | Binary yes/no of whether ctDNA is detected in the sample | Required |
| Mutation frequency in ctDNA | Per sample | Amount of ctDNA (VAF) per variant measured | Required* |
| Measure of ctDNA level | Per sample | Continuous metric (e.g., between 0 and 100%) summarizing the per sample fraction of ctDNA derived from tumor, which could be the mean, max, or median VAF of detected somatic mutations, or alternative estimates of tumor fraction (multi-omic, methylation, etc.) | Required |
| Timing of baseline measurement | Per patient | Days relative to treatment initiation | Required |
| Timing of on-treatment measurement | Per sample | Days relative to treatment initiation for each sample | Required |
| Percent of patients with detected baseline ctDNA | Baseline characteristics | How many patients have baseline ctDNA detected, including pre-specified level of ctDNA | Required |
| Prior therapy | Per patient | Number of lines of prior therapy, types of therapy, and time since the last therapy (i.e., days from prior line ending to treatment initiation) | Recommended |
| Concurrent therapy | Per patient | Report any medications/ surgery/ radiation that the patient receives beyond systemic chemotherapy | Recommended |
| Molecular response | Per patient | Yes/no of whether the patient has a molecular response | Recommended |
| Molecular response approach | General | Describe how the molecular response was defined including when the response was assessed - consider including the following thresholds: 50% reduction, 90% reduction, 100% reduction (i.e., clearance) | Recommended |

*For assays that measure genetic alterations in multiple genes

## Assay Characteristics

Measuring ctDNA involves assays that assess various genes or other somatic features from a liquid biopsy (here we focus on plasma, but cerebral spinal fluid, urine, and saliva are other examples). The poolability of molecular data in a meta-analysis will depend on the similarity of assays with respect to performance metrics. As feasible, we recommend that an appropriate set of reference materials is used to demonstrate comparability across multiple assays. When selecting an assay for a clinical trial, sponsors should consider sensitivity and specificity at a particular limit of detection (LoD) and clinical cut-off, as well as other assay performance measures, the number and types of genes or somatic features assessed, and approaches to clonal hematopoiesis (CH) variant removal. There are many assays currently in use that detect and quantify ctDNA and technology continues to evolve.

Table 2 outlines proposed minimum requirements for assays to ensure there is transparency in how the diagnostic is used and below we discuss some key technical considerations.

### Minimum Requirements for Assay Analytical Validation

Various factors can influence assay performance including pre-analytical variables (e.g., the volume of plasma collected), the bioinformatics pipeline, and inter-assay variability (e.g., depth and breadth of genomic coverage). When selecting a ctDNA assay, it is critical that the assay follows current recommendations for analytical and clinical validation. BLOODPAC proposed a set of analytical and pre-analytical validation protocols for assessing NGS ctDNA platforms.[11, 12] We recommend diagnostic developers use these or similar protocols to ensure analytical and clinical accuracy and reliability and that clinical trial sponsors report the approach used. We also recommend the cut-off is pre-specified, and the same assay and algorithm be used for the entire trial, including the associated cutoffs.

### Considerations Regarding CH Removal

CH variants are somatic mutations that originate from expansions in hematopoietic progenitor cells.[13, 14] ctDNA is measured as a part of total circulating free DNA (cfDNA), which includes CH variants that can pose a challenge when trying to identify tumor related content or quantify ctDNA levels. It is critically important to be accurate when removing CH variants as they may alter interpretation of changes in ctDNA levels. To account for CH-related mutations, diagnostic developers currently employ one of three main approaches for their assays:[15]

1. An algorithmic approach to removing CH variants that is part of the bioinformatic pipeline, which involves removing genetic mutations commonly found in hematopoietic cells and may leverage other information available from the assay. A challenge with this approach is the possibility of removing variants of interest or not appropriately removing the CH variants given that alterations in some genes (i.e., *TP53, ATM*) may be CH or tumor-derived leading to incorrect CH calls.
2. Tumor informed or bespoke approaches consider the variants found in the sequenced tumor tissue to distinguish ctDNA variants in cfDNA. Apart from limited tissue availability, a

challenge with this approach is that it requires tumor tissue to not only be accessible and removed surgically or through a biopsy but also requires waiting for tumor sequencing to select the appropriate ctDNA variants for measuring/tracking, which may or may not be available in real-time. The analysis is limited to variants present in the tumor tissue specimen at baseline, which comes from a single lesion that may not be representative of genetic alterations at other sites.

3. Peripheral blood mononuclear cells (PBMC) removal approaches use PBMCs collected from blood samples to filter out CH or germline variants. Challenges with this approach include the cost of running the samples twice and sensitivity limitations.

We recommend diagnostic companies explicitly state how they identify and/or exclude CH in their assay and specifically report CH-specific false positive rate. Reporting the probability of detection based on sample-level and allele-level coverage is important for all variants, tumor-derived and CH.

## Assay Sensitivity

In cancer, it is assumed that patients with a sizeable, proliferative tumor have ctDNA in their bloodstream prior to any therapy, reflecting the burden of disease, However, ctDNA detection is impacted by both biological factors, such as tumor location, vascularization and aggressiveness, as well as technical factors, especially assay sensitivity and plasma collection volume.

For prospective trials assessing aNSCLC treated with IO, assays should be sufficiently sensitive such that most patients in the planned trial will have 'detected' ctDNA at baseline. Approximately 70-85% of patients with aNSCLC have detected baseline ctDNA when using an assay with a LoD ~0.1% VAF (1000 ppm),[16, 17] a range that should be considered when selecting an appropriate assay for use. This approximation of detection is a lower range, as more sensitive assays would result in a greater number of patients with detected ctDNA. We recommend that sponsors report their predetermined ctDNA detection level cutoff and the rate of ctDNA detection at baseline.

ctMoniTR findings in patients with aNSCLC treated with anti-PD(L)1 demonstrate that when using an assay with a LoD of as low as 0.3% VAF (3000 ppm), a 50% or 90% decrease in ctDNA is associated with improved OS. Additional data are emerging and will determine the level of sensitivity for other treatment types and settings, including early-stage disease.

## Emerging Technology

To date, much of the work assessing associations between change in ctDNA levels and OS has focused on measuring ctDNA levels by assessing tumor-derived variants (i.e., changes to the genome sequence). There are a variety of emerging approaches for quantifying ctDNA that do not rely only on sequence variants, including assessing changes in cfDNA methylation and cfDNA fragment size distributions as well as physical properties of cfDNA fragments (i.e., the cell free DNA fragmentome). As appropriate, characteristics included herein should be reported for these emerging technologies so that their potential utility relative to currently established approaches can be understood.

**Table 2.** Minimum reporting requirements for assays to ensure there is transparency in how the diagnostic is used and support considerations regarding poolability in the meta-analysis.

| Characteristic | What should be reported | Recommendations |
|---|---|---|
| Limit of detection (LoD) | • LoD as reported by the diagnostics company<br>• Report whether this is LoD50 or LoD95<br>• Reportable range for values below LoD<br>• Approach to defining the LoD | • The LoD of the assay as reported by the assay developer should be 0.3% VAF or lower<br>• Higher sensitivity assays may be worth exploring |
| CH Removal Approach | • Approach to removing CH variants:<br>  ○ Algorithm/machine learning- based removal<br>  ○ Tumor informed<br>  ○ PBMC-analysis-based removal<br>• CH-specific false positive rate at the sample level and variant level | • PBMC removal or tumor informed approaches are preferred<br>• If using algorithmic removal, the assays should clearly report limitations or uncertainties |
| Assay Characteristics | • Assay version (to account for potential modifications over time)<br>• Number of genes and alterations measured and gene names<br>• Detection threshold (cut-off) and approach to determine detected vs. non-detected ctDNA | • Sponsors may choose to conduct all assays at trial completion to avoid time-drift of assay methodology which could add noise (or worse confounding factors) to the trial specific data set |
| Performance Data | • Limit of the Blank (LoB)<br>• Precision<br>• Accuracy<br>• Assay sensitivity and specificity<br>• Pre-analytical approach, including which guidelines were followed<br>• Standardized protocols for sample collection, storage, processing, and handling | • Assay sensitivity should ensure that most patients in the planned trial have 'detected' ctDNA at baseline based on historical data<br>• Pre-analytical assessments should follow established guidelines |

## Timing of Sample Collection for ctDNA Assessment

One of the most critical areas for alignment regarding the ability to combine data from various prospectively designed clinical trials is the timing of blood sample collection for ctDNA analysis. Table 3 prioritizes recommendations for timing of sample collection.

### Baseline Sample Collection

It is critical that sponsors collect a baseline ctDNA measurement before treatment initiation. Ideally, this collection should occur on the same day as the first cycle of therapy (i.e., cycle 1 day 1; C1D1) before infusion. However, some flexibility may be warranted as some patients may visit the healthcare system for laboratory work before their first treatment. When considering appropriate flexibility, sponsors should avoid using the ctDNA assessment from the screening assessment as the baseline because there may be differences in these values.[18] The aligned approach from ctMoniTR was to consider samples collected up to 14 days before treatment initiation as the baseline sample.[19]

### On-treatment Sample Collection

For the on-treatment sample collections, many studies collect samples 3-12 weeks after treatment initiation.[5] The ctMoniTR project assessed on-treatment ctDNA up to 10 weeks from treatment initiation as the 1st on treatment measurement. The project combined data from multiple collection time points within that time window and saw associations with outcomes, suggesting there could be some flexibility on which specific week the samples are collected early in treatment. Ideally, samples would be collected when the patient is present for other reasons such as labs, scans, or infusions. To continue building evidence to compare and contrast radiographic response, sample collections near scans for radiographic response assessment could be helpful. Since many IO infusions occur on similar schedules (i.e., once every 3 or 4 weeks), we highly recommend sample collections for ctDNA assessment occur prior to drug administration on infusion day. Along with this, we recommend ensuring that at least one on-treatment measurement occurs between 2-10 weeks post treatment initiation.

### Frequency of Sample Collection

Whether a “confirmation of response” is necessary for molecular response assessment is a question of interest for which we do not currently have sufficient data. A confirmation is required for radiographic imaging progression in RECIST guidelines[20] and is recommended for biochemical disease progression using prostate specific antigen (PSA) where the 'confirmed' category requires two consecutive measurements to agree on response or non-response. Few studies have assessed plasma during multiple on-treatment timepoints, which makes it challenging to provide recommendations on the dynamics of ctDNA. Clinical trialists are challenged to simplify trials and patients may have clinical progression or toxicity due to treatment, so while collecting samples over continued cycles is ideal, it may not be feasible or practical. To support identifying the most appropriate timing for ctDNA collection, we highly recommend at least 3 subsequent on-treatment

samples are collected (i.e., C2D1, C3D1, and C4D1). This would ideally be on or around the same time as radiological assessment, as is feasible.

### Durability of Response

While demonstrating early associations of change in ctDNA levels with outcomes would be the primary goal of a meta-analysis (**Appendix**), understanding the durability of the molecular response, or how long a decrease in ctDNA or clearance of ctDNA lasts, is also important. It is likely unfeasible for clinical trialists to collect samples for ctDNA assessment every cycle for the entire trial, rather, a single aligned timepoint later in the trial may be more beneficial, for example, using 6-, 9-, or 12-months post-treatment initiation, similar to what was considered when establishing MRD as an intermediate endpoint in multiple myeloma.[21] We recommend sponsors prioritize including a 6-month post-baseline sample collection, as some literature has shown durability of response and this measurement aligns with when PFS6 is assessed, an endpoint often used in studies focused on aNSCLC treated with IO.[22] If feasible, consider also including a 1-year assessment[23] and a measurement at the time of progression.

**Table 3.** Prioritized recommendations for timing of sample collection for ctDNA assessment.

| ctDNA Sample | Definition | High priority | Lower priority |
|---|---|---|---|
| Baseline | Sample measurements before treatment initiation | Collect baseline ctDNA on C1D1 before infusion | Measurements up to 14 days before C1D1 can be considered |
| Molecular Response Assessment | Samples collected after treatment initiation but before progression | Collect samples at the same time as infusions for the subsequent 3 cycles (i.e., C2D1, C3D1, and C4D1) | Continue collecting samples every infusion through progression |
| Durability | Samples collected after a period of time to assess durability of response | Collect a 6-month post-treatment initiation sample | Collect a sample at progression and at 1-year post treatment initiation |

## Patient Inclusion Considerations

When performing the meta-analysis, there may be baseline characteristics (e.g., specific clinical prognostic factors) that should be included in the analysis to assess their impact on the predictive nature of ctDNA. Sometimes, patients with non-measurable disease at baseline by radiographic assessment using RECIST guidelines are excluded from clinical trials. A parallel scenario is non-detected ctDNA at baseline, which may be due to either the limited sensitivity of the assay or true non-detectable ctDNA in the plasma sample. Either way, various reports demonstrate that non-detected ctDNA at baseline is a prognostic biomarker.[8, 17, 24, 25] In the ctMoniTR Project, most patients with non-detected ctDNA at baseline also have non-detected ctDNA on treatment, which makes it challenging to know whether the resulting associations with outcomes are related to the patient's response to treatment. The serial non-detection can also be due to assay limitations (i.e., the

patient's tumor does not have a mutation in the gene panel), the assay's LoD may not support the detection of the mutation, or the volume of plasma was too low (i.e. by chance, insufficient tumor DNA fragments were in the small sample).

Many prospective clinical trials currently under development will consider ctDNA as an exploratory endpoint and inclusion/ exclusion criteria will be tailored to evaluating the primary endpoint (e.g., measurable disease by RECIST assessment). Excluding patients with non-detected ctDNA at baseline could lead to bias and there may be patients who go on to have detected ctDNA on treatment. Additionally, the time it takes for ctDNA results to return may be too long for many patients to wait to start a trial if detected baseline ctDNA were an inclusion criterion. As such, we recommend including patients with non-detected ctDNA in the clinical trial. The meta-analysis plan should include an approach to how these patients' data will be considered (e.g., as a stratification factor).

Prior anti-cancer therapies may impact baseline ctDNA values. When describing patients' baseline measures, it is important to report the history of prior therapy and the time since previous line of therapy. A minimum washout period before ctDNA analysis is unclear but should be accounted for and has the potential to be analyzed in meta-analyses.

## Calculating Molecular Response

Approaches to defining molecular response are evolving and will be finalized prior to undertaking the formal validation meta-analysis. As such, several characteristics might be required for defining response and sponsors will be asked to ensure all necessary data are prospectively collected.

While it is agreed that change in ctDNA from a baseline to on-treatment should define a molecular response, the approach to calculating change is yet to be determined. Most commonly, a change is calculated as a percent change, which in the context of an aggregate analysis, accounts for assay differences making it more poolable. Currently, a greater than 50% or 90% decrease in ctDNA levels or ctDNA clearance is used to determine a molecular responder.[5, 26, 27] One concern with this approach is that patients with small VAFs at baseline (i.e., <1.0%) may have large percent changes as VAF values become exponentially smaller, but these changes may not translate to biologically or clinically relevant differences. Another potential concern is with the reliability of the numerical results at low VAF. As such, a proposed method for calculating molecular response is to use absolute change–though it is unclear how to apply this strategy in the context of an aggregate analysis without comparability across assays. A third proposed method is to consider clearance of ctDNA (i.e., ctDNA that becomes non-detected on treatment). Again, assay differences, including variations in sensitivity, may influence results and contribute to differences in detection levels and this may overly limit the population of patients who qualify as molecular responders.

We recommend that sponsors use a percent change for studies that assess aNSCLC treated with IO given past work, but if a considerable number of patients have ctDNA clearance on treatment, a clearance cutoff could be considered in the context of a clearly documented assay LoD. The meta-

analysis should consider >50% decrease, >90% decrease, and clearance as three approaches to calculate a molecular responder. The primary endpoint for validation will be defined prior to conducting the meta-analysis. For each patient, it is important to record the precise volume of input plasma used for cfDNA extraction, the total cfDNA extracted from the plasma, the tumor fraction estimate, a measure of the error range/ confidence interval around the estimate, and the amount of cfDNA input into DNA sequencing library preparation. These values can support recalculation of ctDNA output and ctDNA change metrics as needed to ensure consistency.

### Comparing Treatment Groups Using Molecular Response

While there is robust evidence demonstrating associations between change in ctDNA levels and outcomes, few studies focused on comparing two trial arms to determine superiority using ctDNA data. While this is something that can be further explored in meta-analyses, it is important to consider what data should be collected in prospective analyses to ensure meaningful results.

There are two main approaches for determining whether one group has a better molecular response over another: 1) defining a cutoff (e.g., >50% decrease) as a "molecular responder" then calculating a molecular response rate (i.e., percentage of patients who are molecular responders) and/or 2) determining the depth of response and comparing whether one arm has a "deeper" response compared to another (i.e., greater reduction in ctDNA levels). Duration of response is another important element for evaluating molecular response, which measures how long ctDNA levels remain reduced. While FDA does not consider durable clinical benefit a meaningful endpoint on its own, it remains a key component of RECIST-based evaluations. Therefore, we recommend the definition of molecular response focus on change in ctDNA at earlier timepoints; however, understanding the length of response is important and, as described above, collecting later blood measurements (i.e., 6 months) as well as at progression is recommended. Considered as a package, this enables both the primary goal of the validation of an early endpoint based molecular response and the additional secondary goal of assessment of the evolution of ctDNA mid- to long-term post initiation of treatment.

## Conclusions

We have outlined a variety of considerations for data collection in prospective clinical trials assessing aNSCLC treated with IO and incorporating ctDNA. The most critical recommendations for alignment of study design and data collection are:

- Collect a baseline ctDNA measurement before treatment initiation, preferably on C1D1 before infusion.
- Collect four on treatment measurements: three during subsequent treatment cycles (i.e., C2D1, C3D1, and C4D1) and one at 6-months post-treatment initiation; align with RECIST measurements as is feasible.
- Use an assay that is sensitive enough to detect ctDNA in at least 70% of patients at baseline.

- Report all relevant information as outlined in Table 1.

As sponsors plan and execute clinical trials assessing aNSCLC treated with IO, there is an opportunity to prospectively incorporate ctDNA with an aligned approach. We focused on aNSCLC treated with IO, however, principles outlined here could be considered for other advanced cancer types and treatment modalities. Trials focused on aNSCLC treated with IO, if using an aligned approach, can support the development and implementation of a meta-analysis plan that can assess how change in ctDNA levels associate with OS in an aggregate trial-level manner. These analyses can lay the groundwork for using ctDNA as an intermediate endpoint to ensure more rapid availability of safe and effective drugs to patients with cancer.

## Citations


1. Use of Circulating Tumor DNA for Early Stage Solid Tumor Drug Development Guidance for Industry. Available from: https://www.fda.gov/media/158072/download
2. Meeting of the Oncologic Drugs Advisory Committee - 07/25/2024 | FDA. Available from: https://www.fda.gov/advisory-committees/updated-agenda-information-july-25-2024-meeting-oncologic-drugs-advisory-committee-meeting
3. Vellanki PJ, Ghosh S, Pathak A, et al: Regulatory Implications of ctDNA in Immuno-oncology for Solid Tumors. J Immunother Cancer 11:5344, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36796877/
4. Stewart MD, Anagnostou V: Liquid Biopsies Coming of Age: Biology, Emerging Technologies, and Clinical Translation- An Introduction to the JITC Expert Opinion Special Review Series on Liquid Biopsies. J Immunother Cancer 11, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36657817/
5. Thompson JC, Scholes DG, Carpenter EL, et al: Molecular Response Assessment Using Circulating Tumor DNA (ctDNA) in Advanced Solid Tumors. Br J Cancer 129:1893–1902, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/37789101/
6. Stutheit-Zhao EY, Sanz-Garcia E, Liu Z, et al: Early Changes in Tumor-Naive Cell-Free Methylomes and Fragmentomes Predict Outcomes in Pembrolizumab-Treated Solid Tumors. Cancer Discov 14:1048–1063, 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38393391/
7. Adalsteinsson VA, Ha G, Freeman SS, et al: Scalable Whole-Exome Sequencing of Cell-Free DNA Reveals High Concordance with Metastatic Tumors. Nat Commun 8, 2017. Available from: https://pubmed.ncbi.nlm.nih.gov/29109393/
8. Fonseca NM, Maurice-Dror C, Herberts C, et al: Prediction of Plasma ctDNA Fraction and Prognostic Implications of Liquid Biopsy in Advanced Prostate Cancer. Nat Commun 15, 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38418825/
9. Tolmeijer SH, Boerrigter E, Sumiyoshi T, et al: Early On-treatment Changes in Circulating Tumor DNA Fraction and Response to Enzalutamide or Abiraterone in Metastatic Castration-Resistant Prostate Cancer. Clin Cancer Res 29:2835–2844, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36996325/
10. Soo RA, Martini JF, van der Wekken AJ, et al: Early Circulating Tumor DNA Dynamics and Efficacy of Lorlatinib in Patients with Treatment-Naive, Advanced, ALK-Positive NSCLC. J Thorac Oncol 18:1568–1580, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/37295609/
11. Lockwood CM, Merker JD, Bain E, et al: Towards Preanalytical Best Practices for Liquid Biopsy Studies: A BLOODPAC Landscape Analysis. Clin Pharmacol Ther, 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39164947/
12. Godsey JH, Silvestro A, Barrett JC, et al: Generic Protocols for the Analytical Validation of Next-Generation Sequencing-Based ctDNA Assays: A Joint Consensus Recommendation of the BloodPAC's Analytical Variables Working Group. Clin Chem 66:1156–1166, 2020. Available from: https://pubmed.ncbi.nlm.nih.gov/32870995/
13. Abbosh C, Swanton C, Birkbak NJ: Clonal Haematopoiesis: A Source of Biological Noise in Cell-Free DNA Analyses. Ann Oncol 30:358–359, 2019. Available from: https://pubmed.ncbi.nlm.nih.gov/30649226/

14. Liu J, Chen X, Wang J, et al: Biological Background of the Genomic Variations of cf-DNA in Healthy Individuals. Ann Oncol 30:464–470, 2019. Available from: https://pubmed.ncbi.nlm.nih.gov/30475948/

15. Chan HT, Chin YM, Nakamura Y, et al: Clonal Hematopoiesis in Liquid Biopsy: From Biological Noise to Valuable Clinical Implications. Cancers (Basel) 12:1–18, 2020. Available from: https://pubmed.ncbi.nlm.nih.gov/32823942/

16. Assaf ZJF, Zou W, Fine AD, et al: A Longitudinal Circulating Tumor DNA-based Model Associated with Survival in Metastatic Non-Small-Cell Lung Cancer. Nature Medicine 29:859–868, 2023. Available from: https://www.nature.com/articles/s41591-023-02226-6

17. Gray JE, Ahn MJ, Oxnard GR, et al: Early Clearance of Plasma Epidermal Growth Factor Receptor Mutations as a Predictor of Outcome on Osimertinib in Advanced Non-Small Cell Lung Cancer; Exploratory Analysis from AURA3 and FLAURA. Clin Cancer Res 29:3340–3351, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/37379430/

18. Vega DM, Markovets A, Abbosh C, et al: Abstract 1052: Background Variability in Plasma ctDNA Levels in Patients with Advanced Lung Cancer in the Absence of Treatment. Cancer Res 83:1052–1052, 2023. Available from: https://aacrjournals.org/cancerres/article/83/7_Supplement/1052/721446/Abstract-1052-Background-variability-in-plasma

19. 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 6:e2100372, 2022 Available from: http://www.ncbi.nlm.nih.gov/pubmed/35952319

20. Eisenhauer EA, Therasse P, Bogaerts J, et al: New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). Eur J Cancer 45:228–247, 2009. Available from: https://pubmed.ncbi.nlm.nih.gov/19097774/

21. Ola Landgren C, and Devlin S: Evaluating Minimal Residual Disease as an Intermediate Clinical Endpoint for Multiple Myeloma: The EVIDENCE Meta-Analysis Presentation at the April 12, 2024, Meeting of the Oncologic Drugs Advisory Committee Meeting. Available from: https://www.fda.gov/media/177737/download

22. Kok PS, Cho D, Yoon WH, et al: Validation of Progression-Free Survival Rate at 6 Months and Objective Response for Estimating Overall Survival in Immune Checkpoint Inhibitor Trials: A Systematic Review and Meta-analysis. JAMA Netw Open 3:e2011809, 2020. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7489825/

23. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al: Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med 378:2078–2092, 2018. Available from: https://pubmed.ncbi.nlm.nih.gov/29658856/

24. Reichert ZR, Morgan TM, Li G, et al: Prognostic Value of Plasma Circulating Tumor DNA Fraction Across Four Common Cancer Types: A Real-World Outcomes Study. Ann Oncol 34:111–120, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36208697/

25. Zhang Q, Luo J, Wu S, et al: Prognostic and Predictive Impact of Circulating Tumor DNA in Patients with Advanced Cancers Treated with Immune Checkpoint Blockade. Cancer Discov 10:1842–1853, 2020. Available from: https://pubmed.ncbi.nlm.nih.gov/32816849/

26. Sivapalan L, Murray JC, Canzoniero JV, et al: Liquid Biopsy Approaches to Capture Tumor Evolution and Clinical Outcomes During Cancer Immunotherapy. J Immunother Cancer 11, 2023. Available from: https://pubmed.ncbi.nlm.nih.gov/36657818/

27. Anagnostou V, Ho C, Nicholas G, et al: ctDNA Response After Pembrolizumab in Non-Small Cell Lung Cancer: Phase 2 Adaptive Trial Results. Nature Medicine 29:2559-2569, 2023. Available from: https://www.nature.com/articles/s41591-023-02598-9

# Appendix: Proposed Meta-Analysis Strategy for Validation of ctDNA as an Intermediate Endpoint in aNSCLC Treated with Immunotherapy and Summary of Additional Substudies to Consider

## Background

As we develop recommendations for data collection in prospectively designed trials, it is critical that the approach to conduct the meta-analysis be considered concurrently. Although a fully detailed meta-analysis protocol is outside of the scope of this white paper, we felt it was important to provide considerations. Finalizing the analysis plan will require extensive feedback from different statisticians, including statisticians from FDA's Center for Drug Evaluation and Research and from regulators to ensure that the results would be sufficient for the validation of ctDNA as an intermediate endpoint. Herein, we provide considerations for a meta-analysis that could be used to aggregate randomized controlled trials that assess immunotherapy treatment in aNSCLC based on ctDNA incorporated to future trials following the recommendations provided in the white paper.

This concept analysis plan provides considerations for potential statistical methods for trial- and patient-level metanalyses to validate change in ctDNA levels as an intermediate endpoint. The proposal serves to validate the trial design considerations discussed in the white paper, as well as identify the key considerations to establish an analysis plan. As a guiding principle for the primary goal, we consider the adoption of the simplest binary scenario for molecular response (MR). The primary analysis in this concept analysis plan would use a percent change cutoff at a single early timepoint for assessing associations with overall survival. The collaboration participants and regulators will agree on both the cutoff and the timepoint prior to conducting the analysis. Secondary objectives will include other cutoffs and timepoints. An initial list of sensitivity analyses is listed and will be prioritized as part of the final analysis plan. Additional substudies are also provided in summary form. These may be promoted to secondary analyses as part of the finalization process.

Trials to consider for the meta-analysis must be randomized controlled trials that meet assay and clinical specifications described throughout the white paper. Studies should be included whether or not they show a treatment effect on overall survival. Trial selection should be transparent and unbiased (i.e., based on trial quality, relevance, and consistency rather than outcome driven).

## Criteria to Establish for Study Inclusion in the Analysis

- Minimum number of patients per arm
- Minimum number of patients per arm have a MR
- Minimum number of pairwise comparisons to support the study level analysis
- If survival follow-up is ongoing at the time of data cut-off, determine a minimum degree of maturity

## Data Collection

### *ctDNA Timing*

- Baseline measurement (relative to treatment initiation, days)
- Each on-treatment measurement (relative to treatment initiation, days)

### *Patient Characteristics*

- Age
- Sex
- Race
- Smoking status
- Stage (advanced stave IV vs else)
- ECOG Performance Status
- Number of prior lines of therapy
- Histology
- PD-L1 expression
- Others to be pre-defined

### *Clinical Characteristics*

- Overall survival
- Progression free survival
- Confirmed (Yes/No)
- BICR used (Yes/No)
- Radiographic measurements throughout the study (i.e., RECIST categories, sum of diameter calculations, timing for RECIST measurements relative to treatment initiation in days)

### *Assay Characteristics*

- Limit of detection (LoD)
- Percent of patients with detected baseline ctDNA
- Clonal hematopoiesis (CH) removal approach
- Sample volume
    - Serum at blood draw
    - Input volume for ctDNA assay
- Performance Parameters
    - Limit of the blank (if applicable)
    - Precision
    - Accuracy
    - Sensitivity/ Specificity
    - Pre-analytical approach including guidelines followed

## Descriptive Analyses

Various tabular and graphical summaries to describe:

- Study design features: sample size, arms under study, patient characteristics, median duration of follow up
- Assay description
    - Limit of detection
    - Percent of patients with detected baseline ctDNA
    - Sample volume
- ctDNA data completeness
- ctDNA distribution at baseline and primary timepoint
- OS summary statistics
- PFS and RECIST summary statistics

## Primary Endpoints for Evaluation of ctDNA as an Intermediate Endpoint

OS is the clinical outcome of interest analyzed as a time to event variable. KM estimate at 2 or 3 years will be a secondary or sensitivity analysis. Molecular response will be defined prior to the conduct of the analysis as described above and is denoted as MR below.

## Primary Analysis for Individual Patient Level Assessment

Degree of separation between MR and nMR relative to OS in a Cox Proportional hazard model accounting for all relevant patient-level covariates, MR (y/n), treatment, treatment by MR interactions (depending on the treatments included in the studies) and a stratification term for study. The primary endpoint will be the HR for MR relative to nMR, the confidence interval and the p-value serve to assess the strength of evidence. A specific threshold should be developed ahead of time (e.g., HR at least 0.7 or better and 95% CI excludes 1).

## Secondary Analyses for Individual Patient Level Assessment

Secondary analyses will explore the relationship between various cutpoints and time points of molecular response and overall survival, while sensitivity analyses will assess the robustness of these findings by evaluating different patient subgroups, assay types, and MR thresholds.

## Primary Analysis for Trial Level Assessment

Weighted linear regression model of log HR OS (test treatment vs. control in each study) vs. Log odds ratio of MR to nMR

HR OS based on proportional Cox hazard with adjustment for covariates as described above.

The regression will be weighted by the inverse variances of the log odds ratio for log OR MR.

The linear regression may include additional terms for covariates such as study and paired types of treatment-control pairs.

Study-level association metrics of $R^2$ and associated confidence interval will be calculated. Criteria will be pre-specified (e.g., $R^2$ at least 0.7 and the lower end of the confidence interval is above 0.5).

### Secondary Analyses for Trial Level Assessment

Secondary analyses will explore the relationship between various definitions and time points of molecular response and overall survival, while sensitivity analyses will assess the robustness of these findings by additional statistical model (e.g., weighting by study size) and as above, evaluating different patient subgroups, assay types, and variant allele frequency thresholds.

### Substudies

### Additional substudies may be valuable as part of a supportive package for the main analysis to validate ctDNA as an intermediate endpoint. An initial list is provided below:

| Aspect Evaluated | Description |
|---|---|
| Baseline ctDNA as a prognostic factor | Include assessment of stratification in tertiles or quintiles of outcomes by baseline values of ctDNA. |
| Time course of ctDNA: depth and duration of response | Enables evaluation of response over time, which may be key to future development of more refined tools/definition in aNSCLC and in other settings.<br>If sufficient data are provided, we can further evaluate any 'lead time' in ctDNA to identify when molecular progression occurs. |
| Reproduce the validation for PFS to OS assessment and/or an earlier RECIST based assessment of response | Serves as context setting for the results achieved with ctDNA. |
| Value of ctDNA beyond other intermediate endpoints such as PFS or ORR | Addresses how ctDNA can be positioned alongside other intermediate endpoints (e.g., Is molecular response more or less predictive than PFS or other definitions of RECIST relative to OS? Does MR have a role in further stratifying patients with stable disease?). |

# Advancing Evidence Generation for Circulating Tumor DNA: Lessons Learned from A Multi-Assay Study of Baseline Circulating Tumor DNA Levels across Cancer Types and Stages

**Brittany A. McKelvey [1,*], Hillary S. Andrews [1], Frederick L. Baehner [2], James Chen [3], Carin R. Espenschied [4], David Fabrizio [5], Vanessa Gorton [6], Claire Gould [7], Justin Guinney [3], Greg Jones [8], Xiangyang Lv [9], Michael S. Nahorski [8], Melanie R. Palomares [2], Gary A. Pestano [7], Mark Sausen [10], Alain Silk [3], Nicole Zhang [4], Zhihong Zhang [9], Mark D. Stewart [1] and Jeff D. Allen [1]**

1 Friends of Cancer Research, Washington, DC 20036, USA
2 Exact Sciences Corp., Madison, WI 53719, USA
3 Tempus AI, Inc., Chicago, IL 60654, USA
4 Guardant Health, Inc., Redwood City, CA 94063, USA
5 Foundation Medicine, Inc., Cambridge, MA 02141, USA
6 Predicine, Hayward, CA 94545, USA
7 Biodesix Inc., Louisville, CO 80027, USA
8 NeoGenomics Laboratories, Fort Myers, FL 33912, USA
9 Burning Rock, Irvine, CA 92617, USA
10 Personal Genome Diagnostics, Labcorp, Baltimore, MD 21224, USA
* Correspondence: bmckelvey@focr.org; Tel.: +1-202-944-6717

**Abstract:** Circulating tumor DNA (ctDNA) holds promise as a biomarker for predicting clinical responses to therapy in solid tumors, and multiple ctDNA assays are in development. However, the heterogeneity in ctDNA levels prior to treatment (baseline) across different cancer types and stages and across ctDNA assays has not been widely studied. Friends of Cancer Research formed a collaboration across multiple commercial ctDNA assay developers to assess baseline ctDNA levels across five cancer types in early- and late-stage disease. This retrospective study included eight commercial ctDNA assay developers providing summary-level de-identified data for patients with non-small cell lung cancer (NSCLC), bladder, breast, prostate, and head and neck squamous cell carcinoma following a common analysis protocol. Baseline ctDNA levels across late-stage cancer types were similarly detected, highlighting the potential use of ctDNA as a biomarker in these cancer types. Variability was observed in ctDNA levels across assays in early-stage NSCLC, indicative of the contribution of assay analytical performance and methodology on variability. We identified key data elements, including assay characteristics and clinicopathological metadata, that need to be standardized for future meta-analyses across multiple assays. This work facilitates evidence generation opportunities to support the use of ctDNA as a biomarker for clinical response.

**Keywords:** ctDNA; cancer; biomarker



**Citation:** McKelvey, B.A.; Andrews, H.S.; Baehner, F.L.; Chen, J.; Espenschied, C.R.; Fabrizio, D.; Gorton, V.; Gould, C.; Guinney, J.; Jones, G.; et al. Advancing Evidence Generation for Circulating Tumor DNA: Lessons Learned from A Multi-Assay Study of Baseline Circulating Tumor DNA Levels across Cancer Types and Stages. *Diagnostics* **2024**, *14*, 912. https://doi.org/10.3390/diagnostics14090912

Academic Editor: Joaquin Cubiella

Received: 14 March 2024
Revised: 16 April 2024
Accepted: 25 April 2024
Published: 27 April 2024



## 1. Introduction

The measurement of circulating tumor DNA (ctDNA) has emerged as a promising surrogate for disease burden and, by extension, a research tool to rapidly evaluate clinical response across a myriad of therapeutic interventions. Emerging data continue to build momentum around the various clinical and regulatory applications of ctDNA in oncology, including predicting a patient's response to therapy [1–5]. The use of ctDNA to predict clinical response could enable faster identification and development of more effective drugs and, importantly, support regulatory decision-making as an early endpoint predicting long-term clinical outcomes [6–9]. Early endpoints that are "reasonably likely to predict a clinical benefit" are increasingly important in oncology drug development to shorten

development timelines and get effective drugs to patients faster [10]. The U.S. Food and Drug Administration's (FDA) Draft Guidance on the Use of Circulating Tumor DNA for Early-Stage Solid Tumor Drug Development highlights the use of ctDNA as an early endpoint in clinical trials; however, it also states that further data are needed to support its use [11].

Although advancements in technologies are leading to more sensitive and precise tools for detecting and measuring ctDNA, all technologies have inherent limitations and variability [12]. Further, ctDNA may not be detected at sufficient levels to allow informative analysis across all cancer types and stages. Thus, it is important to understand the extent to which heterogeneity in ctDNA levels across different cancer types and stages stems from tumor-specific factors, such as tumor shed rates, and technical factors, such as the dynamic range of the assay for interpreting ctDNA measurement. Several efforts have assessed the landscape of ctDNA detection across cancer types in large real-world evidence cohorts [13–15]. However, these data are specific to a single technology, laboratory, or assay and are focused largely in the advanced or metastatic setting where tumor biology may be fundamentally different from earlier-stage cancer in which the application of ctDNA as an early endpoint may be especially valuable. To evaluate the technical and biological variability across cancer types and assays, a multi-assay study was conducted to investigate baseline ctDNA levels (ctDNA levels prior to current cancer treatment) in multiple cancer types and stages. We generated descriptive statistics to compare trends in baseline ctDNA levels across assays by cancer type and stage through a collaborative effort with multiple commercial assay developers. While informative, our findings identified key considerations required to support broad data harmonization efforts to generate evidence for the use of ctDNA as an early endpoint across assays and clinical settings.

## 2. Materials and Methods

Each assay developer retrospectively aggregated data from their database following a common data analysis protocol, which specified data elements and analyses to generate summary-level statistics across five cancer types (see Supplementary Materials, Tables S1–S3), with each assay dataset defined as a cohort. Patients included in this analysis were adult patients, aged 18 or older at the date of ctDNA sample collection, diagnosed with cancer, and had either not yet initiated anti-cancer therapy or had not received anti-cancer therapy at the time of baseline sampling (see Supplementary Materials, Section S3). Non-small cell lung cancer (NSCLC), bladder, breast, prostate, and head and neck squamous cell carcinoma (HNSCC) cancers were analyzed due to the availability of baseline ctDNA data from at least two assay developers. Patients were included if they had known early- or late-stage cancer at the time of baseline sampling. Summary-level clinical and demographic characteristics were reported for each cohort if known.

The pre-analytic cell free DNA (cfDNA) minimal technical data elements (MTDEs) [16] proposed by the Blood Profiling Atlas in Cancer (BloodPAC) Consortium were used to ensure that pre-analytical variability was similarly controlled across cohorts to reduce the impact of pre-analytical factors. Assay characteristics were reported and aggregated across developers. No patient-level identifiers and, thus, no protected health information were revealed or exchanged in this process.

Summary-level data on baseline ctDNA levels for specific cancer types and stages were reported by cohort. Following the ctDNA to Monitor Treatment Response (ctMoniTR) project [9], summary-level statistics of sample size, median, mean, standard deviation (SD), Interquartile Range (IQR), minimum and maximum for each of the median variant allele frequency (VAF), maximum VAF, and mean VAF were reported for baseline ctDNA levels. Descriptive statistics were used.

## 3. Results

### *3.1. Assay Characteristics*

Eight commercial assays measuring baseline ctDNA were blinded and included in the analysis (labeled Cohort A-I). Five assays (62.5%) were tumor-informed (i.e., mutations identified in the primary tumor tissue that are tracked in the plasma), and three (37.5%) were tumor-naïve (i.e., mutations were detected de novo from the plasma). All but one assay (87.5%) used next-generation sequencing (NGS); the remaining assay used droplet digital PCR (ddPCR). Half (4/8) of the assays did not conduct clonal hematopoiesis of indeterminate potential (CHIP) filtering, three (37.5%) used bioinformatic methods, and one (12.5%) used germline sequencing methods to filter for CHIP variants. All assays assessed single nucleotide variants (SNVs) with a median limit of detection (LOD) of 0.2% VAF (range, 0.0011–0.5%).

### *3.2. Sample Characteristics*

Across the eight cohorts, data from early- and late-stage samples were provided for NSCLC, with 2357 early-stage and 62,994 late-stage samples and 87,209 total samples across all five late-stage cancer types (Table 1). Most cohorts did not have data available for AJCC staging, prior anti-cancer treatments, recurrence or progression status, and the type of recurrence. The timing of ctDNA sampling relative to diagnosis varied across cohorts, with long durations observed in late-stage cancers.

**Table 1.** Clinicopathological characteristics and baseline ctDNA VAFs by assay cohort [1].

| A: Early- and Late-Stage Non-Small Cell Lung Cancer (NSCLC) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | NSCLC | | | | | | |
| | Early Stage | | | | Late Stage | | | | | | |
| **Cohort** | **A** | **B** | **D** | **E** | **I** | **A** | **B** | **C** | **D** | **F** | **G** |
| **N** (samples) | 245 | 1873 | 679 | 78 | 131 | 1232 | 31,889 | 23,157 | 2452 | 264 | 4000 |
| **Age** (median, years) | 70 | 70 | 70 | Unkn | 63 | 69 | 67 | 68 | Unkn | 70 | 73 |
| **Gender** (% female) | 48 | 49 | 49 | 49 | 35 | 49 | 50 | 52 | 49 | 52 | 53 |
| **Clinical Stage** | | | | | | | | | | | |
| I | 5 | 19 | 15 | 53 | 48 | 0 | 0 | 0 | 0 | 0 | 0 |
| II | 2 | 15 | 17 | 28 | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| III | 11 | 29 | 68 | 19 | 35 | 5 | 1 | 0 | 0 | 15 | 0 |
| IV | 0 | 0 | 0 | 0 | 0 | 13 | 7 | 0 | 100 | 82 | 0 |
| Unknown | 82 | 37 | 0 | 0 | 0 | 82 | 92 | 100 | 0 | 3 | 100 |
| **Prior Anti-Cancer Treatments** | | | | | | | | | | | |
| Known Tx | 18 | 21 | 0 | 0 | 0 | 18 | 3 | 0 | 0 | 0 | 0 |
| None | 1 | 42 | 0 | 0 | 100 | 11 | 6 | 0 | 0 | 0 | 0 |
| Unknown | 81 | 37 | 100 | 100 | 0 | 71 | 91 | 100 | 100 | 100 | 100 |
| **Recurrence/Progression Status** | | | | | | | | | | | |
| No prior cancer | 2 | 11 | 0 | 0 | 100 | 0 | 1 | 0 | 0 | 0 | 0 |
| Unknown | 95 | 86 | 100 | 100 | 0 | 97 | 98 | 100 | 100 | 100 | 100 |
| **Timing of Sampling**, days from diagnosis to sampling, (median (IQR)) | 461 (145.5, 891.5) | 40 (16, 134.3) | 22 (13, 36) | <84 | 1 (1, 1) | 602.5 (334, 850.8) | 29 (13, 31.8) | 18 (7, 52) | 17 (9, 36) | <84 | Unkn |
| **Frequency of ctDNA Detected in Samples** (%) | 23.2 | 3.1 | 89.7 | 51.3 | 85.5 | 37.6 | 7.9 | 92.4 | 92.4 | 77.3 | 99.6 |
| **Median VAF** (IQR) | 1.8 (0.7, 2.3) | 1.6 (0.9, 2) | 0.78 (0.32, 1.44) | 0.09 (0.02, 0.42) | 0.001 (0.001, 0.2) | 2.8 (1.5, 4.6) | 2.6 (1.7, 5.8) | 0.8 (0.5, 2.2) | 1.33 (0.75, 3.47) | 3.49 (2.67, 8.21) | 1.3 (0.6, 3.2) |

**Table 1.** *Cont.*

| B: Late-Stage Breast, Bladder, Prostate, and HNSCC Cancers | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Late Stage | | | | | | | | | | | |
| | Breast | | | Bladder | | | Prostate | | | HNSCC | | |
| **Cohort** | **C** | **D** | **G** | **C** | **D** | **G** | **C** | **D** | **G** | **C** | **D** | **G** |
| **N** (samples) | 2572 | 1020 | 6940 | 500 | 282 | 577 | 1100 | 633 | 9502 | 274 | 136 | 546 |
| **Age** (median, years) | 62 | 61 | 64 | 72 | 71 | 73 | 70 | 68 | 74 | 64 | 62 | 64 |
| **Gender** (% female) | 98 | 100 | 99 | 29 | 26 | 25 | 0 | 0 | 0 | 22 | 24 | 23 |
| **Clinical Stage** | | | | | | | | | | | | |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| II | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| III | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| IV | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 |
| Unknown | 100 | 0 | 100 | 100 | 0 | 100 | 100 | 0 | 100 | 100 | 0 | 100 |
| **Prior Anti-Cancer Treatments** | | | | | | | | | | | | |
| Known Tx | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Unknown | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **Recurrence/Progression Status** | | | | | | | | | | | | |
| No prior cancer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Unknown | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **Timing of Sampling**, days from diagnosis to Ssmpling, (median (IQR)) | 262 (16, 1220) | 35 (20, 75) | Unkn | 70 (11, 557) | 241 (35, 741) | Unkn | 126.5 (9, 1130.5) | 42 (21, 1255) | Unkn | 97.5 (9, 728.5) | 34 (24, 135) | Unkn |
| **Frequency of ctDNA Detected in Samples** (%) | 89.6 | 92.1 | 99.5 | 93 | 90.1 | 99.3 | 84.5 | 86.3 | 99.6 | 88.3 | 87.5 | 99.4 |
| **Median VAF** (IQR) | 0.8 (0.5, 3.4) | 1.35 (0.79, 5.19) | 1.4 (0.73, 5.22) | 0.9 (0.6, 2.5) | 2.15 (1.44, 6.36) | 1.5 (0.8, 3.4) | 0.6 (0.35, 2.2) | 1.06 (0.56, 6.14) | 1.4 (0.7, 4.8) | 0.7 (0.4, 1.2) | 0.78 (0.26, 1.53) | 1.2 (0.58, 4.19) |

[1] Tx: treatment, Unkn: unknown, IQR: Interquartile Range. Cohorts in red are tumor-informed assays, and cohorts in black are tumor-naïve assays. Median VAF—the median of VAF values from all somatic tumor-derived variants.

### 3.3. Baseline ctDNA Levels

In comparing early- versus late-stage NSCLC, the frequency of ctDNA detection varied across cohorts, with late-stage NSCLC having a higher proportion of samples with detected ctDNA than early-stage in data from assays that had both early- and late-stage data available (Table 1, Figure 1). For those samples with detected ctDNA, late-stage NSCLC samples generally appeared to have higher levels as compared to early-stage samples, with cohort variability observed. Across the late-stage cancer types evaluated, baseline ctDNA was similarly detected across most samples across cohorts (Table 1, Figure 2). For the three assays with data available across all five late-stage cancer types, baseline ctDNA levels were similar across cancer types and assays.

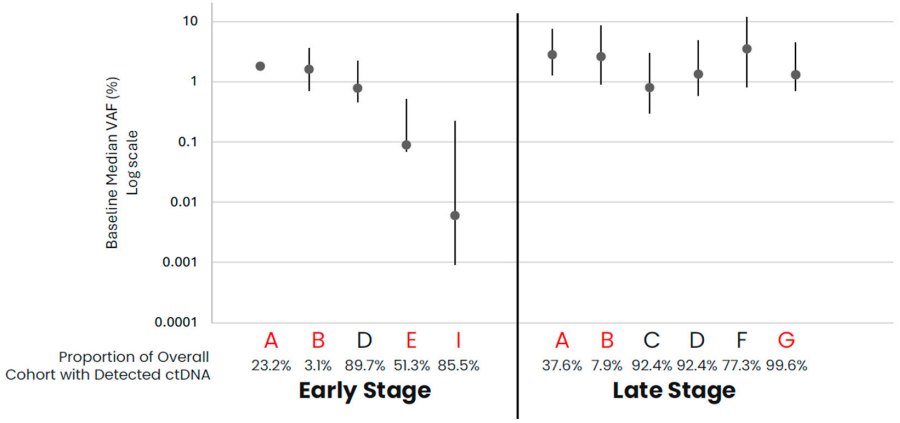



**Figure 1.** NSCLC baseline ctDNA levels for samples with detected ctDNA. Median VAF (IQR) ctDNA levels for samples with detected ctDNA by cohort, with the proportion of total cohort samples with detected ctDNA shown below the graph. Cohorts in red are tumor-informed assays, and cohorts in black are tumor-naïve assays. Median VAF—the median of VAF values from all somatic tumor-derived variants.

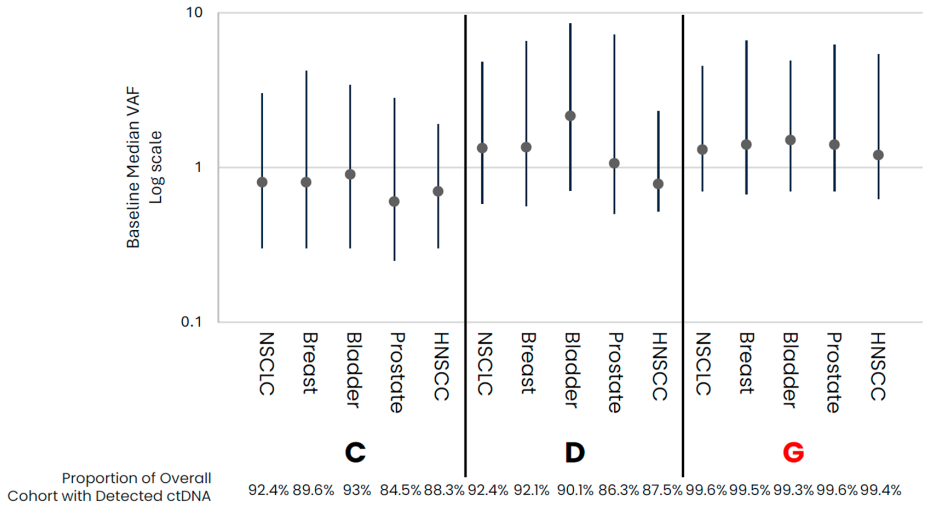



**Figure 2.** Late-Stage baseline ctDNA levels for samples with detected ctDNA. Median (IQR) VAF ctDNA levels for samples with detected ctDNA by cohort, with the proportion of total cohort samples with detected ctDNA shown below the graph. Colored points highlight the different cancer types. Cohorts in red are tumor-informed assays, and cohorts in black are tumor-naïve assays. Median VAF—the median of VAF values from all somatic tumor-derived variants.

## 4. Discussion

This collaborative effort evaluated baseline ctDNA levels by cancer type and stage across different assays to identify overall trends and considerations to support future data harmonization efforts to generate evidence for the use of ctDNA as an early endpoint. Overall, baseline ctDNA levels across late-stage NSCLC, breast, bladder, prostate, and HNSCC cancers were similarly detected, suggesting the potential opportunity to use ctDNA as a clinical biomarker in these cancer types. Conversely, more variability in ctDNA levels across assays was observed in early-stage NSCLC than in late-stage disease, highlighting the critical need to consider factors such as assay analytical performance and methodology for evaluating ctDNA in this setting [17].

Assay characteristics, including the intended use, features assessed, and analytical performance, were variable, leading to difficulties in interpreting aggregated data. The development of common data standards could help allow more robust comparisons across assay datasets [18]. The heterogeneity in approaches to identifying SNVs (e.g., tumor-informed or naïve) and CHIP filtering can cause variability between assays for samples determined to have detected ctDNA. For example, our study explored mean, median, and maximum VAF (median reported herein) and observed biases in mean and maximum VAF values in some cohorts due to conflation by high VAF values derived from suspected germline variants. However, median VAF may also misrepresent data when ctDNA levels are low (e.g., in the stochastic range) and bias against the lower range of detection. Therefore, setting standards for how ctDNA levels are reported across assays as well as a clear understanding of the methodology for obtaining ctDNA values are critical.

Real-world data are a valuable source of data for analyses but provide challenges in meta-analyses due to data missingness and heterogeneity [19]. The availability of clinicopathological data was generally lacking across cohorts in this study. Each developer could confidently categorize their samples as either early- or late-stage disease. Many could not provide the AJCC clinical staging, which may impact observed ctDNA levels given differences in tumor shedding by stage, and data on prior anti-cancer treatments and recurrence or progression status were mostly unknown. The lack of available clinical data was not surprising given that assay developers included in this analysis were clinical laboratories providing testing as a service to health systems and may not have routine access to comprehensive clinical data for each sample tested. However, an understanding of prior treatment is critical to define baselines, as samples may be included from patients who are treatment-naïve, as well as patients who have received prior anti-cancer treatment and subsequently recurred or progressed. Due to unknown clinicopathological factors, treatment or surgical intervention status, and sample collection timing from diagnosis, significant cohort heterogeneity may complicate comparisons across cohorts.

The timing from diagnosis to sampling was heterogeneous, especially in late-stage cancers, which could be affected by the intended use of the test when ctDNA analysis is conducted during the patient journey. This variability, along with other anti-cancer treatments or modalities that could impact ctDNA levels, highlights the importance of defining minimal criteria for the length of time between diagnosis and sampling. This may potentially avoid variability surrounding long timeframes. As a result, it is important to identify and standardize key data elements, including assay characteristics and clinicopathological data, to facilitate robust evidence generation to support the use of ctDNA as an early endpoint, leading to more harmonized and effective use of ctDNA in future clinical research and care.

## 5. Conclusions

To support the future use of ctDNA as an early endpoint, meta-analyses across assays, supported by appropriate clinicopathological metadata, are needed for multiple cancer types and stages. This collaborative effort has enabled the evaluation of baseline ctDNA levels by cancer type and stage across different assays to identify overall trends and considerations. This effort supports future data harmonization efforts to validate the use

of ctDNA as an early endpoint, highlighting the potential opportunity to use ctDNA as a clinical biomarker in late-stage NSCLC, breast, bladder, prostate, and HNSCC cancers due to the similar detection of baseline ctDNA levels across these cancer types. However, more variability in ctDNA levels across assays was observed in early-stage NSCLC than in late-stage types, underscoring the importance of evaluating factors such as assay analytical performance and methodology in this setting.

Given the heterogeneity of data from real-world sources, routine collection and analysis of ctDNA from patients in oncology clinical trials may provide more comprehensive and standardized clinical data and assure within-cohort control over technical variability. The development of common data standards and an understanding of assay technological features and key performance characteristics can improve the poolability of data generated using different assays. The learnings from this study, such as the need to address the heterogeneity in approaches to identifying SNVs and the challenges posed by assay characteristic variability, underscore the complexity of interpreting aggregated data and the importance of developing methodological approaches to combine data from different trials and assays. These highlighted data needs can facilitate future pooled analyses to generate robust evidence to support the use of ctDNA as a biomarker and early endpoint, setting the stage for a more harmonized and effective approach to oncology drug development and patient care.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/diagnostics14090912/s1: Supplementary Protocol: Introduction, Objective, Study Cohort, Data Collection, Statistical Considerations, Data Dictionary and Table S1: Study variable definitions; Table S2: Pre-analytic technical specification elements; Table S3: ctDNA results.

**Author Contributions:** Conceptualization, B.A.M., H.S.A., M.D.S. and J.D.A.; methodology, B.A.M., H.S.A., F.L.B., J.C., C.R.E., D.F., V.G., C.G., J.G., G.J., X.L., M.S.N., M.R.P., G.A.P., M.S., A.S., N.Z., Z.Z., M.D.S. and J.D.A.; formal analysis, B.A.M., H.S.A. and M.D.S.; resources, M.D.S. and J.D.A.; data curation, B.A.M. and M.D.S.; writing—original draft preparation, B.A.M.; writing—review and editing, B.A.M., H.S.A., F.L.B., J.C., C.R.E., D.F., V.G., C.G., J.G., G.J., X.L., M.S.N., M.R.P., G.A.P., M.S., A.S., N.Z., Z.Z., M.D.S. and J.D.A.; visualization, B.A.M., H.S.A., F.L.B., J.C., C.R.E., D.F., V.G., C.G., J.G., G.J., X.L., M.S.N., M.R.P., G.A.P., M.S., A.S., N.Z., Z.Z., M.D.S. and J.D.A.; project administration, B.A.M. and M.D.S.; funding acquisition, J.D.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Friends of Cancer Research.

**Institutional Review Board Statement:** This study involved the secondary use of de-identified data, which did not include any patient-level identifiers or protected health information. Accordingly, this research did not involve human subjects as defined under applicable regulations and, thus, was not subject to IRB review.

**Informed Consent Statement:** Patient consent was either waived or not required under 45 CFR 46.104(d)(4)(ii) which defines use of biospecimens in such a manner that the identity of the human subjects cannot readily be ascertained.

**Data Availability Statement:** The patient-level datasets presented in this article are not readily available due to patient privacy and legal restrictions.

**Conflicts of Interest:** F.L.B. is an employee of Exact Sciences; J.C. is an employee of Tempus AI, Inc.; C.R.E. is an employee and stockholder with Guardant Health; D.F. is an employee of Foundation Medicine; V.G. is an employee of Predicine; J.G. is an employee of Tempus AI, Inc.; G.J. is an employee and stockholder with NeoGenomics; X.L. is an employee of Burning Rock; M.N. is employee and shareholder of NeoGenomics; G.A.P. is an employee and holds equity in Biodesix, Inc.; M.S. is an employee of Personal Genome Diagnostics (Labcorp) and holds equity in Labcorp; A.S. is an employee of Tempus AI, Inc.; N.Z. is an employee of Guardant Health; Z.Z. is an employee of Burning Rock; and all other authors report no conflicts of interest.

## References


1. Medina, J.E.; Dracopoli, N.C.; Bach, P.B.; Lau, A.; Scharpf, R.B.; Meijer, G.A.; Andersen, C.L.; Velculescu, V.E. Cell-free DNA approaches for cancer early detection and interception. *J. Immunother. Cancer* **2023**, *11*, e006013. [CrossRef] [PubMed]
2. Semenkovich, N.P.; Szymanski, J.J.; Earland, N.; Chauhan, P.S.; Pellini, B.; Chaudhuri, A.A. Genomic approaches to cancer and minimal residual disease detection using circulating tumor DNA. *J. Immunother. Cancer* **2023**, *11*, e006284. [CrossRef] [PubMed]
3. Vellanki, P.J.; Ghosh, S.; Pathak, A.; Fusco, M.J.; Bloomquist, E.W.; Tang, S.; Singh, H.; Philip, R.; Pazdur, R.; Beaver, J.A. Regulatory implications of ctDNA in immuno-oncology for solid tumors. *J. Immunother. Cancer* **2023**, *11*, e005344. [CrossRef] [PubMed]
4. Sivapalan, L.; Murray, J.C.; Canzoniero, J.V.; Landon, B.; Jackson, J.; Scott, S.; Lam, V.; Levy, B.P.; Sausen, M.; Anagnostou, V. Liquid biopsy approaches to capture tumor evolution and clinical outcomes during cancer immunotherapy. *J. Immunother. Cancer* **2023**, *11*, e005924. [CrossRef] [PubMed]
5. Pantel, K.; Alix-Panabières, C. Crucial roles of circulating tumor cells in the metastatic cascade and tumor immune escape: Biology and clinical translation. *J. Immunother. Cancer* **2022**, *10*, e005615. [CrossRef] [PubMed]
6. Ricciuti, B.; Jones, G.; Severgnini, M.; Alessi, J.V.; Recondo, G.; Lawrence, M.; Forshew, T.; Lydon, C.; Nishino, M.; Cheng, M.; et al. Early plasma circulating tumor DNA (ctDNA) changes predict response to first-line pembrolizumab-based therapy in non-small cell lung cancer (NSCLC). *J. Immunother. Cancer* **2021**, *9*, e001504. [CrossRef] [PubMed]
7. Seremet, T.; Jansen, Y.; Planken, S.; Njimi, H.; Delaunoy, M.; El Housni, H.; Awada, G.; Schwarze, J.K.; Keyaerts, M.; Everaert, H.; et al. Undetectable circulating tumor DNA (ctDNA) levels correlate with favorable outcome in metastatic melanoma patients treated with anti-PD1 therapy. *J. Transl. Med.* **2019**, *17*, 303. [CrossRef] [PubMed]
8. Nabet, B.Y.; Esfahani, M.S.; Moding, E.J.; Hamilton, E.G.; Chabon, J.J.; Rizvi, H.; Steen, C.B.; Chaudhuri, A.A.; Liu, C.L.; Hui, A.B.; et al. Noninvasive Early Identification of Therapeutic Benefit from Immune Checkpoint Inhibition. *Cell* **2020**, *183*, 363–376.e13. [CrossRef] [PubMed]
9. Vega, D.M.; Nishimura, K.K.; Zariffa, N.; Thompson, J.C.; Hoering, A.; Cilento, V.; Rosenthal, A.; Anagnostou, V.; Baden, J.; Beaver, J.A.; 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. [CrossRef] [PubMed]
10. Beaver, J.A.; Howie, L.J.; Pelosof, L.; Kim, T.; Liu, J.; Goldberg, K.B.; Sridhara, R.; Blumenthal, G.M.; Farrell, A.T.; Keegan, P.; et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics. *JAMA Oncol.* **2018**, *4*, 849–856. [CrossRef] [PubMed]
11. Use of Circulating Tumor DNA for Early-Stage Solid Tumor Drug Development Guidance for Industry DRAFT GUIDANCE. Available online: https://www.fda.gov/vaccines-blood-biologics/guidance (accessed on 19 October 2023).
12. Desai, A.; Lovly, C.M. Challenges in the implementation of ultrasensitive liquid biopsy approaches in precision oncology. *J. Immunother. Cancer* **2023**, *11*, e006793. [CrossRef] [PubMed]
13. Zhang, Y.; Yao, Y.; Xu, Y.; Li, L.; Gong, Y.; Zhang, K.; Zhang, M.; Guan, Y.; Chang, L.; Xia, X.; et al. Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat. Commun.* **2021**, *12*, 11. [CrossRef] [PubMed]
14. Zill, O.A.; Banks, K.C.; Fairclough, S.R.; Mortimer, S.A.; Vowles, J.V.; Mokhtari, R.; Gandara, D.R.; Mack, P.C.; Odegaard, J.I.; Nagy, R.J.; et al. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin. Cancer Res.* **2018**, *24*, 3528–3538. [CrossRef] [PubMed]
15. Bettegowda, C.; Sausen, M.; Leary, R.J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B.R.; Wang, H.; Luber, B.; Alani, R.M.; et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **2014**, *6*, 224. [CrossRef]
16. Febbo, P.G.; Martin, A.; Scher, H.I.; Barrett, J.C.; Beaver, J.A.; Beresford, P.J.; Blumenthal, G.M.; Bramlett, K.; Compton, C.; Dittamore, R.; et al. Minimum Technical Data Elements for Liquid Biopsy Data Submitted to Public Databases. *Clin. Pharmacol. Ther.* **2020**, *107*, 730–734. [CrossRef] [PubMed]
17. Pennell, N.A.; Mutebi, A.; Zhou, Z.-Y.; Ricculli, M.L.; Tang, W.; Wang, H.; Guerin, A.; Arnhart, T.; Dalal, A.; Sasane, M.; et al. Economic Impact of Next-Generation Sequencing Versus Single-Gene Testing to Detect Genomic Alterations in Metastatic Non–Small-Cell Lung Cancer Using a Decision Analytic Model. *JCO Precis. Oncol.* **2019**, *3*, 1–9. [CrossRef] [PubMed]
18. Henriksen, T.V.; Reinert, T.; Christensen, E.; Sethi, H.; Birkenkamp-Demtröder, K.; Gögenur, M.; Gögenur, I.; Zimmermann, B.G.; Dyrskjøt, L.; Andersen, C.L.; et al. The effect of surgical trauma on circulating free DNA levels in cancer patients—Implications for studies of circulating tumor DNA. *Mol. Oncol.* **2020**, *14*, 1670–1679. [CrossRef] [PubMed]
19. Lerro, C.C.; Bradley, M.C.; Forshee, R.A.; Rivera, D.R. The Bar Is High: Evaluating Fit-for-Use Oncology Real-World Data for Regulatory Decision Making. *JCO Clin. Cancer Inform.* **2024**, *8*, e2300261. [CrossRef] [PubMed]

# Complex Biomarkers

## Harmonizing Measurement and AI Applications

# Analysis of 20 Independently Performed Assays to Measure Homologous Recombination Deficiency (HRD) in Ovarian Cancer: Findings From the Friends' HRD Harmonization Project

Hillary S. Andrews, PhD[1]; Lisa M. McShane, PhD[2]; Elise C. Kohn, MD[2]; Rebecca Arend, MD[3]; Chris Karlovich, PhD[4]; Kaitlyn Kincaid, MD[3]; A. Douglas Laird, PhD[5]; Ming-Chung Li, PhD[2]; Ethan S. Sokol, PhD[6]; Elizabeth R. Starks, MS[7]; Shulin Bi, MS[8]; Lauren Brunner, PhD[9]; Alyssa Chapman, MS[4]; Li Chen, PhD[4]; Tommaso Coletta, PhD[10]; Yuan Ding, PhD[11]; Bailee D. Oliver, BS[3]; McKenzie Foxall, BS[3]; Mohit Gupta, MS[12]; Zan Halford, MS[13]; Andrea G. Kahn, MD[3]; Nikita Kotlov, MS[14]; Yi-Hsuan Lucy Lai, PhD[15]; Alexander J. Lazar, MD, PhD[16]; Wenjie Li, MD[17]; Brittany A. McKelvey, PhD[1]; Hyunjun Nam, PhD[18]; Sarabjot Pabla, PhD[19]; Pegah Safabakhsh, MS[20]; Daniel Saul, MS[21]; Albrecht Stenzinger, MD[22]; Timothy Taxter, MD[23]; Zhiwei Zhang, PhD[2]; Yingdong Zhao, PhD[2]; ShiPing Zou, PhD[24]; Mark D. Stewart, PhD[1]; and Jeff D. Allen, PhD[1]

DOI https://doi.org/10.1200/OA-24-00042

## ABSTRACT

**PURPOSE** Homologous recombination deficiency (HRD) assays measure DNA damage repair dysfunction to identify patients with high-grade serous ovarian cancer (HGSOC) who may benefit from poly ADP-ribose polymerase inhibitors (PARPis). Numerous assays are available, but only two have undergone prospective clinical validation. Assay variability can affect patient and provider treatment choices; however, the level of assay variability across laboratory developed tests is unknown.

**METHODS** Friends of Cancer Research initiated a research partnership, inviting HRD assay developers to participate in two blinded analyses. In the first, 11 assay developers reported HRD status for the Cancer Genome Atlas HGSOC data set (In Silico; n = 348) and then 17 assay developers reported HRD status for nucleic acids freshly extracted from archival specimens (n = 90) from patients with advanced HGSOC (clinical). HRD status was compared for each analysis.

**RESULTS** The median (IQR) pairwise positive percent agreement (PPA) for the in silico analysis was 74% (51%-89%) and 81% (64%-92%) for pairwise negative percent agreement (NPA); for the clinical analysis PPA was 83% (70%-91%) and NPA was 80% (62%-91%). There was higher positive agreement on HRD status calls among those with a *BRCA1* or *BRCA2* mutation and a higher negative agreement in *CCNE1*-amplified cases. Sample characteristics like tissue block age were not observed to be associated with agreement. A subgroup of tumors largely called HRD across assays with no *BRCA1* or *BRCA2* mutations was associated with better outcomes on standard platinum-based therapy compared with not HRD; however, the subgroup was small, and further research is warranted.

**CONCLUSION** This analysis demonstrates how results from 20 HRD assays compare when assessing HGSOC. The results set the stage to improve alignment and establish standards for acceptable levels of agreement moving forward.

ACCOMPANYING CONTENT

Data Supplement

Accepted October 9, 2024
Published November 4, 2024

JCO Oncology Adv 1:e2400042


## INTRODUCTION

Poly ADP-ribose polymerase (PARP) inhibitors (PARPis) treat a range of malignancies and exert their effects through synthetic lethality, where the accumulation of single-stranded DNA breaks and PARPi trapping is lethal in the context of homologous recombination deficiency (HRD).[1-3] Patients with high-grade serous ovarian cancer (HGSOC) whose tumors are HRD have better outcomes when treated with PARPi maintenance therapy.[4-7] Nearly half of patients with HGSOC have HRD, and approximately half of patients with HRD or about 20% of all patients with HGSOC have a deleterious or potentially deleterious mutation in *BRCA1* or *BRCA2* (*BRCA1/2*),[7-10] which are integral to the homologous recombination repair (HRR) pathway.

## CONTEXT

**Key Objective**
To determine whether homologous recombination deficiency (HRD) status reporting differs across HRD assays assessing ovarian cancer and what contributes to variability.

**Knowledge Generated**
Level of HRD status agreement across 20 assays was established, without consensus on acceptable level of agreement. Characteristics that influenced HRD status agreement in the study included presence of a *BRCA1* or *BRCA2* mutation or a *CCNE1* amplification, whereas those not observed to influence agreement included tumor content and DNA integrity.

**Relevance *(S. Halabi)***
This study has significant implications for clinical care, particularly in the management of high grade serous ovarian cancer. The observed variability across different HRD assays underscores the need for standardization, as inconsistencies in assay results can influence treatment decisions for both patients and providers. These findings highlight the importance of using clinically validated assays to ensure reliable HRD status reporting, enabling more informed treatment choices and potentially improving patient outcomes.*

**Plain Language Summary *(M. Lewis)***
Whether or not 20 different tests of ovarian cancer agreed in their measurement of HRD (a condition in which cells cannot repair their DNA properly) was most determined by the presence of a *BRCA1* or *BRCA2* mutation or a *CCNE1* amplification.†

*Relevance section written by *JCO Oncology Advances* Associate Editor Susan Halabi, PhD, FASCO.
†Plain Language Summary written by *JCO Oncology Advances* Associate Editor Mark Lewis, MD.

Much of the PARPi research and development, including use of HRD assays, focuses on HGSOC.[11] In 2012, three seminal papers described genomic measures including genomic loss of heterozygosity (gLOH),[7] telomeric-allelic imbalance (TAI),[5] and large-scale transitions (LSTs)[6] that correlate with what was then described as BRCAness, which is now known as HRD. Since then, these measures have been key for diagnostic assays used to define and assess HRD.

HRD assays measure genotypic and/or phenotypic changes reflecting impairment of genes in the HRR pathway (causes) and genomic scarring/instability (ie, gLOH, TAI, LST, or consequences).[3] Each assay aggregates selected features, most often into a continuous score to which predetermined assay-specific cutoffs are applied to define the HRD status (ie, HRD or not HRD). Two HRD assays have been prospectively validated in phase III clinical trials for selecting patients with HGSOC who are more likely to benefit from PARPis, and many other assays are available as laboratory-developed tests.[12-15]

Assay developers use different methods to assess HRD and different terminology to describe HRD versus not HRD, which may lead to variability in interpreting and reporting HRD. Discordance in assay outputs may create challenges for assay interoperability in oncology drug development, biomarker selected trial design, and clinical decision-making. Evaluating concordance across assays and identifying potential sources of discordance is an important first step to address these challenges.

The value of multiassay comparisons to promote alignment has been shown for tumor mutational burden[16] and PDL1.[17] Friends of Cancer Research (*Friends*) established a consortium to assess agreement and identify potential sources of discordance in HRD status interpretation from multiple HRD assay developers assessing two different sets of ovarian cancer samples. Insights gleaned from these analyses may inform development, regulatory review, and clinical use of HRD assays and provide meaningful information for other oncology biomarker assay developments in the future.

## METHODS

### Overall Approach

*Friends* convened a working group inviting representatives from commercial and academic HRD assay developers, academia, pharmaceutical companies, National Cancer Institute (NCI), and the US Food and Drug Administration. HRD assay developers were surveyed regarding factors incorporated into their algorithms. Assay developers analyzed (1) the Cancer Genome Atlas (TCGA) ovarian cancer data (in silico; n = 348 patients)[10] and/or (2) freshly extracted nucleic acids from archival ovarian cancer samples (clinical; n = 90 patients) to determine HRD status (Table 1). The results from assay developers were blinded and deidentified. Before analysis, the

**TABLE 1.** Patient Characteristics

| Characteristic | No. (%) |
|---|---|
| In silico samples (n = 348) | |
| *BRCA1/2* mutation | 83 (24) |
| Clinical samples (n = 90) | |
| *BRCA1/2* mutation | 23 (26) |
| *CCNE1* amplification | 14 (16) |
| Race | |
| White | 66 (73) |
| Black | 23 (26) |
| Other | 1 (1) |
| Stage | |
| IIIA | 9 (10) |
| IIIB | 7 (8) |
| IIIC | 68 (76) |
| IVA | 2 (2) |
| IVB | 4 (4) |
| Surgical treatment | |
| Primary debulking | |
| R0/NRD | 18 (20) |
| Optimal | 36 (40) |
| Suboptimal | 9 (10) |
| Unknown | 6 (7) |
| Interval debulking | |
| R0/NRD | 14 (16) |
| Optimal | 6 (7) |
| Suboptimal | 1 (1) |
| Maintenance therapy | |
| Primary debulking | |
| No maintenance[a] | 54 (60) |
| Avastin | 4 (4) |
| PARPi[b] | 11 (12) |
| Interval debulking | |
| No maintenance | 15 (17) |
| Avastin | 2 (2) |
| PARPi[b] | 4 (4) |
| Platinum status | |
| Refractory/resistant | 14 (16) |
| Sensitive | 72 (80) |
| Missing | 4 (4) |
| Age at diagnosis | |
| ≤65 year | 52 (58) |
| >65 year | 38 (42) |
| Year of diagnosis | |
| 2012-2017 | 40 (44) |
| 2018-2022 | 50 (56) |

Abbreviations: R0/NRD, no residual disease; PARPi, poly ADP-ribose polymerase inhibitors.
[a]One patient included in the no maintenance category was reported to have letrozole as maintenance therapy.
[b]PARPi was either niraparib or olaparib.

working group agreed on key variables for reporting (Supplementary Methods; Data Supplement, Fig S1).

### Assay Factors Analysis

The 20 assay developers identified factors used to determine HRD, including whether the assay evaluated gLOH, TAI, LST, *BRCA1/2* mutations, and methylation and/or mutations in HRR pathway genes other than *BRCA1/2* (including number of HRR genes assessed). Assays included in the clinical analysis were also qualified as research use only (RUO) or used in clinical settings (ie, clinical use assays). This information was reported to demonstrate the variability in assay approaches and was used to assess associations with concordance. Results reported herein use the terminology HRD versus not HRD (homologous recombination proficiency [HRP] is the biological scenario where cells effectively repair DNA damage by HRR and HRD assays do not consistently identify HRP.[18,19]).

### In Silico Samples

Assay developers measured and reported HRD status and the contributing factor(s) for each sample. In silico analysis participants (n = 11) accessed deidentified segmented files,[20] MAF files, and *BRCA1/2* germline mutation files[21] for 348 TCGA ovarian cancer samples.[10] Some of the assay developers modified their pipelines to allow for use of TCGA inputs. Samples with *BRCA1/2* mutations were defined as samples identified in the TCGA germline mutation file and samples in which any assay developer identified a pathogenic somatic *BRCA1/2* mutation (Table 1).

### Clinical Samples

Institutional review board approval from the University of Alabama at Birmingham (UAB) was obtained for use of deidentified biospecimens accompanied by clinical information. Archival formalin-fixed paraffin-embedded (FFPE) ovarian cancer specimens from patients with stage III or IV HGSOC newly diagnosed at UAB between 2011 and 2022 were identified (n = 386). Patients included in the analyses were chemotherapy-naïve at tissue collection (n = 142).

Tissue curls (n = 99) were shipped to the Molecular Characterization Laboratory (MoCha) at the NCI Frederick National Laboratory for DNA and RNA extraction using the AllPrep FFPE Nucleic acid Extraction kit and the QIAcube automated platform (QIAGEN, Germantown, MD) as previously described.[16] The method of extraction did not necessarily align with each assay developers' standard methods. MoCha shipped identical aliquots of DNA and/or RNA from samples passing quality control to each assay developer for independent sequencing and HRD measurement (n = 90 cases/17 assays developers; Fig 1). Digitized H&E-stained slides were used to calculate tumor content for associations with concordance (Supplementary Methods).

### Statistical Analyses

Statisticians from the NCI Biometric Research Program analyzed HRD status calls to assess the level of agreement between assays and considered specific factors measured by each assay to identify potential sources of variation. Data from the in silico and clinical analyses were assessed consecutively, and the results were not directly compared, as not all assay developers participated in both studies. HRD positivity was defined as the percentage of samples called HRD per assay developer. Indeterminate and failed cases were excluded from both the numerator and denominator for calculation of each assay's percent HRD.

Statistical tests to compare agreement measures between different groups of samples (eg, those harboring *BRCA1/2* mutations versus wild type [WT]) were performed using a nonparametric bootstrapping procedure, resampling patients but treating assays as fixed. Two-sided *P* values were computed. *P* values <.05 were reported as significant, but no correction for multiplicity was applied so the results should be interpreted as descriptive.

Additional details about definitions of agreement metrics and statistical approaches are provided in the Supplementary Methods.

## RESULTS

### Assay Factors

Assay developers volunteered to participate in the clinical analysis only (n = 9), the in silico analysis only (n = 3), or both analyses (n = 8). All participating assay developers assessed alterations in *BRCA1* and *BRCA2* when defining HRD status. Sixty percent of the assays considered mutations (median of 22 putative HRR genes, range, 14-46) beyond *BRCA1* and *BRCA2*. For consequences, 60% of the assays included assessment of gLOH, 45% included TAI, and 45% included LST or some combination thereof. Some stakeholders incorporated artificial intelligence (AI) on broad copy number features into their algorithms to determine HRD status without directly assessing recognized measures of gLOH, TAI, or LST.

The approach to calculating and reporting HRD status differed by assay. The range of continuous values reflecting degree of HRD differed with developers using 0% to 100%, 0 to 100, 0 to 1, 0 to 90, and −30 to +30. Among those who used similar ranges, the cutoff value for defining HRD differed (eg, 35, 40, 42, 50). When reporting results clinically, the terminology to describe the findings included HRD-positive versus HRD-negative, HRD-detected versus not detected, HRD versus HRP, and HRD versus not HRD.

The median (range) percent HRD across assays for in silico was 49% (9%-67%) and 52% (23%-74%) for clinical (Data Supplement, Fig S2). For indeterminate and failed cases, 10 assays reported 0 cases, four assays reported 1-5 cases, and three assays reported more than five cases.

### HRD Concordance

Variability in HRD status across assays was observed for both analyses (Figs 2A and 2B). Pairwise agreement percentages for HRD status are shown in Figure 3, with

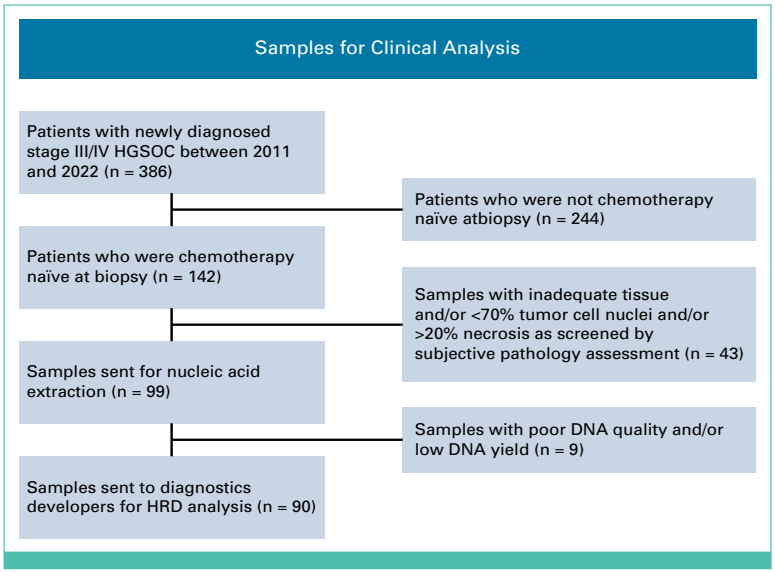



**FIG 1.** Identification of clinical samples for inclusion in this study. HGSOC, high-grade serous ovarian cancer; HRD, homologous recombination deficiency.

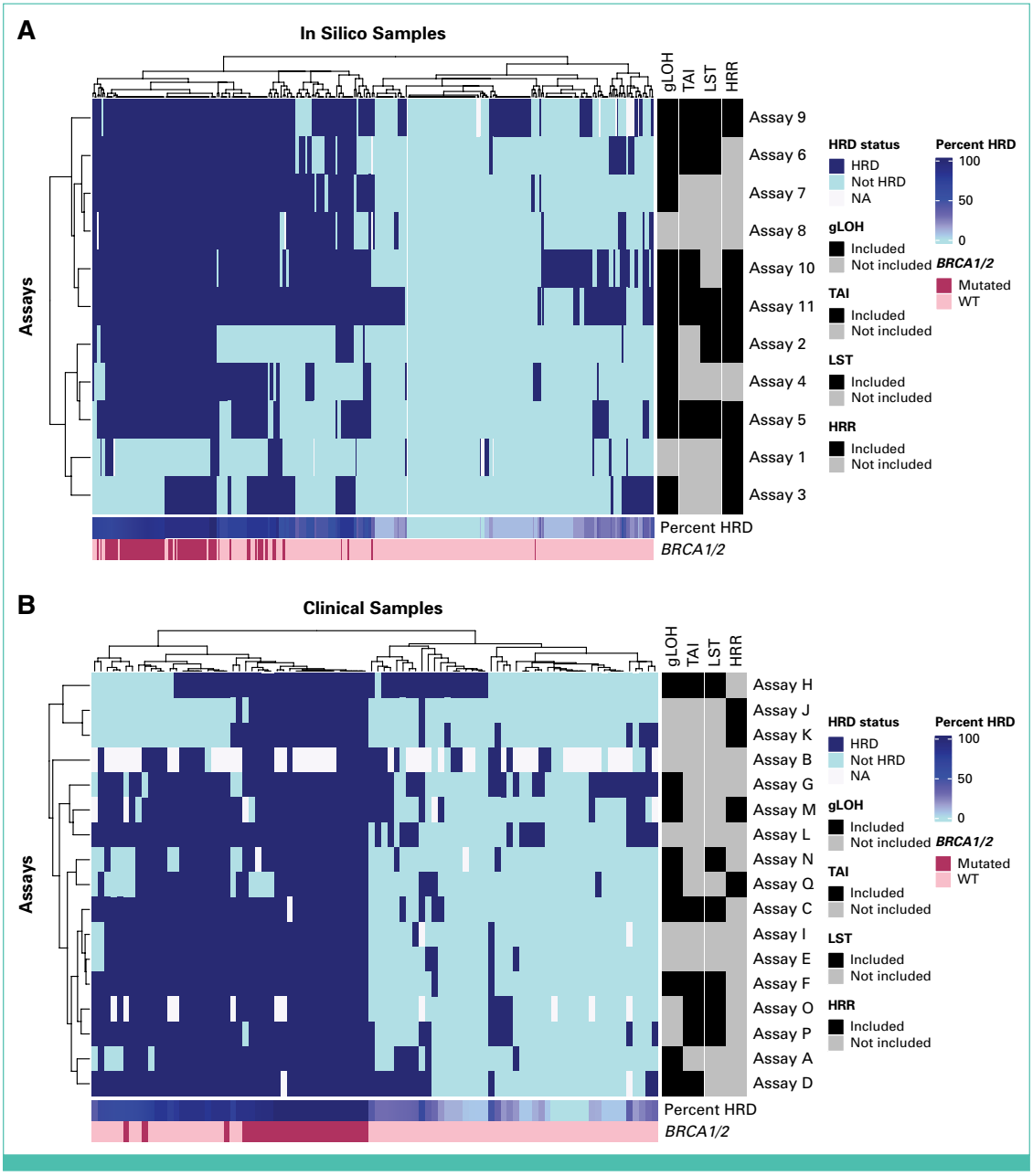



**FIG 2.** Variability in HRD status calls across assays. For both (A) the in silico analysis and (B) the clinical analysis, the tile plots depict assays and samples ordered according to clustering by relatedness of HRD assessments. Assay factors (ie, gLOH, TAI, LST, HRR [mutations in HRR pathway genes in addition to *BRCA1* and *BRCA2*]) are depicted as yes/no on the basis of whether assay developer included the factor to determine HRD status in their assay algorithm. For HRD status, NA includes indeterminate, failed, or not applicable samples. Percent HRD is included as a gradient depicting the percent HRD for the patient sample (ie, percentage of assays called the sample HRD). *BRCA1/2* indicates whether that sample had a *BRCA1* or *BRCA2* mutation (mutated) or was WT. gLOH, genomic loss of heterozygosity; HRD, homologous recombination deficiency; HRR, homologous recombination repair; LST, large-scale transitions; TAI, telomeric-allelic imbalance; WT, wild type.

median (IQR) computed over all possible pairs. The median pairwise positive percent agreement (PPA) for the in silico samples was 74% (IQR, 51%-89%) and for the clinical samples was 83% (IQR, 70%-91%). The median pairwise negative percent agreement (NPA) for the in silico samples was 81% (IQR, 64%-92%) and for the

clinical samples was 80% (IQR, 62%-91%; Data Supplement, Table S1).

### Assessing Drivers of Variability

Assay, patient, tumor, and sample characteristics were assessed for potential contributions to variability across assays. Assay factors assessed to define HRD including gLOH, TAI, LST, and HRR (ie, mutations in HRR pathway genes) and assay use are indicated on the right on the tile plots (Fig 2, Data Supplement, Fig S3). Similar levels of agreement were observed when comparing RUO only assays (n = 10) to clinical used assays (n = 7; Data Supplement, Table S1).

For patient clinical and demographic characteristics, agreement was similar for White and non-White patients on most metrics; however, average pairwise negative agreement (ANA; ie, on non-HRD) was better for non-White patients with a median of 86% (IQR, 81%-89%) compared with White patients who had a median of 70% (IQR, 55%-80%; $P < .01$). Surgical debulking status did not exhibit an association with agreement on HRD status (Data Supplement, Table S1, Data Supplement, Fig S3).

Tumors with *BRCA1/2* mutations were more consistently categorized as HRD compared with cases with WT *BRCA1* and *BRCA2* (WT *BRCA1/2*) in both the in silico (n = 83) and clinical (n = 23) analyses (Figs 2 and 3). Median PPA and average pairwise positive agreement were higher, and NPA and ANA lower, for tumors with *BRCA1/2* mutations compared with WT in the clinical analysis ($P < .0001$ for all four agreement measures) and the in silico analysis ($P < .05$ for all agreement measures except NPA where $P = .3761$). *CCNE1* gene amplification is associated with non-HRD in HGSOC,[22] and a greater negative (non-HRD) agreement was observed for assessment of cases in the clinical analysis with *CCNE1* amplification (n = 14; median NPA 91%, IQR, 73-100) compared with those without *CCNE1* amplification (n = 76; median NPA 75%, IQR, 61-89; $P < .01$; Data Supplement, Table S1; Data Supplement, Fig S3).

Clinical specimen characteristics were assessed for associations with assay agreement. The median tumor purity as assessed by computational pathology was 78% (IQR, 66%-84%), and average concordance score of HRD status (ACS [HRD]) was not observed to be associated with tumor purity (Data Supplement, Fig S4A). DNA quality and the age of the block were not observed to be associated with concordance (Data Supplement, Figs S4B and S4C).

Examination of concordance of (binary) HRD status (ACS [HRD]) in relation to continuous HRD scores revealed that the lowest point on the blue lowess smoothing curve for many assays was near their HRD cutoff, suggesting reduced concordance of HRD status calls in that range (Fig 4).

### Causes Versus Consequences

For the clinical analysis, assay developers reported whether HRD status was determined by causes, consequences, or both (Data Supplement, Fig S5). Of note, hierarchical clustering with L1 distance and complete linkage led to visual identification of two clusters of samples: one including more HRD (HRD cluster; denoted in red) and one with more not HRD (not HRD cluster; denoted in blue; Data Supplement, Fig S3). Within the HRD cluster, most patients with a *BRCA1/2* mutation were clustered together. Among those within the HRD cluster with WT *BRCA1/2*, many were defined as HRD by consequences (Data Supplement, Fig S5). Aligning with this finding, assays J and K, which assessed only mutations in HRR genes to define HRD, did not call these samples HRD. Ultimately, the samples fell into three groups: HRD cluster with mutated *BRCA1/2*, HRD cluster with WT *BRCA1/2*, and the not HRD cluster.

### HRD Associations With Platinum Response Status and Long-Term Outcomes

All patients in the clinical analysis received platinum-based therapy with or without a taxane and/or bevacizumab. Only 15 patients (17%) received PARPi maintenance therapy, which was too few to draw conclusions about outcome-HRD associations. Clinical sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of each of the HRD assays for association with responsiveness to platinum-based chemotherapy were assessed (Data Supplement, Fig S6). Most HRD cases had platinum-sensitive HGSOC (n = 72, 80%), and the median (IQR) PPV across assays was 91% (IQR, 89-95), whereas median (IQR) NPV was 24% (IQR, 22-27).

The HRD cluster was numerically associated with better overall survival (OS; hazard ratio [HR] = 0.67); however, the curves crossed, and there was no statistically significant difference based on a log-rank test (Figs 5A and 5B). The median OS for the HRD cluster and not HRD cluster were 85.9 and 91.6 months, respectively, near where the curves cross. Similar trends were seen when segregating samples by HRD status according to individual assay results (Data Supplement, Fig S7); although, the degree of separation of the survival curves varied across assays.

A separate analysis explored the effects of the presence of *BRCA1/2* mutation in the HRD cluster on outcomes (Figs 5C and 5D). When comparing OS for the subgroup with a *BRCA1/2* mutation with that for the WT *BRCA1/2* subgroup in the HRD cluster, the lines on the graph overlapped, suggesting that HRD in the absence of *BRCA1/2* mutations identifies patients with similar OS after platinum-based therapy as those with *BRCA1/2* mutations. Interestingly, the median OS for the HRD cluster with WT *BRCA1/2* was not reached, whereas the median OS for those with *BRCA1/2* mutations was 72.6 months. Assessments were also made for *BRCA1/2* mutated versus WT, demonstrating a similar association with outcome as seen for HRD; patients with

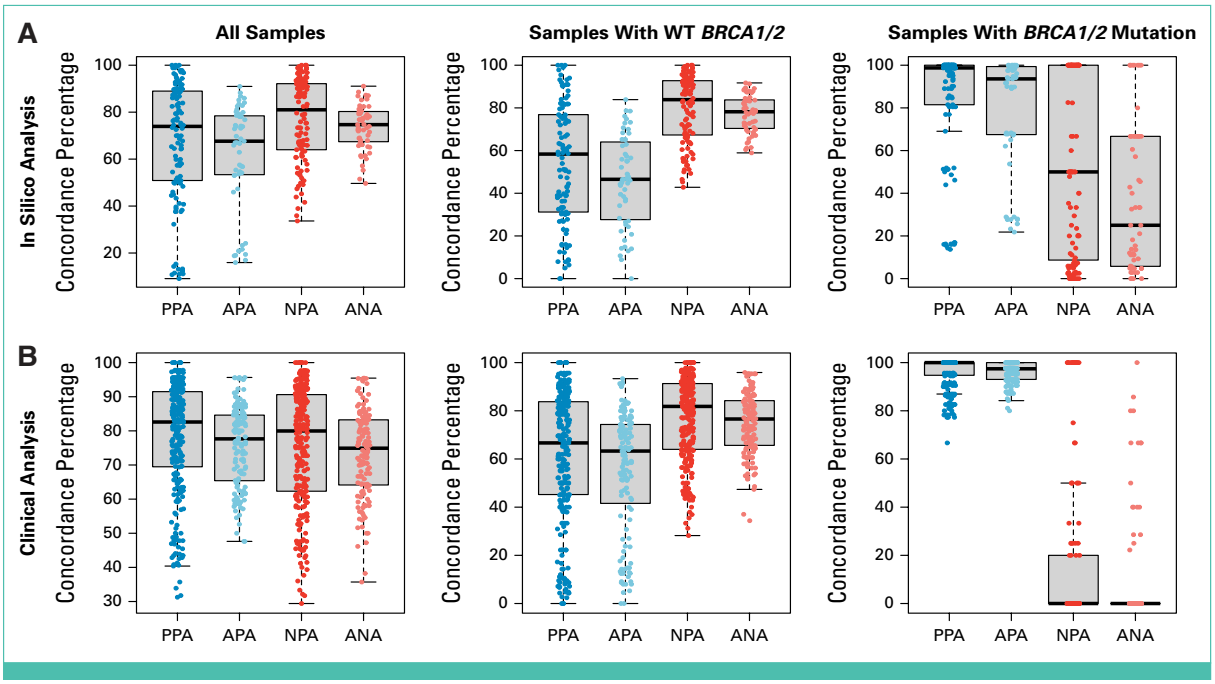



**FIG 3.** Pairwise positive/negative agreement varied across assays. In both (A) the in silico analysis and (B) the clinical analysis, PPA is higher when only samples with *BRCA1/2* mutations are considered, NPA is lower. The dots in the graphs depict the indicated concordance percentages from all possible pairings of assays for in silico and clinical analyses. ANA, average pairwise negative agreement; APA, average pairwise positive agreement; NPA, negative percent agreement; PPA, positive percent agreement.

*BRCA1/2* mutations had numerically (HR = 0.78), but not statistically significantly, better outcome compared with WT *BRCA1/2* (Data Supplement, Fig S8). However, again, median OS was not reached for those with WT *BRCA1/2*, whereas it was 85.9 months for those with *BRCA1/2* mutations.

## DISCUSSION

Assessing variability in assay outputs provides foundational information to support the development of assays that provide optimal clinical performance for predicting treatment response.[16,17,23] A volunteer group of HRD assay developers (n = 20) participated in two blinded assessments of variability of HRD status calls across HGSOC data sets, including an in silico analysis (n = 348) and a clinical analysis (n = 90). PPA and NPA were similar within each data set, and they were also similar between in silico and clinical data sets. Positive agreement increased, whereas negative agreement decreased, when patients with *BRCA1/2* mutations were analyzed. Various factors that may affect concordance were considered. Although the sample size for the clinical analysis (n = 90) was not powered to reach definitive conclusions, the analysis enabled the identification of trends and areas for future investigation.

Several factors including clinical, sample, and assay characteristics were evaluated to determine effects on assay agreement. Alterations in *BRCA1/2* drive HRD, and all assays assessed alterations in *BRCA1/2* when determining HRD, aligning with stronger positive agreement on HRD status where a mutation in *BRCA1/2* was present. *CCNE1* amplification tends to be mutually exclusive from HRD, and there was better negative agreement of HRD status in *CCNE1*-amplified cases. Previous work demonstrates that tissue sample quality and intratumoral heterogeneity may influence HRD assay agreement.[24] Here, sample characteristics such as tumor purity, DNA quality, and age of the block were not observed to be associated with agreement; however, sample selection limited analysis to high-quality samples. Agreement in a real-world setting with lower purity and DNA quality is beyond the scope of this project.[25,26]

Differences in how HRD assays determine HRD status may influence concordance. Variable combinatorial inclusion of gLOH, TAI, and LST made it challenging to identify any single factor explaining differences in assay agreement. However, patients in the HRD cluster with WT *BRCA1/2* were determined to not have HRD by assays that only assessed mutations in HRR genes. The assays assessed a median of 22 genes; however, recent research shows that few genes beyond *BRCA1/2* strongly associate with an HRD phenotype.[27,28] In this study, patients with WT *BRCA1/2* who were in the HRD cluster were primarily identified by consequences and had similar outcomes to those with *BRCA1/2* mutations,

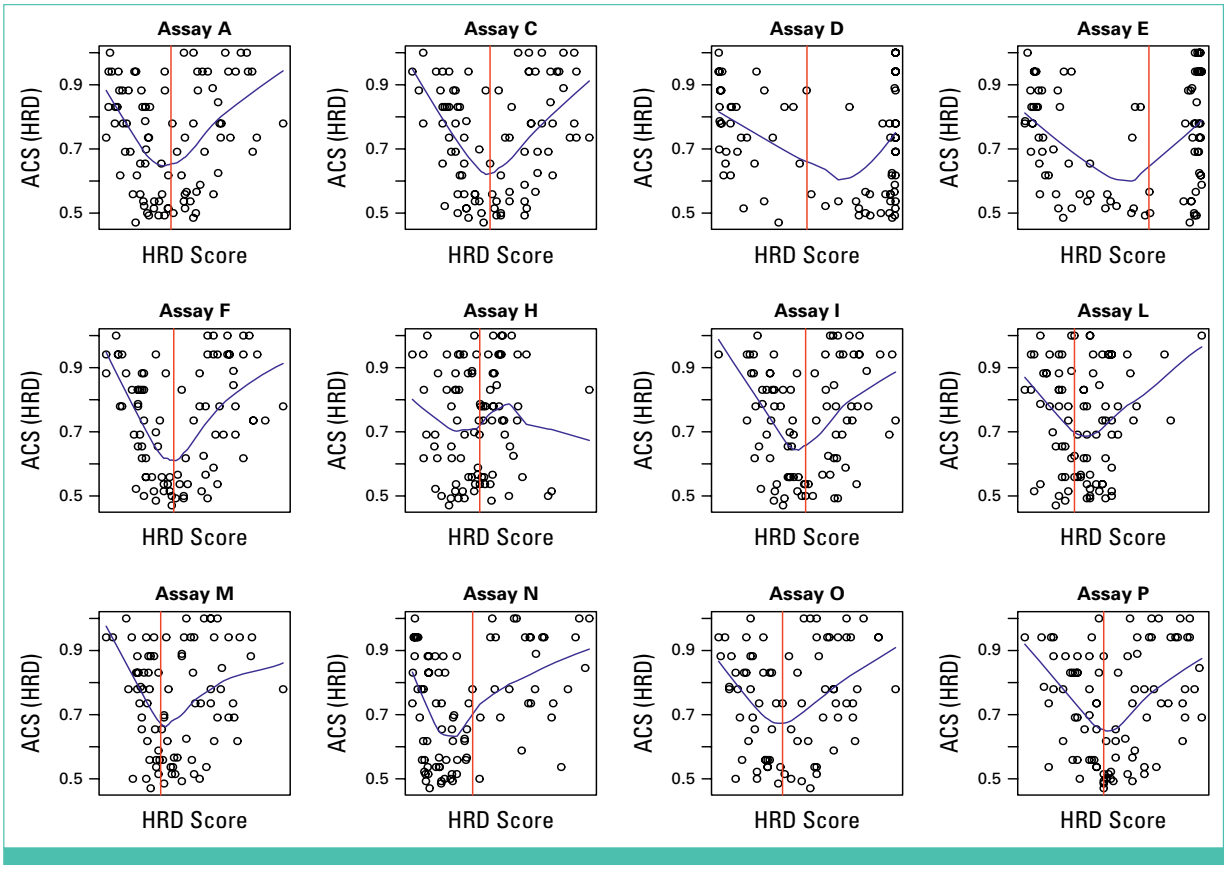


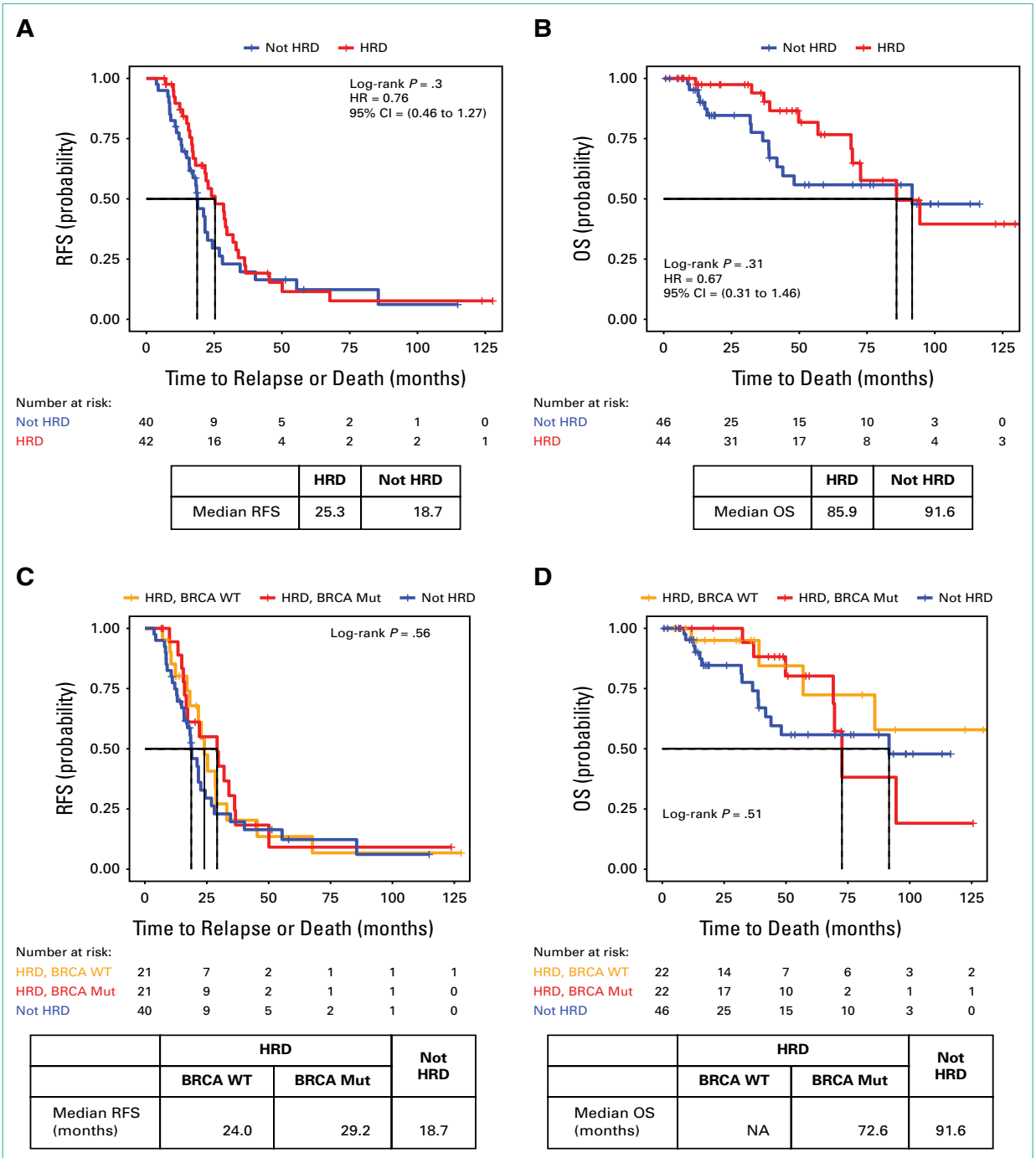

**FIG 4.** Assay concordance at or near individual assay's HRD cutoff. Each sample is represented by a dot on the graph representing its ACS (HRD). The red line represents the assay's HRD cutoff and the blue line is a lowess smoothing curve to demonstrate trend. The HRD score was provided by each individual assay and is intentionally not included to prevent unblinding. ACS, average concordance score; HRD, homologous recombination deficiency.

suggesting that patients identified to have HRD by consequences may have better clinical outcomes compared with those without HRD, but further studies are needed to confirm trends. Additionally, previous work demonstrated inconsistency in HRD assay cutoff thresholds for defining HRD,[3] and the assays herein had lower concordance near cutoffs.

The current clinical use of HRD assays is to identify patients who would benefit from PARPi treatment. Most patients in the analysis did not receive PARPi (n = 75, 83%), and treatment was not randomized, so the predictive value of the assays for PARPi benefit could not be evaluated. The hazard ratio (HR = 0.67) suggested the HRD cluster numerically trended toward more favorable OS, consistent with other studies,[29] but the difference was not statistically significant. The high percentage of platinum responders in this study aligns with a high PPV in the platinum sensitivity analyses, yet this favorable scenario might have made further refinement of prognosis by HRD more challenging.

The analyses did not compare to a gold standard because there is not an agreed on gold standard for the biological concept of HRD. The ideal gold standard to improve concordance should consider biological characterization as HRD since specific therapeutic options may change with time, but the assays will likely continue to be used to measure HRD status. However, for clinical decision making, a fit-for-purpose assay should reliably, reproducibly, and accurately address the assay's intended clinical use.

The results reported herein quantify concordance among assays, but the multistakeholder working group did not reach consensus on how to define the quality of concordance, which highlights the need to define the level of acceptable disagreement between assays. From a clinical perspective, it is critical that patients and providers have consistent results when making treatment decisions and concordance closer to 100% would be ideal. However, assays are not perfect, and even interlaboratory differences in results when running the same assay occur.[30] Pfarr et al assessed >100 samples in




| | HRD | Not HRD |
|---|---|---|
| Median RFS | 25.3 | 18.7 |

| | HRD | Not HRD |
|---|---|---|
| Median OS | 85.9 | 91.6 |

| | HRD | | Not HRD |
|---|---|---|---|
| | BRCA WT | BRCA Mut | |
| Median RFS (months) | 24.0 | 29.2 | 18.7 |

| | HRD | | Not HRD |
|---|---|---|---|
| | BRCA WT | BRCA Mut | |
| Median OS (months) | NA | 72.6 | 91.6 |

**FIG 5.** Survival analysis of patients in the HRD cluster (red) compared with the not HRD cluster (blue), as well as comparing patients within the HRD cluster who had WT *BRCA1/2* (BRCA WT) versus those with mutated *BRCA1/2* (BRCA Mut). RFS and OS were estimated by Kaplan-Meier curves. A two-sided log-rank test was used to compare RFS and OS between the HRD cluster and non-HRD cluster. RFS was defined as time from start of primary platinum-based therapy to date of relapse or death, whichever comes first, or censored if the patient had not relapsed and was still alive at last follow up. OS was defined as time from start of primary platinum-based therapy to date of death due to any cause or censored if alive at last follow up. HR, hazard ratio; HRD, homologous recombination deficiency; NA, not applicable; OS, overall survival; RFS, recurrence-free survival; WT, wild type.

seven pairs of laboratories using the same assays; Pearson correlations between paired continuous HRD scores using the same assay in different laboratories exhibited a range of 0.62-0.98.[30] Other comparisons have seen greater agreement when considering fewer laboratories.[31-35] It would be expected that concordance decreases further with more assays. Importantly, there could be discordance between assays, yet the assays might similarly identify patients who benefit from treatment (eg, PARPis).

When establishing approaches for improved alignment, it is important to consider that HRD testing continues to advance. Novel techniques like incorporating AI or circulating tumor DNA are being employed for assay development. While assessing HRD in other cancer types to determine benefit from PARPi treatment shows promise, this analysis focused on ovarian cancer, and whether the trends in assay agreement seen here would apply to other solid tumors is unknown. Adjustments to assay algorithms may be necessary for different cancer types as evidenced by the HRR gene panel (ie, does not assess consequences) used for identifying patients with metastatic castrate-resistant prostate cancer who might benefit from PARPi monotherapy.[36] The variability in assay results and HRD reporting suggests that harmonizing clinical validation approaches of HRD is critical to ensure that patients and providers have access to consistent, clinically reliable, and meaningful information to guide treatment decisions.


## AFFILIATIONS

[1]Friends of Cancer Research, Washington, DC
[2]National Cancer Institute, Bethesda, MD
[3]University of Alabama Birmingham, O'Neal Comprehensive Cancer Center, Birmingham, AL
[4]Frederick National Laboratory for Cancer Research, Frederick, MD
[5]Pfizer, Inc, South San Francisco, CA
[6]Foundation Medicine, Boston, MA
[7]Invitae Corporation, San Francisco, CA
[8]Amoy Diagnostics Co, Ltd, Shanghai, China
[9]PathAI, Boston, MA
[10]SOPHiA GENETICS, Rolle, Switzerland
[11]Illumina, Inc, San Diego, CA
[12]ThermoFisher, South San Fransisco, CA
[13]Diagnostic Laboratory Services, Inc, Aiea, HI
[14]BostonGene, Waltham, MA
[15]ACT Genomics Co, Ltd, Taipei, Taiwan
[16]Departments of Pathology & Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX
[17]Burning Rock Dx, Guangzhou, China
[18]NeoGenomics Laboratories, San Diego, CA
[19]Labcorp Oncology (OmniSeq), Buffalo, NY
[20]Guardant Health, Palo Alto, CA
[21]Bionano Genomics, Inc, San Diego, CA
[22]Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[23]Tempus AI, Chicago, IL
[24]Pillar Biosciences, Inc, Natick, MA

## CORRESPONDING AUTHOR

Hillary S. Andrews, PhD; twitter: @hillstirsci; e-mail: handrews@focr.org.


## PRIOR PRESENTATION

Presented in part at the Association for Molecular Pathology's 2022 Annual Meeting and Expo, Phoenix AZ, November 1-5, 2022; the American Association for Cancer Research Special Conference: Ovarian Cancer, Boston, MA, October 5-7, 2023; the Society for Gynecologic Oncology Winter Meeting, Whistler, British Columbia, CA, January 30-February 1, 2024; the Friends of Cancer Research Future of Diagnostic Tests Public Meeting, Washington, DC, February 1, 2024; and by The Cancer Letter, May 17, 2024.


## SUPPORT

Supported by Friends of Cancer Research.


## AUTHOR CONTRIBUTIONS

**Conception and design:** Hillary S. Andrews, Lisa M. McShane, Elise C. Kohn, Rebecca Arend, Chris Karlovich, A. Douglas Laird, Ethan S. Sokol, Elizabeth R. Starks, Mohit Gupta, Alexander J. Lazar, Brittany A. McKelvey, Hyunjun Nam, Timothy Taxter, Mark D. Stewart, Jeff D. Allen
**Administrative support:** Rebecca Arend, Albrecht Stenzinger
**Provision of study materials or patients:** Rebecca Arend
**Collection and assembly of data:** Hillary S. Andrews, Lisa M. McShane, Elise C. Kohn, Rebecca Arend, Kaitlyn Kincaid, Ethan S. Sokol, Elizabeth R. Starks, Lauren Brunner, Alyssa Chapman, Li Chen, Tommaso Coletta, Yuan Ding, Bailee D. Oliver, McKenzie Foxall, Mohit Gupta, Zan Halford, Andrea G. Kahn, Yi-Hsuan Lucy Lai, Alexander J. Lazar, Hyunjun Nam, Sarabjot Pabla, Pegah Safabakhsh, Daniel Saul, Timothy Taxter, ShiPing Zou
**Data analysis and interpretation:** Hillary S. Andrews, Lisa M. McShane, Elise C. Kohn, Rebecca Arend, A. Douglas Laird, Ming-Chung Li, Ethan S. Sokol, Shulin Bi, Mohit Gupta, Nikita Kotlov, Alexander J. Lazar, Wenjie Li, Brittany A. McKelvey, Hyunjun Nam, Sarabjot Pabla, Pegah Safabakhsh, Albrecht Stenzinger, Timothy Taxter, Zhiwei Zhang, Yingdong Zhao, ShiPing Zou, Mark D. Stewart Elizabeth R. Starks, Zan Halford
**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. Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

**Hillary S. Andrews**
**Consulting or Advisory Role:** Avalere Health

**Rebecca Arend**
**Employment:** Signatera (I)

**Consulting or Advisory Role:** VBL Therapeutics, GlaxoSmithKline, Merck, Seagan, Sutro Biopharma, KIYATEC, Caris Life Sciences, Leap Therapeutics
**Research Funding:** Champions Oncology, Exelixis, GlaxoSmithKline, Immunogen, Merck
**Travel, Accommodations, Expenses:** Caris Life Sciences, GlaxoSmithKline, VBL Therapeutics

**Chris Karlovich**
**Research Funding:** Illumina (Inst)
**Uncompensated Relationships:** Illumina

**A. Douglas Laird**
**Employment:** Pfizer
**Stock and Other Ownership Interests:** Pfizer

**Ethan S. Sokol**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Roche

**Elizabeth R. Starks**
**Employment:** LabCorp, InVitae
**Stock and Other Ownership Interests:** InVitae
**Travel, Accommodations, Expenses:** LabCorp, InVitae

**Shulin Bi**
**Employment:** Amoy Diagnostics
**Honoraria:** Amoy Diagnostics

**Lauren Brunner**
**Employment:** GlaxoSmithKline, PathAI
**Stock and Other Ownership Interests:** PathAI
**Travel, Accommodations, Expenses:** GlaxoSmithKline
**Travel, Accommodations, Expenses:** PathAI

**Li Chen**
**Research Funding:** Illumina (Inst)
**Expert Testimony:** Illumina (Inst)

**Yuan Ding**
**Employment:** Illumina, Lilly (I)
**Stock and Other Ownership Interests:** Illumina
**Patents, Royalties, Other Intellectual Property:** Magnetic-bead based DNA extraction and normalization (Inst)

**Nikita Kotlov**
**Employment:** BostonGene
**Stock and Other Ownership Interests:** BostonGene
**Patents, Royalties, Other Intellectual Property:** Co-author of 1 BostonGene patent
**Travel, Accommodations, Expenses:** BostonGene

**Alexander J. Lazar**
**Leadership:** Archer, Iterion Therapeutics, Nucleai
**Stock and Other Ownership Interests:** Archer, Beta Cat Pharmaceuticals, Paige.AI
**Honoraria:** Novartis, Bristol Myers Squibb, Janssen Oncology, Genentech/Roche
**Consulting or Advisory Role:** Novartis, AbbVie, Bayer, BMS, Deciphera
**Research Funding:** MedImmune, AstraZeneca, Roche, Novartis
**Patents, Royalties, Other Intellectual Property:** Elsevier
**Travel, Accommodations, Expenses:** Bristol Myers Squibb, Novartis

**Hyunjun Nam**
**Employment:** NeoGenomics Laboratories
**Stock and Other Ownership Interests:** NeoGenomics Laboratories
**Travel, Accommodations, Expenses:** NeoGenomics Laboratories

**Sarabjot Pabla**
**Employment:** Omniseq, LabCorp
**Patents, Royalties, Other Intellectual Property:** Patent pending at Omniseq (Inst)

**Pegah Safabakhsh**
**Employment:** Guardant Health
**Stock and Other Ownership Interests:** Guardant Health
**Travel, Accommodations, Expenses:** Guardant Health

**Daniel Saul**
**Employment:** BioNano Genomics
**Stock and Other Ownership Interests:** BioNano Genomics, Illumina

**Albrecht Stenzinger**
**Consulting or Advisory Role:** AstraZeneca, Novartis, Bristol Myers Squibb, Bayer, Illumina, Janssen, Lilly, Takeda, MSD, Amgen, Incyte, SERVIER, Astellas Pharma
**Speakers' Bureau:** Bristol Myers Squibb, AstraZeneca, MSD, Roche, Bayer, Illumina, Thermo Fisher Scientific, Novartis, Astellas Pharma, SERVIER
**Research Funding:** Chugai Pharma, Bristol Myers Squibb, Bayer, Incyte

**Timothy Taxter**
**Employment:** Tempus
**Stock and Other Ownership Interests:** Tempus

**Zhiwei Zhang**
**Employment:** Gilead Sciences
**Stock and Other Ownership Interests:** Gilead Sciences
**Travel, Accommodations, Expenses:** Gilead Sciences

No other potential conflicts of interest were reported.

## REFERENCES


1. Krastev DB, Wicks AJ, Lord CJ: PARP inhibitors—Trapped in a toxic love affair. Cancer Res 81:5605-5607, 2021
2. Rose M, Burgess JT, O'Byrne K, et al: PARP inhibitors: Clinical relevance, mechanisms of action and tumor resistance. Front Cell Dev Biol 8:564601, 2020
3. Stewart MD, Merino Vega D, Arend RC, et al: Homologous recombination deficiency: Concepts, definitions, and assays. Oncologist 27:167-174, 2022
4. Huang D, Kraus WL: The expanding universe of PARP1-mediated molecular and therapeutic mechanisms. Mol Cell 82:2315-2334, 2022
5. Birkbak NJ, Wang ZC, Kim J-Y, et al: Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer Discov 2:366-375, 2012
6. Popova T, Manié E, Rieunier G, et al: Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. Cancer Res 72:5454-5462, 2012
7. Abkevich V, Timms KM, Hennessy BT, et al: Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer 107:1776-1782, 2012
8. Sosinsky A, Ambrose J, Cross W, et al: Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme. Nat Med 30:279-289, 2024
9. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, et al: Homologous recombination deficiency: Exploiting the fundamental vulnerability of ovarian cancer. Cancer Discov 5:1137-1154, 2015
10. The Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. Nature 474:609-615, 2011
11. Wu Y, Xu S, Cheng S, et al: Clinical application of PARP inhibitors in ovarian cancer: From molecular mechanisms to the current status. J Ovarian Res 16:6, 2023
12. Mirza MR, Monk BJ, Herrstedt J, et al: Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. N Engl J Med 375:2154-2164, 2016
13. Coleman RL, Oza AM, Lorusso D, et al: Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): A randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 390:1949-1961, 2017
14. Pujade-Lauraine E, Ledermann JA, Selle F, et al: Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): A double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol 18:1274-1284, 2017
15. Pozzorini C, Andre G, Coletta T, et al: GIInger predicts homologous recombination deficiency and patient response to PARPi treatment from shallow genomic profiles. Cell Rep Med 4:101344, 2023
16. Vega DM, Yee LM, McShane LM, et al: Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: Phase II of the Friends of cancer research TMB harmonization project. Ann Oncol 32:1626-1636, 2021

17. Tsimafeyeu I, Imyanitov E, Zavalishina L, et al: Agreement between PDL1 immunohistochemistry assays and polymerase chain reaction in non-small cell lung cancer: CLOVER comparison study. Sci Rep 10:3928, 2020
18. Ngoi NYL, Tan DSP: The role of homologous recombination deficiency testing in ovarian cancer and its clinical implications: Do we need it? ESMO Open 6:100144, 2021
19. Miller RE, Leary A, Scott CL, et al: ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. Ann Oncol 31: 1606-1622, 2020
20. Weinstein JN, Collisson EA, Mills GB, et al: The Cancer Genome Atlas pan-cancer analysis project. Nat Genet 45:1113-1120, 2013
21. Maxwell KN, Wubbenhorst B, Wenz BM, et al: BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun 8:319, 2017
22. Bai S, Taylor SE, Jamalruddin MA, et al: Targeting therapeutic resistance and multinucleate giant cells in CCNE1-amplified HR-Proficient ovarian cancer. Mol Cancer Ther 21:1473-1484, 2022
23. Merino DM, McShane LM, Fabrizio D, et al: Establishing guidelines to harmonize tumor mutational burden (TMB): In silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of cancer research TMB harmonization project. J Immunother Cancer 8:e000147, 2020
24. Takaya H, Nakai H, Sakai K, et al: Intratumor heterogeneity and homologous recombination deficiency of high-grade serous ovarian cancer are associated with prognosis and molecular subtype and change in treatment course. Gynecol Oncol 156:415-422, 2020
25. O'Sullivan Coyne G, Karlovich C, Wilsker D, et al: PARP inhibitor applicability: Detailed assays for homologous recombination repair pathway components. OncoTargets Ther 15:165-180, 2022
26. Chiang Y-C, Lin P-H, Cheng W-F: Homologous recombination deficiency assays in epithelial ovarian cancer: Current status and future direction. Front Oncol 11:675972, 2021
27. Westphalen CB, Fine AD, André F, et al: Pan-cancer analysis of homologous recombination repair-associated gene alterations and genome-wide loss-of-heterozygosity score. Clin Cancer Res 28: 1412-1421, 2022
28. Kahn RM, Selenica P, Boerner T, et al: Pathogenic germline variants in non-BRCA1/2 homologous recombination genes in ovarian cancer: Analysis of tumor phenotype and survival. Gynecol Oncol 180:35-43, 2024
29. Stronach EA, Paul J, Timms KM, et al: Biomarker assessment of HR deficiency, tumor BRCA1/2 mutations, and CCNE1 copy number in ovarian cancer: Associations with Clinical Outcome Following Platinum Monotherapy. Mol Cancer Res 16:1103-1111, 2018
30. Pfarr N, von Schwarzenberg K, Zocholl D, et al: High concordance of different assays in the determination of homologous recombination deficiency-associated genomic instability in ovarian cancer. JCO Precis Oncol 10.1200/PO.23.00348
31. Pepe F, Guerini-Rocco E, Fassan M, et al: In-house homologous recombination deficiency testing in ovarian cancer: A multi-institutional Italian pilot study. J Clin Pathol 77:478-485, 2023
32. Magliacane G, Brunetto E, Calzavara S, et al: Locally performed HRD testing for ovarian cancer? Yes, we can. Cancers (Basel) 15:43, 2022
33. Denkert C, Romey M, Swedlund B, et al: Homologous recombination deficiency as an ovarian cancer biomarker in a real-world cohort: Validation of decentralized genomic profiling. J Mol Diagn 24: 1254-1263, 2022
34. Capoluongo ED, Pellegrino B, Arenare L, et al: Alternative academic approaches for testing homologous recombination deficiency in ovarian cancer in the MITO16A/MaNGO-OV2 trial. ESMO Open 7:100585, 2022
35. Roma C, Esposito Abate R, Sacco A, et al: Harmonization of homologous recombination deficiency testing in ovarian cancer: Results from the MITO16A/MaNGO-OV2 trial. Eur J Cancer 206: 114127, 2024
36. de Bono J, Mateo J, Fizazi K, et al: Olaparib for metastatic castration-resistant prostate cancer. N Engl J Med 382:2091-2102, 2020

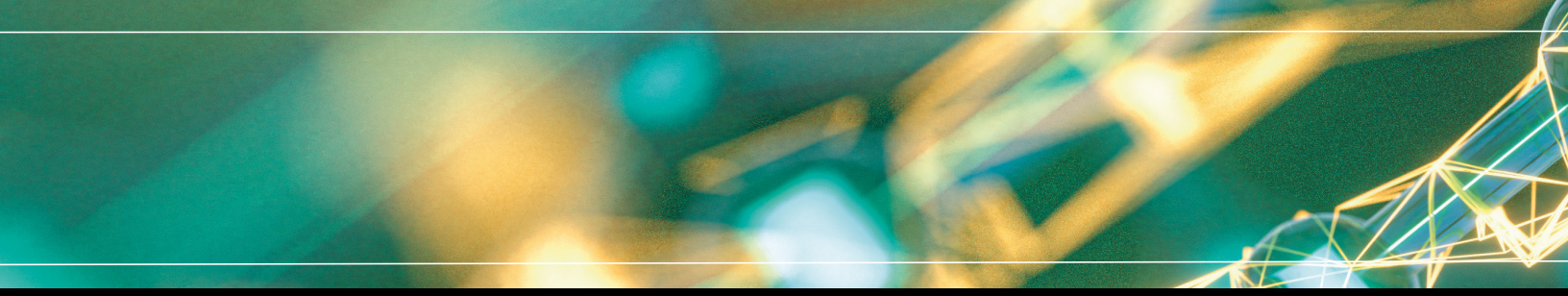
## P2-02-18: Agreement Across 10 Artificial Intelligence Models in Assessing HER2 in Breast Cancer Whole Slide Images: Findings from the Friends of Cancer Research Digital PATH Project

Presenting Author(s): Brittany McKelvey and Co-Author(s): Brittany McKelvey, Pedro A. Torres-Saavedra, Jessica Li, Glenn Broeckx, Frederik Deman, Siraj Ali, Hillary S. Andrews, Salim Arslan, Santhosh Balasubramanian, J. Carl Barrett, Peter Caie, Ming Chen, Daniel Cohen, Tathagata Dasgupta, Brandon Gallas, George Green, Mark Gustavson, Sarah Hersey, Ana Hidalgo Sastre, Shahanawaz Jiwani, Wonkyung Jung, Kimary Kulig, Vladimir Kushnarev, Xiaoxian Li, Meredith Lodge, Joan Mancuso, Mike Montalto, Satabhisa Mukhopadhyay, Matthew Oberley, Pahini Pandya, Oscar Puig, Edward Richardson, Alexander Sarachakov, Or Shaked, Mark Stewart, Lisa M. McShane, Roberto Salgado, Jeff Allen

Abstract Number: SESS-1790

Recent successes of HER2 antibody-drug conjugates (ADCs) have expanded patient eligibility for HER2-targeted therapy; therefore, accurate and consistent identification of patients who may benefit from ADCs is more critical than ever. Previous studies of agreement between pathologists highlight areas of discordance, but little is known about the reproducibility of assessments by emerging artificial intelligence (AI) models, particularly at low levels of HER2 expression. These models have the potential to deliver more quantitative and reproducible HER2 assessments than visual scoring by pathologists, but large-scale comparative evaluations to understand their variability are lacking.
Friends of Cancer Research created a research partnership to describe and evaluate the agreement of HER2 biomarker assessment across independently developed AI models. Both H&E and HER2 IHC whole-slide images (WSIs, N=1,124) from 733 patients diagnosed with breast cancer in 2021 were obtained from a single laboratory (ZAS Hospital, Antwerp, Belgium). Available pathology and specimen data include three pathologists' HER2 readings and details on slide processing and digitization. Ten AI models assessed HER2 status on all cases. Blinded, independent analyses were performed by statisticians from the National Cancer Institute.
Of the 10 AI models, seven used HER2 IHC WSIs, two used H&E WSIs, and one used both stains as inputs to determine HER2 score and/or status. The primary analysis focused on the seven models (6 using IHC, 1 using IHC and H&E) providing HER2 scores based on the ASCO/CAP 2018 categories (0, 1+, 2+, 3+). Absent a defined reference standard, agreement was evaluated for all possible pairings of models across all samples, resulting in a median (interquartile range, IQR) pairwise overall percent agreement (OPA) of 65.1% (60.3-69.1%) and unweighted Cohen's kappa of 0.51 (0.45-0.55). When defining binary HER2 scores as 3+ vs. not 3+, the median (IQR) pairwise agreement measures were: OPA 97.3% (95.9-97.9%), average positive agreement (APA) 87.3% (84.1-90.9%), average negative agreement (ANA) 98.5% (97.7-98.8%), and kappa 0.86 (0.82-0.90). Conversely, when defining HER2 scores as 0 vs. not 0, the median (IQR) pairwise measures were: OPA 85.6% (82.4-88.0%), APA 91.3% (87.4-92.6%), ANA 65.2% (59.9-69.7%), and kappa 0.57 (0.51-0.61). Ongoing analyses aim to assess the association of between-model agreement with patient, specimen, and model

characteristics as well as the agreement between models and pathologist readings. These findings highlight variability in HER2 biomarker scoring across models, with the least variability and a higher level of agreement in reporting 3+ cases and larger inter-model variations in evaluating HER2 low tumors, similar to agreement measures between pathologists observed in published studies. Further work is needed to understand the variability in ascribing lower HER2 scores and to evaluate performance in the context of clinical application, especially given the evolving treatment landscape and clinical implications of HER2 scores. This ongoing research partnership will enable a greater understanding of the variability in AI models and support best practices for using these models for measuring and reporting AI driven biomarker assessments in drug development and clinical practice. This dataset also has potential value for creating reference sets for future model development.



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
X @CANCERRESRCH

FRIENDS of CANCER RESEARCH