Regulatory Advancements for Patients

FRIENDS
of CANCER
RESEARCH

INTRODUCTION

PATIENT-FOCUSED DRUG DEVELOPMENT: MAXIMIZING PARTICIPATION AND REPRESENTATIVENESS IN CLINICAL TRIALS

REAL-WORLD EVIDENCE: CHARACTERIZING ENDPOINTS FOR REAL WORLD DATA CAPTURE AND SUPPORTING RAPID LEARNING

INNOVATIVE DRUG DEVELOPMENT: MODERNIZING CLINICAL TRIALS AND REGULATORY REVIEW

COMPLEX BIOMARKERS: HARMONIZING USE TO SUPPORT IMPLEMENTATION

Introduction

In our 25th year, Friends of Cancer Research (*Friends*) continues to serve as a catalyst for generating innovative science, policy, and regulatory proposals to facilitate meaningful improvements in oncology drug development, legislative and regulatory policy, and patient care. *Friends* unites scientists, experts, advocates, and patients throughout the year to generate collaborative solutions developed through working groups, roundtables, and scientific conferences.

Several recently launched research partnerships enable *Friends* to lead change by supporting the development of evidence-based solutions to current challenges. These research partnerships include projects that identify clinically useful endpoints in real-world data (RWE Pilot), investigate the use of circulating tumor DNA (ctDNA) as an endpoint to measure treatment response (ctMoniTR project), and harmonize complex biomarkers to optimize test reliability and accuracy (HRD Harmonization). This year also saw the completion of the TMB Harmonization Project, highlighted below in the Project Spotlight. Outputs from these research partnerships, in addition to our working groups, roundtables, and scientific conferences, are captured in this scientific report.

The 2021 Scientific Report contains the full text of our 2021 white papers and publications to serve as a resource for those in the drug development and regulatory space and can be found online using the QR code on the front cover.

2021 Key Themes

❶ **PATIENT-FOCUSED DRUG DEVELOPMENT**: Maximizing participation and representativeness in clinical trials

❷ **REAL-WORLD EVIDENCE**: Characterizing endpoints for real world data capture and supporting rapid learning

❸ **INNOVATIVE DRUG DEVELOPMENT**: Modernizing clinical trials and regulatory review

❹ **COMPLEX BIOMARKERS**: Harmonizing use to support implementation

2021 By the Numbers

Our impact and reach for the year

22
working
groups

400+
expert
advisors

100+
organizations

19
roundtables &
public meetings

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

16 white papers
publications **&**

Patient-Focused Drug Development: Maximizing participation and representativeness

Clinical trial eligibility criteria are carefully designed to establish an appropriate patient population that matches the intended use population of the new therapy. Overly restrictive criteria may unnecessarily exclude patients from participating in clinical trials leading to slower trial enrollment, barriers to promoting equity in research, and a potentially limited generalizability of trial results.

In 2017, *Friends* partnered with the American Society of Clinical Oncology (ASCO) to develop recommendations for broadening eligibility criteria frequently used in cancer clinical trials. These guidelines were implemented by the National Cancer Institute (NCI) to modernize the Cancer Therapy Evaluation Program (CTEP) protocol and were used to inform FDA guidance eligibility criteria for cancer clinical trials. Building on this momentum, in 2021 *Friends* and ASCO reconvened to evaluate the impact of the 2017 guidelines, identify additional opportunities for modernization, and develop recommendations for broadening other commonly restrictive eligibility criteria. In a proof of principle study, implementing guidelines for broadened eligibility criteria relating to brain metastases, organ function, primary/concurrent malignancies, HIV status, and age demonstrated the potential to double the number of advanced non-small cell lung cancer patients eligible to participate in clinical trials.

- Several eligibility criteria commonly used in cancer clinical trials unnecessarily restrict patients from participating in research, and broadening these criteria provides access to more patients

Broadening common eligibility criteria has potential to **double the number of eligible lung cancer patients** for participation in clinical trials

17-21%
patients
are unable to enroll in trials due to restrictive eligibility criteria

SOURCE: CLIN CANCER RES. 2021;27(9):2394-2399

Real-World Evidence: Characterizing endpoints for real world data capture and supporting rapid learning

Clinical trials are often composed of homogenous patient populations that are not always reflective of the broader patient population eligible to receive the drug. Real-world data (RWD), which is data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources, can be used to generate real-world evidence (RWE) to help evaluate the effectiveness of therapies in broader patient populations. FDA released four guidance documents in 2021 related to the use of RWD and RWE in regulatory decision-making, however, alignment in how this information is collected and analyzed is needed to fully realize its potential.

To aid alignment on the use of RWD, *Friends* established multi-stake-holder partnerships to determine how RWD can be leveraged to support drug development. Our recent publications demonstrate the ability to align on measures of real-world endpoints to capture critical information on patient outcomes from RWD sources. Work from our RWE Pilot 2.0 highlights alignment on a methodological framework for rw-endpoints allowing for the evaluation of treatment outcomes in patients with advanced non-small cell lung cancer. *Friends* was able to quickly apply this knowledge as part of a multi-stakeholder collaboration to assess the effectiveness of certain treatments for COVID-19 patients in the real-world setting. These findings will advance future use of RWD to generate meaningful evidence and inform future policy decisions.

- Use of a common methodological framework among data partners allows for generation of high-quality RWE to support robust assessment of patient outcomes

- Diverse RWD sets can be harmonized to achieve harmonized measures of rw-endpoints that correlate with endpoints commonly used in clinical trials

4
FDA guidance documents related to RWE/RWD released in 2021

142
regulatory submissions contained RWD in 2011–2020

SOURCE: BLOOD. 2021;138(SUPPLEMENT 1):5037

50% women

27% people of color

44% 65+ years

Representativeness of clinical trials supporting oncology approvals in 2020 – 50% of all participants were women, 73% were White, 5% were Black or African American, 14% were Asian, 6% were Hispanic, 44% were 65 years and older, and 41% were from sites in the United States
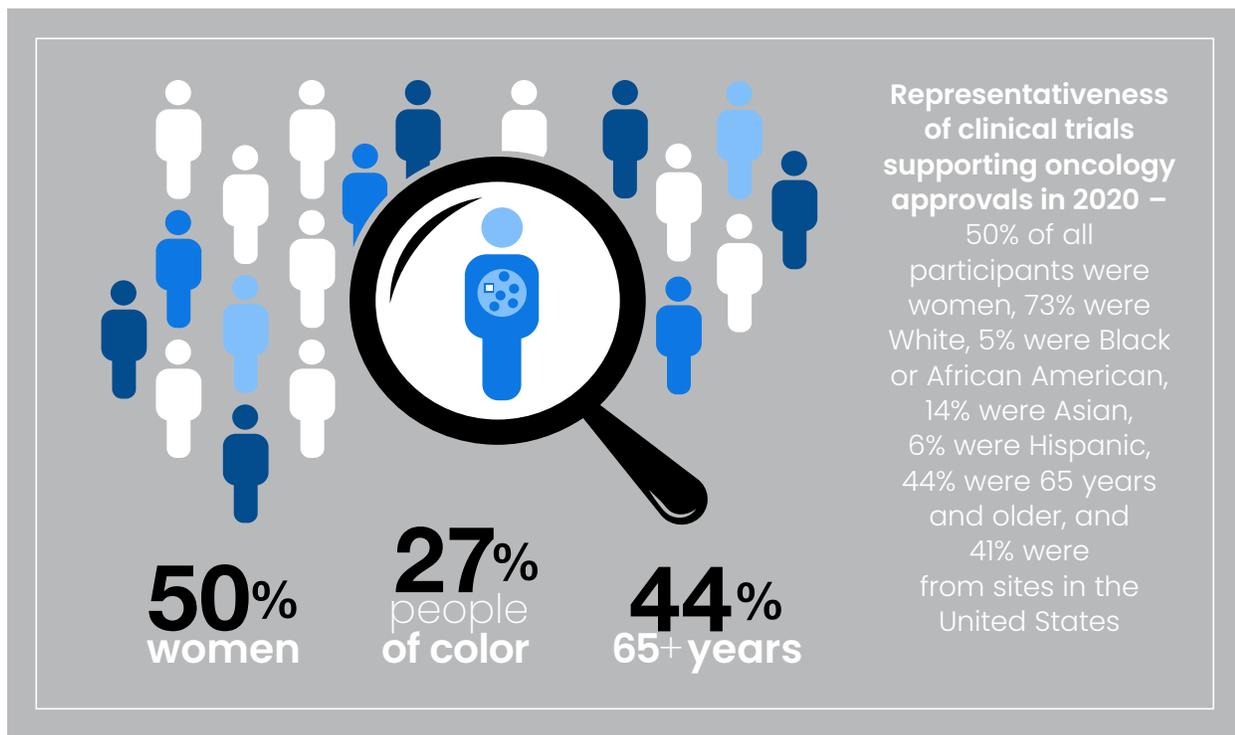
SOURCE: FDA. 2020 DRUG TRIALS SNAPSHOTS: SUMMARY REPORT

Innovative Drug Development: Modernizing clinical trials and regulatory review

The landscape of oncology therapies continues to evolve with emerging targets and novel technologies. While these innovations help to improve patient outcomes, they can be coupled with increased complexity in drug development and potential novel toxicities. In 2021, *Friends* hosted several roundtable discussions with key stakeholders to develop consensus recommendations that address some of these emerging challenges.

In the US, FDA reviews drug applications to determine if the benefits of the therapy for a specific indication outweigh the risks. Once reviewed and approved, the drug can be made available to the public. Drug development is a time-consuming process, but expedited programs like Breakthrough Therapy Designation (BTD) can help get promising treatments to patients faster. An analysis by *Friends* demonstrated that BTD has shortened the time to approval for new oncology products by a median of 2.8 years compared to those approved without a BTD since its inception in 2012. An understanding of the value of this pathway and how it can be improved supports continued benefit to patients.

100%

**of first-
in-class
oncology** drug
approvals
**2013-2021
were
expedited**

90%

**of new
oncology
approvals** had an
**expedited
approval**

In addition to supporting pathways that allow for innovative drugs to be approved more quickly, in 2021 *Friends* supported work that helps modernize clinical trials. Novel therapies may be accompanied by toxicities that are not well characterized. Cytokine release syndrome (CRS), which is an emerging toxicity associated with CAR-T therapies, was used as a case study to support efforts to align definitions and data capture to support management strategies and clinical guideline development. Another opportunity with new drug development is optimizing drug dosing to minimize or prevent unnecessary side effects. Historically, the Maximum Tolerated Dose (MTD), or the highest dose that has tolerable side effects, has been considered the optimal dose. With the advent of targeted therapy, a lower dose may be optimal for patients and may have fewer side effects.

- Clarifying the data necessary to receive BTD and increasing coordination within and between various departments at FDA and the sponsor's organization may facilitate more optimal use of the pathway

- To support the evolving understanding and more effective evidence-based risk monitoring and patient care, there should be consistency in evaluating and reporting novel toxicities, such as CRS

- Opportunities and strategies for improving dose-finding studies in oncology will improve patient outcomes

expedited
pathways
3.9
years
sooner

traditional
10.1 years to approval

expedited
6.2 years to approval

Complex Biomarkers: Harmonizing use to support implementation

Increasingly in oncology, diagnostic tests are used to improve patient care. Many innovative therapies focus on specific targets, which often require the use of a diagnostic test to identify the appropriate patient population who may benefit from use. Other diagnostic tests are used to track patient response to treatment and treatment outcomes. However, diagnostic tests often use different technologies, definitions, and methodological approaches for measuring biomarkers, which can lead to inconsistent results and negatively impact patient care.

Friends works to ensure diagnostic tests are harmonized for effective implementation and optimal use in drug development and clinical care. In 2021, *Friends* published the results of the final phase of the Tumor Mutational Burden (TMB) Harmonization Project (see project spotlight, pg. 10), calling attention to the variability across assays measuring TMB and identifying solutions for harmonization, including the development of a calibration tool to help optimize development of TMB assays. In 2021, *Friends* also convened a working group to discuss strategies to support current and future use of circulating tumor DNA (ctDNA) as a drug development tool for regulatory decision making in early stage cancer.

- Harmonization of biomarkers is critical for drug development and regulatory decision-making to support improved patient care

Number of Clinical Trials that Included ctDNA Evaluation

Year	Number
2012	2
2015	11
2018	30
2021	37

Year Registered to ClinicalTrials.gov
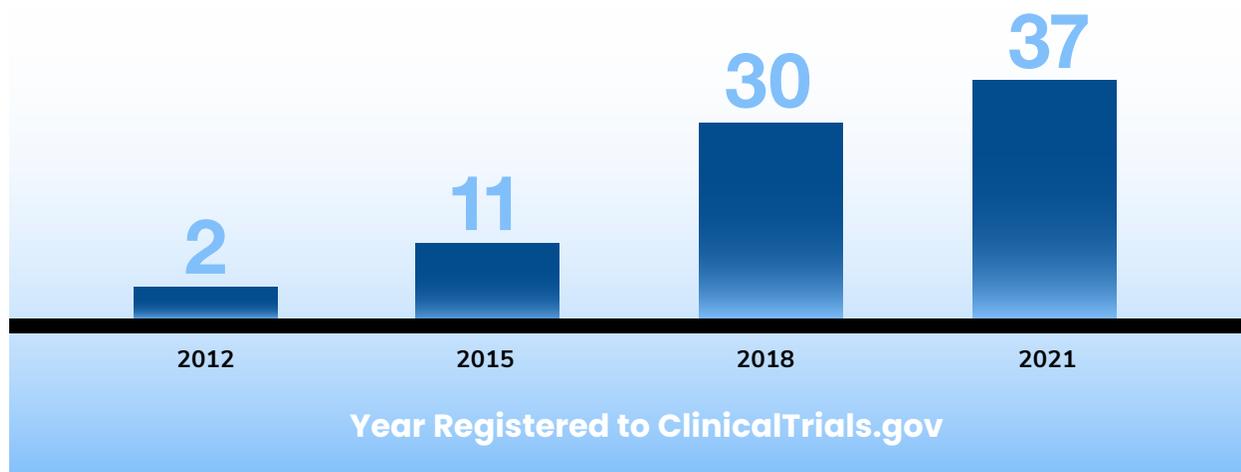
Project Spotlight: *Friends* TMB Harmonization Project

The TMB Harmonization Project is a collaborative effort that establishes an approach for harmonizing diagnostic tests to provide consistent identification of patients who are likely to respond to certain therapies.

Background
Tumors with a high number of mutations are more sensitive to immunotherapy (IO): the higher the number of tumor mutations, the better the patient's outcome. Therefore, measuring the tumor mutational burden (TMB) helps to identify which patients may benefit from treatment with IO. However, different methods and technologies are used to determine TMB, which can result in variability TMB measurements and reporting. To optimize the use of TMB assays and ensure patients and providers receive accurate and reliable information to make appropriate treatment decisions, identifying sources of discordance and developing best practices to support assay alignment is critical.

Approach
Friends initiated a unique collaboration with key stakeholders including pharmaceutical companies, diagnostics developers, FDA, and academics in September 2017 to discuss variability in how TMB is defined, analyzed, and used in clinical practice, and the need for establishing industry standards. Over the next six months, *Friends* hosted discussions with project participants to develop consensus on a methodological approach to compare TMB assays, to develop a calibration tool to promote reproducibility and comparability across assays, and to provide recommendations for a clinical cutoff to support evaluation of TMB for clinical trial enrollment using a common strategy. Throughout the duration of the project, the group presented findings through public workshops, conferences, and manuscripts.
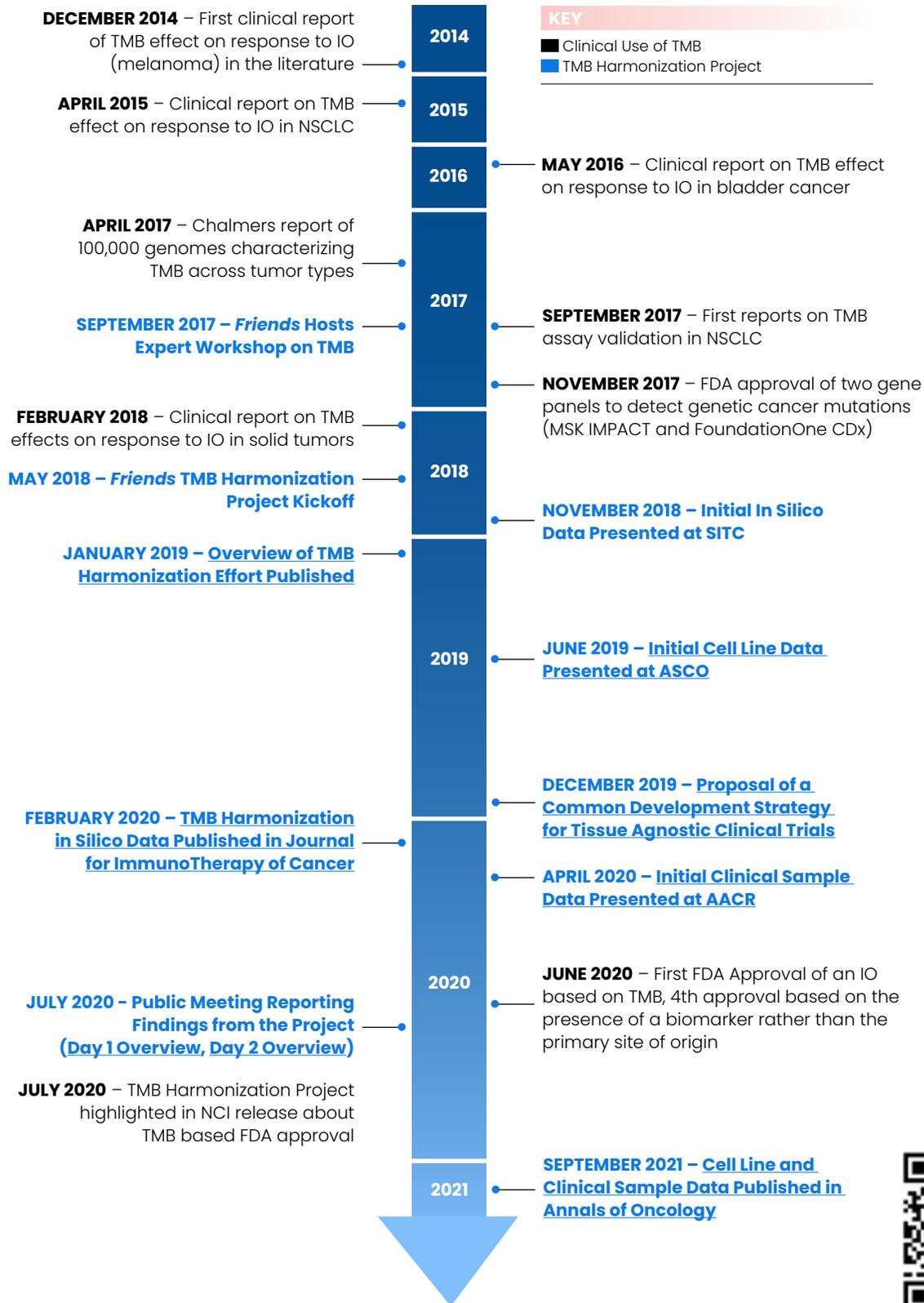
Findings
Diagnostic developers used data from The Cancer Genome Atlas (TCGA), matched normal-tumor cell lines, and tumor samples to calculate TMB in each sample using their own analysis pipeline. An agreed upon method for estimating TMB was established as the "gold-standard", which enabled comparisons across assays and to the gold standard. The analysis showed that as TMB values increased, so did variability between TMB assays. The group developed a publicly available calibration tool that can assist with assay development, reduce variability, and facilitate interpreting data from across different studies. Additionally, the group agreed on a lower bound cutoff of 10 mutations/megabase should be considered when evaluating TMB for clinical trial enrollment in a pan-tumor indication.

Next Steps
The findings from the TMB Harmonization Project demonstrate the power of a multi-stakeholder project to support assay harmonization. As more complex biomarkers become more routine in oncology care, the non-systematic approach to biomarker development can lead to challenges for regulators, payors, patients, and physicians. This work set the foundation for an ongoing *Friends*' project, the Homologous Recombination Deficiency (HRD) Harmonization Project and provides a foundation to support efforts to modernize diagnostic regulations at the FDA.

TMB Development and Milestones Timeline

DECEMBER 2014 – First clinical report of TMB effect on response to IO (melanoma) in the literature

2014

APRIL 2015 – Clinical report on TMB effect on response to IO in NSCLC

2015

2016

MAY 2016 – Clinical report on TMB effect on response to IO in bladder cancer

APRIL 2017 – Chalmers report of 100,000 genomes characterizing TMB across tumor types

2017

SEPTEMBER 2017 – *Friends* Hosts Expert Workshop on TMB

SEPTEMBER 2017 – First reports on TMB assay validation in NSCLC

NOVEMBER 2017 – FDA approval of two gene panels to detect genetic cancer mutations (MSK IMPACT and FoundationOne CDx)

FEBRUARY 2018 – Clinical report on TMB effects on response to IO in solid tumors

2018

MAY 2018 – *Friends* TMB Harmonization Project Kickoff

NOVEMBER 2018 – Initial In Silico Data Presented at SITC

JANUARY 2019 – Overview of TMB Harmonization Effort Published

2019

JUNE 2019 – Initial Cell Line Data Presented at ASCO

DECEMBER 2019 – Proposal of a Common Development Strategy for Tissue Agnostic Clinical Trials

FEBRUARY 2020 – TMB Harmonization in Silico Data Published in Journal for ImmunoTherapy of Cancer

APRIL 2020 – Initial Clinical Sample Data Presented at AACR

2020

JUNE 2020 – First FDA Approval of an IO based on TMB, 4th approval based on the presence of a biomarker rather than the primary site of origin

JULY 2020 - Public Meeting Reporting Findings from the Project (Day 1 Overview, Day 2 Overview)

JULY 2020 – TMB Harmonization Project highlighted in NCI release about TMB based FDA approval

2021

SEPTEMBER 2021 – Cell Line and Clinical Sample Data Published in Annals of Oncology

FRIENDS of CANCER RESEARCH

Continuing to Broaden Eligibility Criteria to Make Clinical Trials More Representative and Inclusive: ASCO–Friends of Cancer Research Joint Research Statement

Edward S. Kim[1], Thomas S. Uldrick[2], Caroline Schenkel[3], Suanna S. Bruinooge[3], R. Donald Harvey[4], Allison Magnuson[5], Alexander Spira[6], James L. Wade[7], Mark D. Stewart[8], Diana Merino Vega[8], Julia A. Beaver[9], Andrea M. Denicoff[10], Gwynn Ison[9], S. Percy Ivy[10], Suzanne George[11], Raymond P. Perez[12], Patricia A. Spears[13], William D. Tap[14], and Richard L. Schilsky[3]

ABSTRACT

Purpose: Restrictive clinical trial eligibility criteria (EC) limit the number of patients who can enroll and potentially benefit from protocol-driven, investigational treatment plans and reduce the generalizability of trial results to the broader population. Following publication of expert stakeholder recommendations for broadening EC in 2017, the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*) convened working groups to produce additional recommendations and analyze the potential impact on clinical trials using real-world data.

Experimental Design: Multistakeholder working groups were appointed by an ASCO-*Friends* leadership group to propose recommendations for more inclusive EC related to: washout periods, concomitant medications, prior therapies, laboratory reference ranges and test intervals, and performance status.

Results: The four working groups, ASCO Board of Directors, and *Friends* leadership support the recommendations included in this statement to modernize EC related to washout periods, concomitant medications, prior therapies, laboratory references ranges and test intervals, and performance status to make trial populations more inclusive and representative of cancer patient populations.

Conclusions: Implementation of the recommendations is intended to result in greater ease of determining patient eligibility. Increased opportunities for patient participation in research will help address longstanding underrepresentation of certain groups in clinical trials and produce evidence that is more informative for a broader patient population. More patients eligible will also likely speed clinical trial accrual.

See related commentary by Giantonio, p. 2369

Introduction

Accelerating advances in cancer treatment requires efficient clinical trials that produce clinically meaningful outcomes and generalizable knowledge. Clinical trials are not possible without patients, whose eligibility to participate is determined by inclusion and exclusion criteria. Trial eligibility criteria (EC) are designed to protect participant safety and define an appropriate study population. Following approval, patient safety may be compromised if a trial generates insufficient evidence to inform care for specific patient groups, for example, those underrepresented among trial participants. Furthermore, restrictive

EC limit clinical treatment options for patients who weigh the potential risks, benefits, and alternatives of a protocol-driven investigational treatment plan and opt to participate in studies.

Exclusion of certain patient populations or disease characteristics is common in oncology clinical trials and is often not founded on current evidence-based scientific justification. This leads to underrepresentation of older adults (1), racial/ethnic (2–4) and sexual/gender minorities (5–7), and patients with well-managed comorbidities (8). An estimated 17%–21% of patients are not able to enroll on clinical trials due to restrictive EC, among other reasons (9, 10). In the era of biomarker-driven therapies where the pool of potential study participants may be very low due to low biomarker prevalence, the negative impact of excessively restrictive EC is magnified (11).

The desire to mitigate safety concerns and ensure trial integrity is paramount, but EC are often replicated from earlier trials and may date back to concerns about cytotoxic chemotherapy. A 2017 review by the FDA concluded that clinical trial EC can be expanded without compromising patient safety (12). To ensure that only criteria relevant to safety concerns about the specific agent are included and extraneous EC are excluded, scientific rationale should be included to justify any exclusion criteria.

ASCO-*Friends* Eligibility Criteria Initiative

Eliminating overly restrictive EC is a priority for the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*), as well as many other patient groups (such as the American Cancer Society Cancer Action Network), researchers, sponsors, regulators, and the National Academy of Medicine (9, 13–18). Enacting

[1]City of Hope Orange County and National Medical Center, Los Angeles, California. [2]Fred Hutchinson Cancer Research Center, Seattle, Washington. [3]American Society of Clinical Oncology, Alexandria, Virginia. [4]Emory University School of Medicine, Atlanta, Georgia. [5]University of Rochester, Rochester, New York. [6]Virginia Cancer Specialists PC, Fairfax, Virginia. [7]Cancer Care Specialists of Illinois, Decatur, Illinois. [8]Friends of Cancer Research, Washington, DC. [9]U.S. Food and Drug Administration, Silver Spring, Maryland. [10]National Cancer Institute, Bethesda, Maryland. [11]Dana-Farber Cancer Institute, Boston, Massachusetts. [12]Bristol-Myers Squibb, Lawrenceville, New Jersey. [13]UNC Patient Advocates for Research Council, Raleigh, North Carolina. [14]Memorial Sloan Kettering Cancer Center, New York, New York.

E.S. Kim and T.S. Uldrick contributed equally to this article.

Corresponding Author: Edward S. Kim, City of Hope National Medical Center, 1500 East Duarte Road, Los Angeles, CA 91010. Phone: 626-866-0478; Fax: 626-765-3379; E-mail: edwkim@coh.org

changes will optimize trial enrollment and ensure that benefits to patients and the broader scientific community are maximized. In addition, broadening EC is desirable to improve accrual and prevent trial delays and failures, which are a significant strain on human and financial resources during development of new therapies (19–21).

Through this work, ASCO and *Friends* propose a new cancer clinical trial paradigm, in which:

(i) Patients are eligible for a trial by default and excluded only when there is scientific rationale and/or evidence demonstrating that enrollment would compromise the patient's safety.
(ii) In all cases, protocol development begins with informed consent as the only eligibility criteria. Any inclusion/exclusion criteria are tailored to the scientific objectives of the study, based on the investigational treatment and study population, and address only substantiated participant risks.
(iii) Trial participants more closely resemble the population intended to receive the therapy and no group is excluded without scientific justification based on current evidence.

ASCO, *Friends*, and FDA first formed a collaboration to address overly restrictive cancer clinical trial EC in 2016, which led to publication of recommendations for more inclusive EC for brain metastases, minimum age for enrollment, human immunodeficiency virus (HIV) status, organ dysfunction, and prior or concurrent malignancies (13–17).

In 2019, project leadership consulted with stakeholder experts, including ASCO's Cancer Research and Health Equity Committees, to select additional categories of common EC that pose significant barriers to clinical trial enrollment. These topics were selected with an eye for how many patients they impact and how they affect special populations, as well as their potential impact on evaluation of safety and efficacy if relaxed.

Representatives from academic and community research sites, regulatory agencies (FDA and NCI), patient advocacy groups, NCI Network Groups, and the pharma-biotech industry were invited to join the project work groups. The work groups finalized their consensus recommendations after convening with additional patient and industry representatives to discuss their draft recommendations.

ASCO and *Friends* herein recommend broadening approaches to clinical trial enrollment related to the following five EC:

(i) Washout periods
(ii) Concomitant medications
(iii) Prior therapies
(iv) Laboratory reference ranges and test intervals
(v) Performance status (PS)

ASCO-*Friends* Recommendations

This statement provides a high-level summary of additional ASCO-*Friends* recommendations for more inclusive clinical trial EC (**Table 1**). Detailed discussion of each recommendation and supporting rationale is presented in separate manuscripts.

There are three common themes across these recommendations. First, clinical trial designers should launch every trial with a goal of inclusion and should add exclusions only where safety concerns warrant exclusion of patients with certain characteristics. Protocols should be living documents; that is, over the course of new agent development from first-in-human through phase III studies, EC should be examined critically and revised to allow for the enrollment of patients who may have previously been excluded because of safety concerns, but for whom new information provides sufficient evidence to support their inclusion.

Second, inclusion of all populations who are anticipated to benefit from the therapy based on the mechanism of action early in clinical development is both equitable and necessary. This will ensure that patients who may ultimately benefit from the treatment being studied are not excluded because of lack of safety data for that population. If representative populations are not included, dose, tolerance, risk of adverse events, and therapeutic benefit remain unknown. The inclusion of exploratory cohorts with broader eligibility in early-phase trials will help to inform and enable revisions to the protocol EC based on these earlier risk-benefit analyses. These exploratory cohorts should help sponsors strike a balance between more rapid patient accrual with broader criteria, time associated with enacting protocol amendments later in development, and number of postmarketing requirements and commitments to expedite trial completion and submission of more complete study findings to regulatory agencies, ultimately leading to broader knowledge in clinical use. At minimum, participants in trials leading to marketing authorization should be inclusive of the patients in the intended use population.

Finally, study design should consider both internal and external validation. In phase I studies, safety is paramount and EC are based on existing knowledge. More stringent EC may also be appropriate in early-phase studies conducted to establish principles of management or to explore a biological question. Including an exploratory cohort in early-phase trials through broadened EC will provide safety information to expand participation in the next phase of study. Registration trials can include participants that resemble the entire population of patients who may use the therapy after approval more closely, that is, improving external validity. Including broader populations also helps fulfill the principle of distributive justice, ensuring appropriate representation of groups who are underrepresented in research, where safety permits.

Washout periods

A washout period is a time between most recent treatment and trial enrollment that is intended to prevent confounding the interpretation of the effect of a new treatment by a persistent effect of an immediately prior

Table 1. Summary of Work Group Recommendations.

Eligibility criteria category	Recommendation
Washout periods	1. Time-based washout periods should be removed from protocol eligibility criteria in most cases. Any inclusion of time-based washout periods should be scientifically justified and clearly specified. 2. Relevant clinical and laboratory parameters should be used in place of time-based washout periods to address safety considerations. 3. Potential trial participants should have recovered from clinically significant adverse events of their most recent therapy/intervention prior to enrollment.
Concomitant medications	1. Concomitant medications use should only exclude patients from trial participation when clinically relevant known or predicted drug–drug interactions or potential overlapping toxicities will impact safety or efficacy.

Prior therapies	1. Patients are eligible for clinical trials regardless of the number or type of prior therapies and without a requirement to have received a specific therapy prior to enrollment unless a scientific or clinically based rationale is provided as justification.
	2. Prior therapy (either limits on the number and type of prior therapies or requirements for specific therapies before enrollment) could be used to determine eligibility in the following cases:
	  a. If the agents being studied target a specific mechanism or pathway that could potentially interact with a prior therapy.
	  b. If the study design requires that all patients begin protocol-specified treatment at the same point in the disease trajectory.
	  c. In randomized clinical studies, if the therapy in the control arm is not appropriate for the patient due to previous therapies received.
	3. Trial designers should consider conducting evaluation separately from the primary endpoint analysis for participants who have received prior therapies.

Laboratory reference ranges and test intervals	1. Laboratory test results should only be used as exclusion criteria when scientifically justified and when abnormal test results confer safety concerns. 2. Laboratory reference values should account for potential normal variations due to race, ethnicity, age, sex, and gender identity (i.e., due to surgical and/or hormonal changes). 3. Routine reassessment of laboratory test-based exclusion criteria should be conducted during the course of clinical research and drug development as investigational agents progress from earlier- to later-phase clinical trials. 4. Increasing the intervals between protocol-specified tests should be considered to help reduce patient burden and increase ability to rely on routine clinical testing, especially in later cycles of treatment and over the evolution of the protocol from earlier- to later-phase clinical trials.

Performance status	1. Patients with reduced PS (e.g., ECOG PS 2) should be included unless there is a scientific and/or clinical rationale for exclusion justified by established safety considerations.
	  a. ECOG PS eligibility criteria should be based on the patient population in which the intervention is expected to be used in clinical practice.
	  b. PS eligibility criteria should be continually reevaluated and modified throughout the clinical development process to reflect accumulated safety data of the investigational treatment. Decisions about PS eligibility criteria should be based on early clinical safety and efficacy data about the specific investigational agent or based on known data from other drugs in the same class with similar mechanism of action. Later-phase trials (e.g., phase II/III) should generally mirror the intended use population and ECOG PS 2 patients should be included, unless safety concerns have manifested in earlier-phase trials. The rationale for exclusion should be justified and stated explicitly.
	  c. Incorporating the rationale for inclusion of a broader population into the protocol could help encourage investigators to enroll these patients.
	  d. Performance status data should still be collected for use as a stratification factor, regardless of how it is incorporated into eligibility criteria.
	2. Consider alternate trial designs, such as prespecified cohorts with lower PS that are exempt from the primary analysis, to encourage inclusion of these patients. These cohorts would generally be small in size and exploratory in nature and could be enrolled in an incremental way to enable an early stopping rule based upon safety data. Consideration of the data analysis approach for the broader eligibility cohort and subgroup analysis should be determined during the study design phase. Early discussion with FDA about enrollment of a broader population may have implications for marketing and post-marketing research requirements.
	3. Additional assessments of functional status should be considered to better characterize the functional status of ECOG PS 2 patients and patients ages ≥65, such as activities of daily living (ADLs) and instrumental ADLs.

treatment. Washout/waiting time periods prior to enrollment are common for all modalities of cancer treatment. In many cases, washout periods are associated with theoretical concerns (e.g., prevention of untoward adverse events, drug interactions, and incorrect adverse event attribution) that lack scientific rationale and/or are clinically irrelevant.

Concomitant medications

On average, patients with cancer take five chronic noncancer medications, in addition to drugs that manage adverse effects of their cancer treatment (22). Exclusion of concomitant medications during trials is intended to prevent adverse drug interactions that may affect pharmacokinetic assessment or patient safety, reduce the risk of drug-related adverse events, and, rarely, prevent the use of drugs that are known or predicted to antagonize the anticancer efficacy of investigational therapies. While some medications may be necessarily prohibited early in the development of an investigational agent while knowledge is gained, persistent prohibition reduces the applicability of a therapy to a broader population of patients both in trials and following approval.

Prior therapies

Many cancer trial protocols disallow patients based upon receipt of previous cancer-directed therapies. This may take the form of blanket EC (e.g., any history of prior therapy excluded) or conditional criteria (e.g., specific treatments or a specified number or type of prior treatment lines excluded). In other situations, particularly earlier in drug development, clinical trials commonly exclude patients if they have not received a specific therapy prior to enrollment. Improved molecularly driven therapies and immunotherapies may alter the risk-benefit consideration of study participation in relation to treatment with standard therapies with low efficacy or high toxicity, and in some cases participation in a clinical trial without a requisite receipt of prior standard-of-care therapy may be warranted with appropriate informed consent. As with any other EC, clinical trial designers and sponsors should rigorously justify any restrictions based on prior therapies.

Laboratory reference ranges and test intervals

Laboratory tests that predict and assess toxicity are critical for determining whether a patient can safely enroll on a clinical trial. However, some laboratory reference ranges and test intervals that are included as trial EC are arbitrary, with minimal justification for their use, particularly for investigations of targeted therapies and immunotherapies that may have more favorable or unique toxicity profiles. Reference ranges or intervals that lack scientific rationale and/or differ from routine clinical care often result in biased clinical trial outcomes (as healthier, more homogeneous trial participants may not represent the patients actually treated with a drug once it is approved) and may hinder clinical trial accrual. In addition, nonroutine testing, requirements for central testing, and/or strict adherence to time intervals often increase trial expenses for participants, sponsors, and research sites, and may increase risks associated with certain tests and biopsies (23). Because each clinical trial has distinct therapies with differing toxicity and pharmacokinetic considerations, it is not feasible to provide specific laboratory test value thresholds for broad applicability. Nevertheless, incorporation of principles in **Table 2** may help ensure safety, while minimizing unnecessary participant exclusions.

PS

PS is one of the most common EC utilized in oncology, with many trials limited to patients with good PS [i.e., Eastern Cooperative Oncology Group (ECOG) PS 0 or 1; ref. 13]. This practice restricts therapeutic options for a significant proportion of patients (12), contributes to the pervasive age disparity observed in oncology clinical trials (24), and limits the generalizability of research results in clinical practice. PS as an eligibility criterion should be reconsidered to be more inclusive while maintaining patient safety and study integrity.

Discussion

ASCO and *Friends* are engaged in additional activities to maximize the likelihood that these recommendations are implemented and representative participant populations are accrued to trials.

Our strategies involve four primary elements:

(i) Dissemination—Stakeholders are aware of the EC recommendations and endorse the new cancer clinical trial paradigm outlined above.
(ii) Implementation—More inclusive EC are incorporated into cancer clinical trial protocols.

(iii) Equity—Investigators discuss clinical trial participation with all patients who would qualify and seek to enroll all eligible participants.
(iv) Evaluation—Clinical trial sponsors and investigators monitor the impact of implementing the recommendations, continuously assess accrual during clinical trial conduct to address any challenges that may delay efficient enrollment and completion, and identify additional opportunities to broaden EC to ensure that cancer clinical trial populations mirror the entire population who will be prescribed the treatment.

In efforts to broaden EC, ASCO and *Friends* gathered feedback, reviewed evidence, and conducted analysis of the most common and restrictive criteria. An analysis of 21 Southwest Oncology Group studies showed that 60% of EC are related to comorbidities (including prior treatment exclusions, prior malignancy exclusions, PS, organ function status, HIV status, and brain metastases, among other criteria; ref. 25). Recommendations in this statement and the previous ASCO-*Friends* statement address all of these EC (13).

Research suggests that adoption of the 2017 ASCO-*Friends* recommendations could lead to more inclusive protocols. Data presented at the 2019 ASCO annual meeting demonstrated in a cohort of 10,500 patients with advanced non–small cell lung cancer that implementation of ASCO-*Friends* recommendations could avoid exclusion of nearly half the cohort due to broadened inclusion criteria for brain metastases, prior/concurrent malignancies, and/or reduced kidney function (26).

Publication of these recommendations and analysis of their potential impact will accomplish little if protocols are not updated and investigators do not enroll representative participant populations. Support from trial sponsors, physician investigators, institutional review boards, contract research organizations, and research staff is essential to ensuring that broadened EC are applied appropriately. Eligibility for clinical trials should be recognized as a distributive justice issue for individual patients and for vulnerable populations (27). To the fullest extent possible, FDA, NCI, NIH, and other regulatory bodies, and sponsors should leverage the incentives for broader enrollment that they can offer.

ASCO and *Friends* have partnered with various stakeholders to disseminate and encourage implementation, including working closely with FDA, NCI, and NCI Network Groups. FDA finalized four guidance documents in July 2020 to encourage sponsors to apply the 2017 ASCO-*Friends* recommendations (28–31). NCI revised its protocol template to incorporate the recommendations, including implementation in active protocols and future NCI-funded trials (32).

The general EC in ASCO's TAPUR (Targeted Agent and Profiling Utilization Registry) study mirrors ASCO-*Friends* recommendations by not excluding patients who: are 12 years and older; have new or progressive brain metastases or previously treated or untreated brain metastases, if they are clinically stable; have a prior malignancy; are HIV+; and/or are ECOG PS 0–2. For biomarker-selected therapies, the biomarker driving the cancer should be the primary inclusion criteria, as these therapies often do not pose the same risks as cytotoxic chemotherapy.

Conclusions

EC for washout periods, concomitant medications, prior therapies, laboratory references ranges and test intervals, and PS can and should be modernized to be inclusive of broader, more representative patient populations. These considerations, along with previously proposed

Table 2. Benefits and risks/challenges of expanded eligibility criteria (Adapted from Kim and colleagues, 2017).

Benefit and risk/challenge	Patients	Physicians	Sponsors and investigators
Benefits	Earlier access to investigational agents and expanded trial and treatment options	More complete safety data, which can inform clinical use and enable safe delivery if investigational agent becomes commercially available	Ability to generalize to real-world patients and potentially reduce postmarketing requirements; efficacy in traditionally understudied population(s) could potentially result in expanded marketing claims and provide a differentiating factor between drugs of same class
	Increased confidence in treatment decision-making due to availability of efficacy and safety (i.e., side effect) data from a representative group of trial participants	Availability of efficacy and safety data informs weighing of available treatment options across a broader array of patients and increases confidence in therapy selection	Quicker accrual, fewer trial delays and failures, and more patients may be eligible at each site. All these factors may also reduce cost and time of clinical trial conduct.
	If early trial data in expanded populations demonstrates concerns with efficacy or safety, future patients will have better information to avoid more toxic or less efficacious therapies or know how to modify therapy delivery to avoid toxicities.	Earlier identification of drugs that may not be efficacious in a specific patient population or that may cause more harm than good or earlier knowledge about dose modification of an investigational therapy to improve efficacy or safety/tolerability	Identification of potential safety issues earlier during closely monitored clinical trials may facilitate earlier development of mitigation strategies, enabling broader uptake after approval, and avoidance of post-marketing harms in a larger number of patients due to length of time required for the passive, postmarketing safety surveillance system to identify safety concerns
Risks/challenges	Patients with comorbidities may have a potentially higher risk of experiencing an adverse event as a result of the investigational drug or their disease	Limited data from small cohorts enrolled with broadened criteria may not be adequate for clinical decision-making	More variability in outcomes may require larger sample sizes and inferences may not be as precise
	Additional procedures for increased safety monitoring in some situations may incur additional costs to patients	Additional procedures for increased safety monitoring in some situations may incur additional costs and increased complexity of patient care	Potential safety concerns may require separate cohorts or analysis plans and early stopping rules for excess toxicity
		Additional resources may be required to ensure staff are able to manage safety monitoring	May complicate attribution of adverse events
			Increased costs associated with additional cohorts, statistical requirements, additional testing, additional data for analysis, or special expertise to manage specific patient needs

modifications, may result in greater efficiency of trial conduct and faster clinical trial accrual. Implementation will increase opportunities for patient participation and generation of generalizable evidence to better inform use of new therapies in populations encountered in clinical practice.

Authors' Disclosures

Disclaimer

The opinions expressed in this article are those of the authors and do not necessarily reflect the views or policies of the authors' affiliated institutions.

Acknowledgments

References

1. Abbasi J. Older patients (still) left out of cancer clinical trials. JAMA 2019;322: 1751–3.
2. Vastola ME, Yang DD, Muralidhar V, Mahal BA, Lathan CS, McGregor BA, et al. Laboratory eligibility criteria as potential barriers to participation by black men in prostate cancer clinical trials. JAMA Oncol 2018;4:413–4.
3. 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.
4. Duma N, Vera Aguilera J, Paludo J, Wang Y, Leventakos K, Mansfield AS, et al. Representation of minorities in oncology clinical trials: review of the past 14 years. J Oncol Pract 2017;14:e1–e10.
5. Griggs J, Maingi S, Blinder V, Denduluri N, Khorana AA, Norton L, et al. American Society of Clinical Oncology position statement: strategies for reducing cancer health disparities among sexual and gender minority populations. J Clin Oncol 2017;35:2203–8.
6. Hamel LM, Penner LA, Albrecht TL, Heath E, Gwede CK, Eggly S. Barriers to clinical trial enrollment in racial and ethnic minority patients with cancer. Cancer Control 2016;23:327–37.
7. Obedin-Maliver J. Time to change: supporting sexual and gender minority people-an underserved, understudied cancer risk population. J Natl Compr Canc Netw 2017;15:1305–8.
8. Unger JM, Hershman DL, Fleury ME, Vaidya R. Association of patient comorbid conditions with cancer clinical trial participation. JAMA Oncol 2019;5:326–33.
9. American Cancer Society Cancer Action Network. Barriers to patient enrollment in therapeutic clinical trials for cancer; 2018. Available from: https://www.fightcancer.org/sites/default/files/National%20Documents/Clinical-Trials-Landscape-Report.pdf.
10. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. J Natl Cancer Inst 2019; 111:245–55.
11. Kim ES, Bernstein D, Hilsenbeck SG, Chung CH, Dicker AP, Ersek JL, et al. Modernizing eligibility criteria for molecularly driven trials. J Clin Oncol 2015; 33:2815–20.
12. Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. J Clin Oncol 2017;35:3745–52.
13. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.
14. Uldrick TS, Ison G, Rudek MA, Noy A, Schwartz K, Bruinooge S, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research HIV working group. J Clin Oncol 2017;35:3774–80.
15. Lin NU, Prowell T, Tan AR, Kozak M, Rosen O, Amiri-Kordestani L, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research brain metastases working group. J Clin Oncol 2017;35:3760–73.
16. Gore L, Ivy SP, Balis FM, Rubin E, Thornton K, Donoghue M, et al. Modernizing clinical trial eligibility: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research minimum age working group. J Clin Oncol 2017;35:3781–7.
17. Lichtman SM, Harvey RD, Damiette Smit M-A, Rahman A, Thompson MA, Roach N, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research organ dysfunction, prior or concurrent malignancy, and comorbidities working group. J Clin Oncol 2017;35:3753–59.
18. Nass SJ, Moses HL, Mendelsohn John, editors. A national cancer clinical trials system for the 21st century: reinvigorating the NCI cooperative group program, in institute of medicine. Washington (DC): National Academies Press; 2010.
19. Sertkaya A, Birkenbach A, Berlind A, Eyraud J. Examination of clinical trial costs and barriers for drug development; 2014. Available from: https://aspe.hhs.gov/system/files/pdf/77166/rpt_erg.pdf.
20. Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N, et al. Adult cancer clinical trials that fail to complete: an epidemic? J Natl Cancer Inst 2014;106:dju229.
21. Malik L, Lu D. Eligibility criteria for phase I clinical trials: tight vs loose? Cancer Chemother Pharmacol 2019;83:999–1002.
22. Turner JP, Shakib S, Singhal N, Hogan-Doran J, Prowse R, Johns S, et al. Prevalence and factors associated with polypharmacy in older people with cancer. Support Care Cancer 2014;22:1727–34.
23. Winkfield KM, Phillips JK, Joffe S, Halpern MT, Wollins DS, Moy B. Addressing financial barriers to patient participation in clinical trials: ASCO policy statement. J Clin Oncol 2018;36:3331–9.
24. Xu Y, Zhang Y, Wang X, Kang J, Liu X. Prognostic value of performance status in metastatic renal cell carcinoma patients receiving tyrosine kinase inhibitors: a systematic review and meta-analysis. BMC Cancer 2019;19:168.
25. Unger JM, Barlow WE, Martin DP, Ramsey SD, LeBlanc M, Etzioni R, et al. Comparison of survival outcomes among cancer patients treated in and out of clinical trials. J Natl Cancer Inst 2014;106:dju002.
26. Harvey RD, Rubinstein WS, Ison G, et al. Impact of broadening clinical trial eligibility criteria for advanced non-small cell lung cancer patients: real-world analysis. J Clin Oncol 37:18s, 2019 (suppl; abstr LBA108).
27. Burke NJ. Rethinking the therapeutic misconception: social justice, patient advocacy, and cancer clinical trial recruitment in the US safety net. BMC Med Ethics 2014;15:68.
28. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: brain metastases; 2020. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-brain-metastases.
29. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: patients with HIV, hepatitis B virus, or hepatitis C virus infections; 2020. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-patients-hiv-hepatitis-b-virus-or-hepatitis-c-virus.
30. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: patients with organ dysfunction or prior or concurrent malignancies; 2020. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-patients-organ-dysfunction-or-prior-or-concurrent.
31. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: minimum age for pediatric patients; 2020. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cancer-clinical-trial-eligibility-criteria-minimum-age-considerations-inclusion-pediatric-patients.
32. National Cancer Institute Cancer Therapy Evaluation Program (NCI CTEP): Broadened Inclusion/Exclusion Criteria; 2018. Available from: https://ctep.cancer.gov/protocoldevelopment/docs/NCI_ASCO_Friends_Eligibility_Criteria.pdf.

Modernizing Clinical Trial Eligibility Criteria: Recommendations of the ASCO-Friends of Cancer Research Laboratory Reference Ranges and Testing Intervals Work Group

Alexander I. Spira[1], Mark D. Stewart[2], Suzanne Jones[3], Elaine Chang[4], Anitra Fielding[5], Nicole Richie[6], Laura S. Wood[7], Michael A. Thompson[8], Lee Jones[9], Abhilasha Nair[4], Brandon A. Mahal[10], and David E. Gerber[11]

ABSTRACT

Purpose: In clinical research, eligibility criteria promote patient safety and optimize the evidence generated from clinical trials. However, overly stringent eligibility criteria, including laboratory requirements, may limit enrollment, resulting in delayed trial completion and potentially limiting applicability of trial results to a general practice population.

Experimental Design: Starting in 2018, a working group consisting of experts in direct patient care, the FDA, industry, and patient advocacy developed recommendations to guide the optimal use of laboratory reference ranges and testing intervals in clinical trial eligibility criteria and study procedures. The working group evaluated current eligibility criteria across different clinical trial phases and performed a literature review to evaluate the impact of and justification for laboratory test eligibility requirements and

testing intervals in clinical trials. Recommendations were developed on the basis of the goals of promoting safety and optimizing the evidence generated, while also expanding eligibility and applicability, and minimizing excess burden of trial participation.

Results: In general, we found little variation over time and trial phase in laboratory test requirements, suggesting that these eligibility criteria are not refined according to ongoing clinical experience. We propose recommendations to optimize the use of laboratory tests when considering eligibility criteria.

Conclusions: Tailoring the use of laboratory test requirements and testing intervals may increase the number and diversity of patients in clinical trials and provide clinical data that more closely represent the general practice populations.

See related commentary by Giantonio, p. 2369

Introduction

Clinical trial enrollment has become more challenging over the years, in part, due to increasing number and complexity of eligibility criteria and study requirements. From 2001 to 2015, trial endpoints, eligibility criteria, and procedures steadily increased (1, 2). An evaluation of Eastern Cooperative Oncology Group lung cancer protocols revealed a median increase in number and complexity of eligibility criteria from 17 in 1986–1995 to 27 in 2006–2016 (3). Appropriate and relevant eligibility criteria are necessary to ensure the safety of patients participating in a clinical study and to allow for interpretability of the clinical study results (4). However, overly stringent eligibility criteria

may unnecessarily limit enrollment, resulting in delayed trial completion, and limiting generalizability of the research results to a broader practice population. Eligibility for clinical trials should be recognized as a distributive justice issue for individual patients and for vulnerable populations (5). Balancing the need for modernized eligibility criteria with patient safety requires careful review and planning of clinical trial protocols and eligibility criteria.

In 2016, the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*) initiated a joint project to evaluate eligibility criteria in oncology clinical trials and to investigate potential strategies that could expand trial eligibility while maintaining patient safety (6). This initial effort resulted in the development of key recommendations that catalyzed efforts to improve the applicability and accessibility of clinical studies to patients with brain metastases, human immunodeficiency virus infection, younger age, organ dysfunction, and prior/concurrent malignancies (6–10). However, additional barriers and opportunities remain. Follow-up activities were conducted to identify and prioritize additional criteria that may hinder the rate of trial accrual and unnecessarily restrict patient access to investigational therapies.

Laboratory tests represent one of the most commonly employed categories of eligibility criteria in clinical trials. For instance, minimum renal and hepatic function may be required for therapies that are either metabolized by or pose toxicity to these organ systems. Similarly, threshold blood counts provide a margin of safety for myelosuppressive treatments. Despite this clear rationale, there is obvious potential for unintended consequences. For instance, in oncology, the majority of patients are older, a population in which some degree of organ dysfunction is quite common, but rarely has clinical consequences. It follows that a recent study found that strict renal and hepatic function

[1]Virginia Cancer Specialists, US Oncology Research, Fairfax, Virginia. [2]Friends of Cancer Research, Washington, DC. [3]Sarah Cannon Research Institute, Nashville, Tennessee. [4]FDA, Silver Spring, Maryland. [5]AstraZeneca, Gaithersburg, Maryland. [6]Genentech, Inc., South San Francisco, California. [7]Cleveland Clinic Taussig Cancer Center, Cleveland, Ohio. [8]Aurora Cancer Care, Advocate Aurora Health, Milwaukee, Wisconsin. [9]Patient and Research Advocate, Arlington, Virginia. [10]Miller School of Medicine, Sylvester Comprehensive Cancer Center, University of Miami, Miami, Florida. [11]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas.

Translational Relevance

Stringent eligibility criteria, including laboratory test thresholds, may restrict clinical trial enrollment and limit the relevance of study results. The American Society of Clinical Oncology and Friends of Cancer Research worked with stakeholders throughout the cancer research community to develop evidence-based, consensus recommendations to modernize the use of clinical trial laboratory test–related eligibility and intervals. These recommendations may help to facilitate accrual and render trial populations more representative of the disease population, improving the generalizability of the research results.

Table 1. Oncology clinical trial distribution by trial phase and therapy.

Trial characteristic	Solid cancer trials, n (%) (n = 107)	Hematology–oncology trials, n (%) (n = 16)
Trial phase		
I	71 (66%)	11 (69%)
I/II	19 (18%)	4 (25%)
II	8 (8%)	1 (6%)
III	9 (8%)	0 (0%)
Therapy category		
Targeted/small molecule	37 (35%)	5 (31%)
Immunotherapies	46 (44%)	7 (44%)
Chemotherapy	14 (13%)	1 (6%)
Combination	8 (8%)	3 (19%)

requirements were one of the most common reasons for excluding potential patients from clinical trials (11). While not every patient will be a candidate for a clinical trial, the exclusion of patients for what can often be arbitrary reasons, thereby diminishes the desire for those involved to enroll on clinical trials. Laboratory abnormalities may also represent reversible manifestations of the underlying malignancy. ASCO and *Friends* established a working group to understand current practices related to clinical trial laboratory test requirements and intervals. The group also assessed whether reasonable changes could be recommended while preserving patient safety and study scientific integrity. The scope of work did not encompass tumor tissue requirements or biomarker testing for clinical trial enrollment, as they require additional considerations beyond the use of laboratory tests as eligibility criteria (3, 12).

Process

To inform our recommendations related to laboratory test requirements and testing intervals, we reviewed eligibility criteria from a sampling of recently submitted or active cancer clinical trial protocols from diverse sources. Specifically, we included protocols from (i) a clinical practice setting (Sarah Cannon Research Institute, Nashville, TN; industry-sponsored trials activated January 2018–May 2019; $N = 97$), (ii) an industry sponsor (AstraZeneca; late-phase oncology trials active in 2018; $N = 13$), and (iii) a regulatory authority (FDA; applications submitted May 2018–May 2019; $N = 13$). The following information was collected and summarized: disease under study; trial phase; class of therapy (targeted/small molecule, immunotherapy, chemotherapy, or combination therapy); eligibility thresholds for bone marrow, renal, and hepatic function; requirements for transfusion- and growth factor–free periods; and coagulation parameters.

Separately, we reviewed 2019 oncology FDA approvals and identified 26 approvals on the basis of randomized phase III clinical trials. Published articles supporting 23 of the 26 approvals were retrieved (as of March 2020) and the eligibility criteria specifics for each trial were extracted from the article supplementary material (Supplementary Table S1).

Findings

Evaluation of eligibility criteria in clinical trial protocols

Table 1 broadly describes the characteristics of the clinical trials included in our assessment of laboratory test criteria. More than a 100 industry-sponsored trials were represented in the trial review and 13% of the trials only enrolled patients with a hematologic malignancy.

Figure 1 displays the laboratory test–based eligibility criteria for the 107 solid tumor trials included in our analysis. In general, we observed the greatest heterogeneity for renal function, even within a single-drug class. For instance, among immune checkpoint inhibitor trials, creatinine clearance (CrCl) requirements were almost equally distributed among 30, 40, 50, and 60 mL/minute. The justification for such variation is not readily clear, as these drugs tend to undergo similar metabolism and excretion and have similar rates of nephrotoxicity. It is also noteworthy that the most common minimum platelet count requirement was 100,000/μL for all three drug classes, even though thrombocytopenia occurs almost universally with cytotoxic chemotherapy, but in well under 5% of patients treated with immune checkpoint inhibitors. Similarly, hemoglobin eligibility requirement was 9 g/dL for almost all trials, with anemia a common toxicity with cytotoxic agents, but a rare event with immunotherapy.

Hepatic function exceptions for patients with suspected Gilbert syndrome and liver metastases were employed for most clinical trials (66% and 71%, respectively). The guidelines for patients with Gilbert syndrome ranged widely: some trials allowing for a total bilirubin of up to 3× to 5× upper limit of normal (ULN) and a direct bilirubin up to 1.5× ULN; in some cases, no threshold was specified. In addition, the existence of such exceptions raises the question whether laboratory test thresholds could be relaxed more broadly. That is, whether a therapy is considered safe in a patient with elevated hepatic transaminase levels due to liver metastases, might it also be safe in a patient with liver dysfunction due to another reason? As expected, we found that bone marrow function (i.e., minimum blood counts) criteria have different thresholds, if included in hematology malignancy trials (**Fig. 2**).

Importantly, our findings are almost identical to earlier reviews by the FDA and by the ASCO and *Friends* working group (13). This lack of variation over time suggests the possibility that laboratory test–based eligibility criteria template language may be carried forward despite the accumulation of additional clinical experience, on trials or after approval. We noted a similar phenomenon when tracking clinical development across trial phases. Our review of published material (Supplementary Table S1) of 23 of the 26 oncology drugs approved by FDA on the basis of randomized phase III trials in 2019 demonstrated a lack of variation in laboratory test requirements between early-phase and later phase clinical trials of the same agent. Again, this observation may suggest that these eligibility criteria remain static, not taking into account new or developing knowledge.
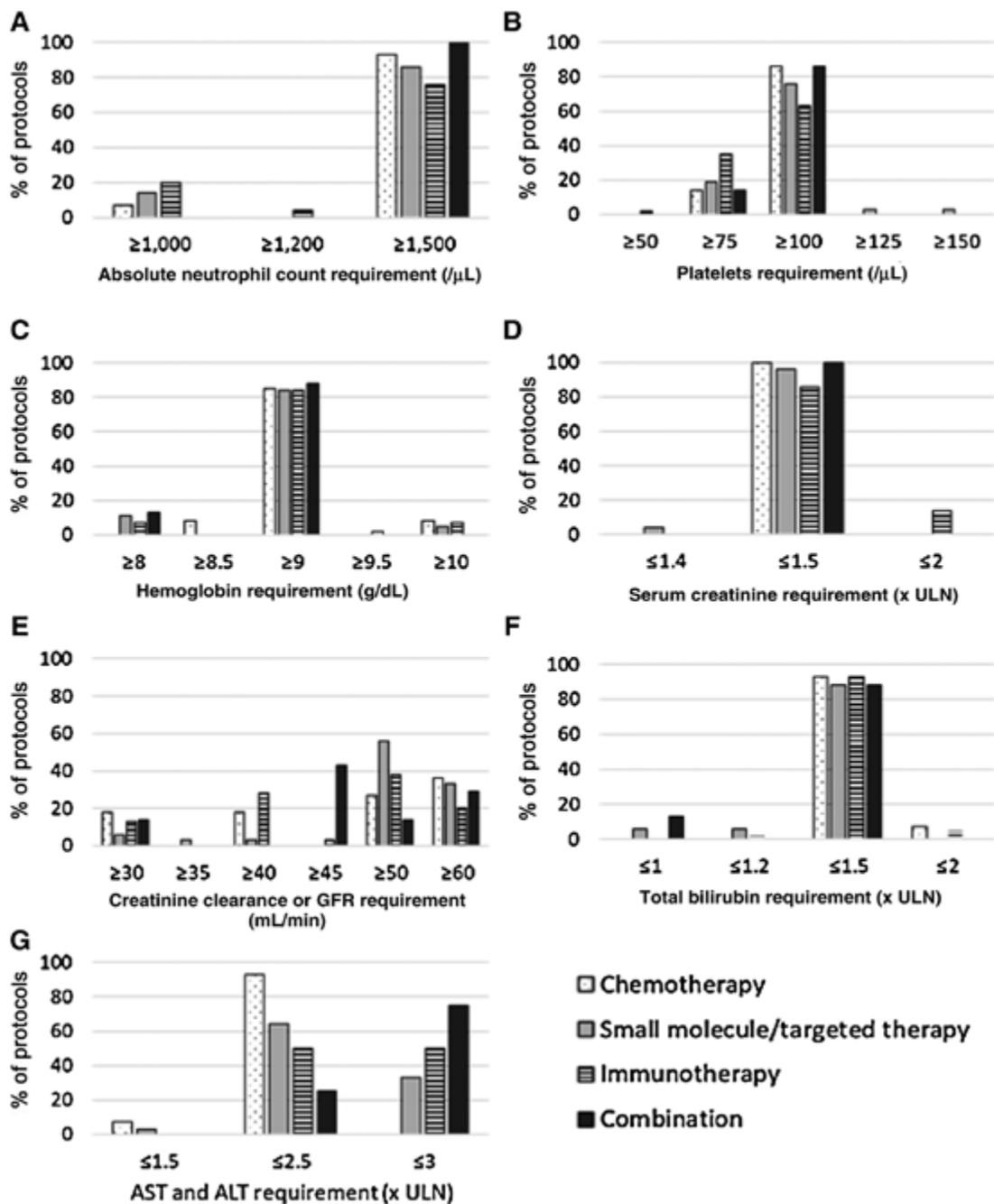
Figure 1.
Frequency of laboratory value requirements according to therapy type for 107 oncology clinical trial protocols for solid tumors. Protocol-specified accepted laboratory test values and number of protocols with each requirement for ANC (**A**), platelet count (**B**), hemoglobin (**C**), serum creatinine (**D**), CrCl or glomerular filtration rate (GFR; **E**), total bilirubin (**F**), and aspartate aminotransferase (AST) and ALT (**G**).

Implications of laboratory eligibility criteria

How do laboratory eligibility criteria impact clinical trial enrollment? A recent study examining 10,500 electronic health records of patients with advanced non–small cell lung cancer (NSCLC) found that expanded criteria that would allow patients with advanced NSCLC and brain metastases, previous or concurrent cancers, and limited kidney function to enroll in clinical trials would nearly double the percentage of patients potentially eligible to enroll in clinical trials (14).

Instances may still exist where strict eligibility criteria are required for patient safety. For example, a drug that causes hemolytic anemia or risk of bleeding may require patients to have a higher hemoglobin criteria for entry; however, a drug without any known effect on this parameter may not require this and could be adequately managed expectantly according to best oncologic care.

Differences between and within drug classes

Laboratory-based criteria should reflect treatment considerations, including organ function adequate for drug metabolism and elimination, and provide a sufficient margin in the event of hepatic or renal toxicity of investigational treatments. Therapies that may be hepatically metabolized or renally excreted would be expected to have more narrow enrollment criteria than those which are eliminated via other means.

Among medical therapies, substantial differences in metabolism/excretion and toxicity profiles render broad recommendations challenging. In some instances, multiple drugs in a class would be expected to have comparable profiles, as is the case for PD-1/PD-L1 immune checkpoint inhibitors. Minor pharmacologic differences within the class, such as IgG subtype (IgG1 vs. IgG4) or antibody species (human vs. humanized), do not translate into meaningful variation in laboratory requirements. In contrast, ALK inhibitors approved for ALK-positive lung cancer differ substantially in pharmacodynamics properties, resulting in truly distinct metabolic and toxicity profiles (17). With this in mind, there will need to be some variability, but data and experience from similar in-class molecules should be used to inform selection of laboratory requirements for eligibility criteria. Furthermore, as investigational therapies advance from early-phase to late-phase development, those criteria should be adjusted on the basis of earlier experience and observations. The current "cut and paste" approach should be challenged and clinical trial protocols continuously reevaluated as recommended in FDA guidance (18).

Laboratory test value variability

Importantly, laboratory test values may differ substantially between testing facilities and among populations. For instance, the lower limit of normal for hemoglobin is 9.6 g/dL in Black women, which falls below the eligibility threshold for some clinical trials (19). In addition, study criteria that use absolute neutrophil count (ANC) > 1,500/µL can contribute to significant racial disparities in studies as a result of benign ethnic neutropenia (20). Lowering the ANC cutoff level could increase the number of eligible minority patients that may have benign ethnic neutropenia. Across populations, among 38 standard laboratory tests analyzed among more than 3,000 healthy individuals in the National Health and Nutrition Examination Survey, only five (glucose, phosphorus, potassium, total bilirubin, and uric acid) did not show significant racial/ethnic difference in distribution (20). For instance, the normal range of serum creatinine for White females was 0.50–1.10 mg/dL, but 0.43–0.88 mg/dL for Asian females. Furthermore, formulas used to assess CrCl often vary widely (21). Black participants had significantly higher normal ranges in CPK, globulin, and total protein, and lower normal ranges in hematocrit, hemoglobin, total cholesterol, triglycerides, and white blood cell than Whites. There are also differences according to gender. For alanine aminotransferase (ALT), upper reference ranges vary from 35 to 79 U/L for men, and 31 to 55 U/L for women (22). Other laboratory tests with significant differences between males and females include total bilirubin, cholesterol, bicarbonate, calcium, and total protein (20). To the best of our knowledge, we cannot identify the rationale for one of the most common liver dysfunction criteria, transaminases of 2×–2.5× ULN for most patients, and sometimes up to 5× with liver metastases. The number of patients this excludes from studies is unknown, but is felt to represent a significant burden especially in patients who may have adequate synthetic and clearance function, but have elevated transaminases because of liver metastases. Current FDA guidance suggests that patients with transaminase elevation up to 20× ULN may have similar tolerance to therapies as those with normal levels (18, 23).

Advanced age also represents a key consideration in laboratory test interpretation, as many patients with common cancers are elderly. Alkaline phosphatase increases by 20% between the 3rd and 8th decade. CrCl increases by 10 mL/minute/1.73 m^2 per decade. Post-prandial glucose increases by 30–40 mg/dL per decade after age 40 years (24). Between the 6th and 8th decades, platelet count decreases by approximately 20,000/mcl (25).

Laboratory test results in cancer populations

Across cancer types, laboratory abnormalities are more common in oncology populations. Anemia, when defined as hemoglobin < 11 g/dL, occurs in up to 40%–60% of patients with common malignancies (26). This is especially true in patients who have already received several treatments for their malignancy, and can be supported easily with transfusions or other care. In terms of renal function, 50% of patients with cancer have CrCl < 90 mL/minute and 20% have CrCl < 60 mL/minute (27). For drugs that are known not to be renally metabolized, this may not be relevant, and only reflect the general performance status of the patient. Furthermore, the formulas used to estimate glomerular filtration rate (e.g., Cockcroft–Gault) often underestimate true CrCl, especially in females and in those that are older with less body mass. More direct measures (e.g., 24-hour urine CrCl) should often be used. Furthermore, the prevalence of laboratory abnormalities is greatest in patients with more advanced cancer, which tend to represent the cases for which a clinical trial may be most appropriate and potentially most beneficial (28, 29).

Recommendations

The group concluded that laboratory tests should be used as exclusionary criteria only when clearly necessary due to safety or efficacy concerns. As demonstrated previously, laboratory-based eligibility criteria are frequently carried forward from earlier protocols to new trials, without critical scientific evaluation of the need and impact of these decisions. Because each clinical trial focuses on specific patient populations and studies specific therapies with differing toxicity and pharmacokinetics considerations, it is not feasible to provide specific laboratory test value thresholds for broad applicability. Nevertheless, the incorporation of the key principles (**Table 2**) may help ensure safety and optimize efficacy, while minimizing unnecessary patient exclusions.

Conclusion

Overall, this working group found that laboratory test–related eligibility criteria (i) may account for exclusion of a meaningful proportion of patients from clinical trials, (ii) rarely change over time or over the course of a therapeutic agent's clinical development, (iii) are highly similar between drug classes that have substantially different pharmacologic and toxicity profiles, and (iv) may have varying impact on patients according to age, gender, and race/ethnicity. We have

Table 2. Recommendations for broadening laboratory reference ranges and testing intervals.

1. Laboratory tests should only be used as exclusionary criteria when scientifically justified and when abnormal test results confer safety concerns.

Laboratory test requirements should be customized to the therapy/therapies under investigation. Ultimately, laboratory test requirements should reflect study therapy pharmacokinetics and pharmacodynamics and anticipated toxicities. For instance, if a therapy does not undergo hepatic metabolism and is not expected to cause hepatic toxicity, strict hepatic function eligibility criteria may not be necessary, or at a minimum, there should be very broad entry criteria. Wherever data are available from similar agents and previous experience should be used as a guide. For example, in some instances (e.g., PD-1/PD-L1 checkpoint inhibitors), pharmacology and toxicity profiles are similar across agents, allowing use of comparable laboratory-related eligibility criteria. In other instances (e.g., ALK inhibitors), each individual drug may have different requirements depending on its individual pharmacokinetic/pharmacodynamic profile. Importantly, restrictions from earlier clinical trials should not be carried forward automatically, but should be modified to reflect the experiences of patients in earlier trials and in postmarket use.

Laboratory test–related eligibility criteria should not be used as a surrogate for performance status or the presence of comorbidities. Because of the older age of most patients with cancer and the likelihood of identifying laboratory anomalies of no clinical significance, the use of laboratory tests to identify sufficiently healthy individuals is likely to result in unnecessary exclusion of potential patients. Instead, clinical trial protocols should specify functional status and comorbidity requirements in line with previous recommendations, as appropriate (10).

Consider adjusting laboratory-based eligibility criteria broadly rather than in specific clinical scenarios. A frequent clinical trial practice is to relax laboratory-related eligibility criteria in populations more likely to have baseline laboratory abnormalities (e.g., allowing lower levels of renal function in patients with genitourinary malignancies, or allowing greater degrees of hepatic dysfunction in patients with primary or metastatic liver cancer). If these population subgroups can be treated effectively and safely, consideration should be given to applying similar laboratory-related eligibility criteria more broadly.

Laboratory-based eligibility criteria should be limited to the clinical concern. As an example, in clinical trials of therapies that may prolong the QTc interval, low levels of electrolytes, such as potassium, calcium, and magnesium, may increase risk of cardiac arrhythmias. A common response to this concern is to require levels of these electrolytes to be within normal limits. This results in unnecessary exclusion of patients whose electrolyte levels are slightly above the normal range, even though there is no increased risk of QTc prolongation. In these cases, precise protocol writing (e.g., requirements for laboratory tests to be above the lower limit of normal rather than within normal limits) with an understanding of the intent of the criteria and the normal variations among people as outlined above is of utmost importance. Furthermore, opportunities to allow for correction to the near-normal range should be allowed. While safety is of utmost concern, protocols should reflect the intended use population for the treatment being evaluated and not situations where the trial data cannot realistically be applied to post-approval scenarios.

Interlaboratory variation should be accounted for when selecting laboratory-based eligibility criteria. It is important to consider thresholds rather than specific normal values. ULN's can vary across laboratories, and criteria should reflect multiples of ULN, rather than absolute numbers (akin to NCI CTCAE criteria). Across academic medical centers, there are substantial differences in serum creatinine determination, with laboratory site accounting for 50% and time of assay performance accounting for another 15% of this variation (23). CrCl should be accounted for by accurate measurements, and options for direct measurements (24-hour urine CrCl) be allowed, rather than formulas that simply estimate the clearance (e.g., Cockcroft–Gault).

2. Laboratory reference values should account for potential normal variations due to race, ethnicity, age, sex, and gender identity (i.e., due to surgical and hormonal changes).

The impact on trial eligibility, enrollment, and relevance should be assessed when selecting laboratory-based eligibility criteria. Laboratory abnormalities occur frequently without clinical significance. Reference intervals generally include 95% of test results obtained from a presumably healthy population. The chance that a healthy person has a test result falling outside this range is 5% for a single test, but rises to 64% for 20 tests (e.g., complete blood count and metabolic panel; ref. 30). As noted previously, the likelihood of test results outside reference ranges is far greater among individuals with cancer and may not be of clinical significance with respect to the treatment being studied.

Demographic differences in laboratory test results, and their implication across populations, should be understood. Given the differences among ethnicities, those criteria that are included should be sufficiently broad to allow for these natural variations (20, 26). It should be noted that persons who have undergone surgery or take medications to align with their gender identity may have altered "normal" laboratory values despite being healthy (31, 32).

3. Routine reassessment of laboratory test–based exclusion criteria should be conducted during the course of clinical research and drug development as investigational agents progress from earlier to later phase clinical trials.

Eligibility criteria should be expanded on the basis of earlier clinical experience and in the absence of safety concerns. Phase I, first-in-human trials should incorporate strict laboratory-related eligibility criteria as a precautionary measure, as the clinical pharmacology and toxicity profile of the novel therapy are not known. However, once these characteristics have been established, laboratory-related eligibility criteria should be adjusted to reflect this experience, enabling appropriate access to therapies under investigation. Currently, the initial criteria are often carried forward to phase II and phase III trials, resulting in unnecessarily strict requirements and exclusion of potential patients, and limiting applicability of results. Similarly, criteria and experience from drugs of a similar class may be used to formulate eligibility entry criteria.

Broadening eligibility criteria by employing less stringent requirements for laboratory eligibility requirements should be accounted for when assessing on-treatment abnormal laboratory values. In addition to grading of laboratory abnormalities using CTCAE, which accounts for the most severe laboratory value aberration, interpretation of results should take into account CTCAE attribution. If patients have baseline laboratory anomalies prior to starting treatment, they may have more frequent and more severe laboratory abnormalities after initiating therapy. To account for this possibility, one approach is to focus on the degree of change in laboratory values, as conveyed by shift tables (33). Shift tables display baseline laboratory values and the shift at postdose, which helps determine the potential impact of the investigational therapy on these results.

(Continued on the following page)

Table 2. Recommendations for broadening laboratory reference ranges and testing intervals. (Cont'd)

---

4. Increasing the intervals between protocol-specific tests should be considered to help reduce patient burden and increase ability to rely on routine clinical testing, especially in later cycles of treatment and over the evolution of protocols from earlier to later phase clinical trials.

Restrictive test intervals could result in reduced interest in and commitment to clinical trials among patients, clinicians, and investigators. Oncology patients, in general, are spending an inordinate amount of time for treatment of their cancer. The average informed consent form for oncology trials is more than 4,000 words and describes hundreds of procedures (34). Unnecessary testing and procedures can lead to more patients choosing not to participate in trials or dropping out over the course of a study. Minimizing testing frequency to reflect what is truly needed to assess safety and efficacy may improve interest, enrollment, and adherence on clinical trials.

Abbreviation: CTCAE, Common Terminology Criteria for Adverse Events.

outlined a number of areas in which modifying current clinical trial eligibility and following the principles of distributive justice may optimize trial participation and efficiency, and applicability of study results to better inform appropriate uses of new therapies. Recommendations outlined in this article can help guide appropriate use of laboratory tests and testing intervals as exclusionary criteria in protocols. This would enable increased clinical trial accrual and provide more relevant data that better mirror the oncology patient populations that ultimately will be treated with these agents. While it is reasonable to establish some minimum criteria for safety, they should be appropriately broad without compromising safety. This will allow oncologists to have more evidence-based discussions with patients and caregivers regarding the potential risks and benefits, ultimately improving shared decision-making in cancer care.

Authors' Disclosures

Disclaimer

The opinions expressed in this article are those of the authors and do not necessarily reflect the views or policies of the authors' affiliated institutions.

Acknowledgments

References

1. Getz KA, Wenger J, Campo RA, Seguine ES, Kaitin KI. Assessing the impact of protocol design changes on clinical trial performance. Am J Ther 2008;15: 450–7.
2. Getz KA, Campo RA. New benchmarks characterizing growth in protocol design complexity. Ther Innov Regul Sci 2017;52:22–8.
3. Garcia S, Bisen A, Yan J, Xie X-J, Ramalingam S, Schiller JH, et al. Thoracic oncology clinical trial eligibility criteria and requirements continue to increase in number and complexity. J Thorac Oncol 2017;12:1489–95.
4. Duke-Margolis Center for Health Policy. Evaluating inclusion and exclusion criteria in clinical trials. Available from:https://healthpolicy.duke.edu/sites/default/files/atoms/files/discussion_guide_evaluating_inclusion_and_exclusion_criteria_in_clinical_trials.pdf.
5. Burke NJ. Rethinking the therapeutic misconception: social justice, patient advocacy, and cancer clinical trial recruitment in the US safety net. BMC Med Ethics 2014;15:68.
6. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.
7. Uldrick TS, Ison G, Rudek MA, Noy A, Schwartz K, Bruinooge S, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research HIV Working Group. J Clin Oncol 2017;35:3774–80.
8. Lin NU, Prowell T, Tan AR, Kozak M, Rosen O, Amiri-Kordestani L, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research Brain Metastases Working Group. J Clin Oncol 2017;35:3760–73.
9. Gore L, Ivy SP, Balis FM, Rubin E, Thornton K, Donoghue M, et al. Modernizing clinical trial eligibility: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Minimum Age Working Group. J Clin Oncol 2017;35:3781–7.
10. Lichtman SM, Harvey RD, Damiette Smit M-A, Rahman A, Thompson MA, Roach N, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research Organ Dysfunction, Prior or Concurrent Malignancy, and Comorbidities Working Group. J Clin Oncol 2017;35:3753–9.
11. Malik L, Lu D. Eligibility criteria for phase I clinical trials: tight vs loose? Cancer Chemother Pharmacol 2019;83:1–4.
12. Lim C, Sung M, Shepherd FA, Nouriany N, Sawczak M, Paul T, et al. Patients with advanced non-small cell lung cancer: are research biopsies a barrier to participation in clinical trials? J Thorac Oncol 2016;11:79–84.
13. Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. J Clin Oncol 2017;35:3745–52.
14. Harvey RD, Rubinstein WS, Ison G, Khozin S, Chen Li, Miller RS, et al. Impact of broadening clinical trial eligibility criteria for advanced non-small cell lung cancer patients: real-world analysis. J Clin Oncol 2019;37:LBA108.
15. LIU Qi, Sharon E, Zineh I, Xie D(J), Baxi SS, Liu C, et al. Organ dysfunction (dys) and clinical outcomes in patients (pts) treated with immune checkpoint inhibitors (ICIs). J Clin Oncol 2019;37:2569.
16. Kutluk Cenik B, Sun H, Gerber DE. Impact of renal function on treatment options and outcomes in advanced non-small cell lung cancer. Lung Cancer 2013;80:326–32.
17. Rashdan S, Gerber DE. A crowded, but still varied, space: brigatinib in anaplastic lymphoma kinase-rearranged non-small cell lung cancer. Transl Cancer Res 2017;6:S78–S82.
18. FDA. Cancer clinical trial eligibility criteria: patients with organ dysfunction or prior or concurrent malignancies guidance for industry. Available from: https://www.fda.gov/media/123745/download.
19. Lim E, Miyamura J, Chen JJ. Racial/ethnic-specific reference intervals for common laboratory tests: a comparison among Asians, Blacks, Hispanics, and White Hawaii. J Med Public Health 2015;74:302–10.

20. Vastola ME, Yang DD, Muralidhar V, Mahal BA, Lathan CS, McGregor BA, et al. Laboratory eligibility criteria as potential barriers to participation by black men in prostate cancer clinical trials. JAMA Oncol 2018;4:413–4.

21. Joffe M, Hsu C-Y, Feldman HI, Weir M, Landis JR, Hamm LL. Variability of creatinine measurements in clinical laboratories: results from the CRIC study. Am J Nephrol 2010;31:426–34.

22. Neuschwander-Tetri BA, Ünalp A, Creer MH Nonalcoholic Steatohepatitis Clinical Research Network. Influence of local reference populations on upper limits of normal for serum alanine aminotransferase levels. Arch Intern Med 2008;168:663–6.

23. Talal AH, Venuto CS, Younis I. Assessment of hepatic impairment and implications for pharmacokinetics of substance use treatment. Clin Pharmacol Drug Dev 2017;6:206–12.

24. Cavalieri TA, Chopra A, Bryman PN. When outside the norm is normal: interpreting lab data in the aged. Geriatrics 1992;47:66–70.

25. Segal JB, Moliterno AR. Platelet counts differ by sex, ethnicity, and age in the United States. Ann Epidemiol 2006;16:123–30.

26. Knight K, Wade S, Balducci L. Prevalence and outcomes of anemia in cancer: a systematic review of the literature. Am J Med 2004;116:11S–26S.

27. Launay-Vacher V, Oudard S, Janus N, Gligorov J, Pourrat X, Rixe O, et al. Prevalence of renal insufficiency in cancer patients and implications for anti-cancer drug management: the renal insufficiency and anticancer medications (IRMA) study. Cancer 2007;110:1376–84.

28. Dunphy EP, Petersen IA, Cox RS, Bagshaw MA. The influence of initial hemoglobin and blood pressure levels on results of radiation therapy for carcinoma of the prostate. Int J Radiat Oncol Biol Phys 1989;16:1173–8.

29. Cappell MS, Goldberg ES. The relationship between the clinical presentation and spread of colon cancer in 315 consecutive patients. A significant trend of earlier cancer detection from 1982 through 1988 at a university hospital. J Clin Gastroenterol 1992;14:227–35.

30. O'Kane MJ, Lopez B. Explaining laboratory test results to patients: what the clinician needs to know. BMJ 2015;351:h5552.

31. Roberts TK, Kraft CS, French D, Ji W, Wu AHB, Tangpricha V, et al. Interpreting laboratory results in transgender patients on hormone therapy. Am J Med 2014;127:159–62.

32. Adriaansen MJ, Perry WNC, Perry H, Steel R. Binary male-female laboratory reference ranges do not reflect reality for transgender individuals on sex-hormone therapy. New Zeal J Med Lab Sci 2017;71:101–5.

33. FDA. Reviewer guidance: conducting a clinical safety review of a new product application and preparing a report on the review. Available from:https://www.fda.gov/media/71665/download.

34. Nathe JM, Krakow EF. The challenges of informed consent in high-stakes, randomized oncology trials: a systematic review. MDM Policy Pract 2019;4:2381468319840322.

Modernizing Clinical Trial Eligibility Criteria: Recommendations of the ASCO–Friends of Cancer Research Prior Therapies Work Group

Raymond U. Osarogiagbon[1], Diana Merino Vega[2], Lola Fashoyin-Aje[3], Suparna Wedam[3], Gwynn Ison[3], Sol Atienza[4], Peter De Porre[5], Tithi Biswas[6], Jamie N. Holloway[7], David S. Hong[8], Madison M. Wempe[2], Richard L. Schilsky[9], Edward S. Kim[10], and James L. Wade III[11]

ABSTRACT

Purpose: Restrictive eligibility criteria induce differences between clinical trial and "real-world" treatment populations. Restrictions based on prior therapies are common; minimizing them when appropriate may increase patient participation in clinical trials.

Experimental Design: A multi-stakeholder working group developed a conceptual framework to guide evaluation of prevailing practices with respect to using prior treatment as selection criteria for clinical trials. The working group made recommendations to minimize restrictions based on prior therapies within the boundaries of scientific validity, patient centeredness, distributive justice, and beneficence.

Recommendations: (i) Patients are eligible for clinical trials regardless of the number or type of prior therapies and without requiring a specific therapy prior to enrollment unless a scientific or clinically based rationale is provided as justification. (ii) Prior therapy (either limits on number and type of prior therapies or requirements for specific therapies before enrollment) could be used to determine eligibility in the following cases: a) the agents being studied target a specific mechanism or pathway that could potentially interact with a prior therapy; b) the study design requires that all patients begin protocol-specified treatment at the same point in the disease trajectory; and c) in randomized clinical studies, if the therapy in the control arm is not appropriate for the patient due to previous therapies received. (iii) Trial designers should consider conducting evaluation separately from the primary endpoint analysis for participants who have received prior therapies.

Conclusions: Clinical trial sponsors and regulators should thoughtfully reexamine the use of prior therapy exposure as selection criteria to maximize clinical trial participation.

See related commentary by Giantonio, p. 2369

Introduction

An expanding number of innovative approaches to cancer treatment, such as targeted anticancer therapies, are reframing our approach to patient selection for the evaluation of experimental agents in clinical trials (1). For example, targeted anticancer therapies are efficacious in molecularly defined patient subsets, irrespective of previous exposure to other types of treatment; they also tend to have more favorable side effect profiles than traditional cytotoxic chemotherapeutic agents, thus patients treated with targeted agents are often physically well enough to receive subsequent therapies (2–8). With ever-increasing understanding of drug interactions and novel toxicity profiles of innovative therapies, it is important to rethink the use of prior therapy as eligibility criteria for patient exclusion or inclusion into controlled studies.

The prior therapy selection criteria are a key aspect of clinical trial design and vary substantially, depending on the goals of the study. Updating their use would promote patient access to experimental drugs, improve patient participation, clinical trials accrual, and increase the applicability of trial results to a more general population of patients. The rationale for broadening eligibility criteria to make clinical trial participants more representative of the general population has been described previously in a joint effort by the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*; ref. 9). Key recommendations to modernize clinical trial eligibility criteria associated with minimum age (10), organ dysfunction, prior or concurrent malignancies (11), brain metastases (12), and human immunodeficiency virus infection (13) have been published and led to new guidance documents development by the FDA (14–17).

Building on the success of this initial effort, ASCO and *Friends* convened a multi-stakeholder working group of experts from multiple disciplines to evaluate the current practice of using prior therapy as entry criteria for clinical trials, and developed recommendations to support the design and conduct of clinical trials.

Process

The ASCO-*Friends* Prior Therapies Working Group included clinical investigators, clinical pharmacologists, patient advocates, and regulatory and industry representatives. The working group developed a framework for eliminating barriers to participation in clinical trials based on restrictive criteria on the types and timing of prior cancer therapy, to maximize inclusivity in clinical trials. All working group members acknowledged that the use of prior therapies as a criterion for eligibility is deeply tied to clinical

[1]Multidisciplinary Thoracic Oncology Program, Baptist Cancer Center, Memphis, Tennessee. [2]Friends of Cancer Research, Washington, DC. [3]U.S. Food and Drug Administration, Silver Spring, Maryland. [4]Advocate Aurora Health, Milwaukee, Wisconsin. [5]Janssen Pharmaceutica, Beerse, Belgium. [6]University Hospitals Seidman Cancer Center, Cleveland, Ohio. [7]Patient Advocate, Arlington, Virginia. [8]MD Anderson Cancer Center, Houston, Texas. [9]American Society of Clinical Oncology, Alexandria, Virginia. [10]Levine Cancer Institute, Atrium Health, Charlotte, North Carolina. [11]Cancer Care Specialists of Central Illinois, Decatur, Illinois.

Corresponding Author: Raymond U. Osarogiagbon, Baptist Cancer Center, 80 Humphreys Center Dr, Ste 330, Memphis, TN 38120. Phone: 901-722-0540; E-mail: rosarogi@bmhcc.org

Translational Relevance

With rapid-cycle discovery of new drugs with well-characterized targets and mechanisms of action, the therapeutic index and efficacy of new candidate cancer drugs are significantly better than in the past, reconfiguring the safety and efficacy calculus in using prior therapy exposure to select patients for clinical trials. Concurrently, there is growing awareness that cancer drugs approved in restrictive clinical trials are often used in real-world practice for groups of patients ineligible for such trials, even without evidence for their safety or efficacy. Using a patient-centered conceptual model that considers safety, efficacy, and the ethical principles of distributive justice and beneficence, a multi-stakeholder working group has proposed three recommendations to promote more thoughtful consideration of the use of prior therapy as a selection criterion for oncology trials. The overarching objective is to minimize this potential barrier to clinical trial access for willing patients.

trial design. For example, a trial may require patients to be treatment naïve if investigating an agent as first-line therapy, while another trial may require patients to have received a specific prior therapy or a limited number of prior therapies if investigating an agent as a later line of therapy, or if a prior therapy is known to interact with the investigational agent.

The working group, thus, developed a conceptual framework with which to guide subsequent discussions; and reviewed a sample of the most recently registered, ongoing clinical trials in five disease areas selected from the ClinicalTrials.gov website to examine how prior therapies are being used as eligibility criteria. The working group used these insights to guide subsequent discussions that led to the final proposed recommendations.

Conceptual framework

We developed a conceptual framework for evaluating the advantages and limitations of using prior therapy as eligibility criteria in clinical trials (**Fig. 1**). This framework considers both information about the potential or known toxicity of the experimental agent and its efficacy in relation to existing treatments, to determine the appropriateness of restricting clinical trial participation based on previous treatments. By deconstructing the decision-making process into its basic elements of safety and efficacy, the working group was able to develop recommendations within the context of patient centeredness, keeping with the ethical principles of distributive justice and beneficence (18, 19). These considerations guided the thought process behind the ClinicalTrials.gov exercise and the recommendations proposed by this working group.
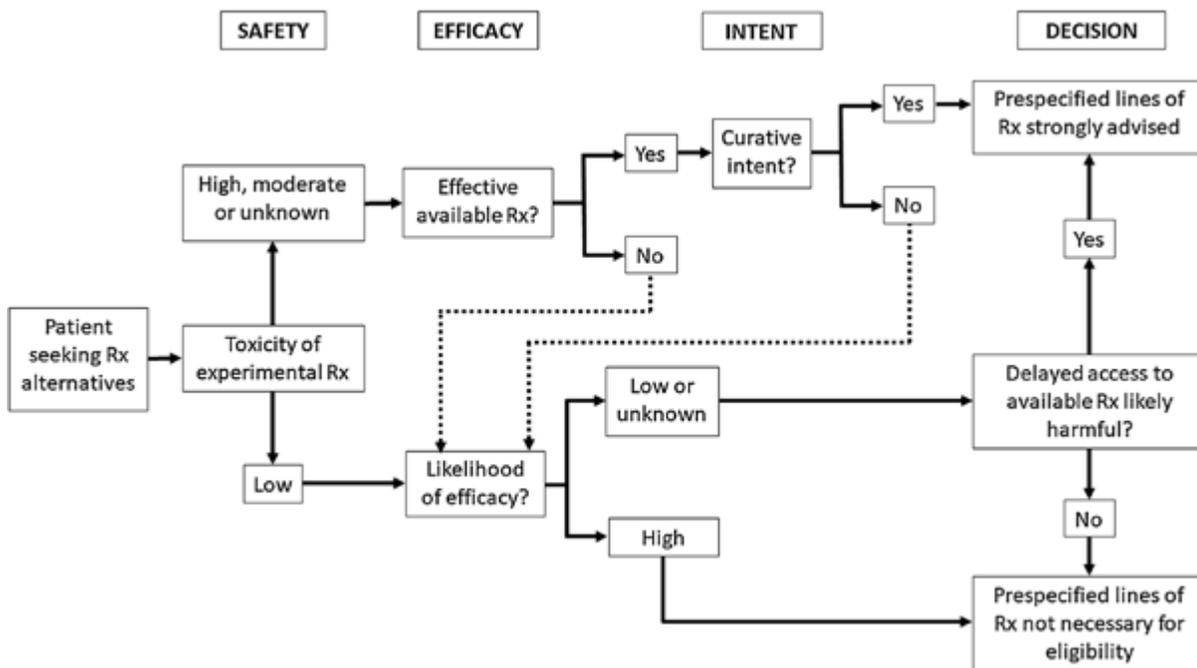
Figure 1.
Conceptual framework to guide the use of prior therapy as selection criteria in clinical trials. We encourage the use of this conceptual framework early in the process of clinical trial design, to minimize the barrier to entry. We encourage shared decision-making between the patient and the health care provider in selecting treatment options, including treatment within a clinical trial. In general, clinical trials should be designed with the aim of achieving greater inclusivity with minimal restrictions placed on trial entry. Decisions regarding whether exposure to existing therapy should be required prior to administering an investigational therapy should consider the risks (i.e., known or unknown safety profile) and the efficacy of the therapy, and the availability of alternative treatments. In a clinical setting, wherein the standard-of-care treatment provides a high likelihood of cure, such as may be the case for some in an adjuvant setting or for some advanced lymphomas or pediatric cancers, it may be appropriate to restrict access to experimental therapy until after disease progression, relapse, or intolerance of such existing treatments. However, in the noncurative setting, a requirement for receipt of such "standard" options is not recommended unless there is sound scientific or clinical rationale to support the restriction. Rx, therapy.

Current use of prior therapies as eligibility criteria, ClinicalTrials.gov assessment

To better understand the scope of the problem, we used the ClinicalTrials.gov website to assess the extent to which prior therapy is currently used as eligibility criteria (20). In July 2019, we accessed the ClinicalTrials.gov website to select the 11–12 most recently registered phase I–III clinical trials in breast cancer, colon cancer, lung cancer, malignant melanoma, and multiple myeloma, for close evaluation of how prior therapy requirements were being used as eligibility criteria. The working group defined inclusion and exclusion criteria based on prior therapies as those criteria that did not have a specified washout period. Any criteria based on prior therapies that included a washout period were categorized as "washout period criteria" and not assessed in this study. Trials were categorized by cancer type, clinical trial phase (with phase I/phase II trials considered phase I due to their emphasis on the exploration of safety endpoints), and by drug class (including immunotherapy, alone or in combination, chemotherapy only, and other).

Findings

The working group reviewed a total of 57 trials to assess whether there were requirements based on prior therapies, specifying whether these were inclusion or exclusion criteria (**Fig. 2**; Supplementary Table S1). Thirty-three clinical trials corresponded to phase I (58%), 15 trials to phase II (26%), and nine to phase III (16%). More than 90% of clinical trials investigated immunotherapies (91%; 52/57), either alone or in combination with other agents or treatment modalities, such as chemotherapy and radiotherapy. Of the remaining five clinical trials (9%), four investigated other therapies, such as retinoid X receptor and hyperbaric oxygen, and one investigated a chemotherapy agent only (Supplementary Table S1). The breakdown of clinical trials by cancer type was 12 trials in breast cancer, 11 in colon cancer, 11 in lung cancer, 11 in melanoma, and 12 in multiple myeloma.

Two-thirds (38/57) of the trials included prior therapy as an eligibility criterion and 19 trials (33%) did not specify prior therapy as an eligibility criterion. Among 38 trials, 19 (50%) specified prior therapy as an exclusion criterion only, 14 (37%) specified prior therapy as both an inclusion and exclusion criterion, and five (13%) specified prior therapy as an inclusion criterion only. When categorized by clinical trial phase, 58% (19/33) of phase I trials specified either an inclusion or exclusion criterion based on prior therapies, while 42% (14/33) did not. Of the 15 phase II trials, 10 (67%) specified either an inclusion or exclusion criterion based on prior therapies, compared with five (33%), which did not. Finally, all nine phase III trials specified either an inclusion or exclusion criterion based on prior therapies, with exclusion criterion only specified in six trials (67%), and both inclusion and exclusion criteria based on prior therapies specified in three trials (33%; **Fig. 2**).

The pattern of use of prior therapies as eligibility criteria in trials categorized by drug class and tumor type is shown in **Fig. 2**. Use was most prevalent in lung cancer trials (82%; 9/11) and least prevalent in melanoma trials (45%; 5/11). The predominance of immunotherapy trials in the survey, which may be indicative of the predominance of immunotherapy trials in recent ClinicalTrials.gov registrations, may limit the extrapolation of our findings to nonimmunotherapy trials.

Recommendations

Taking all available evidence into consideration, the working group proposed the recommendations outlined in **Table 1**, on the basis of the key principles of preserving patient safety, facilitating the assessment
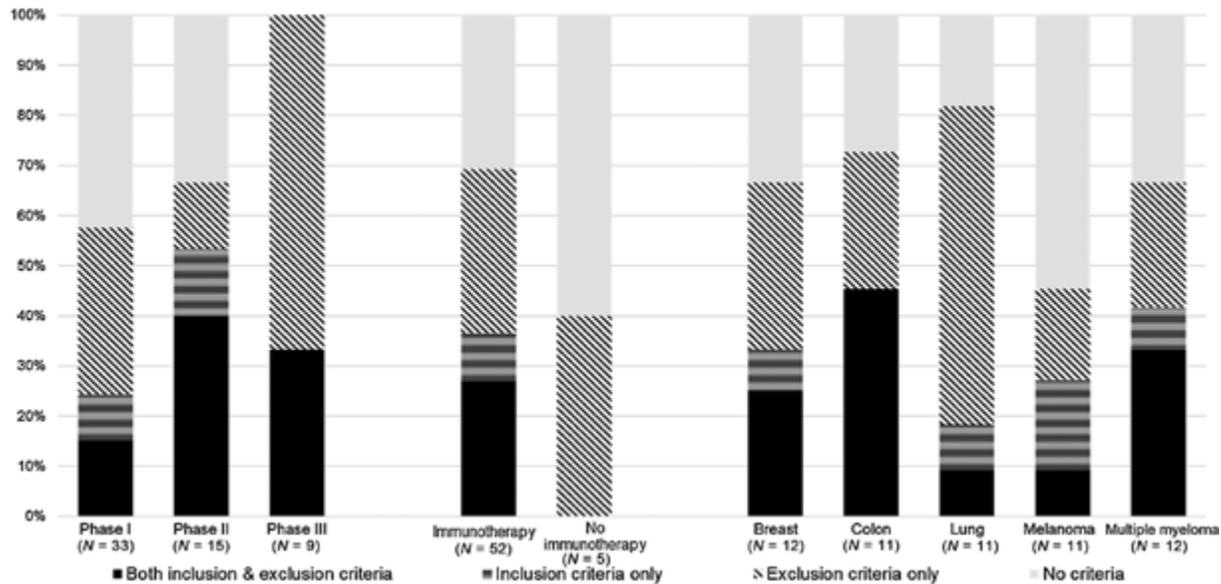
Figure 2.
Frequency of the use of prior therapies as inclusion and/or exclusion criteria in clinical trials as part of the ClinicalTrials.gov exercise categorized by phase, drug class, and tumor type. ClinicalTrials.gov was accessed on July 23, 2019. Trials with any component of phase I trials (e.g., phase I/II) were categorized as phase I trials. The working group defined inclusion and exclusion criteria based on prior therapies as those criteria that did not have a specified washout period. Any criteria based on prior therapies that included a washout period was categorized as "washout period criteria" and not evaluated in this assessment.

Table 1. Recommendations for the modernization of eligibility criteria based on prior therapies.

Recommendation	Comment
(i) Patients are eligible for clinical trials regardless of the number or type of prior therapies and without a requirement to have received a specific therapy prior to enrollment unless a scientific or clinical rationale is provided as justification.	There needs to be a balance between the desire to conduct a tightly controlled experiment with high internal validity and the reality that patients with much broader demographic and disease characteristics than those patients evaluated in clinical trials are prescribed the approved drugs (22). Clinical trials are the most controlled mechanism for evaluation of the safety and efficacy of investigational agents in a carefully selected patient population. However, arbitrary exclusion of populations of patients who may desire access to clinical trials, and may derive benefit from them, runs counter to the principles of patient autonomy and beneficence. The opportunity cost of such arbitrary restrictions to sponsors and designers of clinical trials may also be largely underrecognized. Overly complex eligibility criteria may, in part, increase the burden on research staff, slow down clinical trial accrual, increase the risk of failure to complete clinical trials, and raise the audit and regulatory stakes for enrolling sites and their staff. Indeed, the Pharmaceutical Research and Manufacturers Association reported that 80% of clinical trials do not finish on time, 20% are delayed 6 months or more, and up to two thirds of clinical trials fail to meet their original patient enrollment goals (39). Reducing barriers that hinder recruitment, such as broadening eligibility criteria, would be beneficial.
(ii) Prior therapy (either limits on the number and type of prior therapies or requirements for specific therapies before enrollment) could be used to determine eligibility in the following cases:	To promote greater clinical trial inclusivity in trials, minimally restrictive criteria should be used, with patient safety and autonomy as the primary considerations. However, there may be some specific scenarios in which these criteria may be justifiable and necessary to maintain patient safety and ensure treatment efficacy. In these cases, when entry into a trial is contingent upon exposure to a prior therapy (or lack thereof), scientific and/or clinically sound rationale should be provided.
a. If the agents being studied target a specific mechanism or pathway that could potentially interact with a prior therapy.	The working group identified three cases in which prior therapy could be used to determine patient eligibility, and additional specific scenarios are listed in **Table 2**.
b. If the study design requires that all patients begin protocol-specified treatment at the same point in the disease trajectory.	With evolving evidence that investigational agents with known mechanisms of action, which effectively target specific biologic pathways, are highly effective irrespective of the point in the disease trajectory (2–5, 32–36, 38, 43), trial designers are encouraged to continuously reevaluate the use of prior therapies as eligibility criteria.
c. In randomized clinical studies, if the therapy in the control arm is not appropriate for the patient due to previous therapies received.	
(iii) Trial designers should consider conducting evaluation separately from the primary endpoint analysis for participants who have received prior therapies.	There may be concerns that enrolling a subset of patients who may be considered higher risk due to their exposure or lack of exposure to prior therapies might jeopardize drug development by reducing apparent treatment efficacy and/or increasing the risk of severe adverse events in such high-risk patients. To preserve methodologic rigor in clinical trials while maintaining the desire for minimal barriers to entry, multiple strategies can be considered for data analysis and interpretation. Some of these approaches have been suggested in prior publications in this series (9, 22). These concerns can be addressed at the time of trial design by prespecifying how data from this subset of higher risk patients will be handled in executing the trial and the statistical analyses. For example, in early-phase trials, an expanded cohort with perceived high risk due to prior therapy history can be recruited and closely monitored for safety signals, which can be used to prompt closure of that subset without jeopardizing the whole program (9–12). The information generated from this expansion cohort can then be used to inform the criteria for later-phase trials. In addition, patient enrollment into the arms of randomized clinical trials can be stratified on the basis of prior therapy history, with all patients included in the intention-to-treat analysis, but with prespecified analyses restricted to a "modified intention-to-treatment" subset. As suggested by Jin and colleagues, in hierarchical testing, the primary analysis could be restricted to the lower-risk modified intention-to-treat population, with subsequent analyses to include the whole population (22). Another alternative would be to enroll a parallel cohort of patients who do not meet the prior therapy restriction, which would not be part of the intention-to-treat population, but whose data, analyzed separately, would still provide descriptive safety and efficacy information. This alternative, however, might be considered less desirable because toxicity data from such a trial design would be difficult to interpret due to the absence of a control group (22) and because of the time still required to accrue the intention-to-treat population. The rapidly evolving development of adaptive clinical trial designs and statistical analysis methodologies may accommodate the goal of expanding clinical trial participation irrespective of prior therapy history.

Table 2. Scenarios in which selection criteria based on prior exposure to therapy may be applied, rationale, pros, and cons.

Scenario	Rationale	Pros	Cons
Limitation to treatment-naïve patients (i.e., first-line treatment trials)	Comparison with standard first-line treatment Typically, drug registration trials	• Need clearly defined target treatment population • High risk of poor efficacy or greater toxicity when pretreated patients are included • Minimize expense of including patients deemed higher risk for failure • Need to establish a market niche	• Exclusion of healthy individuals, often long-term beneficiaries of prior therapy with recent disease progression (e.g., long-term cancer survivors with subsequent disease) • Progression to metastatic disease after prior systemic adjuvant therapy is a common scenario with arbitrary time interval rules for inclusion/exclusion
Limitation to previously treated patients ("salvage therapy" trials)	Assurance of prior exposure to standard of care. Possibly, drug registration trial	• Secure means of testing promising new agents in the face of existing, highly curative standard treatments • Establishment of market niche • Potentially enable comparison with historically treated population outcomes	• Risks of requiring prior therapy in the face of potentially more effective or safer novel treatment, especially in rare or clinically aggressive disease, when available standard treatment is modestly effective or highly toxic • Violation of patient autonomy/threat to principle of beneficence • Comparisons across clinical trials and over different time frames and populations, even of ostensibly similar groups, is rife with unknown bias, is statistically unsound, and should be discouraged. Furthermore, restricting access to clinical trials in relatively uncommon diseases is particularly wasteful
Strict limitation of the number of allowed prior therapies (typically one or two)	Assurance of prior exposure to standard of care. Often, drug registration trials	• Patients typically still are good candidates for treatment despite prior therapy. • Clearly defined population with relative homogeneity in terms of disease refractoriness and susceptibility to toxicity • Establishment of market niche for registration • Enable comparison of outcomes in a noncomparative trial with historically treated populations	• "Indication creep" occurs in real-world practice, creating exposures with unknown safety or efficacy • Optimal time to identify adverse safety and/or efficacy signals is under the auspices of a clinical trial with greater standardized data collection and evaluation than in routine practice • Potential missed opportunity to detect new efficacy signals that can expand market share • Expansion of eligible cohorts promotes accrual, timely trial completion, with associated cost savings • Given greater success rates and less toxicity of drugs with well-defined mechanisms of action, ethical dilemma created with potential loss of opportunity for access to beneficial treatment, in violation of patient autonomy and the principle of beneficence • Comparisons with historical populations spurious and invalid
Exclusion based on prior exposures to specific treatments	Concerns about interference with action of trial agent (effectiveness and/or toxicity)	• Typically, exclusion of exposure to drugs of the same class, with similar mechanism of action and, therefore, likelihood of adverse interaction with effectiveness or safety of trial agent	• Potential missed opportunity for detection of differential activity • Increasing population of long-term survivors who have had remote prior exposure and no residual effects from prior therapy

of drug efficacy, and promoting patient centeredness. The recommendations reflect the general position that as a default, minimal eligibility restrictions based on prior therapy should be implemented. The working group encourages critical thinking about the appropriate use of eligibility criteria based on prior therapies by considering the need to balance the goals of scientific rigor, and trial efficiency, with the goal of broader clinical trial inclusivity.

Discussion

Traditionally, clinical trials specify prior therapies that are either required for inclusion or exclusion of patients from participation. Tightly controlled eligibility criteria are thought to optimize conditions to test the safety and efficacy of an investigational therapy (21, 22). The working group characterized scenarios under which the use of

prior therapies as exclusion criteria (i.e., no prior therapy or no prior therapy of a certain type allowed) or as inclusion criteria (i.e., a defined number and/or type of prior therapies required for eligibility), may be appropriate. We further outlined the rationale for, and the pros and cons of, these scenarios (**Table 2**). We avoided scenarios in which prior therapies might be specified to exclude overlapping exposure to prior and new therapy (i.e., washout periods), deferring to the ASCO-*Friends* Working Group on Washout Periods and Concomitant Medications, which was charged with the task of addressing this important topic.

For example, to "optimize for safety," prevailing practice may be to specify a maximum number of prior therapy exposures, to minimize the risk that heavily pretreated patients may be more likely to experience excessive toxicity. Another common practice may be to limit prior exposure to specific types of prior therapies whose toxicities potentially overlap those of the experimental therapy, such as would be the case for potentially irreversible toxicities, like bone marrow toxicity, neuropathy, or cardiotoxicity. However, the concerns for excess toxicity may be more appropriately addressed by requiring resolution or improvement in the toxicities of concern, rather than by implementing broad exclusions based upon exposure to prior therapies.

Eligibility criteria may also be designed to "optimize for efficacy" by defining a study population that is comparable with historical trials to permit an evaluation of improvements in outcomes with the new therapeutic in noncomparative trials. Trial designers may seek to define a study population that is most likely to respond to the treatment being studied. For example, positioning a second-generation agent targeting resistance mutations where either relapse after a first-generation compound has selected for the acquired resistance mutation or designing upfront treatment for an *ab origin* mutation (23, 24), or limiting the number of prior treatments to minimize the risk that heavily pretreated patients with refractory disease will bias trial results against treatment response (25, 26). This raises the fundamental question whether restricting patients from enrollment in a clinical trial solely based on prior therapy is justifiable to show the best outcome of the trial treatment for a specified patient population or whether it is more advantageous to open up patient eligibility to enable quantification of outcomes across the spectrum of potential clinical use scenarios (22, 27).

In addition to safety and efficacy considerations, the intent of the trial is an important consideration. A trial designed to evaluate safety and effectiveness of an investigational agent for the purposes of gaining marketing licensure may seek to enroll patients with an unmet medical need, for example, patients with advanced refractory disease who have exhausted currently available treatment options, for whom clinical trials may be the only potential treatment option (28, 29). A trial may be designed to compare a new investigational agent against a standard-of-care treatment in a particular treatment setting, such as a specific line of therapy or treatment with a particular class of drugs (30, 31). In this case, trial designers often insist that a study population be naïve to any treatment or restricted to a population that has received a minimum or maximum number, or certain specific types, of prior therapies.

Advances in the understanding of the biological underpinnings of cancer have facilitated the development of therapies that are based on key tumor characteristics, such as gene and protein expression profiles, have relatively well-understood mechanisms of action, are often effective irrespective of prior drug exposures, and have a wider therapeutic index (2–5, 32–38). This has mostly rendered obsolete the clinical

rationale for eligibility criteria that specify requirements for prior therapy, simultaneously raising the ethical dilemma of the opportunity cost to the patient, of arbitrary patient selection criteria based on prior therapy (18, 19, 39). Patients increasingly seek access to promising drugs in development, particularly those that treat rare or clinically aggressive cancers. Mandating prior exposure to marginally effective or excessively toxic treatments, in theory, may delay or prevent access to potentially more beneficial novel treatment. Conversely, blanket exclusion of patients who have received any prior treatment may prevent otherwise healthy patients with disease progression from gaining access to potentially health-preserving new treatments.

Finally, the implementation of prior therapy criteria in the absence of scientific or clinical rationale may unnecessarily restrict the post-approval target population and delay evaluation of a new drug's efficacy and safety in the wider population that may ultimately receive the drug once it is approved (40). Limiting patient access to clinical trials based either on exposure to prior therapies or the requirement for patients to have progressed after specific therapies limits patient access to clinical trials and may significantly slow trial accrual or compromise completion of these trials.

Conclusion

The discovery of highly effective anticancer treatments, and the technologies that enable the selection of patients with targetable genomic alterations, has resulted in the traditional line-of-treatment demarcations becoming increasingly blurred. Ultimately, the inclination to conduct clinical trials in homogeneous populations for more robust comparisons (internal validity) must be finely balanced against the pragmatic need to test novel therapies in the "real-world" populations that will eventually be exposed to approved treatments (external validity; ref. 40), as well as the concept of patient centeredness, and the ethical principles of distributive justice and beneficence. The Institute of Medicine, now National Academy of Medicine, defined patient centeredness as "responsiveness to the needs, values, and expressed preferences of the individual patient" (41). In 2010, the same body recommended that the NCI, Cooperative Groups, and physicians should take steps to increase the speed, volume, and diversity of patient accrual and to ensure high-quality performance at all sites participating in cooperative group trials. As an example, they recommended that they should "encourage patient eligibility criteria that allow the broadest participation possible" (42).

Clinical trial designers and sponsors should clearly justify any restrictions based on prior therapies. The working group's overarching consideration in making these recommendations was to promote patient-centered clinical trials with the minimum entry criteria needed to ensure participant safety and broad access. We hope these recommendations will be widely adopted by key stakeholders, especially designers, sponsors, and regulators of clinical trials.

References

1. Heymach J, Krilov L, Alberg A, Baxter N, Chang SM, Corcoran RB, et al. Clinical cancer advances 2018: annual report on progress against cancer from the American Society of Clinical Oncology. J Clin Oncol 2018;36: 1020–44.
2. Fontes Jardim DL, Schwaederle M, Wei C, Lee JJ, Hong DS, Eggermont AM, et al. Impact of a biomarker-based strategy on oncology drug development: a meta-analysis of clinical trials leading to FDA approval. J Natl Cancer Inst 2015;107: djv253.
3. Zhang S, Anjum R, Squillace R, Nadworny S, Zhou T, Keats J, et al. The potent ALK inhibitor brigatinib (AP26113) overcomes mechanisms of resistance to first- and second-generation ALK inhibitors in preclinical models. Clin Cancer Res 2016;22:5527–38.
4. Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. Nat Rev Cancer 2017;17:637–58.
5. Recondo G, Facchinetti F, Olaussen KA, Besse B, Friboulet L.Making the first move in EGFR-driven or ALK-driven NSCLC: first-generation or next-generation TKI? Nat Rev Clin Oncol 2018;15:694–708.
6. Leonetti A, Facchinetti F, Rossi G, Minari R, Conti A, Friboulet L, et al. BRAF in non-small cell lung cancer (NSCLC): pickaxing another brick in the wall. Cancer Treat Rev 2018;66:82–94.
7. Proto C, Ferrara R, Signorelli D, Lo Russo G, Galli G, Imbimbo M, et al. Choosing wisely first line immunotherapy in non-small cell lung cancer (NSCLC): what to add and what to leave out. Cancer Treat Rev 2019;75:39–51.
8. Kummar S, Gutierrez M, Doroshow JH, Murgo AJ. Drug development in oncology: classical cytotoxics and molecularly targeted agents. Br J Clin Pharmacol 2006;62:15–26.
9. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.
10. Gore L, Ivy SP, Balis FM, Rubin E, Thornton K, Donoghue M, et al. Modernizing clinical trial eligibility: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research minimum age working group. J Clin Oncol 2017;35:3781–7.
11. Lichtman SM, Harvey RD, Damiette Smit MA, Rahman A, Thompson MA, Roach N, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research organ dysfunction, prior or concurrent malignancy, and comorbidities working group. J Clin Oncol 2017;35:3753–9.
12. Lin NU, Prowell T, Tan AR, Kozak M, Rosen O, Amiri-Kordestani L, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research brain metastases working group. J Clin Oncol 2017;35:3760–73.
13. Uldrick TS, Ison G, Rudek MA, Noy A, Schwartz K, Bruinooge S, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology-Friends of Cancer Research HIV working group. J Clin Oncol 2017;35:3774–80.
14. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: brain metastases. Available from: https://www.fda.gov/media/121317/download.
15. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: patients with HIV, hepatitis B virus, or hepatitis C virus infections. Available from: https://www.fda.gov/media/121319/download.
16. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: minimum age for pediatric patients. Available from: https://www.fda.gov/media/121318/download.
17. U.S. Food and Drug Administration. Guidance document: cancer clinical trial eligibility criteria: patients with organ dysfunction or prior or concurrent malignancies. Available from: https://www.fda.gov/media/123745/download.
18. Laage T, Loewy JW, Menon S, Miller ER, Pulkstenis E, Kan-Dobrosky N, et al. Ethical considerations in adaptive design clinical trials. Ther Innov Regul Sci 2017;51:190–9.
19. Markman M. Survival and participation in a cancer clinical trial: a provocative relationship. Oncology 2016;91:177–8.
20. U.S. National Library of Medicine. ClinicalTrials.gov. Available from: clinical trials.gov.
21. Schilsky RL. End points in cancer clinical trials and the drug approval process. Clin Cancer Res 2002;8:935–8.
22. Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. J Clin Oncol 2017;35:3745–52.
23. Mok TS, Wu Y-L, Ahn M-J, Garassino MC, Kim HR, Ramalingam SS, et al. Osimertinib or platinum–pemetrexed in EGFR T790M–positive lung cancer. N Engl J Med 2016;376:629–40.
24. Yang JC-H, Ahn M-J, Kim D-W, Ramalingam SS, Sequist L V, Su W-C, et al. Osimertinib in pretreated T790M-positive advanced non–small-cell lung cancer: AURA study phase II extension component. J Clin Oncol 2017;35: 1288–96.
25. Middleton G, Crack LR, Popat S, Swanton C, Hollingsworth SJ, Buller R, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Ann Oncol 2015;26:2464–9.
26. Moore KN, Mannel RS. Is the NCI MATCH trial a match for gynecologic oncology? Gynecol Oncol 2016;140:161–6.
27. Schilsky RL. Publicly funded clinical trials and the future of cancer care. Oncologist 2013;18:232–8.
28. Gadgeel SM, Gandhi L, Riely GJ, Chiappori AA, West HL, Azada MC, et al. Safety and activity of alectinib against systemic disease and brain metastases in patients with crizotinib-resistant ALK-rearranged non-small-cell lung cancer (AF-002JG): results from the dose-finding portion of a phase 1/2 study. Lancet Oncol 2014;15:1119–28.
29. Lin JJ, Zhu VW, Schoenfeld AJ, Yeap BY, Saxena A, Ferris LA, et al. Brigatinib in patients with alectinib-refractory ALK-positive NSCLC. J Thorac Oncol 2018;13: 1530–8.
30. Soria J-C, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in untreated EGFR-mutated advanced non–small-cell lung cancer. N Engl J Med 2017;378:113–25.
31. Camidge DR, Kim HR, Ahn M-J, Yang JC-H, Han J-Y, Lee J-S, et al. Brigatinib versus crizotinib in ALK-positive non–small-cell lung cancer. N Engl J Med 2018;379:2027–39.

Disclaimer

Acknowledgments

32. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372: 2509–20.

33. Prowell TM, Theoret MR, Pazdur R. Seamless oncology-drug development. N Engl J Med 2016;374:2001–3.

34. Ding MQ, Chen L, Cooper GF, Young JD, Lu X. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. Mol Cancer Res 2018;16:269–78.

35. Hobbs BP, Barata PC, Kanjanapan Y, Paller CJ, Perlmutter J, Pond GR, et al. Seamless designs: current practice and considerations for early-phase drug development in oncology. J Natl Cancer Inst 2019;111:118–28.

36. Savard M-F, Khan O, Hunt KK, Verma S. Redrawing the lines: the next generation of treatment in metastatic breast cancer. Am Soc Clin Oncol Educ B 2019;39:e8–21.

37. Solomon BJ, Besse B, Bauer TM, Felip E, Soo RA, Camidge DR, et al. Lorlatinib in patients with ALK-positive non-small-cell lung cancer: results from a global phase 2 study. Lancet Oncol 2018;19:1654–67.

38. Drilon A, Laetsch TW, Kummar S, Dubois SG, Lassen UN, Demetri GD, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. N Engl J Med 2018;378:731–9.

39. Printz C. Trial and error. Cancer 2013;119:471–2.

40. Masuda N, Ohe Y, Gemma A, Kusumoto M, Yamada I, Ishii T, et al. Safety and effectiveness of alectinib in a real-world surveillance study in patients with ALK-positive non–small-cell lung cancer in Japan. Cancer Sci 2019;110: 1401–7.

41. Institute of Medicine (US) Committee on Quality of Health Care in America. Improving the 21st-century health care system. In crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.

42. Mendelsohn J, Moses HL, Nass SJ. A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. Washington, DC: National Academies Press; 2010.

43. Renfro LA, An M-W, Mandrekar SJ. Precision oncology: a new era of cancer clinical trials. Cancer Lett 2017;387:121–6.

Modernizing Clinical Trial Eligibility Criteria: Recommendations of the ASCO-Friends of Cancer Research Washout Period and Concomitant Medication Work Group

R. Donald Harvey[1], Kathryn F. Mileham[2], Vishal Bhatnagar[3], Jamie R. Brewer[3], Atiqur Rahman[3], Cassadie Moravek[4], Andrew S. Kennedy[5], Elizabeth A. Ness[6], E. Claire Dees[7], S. Percy Ivy[8], Scot W. Ebbinghaus[9], Caroline Schenkel[10], and Thomas S. Uldrick[11]

ABSTRACT

Purpose: Washout periods and concomitant medication exclusions are common in cancer clinical trial protocols. These exclusion criteria are often applied inconsistently and without evidence to justify their use. The authors sought to determine how washout period and concomitant medication allowances can be broadened to speed trial enrollment and improve the generalizability of trial data to a larger oncology practice population without compromising the safety of trial participants.

Experimental Design: A multistakeholder working group was convened to define problems associated with excessively long washout periods and exclusion of patients due to concomitant medications. The group performed a literature search and evaluated study data from the Pancreatic Cancer Action Network (PanCAN), Emory University

School of Medicine (Atlanta, GA), and the FDA to understand recent approaches to these eligibility criteria. The group convened to develop consensus recommendations for broadened eligibility criteria.

Results: The data analysis found that exclusion criteria based on washout periods and concomitant medications are frequently inconsistent and lack scientific rationale. Scientific rationale for appropriate eligibility criteria are presented in the article; for washout periods, rationale is presented by treatment type.

Conclusions: Arbitrary or blanket washout and concomitant medication exclusions should be eliminated. Where there is evidence to support them, clinically relevant washout periods and concomitant medication–related eligibility criteria may be included.

See related commentary by Giantonio, p. 2369

Introduction

Patient access to evidence-based experimental treatments is associated with improved outcomes in the cancer population (1). Expediting enrollment into therapeutic clinical trials in cancer is dependent on removing barriers to patient participation, such as overly restrictive eligibility criteria. Trials that adopt criteria safely reflecting populations most commonly seen in daily practice are more likely to accrue rapidly and be applicable to greater numbers of patients.

Approximately 20% of patients are ineligible for trials on the basis of commonly employed eligibility criteria (2). This makes a strong case for critical analysis of areas where eligibility criteria may be expanded safely. Prior work by American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*) recommended numerous areas where expanded eligibility should be employed (3). This list was

[1]Emory University School of Medicine, Atlanta, Georgia. [2]Levine Cancer Institute, Atrium Health, Charlotte, North Carolina. [3]FDA, Washington, D.C. [4]Pancreatic Cancer Action Network, Manhattan Beach, California. [5]Sarah Cannon Research Institute, Nashville, Tennessee. [6]NCI Center for Cancer Research, Bethesda, Maryland. [7]UNC Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina. [8]NCI Cancer Therapy Evaluation Program, Bethesda, Maryland. [9]Merck & Co., North Wales, Pennsylvania. [10]American Society of Clinical Oncology, Alexandria, Virginia. [11]Fred Hutchinson Cancer Research Center, Seattle, Washington.

Corresponding Author: R. Donald Harvey, Winship Cancer Institute of Emory University, 1365 C Clifton Road, NE, CPL017B, Atlanta, GA 30322. Phone: 404-778-4381; E-mail: rdharve@emory.edu

extensive, but a number of barriers remain. Our working group was formed to evaluate two commonly perceived barriers: washout periods from recent therapies/interventions and prohibited concomitant medications.

A washout period is defined as a time between treatment periods that is intended to prevent misinterpreting observations about study-related treatments that were actually due to prior therapies. Generally, washout/waiting periods prior to enrollment are employed in cancer trials following surgery, radiation, cytotoxic chemotherapy, small-molecule/tyrosine kinase inhibitors, monoclonal antibodies (with and without drug conjugates), and immunotherapies.

Prohibited concomitant medications create eligibility and timing challenges, because patients receiving anticancer therapies often have comorbidities that require drug therapy, such as pain, diabetes, or gastrointestinal or cardiovascular disorders. While some medications may be necessarily prohibited early in investigational agent development, prolonged prohibition across trial phases reduces the applicability of a therapy to a broader patient population in trials and following approval.

Current applications of washout period and concomitant medication eligibility criteria are discussed in **Table 1**. Reducing and/or eliminating a need to include time-based washout periods and prohibit concomitant medications may facilitate both clinical trial participation and greater generalizability of the research findings to a larger oncology practice population.

Process

The multistakeholder group identified concerns regarding washout periods and prohibited medications, with a focus on broadening eligibility criteria as much as possible to increase

Translational Relevance

Washout periods for prior treatments and interventions limit timely accrual and evidence generation and may prevent patient enrollment without adding safety measures or preventing misinterpretation of efficacy results. Exclusion of patients who require concomitant medications for comorbidity or supportive care management prevents early understanding of investigational agent tolerability and dosing in those likely to receive the treatment after approval. Less restrictive requirements for prior therapy washout periods and concomitant medication use, in many instances, should be considered and may facilitate both clinical trial participation and greater generalizability of the research findings to a larger oncology practice population.

efficiency of enrollment and potentially diversify enrolled populations to include greater numbers of patients with comorbidities and chronic medication management needs. The group's observations of current and ideal eligibility criteria and trial design related to washout periods and concomitant medications are described in **Table 2**.

A literature search was performed to understand the historical rationale and background of common eligibility criteria, particularly for washout periods. Because of the relative lack of information obtained, additional data were pursued from three datasets: a series of trials in the Pancreatic Cancer Action Network (PanCAN) portfolio, a sampling of trials performed at the Winship Cancer Institute of Emory University (Atlanta, GA), and a review of new approvals in 2018 by the FDA.

Data Analysis

PanCAN trial dataset

Eligibility criteria for industry-, institutional-, and NCI-sponsored metastatic pancreatic adenocarcinoma treatment studies were reviewed to evaluate the need for specific recommendations related to washout periods and concomitant medications. Eligibility criteria from 16 phase III (including one seamless phase II/III) trials in the PanCAN database between 2010 and 2019 were evaluated (**Table 3**). Eligibility criteria from corresponding phase I and II trials studying treatments that advanced to phase III trials listed in PanCAN's database or on clinicaltrials.gov were also evaluated. In total, 34 trials studying 15 unique investigational agents were evaluated.

Studies were evaluated for washout periods for prior radiotherapy, chemotherapy, monoclonal antibodies, immunotherapy, and investigational agents. Washout periods for surgery, corticosteroids, blood cell stimulating drugs, antibiotics, and hormone therapy were also noted when indicated. When treatment-specific washout periods were not available as a result of inadequate details about entry criteria, more general exclusion criteria that would likely include these specific treatments were included (e.g., "washout from all prior systemic treatment").

Results showed a lack of consistency in washout periods from trial to trial, regardless of study phase and type of therapy, with most trials not mentioning a washout period in eligibility criteria. There was also a lack of consistency when reviewing how washout periods for therapies change over time as an investigational agent moves from earlier phase to later phase trials. While the washout periods often stayed the same for many types of therapies as an investigational agent moved to later phase testing, in some instances the washout periods decreased, increased, or were removed altogether. A rationale for washout periods was rarely provided. Our review demonstrated that about 50% of studies included time-based washout periods from 14 to 28 days.

Table 1. Definitions and applications of washout periods and prohibited concomitant medications.

Washout periods

Definition: a washout period is defined as a time between treatment periods that is intended to prevent clouding of information from one intervention to the next.

Application: washout/waiting time periods prior to enrollment are identified in protocols following surgery, radiation, cytotoxic chemotherapy, small-molecule/tyrosine kinase inhibitors, monoclonal antibodies (with and without drug conjugates), and immunotherapies.

Historical rationale: each aspect of a protocol-required washout period may have a different historical rationale, including prevention of untoward adverse events (e.g., wound healing after surgery and cytopenias), drug interactions (e.g., tyrosine kinase inhibitors overlapping with investigational agents), and incorrect adverse event attribution (e.g., late effects with immunotherapies). While in many cases these may be associated with theoretical concerns, they are often irrelevant to clinical practice.

Example: protocol-based treatment vs. clinical practice, a protocol may require a 21-day washout period from a daily oral EGFR-directed tyrosine kinase inhibitor; whereas in practice, a patient would be rapidly transitioned to next-line therapy after knowledge of progressive disease, with the only interval between doses being that required for insurance approval. These agents have short half-lives, and in some instances, discontinuation may be associated with a disease flare, making rapid transitions to next-line therapies critical (19, 20).

Concomitant medications

Definition: a concomitant medication is any drug or dietary supplement that a study participant uses in addition to the treatment under investigation.

Application: on average, patients with cancer take five chronic noncancer medications, not including those that may be used to manage adverse events associated with anticancer therapy (21). As patients age, the prevalence of comorbidities and associated polypharmacy increases (22).

Historical rationale: exclusion of concomitant medications is intended to prevent adverse drug interactions that may affect pharmacokinetics assessment, increase adverse event risks, and in rarer cases, reduce anticancer agent efficacy.

Example: protocol-based treatment vs. clinical practice, protocols often prohibit patients from taking ondansetron in any dose or route due to fears of QTc prolongation with an investigational agent; however, oral ondansetron is used widely and commonly in practice. The risk of QTc prolongation is solely due to high-dose intravenous ondansetron use and has not been shown with the oral route (23).

Table 2. Working group observations related to washout period- and concomitant medication–based trial design.

Current state

Real-time learning of adverse event profiles and pharmacology applicable to washout periods and concomitant medication prohibition is often not reflected in updated protocols.

A lack of data exists regarding patients not enrolled on trials due to extensive washout periods or inability to change or discontinue a prohibited medication.

Washout periods are essentially nonspecific surrogates for a clinical (e.g., adverse event) or laboratory (e.g., absolute neutrophil count) measurement that are included to ensure participant safety and prevent confounding of observations (safety or efficacy) on trial.

Lack of rationale for or specificity regarding washout period and concomitant medication exclusions can cause patient confusion about why they are ineligible for certain trials.

Optimal state

Although postmarketing development of drugs occurs, it is optimal and possible to have complete data on concomitant medication allowances at approval.

Evaluating potential safety and pharmacology interactions, such as QT interval prolongation studies and drug–drug interaction studies, early in drug development can liberalize concomitant medication allowances during later phases of drug development.

Nonclinical tools, such as *in silico* modeling, should be optimized to potentially minimize exclusion of medications and/or reduce required sample sizes in trials.

In reviewing the concomitant medications data, the most commonly excluded concomitant medications were infectious disease treatments and anticoagulants. As with washout periods, rationale for the exclusion of these concomitant medications was rarely provided.

Emory dataset

A series of 102 trials, across phases, was retrospectively evaluated for both washout periods and allowance of concomitant medications (**Table 4**). The majority were early-phase trials with pharmaceutical sponsors, and primarily included investigational oral small molecules alone or in combinations. Each trial was assessed for required washout periods for surgery, radiation, chemotherapy, monoclonal antibodies, immunotherapy, and investigational agents. Of the 102 trials, 36 were silent for a washout period from surgery. The remainder are listed in **Table 4**. Overall, washout periods varied; however, many categories had similar proportions in the ≤14 and

Table 3. Summary of PanCAN data review.

Washout periods as I/E criteria			
	14 days	**21 days**	**28+ days**	**No washout period I/E criteria**
Radiation	11.76%	2.94%	26.47%	58.82%
Chemotherapy	23.53%	5.88%	5.88%	64.71%
Monoclonal antibodies	11.76%	5.88%	2.94%	79.41%
Immunotherapy	14.71%	5.88%	2.94%	76.47%
Investigational agents	20.59%	5.88%	20.59%	52.94%
Surgery	8.82%	14.71%	47.06%	29.41%

Change in washout periods with later-phase trials					
	Shorter	**Same**	**Longer**	**Not allowed**	**Silent**	**Added**
Radiation	6.67%	33.33%	0.00%	40.00%	13.33%	6.67%
Chemotherapy	0.00%	33.33%	6.67%	33.33%	13.33%	13.33%
Monoclonal antibodies	0.00%	40.00%	0.00%	33.33%	0.00%	26.67%
Immunotherapy	0.00%	40.00%	0.00%	33.33%	6.67%	20.00%
Investigational agents	6.67%	13.33%	0.00%	26.67%	20.00%	33.33%
Surgery	0.00%	66.67%	20.00%	0.00%	6.67%	6.67%

Most commonly excluded concomitant medications	
Antibiotics	35.29%
Other anti-infectives	29.41%
Antifungals	26.47%
Anticoagulants	17.65%
Corticosteroids	2.94%
Growth factors	2.94%

Abbreviation: I/E, inclusion/exclusion.

Table 4. Summary of Emory data review.

Trial characteristics (*N* = 102)	
Phase	%
I	37%
I/II	22%
II	28%
III (2 seamless trials)	13%
Sponsor	
Pharmaceutical	77%
National Cancer Institute (NCI)	11%
Academic center	11%
Performance status allowed	
0–1	42%
0–2	55%
0–3	3%
Investigational agent type	
Small molecule[a]	66%
Monoclonal antibody[a]	21%
Chemotherapy[a]	8%
Antibody–drug conjugate	5%

Trial washout periods for prior treatments			
	≤14 days	21 days	≥28 days
Radiation (*n* = 87)	47%	9%	27%
Chemotherapy (*n* = 93)	34%	20%	37%
Monoclonal antibody (non-IO; *n* = 78)	24%	7%	45%
Immunotherapy (*n* = 75)	30%	12%	31%
Investigational agent (*n* = 88)	19%	16%	46%

Exclusions for concomitant medications			
CYP isozyme	Inducers	Inhibitors	Substrates
3A4/5	39%	40%	9%
2D6	2%	2%	2%
2C8/9	2%	3%	3%
1A2	4%	10%	2%
2C19	2%	3%	1%

[a]Includes combinations.

≥28 day timeframes, suggesting periods were not uniformly selected regardless of investigational agent mechanism of action (MOA).

Exclusions for concomitant medications were also evaluated, and common classes leading to ineligibility included corticosteroids (60%), antifungal agents (36%), anticoagulants (15%), human immunodeficiency virus therapy (13%), other anti-infectives (12%), and gastrointestinal medications (11%). Drug–drug interactions leading to exclusions were also evaluated, with a focus on agents that are metabolized by or affect the cytochrome P450 (CYP) enzyme system. Of 102 trials, 49 excluded some type of CYP agent. The most common isozyme leading to exclusions was CYP 3A4/3A5, with similar numbers for agents that induce and inhibit the pathway. The frequency of this exclusion aligns with this isozyme's role in the metabolism of approximately 60% of orally administered drugs (4, 5).

FDA data

The FDA analysis focused on new molecular entities (NME) that were approved in 2018 across all therapeutic areas within the Office of Hematology and Oncology Products (6). The rationale for this selection method of recently approved NMEs was to obtain a sample of products spanning a diverse range of molecules, novel targets, and therapeutic areas. The FDA working group members reviewed characteristics of registrational trials specific to concomitant medications and washouts, as outlined in the publicly available FDA product reviews and product labeling. For washouts, the FDA analysis included whether trials included periods for chemotherapy agents, monoclonal antibodies, immuno-oncology agents, prior investigational agents, and radiotherapy. For concomitant medications, the FDA analysis focused on whether CYP exclusions, drug–drug interactions, and concomitant medication allowances were included in registrational trial protocols.

The FDA analysis evaluated a variety of products, including therapies for solid and hematologic malignancies. A variety of types of molecular entities were reviewed for this analysis, including small molecules, monoclonal antibodies, radiolabeled analogues, and enzymes. Of the 19 NMEs approved in 2018, there was a wide range of washout periods specified in the registrational trials. Frequently, protocols included blanket language encompassing prior chemotherapy, radiation, and surgery. The most frequently used washout period ranged between 14 and 28 days, however, some protocols did not specify any washout period, and the longest washout period was 3 months. Overall, there was heterogeneity in washout periods specified in registrational protocols, even among similar therapeutic classes and diseases, and absence of rationale was common.

Prohibited concomitant medications were also specified in a heterogeneous manner. Many trials of small molecules prohibited the use of CYP3A4 substrate medications, and washout periods varied greatly. For example, one trial used clear language regarding CYP3A4: "the concomitant use of drugs or foods that are strong inhibitors or inducers of CYP3A are not allowed," whereas another protocol used less definitive language: "coadministration with moderate/strong CYP3A4 inhibitors was not recommended. However, such medications could be used with caution and only if considered medically necessary..." As with washout periods, this analysis revealed a dearth of rationale for prohibited concomitant medications included in these registrational trials.

Recommendations

The consensus recommendations below are made in consideration of the benefits and risks to broadening criteria described above. These recommendations should inform sponsors and investigators as they draft study eligibility criteria, but are not intended as template language for trial protocols. Eligibility criteria should be tailored to the investigational treatment and patient population. For that reason, the recommendations are inclusive, rather than specific and prescriptive. Recommended language such as "clinically significant expected adverse event" should be replaced or supported by disease- and drug-specific, evidence-based examples.

Washout periods

(i) Time-based washout periods should be removed from protocol eligibility criteria in most cases. Any inclusion of time-based washout periods should be scientifically justified and clearly specified.

(ii) Relevant clinical and laboratory parameters should be used in place of time-based washout periods to address safety considerations.

(iii) Potential trial participants should have recovered from clinically significant adverse events of their most recent therapy/intervention prior to enrollment.

Table 5. Historical rationale for common time-based washout period eligibility criteria and key considerations for scientifically justified washout eligibility criteria, by treatment type: chemotherapy, small-molecule inhibitors, monoclonal antibodies, and antibody–drug conjugates.

Treatment type	Key shortcomings of common/historical washouts	Key considerations for scientifically justified washouts
Chemotherapy	• Many protocols include requirements for washout periods from prior therapy, often ranging from 14 to 28 days, yet a literature search yielded little in the way of published rationale for time-based washout periods from cytotoxic chemotherapy. • Treatment delays are a risk to patients who demonstrate radiographic progression, and screening periods may be employed to establish required intervals between radiographic evaluation.	• The typical 28-day washout period on the basis of the anticipated time for patients to recover from side effects of prior chemotherapy is no longer scientifically justified in many cases. – For example, in the era of growth factors, 3–4 weeks are not necessarily required for myelosuppression recovery.

Small-molecule inhibitors (including, but not limited to TKIs, serine and threonine kinase inhibitors, cyclin-dependent kinase inhibitors, MEK inhibitors, and tropomyosin kinase inhibitors)	• EC are not routinely updated to reflect differing MOAs, elimination half-lives, and toxicity profiles of targeted therapies.
	• Much of the trial language surrounding kinase inhibitors is the same as cytotoxic agents, antibodies, or other cancer treatments with prolonged washouts without justification.
	• The rationale for the differences with agents with minimal acute and chronic toxicity profiles is not well elucidated.
	• An approach to ensure patient safety from treatment withdrawal complications has yet to infiltrate protocol design, despite extensive documentation of effects, such as TKI withdrawal disease flare.
	  – For example, gastrointestinal stromal tumors have a unique biology with rapid disease progression when imatinib is removed after prolonged benefit (9).
		• Many targeted agents have rapid time-to-peak concentration, as well as abbreviated elimination half-lives, a unique property (e.g., compared with monoclonal antibodies).
		• The MOA of a given TKI on the tumor and the effects of any specific TKI on other factors related to the natural history of a given cancer or anticipated clinical course of a trial participant must be understood prior to initiation of treatment. This is imperative for the safety of the patient not only for treatment-related side effects, but also for treatment withdrawal effects.
		  – For example, when outcomes of patients with advanced renal cell carcinoma treated with TKIs before and after cytoreductive nephrectomy are compared, complication rates are variable, but most note potential delayed wound healing and exacerbation of underlying medical conditions specific to perioperative VEGF-targeting TKIs (8).

Monoclonal antibodies (therapeutic tumor-targeted proteins with variable fragments engineered for epitope binding and based on IgG1 or IgG4 backbones)	• Monoclonal antibody therapies have more pharmacologic consistency than other agents (e.g., oral therapies), allowing for more predictable distribution and elimination, with typical half-lives ranging from 14 to 21 days (12). Despite this consistency, washout periods in EC are highly variable, suggesting history rather than pharmacology-driven timing.	• Concerns of clouding investigational therapeutic efficacy are minimal when the most recent therapy has failed the patient. • Because of the target specificity, concrete consideration of adverse events associated with monoclonal antibodies and their impact on next treatments may be determined in the absence of an arbitrary time period.
ADCs (a subset of monoclonal antibodies that comprise a monoclonal antibody, a linker, and a therapeutic payload)	• Payloads utilized to date have been agents such as maytansinoids and topoisomerase inhibitors that are in actuality chemotherapeutic agents, with cytopenias and other conventional acute adverse effects. Washout periods following these agents have varied and have often not been specified for this class; however, their growing use warrants discussion.	• For eligibility purposes, ADCs may be considered for washout periods as two different drugs, the monoclonal antibody and the payload. • The targeted component of the monoclonal antibody portion of the ADC can be considered for its specificity and contribution to a potential adverse event for an investigational agent or regimen. • Like cytotoxic chemotherapy, recovery from toxicities following ADCs are best measured by laboratory and clinical parameters, rather than timeframes. Rarely will a simple time period be justified, adequate, or necessary for ensuring safe and clear management of patients enrolled on trials.

Abbreviations: ADC, antibody–drug conjugate; EC, eligibility criteria; TKI, tyrosine kinase inhibitor.

Table 6. Historical rationale for common time-based washout period eligibility criteria and key considerations for scientifically justified washout eligibility criteria, by treatment type: radiotherapy, surgery, and immunotherapies.

Treatment type	Key shortcomings of common/historical washouts	Key considerations for scientifically justified washouts
Radiotherapy		• CNS edema postradiation: to realize all the potential benefits of enrolling patients with brain metastases and gather real-world experience of such patients, eligibility requirements should establish a 14-day washout after stereotactic radiotherapy or whole-brain radiotherapy for patients as a standard (13). • Myelosuppression risk: postradiotherapy myelosuppression risk is based on the percentage of active bone marrow irradiated, so the percentage of total bone marrow activity by bony site is helpful in determining the RR of marrow acute side effects from radiotherapy (14). • Acute mucosal membrane reactions to radiation: defined washout period times following standard palliative radiotherapy to mucosal or other surfaces are better replaced by clinical observation, particularly because adverse events will be low-grade and self-limited in nature in most patients.
Surgery	• As noted in the PanCAN dataset, eligibility washout timeframes following surgery vary greatly, and are often not mentioned, even within a single cancer type (**Table 1**). • Differing approaches (laparoscopic vs. open), invasiveness, anesthesia employed, and anatomic location are some of the variables that may impact recovery from the variety of surgeries that patients with cancer may undergo prior to trial enrollment. This heterogeneity suggests that the underlying rationale for including a specified number of days or weeks, rather than more specific parameters for recovery following a procedure, is arbitrary and should be removed from protocols.	• Specific clinical and medical assessment should be employed to ensure potential trial volunteers are functionally prepared and healed to safely receive investigational therapies. • For postsurgery treatment as with other treatments, arbitrary time periods do not reflect or replace clinical judgment, are part of a combination of EC that often overlap to ensure safety (e.g., laboratory values and performance status), and cannot be expected to be broadly applicable across multiple patients and procedures.

Immunotherapies	• Trials should not default to historical washout periods based on time or pharmacokinetic parameters (e.g., half-life), as this approach is both impractical and may not be in the patient's best interests, particularly since a new regimen on a trial has most likely been selected because of cancer progression.	• Pharmacologically, this class of agents includes a variety of molecules designed to modulate antitumor immune responses, and that often have an extended period of time for onset of both clinical activity and adverse events.
		• The tempo of median onset and resolution of irAEs have to be considered when patients transition from immunotherapies on trials or in the clinic to investigational agents.
		  • Median time to resolution of irAEs of 12 weeks has been generally consistent among immune checkpoint inhibitor agents (e.g., initial reports of ipilimumab; ref. 10).
		• Data support rapid subsequent trial enrollment when coupled with an initial understanding of investigational agent adverse event profiles and experience in adverse event attribution.
		  • A recent study showed that up to 25% of patients may experience new or worsening irAEs (most commonly hypothyroidism) after 6 or more months of therapy, but only 2.5% will experience a deepening of response after 6 months (23).
		• Late occurring irAEs that may cloud attribution to a single drug or regimen on study have to be accounted for prior to enrollment.
		• A thorough history of agent(s) given, timing of treatment, irAEs experienced, and understanding of the timing of common late effects may assist in differentiating late effects from prior therapies versus new effects from investigational ones.
		• It may be more useful to stratify study participants based on prior immunotherapy use and to avoid washout periods in the absence of unresolved irAEs that threaten participant safety.

Abbreviations: CNS, central nervous system; EC, eligibility criteria; irAE, immune-related adverse events.

Concomitant medications

(i) Concomitant medication use should only exclude patients from trial participation when clinically relevant known or predicted drug–drug interactions or potential overlapping toxicities will impact the safety of trial participants or compromise efficacy.

Scientific Rationale for Washout Periods by Treatment Type

Arbitrary time periods do not reflect or replace clinical judgment, are part of a combination of eligibility criteria that often overlap to ensure safety (e.g., laboratory values and performance status), and cannot be expected to be broadly applicable across multiple patients and procedures. Sponsors and investigators should provide the scientific rationale for washout periods when developing and implementing protocols, rather than relying on historic precedent that may not be appropriate for the treatment or disease being studied.

The group reviewed the rationale for common time period–based washout eligibility criteria for seven treatment types [chemotherapy, small-molecule inhibitors (1, 2 , 4, 5, 8, 9), immunotherapies (3, 10, 11), monoclonal antibodies (12), antibody–drug conjugates, radiotherapy (6, 7, 13–15), and surgery], where it was available. **Tables 5** and **6** outline the shortcomings of these common eligibility criteria and present key patient responses and safety considerations (e.g., potential risk of and recovery from clinically significant adverse events) that should guide clinical assessment of patient readiness for initiation of a new treatment.

Scientific Rationale for Excluding Certain Medications

As with washout periods, exclusion of concomitant medications during protocol-driven treatment should be supported by scientific rationale. Clearance and elimination of many investigational agents are predictable based on agent type, molecular weight, and/or other physicochemical characteristics. These more predictable agents (e.g., monoclonal antibodies) have known pharmacokinetic properties, and have a very low *a priori* likelihood of being involved in drug interactions. Other drugs under investigation, such as many oral small molecules, have a higher likelihood of being substrates, inducers, or inhibitors of metabolic clearance or transporter pathways, and therefore, must be approached more conservatively when considering which concomitant medications should be allowed. Although the preclinical ability to predict interactions has improved over time, no model or approach has sufficiently replaced dedicated studies in patients (16). Another consideration is actual oral bioavailability of a novel formulation and the effects of coadministration of agents that affect gastric pH (antacids, H2 antagonists, and proton pump inhibitors) and/or gastric emptying (food). Because these are unknown, many trials require patients to fast for 2–8 hours prior to and up to 4 hours following ingestion of an investigational agent, as well as prohibit agents that affect gastric pH. Also, as these drugs are available over the counter and prescribed in up to 55% of patients with cancer, it is important to mitigate the effect on investigational agents as early as possible in development and allow for their use in a general population (17).

Because the presence of concomitant medications can result in drug–drug interactions that affect the safety profile and interpretation of efficacy of an investigational drug, there are understandable concerns regarding loosening restrictions on concomitant medications in clinical trials. Unfortunately, polypharmacy tends to be common in patients with cancer, who also tend to be an older population, with multiple comorbid conditions that may require medical management. A review conducted by LeBlanc and colleagues reported the number of prescribed drugs in patients ranged from 3 to 9.1 (18). Without prior nonclinical knowledge of the potential effects of concomitant medications on investigational drug's pharmacokinetics and pharmacodynamics, many concomitant drugs are prohibited in early-phase clinical trials to ensure patient safety, reduce variability of responses, and ensure optimal conditions for proof of concept. This stringent exclusion of concomitant medications is often duplicated in later phases of drug development without much consideration of how growing nonclinical or clinical knowledge may support broader inclusion of concomitant medications.

Clinical pharmacology studies should be conducted as early as possible in drug development to inform concomitant medication use in eligibility criteria. Formulations of oral investigational agents should be optimized as early as possible in drug development to minimize absorption interactions and pharmacokinetic variability and inform allowance of concomitant medications as early as possible. Concomitant medication allowances should be broadened in later phase trials so that safety is assessed in the premarket setting.

Conclusion

Washout periods and concomitant medication exclusions are common in cancer clinical trial protocols. These exclusion criteria are often applied inconsistently (across trials and between protocol-driven vs. off-protocol treatment) and without evidence to justify their use. Arbitrary or blanket washout period and concomitant medication exclusions should be eliminated. Where there is evidence to support them, clinically relevant washout periods and concomitant medication–related eligibility criteria may be included.

Information gained from preclinical studies and earlier trials about investigational agent adverse event profiles and pharmacology should be incorporated as early as possible in drug development to minimize washout periods and liberalize concomitant medication allowances.

Authors' Disclosures

K.F. Mileham reports personal fees from AstraZeneca, Merck, BMS, Takeda, and Bayer outside the submitted work. C. Moravek reports personal fees and nonfinancial support from AstraZeneca Pharmaceuticals outside the submitted work. E.C. Dees reports personal fees from Novartis, STRATA, and Sanofi and other from MERYX, Pfizer, Apollomics, G1Therapeutics, and Lilly outside the submitted work. S.W. Ebbinghaus reports other from Merck outside the submitted work. T.S. Uldrick reports other from Merck, Roche, and Celgene/BMS outside the submitted work, as well as a patent for US 10,001,483 B2 licensed to Celgene and NCI. No disclosures were reported by the other authors.

Disclaimer

The opinions expressed in this article are those of the authors and do not necessarily reflect the views or policies of the authors' affiliated institutions.

References

1. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 2016;35:185–98.

2. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. J Natl Cancer Inst 2019;111: 245–55.

3. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.

4. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. Drug Metab Rev 2002;34:83–448.

5. Zimmerlin A, Trunzer M, Faller B. CYP3A time-dependent inhibition risk assessment validated with 400 reference drugs. Drug Metab Dispos 2011;39: 1039–46.

6. Blumenthal GM, Pazdur R. Approvals in 2018: a histology-agnostic new molecular entity, novel end points and real-time review. Nat Rev Clin Oncol 2019;16:139–41.

7. Evans SR. Clinical trial structures. J Exp Stroke Transl Med 2010;3:8–18.

8. Carvalho FLF, Zheng C, Witmer K, O'neill J, Lynch JH, Kowalczyk KJ. Complications associated with perioperative use of tyrosine kinase inhibitor in cytoreductive nephrectomy. Sci Rep 2019;9:15272.

9. Le Cesne A, Ray-Coquard I, Bui BN, Adenis A, Rios M, Bertucci F, et al. Discontinuation of imatinib in patients with advanced gastrointestinal stromal tumours after 3 years of treatment: an open-label multicentre randomised phase 3 trial. Lancet Oncol 2010;11:942–9.

10. Weber JS, Dummer R, de Pril V, Lebbé C, Hodi FS. Patterns of onset and resolution of immune-related adverse events of special interest with ipilimumab. Cancer 2013;119:1675–82.

11. Hall KH, Liu Y, Jiang C, Harvey RD. New and worsening long-term immune-related adverse events with PD-1/PD-L1 pathway agents in patients with cancer. Pharmacotherapy 2020;40:133–41.

12. Ryman JT, Meibohm B. Pharmacokinetics of monoclonal antibodies. CPT Pharmacometrics Syst Pharmacol 2017;6:576–88.

13. Lin NU, Prowell T, Tan AR, Kozak M, Rosen O, Amiri-Kordestani L, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Brain Metastases Working Group. J Clin Oncol 2017;35:3760–73.

14. Hayman JA, Callahan JW, Herschtal A, Everitt S, Binns DS, Hicks RJ, et al. Distribution of proliferating bone marrow in adult cancer patients determined using FLT-PET imaging. Int J Radiat Oncol Biol Phys 2011; 79:847–52.

15. Klopp AH, Moughan J, Portelance L, Miller BE, Salehpour MR, Hildebrandt E, et al. Hematologic toxicity in RTOG 0418: a phase 2 study of postoperative IMRT for gynecologic cancer. Int J Radiat Oncol Biol Phys 2013;86: 83–90.

16. Obach RS, Walsky RL, Venkatakrishnan K. Mechanism-based inactivation of human cytochrome P450 enzymes and the prediction of drug-drug interactions. Drug Metab Dispos 2007;35:246–55.

17. Planchard D. Can an acidic beverage reduce interactions between proton pump inhibitors and erlotinib? J Clin Oncol 2016;34:1292–4.

18. LeBlanc TW, McNeil MJ, Kamal AH, Currow DC, Abernethy AP. Polypharmacy in patients with advanced cancer and the role of medication discontinuation. Lancet Oncol 2015;16:e333–41.

19. Chaft JE, Oxnard GR, Sima CS, Kris MG, Miller VA, Riely GJ. Disease flare after tyrosine kinase inhibitor discontinuation in patients with *EGFR*-mutant lung cancer and acquired resistance to erlotinib or gefitinib: implications for clinical trial design. Clin Cancer Res 2011;17:6298–303.

20. Riely GJ, Kris MG, Zhao B, Akhurst T, Milton DT, Moore E, et al. Prospective assessment of discontinuation and reinitiation of erlotinib or gefitinib in patients with acquired resistance to erlotinib or gefitinib followed by the addition of everolimus. Clin Cancer Res 2007;13:5150–5.

21. Turner JP, Shakib S, Singhal N, Hogan-Doran J, Prowse R, Johns S, et al. Prevalence and factors associated with polypharmacy in older people with cancer. Support Care Cancer 2014;22:1727–34.

22. Balducci L, Goetz-Parten D, Steinman MA. Polypharmacy and the management of the older cancer patient. Ann Oncol 2013;24:vii36–vii40.

23. Harvey RD. Managing drug interactions with oral anticancer agents: signals, noise, and echoes. J Oncol Pract 2019;15:91–2.

Modernizing Clinical Trial Eligibility Criteria: Recommendations of the ASCO-Friends of Cancer Research Performance Status Work Group

Allison Magnuson[1], Suanna S. Bruinooge[2], Harpreet Singh[3], Keith D. Wilner[4], Shadia Jalal[5], Stuart M. Lichtman[6], Paul G. Kluetz[3], Gary H. Lyman[7], Heidi D. Klepin[8], Mark E. Fleury[9], Brad Hirsch[10], Allen Melemed[11], Fernanda I. Arnaldez[12], Upal Basu Roy[13], Caroline Schenkel[2], Shimere Sherwood[14], and Elizabeth Garrett-Mayer[2]

ABSTRACT

Purpose: Performance status (PS) is one of the most common eligibility criteria. Many trials are limited to patients with high-functioning PS, resulting in important differences between trial participants and patient populations with the disease. In addition, existing PS measures are subjective and susceptible to investigator bias.

Experimental Design: A multidisciplinary working group of the American Society of Clinical Oncology and Friends of Cancer Research evaluated how PS eligibility criteria could be more inclusive. The working group recommendations are based on a literature search, review of trials, simulation study, and multistakeholder consensus. The working group prioritized inclusiveness and access to investigational therapies, while balancing patient safety and study integrity.

Results: Broadening PS eligibility criteria may increase the number of potentially eligible patients for a given clinical trial, thus shortening accrual time. It may also result in greater participant diversity, potentially reduce trial participant and patient disparities, and enable clinicians to more readily translate trial results to patients with low-functioning PS. Potential impact on outcomes was explored through a simulation trial demonstrating that when the number of Eastern Cooperative Oncology Group PS2 participants was relatively small, the effect on the estimated HR and power was modest, even when PS2 patients did not derive a treatment benefit.

Conclusions: Expanding PS eligibility criteria to be more inclusive may be justified in many cases and could result in faster accrual rates and more representative trial populations.

See related commentary by Giantonio, p. 2369

Introduction

An important goal of the American Society of Clinical Oncology, Friends of Cancer Research, and the oncology community at large is broadening clinical trial eligibility criteria to enhance trial access and accrual, and to ensure trial populations better reflect patients with the disease (1). Performance status (PS) is one of the most common inclusion/exclusion criteria in oncology trials. Many trials are limited to high-functioning participants (i.e., "good" PS) and exclude low-functioning patients (i.e., "poor" PS; ref. 2).

Two main PS scales are utilized in oncology clinical trials: Eastern Cooperative Oncology Group (ECOG; ref. 3) and Karnofsky (KPS) scales (4). Multiple trials in various tumor types and settings have demonstrated that low-functioning PS (i.e., ECOG PS, 2–4 and KPS ≤ 70) is correlated with lower overall survival and progression-free survival compared with high-functioning PS (ECOG PS, 0–1 and KPS, 80–100; refs. 5–13). Because of this, PS is included as a common eligibility criteria and stratification factor. However, this practice prevents trial enrollment for many patients and limits generalizability of trial results. Select trials that have focused exclusively on participants with low-functioning PS demonstrated patient and clinician interest and enrollment (14–17). The underlying etiology for low-functioning PS is also important; for patients whose low-functioning PS is due to disease burden, investigational treatment may result in improved PS with tumor control and symptom alleviation, especially with highly effective treatments. However, current PS scales do not differentiate causes of low-functioning PS.

In addition, there are limitations to PS assessments. PS is inherently subjective, which can affect interrater reliability (18) and invite potential bias particularly for patients at the borderline between values. For example, studies have demonstrated that clinicians assign patients aged >65 years higher numeric PS scores than younger patients, despite no difference in objectively measured physical activity (19). In addition, PS is less predictive of cancer-related outcomes for older adults (20, 21).

Materials and Methods

Because clinical trials frequently exclude PS2 patients, the working group chose to focus on this category. To understand the potential effect of including PS2 patients, the working group conducted a simulation study, where randomized trials of a hypothetical agent were simulated under various conditions. We also examined the literature to identify the potential risks and benefits of including PS2 patients on therapeutic clinical trials and evidence of the effectiveness

[1]University of Rochester Medical Center, Rochester, New York. [2]American Society of Clinical Oncology, Alexandria, Virginia. [3]FDA, Silver Spring, Maryland. [4]Pfizer Inc, New York, New York. [5]Indiana University School of Medicine, Indianapolis, Indiana. [6]Memorial Sloan Kettering Cancer Center, New York, New York. [7]Fred Hutchinson Cancer Research Center, Seattle, Washington. [8]Wake Forest Baptist Medical Center, Winston-Salem, North Carolina. [9]American Cancer Society Cancer Action Network, Washington, D.C. [10]SignalPath, Raleigh, North Carolina. [11]Chimerix, Durham, North Carolina. [12]MacroGenics, Inc, Rockville, Maryland. [13]LUNGevity Foundation, Chicago, Illinois. [14]Association for Clinical Oncology, Alexandria, Virginia.

Corresponding Author: Allison Magnuson, University of Rochester Medical Center, Rochester, NY 14620. Phone: 585-276-7155; E-mail: allison_magnuson@urmc.rochester.edu

Performance status (PS) is one of the most common eligibility criteria, often resulting in exclusion of patients from trial participation and leading to clinical trial populations that are not reflective of populations afflicted with the disease. Existing PS tools are inherently subjective and invite bias. In addition, PS is less predictive of outcomes for older adults. Broadening PS eligibility criteria to be more inclusive can increase the number and diversity of participants in clinical trials. Trial sponsors should justify any exclusion of low-functioning PS patients and limit exclusions to circumstances of participant safety and trial integrity. A multidisciplinary working group of the American Society of Clinical Oncology and Friends of Cancer Research outlined several strategies to encourage broader trial eligibility criteria. Implementation of these recommendations will require cooperation of multiple stakeholders, and providing incentives for expanded PS eligibility may support this effort.

of PS2 as a prognostic factor, reviewed past and current clinical trials to determine how often PS2 was included in inclusion/exclusion criteria, and developed consensus recommendations on how PS eligibility criteria could be revised while ensuring the safety of participants and integrity of the trial, and additional areas for research.

Benefits
Increase number of patients eligible and shorten enrollment time

Small, mainly single institution studies have demonstrated that of patients deemed ineligible for a clinical trial, exclusion was related to poor PS in a significant proportion of patients, with variability across disease type, investigational therapy, and therapy line (22, 23). Even if other objective eligibility measures can be addressed, PS may remain a broad factor that excludes many patients (**Table 1**).

Improve assessment accuracy, particularly in older adults

Most patients with cancer are aged ≥65 years, however, existing PS scales are inadequate in this population (20). Restrictive PS eligibility criteria contribute to the pervasive age disparity between trial participants and the overall cancer population, raising concerns about whether PS is unjustly limiting older populations' ability to participate in trials (24–26). Multiple studies have demonstrated that other tools, such as the geriatric assessment, are better than PS at evaluating older adults' overall health status (27) and at predicting chemotherapy toxicity (20). While a full geriatric assessment may not be practical due to length, subcomponents may provide a better functional assessment, such as instrumental activities of daily living that measure functional independence.

Improve generalizability

Benefits for patients with high-functioning PS may not reflect outcomes for patients with low-functioning PS (28, 29). Many eligibility restrictions from registration trials, such as line of therapy or cancer stage, are incorporated explicitly into the labeled indications with the exception of PS limitations. Therefore, therapies tested only in participants with high-functioning PS are administered to patients with lower functioning PS. This extrapolation may occur more readily with targeted and immunotherapies given greater efficacy (30). Therefore, evaluation of an investigational agent in participants reflective of

Table 1. Risks and benefits of expanding enrollment to patients with worse PS.

	Patients/prescribing physicians	Sponsors/investigators
Benefits	• Earlier access to investigational agents for a larger population of patients • More complete safety and efficacy data to help inform standard-of-care decision-making in the "real world" once the agent is commercially available	• Greater ability to generalize to "real-world" populations • Larger population of potentially eligible patients may afford faster clinical trial accrual times • Efficacy/tolerability in an understudied population provides more informative drug labeling and may facilitate more use in these patients • Higher overall AEs may make PS2 population more sensitive to demonstration of a potential comparative tolerability benefit • Where poor PS is because of advanced disease, benefits in a clinical outcome (survival, symptom, or functional improvement) may be easier to demonstrate for a highly effective drug
Risks	• Potentially higher rates of AEs	• Potentially greater variability in outcomes if not stratified/balanced between treatment groups • Potentially higher rates of AEs/more complicated attribution of AEs; if PS balanced between treatment groups, it should be able to account for this • Diminished treatment effect if PS2 patients do not have the same treatment benefit as patients with good PS

the patient population is important. More inclusive PS eligibility will also likely increase enrollment of older adults (24, 31) and address the lack of evidence noted above (32, 33).

Risks
Increased adverse events

Rates of adverse events (AEs) may be greater in PS2 participants as compared with PS0 and PS1 participants, and this may influence patient's outcomes and ability to comply with study procedures. As a result, investigators and sponsors may be reluctant to consider trial enrollment. PS2 patients risk AEs with standard therapy options as well, and thus participation on a trial may not necessarily pose a greater risk of AEs compared with standard therapy for a particular patient. Because targeted therapies often have higher response rates, PS2 patients may experience a greater therapeutic index in a targeted therapy trial than standard of care (e.g., cytotoxic chemotherapy), even if their absolute rate of AEs is higher than in patients with PS0 and PS1. Where the comparative tolerability between an investigational agent and standard therapy is less clear, including PS2 patients (who may be more sensitive to toxicity) may unmask subtle differences.

Importantly, having a subset of PS2 patients will add important safety data to facilitate decision-making for patients in the post-approval setting (**Table 1**). Determining appropriate timing for including PS2 participants is challenging. When possible, inclusion of a small number of PS2 participants in early-phase trials is recommended to guide separate expansion cohorts for phase II or broader inclusion into registration trials.

Even when clinical trial eligibility allows PS2 patients to enroll, relatively few PS2 participants are actually enrolled (34, 35). This may relate to clinicians' lack of familiarity with the investigational agent and concerns about the tolerability and safety. Enhanced information about safety, tolerability, and efficacy from earlier phase trials with the agent may help to counteract this. In addition, when clinically appropriate, allowing physician discretion in the treatment approach as a component of the clinical trial may help to mitigate this issue (36, 37).

Potential impact on trial outcome data

In trials of novel therapies including PS2 participants, data suggest that outcomes may be inferior compared with participants with PS 0–1, even though low proportions of PS2 participants were included (38–40). This information alone should not be used as a justification for excluding PS2 patients. Instead, similar to other high-risk prognostic markers identified in oncology, PS information could be considered as a stratification factor. When safe, inclusion of participants with low-functioning PS provides valuable evidence to guide clinical care for most patients. Outcomes in low-functioning PS participants can also better inform statistical considerations for future trials.

The risk of inferior outcomes from low-functioning PS participants is a potential concern to sponsors, especially if compared with historical cohorts including high-functioning PS. The FDA has addressed a similar concern in a March 2019 final guidance on enrollment of patients with brain metastases stating, "to mitigate uncertainties about including patients with brain metastases in clinical trials, consider enrolling these patients in a separate subgroup within the trial" (41). In addition, FDA commentary has further indicated a willingness to restrict primary efficacy analysis to the participant subset who meet more conventional eligibility criteria when a sponsor enrolls a broader range of participants (42). FDA also notes that including a broader group of participants could offer benefits, such as additional information in drug labeling and/or reduced postmarketing commitments.

Simulation study methods

To explore the effects on inferences comparing trials that include versus exclude participants with PS2, simulations were conducted under a variety of trial settings with three levels of PS: PS0, PS1, and PS2. **Figure 1** presents results based on: (i) total sample size of 500 participants, (ii) 1:1 randomization to two treatment groups, (iii) accrual time of 24 months, (iv) a time-to-event endpoint, and (v) follow-up until 283 events are observed, achieving power of 85% based on an HR of 0.70 versus a null hypothesis of 1.0 and a two-sided alpha of 0.05. Participants were assumed to vary in their median survival: 12-, 9-, and 6-month median survival in PS0, PS1, and PS2 participants, respectively. Differences in drop-outs due to AEs or other factors varied: 5%, 10%, and 20% of PS0, PS1, and PS2, respectively, and AEs were assumed to have censored event times within the first 4 months. Simulations assumed 45% PS0, 45% PS1, and 10% PS2 participants, and the true HRs reflecting treatment benefit were varied across PS groups. Scenario 1 assumes all three PS groups have the same treatment effect, HR = 0.7. Scenario 2 assumes PS0 and PS1 participants derive benefit, but PS2 participants do not (PS0 and PS1 HR,

0.7 and PS2 HR, 1.0). Scenario 3 assumes PS2 participants derive greater benefit compared with PS0 and PS1 participants (PS0 and PS1 HR, 0.7 and PS2 HR, 0.5). Outcome measures that were assessed to determine the differences in inferences due to the variability in HRs across the groups were (i) the estimated HR, (ii) power, and (iii) time to complete the study because fewer patients would be excluded (measured as the time from the first enrolled participant to the last event required for analysis). Inferences from simulated trials (10,000/scenario) were analyzed under two different approaches: (i) excluding PS2 participants ($N = 450$ PS0 and PS1 patients included in analysis) and (ii) including the PS2 participants ($N = 500$ for analysis). When excluding PS2 participants, the analysis was undertaken when there were 283 events among the PS0 and PS1 participants.

The simulation study demonstrated the following conclusions for including PS2 participants:

(i) when the number of PS2 participants is relatively small (e.g., 10%), the effect on the estimated HR and power is relatively modest, even when the PS2 participants do not have a true treatment benefit (**Fig. 1A** and **B**).
(ii) including PS2 participants is likely to shorten duration of the trial by increasing the number of potentially eligible trial participants (**Fig. 1C**) and due to the higher event rate in PS2 participants relative to PS0–1 participants.

These conclusions may not be generalized to all trial settings. Single-arm trials need attention given that previous trial results (to which the study results will be compared) may not have included PS2 participants. Similarly, trials with smaller (or larger) sample size may have more dramatic or muted effects depending on other trial parameters, such as the fraction of PS2 participants.

Mechanisms for addressing risks associated with expanding PS eligibility criteria

- Assessing safety concerns should take into account the potential increased risk in AE rates between standard-of-care and experimental intervention, rather than the absolute rate of expected AEs.
- Reassess and revise PS eligibility criteria at each phase of drug development, in accordance with growing knowledge about the investigational agent. Early-phase data (AE rates and durable objective responses) for PS2 participants can decrease uncertainty of subsequent randomized trials. For example, trials could:
  (i) include an exploratory PS2 cohort in early-phase trials to collect data without compromising internal validity and to inform inclusion in later phase trials, incorporating early stopping rules for unacceptable toxicity, or
  (ii) if tolerability/safety is acceptable during early phase for PS0–1 participants, expand to include PS2 participants in later phases.
- Consider alternate trial designs and settings. Examples may include:
  (i) trials specifically for PS2 participants and, where appropriate, PS3 participants. This may be most ideal for studies of modified ("deintensified") regimens where the overall goal is to develop a more tolerable therapy.
  (ii) flexibility in the dosing schema, particularly for palliative trials. For example, enable investigator discretion to allow participants to initiate treatment at a reduced dosage with escalation to full dosage based on tolerability (37). This may be most appropriate for studies in advanced cancer where the goal of therapy is palliation.
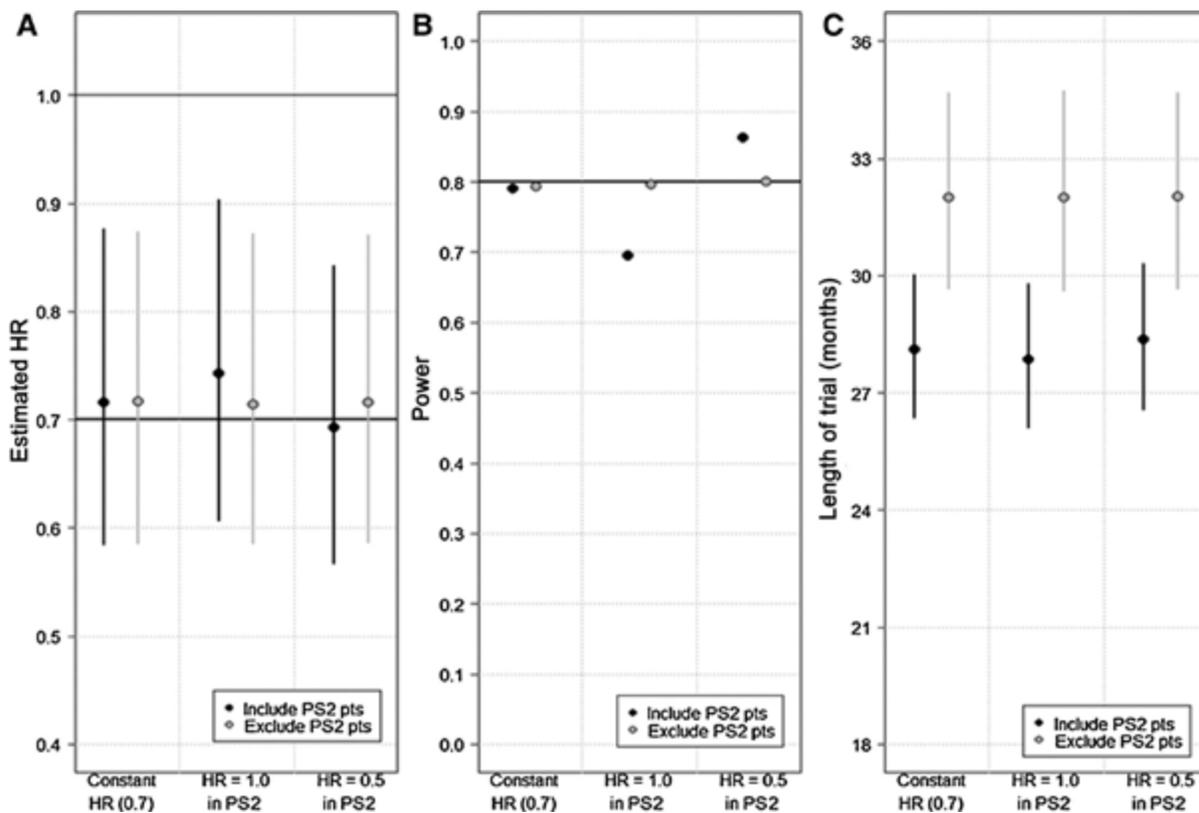
Figure 1.
Simulation study results depicting changes in estimated HR (**A**), power (**B**), and length of trial from first accrual to last required event for analysis (**C**). Within each panel, six analyses are depicted. **A** and **C**, The median from the simulations is plotted as a circle with lines extending vertically to the 5th and 95th percentiles. For each analysis, the HR of PS0 and PS1 patients (pts) remains constant at 0.7 and the HR of PS2 patients is varied. Red points/lines depict results when PS2 patients are included in the final analysis ($N = 500$); blue points/lines depict results when PS2 patients are excluded ($N = 450$). Regardless of sample size, the trial end is assumed to be when the required number of events (283 events) have been accrued per the power calculation.

(iii) Consider expansion cohorts to enhance enrollment of PS2 patients. This may be the most effective strategy for therapies with novel mechanisms or less well-defined AE profiles, whereby initial enrollment includes patients with high-functioning PS and once safety and tolerability are better understood, expansion to include PS2 patients occurs.

(iv) A postmarketing study that focuses on subgroups not well represented in premarket studies (43, 44). This may be most effective strategy for approved therapies where limited data currently exists for patients with low-functioning PS.

• Discuss study design and statistical analysis approaches for broader eligibility and implications for postmarketing research with FDA during trial design, where appropriate. This may include performing simulations under a variety of assumptions regarding fraction of PS2 patients and heterogeneity of efficacy and safety across PS groups.

Recommendations for inclusion of PS2 participants are included in **Table 2**. Although discussion has focused on inclusion of PS2 participants, PS3 participants should also be considered. With targeted therapies for rare alterations, inclusion of PS3 participants may be considered to expand the eligible patient population,

if the agent has demonstrated favorable toxicity and efficacy signals.

Areas of Need for Future Research
Methods to incorporate functional status assessment

Alternate methods for assessing physical function exist, such as patient-reported outcome measures (45), objective performance measures (e.g., gait speed; ref. 46), and activity monitoring devices (e.g., wearable devices; ref. 47). Further research is needed to understand how to incorporate and use these alternative methods in oncology trials. Enhancing the objectivity of PS assessments may more accurately characterize functional capacity and improve trial suitability assessment, particularly if low-functioning PS is related to disease burden versus other factors around the time of diagnosis. Incorporating these methods may also reduce bias of PS assessments.

Associations between PS and safety/toxicity in targeted therapies and immunotherapy

The majority of newly approved investigational agents have targeted mechanisms of action, however, the safety and efficacy of many of these therapies remain unclear in the PS2 population given their

Table 2. Recommendations.

Number	Recommendation
1.	Patients with ECOG PS2 (or KPS 60–70) should be included, unless there is a scientific and/or clinical rationale for exclusion justified by established safety considerations.
	a. PS eligibility criteria should be based on the patient population in which the intervention is expected to be applied in clinical practice.
	b. PS eligibility criteria should be continually reevaluated and modified throughout the drug development process to reflect accumulated safety data of the investigational treatment. Decisions about PS eligibility criteria should be based on early clinical safety and efficacy data about the specific investigational agent or based on known data from other drugs in the same class with similar mechanism of action. Later-phase trials (e.g., phase II/III) should generally mirror the intended use population, and ECOG PS2 (or KPS 60–70) patients should be included unless safety concerns have manifested in earlier-phase trials. The rationale for exclusion should be justified and stated explicitly.
	c. Incorporating the rationale for inclusion of a broader population into the protocol could help encourage investigators to enroll these patients.
	d. PS data should still be collected for use as a stratification factor, regardless of how it is incorporated into eligibility criteria.
2.	Consider alternative trial designs, such as prespecified cohorts with lower functioning PS that are exempt from the primary analysis, to encourage inclusion of these patients. These cohorts would generally be small in size and exploratory in nature and could be enrolled in an incremental way to enable an early stopping rule based upon safety data. Consideration of the data analysis approach for the broader eligibility cohort and subgroup analysis should be determined during the study design phase and its implications for marketing and postmarketing requirements discussed with FDA when appropriate.
3.	Additional assessments of functional status should be considered to better characterize the functional status of ECOG PS2 patients and patients aged ≥65 years, such as ADLs and instrumental ADLs.

Abbreviation: ADLs, activities of daily living.

underrepresentation on clinical trials leading to approval (48). Understanding safety and efficacy of novel therapies in PS2 patients, particularly for patients with low-functioning PS due to disease burden, is a critical area of need, as a targeted therapy or immunotherapy with a high objective response rate may afford improvement in PS by improving disease-related symptoms.

Conclusion

Broadening PS eligibility criteria to be more inclusive can increase the number and diversity of trial participants. More effective biomarker-driven therapies warrant reconsideration of this traditional approach. Trial sponsors should justify exclusion of PS2 patients and limit exclusions to those affecting patient safety and trial integrity. Several strategies can encourage broader inclusion of PS2, and in select cases PS3, participants. Implementation of these recommendations will require cooperation of multiple stakeholders and can result in incentives following FDA approval.

Authors' Disclosures

A. Magnuson reports receiving National Institute on Aging K76 Career Development Award, NCI Loan Repayment Award, and honoraria for educational lecture provided at the American Society of Radiation Oncology 2020 Annual meeting.

K.D. Wilner reports other from Pfizer Inc outside the submitted work. H.D. Klepin reports personal fees from Pfizer outside the submitted work and is a contributor to UpToDate. M.E. Fleury reports grants from Genentech, Bristol Myers Squibb, Merck, and IQVIA outside the submitted work. A. Melemed reports full-time employment with Chimerix, a biopharmaceutical company, and was previously employed by Eli Lilly and Company as a full-time employee. F.I. Arnaldez reports other from MacroGenics, Inc. outside the submitted work. No disclosures were reported by the other authors.

Disclaimer

The opinions expressed in this article are those of the authors and do not necessarily reflect the views or policies of the authors' affiliated institutions.

References

1. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.
2. Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. J Clin Oncol 2017;35:3745–52.
3. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55.
4. Karnofsky D, Burchenal J. The clinical evaluation of chemotherapeutic agents in cancer. Evaluation of chemotherapeutic agents. MacLeod CM, editor. New York, NY: Columbia University Press; 1949.
5. Kawaguchi T, Takada M, Kubo A, Matsumura A, Fukai S, Tamura A, et al. Performance status and smoking status are independent favorable prognostic

factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. J Thorac Oncol 2010;5:620–30.
6. Carey MS, Bacon M, Tu D, Butler L, Bezjak A, Stuart GC. The prognostic effects of performance status and quality of life scores on progression-free survival and overall survival in advanced ovarian cancer. Gynecol Oncol 2008;108:100–5.
7. Ishii H, Okada S, Nose H, Yoshimori M, Aoki K, Okusaka T. Prognostic factors in patients with advanced pancreatic cancer treated with systemic chemotherapy. Pancreas 1996;12:267–71.
8. Sleijfer S, Ouali M, van Glabbeke M, Krarup-Hansen A, Rodenhuis S, Le Cesne A, et al. Prognostic and predictive factors for outcome to first-line ifosfamide-containing chemotherapy for adult patients with advanced soft tissue sarcomas: an exploratory, retrospective analysis on large series from the European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group (EORTC-STBSG). Eur J Cancer 2010;46:72–83.

9. Kodaira M, Takahashi S, Yamada S, Ueda K, Mishima Y, Takeuchi K, et al. Bone metastasis and poor performance status are prognostic factors for survival of carcinoma of unknown primary site in patients treated with systematic chemotherapy. Ann Oncol 2010;21:1163–7.

10. Tas F, Sen F, Odabas H, Kilic L, Keskin S, Yildiz I. Performance status of patients is the major prognostic factor at all stages of pancreatic cancer. Int J Clin Oncol 2013;18:839–46.

11. Arboe B, Halgren Olsen M, Duun-Henriksen AK, Gørløv JS, Nielsen KF, Thomsen RH, et al. Prolonged hospitalization, primary refractory disease, performance status and age are prognostic factors for survival in patients with diffuse large B-cell lymphoma and transformed indolent lymphoma undergoing autologous stem cell transplantation. Leuk Lymphoma 2018;59:1153–62.

12. Song T, Wan Q, Yu W, Li J, Lu S, Xie C, et al. Pretreatment nutritional risk scores and performance status are prognostic factors in esophageal cancer patients treated with definitive chemoradiotherapy. Oncotarget 2017;8:98974–84.

13. Wang JR, Habbous S, Espin-Garcia O, Chen D, Huang SH, Simpson C, et al. Comorbidity and performance status as independent prognostic factors in patients with head and neck squamous cell carcinoma. Head Neck 2016;38:736–42.

14. Langer C, Li S, Schiller J, Tester W, Rapoport BL, Johnson DH. Randomized phase II trial of paclitaxel plus carboplatin or gemcitabine plus cisplatin in Eastern Cooperative Oncology Group performance status 2 non-small-cell lung cancer patients: ECOG 1599. J Clin Oncol 2007;25:418–23.

15. Lilenbaum R, Villaflor VM, Langer C, O'Byrne K, O'Brien M, Ross HJ, et al. Single-agent versus combination chemotherapy in patients with advanced non-small cell lung cancer and a performance status of 2: prognostic factors and treatment selection based on two large randomized clinical trials. J Thorac Oncol 2009;4:869–74.

16. Zukin M, Barrios CH, Rodrigues Pereira J, De Albuquerque Ribeiro R, de Mendonça Beato CA, do Nascimento YN, et al. Randomized phase III trial of single-agent pemetrexed versus carboplatin and pemetrexed in patients with advanced non-small-cell lung cancer and Eastern Cooperative Oncology Group performance status of 2. J Clin Oncol 2013;31:2849–53.

17. Spigel DR, Hainsworth JD, Joseph MJ, Shipley DL, Hagan MK, Thompson DS, et al. Randomized phase 2 trial of pemetrexed, pemetrexed/bevacizumab, and pemetrexed/carboplatin/bevacizumab in patients with stage IIIB/IV non-small cell lung cancer and an Eastern Cooperative Oncology Group performance status of 2. Cancer 2018;124:1982–91.

18. Chow R, Bruera E, Temel JS, Krishnan M, Im J, Lock M. Inter-rater reliability in performance status assessment among healthcare professionals: an updated systematic review and meta-analysis. Support Care Cancer 2020;28:2071–8.

19. Broderick JM, Hussey J, Kennedy MJ, O' Donnell DM. Patients over 65 years are assigned lower ECOG PS scores than younger patients, although objectively measured physical activity is no different. J Geriatr Oncol 2014;5:49–56.

20. Hurria A, Togawa K, Mohile SG, Owusu C, Klepin HD, Gross CP, et al. Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study. J Clin Oncol 2011;29:3457–65.

21. Ghosn M, Ibrahim T, El Rassy E, Nassani N, Ghanem S, Assi T. Abridged geriatric assessment is a better predictor of overall survival than the Karnofsky performance scale and physical performance test in elderly patients with cancer. J Geriatr Oncol 2017;8:128–32.

22. Cancer Action Network, American Cancer Society. Barriers to patient enrollment in therapeutic clinical trials for cancer: a landscape report. Available from: https://www.fightcancer.org/policy-resources/clinical-trial-barriers.

23. Lara PN, Higdon R, Lim N, Kwan K, Tanaka M, Lau DHM, et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. J Clin Oncol 2001;19:1728–33.

24. Ludmir EB, Mainwaring W, Lin TA, Miller AB, Jethanandani A, Espinoza AF, et al. Factors associated with age disparities among cancer clinical trial participants. JAMA Oncol 2019 Jun 3 [Epub ahead of print].

25. Canouï-Poitrine F, Lièvre A, Dayde F, Lopez-Trabada-Ataz D, Baumgaertner I, Dubreuil O, et al. Inclusion of older patients with cancer in clinical trials: the SAGE Prospective Multicenter Cohort Survey. Oncologist 2019;24:e1351–9.

26. Department of Health, Education, and Welfare; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: ethical principles and guidelines for the protection of human subjects of research. J Am Coll Dent 2014;81:4–13.

27. Repetto L, Fratino L, Audisio RA, Venturino A, Gianni W, Vercelli M, et al. Comprehensive geriatric assessment adds information to Eastern Cooperative Oncology Group performance status in elderly cancer patients: an Italian Group for Geriatric Oncology Study. J Clin Oncol 2002;20:494–502.

28. Azad AA, Eigl BJ, Leibowitz-Amit R, Lester R, Kollmannsberger C, Murray N, et al. Outcomes with abiraterone acetate in metastatic castration-resistant prostate cancer patients who have poor performance status. Eur Urol 2015;67:441–7.

29. Blackhall F, Ross Camidge D, Shaw AT, Soria J-C, Solomon BJ, Mok T, et al. Final results of the large-scale multinational trial PROFILE 1005: efficacy and safety of crizotinib in previously treated patients with advanced/metastatic ALK-positive non-small-cell lung cancer. ESMO Open 2017;2:e000219.

30. Parikh RB, Galsky MD, Gyawali B, Riaz F, Kaufmann TL, Cohen AB, et al. Trends in checkpoint inhibitor therapy for advanced urothelial cell carcinoma at the end of life: insights from real-world practice. Oncologist 2019;24:e397–9.

31. Talarico L, Chen G, Pazdur R. Enrollment of elderly patients in clinical trials for cancer drug registration: a 7-year experience by the US Food and Drug Administration. J Clin Oncol 2004;22:4626–31.

32. Hurria A, Levit LA, Dale W, Mohile SG, Muss HB, Fehrenbacher L, et al. Improving the evidence base for treating older adults with cancer: American Society of Clinical Oncology Statement. J Clin Oncol 2015;33:3826–33.

33. Levit LA, Singh H, Klepin HD, Hurria A. Expanding the evidence base in geriatric oncology: action items from an FDA-ASCO workshop. J Natl Cancer Inst 2018;110:1163–70.

34. Solomon BJ, Mok T, Kim D-W, Wu Yi-L, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. N Engl J Med 2014;371:2167–77.

35. Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee J-L, Fong L, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. N Engl J Med 2017;376:1015–26.

36. von Gruenigen VE, Huang HQ, Beumer JH, Lankes HA, Tew W, Herzog T, et al. Chemotherapy completion in elderly women with ovarian, primary peritoneal or fallopian tube cancer–an NRG oncology/Gynecologic Oncology Group study. Gynecol Oncol 2017;144:459–67.

37. Seymour MT, Thompson LC, Wasan HS, Middleton G, Brewster AE, Shepherd SF, et al. Chemotherapy options in elderly and frail patients with metastatic colorectal cancer (MRC FOCUS2): an open-label, randomised factorial trial. Lancet 2011;377:1749–59.

38. Mok TS, Wu Yi-L, Thongprasert S, Yang C-H, Chu Da-T, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947–57.

39. Zhou C, Wu Yi-L, Chen G, Feng J, Liu X-Q, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011;12:735–42.

40. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012;13:239–46.

41. Enhancing the diversity of clinical trial populations - eligibility criteria, enrollment practices, and trial designs guidance for industry - guidance document. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial.

42. Beaver JA, Ison G, Pazdur R. Reevaluating eligibility criteria - balancing patient protection and participation in oncology trials. N Engl J Med 2017;376:1504–5.

43. Hurria A, Dale W, Mooney M, Rowland JH, Ballman KV, Cohen HJ, et al. Designing therapeutic clinical trials for older and frail adults with cancer: U13 conference recommendations. J Clin Oncol 2014;32:2587–94.

44. Nipp RD, Yao N(A), Lowenstein LM, Buckner JC, Parker IR, Gajra A, et al. Pragmatic study designs for older adults with cancer: report from the U13 conference. J Geriatr Oncol 2016;7:234–41.

45. Kluetz PG, Slagle A, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, et al. Focusing on core patient-reported outcomes in cancer clinical trials: symptomatic adverse events, physical function, and disease-related symptoms. Clin Cancer Res 2016;22:1553–8.

46. Liu MA, DuMontier C, Murillo A, Hshieh TT, Bean JF, Soiffer RJ, et al. Gait speed, grip strength, and clinical outcomes in older patients with hematologic malignancies. Blood 2019;134:374–82.

47. Kelly CM, Shahrokni A. Moving beyond Karnofsky and ECOG performance status assessments with new technologies. J Oncol 2016;2016:6186543.

48. Cheng S, Qureshi M, Pullenayegum E, Haynes A, Chan KK. Do patients with reduced or excellent performance status derive the same clinical benefit from novel systemic cancer therapies? A systematic review and meta-analysis. ESMO Open 2017;2:e000225.

Impact of Broadening Trial Eligibility Criteria for Patients with Advanced Non–Small Cell Lung Cancer: Real-World Analysis of Select ASCO-*Friends* Recommendations

R. Donald Harvey[1], Suanna S. Bruinooge[2], Li Chen[3], Elizabeth Garrett-Mayer[2], Whitney Rhodes[3], Edward Stepanski[3], Thomas S. Uldrick[4], Gwynn Ison[5], Sean Khozin[6], Wendy S. Rubinstein[5], Caroline Schenkel[2], Robert S. Miller[7], George A. Komatsoulis[7], Richard L. Schilsky[2], and Edward S. Kim[8]

ABSTRACT

Purpose: Cancer clinical trials often accrue slowly or miss enrollment targets. Strict eligibility criteria are a major reason. Restrictive criteria also limit opportunities for patient participation while compromising external validity of trial results. We examined the impact of broadening select eligibility criteria on characteristics and number of patients eligible for trials, using recommendations of the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research.

Experimental Design: A retrospective, observational analysis used electronic health record data from ASCO's CancerLinQ Discovery database. Study cohort included patients with advanced non–small cell lung cancer treated from 2011 to 2018. Patients were grouped by traditional criteria [no brain metastases, no other malignancies, and creatinine clearance (CrCl) ≥ 60 mL/minute] and broadened criteria (including brain metastases, other malignancies, and CrCl ≥ 30 mL/minute).

Results: The analysis cohort included 10,500 patients. Median age was 68 years, and 73% of patients were White. Most patients had stage IV disease (65%). A total of 5,005 patients (48%) would be excluded from trial participation using the traditional criteria. The broadened criteria, however, would allow 98% of patients (10,346) to be potential participants. Examination of patients included by traditional criteria (5,495) versus those added (4,851) by broadened criteria showed that the number of women, patients aged 75+ years, and those with stage IV cancer was significantly greater using broadened criteria.

Conclusions: This analysis of real-world data demonstrated that broadening three common eligibility criteria has the potential to double the eligible patient population and include trial participants who are more representative of those encountered in practice.

See related commentary by Giantonio, p. 2369

Introduction

Numerous cancer trials accrue slowly or miss enrollment targets (range, 9%–49% of trials; refs. 1–7) due to strict eligibility criteria. A 2016 analysis of corrective actions for poor-accruing trials found that eligibility criteria were among primary causes of enrollment delays, with broadening eligibility criteria as the primary remedy for phase II trials (8). Among 231 phase I trials (1991–2016), common reasons for exclusion were performance status (PS) ≥ 1, brain metastases, and strict renal/hepatic function requirements (9). These restrictions were associated with fewer eligible patients, longer enrollment periods (26 vs. 17 months), and increased study terminations.

Broadening eligibility criteria also addresses study design factors that limit study participation, thereby causing inequities in trial access particularly among certain populations and creating concerns about external validity of results. A 2019 review of trial participation indicated 21.5% of patients were excluded primarily because of strict eligibility criteria (10). An examination of real-world data (RWD) from Denmark showed that clinical characteristics excluded 61% of patients with melanoma from pivotal trials during 2010–2015; brain metastases and/or PS ≥ 2 affected 75% of those excluded (11). These patients (when treated following approval with study agents) showed improvement in outcomes versus historical controls.

In October 2017, the American Society of Clinical Oncology (ASCO) and Friends of Cancer Research (*Friends*) published recommendations to broaden eligibility criteria (12) for: brain metastases (13), organ function, primary/concurrent malignancies (14), human immunodeficiency virus status (15), and minimum enrollment age (16). This investigation quantified and characterized among patients with advanced non–small cell lung cancer (aNSCLC) trial eligibility using the ASCO-*Friends*' broadened versus traditional eligibility criteria.

Materials and Methods

This retrospective, observational study examined common eligibility criteria in cancer trials: (i) brain metastases, (ii) renal function, and (iii) prior/concurrent malignancies. Data for the analysis were obtained from CancerLinQ Discovery a safe-harbor deidentified dataset compiled from electronic health records (EHR) of 50 U.S. oncology

[1]Winship Cancer Institute of Emory University, Druid Hills, Georgia. [2]ASCO, Alexandria, Virginia. [3]ConcertAI, Boston, Massachusetts. [4]Fred Hutchinson Cancer Research Center and University of Washington, Seattle, Washington. [5]FDA, Silver Spring, Maryland. [6]Janssen Research and Development, New York, New York. [7]ASCO CancerLinQ, Alexandria, Virginia. [8]Levine Cancer Institute, Atrium Health, Charlotte, North Carolina.

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

R.D. Harvey and S.S. Bruinooge contributed equally as co-first authors of this article.

R.L. Schilsky and E.S. Kim contributed equally as co-senior authors of this article.

Corresponding Author: Suanna S. Bruinooge, American Society of Clinical Oncology, 2318 Mill Road, Suite 800, Alexandria, VA 22314. Phone: 571-483-1313; E-mail: suanna.bruinooge@asco.org

Translational Relevance

Overly restrictive clinical trial eligibility criteria make it challenging to translate research findings to all populations likely to receive a new treatment following approval. Less restrictive eligibility criteria over the course of drug development may generate data on a broader population and improve speed of accrual. This may be accomplished by progressive broadening of eligibility criteria across trial phases. We show that expansion of three common eligibility criteria, renal function measures, presence of brain metastases, and history of prior malignancy, increases patients potentially eligible in the dataset analyzed by almost 2-fold. While this analysis was conducted in a population with advanced non–small cell lung cancer, the findings are likely applicable to other advanced malignancies. These expanded eligibility criteria should be widely adopted while pursuing additional expanded inclusion criteria to generate findings more relevant to patients treated in routine clinical practice, maximize patient access to trials, and expedite trial enrollment.

practices (17). We included standardized data and data curated by trained clinical data abstractors from 2011 to 2018 records (18–20).

Study criteria included patients with aNSCLC diagnosis (stage IIIb, IIIc, or IV; see Supplementary Materials and Methods), receipt of systemic therapy, and ≥2 documented clinical visits. Patients with missing serum creatinine laboratory values were excluded.

Criteria for traditional and broadened eligibility criteria are outlined in **Table 1**. Traditional criteria excluded patients with creatinine clearance (CrCl) ≤ 60 mL/minute (21). Broadened criteria included patients with CrCl ≥ 30 mL/minute. A minority of cases (33%) included CrCl in EHR data. CrCl was calculated for 7,031 cases using the Cockcroft–Gault equation (22, 23).

Patients with additional cancer diagnosis codes unrelated to NSCLC were classified as having a prior/concurrent cancer. These patients would be excluded by traditional criteria and included by broadened criteria. All diagnosis codes related to lung cancer metastases sites (e.g., adrenal gland, bone, brain, and other) were considered metastases,

rather than another cancer. From a clinical perspective, metastases may be more likely than second primary cancers at these anatomic sites. Miscoding of metastases as primary cancers is not infrequent.

Data curation identified patients with brain metastases, including coding primary brain neoplasms as brain metastases. All patients with brain metastases were excluded under traditional criteria and included under broadened criteria.

PS values were presented using the Eastern Cooperative Oncology Group (ECOG) scale as documented in EHR or converted from Karnofsky (24). If multiple PS values existed, we used the value closest to date of therapy initiation.

Descriptive statistics summarize the two populations, including proportions, means, and interquartile ranges (IQR). Comparisons were made between patients included on the basis of traditional criteria versus patients excluded on the basis of traditional criteria, but included using broadened criteria (i.e., independent, nonoverlapping patient groups) using χ^2 tests. Alpha was set at 0.01 due to the large sample size. Data management and analyses were conducted in Python 3.7.0 and R 3.5.1.

Results

A total of 10,500 cases were included in the analysis (**Fig. 1**; **Table 2**). Median age was 68 years (IQR, 60–74), and 56% were males. A total of 75% of patients were White. Most patients had stage IV disease (65%).

Of the total cohort, 1,509 (14%) patients had prior/concurrent cancers (**Table 3**), most commonly prostate (154 patients, 2%), colorectal (120, 1%), and breast (31, 0.3%) cancers. These 1,509 patients would be excluded under traditional criteria, but included using broadened criteria. All cases were coded for presence/absence of brain metastases. A total of 21% of patients (2,226) had brain metastases and would be excluded by traditional eligibility criteria.

Overall, 5,005 patients (48%) were excluded by one or more of three traditional criteria, leaving only 5,495 (52%) eligible. More than 20% of patients (2,252) were excluded by traditional eligibility criteria due to CrCl ≤ 60 mL/minute alone. Use of the broadened criteria would only exclude 154 patients (1.5%), leaving nearly all patients (10,346, 98%) potentially eligible.

Table 1. Comparison of definitions for traditional clinical trial eligibility criteria, ASCO-*Friends*' broadened criteria, and criteria used in study.

	Traditional eligibility criteria	ASCO-*Friends*' broadened criteria	Criteria used in study
Prior and concurrent cancer: in addition to NSCLC	Exclude patients with another primary cancer in 2 years prior to trial enrollment	Include patients with another primary cancer that does not interfere with safety or efficacy of study therapy	Included all cases with another primary cancer diagnosis: (i) counted primary diagnostic codes at sites of likely NSCLC metastases as metastases
Brain metastases	Exclude patients with brain metastases	Include patients with treated and/or stable brain metastases, as well as patients with active brain metastases	Included all patients with brain metastases (i) irrespective of treatment status and clinical stability (ii) counted primary brain diagnostic codes as metastases
Renal function	Exclude patients if CrCl ≤ 60 mL/minute	Include patients if CrCl ≥ 30 mL/minute for study therapy without kidney toxicity	Included patients if CrCl ≥ 30 mL/minute: (i) used Cockcroft–Gault formula to calculate CrCl for patients without evidence of CrCl measure
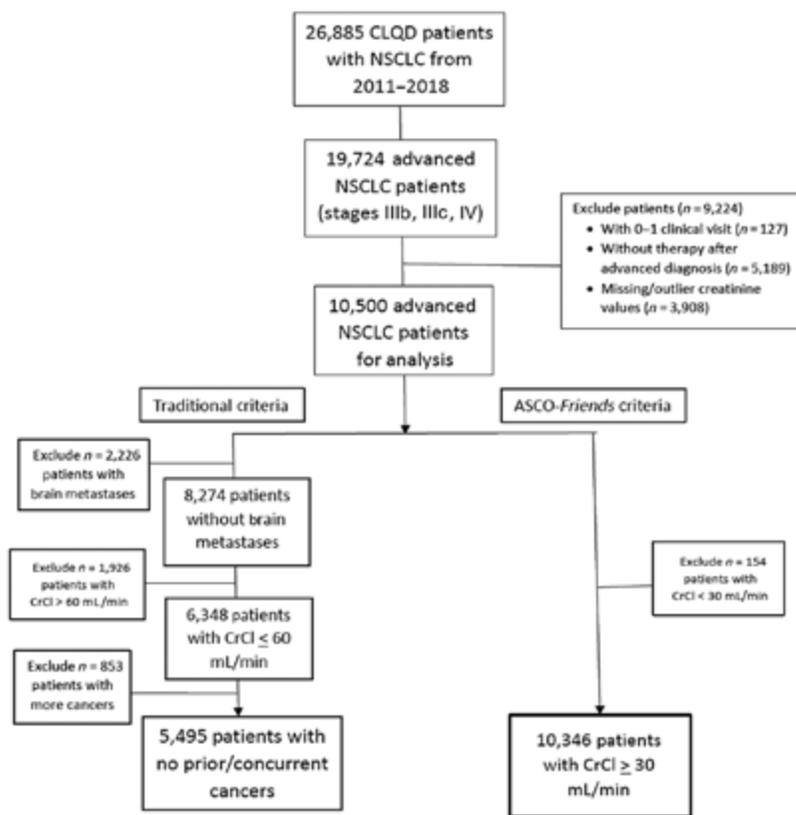
Figure 1.
Cohort comparison with traditional and broadened eligibility criteria.

Patients included by traditional criteria (5,495) versus patients added (4,851) by broadened criteria and excluded by traditional criteria (**Table 2**; columns E and G) differ in important ways (**Fig. 2**). The percentage of women was significantly different; 40% under traditional criteria versus 48% ($P < 0.001$) with broadened criteria. The percentage of patients aged 75+ years was significantly greater with broadened criteria (29% vs. 16%; $P < 0.001$). Comparing by stage, 59% of stage IV patients were included with traditional versus 72% ($P < 0.001$) using broadened criteria. Percentage of patients with ECOG PS 2+ was similar (18% vs. 20%; $P = 0.03$).

Discussion

Broadening three common eligibility criteria can potentially double the number of patients with aNSCLC eligible for trials. Prior analyses of eligibility of patients with aNSCLC demonstrated 60% were ineligible, with common exclusions being brain metastases and poor PS (25).

Support for expanding eligibility criteria examined also comes from analysis of Kaiser Permanente data, which showed 8% of patients would be excluded from trials because of another invasive cancer within 5 years (14). The analysis also revealed 28% of patients with lung, 20% with breast, 25% with colorectal, and 46% with bladder cancers would be excluded because of CrCl < 60 mL/minute. Renal function is of critical importance in aNSCLC, because carboplatin, pemetrexed, and cisplatin are renally cleared. Expanding CrCl eligibility to ≥30 mL/minute would substantially impact the eligible population, adding 20% of patients. It is important to recognize those

instances when including patients with CrCl < 60 mL/minute should be avoided, specifically in studies of drugs cleared by the kidneys and without established dose adjustments where drugs cause direct renal toxicity. In other cases, the change to include those with CrCl > 30 mL/minute should be employed once safety is established (perhaps in an exploratory cohort in early development) and certainly in late-phase trials.

In this analysis, the population who met the broadened eligibility criteria are more representative of patients with aNSCLC than the traditional eligibility criteria population. The broadened population included more women, older patients, and/or patients with stage IV disease. Although broadened criteria resulted in a small increase in patients with PS 2+, analysis of PS was inconclusive. Most records (58.5%) lacked structured PS data. Translation of data from highly selected trial populations to patients seen in real-world practice identifies important knowledge gaps and increases confidence in applying trial results to typical patients.

While our analysis demonstrated an increase in the number of patients with aNSCLC potentially eligible for trials, their inclusion could also potentially affect interpretation of safety and efficacy data because of increased heterogeneity. Similar studies including broader populations, however, demonstrated similar safety and survival rates between restricted and broadened populations (9, 25). Our analysis is limited by characteristics of our data source. The population of patients included in CancerLinQ has not been compared with the U.S. cancer population, although CancerLinQ participating practices are geographically diverse and mostly outside academic settings. It was also difficult to match eligibility criteria to EHR data. We simplified the

	(A) All patients with aNSCLC	(B) Patients excluded due to brain metastases	(C) Patients excluded due to other prior or concurrent malignancies	(D) Patients excluded due to CrCl ≤ 60 mL/minute	(E) Patients included by 3 traditional criteria	(F) Patients included by 3 broadened criteria	(G) Patients included by broadened and excluded from traditional	(H) Patients excluded by both broadened and traditional	P^a
	10,500 (100)	2,226 (21)	1,509 (14)	2,254 (22)	5,495 (52)	10,346 (99)	4,851 (46)	154 (2)	
Age (median and IQR) at treatment index									
	68 (60–74)	65 (57–71)	69 (62–75)	76 (70–81)	66 (59–72)	68 (60–74)	69 (62–76)	78 (72–84)	
Age (at treatment index)									<0.001
≤49 years	426 (4)	140 (6)	48 (3)	7 (3)	254 (5)	425 (4)	171 (4)	1 (1)	
50–64 years	3,794 (36)	1010 (45)	484 (32)	246 (11)	2,268 (41)	3,781 (37)	1,513 (31)	13 (8)	
65–74 years	3,881 (37)	752 (34)	595 (39)	794 (35)	2,089 (38)	3,840 (37)	1,751 (36)	41 (27)	
75+ years	2,399 (23)	324 (15)	382 (25)	1,207 (54)	884 (16)	2,300 (22)	1,416 (29)	99 (64)	
Sex									<0.001
Female	4,647 (44)	1,086 (49)	624 (41)	1,203 (53)	2,216 (40)	4,550 (44)	2,334 (48)	97 (63)	
Male	5,853 (56)	1,140 (51)	885 (59)	1,051 (47)	3,279 (60)	5,796 (56)	2,517 (52)	57 (37)	
Race									0.49
White	6,813 (73)	1,365 (70)	1,009 (76)	1,505 (75)	3,567 (74)	6,716 (74)	3,149 (73)	97 (68)	
Black	1,255 (14)	241 (12)	163 (12)	320 (16)	660 (14)	1,223 (13)	563 (13)	32 (22)	
Other	1,206 (13)	357 (18)	158 (12)	180 (9)	615 (13)	1,192 (13)	577 (13)	14 (10)	
Unknown	1,226	263	179	249	653	1,215	562	11 (<1)	
Stage at index (closest to first-line advanced treatment)									<0.001
Stage IIIb/c	3,355 (35)	334 (16)	405 (30)	710 (34)	2,086 (41)	3,316 (35)	1,230 (28)	39 (27)	
Stage IV	6,354 (65)	1,714 (84)	965 (70)	1,352 (66)	3,039 (59)	6,251 (65)	3,212 (72)	103 (73)	
Unknown	791	178	139	192	370	779	409	12	
PS (on or ≤1 year prior to treatment index date; native ECOG or translated from Karnofsky)									0.06^b
0	1,215 (28)	282 (29)	188 (30)	231 (22)	637 (29)	1,199 (28)	562 (27)	16 (18)	
1	2,121 (49)	490 (50)	289 (46)	495 (48)	1,058 (49)	2,079 (49)	1,021 (49)	42 (47)	
2	831 (19)	174 (18)	126 (20)	248 (24)	384 (18)	805 (19)	421 (20)	26 (29)	
3	182 (4)	37 (4)	27 (4)	59 (6)	83 (4)	178 (4)	95 (5)	4 (4)	
4	12 (<1)	1 (<1)	1 (<1)	2 (<1)	8 (<1)	11 (<1)	3 (<1)	1 (1)	
Unknown	6,139	1,242	878	1,219	3,325	6,074	2,749	65	
Smoking status (at treatment index)									<0.001
Never smoked	1,642 (16)	414 (19)	295 (21)	408 (19)	732 (14)	1,616 (16)	884 (19)	26 (17)	
Former smoker	5,225 (52)	978 (46)	708 (49)	1,195 (56)	2,802 (54)	5,145 (52)	2,343 (51)	80 (52)	
Current smoker	3,093 (31)	746 (35)	430 (30)	543 (25)	1,650 (32)	3,057 (31)	1,407 (30)	36 (23)	
Unknown	540	88	76	108	311	528	217	12	

Note: N (%) are shown in table cells, except for age, which is presented as median and IQR. When calculating percentages, "unknown" cell counts were excluded.
[a] χ^2 tests were used to calculate P values with all categories included, except "unknown."
[b] Comparing PS < 2 versus PS ≥ 2, P = 0.03.

Table 3. Numbers of patients excluded by traditional versus broadened clinical trial eligibility criteria.

Original cohort	10,500 (100%)
Traditional criteria	
Pts excluded due to brain metastases	2,226 (21.2%)
Pts excluded due to prior/concurrent cancers	1,509 (14.4%)
Pts excluded because CrCl ≤ 60 mL/minute	2,254 (21.5%)
Pts excluded by one or more of 3 traditional criteria	5,005 (47.7%)
ASCO-*Friends*' broadened criteria	
Pts excluded by brain metastases and prior/ concurrent cancers	0 (0%)
Pts excluded by CrCl ≤ 30 mL/minute cutoff	154 (1.5%)

Abbreviation: Pt, patient.

definition of brain metastases to present or absent, rather than using ASCO-*Friends*' recommendations' consideration of treated and/or stable metastases. Classification of diagnosis codes for another primary cancer at metastases sites as NSCLC metastases had a small impact (<1%). Record curation helped ensure very few patients were mischaracterized.

Conclusions

Broadening three common eligibility criteria would have allowed nearly twice as many patients (5,495 vs. 10,346) within this RWD analysis to meet eligibility criteria for clinical trials. With broadened eligibility, patients are more representative of the range of patients with aNSCLC, thus increasing the likelihood for definitive trials. Per ASCO and *Friends*' recommendations, eligibility criteria should be carefully selected and reflect safety concerns and compelling scientific rationale
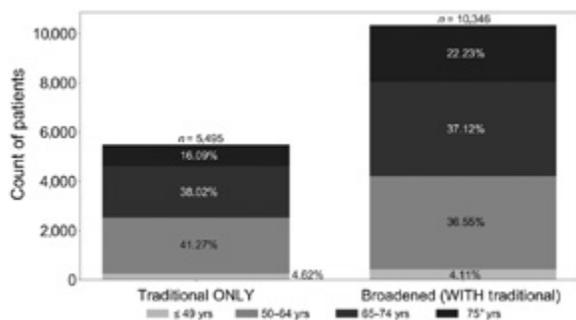
Figure 2.
Effect of traditional versus broadened eligibility criteria by ages represented ($N = 10,500$ patients with aNSCLC).

specific to the investigational therapy. Broadening eligibility criteria will enable improved equitable patient involvement in research and likely accelerate trial enrollment.

References

1. Cheng S, Dietrich M, Finnigan S, Sandler A, Crites J, Ferranti L, et al. A sense of urgency: evaluating the link between clinical trial development time and the accrual performance of CTEP-sponsored studies. J Clin Oncol 27:325s, 2009 (suppl; abstr CRA6509).
2. Institute of Medicine. A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. Washington, DC: The National Academies Press; 2010, p.143.
3. Cheng SK, Dietrich MS, Dilts DM. A sense of urgency: evaluating the link between clinical trial development time and the accrual performance of Cancer Therapy Evaluation Program (NCI-CTEP) sponsored studies. Clin Cancer Res 2010;16:5557–63.
4. Schroen AT, Petroni GR, Wang H, Thielen MJ, Gray R, Benedetti J, et al. Achieving sufficient accrual to address the primary endpoint in phase III clinical trials from U.S. Cooperative Oncology Groups. Clin Cancer Res 2012;18:256–62.
5. Korn EL, Freidlin B, Mooney M, Abrams JS. Accrual experience of National Cancer Institute Cooperative Group phase III trials activated from 2000 to 2007. J Clin Oncol 2010;28:5197–201.
6. Bennette CS, Ramsey SD, McDermott CL, Carlson JJ, Basu A, Veenstra DL. Predicting low accrual in the National Cancer Institute's Cooperative Group clinical trials. J Natl Cancer Inst 2016;108:djv324.
7. Demmy TL, Yasko JM, Collyar DE, Katz ML, Krasnov CL, Borwhat MJ, et al. Managing accrual in Cooperative Group clinical trials. J Clin Oncol 2004;22: 2997–3002.
8. Massett HA, Mishkin G, Rubinstein L, Ivy SP, Denicoff A, Godwin E, et al. Challenges facing early phase trials sponsored by the National Cancer Institute: an analysis of corrective action plans to improve accrual. Clin Cancer Res 2016; 22:5408–16.
9. Malik L, Lu D. Eligibility criteria for phase I clinical trials: tight vs loose? Cancer Chemother Pharmacol 2019;83:999–1002.
10. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. JNCI 2019;111:245–55.
11. Donia M, Ellebaek E, Øllegaard TH, Duval L, Aaby JB, Hoejberg L, et al. The real-world impact of modern treatments on the survival of patients with metastatic melanoma. Eur J Cancer 2019;108:25–32.
12. Kim ES, Bruinooge SS, Roberts S, Ison G, Lin NU, Gore L, 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;35:3737–44.
13. Lin NU, Prowell T, Tan AR, Kozak M, Rosen O, Amiri-Kordestani L, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Brain Metastases Working Group. J Clin Oncol 2017;35:3760–73.
14. Lichtman SM, Harvey RD, Damiette Smit M-A, Rahman A, Thompson MA, Roach N, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Organ Dysfunction, Prior or Concurrent Malignancy, and Comorbidities Working Group. J Clin Oncol 2017;35:3753–9.
15. Uldrick TS, Ison G, Rudek MA, Noy A, Schwartz K, Bruinooge S, et al. Modernizing clinical trial eligibility criteria: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research HIV Working Group. J Clin Oncol 2017;35:3774–80.
16. Gore L, Ivy SP, Balis FM, Rubin E, Thornton K, Donoghue M, et al. Modernizing clinical trial eligibility: recommendations of the American Society of Clinical Oncology–Friends of Cancer Research Minimum Age Working Group. J Clin Oncol 2017;35:3781–7.
17. ASCO CancerLinQ. LLC: Lung Data (2011–2018), November 30, 2018 dataset. Available from: www.cancerlinq.org/solutions/researchers.
18. ASCO CancerLinQ. Available from: www.cancerlinq.org/solutions/researchers.
19. Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. J Clin Oncol 2014;32:2373–9.
20. Shah A, Stewart AK, Kolacevski A, Michels D, Miller R. Building a rapid learning health care system for oncology: why CancerLinQ collects identifiable health information to achieve its vision. J Clin Oncol 2016;34:756–63.
21. Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. J Clin Oncol 2017;35:3745–52.
22. Coresh J, Stevens LA. Kidney function estimating equations: where do we stand? Curr Opin Nephrol Hypertens 2006;15:276–84.
23. Stevens LA, Li S, Kurella Tamura M, Chen S-C, Vassalotti JA, Norris KC, et al. Comparison of the CKD Epidemiology Collaboration (CKD-EPI) and Modification of Diet in Renal Disease (MDRD) study equations: risk factors for and complications of CKD and mortality in the Kidney Early Evaluation Program (KEEP). Am J Kidney Dis 2011;57:S9–S16.
24. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55.
25. Kawachi H, Fujimoto D, Morimoto T, Ito M, Teraoka S, Sato Y, et al. Clinical characteristics and prognosis of patients with advanced non–small-cell lung cancer who are ineligible for clinical trials. Clin Lung Cancer 2018; 19:721–4.

The Friends of Cancer Research Real-World Data Collaboration Pilot 2.0: Methodological Recommendations from Oncology Case Studies

Donna R. Rivera[1], Henry J. Henk[2], Elizabeth Garrett-Mayer[3], Jennifer B. Christian[4], Andrew J. Belli[5], Suanna S. Bruinooge[3], Janet L. Espirito[6], Connor Sweetnam[7], Monika A. Izano[7], Yanina Natanzon[8], Nicholas J. Robert[6], Mark S. Walker[8], Aaron B. Cohen[9], Marley Boyd[6], Lindsey Enewold[10], Eric Hansen[5], Rebecca Honnold[11], Lawrence Kushi[12], Pallavi S. Mishra Kalyani[1], Ruth Pe Benito[11], Lori C. Sakoda[12], Elad Sharon[10], Olga Tymejczyk[9], Emily Valice[12], Joseph Wagner[4], Laura Lasiter[13] and Jeff D. Allen[13],*

The purpose of this study was to evaluate the potential collective opportunities and challenges of transforming real-world data (RWD) to real-world evidence for clinical effectiveness by focusing on aligning analytic definitions of oncology end points. Patients treated with a qualifying therapy for advanced non-small cell lung cancer in the frontline setting meeting broad eligibility criteria were included to reflect the real-world population. Although a trend toward improved outcomes in patients receiving PD-(L)1 therapy over standard chemotherapy was observed in RWD analyses, the magnitude and consistency of treatment effect was more heterogeneous than previously observed in controlled clinical trials. The study design and analysis process highlighted the identification of pertinent methodological issues and potential innovative approaches that could inform the development of high-quality RWD studies.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ The use of real-world evidence (RWE) in drug development is expanding and various applications are being investigated. Efforts to develop common methodological frameworks and align on key variable definitions are needed to support harmonized data collection and standards.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ A common collaborative research protocol was used across distinct real-world data (RWD) assets to assess the level of standardization capable across datasets and the utility of different real-world end points.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Comparison of results highlights areas of concordance, suggesting that real-world time to next treatment line and real-world time to treatment discontinuation may be useful early clinical end points that may be used in prospective studies, although concerns regarding data missingness and potential biases are acknowledged.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ This study illustrates the power of multistakeholder collaboration to identify both the challenge and the importance of methodological rigor in RWD efforts to support generation of high-quality RWE.

The 21st Century Cures Act was enacted in 2016, to evaluate the health policy goals of trial design modernization and use of real-world evidence (RWE) creating numerous synergistic efforts.[1] These efforts aim to further understand ways in which RWE may complement or supplement randomized controlled trial data for regulatory purposes.[2–6] Evaluating applied examples of real-world data (RWD) and the generation of RWE are central to this effort. We build on the foundational work from the Duke-Margolis Center for Health Policy[2,3,6] and the National Academies of Science, Engineering, and Medicine (NASEM)[5] by highlighting

[1]US Food and Drug Administration, Bethesda, Maryland, USA; [2]OptumLabs, Cambridge, Massachusetts, USA; [3]Americal Society of Clinical Oncology, Alexandria, Virginia, USA; [4]IQVIA, Durham, North Carolina, USA; [5]COTA, Inc, Boston, Massachusetts, USA; [6]Ontada, Irving, Texas, USA; [7]Syapse, San Francisco, California, USA; [8]ConcertAI, Cambrdige, Massachusetts, USA; [9]Flatiron Health, New York, New York, USA; [10]National Cancer Institute, Bethesda, Maryland, USA; [11]Tempus, Chicago, Illinois, USA; [12]Kaiser Permanente, Oakland, California, USA; [13]Friends of Cancer Research, Washington, DC, USA. *Correspondence: Jeff D. Allen (jallen@focr.org)

considerations for study concept, design, and analysis, such as extracting key patient characteristic data and standardizing key variable definitions, with the goal of implementing a shared research protocol across distinct RWD assets. Friends of Cancer Research (*Friends*) collaborated with 10 data partners using oncology RWD from administrative claims, electronic health records (EHRs), prior authorization systems, and/or cancer registries, to conduct Pilot 2.0 evaluating outcomes for patients with advanced non-small cell lung cancer (aNSCLC) receiving systemic frontline therapies.[7]

Friends' Pilot 2.0 builds upon the results from Pilot 1.0, which included 6 data partners that evaluated the performance of real-world end points across multiple data sources.[4] These studies evaluated immunotherapy utilization for the treatment of aNSCLC to evaluate outcomes including overall survival (OS), which have previously indicated treatment benefit.[8,9] Furthermore, this subsequent study aimed to provide specific considerations in the development and design of RWD studies based on shared learnings arising from the observed variability among data sources. Methodological solutions were explored to address variability and enhance the alignment of target study populations for appropriate comparison of study outcomes. We discuss a potential strategy for standardization, including development of a common lexicon to describe and evaluate RWD quality, and share specific lessons learned from our experience implementing a common research protocol across varied RWD sources.

METHODS
Data partners and data sources
RWD partners that participated in this study represent data from a range of settings, including community oncology centers, academic medical centers, health systems, and integrated delivery system networks in the United States. The contributing partners included: ASCO CancerLinQ/ConcertAI,[10–13] Cancer Research Network,[14] COTA,[15] IQVIA,[16] Ontada (formerly McKesson),[17] SEER-Medicare,[18] Syapse,[19] and Tempus.[20] In addition, Flatiron Health, Mayo Clinic, OptumLabs, and Aetion also contributed to the early phase study design. Data curation included approaches that were unique to each participant based on availability, including natural language processing, artificial intelligence tools, technology enabled abstraction, and chart review. Common definitions were established (**Table 1**) and parallel analyses were performed by each group and summary results were submitted to FOCR.

Population
Each cohort selected patients with aNSCLC treated in the first line setting for advanced/metastatic disease with platinum doublet chemotherapy (PDC), PD-(L)1 monotherapy, or PD-(L)1 therapy in combination with platinum doublet chemotherapy (combination), as per the defined eligibility criteria (**Figure 1**). Patients were documented as having been physically present at a practice or as having had an encounter (defined as a physician visit, i.v. administration, or vitals documentation) in the database on at least 2 separate occasions on or after January 1, 2011, until data cutoff date (March 31, 2018). For the claims-based data source, patients were required to be enrolled on or after January 1, 2011, and before the data cutoff date (March 31, 2018). Determination of the end of follow-up (censor date) varied by participating institutions and was based on the most recent date for which complete information was available for the outcome of interest.

Population eligibility was limited to two primary factors that were known to be captured well across all data sources: (i) diagnosis: cancer type (aNSCLC) and (ii) treatment: documented receipt of a qualifying treatment regimen for advanced disease. Evidence of advanced disease was defined as stage IIIB, IIIC, or IV NSCLC at initial diagnosis or early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression to metastatic disease (locally advanced or metastatic disease who had not received prior systemic therapy). The study maintained broad eligibility criteria, reflecting a real-world population. As such and due to varying levels of data availability between RWD sources, clinical characteristics often defined as eligibility criteria for clinical trials, such as organ function (renal and hepatic), PD-(L)1 status, and evidence of brain metastases, were not included.

Although histology was available from all data sources, it was included as a covariate in regression models but not as part of the inclusion/exclusion criteria due to sample size concerns. Adequate organ function is often considered in the use of PD-(L)1 therapy in routine clinical practice; however, laboratory values of organ function were not available from all data sources. Among the four data sources with available lab results, < 1% of patients were identified with severe hepatic or renal dysfunction, suggesting the treatment regimens studied were rarely used in patients with organ impairment and this exclusion, if applied, was expected to have little impact on study findings. Similarly, availability of PD-(L)1 status varied across the data sources and could not be included as a required covariate. Last, brain metastases may not be adequately captured in RWD sources, and lack of affirmative evidence of brain metastases was considered inadequate as a proxy for absence of brain metastases. Consequently, we adjusted for evidence of brain metastases but did not consider presence or absence of brain metastases in patient selection.

Frontline treatment
Treatment groups were defined based on exposure to PDC regimens (cisplatin/carboplatin, oxaliplatin, or nedaplatin with pemetrexed, paclitaxel, nab-paclitaxel, or gemcitabine), PD-(L)1 therapy (atezolizumab, nivolumab, or pembrolizumab), or combination therapy in the frontline setting. Treatment regimen was identified within each data source based on medication orders, medication administration records, medical claims, or infusion databases. Informed by expert clinical input, frontline regimen was defined as the first chemotherapy regimen given subsequent to the date of advanced diagnosis, and included all administered agents initiated within 30 days following the day of first infusion. All therapies were eligible for capture from the date of study initiation; however, it should be noted that approval of PD-(L)1 immunotherapy for aNSCLC did not occur until October 2015.

Study end points
The pilot included assessment of three end points: real-world overall survival (rwOS), real-world time to treatment discontinuation (of frontline regimen; rwTTD), and real-world time to next treatment line (rwTTNT). Overall survival (rwOS) was measured as the length of time from the date of first treatment administration in the frontline therapy regimen (index date) to the date of death or disenrollment (defined as the last known recorded clinical activity in structured data); however, completeness and validation of mortality data sources for rwOS varied across groups. End points that could be uniformly operationalized across RWD sources were chosen specifically for their capacity to convey important information associated with treatment benefit. Because disease progression is not uniformly defined nor captured in RWD sources, yet is clinically associated with regimen discontinuation or initiation of a new regimen or modality across therapeutic classes, rwTTD and rwTTNT were selected as measurable parameters to evaluate as end points instead of real-world progression-free survival, where rwTTD was defined as the length of time from the index date to the date of frontline treatment discontinuation.[10] There are notable limitations to interpretability of the TTD end point because the standard chemotherapy (PDC) regimens in the metastatic setting which are expected to continue for four to six cycles, whereas the use of PD-(L)1 therapy may continue indefinitely requiring

Table 1 Harmonized definitions employed in the pilot project

Term	Harmonized definition	Decision impact
Population		
Advanced NSCLC	All data sources had the ability to identify patients diagnosed with NSCLC. Evidence of advanced disease was defined as either stage IIIB, IIIC, or IV NSCLC or early-stage (stages I, II, and IIIA) NSCLC with a recurrence or progression at initial diagnosis.	Including patients diagnosed early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression to advanced or metastatic status improved sample size for analysis but created a less homogeneous population of both newly diagnosed and previously treated (vs. patients newly diagnosed lung cancer).
Frontline	Patients were required to have no evidence of treatment in 180 days before the date of diagnosis and evidence of an eligible treatment within 120 days after diagnosis	Patients who have delays to treatment initiation would not be included.
Histologic subtype	Histology was not required for inclusion	Histology was not universally collected, although subanalysis feasible. Results reflected overall aNSCLC trends but were less specific to a histology subtype.
Eligibility criteria	The study population was not limited to those meeting eligibility criteria common for inclusion in a clinical trial (e.g., kidney function, performance status)	Data on organ function and performance status at or prior to treatment initiation was not often available or difficult to ascertain in RWD sources, although subanalysis was feasible. The population may be less like the RCT population(s).
Regimens		
Drugs	The following medications were included representing traditional chemotherapy or IO given after the date of diagnosis: cisplatin/carboplatin, oxaliplatin, or nedaplatin with pemetrexed, paclitaxel, nab-paclitaxel, or gemcitabine; atezolizumab, nivolumab, or pembrolizumab. Oral agents were not included.	Regimens are subject to misclassification, particularly in the doublet chemotherapy cohort. Patients starting on a PD-(L)1 should not be ALK or EFGR positive.
Frontline (first line regimen) assignment	Frontline regimen was defined as all administered agents received within 30 days following the day of first infusion.	Misclassification or omission of patients with delays to full treatment initiation in the first 30 days was possible. This would not impact the PD-(L)1 monotherapy cohort, as additional therapy would not be expected.
End points		
rwOS	Length of time from the date of treatment initiation to the date of death or end of follow-up; or end of study	Date of initiation may bias toward slightly shorter event times compared with clinical trials which can use date of randomization or enrollment instead. Missing events, on average, tend to make survival outcomes look better than in trials, especially if missingness is not independent of timing of death events.
rwTTNT	Length of time from the date of treatment initiation to the date of the next systemic treatment. When subsequent treatment is not received (e.g., continuing current treatment or disenrollment not due to confirmed death), patients were censored at their last known activity.	Missingness for subsequent treatment, including receiving treatment outside the system of capture is a limitation. This measure is also affected by the clinical guideline recommendations for administration of treatment cycles which can vary by regimen and has to be evaluated for comparability prior to the study to ensure appropriate interpretation.
rwTTD	Length of time from the date of treatment initiation to the date of patient treatment discontinuation the. The study treatment discontinuation date was defined as the last administration or noncancelled order of a drug contained within the regimen. Discontinuation was defined as having a subsequent systemic therapy after the initial regimen, having a gap of more than 120 days with no systemic therapy following the last administration, or having a date of death while on the initial regimen. Patients without a discontinuation were censored at the end of follow-up.	At the patient level, TTD is associated with PFS across therapeutic classes.[21]
rwTTP	Progression was omitted as claims-based algorithms are inadequate and among the EHRs progression events are not consistently captured in structured data. Unlike in clinical trials, there is not a uniform criterion (e.g., RECIST) in the off-protocol setting for determination of disease progression.	As TTP and PFS are accepted outcomes in clinical trials, comparison of these outcomes to randomized trials of similar regimens were limited by the data available.

Table 1 (Continued)

Term	Harmonized definition	Decision impact
Analysis		
Estimation	Kaplan-Meier estimation was used to describe distribution of end points for each dataset for each regimen, and for estimating key time points (e.g., 6-month, 12-month event rates) with confidence intervals.	
Comparisons	Proportional hazards regression, adjusting for prognostic factors available to all groups.	
Additional analyses	For OS, censor all events at 24 months and re-estimate HRs for treatment effect, adjusted for other prognostic variables.	

ALK, anaplastic lymphoma kinase; aNSCLC, advanced non-small cell lung cancer; EHR, electronic health record; HR, hazard ratio; IO, intra-osseous; NSCLC, non-small cell lung cancer; OS, overall survival; PFS, progression-free survival; RCT, randomized controlled trial; RECIST, Response Evaluation Criteria in Solid Tumors; rwOS, real-world overall survival; rwTTD, real-world time to treatment discontinuation; rwTTNT, real-world time to next treatment line; rwTTP, real-world time to treatment progression.

cautious interpretation of this end point in alignment with the clinical context for treatment. As a measure of regimen-specific treatment patterns, rwTTNT was defined as the time from the index date to the initiation of the subsequent regimen or date of death to assess changes in care. To avoid incorrectly identifying patients discontinuing treatment (e.g., leaving the health plan while still on treatment), the operational definition of rwTTD further required identification of patients whose event times were censored due to death or insufficient follow-up. Sufficient follow-up was defined as 120 days with no systemic therapy following the last administration of treatment in the frontline setting. Because mortality is more likely to be under-reported than over-reported in most widely available sources, rwOS, as presented in

Kaplan-Meier curves in this study, likely overestimates the true rwOS distribution in this patient population (and could appear somewhat longer than corresponding data from clinical trials, which tend to have more complete follow-up requirements); nonetheless, rwOS is useful for evaluation in a real-world setting, especially in comparative evaluation of proxy end points in real-world studies, and was included as an end point in Pilot 2.0.

Data standardization and analysis
Collaborators jointly developed a common research protocol *a priori*, including definitions on patient selection criteria, key covariates, and outcomes, which were collected within a uniform reporting template
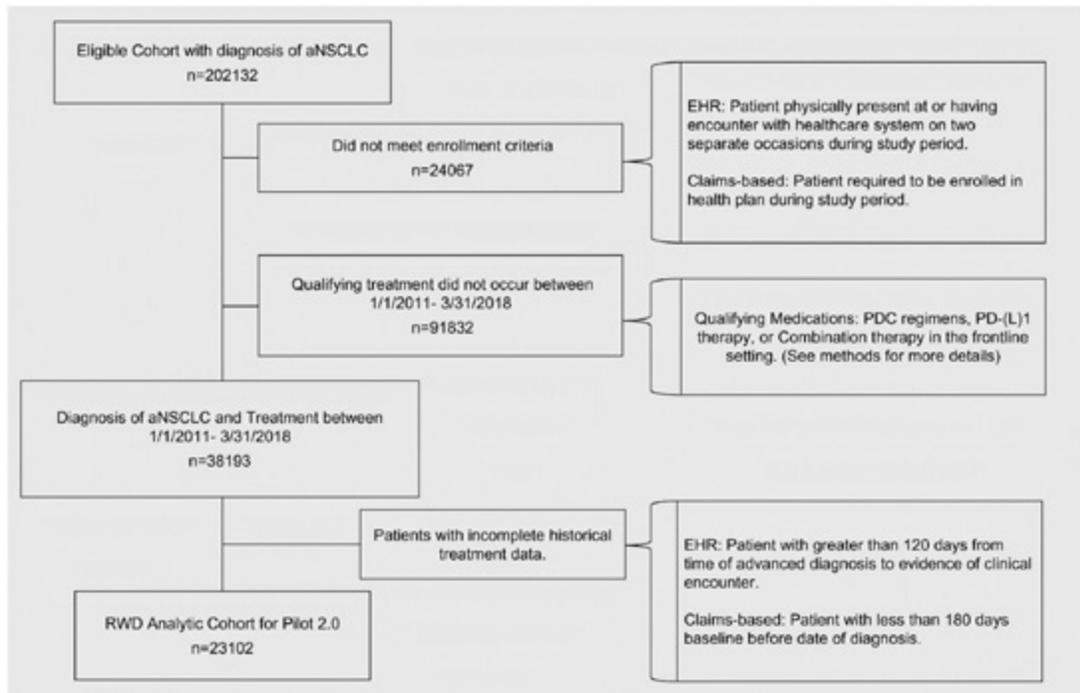
Figure 1 Cohort construction, including data from all data sources. aNSCLC, advanced non-small cell lung cancer; EHR, electronic health record; PDC, platinum doublet chemotherap; PD-(L)1, programmed cell death protein 1/programmed death-ligand 1; RWD, real-world data.

and accompanied by a detailed statistical analysis plan (**Supplementary Protocol**). Each RWD partner operationalized the common research protocol based on characteristics of each data source and technological feasibility, conducting analyses on their respective datasets individually, and reporting summary-level results via the uniform reporting template due to patient privacy, technical complexity, and proprietary nature of the datasets. It was deemed infeasible to aggregate or merge the data for analytic purposes and, instead, collaborators sought to standardize definitions and harmonize processes (**Table S1**). The development of each analytic dataset was subject to data availability and software programming accessible to each RWD partner. Each collaborator has established curation processes designed to evaluate the quality and completeness of their data. Thus, such research-ready databases may not reflect typical EHR data available or health insurance claims data that have not been subjected to such ongoing data curation.

Analytic methods were applied to account for several key sources of variation in the availability of follow-up data; specifically, impact of length of healthcare enrollment, year of treatment initiation, and therapy availability based on US Food and Drug Administration (FDA) approval dates. Kaplan-Meier estimation was used to describe the distribution of each end point (rwOS, rwTTD, and rwTTNT) for each regimen by RWD source. Presentation of the survival curves per regimen was important to assess differential follow-up across data source, and across regimens within sources, and to demonstrate censoring rates which varied across data sources. Proportional hazards regression was used to compare treatment arms for each end point, adjusting for prognostic factors reasonably available to all groups: status at diagnosis (advanced at diagnosis vs. early stage and progressed to aNSCLC), stage, age, year of treatment initiation, gender, race, histology, smoking history, PD-(L)1 expression status, Eastern Cooperative Oncology Group (ECOG) 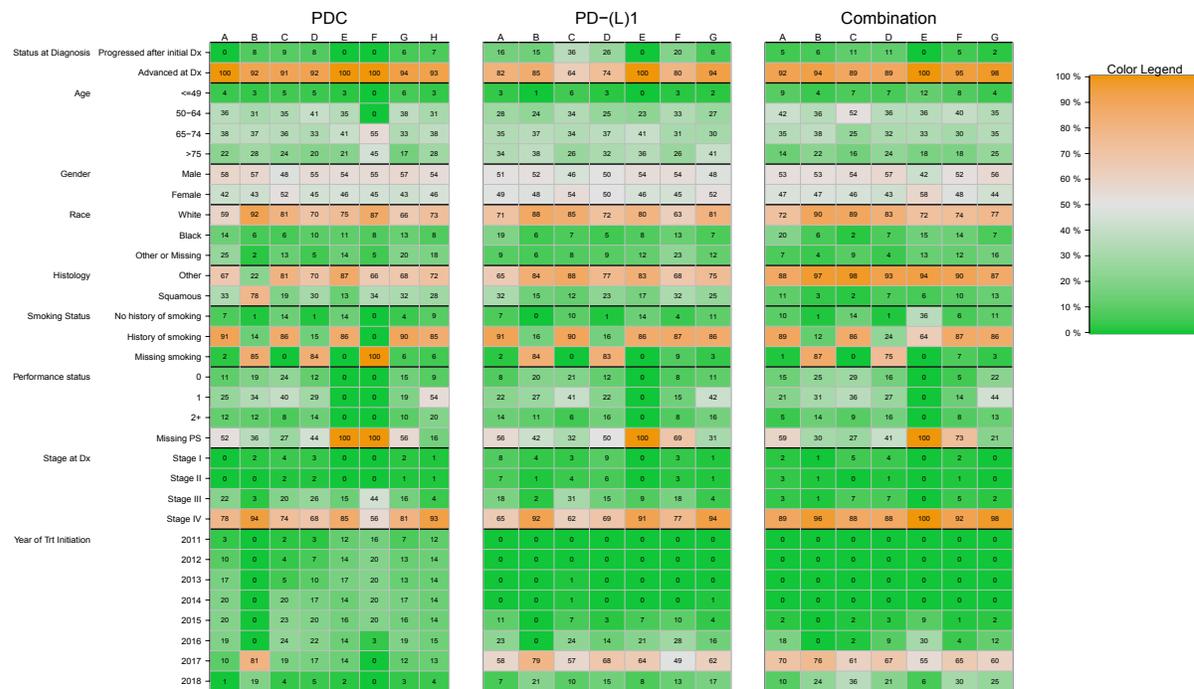performance status, and time to treatment initiation from diagnosis. All prognostic factors were included as nominal categorical covariates, with continuous variables (e.g., age) converted to categorical scales. To address missingness and avoid excluding patients who had missing values for any of the prognostic factors, each covariate included a "missing/unknown" category. Hazard ratios (HRs) from regression models comparing PD-(L)1 to PDC and combination to PDC are presented with 95% confidence intervals (CIs).

The study included patients initiating treatment for aNSCLC in 2011, whereas approval of PD-(L)1 immunotherapy for aNSCLC did not occur until October 2015, skewing the length of available follow-up data toward the PDC cohort. Uptake of PD-(L)1 and combination increased during 2016 and combination during 2017–2018, respectively, limiting follow-up, comparisons, and inferences related to this cohort to a 2-year period. A sensitivity analysis was conducted censoring all patients without an event at 24 months and re-estimating HRs for treatment effect to assess the impact of differential follow-up among the different treatment cohorts (results not shown).

RESULTS

Clinical and demographic characteristics were similarly distributed within treatment groups for each RWD source (**Figure 2**). Geographic coverage varied by data source. There are notable differences in missingness of certain variables across data sources (e.g., smoking and performance status; **Figure 2**). Overall utilization of PD-(L)1 and combination regimens increased over the study period, with the earliest use of PD-(L)1 therapy starting in 2015.

Median rwOS ranged from 10–17 months for PDC across data groups (**Figure 3a**) and was 12–18 months in the PD-(L)1 groups (**Figure 3b**). Kaplan-Meier curves for rwOS, rwTTD, and

		PDC A	B	C	D	E	F	G	H	PD-(L)1 A	B	C	D	E	F	G	Comb. A	B	C	D	E	F	G
Status at Diagnosis	Progressed after initial Dx	0	8	9	8	0	0	6	7	16	15	36	26	0	20	6	5	6	11	11	0	5	2
	Advanced at Dx	100	92	91	92	100	100	94	93	82	85	64	74	100	80	94	92	94	89	89	100	95	98
Age	<=49	4	3	5	5	3	0	6	3	3	1	6	3	0	3	2	9	4	7	7	12	8	4
	50–64	36	31	35	41	35	0	38	31	28	24	34	25	23	23	27	42	36	52	36	36	40	35
	65–74	38	37	36	33	41	55	33	38	35	37	34	37	41	31	30	35	38	25	32	33	30	35
	>75	22	28	24	20	21	45	17	28	34	38	26	32	36	26	41	14	22	16	24	18	18	25
Gender	Male	58	57	48	55	54	55	57	54	51	52	46	50	54	54	48	53	53	54	57	42	52	56
	Female	42	43	52	45	46	45	43	46	49	48	54	50	46	45	52	47	47	46	43	58	48	44
Race	White	59	92	81	70	75	87	66	73	71	88	85	72	80	63	81	72	90	89	83	72	74	77
	Black	14	6	6	10	11	8	13	8	19	6	7	5	8	13	7	20	6	2	7	15	14	7
	Other or Missing	25	2	13	5	14	5	20	18	9	6	8	9	12	23	12	7	4	9	4	13	12	16
Histology	Other	67	22	81	70	87	66	68	72	65	84	88	77	83	68	75	88	97	98	93	94	90	87
	Squamous	33	78	19	30	13	34	32	28	32	15	12	23	17	32	25	11	3	2	7	6	10	13
Smoking Status	No history of smoking	7	1	14	1	14	0	4	9	7	0	10	1	14	4	11	10	1	14	1	36	6	11
	History of smoking	91	14	86	15	86	0	90	85	91	16	90	16	86	87	86	89	12	86	24	64	87	86
	Missing smoking	2	85	0	84	0	100	6	6	2	84	0	83	0	9	3	1	87	0	75	0	7	3
Performance status	0	11	19	24	12	0	0	15	9	8	20	21	12	0	8	11	15	25	29	16	0	5	22
	1	25	34	40	29	0	0	19	54	22	27	41	22	0	15	42	21	31	36	27	0	14	44
	2+	12	12	8	14	0	0	10	20	14	11	6	16	0	8	16	5	14	9	16	0	8	13
	Missing PS	52	36	27	44	100	100	56	16	56	42	32	50	100	69	31	59	30	27	41	100	73	21
Stage at Dx	Stage I	0	2	4	3	0	0	2	1	8	4	3	9	0	3	1	2	1	5	4	0	2	0
	Stage II	0	0	2	2	0	0	1	1	7	1	4	6	0	3	1	3	1	0	1	0	1	0
	Stage III	22	3	20	26	15	44	16	4	18	2	31	15	9	18	4	3	1	7	7	0	5	2
	Stage IV	78	94	74	68	85	56	81	93	65	92	62	69	91	77	94	89	96	88	88	100	92	98
Year of Trt Initiation	2011	3	0	2	3	12	16	7	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2012	10	0	4	7	14	20	13	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2013	17	0	5	10	17	20	13	14	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	2014	20	0	20	17	14	20	17	14	0	0	1	0	0	0	1	0	0	0	0	0	0	0
	2015	20	0	23	24	16	20	16	14	11	0	7	3	7	10	4	2	0	2	3	9	1	2
	2016	19	0	24	22	14	3	19	16	23	0	24	14	21	28	16	18	0	2	9	30	4	12
	2017	10	81	19	17	14	0	12	13	58	79	57	68	64	49	62	70	76	61	67	55	65	60
	2018	1	19	4	5	2	0	3	4	7	21	10	15	8	13	17	10	24	36	21	6	30	25

Color Legend: 100 %, 90 %, 80 %, 70 %, 60 %, 50 %, 40 %, 30 %, 20 %, 10 %, 0 %

Figure 2 Characteristics of treatment groups within participating data sources. Numbers in table represent percent of patients in each category. Coloring ranges from bright green (0%) to bright orange (100%) to highlight areas of differences across data sources for the same treatment.

Figure 3 Kaplan-Meier curves for overall survival (OR) for (**a**) PDC and (**b**) PD-(L)1; hazard ratios with 95% confidence intervals (vertical bars) comparing PD-(L)1 to PDC for (**c**) OS, (**d**) time to treatment discontinuation, and (**e**) time to next treatment (TTNT). PDC, platinum doublet chemotherapy; rwOS, real-world overall survival; rwTTD, real-world time to treatment discontinuation; rwTTNT, real-world time to next treatment.

rwTTNT for all treatment groups are provided in **Figures S1**, **S2** and **S3**. The 1- and 5-year rwOS estimates for PDC were within a similar range across datasets. HRs (with 95% CIs) in **Figure 3c** comparing rwOS for PD-(L)1 vs. PDC, adjusted for a common set of covariates, suggest no evidence of association: HRs range from 0.88 to 1.22 with all 95% CIs overlapping 1. The direction of the association varied among data partners, with 3 having HR estimates greater than 1 (1.06, 1.09, and 1.22), 2 with HR estimates less than 1 (0.88 and 0.88), and 2 with HR estimates at almost exactly 1 (0.99 and 0.99). Although there was consistency in rwOS curves across data sources within each treatment group for PD-(L)1 and for PDC through ~ 6 months, there was more variability in rwOS curves in the 0-to-6-month time period in the combination rwOS curves (**Figure S1**). This may be due to smaller sample sizes in the combination cohorts, leading to more imprecise estimates. For both rwTTD and rwTTNT, HRs were less than 1 (in all but one case; **Figure 3e**, **group C**), demonstrating that patients on PD-(L)1 had longer times on treatment than patients who received frontline PDC in these populations studied. There were differences observed in rwTTD, with HR (range, 0.40 to 0.65; **Figure 3d**); however, these results should be considered in the context of certain therapeutic regimens having a set number of cycles of therapy prior to discontinuation, particularly in the PDC cohort. Associations were also observed for rwTTNT, with smaller effect sizes, with HR (range, 0.51 to 0.80; **Figure 3e**). Similar results were observed when comparing combination to PDC (**Figure S4**). The

evaluation of results and conduct of this RWD study led to a set of methodological best practices when designing a RWD study across a multistakeholder group (**Table 2**). Discussion of the importance of the considerations of confounding by key prognostic factors is discussed in Supplementary Materials (**Figure S5**).

DISCUSSION AND RECOMMENDATIONS

The Friends' RWE Pilot 2.0 was a collaborative effort among participating data partners building upon prior work conducted under Pilot 1.0.[4] The updated collaboration further evaluated the performance of real-world end points across RWD sources in answering a common clinical question, specifically on outcomes among patients with NSCLC who received treatment (PDC, PD-(L)1, or combination) in the first-line advanced or metastatic setting. Through a collaborative common protocol, all RWD partners used the following process steps and engaged in weekly communication for RWD evaluation: (i) common shared protocol and statistical analysis plan (SAP), (ii) standardizing definitions across datasets, (iii) variance in methodological approaches (acceptable variance), (iv) data quality and sensitivity analysis, and (v) transparency by reporting limitations. The study showed broadly similar patterns of outcomes in which the distribution of rwOS was consistent across treatment cohorts among the patient cohorts with similar characteristics (**Figure 2**), but PD-(L)1 containing regimens had longer TTD and TTNT than PDC, demonstrated by estimated HRs in **Figure 3** and **Figure S4**. However, there was

Table 2 Recommendations from the RWE Pilot 2.0 for developing a common RWD protocol to achieve consistency and increase reproducibility using a format that minimizes ambiguity or subjectivity in interpretation of definitions or analysis approaches

Recommendation/ Sub-recommendation	Description
Defining the eligibility criteria	Shared variables that are commonly available across data sources should be used for defining patient inclusion in the study. In the RWE Pilot 2.0, cancer diagnosis (including stage and cancer type) and treatments receipt (platinum doublet and immunotherapy) were the primary criteria. Given that the goal was to make real-world inferences, the eligibilities were based on the population of interest for generalizability. For example, if the goal was to assess treatment differences in patients with advanced age, the age range would be limited to adequately address that question in this study.
Collaborative common RWD protocol	The collaborative protocol should determine a list of core common required data elements, common variables available in a variety of formats that require translation (e.g., age groups; gender and race categories) should be described, definitions (e.g., exposure and end points) should be included, and standardized reporting formats should be agreed on prior to study initiation. Include a standard reporting template complete with table and figure drafts to create understanding around the intended results to be generated.
Define core common key data elements	Establish a core set of data elements with standard definitions enables greater comparability. Variables may have varying levels of availability in RWD, and their relevance for inclusion as a required variable depends on the relation to the study question. Structured data such as age and sex, are minimal common data elements that are typically readily available across independent data sources and requisite for analysis. However, other data elements demand thoughtful consideration and transparency such as (ii) variables available in different formats (e.g., PD-L1 biomarker +/− indicator vs. expression), (ii) variables requiring derivation (e.g., ICD codes vs. laboratory values in the definition of reduced organ function), or (iii) variables requiring extraction from unstructured data (e.g., status of advanced at initial diagnosis vs. progression after initial diagnosis).
Align clinical variables and laboratory values	Key clinical and analytic variables should be identified and aligned as needed, and it should be determined whether strict variable definitions are required for inclusion criteria or if variations are acceptable. Variance in measurement can lead to subsequent impact on outcome calculations. For example, kidney function or genomic testing may be extracted from structured or unstructured data, where a source could have data ranging from the actual lab values to markers of function (e.g., laboratory tests for organ function, CrCl, ICD-9/10 indicating dysfunction) or indicators of testing to specific testing results (e.g., PD-(L) test completed to expression percentage). In areas where variation is accepted, the use of sensitivity analyses to examine variance is useful to guide inappropriate interpretation. Implement a well-developed common protocol for all RWD studies *a priori* to ensure internal and external replication.
Data quality assessment	Development of a template for quantitative evaluation of data distributions, quality, and missingness may provide a quantitative approach to understanding data availability and missingness for improved interpretation. However careful evaluation by a representative team that has deep knowledge of the data curation, extraction, and provenance is necessary. The use of quality indicators for data or consensus on problematic missingness for key covariates may inform the study design.
End point selection	Commonly used end points in clinical trials may not be practical or replicable in RWD. As an example, rwTTP and rwPFS were not included in Pilot 2.0. Challenges with measuring rwTTP and rwPFS exist: claims-based algorithms are limited, relying on proxy measures for progression and consensus definitions among EHRs data sources were prohibitively difficult to establish because of differences in capture and reporting. While uniform criterion (e.g., RECIST) allow protocol directed establishment of progression in clinical trials, progression outcomes are not consistently captured in RWD as there is currently a lacking capability in the off-protocol setting for determination of disease progression. Additional endpoints, rwTTNT and rwTTD, are more readily accessible in RWD. While survival outcomes (rwOS) are easier to define and measure in most RWD sources, sources are often missing mortality information on a large fraction of patients, which affects estimation of rwOS parameters (e.g., median rwOS) and substantially limits interpretation, while incurring additional biases due to missing data. Linking to additional data sources which include more complete mortality data could improve end point ascertainment and should be done if feasible to make estimates based on rwOS more accurate and evaluable to other studies, such as clinical trials.
Defining event times and censoring	When evaluating endpoints, there is a need to it may be most reasonable merge clinical applicability with analytical feasibility. For example, in defining rwTTD, groups had to align on the appropriate time period that would equate with without no treatment receipt to be considered a discontinuation. An additional step in the process would be evaluating the potential to share software code between groups for replicability and additional validation.
Statistical analysis plan	SAP must be written comprehensively with sufficient detail to reduce the risk of deviations in methods used and characterizations of variables in models or tests. In conjunction with the SAP, it is instrumental that the protocol includes table and figure templates to ensure that all groups have the same understanding of the intended results to be generated, and the models required to reduce variance in interpretation. Developing tables within the shared research protocol allowed groups to consider subtle differences in modeling that would not have arisen without having developed them in advance.

Table 2 (Continued)

Recommendation/ Sub-recommendation	Description
Addressing missing data and potential biases	Approaches for quantifying and accounting for missing data in analyses should be considered in the protocol to maintain study integrity while minimizing biases in the interpretation of results.[24] Data missingness should be evaluated by a team that has deep knowledge of the data curation, extraction, and provenance. Imputation should be carefully considered, given the potential for missingness of variables to be related to patient outcomes (i.e., informative missingness) in RWD; choices such as imputation of the data or use of the missing category in modeling have implications for study analyses and inferences. Additionally, use of bias quantification approaches may be useful in appropriate interpretation of the results and understanding study limitations, which in RWD are often limitations of the underlying data.
Assess sample size	Because the number of patients in RWD sources is often based on retrospective data availability, study planning for RWD studies may not consider sample size and the power to detect clinically relevant effects. Even so, it is important to ensure that the sample size is sufficiently large to be able to derive meaningful inferences. If the study is underpowered, modeling may be infeasible or hypothesis tests can tend to find "insignificant" findings with wide confidence intervals, leading to potentially misleading results. In contrast, if the RWD source provides a very large sample, the study may be overpowered and there will be a tendency to over-interpret statistically significant findings. Statistically significant P values do not necessarily imply clinical significance. In that case, interpretation of results could focus effect estimates with their confidence intervals (or similar quantities), and not necessarily alone on P values.
Cautious inference	Even with careful attention to adjustment for population differences, there are inherent selection bias and unmeasured confounding as well as cohort effects that may not be able to be accounted for in a study; these limitations of RWD need to be appropriately addressed in the interpretation of results, inferences, and conclusions of RWE studies. In our study, while there were no obvious differences in the patient characteristics included in Pilot 2.0 across treatment cohorts, the clinical standard of care was likely to differ for the PDC population before and after FDA approval for PD-(L)1 therapies. Similarly, comparisons of results from RWE studies to results from clinical trials need to be cautious given underlying differences in patients treated in clinical trials vs. those in available in RWD sources; these differences are expected due to limited adult clinical trial participation in patients with cancer (3–5%) and strict trial eligibility criteria. This is a strength of RWD in allowing expansion of eligibility criteria to better understand use in a real-world population which is, in turn, a limitation in comparative efforts due to the aforementioned selection bias.
Diverse Multidisciplinary Research Team	Perhaps the most pivotal part of the process in an RWD study is developing a multidisciplinary team, including clinicians, biostatisticians, epidemiologists, and data scientists, to ensure that studies are clinically relevant with appropriate methods utilized to optimally account for potential biases arising from the observational nature of RWD. Teams are encouraged to include patient stakeholders and diverse representation in the conversation, as this is most effectively accomplished as a team science effort.

CrCl, creatinine clearance; EHR, electronic health record; ICD, International Classification of Disease; RECIST, Response Evaluation Criteria in Solid Tumors; RWD, real-world data; RWE, real-world evidence; rwOS, real-world overall survival; rwPFS, real-world progression-free survival; rwTTD, real-world time to treatment discontinuation; rwTTNT, real-world time to next treatment line; rwTTP, real-world time to treatment progression.

notable variability in parameter estimates across data sources, differences in the level of missingness of certain variables of interest to the analysis; and, therefore, differences in subgroup evaluations, as shown in **Figure 3**.

This study demonstrated a successful collaboration aimed at examining research methodologies to provide approaches to measure treatment effects for patients treated in real-world settings and highlights continuing challenges regarding how to best use RWE. For example, different sources of RWD may arrive at similar conclusions regarding relative effectiveness of therapies or heterogeneity may emerge that is not sufficiently able to be overcome or interpreted. Compared with other end points evaluated, rwTTNT showed greater consistency of findings across sources (**Figure S2**). However, all the outcomes reported here were more consistent across data sources for the patients treated with PDC than for those treated with a PD-(L)1 agent or combination, seemingly because of the greater number of chemotherapy recipients (improving precision of estimates for PDC parameters). This suggests that rwTTNT may be less susceptible to the impact of the data variations that exist among RWD sources than other end points, but also highlights the importance of sample size in the stability of parameter estimates. Additionally, whether rwTTNT could be appropriate as an end point in a comparative RWD analysis for a regulatory objective would require additional considerations as this end point does not strictly measure efficacy; further validation would be necessary as it has not been evaluated in a clinical trial setting. However, it could be considered alongside other measures in a pragmatic prospective design. Additional development and validation of these end points is needed, including further exploration around the guidance for clinical use cases for real-world end points as well as ways in which they can be constructed to ensure appropriate interpretability of findings. Although RWD studies may provide an opportunity for increased generalizability and access to expanded populations, study-specific sample size calculations are still necessary to inform study feasibility.

Widely used end points in traditional cancer clinical trials may not be practical or replicable across diverse sources of RWD and pose clear challenges to implementation, particularly when using RWD to construct an external control arm. For example, RWD sources face current challenges with measuring progression-based endpoints (e.g., real-world progression-free survival and

real-world time to treatment progression) including lesser accessibility in claims sources where measurement algorithms are not well-established. Disease progression determination tends to be subjective and may not be uniformly defined across or within and EHR-based data sources and is a future area of investigation for the RWE pilot projects. In contrast, OS is objectively defined, and represents the least variable real-world end point. The use of rwOS, particularly in prospective randomized pragmatic trials, represents an opportunity for use of RWD, particularly where the length and size of such a study would be considered impractical in the traditional clinical trial setting. Despite this potential advantage, it is documented that RWD may be missing mortality information on a large set of patients, which can limit the utility of rwOS in these instances especially in a regulatory context. Understanding the mechanism of missingness (and whether it is random or nonrandom), and potentially incorporating mortality data from external sources are considerations for inferences related to rwOS.[22]

Pilot 2.0 illustrates the importance of considering the potential for selection biases present in RWD, and furthermore in the evaluation of these considerations across diverse data sources. The conditions which cause patient information to be present within a given data source may not be at random, could be associated with the outcome or exposure of interest, and may also be subject to systematic information biases.[23] The likelihood that patient information may be present in a particular source may depend on the practice, treating physician, geography, age, employment, income level, social determinants of health, legal residency status of patients, or other ascertainment practices by the data partner. These factors may be prognostic and therefore associated with the outcome of interest. The evaluation of these types of systematic biases, including selection and information bias, required a multidisciplinary team evaluation to ensure the clinical, statistical, and epidemiological factors are evaluated adequately. Future research to quantify the impact of these biases would improve the ability to interpret the impact of study variance (in between cohorts and among data sources) for patient, clinical, and regulatory decision making. This research also shows the importance of intentionally considering the clinical perspective of how care delivery and treatment may have changed over time (including evaluation of time varying confounding), as well as how the recency of the data and the duration of follow-up may affect the study results regarding treatment patterns described by the data. The heterogeneity present across the Pilot 2.0 data sources also exemplifies the challenges in interpretation of RWD within the context of evidence from traditional clinical trials, especially in any direct comparative or emulative efforts. The real-world sample is less likely to be highly selected with increased comorbidities and increased diversity. Thus, RWD studies may show overall treatment effects that are more modest than those reported in trials; however, they likely could be more representative of the real-world experience of patients. This may provide increased generalizability, especially to the intended treatment population. Establishing population representativeness, including being nationally representative, was beyond the objectives of this study. Acknowledging the benefit of improved representativeness, lack of randomization in retrospective RWD analyses makes causal inference on comparisons challenging and thus differential benefits or harms seen when comparing nonrandomized RWD cohorts should be considered hypothesis generating at present.

Future research on methods for standardization of an approach to categorize data and establish objective measures of data quality that incorporate pertinent RWD assessments is needed. Establishing objective measures of data quality that incorporate assessments of longitudinality, temporality, missingness, and representativeness, perhaps benchmarked against key features of established datasets, would represent an essential advance. This study was limited by inherent factors of retrospective observational research, and by factors unique to the collaborative nature of this effort. First, although data sources were independent, it is unknown to what extent patients are represented in more than one source because of limitations around patient level data access as well as the inability to do any type of matching (comparability). To the extent this occurred, there could be duplication and sources would tend to appear more similar. Second, the data partners conducted their analyses independently, albeit following a carefully developed and detailed analysis plan. Nevertheless, different software packages were used which may have allowed for use of slightly different methods in areas not specifically governed by the analysis plan. Third, there was substantial missingness in certain data types across the data sources. This is likely to have influenced OS estimates for these sources. Additionally, information regarding oral agents was not included and the analysis was not able to account for receipt of certain targeted therapies, which may have been more common in the PDC cohort. Unobserved factors may have influenced receipt of specific treatments. Last, control of potential confounding in this study was limited by the need to implement a uniform analysis across data partners. This approach allowed for comparability of findings across data sources; however, patient characteristics that could have been included as potential confounders in analyses within individual data sources were not included in adjusted analyses in the interest of the broader research goal.

Lessons from the experience of Pilot 2.0 are presented as recommendations for future work in **Table 2**. Friends and the data partners have shown that a diverse group of research enterprises can collaborate effectively to advance the use of oncology RWE. Comparison of results highlights areas of concordance, suggesting that rwTTNT and rwTTD may be useful early clinical end points that may be used in prospective studies, although concerns regarding data missingness and potential biases are acknowledged. In summary, the study illustrates the power of multistakeholder collaboration to identify both the challenge and the importance of methodological rigor in RWD efforts.

AUTHOR CONTRIBUTIONS

D.R.R., H.J.H., E.G.M., J.B.C., A.J.B., S.S.B., J.L.E., C.S., M.A.I., Y.N., N.J.R., M.S.W., A.B.C., M.B., L.E., L.K., P.S.M.K., R.P.B., L.C.S., E.S., O.T., E.V., L.L., and J.D.A. wrote the manuscript. D.R.R., H.J.H., E.G.M., J.B.C., A.J.B., J.L.E., M.A.I., Y.N., N.J.R., A.B.C., M.B., L.E., L.K., P.S.M.K., R.P.B., E.S., O.T., L.L., and J.D.A. designed the research. H.J.H., C.S., E.H., R.H., and J.W. performed the research. D.R.R., H.J.H., E.G.M., A.J.B., J.L.E., M.A.I., Y.N., N.J.R., A.B.C., L.E., R.H., O.T., J.W., and L.L. analyzed the data.

DISCLAIMER

The views represented in this paper represent the individual authors and should not be interpreted as representing any official HHS views or policy.

1. US Food and Drug Administration. *Framework for FDA's Real-World Evidence Program* (US Department of Health & Human Services, Washington, DC, 2018).
2. Daniel, G. *et al. Characterizing RWD Quality and Relevancy for Regulatory Purposes* (Duke-Margolis Center for Health Policy, Washington, DC, 2018)
3. Mahendraratnam, N. *et al. Determining Real-World Data's Fitness for Use and the Role of Reliability* (Duke-Margolis Center for Health Policy, Washington, DC, 2019).
4. Stewart, M. *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 (2019).
5. National Academies of Sciences, Engineering, and Medicine 2019. *Examining the Impact of Real-World Evidence on Medical Product Development: Proceedings of a Workshop Series* (The National Academies Press, Washington, DC, 2019).
6. Mercon, K. *et al. A Roadmap for Developing Study Endpoints in Real-World Settings* (Duke-Margolis Center for Health Policy, Washington, DC, 2020).
7. Bin, M. & Ong, H. Real-world evidence at glance: How a collaboration of "frenemies" produced common definitions for real-world endpoints. *Cancer Lett.* **45**, 5–12 (2019).
8. Howlader, N. *et al*. The effect of advances in lung-cancer treatment on population mortality. *N. Engl. J. Med.* **383**, 640–649 (2020).
9. Marur, S. *et al*. FDA analyses of survival in older adults with metastatic non-small cell lung cancer in controlled trials of PD-1/PD-L1 blocking antibodies. *Semin. Oncol.* **45**, 220–225 (2018).
10. ASCO CancerLinQ. LLC: Lung Data (2011–3/31/2018) <www.cancerlinq.org/solutions/researchers>.
11. Schilsky, R.L. *et al*. Building a rapid learning health care system for oncology: the regulatory framework of Cancer-LinQ. *J. Clin. Oncol.* **32**, 2373–2379 (2014).
12. Shah, A. *et al*. Building a rapid learning health care system for oncology: why CancerLinQ collects identifiable health information to achieve its vision. *J. Clin. Oncol.* **34**, 756–763 (2016).
13. ConcertAI Patient360 <https://www.concertai.com/data-products/> (2021).
14. National Cancer Institute. Cancer Research Network <https://healthcaredelivery.cancer.gov/crn/>.
15. COTA, Inc. COTA Overview <https://cotahealthcare.com/>.
16. IQVIA. Harness the power of Real World Data <https://www.IQVIA.com/solutions/real-world-evidence/real-world-data-and-insights>.
17. Ontada. The US Oncology Network <https://www.usoncology.com> (2020).
18. Warren, J.L. *et al*. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med. Care* **200240**(Suppl. 8):IV-3–IV-18.
19. Lerman, M.H. *et al*. Validation of a mortality composite score in the real-world setting: overcoming source-specific disparities and biases. *JCO Clin. Cancer Inform.* **5**, 401–413 (2021).
20. Fernandes, L.E. *et al*. Real-world evidence of diagnostic testing and treatment patterns in US patients with breast cancer with implications for treatment biomarkers from RNA sequencing data. *Clin. Breast Cancer* 1526–8209 (2020).
21. Blumenthal, G.M. *et al*. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. *Ann. Oncol.* **30**, 830–838 (2019).
22. Curtis, M.D. *et al*. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv. Res.* **53**, 4460–4476 (2018).
23. Tripepi, G., Jager, K.J., Dekker, F.W. & Zoccali, C. Selection bias and information bias in clinical research. *Nephron Clin. Pract.* **115**, c94–c99 (2010).
24. Lash, T.L. *et al*. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* **43**, 1969–1985 (2014).

Real-world Overall Survival Using Oncology Electronic Health Record Data: Friends of Cancer Research Pilot

Laura Lasiter[1], Olga Tymejczyk[2], Elizabeth Garrett-Mayer[3], Shrujal Baxi[2], Andrew J. Belli[4], Marley Boyd[5], Jennifer B. Christian[6], Aaron B. Cohen[2,7], Janet L. Espirito[5], Eric Hansen[4], Connor Sweetnam[8], Nicholas J. Robert[5], Mackenzie Small[6], Mark D. Stewart[1], Monika A. Izano[8], Joseph Wagner[6], Yanina Natanzon[9], Donna R. Rivera[10] and Jeff Allen[1],*

In prior work, Friends of Cancer Research convened multiple data partners to establish standardized definitions for oncology real-world end points derived from electronic health records (EHRs) and claims data. Here, we assessed the performance of real-world overall survival (rwOS) from data sets sourced from EHRs by evaluating the ability of the end point to reflect expected differences from a previous randomized controlled trial across five data sources, after applying inclusion/exclusion criteria. The KEYNOTE-189 clinical trial protocol of platinum doublet chemotherapy (chemotherapy) vs. programmed cell death protein 1 (PD-1) in combination with platinum doublet chemotherapy (PD-1 combination) in first-line nonsquamous metastatic non-small cell lung cancer guided retrospective cohort selection. The Kaplan-Meier product limit estimator was used to calculate 12-month rwOS with 95% confidence intervals (CIs) in each data source. Cox proportional hazards models estimated hazard ratios (HRs) and associated 95% CIs, controlled for prognostic factors. Once the inclusion/exclusion criteria were applied, the five resulting data sets included 155 to 1,501 patients in the chemotherapy cohort and 36 to 405 patients in the PD-1 combination cohort. Twelve-month rwOS ranged from 45% to 58% in the chemotherapy cohort and 44% to 68% in the PD-1 combination cohort. The adjusted HR for death ranged from 0.80 (95% CI: 0.69, 0.93) to 1.15 (95% CI: 0.71, 1.85), controlling for age, gender, performance status, and smoking status. This study yielded insights regarding data capture, including ability of real-world data to precisely identify patient populations and the impact of criteria on end points. Sensitivity analyses could elucidate data set–specific factors that drive results.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Real-world data (RWD) have the potential to complement clinical trial data and fill gaps in knowledge about the performance of approved treatments used in routine care settings, including patient populations excluded from clinical trials, or where limited clinical trial data exist. There is interest in using real-world evidence to support regulatory decisions in rare cancer patient populations, new indications, alternative doses, and schedules.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study sought to evaluate the performance of rwOS and the considerations necessary to assess directionality of treatment associations in a real-world population across five US oncology electronic health record RWD providers with different sources of patient data by aligning the patient population with key inclusion/exclusion criteria from the KEYNOTE-189 study.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Insights were yielded regarding data capture, including ability of RWD to precisely identify patient populations and the impact of criteria on end points.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Sensitivity analyses could elucidate data set–specific factors that drive results.

Randomized controlled trials (RCTs) evaluate the safety and efficacy of medical products in specific patient populations under rigorously controlled conditions. While adherence to structured protocols, use of restrictive eligibility criteria, and patient randomization maximize the internal validity of RCT results, eligibility and protocol-directed care may reduce the relevancy of study results for broader patient populations receiving approved drugs subsequently in routine clinical practice. Limitations regarding the generalizability and transportability of trial findings create a value basis for the evaluation of patient outcomes in real-world settings.[1] Real-world data (RWD) have the potential to complement clinical trial data and fill gaps in knowledge about the performance of approved treatments used in routine care settings, including patient populations excluded from clinical trials, or where limited clinical trial data exist. As such, there is interest in using real-world evidence (RWE) to support regulatory decisions in rare cancer patient populations, new indications, alternative doses, and schedules.[2] Real-world end points measurable across data sources, including administrative claims and electronic health records (EHRs), that are consistently implemented across studies are needed to optimize data collection and accurate interpretation of real-world study findings.[3]

Replication of study findings across multiple data sources using common harmonized data elements and analytical framework is essential to evaluate the potential applications of RWE. In prior work, we established standardized definitions for oncology real-world end points derived from both EHR and claims data.[4] In this study, we assessed the performance of real-world overall survival (rwOS) by evaluating the ability of the end point to reflect expected differences from a previous RCT across multiple data sources, after applying inclusion/exclusion criteria.[5] Although not designed to replicate the clinical trial, this study used KEYNOTE-189, an RCT published in 2018, as a relative benchmark to explore the performance of rwOS across data sources. KEYNOTE-189 demonstrated improved outcomes of pembrolizumab in combination with platinum therapy (cisplatin or carboplatin) and pemetrexed, compared with chemotherapy alone for frontline treatment of patients with nonsquamous metastatic non-small cell lung cancer (metastatic non-small cell lung cancer). By estimating treatment effects in each data set, we sought to distinguish the effects of underlying patient characteristics (e.g., age, biomarker status, and health performance status) from data set–specific considerations (e.g., completeness of key variables). Furthermore, this work generated insights into methodological transparency, data quality, and reporting standards for real-world outcome measures that can inform the interpretation of the results of real-world studies in oncology.

METHODS

Data sources

Five organizations supplying EHR data aligned on a common set of definitions and protocols (**Table 1**, **Table 2**, **and Table S1**). Each data partner conducted data extraction and statistical analyses using deidentified patient data from their respective real-world population and reported aggregated data only. EHR data was sourced through structured (programmatic database extractions) and/or unstructured (chart review) methods conducted in accordance with abstraction rules and quality processes established within each organization.

Study population

Similar to KEYNOTE-189, the real-world population selected for this study included patients with metastatic non-small cell lung cancer (mNSCLC) who initiated frontline treatment in the metastatic setting with combination platinum therapy (chemotherapy) or pembrolizumab plus combination platinum therapy (programmed cell death protein 1 [PD-1] combination), where combination is defined as cisplatin or carboplatin plus pemetrexed (**Figure S1**). Eligible patients had a documented encounter (defined as a physician visit, drug administration, or vitals documentation) in each database on two or more separate occasions on or after January 1, 2011 through March 31, 2018. Frontline treatment was defined as the first regimen subsequent to the date of metastatic diagnosis and included all agents received within 30 days following the day of first administration or noncanceled order after metastatic diagnosis. The index date was the date frontline therapy was initiated. Patients with an index date on or after January 1, 2015 but no later than March 31, 2018 were included in order to improve temporal proximity of treatment groups. Patients with a gap of greater than 120 days from the date of metastatic diagnosis to the first clinical encounter (structured and/or unstructured) were excluded as having possibly incomplete early therapy data. Data cutoff was March 31, 2019, to allow a minimum of 12 months' potential follow-up.

Initial trial-related inclusion and exclusion criteria ("baseline cohort")

The real-world cohorts were identified following key inclusion and exclusion criteria reported by KEYNOTE-189 (**Table S2**). Patients were included in the real-world cohorts if they had evidence of pathologically confirmed mNSCLC (patients metastatic at diagnosis or patients with earlier-stage disease who progressed to metastatic disease), received no previous systemic antineoplastic therapy at any point in the metastatic setting or in the advanced setting within 12 months prior to metastatic diagnosis date, and received at least one dose of either a PD-1 combination or chemotherapy frontline regimen. Patients were included regardless of programmed death-ligand 1 (PD-L1) testing or status. The cohort meeting these criteria is referred to as the "baseline cohort" and included many patients who would likely not have qualified for a typical clinical trial in mNSCLC.

Additional trial-related exclusion criteria ("fully restricted cohort")

To explore potential associations between survival estimates and prognostic factors typically excluded from clinical trials, the following exclusion criteria were applied to the baseline cohort: squamous cell carcinoma or non-small cell lung cancer (NSCLC) not otherwise specified, evidence of inadequate kidney or liver organ function, Eastern Cooperative Oncology Group (ECOG) performance status (PS) ≥2, and evidence of EGFR/ALK (epidermal growth factor receptor / anaplastic lymphoma kinase) sensitizing mutations, all at index date. Patients without organ function, ECOG, or EGFR/ALK data were included. Exact definitions of the criteria, as implemented by each group, are reported in **Table 2**. The cohort meeting all these additional exclusion criteria is referred to as the "fully restricted cohort" and is intended to represent the strict eligibility requirements for patient populations typically enrolled in clinical trials.

End-point definitions

Real-world OS was defined as the length of time from the index date to the date of death. If there was no evidence of death, patients were censored at the last recorded clinical activity prior to data cutoff. The implementation of rwOS by each group is reported in **Table 1**.

Table S1 describes implementation of key descriptive variables and model covariates: disease stage (0, I, II, III, IV, unknown), smoking status (known history of smoking, no known history of smoking), and PD-L1

Table 1 Variations across data sources and implementation of end-point definition

Data set	Description of data source	Population	Derivation of date of death	Censor date
A	Structured and unstructured EHR data and commercial obituary data.	Academic and community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. If dates agree across the three data sources, the date is selected. If discrepancy <7 days exists, EHR data is preferentially captured. If discrepancy >7 days exists, EHR data with accompanying source documentation (e.g., death certificate) is prioritized, otherwise commercial obituary data is captured.	Date of last clinical activity prior to data cutoff, defined as in-person visit event with healthcare provider such as treatment administration or collected test.
B	Structured and unstructured EHR data, commercial obituary data, and Social Security Death Index (SSDI) data.	Academic and community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. If all dates agree across the three data sources, the date is selected. If any two dates agree, that date is selected. If all three dates disagree, the following hierarchy is applied: SSDI, obituary, EHR data. If a day level DoD is available in abstracted EHR data, that date is selected over the consensus structured date. Exact date was used, where available. If only month-level date was available, it was generalized to the end of month. If only year-level date was available, if was generalized to the end of year.	Date of last structured activity, defined as the most recent visit prior to data cutoff.
C	Structured and unstructured EHR data; hospital-based, enterprise-wide, and national cancer registries; commercial obituary data; and digitized obituaries.	Community practice patients in the United States; inpatient and outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. Tumor registry (hospital, enterprise-wide, national) dates were preferentially selected, followed by structured EHR, followed by commercial obituary data.	Date of last contact (physical encounter, medication order, or medication administration, from structured or unstructured EHR data) prior to the data cutoff.
D	Structured EHR data.	Community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	Actual date of death documented from EHR or DMF.	Date of last structured clinical activity prior to data cutoff.
E	Structured EHR data, structured claims data.	Academic and community practice patients in the United States; primarily outpatient; initiated frontline therapy between 2017 and 2018.	Mortality algorithm incorporating EHR and Claims.	Date of last clinical encounter with healthcare provider prior to data cutoff.

DMF, death master file; DoD, date of death; EHR, electronic health record.

expression status (<1%, 1–49%, ≥50%, unknown) and evidence of brain metastases (yes, no) at index date.

Statistical analysis
The Kaplan-Meier product limit estimator was used to calculate 12-month and median rwOS with 95% confidence intervals (CIs) for the baseline and fully restricted cohorts in each data source. Cox proportional hazards models were used to estimate hazard ratios (HRs) and associated 95% CIs for the associations between treatment groups and rwOS in the fully restricted cohort. Unadjusted HRs for the association between treatment groups and rwOS were calculated in the baseline and the fully restricted cohorts. Additionally, unadjusted models for the fully restricted cohort were stratified by age, gender, PD-L1 expression status (<1%, ≥1%, 1–49%, ≥50%), evidence of brain metastases (yes, no), and platinum drug agent (carboplatin, cisplatin). Lastly, we adjusted for age (<65, ≥65 years), gender (male, female), smoking status (known history of smoking, no known history of smoking), and ECOG PS (0, 1, unknown) to account for confounding by prognostic factors not homogenized in

inclusion/exclusion criteria. Forest plots were used to visualize the range of estimates across data sets.

Post hoc analyses
To explore potential associations between estimates of survival and individual prognostic factors, and evaluate the impact of methodological approaches to deriving them from RWD, exclusion criteria were applied sequentially to the baseline cohort in a stepwise manner and unadjusted HRs for the associations between treatment groups and rwOS were calculated after each restriction step:

Step 0: Baseline cohort (as defined above)
Exclude patients with:
Step 1: Squamous cell carcinoma or NSCLC not otherwise specified
Step 2: Inadequate kidney or liver organ function at index date
Step 3: Eastern Cooperative Oncology Group (ECOG) performance status at index date ≥2
Step 4: Evidence of EGFR/ALK sensitizing mutations at index date

Table 2 Definitions of inclusion and exclusion criteria for alignment with Keynote-189

Data set	Advanced diagnosis	Metastatic status	Histology	Organ function	ECOG PS	EGFR/ALK sensitizing mutations
Overall	Diagnosis with advanced disease defined as American Joint Committee on Cancer stage: Stage IIIB, IIIC or IV NSCLC at initial diagnosis or Early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression to metastatic status.	Each group determined metastatic status according to internally consistent protocols.	Patients were placed into one of the following categories: Non-squamous cell carcinoma Squamous cell carcinoma NSCLC histology not otherwise specified (NOS)	Patient is excluded if there is evidence of inadequate kidney/liver function based on laboratory values in the 90 days prior to and including the index date based upon structured lab data. Inadequate renal function is defined as creatinine clearance of <50 ml/min (creatinine clearance calculated using the Cockroft-Gault equation). Inadequate liver function is calculated as a serum total bilirubin of $\geq 1.5\times$ ULN (unless direct bilirubin was measured on the same date and <ULN) and AST or ALT $\geq 5\times$ ULN NOTE: If there are multiple lab values of interest in the time window, the value closest to index date should be used. If there are multiple values for a given lab on the same day, lower creatinine clearance values/ higher bilirubin, AST, and ALT values should be used (i.e. prioritize values that exclude patients)	Patient's ECOG PS at the time of the index date: 0 1 2+ Unknown NOTE: ECOG PS may have been recorded up to 30 days prior to the index date, *OR* up to 7 days after the index date, whichever is closest to the index date. If there are multiple ECOG PS values at the same absolute distance from the index date, priority is given to the ECOG PS value that precedes the index date. For patients with multiple ECOG PS values recorded on the same day, the highest value will be selected.	Approach 1 to the ALK/EGFR exclusion criterion: Test result showing ALK/EGFR rearrangement/mutation at any point before or up to 30 days after the index date. Approach 2 to the ALK/EGFR exclusion criterion: (alone or in combination with approach 1 above) Patients who are prescribed EGFR or ALK targeting therapies in the first 6 months of the index date will be excluded Erlotinib, Afatinib, Osimertinib, Gefitinib, Crizotinib, Ceritinib, Alectinib
A	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Abstracted EHR data Definition: Defined as at least one of the following: stage IV, TNM M1, and/or metastatic progression	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Abstracted EHR data Definition: Used a combination of approach 1 and 2. No deviation from definitions.
B	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Abstracted EHR data. Definition: Stage IV only.	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Abstracted EHR data. Definition: Approach 1 without deviation. Test date was identified as the most recent date across the "specimen collected" date, "specimen received" date, and "result date" variables. If multiple tests were recorded, the one closest to the index date was used. If multiple tests were recorded on the same date, the result showing rearrangement / mutation was used.

Table 2 (Continued)

Data set	Advanced diagnosis	Metastatic status	Histology	Organ function	ECOG PS	EGFR/ALK sensitizing mutations
C	Data source: Unstructured EHR data and structured pathology reports. Definition: Recurrence or progression to metastatic status is based on a medical oncologist stated metastasis or distant recurrence collected from unstructured EHR.	Data source: Unstructured EHR. Definition: Metastatic status is based on a medical oncologist statement.	Data source: Unstructured EHR and structured pathology reports. Definition: Main sources of histology are not codified in the source data but are applied post data collection. Agreed upon vocabulary based on WHO lung cancer histology hierarchy management is utilized to derive NSCLC grouping.	Data source: Structured EHR. Definition: Organ function was not evaluated if value, unit, or range were missing.	Data source: Unstructured and structured EHR. Definition: If Karnofsky performance status (KPS) was available instead of ECOG PS, KPS was converted to ECOG PS.	Data source: Unstructured and structured EHR; commercial laboratory electronic reports. Definition: Used a combination of approach 1 and 2. Definition of ALK rearrangement is based on laboratory report or physician statement. EGFR mutations are defined as any reported mutation on exon 18 through exon 21 regardless of classification. Approach 2 did not deviate.
D	Data source: Structured EHR data. Definition: No deviation.	Data source: EHR data. Definition: Metastatic status identified by having at least one of the following: Stage IV or M1 disease; documented organ site of metastasis; or disease status as metastatic disease	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.
E	Data source: Structured EHR. Definition: To identify staging with ICD9/10 used to classify patients with early stage who progressed.	Data source: Structured EHR. Definition: Diagnosis, staging and M values.	Data source: Structured EHR histology codes and descriptions. Definition: For a subset of patients, NSCLC was indicated, but not Squamous vs. Non-Squamous. These patients were kept since they could not be definitively classified as either.	Data source: Structured EHR. Definition: No deviation.	Data source: Structured EHR. Definition: No deviation.	Data source: Structured and unstructured EHR. Definition: No deviation.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; ECOG PS, Eastern Cooperative Oncology Group performance status; EGFR/ALK, epidermal growth factor receptor/anaplastic lymphoma kinase; EHR, electronic health record; ICD 9/10, International Classification of Diseases, Ninth and Tenth Revisions; NSCLC, non-small cell lung cancer; TNM M1, tumor, node, and metastasis metastasis 1; ULN, upper limit of normal; WHO, World Health Organization.

Table 3 Characteristics of fully restricted cohorts

Characteristic	KEYNOTE-189	A	B	C	D	E
		Chemotherapy/PD-1 combination treated patient characteristics				
Total number of patients/ Treatment, no.	410/206	346/54	1,501/405	232/36	748/132	155/125
Age						
Median, yrs (IQR)	65, 34, 84/64, 34, 84	68, 60, 74/65, 60, 72	67, 59, 74/65, 59, 72	66, 59, 73/64, 58, 71	67, 60, 74/64, 58, 72	68, 59, 74/65, 60, 73
<65 yr, %	48.0/55.8%	38.7/50.0%	42.4/47.2%	46.1/50.5%	41.7/52.3%	37.4/43.2%
Gender, Male, %	62.0/52.9%	45.7/55.6%	45.7/55.6%	52.6/63.9%	46.9/58.3%	47.7/52.8%
ECOG, %						
0	45.4/38.8%	25.4/24.1%	21.0/31.4%	14.7/22.2%	14.7/30.3%	19.4/24.8%
1	53.9/60.7%	46.2/44.4%	33.7/37.5%	29.7/38.9%	63.2/43.2%	35.5/31.2%
Unknown	0.5/0.5%	28.3/31.5%	45.3/31.1%	55.6/38.9%	22.1/26.5%	45.2/44.0%
Smoking status, %						
Evidence of smoking	88.3/87.9%	88.2/87.0%	88.3/89.6%	13.8/36.1%	84.9/84.8%	25.2/25.6%
No evidence	11.7/12.1%	11.8/13.0%	11.7/10.4%	86.2/63.9%	15.1/15.2%	74.8/74.4%
Histology, %						
Non-squamous cell carcinoma	96.1/96.1%	100/100%	100/100%	100/100%	100/100%	100/100%
NOS	2.4/1.9%	Restricted	Restricted	Restricted	Restricted	Restricted
Brain metastases, %						
Evidence of	17.8/17.0%	31.8/29.6%	18.5/14.1%	13.8/19.5%	13.2/9.1%	16.1/20.0%
No evidence	82.2/83.0%	68.2/70.4%	81.5/85.9%	86.2/80.6%	86.8/90.9%	83.9/80.0%
PD-L1 expression status, %						
<1%	31.0/30.6%	11.9/12.8%	23.7/16.3%	50.9/20.8%	50.5/32.3%	NA[a]
>1%	63.4/62.1%	19.3/42.6%	51.9/65.4%	49.1/79.2%	43.5/64.6%	NA
1–49%	31.2/28.2%	14.8/34.0%	39.0/34.6%	26.4/41.7%	39.6/38.5%	NA
>50%	32.2/34.0%	5.9/6.4%	12.9/30.9%	22.6/37.5%	3.8/26.2%	NA
Unknown	5.6/7.3%	67.4/46.8%	24.4/18.3%		6.0/3.1%	NA
Renal function, %						
No Evidence of Inadequate Function		70.5%/64.8%	89.5%/93.1%	100/100%	81.0/77.3%	13.5/16.8%
Unknown%		29.5%/35.2%	10.5/6.9%	0.0/0.0%	19.0/22.7%	86.5/83.2%
Hepatic function, %						
No evidence of inadequate function		70.5%/59.3%	83.3/87.4%	100/100%	79.9/76.5%	49.7/71.2%
Unknown		29.5%/40.7%	16.7/12.6%	0.0/0.0%	20.1/23.5%	50.3/28.8%
Median time from advanced diagnosis to frontline therapy initiation, months (IQR)		1.00, 0.57, 1.63/0.97, 0.62, 1.58	1.2,0.8,1.7/ 1.1,0.7,1.5	1.25, 0.90, 1.80/1.17, 0.70, 1.95	0.83, 0.37, 1.47/0.72, 0.37, 1.38	1.03, 0.53, 1.77/0.93, 0.50, 1.77
Median structured follow-up time from frontline therapy initiation, months (IQR)		10.98, 5.14, 21.28/11.85, 5.60, 14.84	7.2,2.8,17.6/ 10.1,3.5,15.6	12.27, 4.93, 25.04/10.37, 5.36, 17.12	10.08, 3.67, 19.73/12.98, 3.55, 16.10	13.53, 5.93, 14.06/14.06, 6.10, 21.13
Status at initial diagnosis, %						
Advanced at diagnosis		87.6/88.9%	100/100%	88.8/83.3%	95.6/97.0%	96.1/92.0%
Progressed after initial diagnosis		12.4/11.1%		11.2/16.7%	4.4/3.0%	3.9/8.0%
Stage, %						
0		0.0/0.0%		0.0/0.0%	0.0/0.0%	1.3/0.8%

Table 3 (Continued)

Characteristic	KEYNOTE-189	A	B	C	D	E
		Chemotherapy/PD-1 combination treated patient characteristics				
I		6.6/5.6%		6.9/5.6%	0.9/0.0%	0.6/0.0%
II		2.6/0.0%		0.9/2.8%	0.0/0.0%	0.0/0.0%
III		3.2/5.6%		3.4/8.3%	2.9/0.0%	0.0/0.8%
IV		87.6/88.9%	100/100%	87.9/83.3%	95.6/97.0%	95.5/89.6%
Unknown		0.0/0.0%	0.0/0.0%	0.9/0.0%	0.0/0.0%	2.6/8.8%
Index year, %						
2015		35.5/1.9%	36.2/0.0%	34.5/0.0%	29.7/0.0%	0.0/0.0%
2016		36.1/0.0%	36.8/0.2%	35.3/0.0%	33.7/0.0%	0.0/0.0%
2017		23.7/63.0%	22.7/71.6%	23.7/75.0%	30.5/79.5%	83.9/69.6%
2018		4.6/35.2%	4.2/28.1%	6.5/25.0%	6.1/17.4%	16.1/30.4%

ECOG, Eastern Cooperative Oncology Group; NOS, not otherwise specified; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1.
[a]Data were not provided for this analysis.

To evaluate the potential impact of crossover on results, the proportion of crossover in the chemotherapy group, defined as initiation of an immunotherapy (pembrolizumab, nivolumab, or atezolizumab)–containing second line treatment prior to the date of data cutoff was calculated.

RESULTS

Characteristics of patients in each group's fully restricted and baseline cohorts, as well as of patients included in KEYNOTE-189, are described in **Table 3 and Table S3**, respectively. The fully restricted cohorts included 155 to 1,501 chemotherapy and 36 to 542 PD-1 combination patients. Median age in chemotherapy and PD-1 combination groups ranged between 66–68 and 64–65, respectively. The proportion of patients <65 years of age was higher in the PD-1 combination group than in the chemotherapy group by a margin ranging from 4 to 11 percentage points across data sets. The proportion of males in the PD-1 combination groups was 53% to 64% while the chemotherapy group had a slightly lower proportion (46% to 53%). A substantial proportion of patients had unknown ECOG (22–56% across cohorts and data sets), unknown liver function (0–50%), or unknown kidney function (0–87%). History of smoking was identified in >84% of patients in three data sets, and in <40% of patients in the remaining two data sets, due to data missingness. PD-L1 expression status was provided in four data sets, with a proportion of unknown values ranging from 0 to 67% across treatment groups and data sets. A higher proportion of chemotherapy patients had evidence of a PD-L1 expression status <1% compared with PD-1 combination patients, (where percent unknown was low). Patients whose cancer progressed to metastatic after initial diagnosis comprised up to 17% of chemotherapy or PD-1 combination cohorts in four data sets, but were excluded from one data set (i.e., only stage IV patients were included). PD-1 combination patients initiated treatment in 2017–2018, while patients in the chemotherapy group initiated treatment from 2015 to 2018. The proportion of patients initiating either treatment in 2018 was low due to the cutoff date for frontline treatment initiation by March 31, 2018.

Table 4 includes estimates of rwOS at 12 months, as well as unadjusted associations between frontline therapy and rwOS, overall and stratified by key variables of interest, in the fully restricted cohorts. Twelve-month rwOS ranged from 45% to 57% in the chemotherapy group and 44% to 69% in the PD-1 combination group. Unadjusted HR for death, comparing PD-1 combination to chemotherapy, ranged from 0.79 (95% CI: 0.68, 0.92) to 1.10 (95% CI: 0.70, 1.72). In four of the five fully restricted cohorts, HR point estimates were below 1, and four confidence intervals overlapped 1.

Table 5 lists adjusted associations between frontline therapy and rwOS in the fully restricted cohorts, controlling for age (<65 vs. ≥65 years), gender, ECOG (0, 1, or unknown), and smoking status (history, no history, unknown). The adjusted HR ranged from 0.80 (95% CI: 0.69, 0.93) to 1.15 (96% CI: 0.71, 1.85). HR point estimates were below 1 in four of the five fully restricted cohorts and four confidence intervals overlapped 1.

Post hoc analyses explored the unadjusted associations between frontline therapy and rwOS during sequential exclusion step (baseline cohort to fully restricted cohort). Patient numbers ranged from 293 to 2,673 in baseline and 164 to 1,906 in fully restricted cohorts (**Figure S1**). Numerical differences in HRs in analyses with varying exclusion criteria were very small. The finding of statistically significant association or lack thereof was not altered in any of the five data sets across sequential application of criteria.

Between 37% and 44% of the chemotherapy patients in the fully restricted cohort initiated a second line containing immunotherapy prior to the data cutoff (data not shown).

DISCUSSION

This study evaluated the reproducibility and performance of rwOS across five real-world mNSCLC patient data sets receiving chemotherapy or PD-1 combination after applying selected clinical trial inclusion/exclusion criteria.[6] Specifically, the study evaluated whether the application of common inclusion/exclusion criteria and methods across different data sources would result in similar rwOS findings, and if not, whether the observed

Table 4 Unadjusted associations between use of frontline therapy and rwOS

	Keynote 189	A	B	C	D	E
Number of patients	616	346/54	1,501/405	232/36	748/132	155/125
Number of events (Chemotherapy/PD-1 combination)	235	203/25	1094/216	163/22	421/61	164
12-month OS (Chemotherapy/PD-1 combination)	0.49/0.69	0.57/0.67	0.45/0.53	0.53/0.44	0.56/0.62	0.51/0.56
Unadjusted Hazard ratio (HR) for death (95% CI)	0.49 (0.38–0.64)	0.99 (0.65–1.50)	0.79 (0.68–0.92)	1.10 (0.70–1.72)	0.92 (0.70–1.20)	0.97 (0.71–1.33)
Age HR (95% CI)						
<65	0.43 (0.31–0.61)	1.25 (0.71–2.20)	0.75 (0.60–0.94)	1.30 (0.68–2.49)		1.26 (0.79–2.0)
>65	0.64 (0.43–0.95)	0.80 (0.42–1.53)	0.84 (0.69–1.03)	0.94 (0.49–1.81)	1.05 (0.88–1.26)	0.74 (0.48–1.15)
Sex HR (95% CI)						
Male	0.70 (0.50–0.99)	0.81 (0.44–1.48)	0.75 (0.63–0.91)	1.08 (0.49–2.36)	0.96 (0.80–1.15)	1.31 (0.85–2.02)
Female	0.29 (0.19–0.44)	1.24 (0.70–2.22)	0.81 (0.63–1.03)	1.07 (0.62–1.87)		0.72 (0.45–1.14)
PD-L1 expression status HR (95% CI)						
<1%	0.59 (0.38–0.92)	1.84 (0.58–5.81)	1.02 (0.69–1.51)	1.17 (0.34–4.03)		NA
>1%	0.47 (0.34–0.66)	1.45 (0.57–3.67)	0.70 (0.55–0.90)	0.76 (0.33–1.74)	0.80 (0.56–1.14)	NA
1–49%	0.55 (0.34–0.90)	1.55 (0.56–4.30)	0.77 (0.57–1.05)	1.24 (0.44–3.45)	0.77 (0.52–1.13)	NA
>50%	0.42 (0.26–0.68)	0.91 (0.08–10.21)	0.69 (0.44–1.08)	0.31 (0.06–1.53)	0.70 (0.37–1.32)	NA
Brain metastases HR (95% CI)						
Evidence of	0.36 (0.20–0.62)	1.63 (0.80–3.32)	0.96 (0.67–1.39)	0.95 (0.32–2.81)		1.44 (0.73–2.84)
No evidence of	0.53 (0.39–0.71)	0.78 (0.46–1.32)	0.76 (0.65–0.90)	1.12 (0.68–1.83)	1.04 (0.80–1.34)	0.86 (0.60–1.23)
Platinum-based drug HR (95% CI)						
Carboplatin	0.52 (0.39–0.71)	0.94 (0.62–1.44)	0.78 (0.67–0.90)	NA		0.94 (0.69–1.29)
Cisplatin	0.41 (0.24–0.69)	N/A	1.34 (0.19–9.71)	1.03 (0.66–1.63)	0.58 (0.38–0.88)	NA

CI, confidence interval; NA, not available; OS, overall survival; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; rwOS, real-world overall survival.

Table 5 Adjusted associations between frontline therapy and rwOS in the fully restricted cohorts

Covariate	A	B	C	D	E
Overall adjusted HR (95% CI)					
Chemotherapy	Ref.	Ref.	Ref.	Ref.	Ref.
PD-1 combination	0.99 (0.65–1.51)	0.80 (0.69–0.93)	1.15 (0.71–1.85)	0.96 (0.73–1.26)	0.95 (0.69–1.30)
Age HR (95% CI)					
<65	Ref.	Ref.	Ref.	Ref.	Ref.
>65	1.18 (0.90–1.55)	1.16 (1.04–1.30)	1.20 (0.90–1.61)	1.04 (0.86–1.25)	0.72 (0.53–0.98)
Gender HR (95% CI)					
Male	Ref.	Ref.	Ref.	Ref.	Ref.
Female	0.97 (0.74–1.26)	0.80 (0.72–0.90)	0.79 (0.59–1.06)	0.94 (0.79–1.13)	0.84 (0.62–1.15)
ECOG PS HR (95% CI)					
0	Ref.	Ref.	Ref.	Ref.	Ref.
1	1.93 (1.38–2.70)	1.40 (1.21–1.63)	1.51 (0.93–2.46)	1.31 (1.01–1.69)	1.07 (0.69–1.64)
Unknown	1.09 (0.74–1.59)	1.25 (1.08–1.45)	1.48 (0.94–2.34)	1.29 (0.96–1.75)	0.90 (0.53–1.52)
Smoking status HR (95% CI)					
Evidence of history of	Ref.	Ref.	Ref.	Ref.	Ref.
No evidence of history of	0.99 (0.67–1.47)	0.84 (0.70–1.00)	1.11 (0.73–1.68)	1.01 (0.78–1.32)	1.54 (0.61–3.90)
Unknown/missing population	NA	NA	NA	0.76 (0.39–1.48)	0.78 (0.48–1.28)

CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; HR, hazard ratio; NA, not available; OS, overall survival; PD-1, programmed cell death protein 1; Ref., reference; rwOS, real-world overall survival.

results could be explained by differences in the underlying patient populations or data-specific characteristics. While a trial replication or direct comparison was outside the scope of this work, trial-based inclusion/exclusion criteria were used to align the real-world populations and provide a relative benchmark for evaluating the performance of the real-world end point. The stepwise application of inclusion/exclusion criteria, intended to provide additional insights on factors that drove differences between estimates, did not indicate convergence or divergence of results.

The estimated treatment effects varied across data sets. Sample size limitations may have contributed to the lack of statistical significance in some of the real-world data sets compared with KEYNOTE-189. Real-World OS for patients receiving immunotherapy was shorter in real-world cohorts than in the trial, consistent with the findings from other real-world NSCLC analyses.[7–9] The range of observed results across real-world cohorts and as compared with the trial may be partially attributable to outstanding differences in cohort composition, data missingness, interpretation of end-point definitions and their completeness, and differences in routine vs. clinical trial care and data collection patterns.

While several, common inclusion/exclusion criteria were used, differences in their application and in missing data patterns across the five groups may have contributed to variability of observed results. Data missingness is a known and challenging aspect of observational data. Specifically, the proportion of patients missing ECOG PS was up to a quarter and patients lacking laboratory values to ascertain organ function comprised up to 87% of study cohorts, allowing substantial heterogeneity in cohort characteristics. Missing International Classification of Diseases (ICD) or laboratory values may have resulted in differential misclassification of entry criteria and key covariates, potentially biasing observed estimates.

The proportion of patients with evidence of brain metastases in real-world data sets also varied widely, as determined using ICD codes. Given the low sensitivity of ICD codes in identifying brain metastases, differences in this important prognostic factor could lead to variable estimates of the 12-month rwOS. Additional, unmeasured sources of heterogeneity, such as comorbidities, socioeconomic status, health-insurance coverage, and variation in care between community and academic practices may have contributed to the variability in adjusted estimates. Given high rates of crossover in the chemotherapy groups, potentially variable treatment timing, and duration and dosing may have contributed to the relatively strong performance of patients receiving chemotherapy.[10]

Differences in mortality ascertainment may have also contributed to the observed range of results. The sensitivity and specificity of mortality information, obtained from a variety of sources ranging from only structured EHRs to composite approaches of chart review, third party mortality data sources, and the Social Security Death Index data, varied across real-world data sets. Poor completeness of mortality data leads to overestimated survival,[7] and lower statistical power. Additionally, should completeness and/or accuracy of the data vary within cohorts (e.g., due to improved records over time), measures of association may be biased. Finally, granularity of available death dates and handling of partially complete dates may have also varied across groups.

Descriptive comparisons to KEYNOTE-189 should consider differences in patient care in a trial setting. Strict trial protocols dictate regular data collection at baseline and follow-up intervals for RECIST objective response and mortality assessments, whereas

real-world studies observe care as it is delivered and recorded in clinical practice. Consistent reporting of data completeness is critical to inform appropriate analysis, including potential sensitivity analyses, and interpretation of results, both within and across data sets.

The study's ability to apply further inclusion/exclusion criteria, as well as conduct sensitivity analyses, was limited by sample size considerations. Some data sources used only structured data, which limited the extent of the covariate information collected compared with using unstructured data. Since participating groups had different underlying data availability, sensitivity analyses could not be performed consistently across groups, limiting insights into the mechanism and impact of missing data, for example on ECOG PS or kidney and liver organ function. The study partially used noncontemporaneous controls, which could complicate outcome interpretation. Finally, blinding of groups precluded the discussion of data set-specific nuances which could inform observed differences.

In future work, sensitivity analyses can help elucidate data set–specific factors that may drive results. These include only selecting patients who: (i) had stage IV NSCLC at diagnosis; (ii) initiated frontline treatment in 2017 or 2018; (iii) had comparable distribution of potential follow-up across cohorts; (iv) had known ECOG PS and organ function; and (v) had known timing and duration of treatment prior to crossover. It is also important to understand missingness for core variables to inform the selection of proper analytic methods for main and/or sensitivity analyses (imputation-based or model-based approaches vs. complete case analysis). In studies evaluating multiple sources of RWD, sequential application of eligibility criteria can be considered to evaluate consistency of results across data sets and the impact of select clinical characteristics. Designing a shared RWD master protocol *a priori* may assist in understanding differences, promoting efficiency, and increasing reproducibility. Additionally, sensitivity analyses assessing the variability resulting from different mortality assessment approaches and determination of exact death date across data sets (**Table 1**) could inform the relative contribution of these factors to observed differences. Lastly, future analyses that include additional clinical demographic factors and social determinants of health merit future investigation.

This study sought to evaluate the performance of rwOS and the considerations necessary to assess directionality of treatment associations in a real-world population across five US oncology EHR RWD providers with different sources of patient data by aligning the patient population with key inclusion/exclusion criteria from the KEYNOTE-189 study. Such efforts to achieve consistency across RWD sources are necessary to distinguish true treatment effects from ones driven by methodological choices, missing data, confounding, and unmeasurable influences on treatment choice. While an association between frontline treatment and rwOS was not consistently detected, differences in methodologies, delivery of care (protocol vs. observational), capture of critical data elements (required routinely throughout RCT), and residual heterogeneity in real-world patient cohorts help contextualize the observed similarities and differences across the real-world data sets and as compared with the trial. Measuring real-world effectiveness and safety in routine care alongside clinical trials may have an important role in completing the picture of how well a therapy works, for which patients it is most effectively useful, and under what conditions in the future. Building on this research, agreement on minimum reporting and performance standards and capturing of post-baseline events (e.g., frequency and timing of treatment crossover) or subsequent treatments, as well as a process to evaluate real-world end points across data sets could inform best practices that may help unlock the potential of EHR-derived RWD.

AUTHOR CONTRIBUTIONS
L.L., O.T., E.G.-M., S.B., A.J.B., M.B., J.B.C., A.B.C., J.L.E., N.J.R., M.A.I., M.D.S., Y.N., D.R.R., and J.A. designed the research. O.T., E.H., C.S., M.S., M.I., and J.W. performed the research. L.L., O.T., E.G.-M., J.B.C., A.J.B., A.B.C., J.L.E., C.W., N.J.R., M.S., M.A.I., J.W., and Y.N. analyzed the data. L.L., O.T., E.G.-M., S.B., A.J.B., J.B.C., A.B.C., J.L.E., C.S., N.J.R., M.D.S., M.S., M.A.I., Y.N., and D.R.R. wrote the manuscript.

DISCLAIMERS
The authors assume full responsibility for analyses and interpretation of these data.

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

1. Elting, L.S. *et al.* Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. *Cancer* **106**, 2452–2458 (2006).

2. Bolislis, W.R., Fay, M. & Kühler, T.C. Use of real-world data for new drug applications and line extensions. *Clin. Ther.* **42**, 926–938 (2020).

3. National Academies of Sciences, Engineering, and Medicine. *Examining the Impact of Real-world Evidence on Medical Product Development: Proceedings of a Workshop Series* (The National Academies Press, Washington, DC, 2019).

4. Stewart, M. *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 (2019).

5. Gandhi, L. *et al*. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N. Engl. J. Med.* **378**, 2078–2092 (2018).

6. Curtis, M.D. *et al*. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv. Res.* **53**, 4460–4476 (2018).

7. La, J. *et al*. Real-world outcomes for patients treated with immune checkpoint inhibitors in the veterans affairs system. *JCO Clin. Cancer Inform.* **4**, 918–928 (2020).

8. Khozin, S., *et al*. Real-world outcomes of patients with metastatic non-small cell lung cancer treated with programmed cell death protein 1 inhibitors in the year following US regulatory approval. *Oncologist* **24**, 648–656 (2019).

9. Khozin, S. *et al*. Real-world progression, treatment, and survival outcomes during rapid adoption of immunotherapy for advanced non-small cell lung cancer. *Cancer* **125**, 4019–4032 (2019).

10. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. Final Version. E9(R1);2019 <https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf> (2019).

Comparing Findings From a Friends of Cancer Research Exploratory Analysis of Real-World End Points With the Cancer Analysis System in England

Pia Horvat, PhD[1]; Christen M. Gray, PhD[1]; Alexandrina Lambova, BSc[2]; Jennifer B. Christian, PharmD, MPH, PhD[3]; Laura Lasiter, PhD[4]; Mark Stewart, PhD[4]; Jeff Allen, PhD[4]; Paul Clarke, BA[5,6]; Cong Chen, PhD[5,6]; and Adam Reich, BA[1]

PURPOSE This study compared real-world end points extracted from the Cancer Analysis System (CAS), a national cancer registry with linkage to national mortality and other health care databases in England, with those from diverse US oncology data sources, including electronic health care records, insurance claims, unstructured medical charts, or a combination, that participated in the Friends of Cancer Research Real-World Evidence Pilot Project 1.0. Consistency between data sets and between real-world overall survival (rwOS) was assessed in patients with immunotherapy-treated advanced non–small-cell lung cancer (aNSCLC).

PATIENTS AND METHODS Patients with aNSCLC, diagnosed between January 2013 and December 2017, who initiated treatment with approved programmed death ligand-1 (PD-[L]1) inhibitors until March 2018 were included. Real-world end points, including rwOS and real-world time to treatment discontinuation (rwTTD), were assessed using Kaplan-Meier analysis. A synthetic data set, Simulacrum, on the basis of conditional random sampling of the CAS data was used to develop and refine analysis scripts while protecting patient privacy.

RESULTS Characteristics (age, sex, and histology) of the 2,035 patients with immunotherapy-treated aNSCLC included in the CAS study were broadly comparable with US data sets. In CAS, a higher proportion (46.7%) of patients received a PD-(L)1 inhibitor in the first line than in US data sets (18%-30%). Median rwOS (11.4 months; 95% CI, 10.4 to 12.7) and rwTTD (4.9 months; 95% CI, 4.7 to 5.1) were within the range of US-based data sets (rwOS, 8.6-13.5 months; rwTTD, 3.2-7.0 months).

CONCLUSION The CAS findings were consistent with those from US-based oncology data sets. Such consistency is important for regulatory decision making. Differences observed between data sets may be explained by variation in health care settings, such as the timing of PD-(L)1 approval and reimbursement, and data capture.

INTRODUCTION

Randomized controlled trials (RCTs) are the gold standard for assessing safety and efficacy of drugs; however, RCTs have limitations such as controlled settings, selective patient populations, and short-term study outcomes that might not be generalizable to heterogenous real-world populations and settings.[1-3] There is a need to understand the effectiveness of medical interventions in representative real-world populations and for accelerated clinical evidence, which can be met by improved data analytics and unprecedented availability of real-world data (RWD).[4,5] Regulatory bodies including the US Food and Drug Administration, the European Medicines Agency, and health technology assessment agencies, such as the National Institute for Health and Care Excellence, are increasingly acknowledging the importance of RWD.[6-8] However, for RWD to be used widely to supplement or augment clinical trials, the validity of

real-world clinical end points in specific RWD must be established.

A pilot RWD project conducted by Friends of Cancer Research (*Friends*) evaluated the reliability of real-world end points among programmed death-1 or programmed death ligand-1 (PD-[L]1) inhibitor–treated advanced non–small-cell lung cancer (aNSCLC; patients with advanced-stage IIIB-IV or early-stage recurred or progressed non–small-cell lung cancer [NSCLC]) patients by convening six oncology-focused US health care data sets from participating organizations, including Cancer Research Network, Cota Healthcare, Flatiron Health, IQVIA, OptumLabs Data Warehouse, and PCORnet.[5] The findings of the pilot project demonstrated that worthwhile data can be aggregated from diverse research-ready RWD and that real-world end points measured from these data are consistent with each other and are directionally similar to those observed in RCTs,[9] especially in relation to

CONTEXT

Key Objective

The Cancer Analysis System (CAS) is a national cancer registry with linkage to national mortality and other health care databases in England. Real-world end points, including overall survival and time to treatment discontinuation, were analyzed from the CAS database and compared with diverse US oncology data sources used in the Friends of Cancer Research Real-World Evidence Pilot Project 1.0.

Knowledge Generated

The CAS analysis demonstrates the consistent performance and validity of real-world end points across geographically and structurally diverse settings. It also validates the findings from the CAS database by way of comparison with multiple US data sets.

Relevance

Evaluation and comparison of the strengths, limitations, and validity of specific real-world data for addressing defined clinical questions can facilitate development of guidelines on fit-for-purpose real-world data and defining standards for real-world evidence studies intended to inform regulatory decisions.

overall survival. That said, none of the US-based data sets that participated in the pilot captured mortality systematically in all patients and the approaches to measure mortality varied across the US health care data sources.

The present study expands on the work of the pilot study by including the Cancer Analysis System (CAS) in England. CAS is a cancer registry that covers more than 99% of all patients with cancer in England and contains data on patient and tumor characteristics, treatments, hospitalizations, and mortality. It combines data from the Cancer Outcomes and Services Dataset (COSD), collected by Public Health England's National Disease Registration Service, with linkage to other national data sets, including the Systemic Anticancer Therapy (SACT), Hospital Episode Statistics, the National Radiotherapy Dataset, and mortality data from the Office of National Statistics (ONS).[10] The CAS database is a vital resource for understanding patient care and outcomes in England and for supporting drug research and development globally. The present study aims to validate and compare real-world end points extracted from CAS with those from the six US oncology data sets from *Friends*' original pilot. Considering the population and health care system differences between the US data sets and CAS, along with the complete capture of mortality data in CAS, this study will help in the validation of real-world end points across diverse populations and health care settings.

### PATIENTS AND METHODS

Study Design and Objectives

The present study followed the protocol of the US-based pilot project.[5] It used a retrospective observational cohort design in patients with aNSCLC (stage IIIB-IV) leveraging the CAS database. However, because of differences in content and structure of the CAS data and US data sets, some distinctions in the study design should be noted. For example, the CAS study cohort included only incident stage IIIB-IV

histologically confirmed NSCLC patients, unlike several of the US-based data sets, which also included early-stage patients with documented evidence of progression. The overarching objective of the present study was to apply the original *Friends*' pilot study protocol to an ex-US setting and a different type of RWD, namely, a national cancer registry linked to national mortality and other health care databases. The study also had the following secondary objectives:

1. To describe and compare the demographic and clinical characteristics of patients with aNSCLC treated with PD-(L)1 immune checkpoint inhibitors.
2. To assess real-world end points (real-world overall survival [rwOS] and real-world time to treatment discontinuation [rwTTD]) in patients with aNSCLC treated with PD-(L)1 immune checkpoint inhibitors, overall and by clinical and demographic characteristics.
3. To highlight the performance and consistency of real-world end points, particularly rwOS, in CAS (with complete ascertainment of mortality) with the US data sets.

Patient Cohort

The study cohort flow diagram is shown in Figure 1. Histologically confirmed incident stage IIIB-IV NSCLC (referred to as aNSCLC) patients diagnosed between January 1, 2013, and December 31, 2017, who initiated treatment with approved PD-(L)1 immune checkpoint inhibitors (ie, nivolumab, pembrolizumab, or atezolizumab) until March 2018 (most recent available SACT data[11]) were eligible for analysis. Patients in CAS were excluded if they had unconfirmed stage at diagnosis and a concomitant non-melanoma malignancy or had systemic cancer treatment recorded more than 30 days before the aNSCLC diagnosis. Patients retrieved from CAS had a shorter minimum potential follow-up of 3 months (until March 2018) compared with a minimum potential follow-up of 6 months for US-based data sets. Index date was defined as the initiation of therapy containing any PD-(L)1 inhibitor.
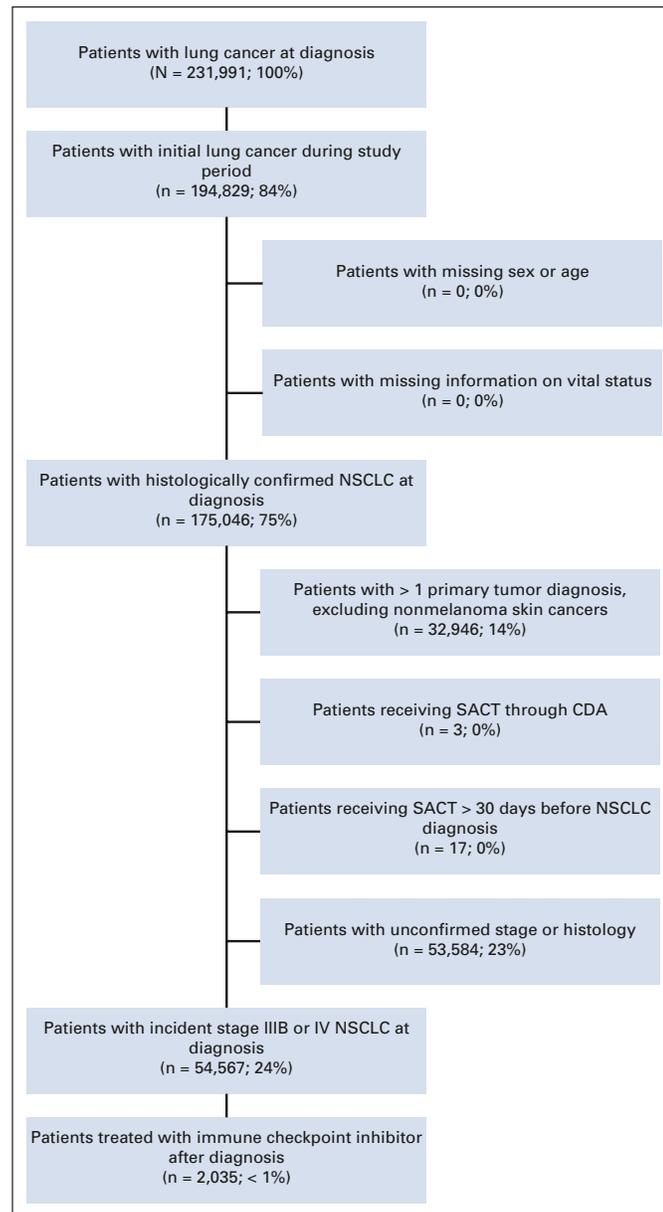
FIG 1. Study cohort flow diagram. CDA, Cancer Drugs Fund; NSCLC, non–small-cell lung cancer; SACT, Systemic Anticancer Therapy.

Flow diagram contents:

- Patients with lung cancer at diagnosis (N = 231,991; 100%)
- Patients with initial lung cancer during study period (n = 194,829; 84%)
  - Patients with missing sex or age (n = 0; 0%)
  - Patients with missing information on vital status (n = 0; 0%)
- Patients with histologically confirmed NSCLC at diagnosis (n = 175,046; 75%)
  - Patients with > 1 primary tumor diagnosis, excluding nonmelanoma skin cancers (n = 32,946; 14%)
  - Patients receiving SACT through CDA (n = 3; 0%)
  - Patients receiving SACT > 30 days before NSCLC diagnosis (n = 17; 0%)
  - Patients with unconfirmed stage or histology (n = 53,584; 23%)
- Patients with incident stage IIIB or IV NSCLC at diagnosis (n = 54,567; 24%)
- Patients treated with immune checkpoint inhibitor after diagnosis (n = 2,035; < 1%)

Data Sources

Data were collected retrospectively from the CAS database. The CAS database comprises several linked databases. For the present study, COSD, SACT, and ONS were used. COSD contains patient demographics (eg, age, sex, ethnicity, and geographic region) and tumor characteristics (eg, staging, morphology, and performance status). The SACT dataset[11] is the national mandatory collection of systemic anticancer therapy from all National Health Service England chemotherapy providers. The CAS database contains patient-level data, which is subject to strict data protection rules. The process to access the CAS data has historically been long and complex. To simplify access to the CAS database for research, Health Data Insight CIC has developed a programming tool called Simulacrum, which contains artificial patient data and allows analytical programs to be developed before running queries on the CAS database. Simulacrum mimics the structure and types of CAS data, enabling faster analyses by enabling debugging and validation of programming code before running analyses on the CAS database. The details of the data sets or health care data organizations,

which participated in the US-based pilot study, are described in the original publication.[5]

End Point Definition

End point definitions used in this analysis were closely aligned with the US pilot study protocol and are given in Appendix Table A1 where end points had to be adapted.[5]

Statistical Analysis

Descriptive statistics (counts [%] or means [with standard deviations]) were performed for demographic and clinical characteristics. The lines of therapy were derived using an algorithm for NSCLC developed with input from clinicians.[12] Continuous variables were summarized using medians and interquartile ranges, and frequencies were calculated for categorical variables. Kaplan-Meier analysis was performed for time-to-event end points (rwOS and rwTTD); these end points were subsequently summarized using median time (in months) with the associated 95% CIs. Spearman's rank-order correlation was used to estimate the correlations between rwOS and rwTTD.

RESULTS

Patient Identification and Demographic Characteristics

Table 1 shows the demographic and clinical characteristics of patients with aNSCLC treated with PD-(L)1 inhibitors in the CAS and six US oncology data sets.[5] Overall, 2,035 patients with aNSCLC were included in the analysis. The median age at advanced NSCLC diagnosis was 67 (interquartile range [IQR]: 60.0-73.0) years and 68.3 (IQR: 61.6-73.7) years at PD-(L)1 inhibitor initiation (Appendix Table A2). Median age was comparable with the US pilot study; however, the proportion of PD-(L)1–treated patients in the 65-74 years age group was higher in CAS (47.7%), compared with 30%-40% in the US data sets. Gender distribution in CAS (54.3% male) was within the range of the US data sets. In terms of ethnicity, the patient population in CAS was more homogeneous (91.6% White) than in the US data sets (65%-87%). The majority of CAS patients (81.7%) were diagnosed with stage IV disease, and 18.3% with stage IIIB. In the four US data sets with data on disease stage at initial diagnosis, 62%-91% of patients were diagnosed with stage IV disease. Distribution of histology was broadly similar to the US data sets (71.2 v range 66%-74% in US data sets), with a higher proportion of nonsquamous histology at diagnosis (Appendix Table A2). In the CAS, a considerably higher proportion (46.7%) of PD-(L)1 inhibitors were given as first-line treatment than in US data sets (18%-30%). Almost similar proportion (44.1%) of PD-(L)1 inhibitors in CAS were given in second line and relatively few (7.9% of patients) in third line or beyond. The majority (93.4%) of CAS patients treated with a PD-(L)1 inhibitor (in any therapy line) did not receive a subsequent therapy line during the study observation period. A subgroup analysis of patients below 50 years showed that the majority of patients received PD-(L)1 inhibitor as their second-line therapy (51.7%; n = 45) and 91% (n = 79) did not receive a subsequent line after their PD-(L)1 inhibitor treatment (Appendix Table A4). The median time from advanced diagnosis to PD-(L)1 inhibitor initiation of 4.9 (IQR: 1.6-11.7) months in CAS was shorter than in US data sets (ranging from 6 to 8 months), likely because of higher first-line PD-(L)1 usage in CAS. At 10.6 (IQR: 3.9-16.0) months, the median follow-up time from PD-(L)1 inhibitor initiation was slightly longer in CAS than in US data sets (6-9 months).

Real-World End Points

Table 2 shows the median time-to-event estimates for real-world end points in CAS. The median rwOS (Fig 2) was 11.4 months (95% CI, 10.4 to 12.7), consistent with the range observed in US data sets (from 8.6 to 13.5 months). The median rwTTD for CAS was 4.9 months (95% CI, 4.7 to 5.1), within the range of US data sets (3.2-7.0 months).

Median rwOS and 95% CI segmented by treatment setting and patient demographic characteristics are given in Table 3. Median rwOS in CAS was higher for the age groups 65-74 and ≥ 75 years (both around 12.2 months). Unexpectedly, survival improved with age among patients in CAS and a similar pattern was reported only in US data set E. In CAS, median rwOS at 13.5 months was higher in female than in male patients (9.95 months), consistent with the pattern seen in most of the US data sets. Within the CAS database, stage III patients had higher median rwOS (15.8 months) than many US data sets (B, C, and D). For CAS patients, median rwOS was higher for nonsquamous than squamous aNSCLC (11.8 months) within the range of US data sets (from 8.7 to 14.2 months). Median rwOS decreased almost linearly with increasing therapy line in CAS, whereas in US data sets, there was greater variability in survival across therapy lines with gains in survival also observed in higher lines. In the first-line setting, the median rwOS in CAS was 13.4 months (95% CI, 11.9 to 15.6). In the second-line setting, it was 10.2 months (95% CI, 8.9 to 11.8). In the US-based data sets, there was a wide range in first-line median rwOS from 9.4 to 20.8 months. In the second-line setting, where the US range for median rwOS was narrower 7.9-11.7 (95% CI, 6.5 to 10.5-10.9 to 12.8), the CAS results were also consistent with the observed range (Appendix Table A3). Unlike the US data sets, mortality information is systematically completed for all patients in CAS through linkage with ONS. Correlations between rwOS and rwTTD for CAS and US data sets are shown in Appendix Table A5 (0.7 in CAS compared with the US range of 0.6-0.9).

DISCUSSION

This study compared the findings from the US-based Friends' pilot project with a UK Cancer Registry (CAS database) to characterize the validity of real-world end points for addressing clinically relevant questions on treatment effectiveness, determination of unmet need, and use of PD-(L)1 inhibitors across diverse health care settings. This

TABLE 1. Description of Demographic and Clinical Characteristics of Patients With aNSCLC Treated With PD-(L)1 Checkpoint Inhibitor

Patient Characteristic	CAS (N = 2,035)	Data Sets A-F (N = 269-6,924)
Demographic characteristics		
Median age at advanced diagnosis, years (IQR)	67.00 (60.00-73.00)	64-70 (14-15)
Median age at PD-(L)1 inhibitor initiations, years (IQR)	68.29 (61.55-73.68)	65-71 (14)
Age categories at PD-(L)1 inhibitor initiation, years, No. (%)		
≤ 49	87 (4.28)	8-219 (3-5)
50-64	567 (27.86)	65-2,048 (24-45)
65-74	971 (47.71)	94-2,504 (33-39)
≥ 75	410 (20.15)	86-2,153 (15-38)
Age categories at PD-(L)1 inhibitor initiation, years, No. (%)		
< 75	1,625 (79.85)	167-4,771 (62-85)
≥ 75	410 (20.15)	86-2,153 (15-38)
Sex, No. (%)		
Female	931 (45.75)	125-3,172 (44-49)
Male	1,104 (54.25)	143-3,752 (51-56)
Unknown or missing	0	0-5
Race or ethnicity, No. (%)		
White	1,865 (91.65)	160-4,969 (65-87)
Black or African American	—	14-594 (8-13)
Asian	27 (1.33)	6-155 (1-19)
Chinese	1,662 (81.67)	
Others	42 (2.06)	1-580 (1-9)
Mixed	—	
Unknown or missing	67 (3.29)	0-2,085
Clinical characteristics		
Stage at initial diagnosis, No. (%)		
0 or occult	—	0-2 (0)
I	—	18-496 (6-7)
II	—	17-426 (6-7)
III[a]	373 (18.33)	17-1,494 (7-23)
IV	1,662 (81.67)	161-4,335 (62-91)
Not reported		10-175
Histology, No. (%)		
Nonsquamous NSCLC	1,448 (71.15)	194-4,679 (66-74)
Squamous cell carcinoma	521 (25.60)	61-1,983 (17-30)
NSCLC NOS	24 (1.18)	10-262 (3-10)
Others specified	42 (2.06)	—
Unknown or missing		4
Line of first PD-(L)1 inhibitor, No. (%)		
1 (no prior therapy received)	950 (46.68)	77-2,074 (18-30)
2	898 (44.13)	87-3,357 (32-56)
3	129 (6.34)	51-1,012 (15-20)
≥ 4	33 (1.62)	44-481 (3-20)
Missing or unknown	25 (1.23)	
Follow-up time, months, median (Q1, Q3)		
From advanced diagnosis to PD-(L)1 inhibitor initiation	4.86 (1.61, 11.70)	6-8 (2.13-4.15)
From advanced diagnosis	16.79 (10.61, 26.05)	14-18 (8.25-10.31)
From PD-(L)1 inhibitor initiation (any LOT)	10.61 (3.91, 16.00)	6-9 (2.12-4.13)

Abbreviations: aNSCLC, advanced non–small-cell lung cancer; CAS, Cancer Analysis System; IQR, interquartile range; LOT, line of therapy; NA, not applicable; NOS, not otherwise specified; NSCLC, non–small-cell lung cancer; PD-(L)1, programmed death-1 or programmed death ligand-1; Q1, quartile 1; Q3, quartile 3.
[a]Only stage IIIB patients were included in the analysis of CAS data.

TABLE 2. Median Time and 95% CI for Real-World Extracted End Points Comparing CAS With US Data Sets

Data Set	rwOS	rwTTD
CAS	11.40 (10.38 to 12.65)	4.86 (4.67 to 5.09)
A	13.50 (12.80 to 14.50)	7.03 (6.27 to 9.97)
B	15.78 (12.2 to 24.59)	3.25 (2.76 to 3.75)
	8.58 (7.56 to 10.36)[a]	
C	8.67 (6.83 to 10.02)	4.70 (3.68 to 5.52)
D	9.15 (8.82 to 9.51)	3.21 (3.21 to 3.44)
E	12.69 (11.70 to 13.87)	3.63 (3.40 to 3.87)
F	12.30 (9.61 to 16.94)	4.60 (3.71 to 6.32)

Abbreviations: CAS, Cancer Analysis System; rwOS, real-world overall survival; rwTTD, real-world time to treatment discontinuation.
[a]Sites with social security or state death data, censored at the estimated earliest date such that data should be available if no death was observed.

study is based on the UK National Health Service, which operates on a fundamentally different health care model (ie, free for all at the point of use). Furthermore, this study helps to assess the performance of real-world end points because of the completeness of patient and mortality information and validate the CAS database by way of comparison with diverse US data sets.

A comparison of the present study with the US-based pilot study highlights a few key differences between CAS and the US data sets in terms of population and database characteristics. CAS captures clinical characteristics related to TNM staging, whereas this information is generally not available in US-based claims data. Thus, unlike most US-based data sets, which also included patients early-stage progressed NSCLC, the present study included only incident stage IIIB-IV NSCLC, with more than 80% of patients diagnosed in stage IV. In CAS, the median follow-up time from PD-(L)1 inhibitor initiation was longer than that in the US data sets, mainly attributed to the large proportion of PD-(L)1 inhibitors in patients within first-line setting. The estimated overall median rwOS in CAS was within the range of estimates observed in the US data sets although the observed US range was relatively wide. Variations in rwOS in the US data sets are likely due to challenges in accessing mortality data, as deaths are not routinely recorded in most electronic health care records or insurance claims. In contrast to CAS where mortality is ascertained through regularly updated linkage with official national mortality statistics, some US-based RWD rely on published obituary or insurance data to supplement the gaps in mortality information.[13] Combining data from multiple research-ready real-world sources under a common framework can enhance the reliability of real-world end points. Therefore, combining different data types and sources allows timely availability of data, addresses missing data, and improves completeness to create robust data sources.[13]

RCTs demonstrate efficacy of PD-(L)1 inhibitors under optimal settings, whereas real-world evidence (RWE) provides realistic estimates of effectiveness in routine clinical practice. There can be differences in the results of RWD and RCTs; however, considering the complexities in health care systems and high degree of variation in treatment response in the real world, RWD may differ from RCTs but still be valid.[14]

The CAS database has several key strengths including the use of synthetic data to facilitate data access while maintaining patient confidentiality, linkage between a national population–based cancer registry of England with a national systemic therapy database (SACT) and other national databases, and most importantly, the inclusion of mortality data obtained via linkage with ONS. Mortality data are reliably captured through regular updates as structured fields and undergo regular evaluations of validity. This includes a tracing process annually, which ensures the completeness of the annual mortality update. Completeness of survival or mortality data emphasizes and validates the overall findings of the study.

This study also highlights some limitations of the CAS database. First, the CAS database currently does not contain biomarker data although this information is expected in the future. Information on patients' biomarker status would be relevant in interpreting the results of this study, given that response to treatment with PD-(L)1 inhibitors has been shown to correlate with the NSCLC molecular profile. Previous studies have failed to demonstrate unequivocal survival benefits of PD-(L)1 inhibitor monotherapy in patients with epidermal growth factor receptor (EGFR) or anaplastic lymphoma kinase (ALK) genetic alterations compared with standard chemotherapy.[15] Second, disease progression and treatment response are also not directly captured. Treatment-related intermediate end points, such as time to next treatment or death, can be derived from the available data and used to make inferences about possible disease progression. However, such intermediate treatment-related end points are generally limited to patients who have received a subsequent therapy. Compared with most electronic health care records, CAS has a longer lag time as it combines data from national cancer registry and other national databases, which contributed to a shorter observed follow-up time in our study. Because of the high completeness of mortality information in CAS, stronger insights can be gleaned into the value of nontraditional end points, such as time to treatment discontinuation (TTD). The strong correlation between overall survival and TTD in these data further elucidates the role of TTD and its potential use as an earlier indicator of clinical benefit. This could prove to be useful in evaluating the effectiveness of therapies in a broader range of RWD sets, particularly in instances where mortality information may be less accurate or incomplete.

Compared with the US-based study, a higher proportion of patients in CAS received treatment with PD-(L)1 inhibitor in the first-line setting. This was largely due to the time period studied, which reflects the later approval and
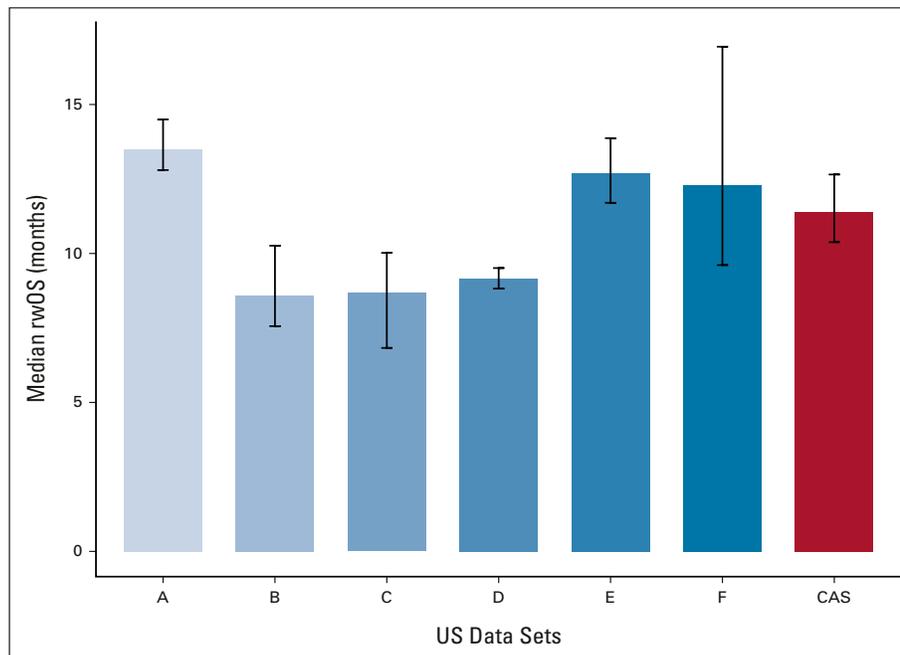
FIG 2. Comparison of median rwOS (with 95% CI) in CAS with six US data sets in patients with aNSCLC treated with PD-(L)1 inhibitors. aNSCLC, advanced non–small-cell lung cancer; CAS, Cancer Analysis System; OS, overall survival; PD-(L)1, programmed death-1 or programmed death ligand-1; rwOS, real-world overall survival.

reimbursement—and in the case of nivolumab also, more restricted use through the Cancer Drugs Fund—of PD-(L)1 inhibitors in England. The earlier approval and reimbursement of pembrolizumab in England, relative to the later approval and conditional reimbursement (through CDF) of nivolumab, has also led to a disproportionate number of patients treated with pembrolizumab in our study. Patients who received other PD-(L)1 inhibitors (ie, nivolumab) through the CDF are under-represented in our analysis because clinical outcomes data were not accessible for these patients.

RCTs are widely used by regulators because of their strong internal validity, despite having limited generalizability to real-world settings.[16] The CAS database represents more than 99% of the population of England and is nationally representative for patients with cancer. Thus, the results of this study are broadly representative of patients with aNSCLC treated with PD-(L)1 inhibitors in England over the observed time period. National coverage makes CAS a valuable resource for assessing real-world end points, and it may also be used for research in rare cancers, where RCTs may be less feasible. Furthermore, studies similar to ours can establish reference ranges for direct comparison of RWD with clinical trial data in oncology paving the way for alternative study designs.[17]

RWD collection practices for treatment and clinical parameters have improved over time in terms of data completeness and quality, and this trend is expected to continue.[18] In 2021, CAS is expected to incorporate biomarker data, which can provide additional insights into clinical characteristics of patients. Extension of initiatives such as *Friends* to multicountry settings will further support the development of best practices for the generation and evaluation of RWE to supplement RCTs in regulatory decision making and inform the development of future regulatory guidance.[14] RWE can support regulatory decision making about new or expanded medication indications in a number of ways, from the more established use of external controls for single-arm trials, which have also been used as the basis of US Food and Drug Administration–accelerated approvals, to the growing interest in the use of pragmatic trials and nonrandomized RWE from health care databases, particularly in situations where real-world outcomes and clinical practice patterns differ significantly from the tightly controlled RCT settings.[6,19] Studies similar to ours that establish research processes and evaluate the strengths, limitations, and validity of specific RWD for addressing defined clinical questions can support development of guidelines on fit-for-purpose RWD and defining standards for RWE studies intended to inform regulatory decisions.[19]

In conclusion, this study corroborates and extends the conclusions of the original pilot study that RWD can

TABLE 3. Median rwOS and 95% CI, Segmented by Treatment Setting and Patient Characteristics Comparing CAS With US Data Sets

Variable	CAS			Data Sets A-F	
	No. of Patients (N = 2,035)	No. of Events (n = 1,244)	Median OS (months) (95% CI)	No. of Patients (N = 269-6,924)	Median OS (months) (95% CI)
Age, years					
≤ 49	87	57	6.54 (4.47 to 9.03)	8-219	9.1-18.1 (2.7 to NA-11.9 to 21.6)
50-64	567	347	10.51 (9.3 to 13.67)	65-2,047	8.3-16.9 (6.8 to 11.4-9.4 to NA)
65-74	971	590	12.19 (10.71 to 13.6)	94-2,504	8.6-13.4 (6.7 to 12.3-12.1 to 14.9)
≥ 75	410	250	12.22 (9.86 to 14.09)	51-2,153	7-13.2 (5.0 to 15.7-11.8 to 14.6)
Sex					
Female	931	532	13.5 (11.79 to 15.51)	187-950	8.4-14.8 (7.3 to 8.9-13.2 to 16.9)
Male	1,104	712	9.95 (8.9 to 11.4)	143-3,172	7.9-13.2 (6.8 to 10.1-12.1 to 14.5)
Stage					
0 or I				9-498	5.7-12.1 (0.5 to 9.7-10.7 to 14.0)
II				13-426	6.4-13.9 (1.5 to NA-4.0 to NA)
III[a]	373	196	15.84 (13.54 to 18.4)	39-1,494	9.6-14.9 (6.7 to NA-9.4 to 22.1)
IV	1,662	1,048	10.51 (9.59 to 11.93)	161-4,334	8.3-12.1 (7.8 to 8.8-8.3 to 20.7)
Unknown				10-91	NA-7.9 (6.2 to 10.6)
Histology					
NSQ	1,448	867	11.83 (10.38 to 13.21)	194-4,678	9.6-14.2 (9.1 to 10.3-12.7 to 15.8)
SQ	521	343	10.51 (8.48 to 12.32)	61-1,983	6.8-13.9 (4.9 to 8.8-8.6 to NA)
NOS	24	14	10.18 (4.99 to NA)	10-220	10.3-20.4 (5.1 to 13.3-6.1 to NA)
Others specified	42	20	21.29 (13.83 to NA)		
LOT number of first PD-(L)1 inhibitor					
1	950	526	13.37 (11.86 to 15.57)	74-2,074	9.4-20.8 (6.4 to 12.2-14.8 to 25.1)
2	898	590	10.15 (8.94 to 11.83)	87-3,357	7.9-11.7 (6.5 to 10.5-10.9 to 12.8)
3	129	95	7.46 (6.01 to 10.55)	51-1,011	9-15.3 (7.8 to 9.9-9.6 to NA)
4+	33	25	4.8 (3.25 to NA)	34-481	5.1-14.2 (2.1 to 11.9-10.1 to 17.2)
Missing	25	8	NA (14.59 to NA)		

Abbreviations: CAS, Cancer Analysis System; LOT, line of therapy; NA, not applicable; NOS, not otherwise specified; NSQ, nonsquamous; OS, overall survival; PD-(L)1, programmed death-1 or programmed death ligand-1; rwOS, real-world overall survival; SQ, squamous.

[a]Only stage IIIB patients were included in the analysis of CAS data.

generate clinically meaningful and timely evidence on the efficacy of new cancer treatments used across diverse real-world settings. It further describes the usefulness of readily extractable end points, such as TTD, to assess clinical benefit.

Despite the variation in local clinical practice and data collection, there was considerable consistency in the findings between CAS and the US data sets. The observed differences in results could be largely explained by underlying differences in health care settings, including the timing of and variation in approvals and reimbursement, and data structure. This supports the premise that RWE can be informative for clinical, payer, policy, and regulatory decision making.

AFFILIATIONS
[1]IQVIA, London, United Kingdom
[2]IQVIA, Sofia, Bulgaria
[3]IQVIA, Durham, NC
[4]Friends of Cancer Research, Washington, DC
[5]Health Data Insight CIC, Cambridge, United Kingdom
[6]National Cancer Registration and Analysis Service, London, United Kingdom

CORRESPONDING AUTHOR
Pia Horvat, PhD, IQVIA, Real World Solutions, 210 Pentonville Rd, London N1 9JY, United Kingdom; e-mail: pia.horvat@iqvia.com.

AUTHOR CONTRIBUTIONS

Conception and design: Pia Horvat, Jennifer B. Christian, Laura Lasiter, Mark Stewart, Jeff Allen, Adam Reich
Provision of study materials or patients: Paul Clarke, Cong Chen
Collection and assembly of data: Paul Clarke, Cong Chen
Data analysis and interpretation: Pia Horvat, Christen M. Gray, Alexandrina Lambova, Jennifer B. Christian, Mark Stewart, Adam Reich
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Kim H-S, Lee S, Kim JH: Real-world evidence versus randomized controlled trial: Clinical research based on electronic medical records. J Korean Med Sci 33:e213, 2018

2. Makady A, Boer A, Hillege H, et al: What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health 20:858-865, 2017

3. Ramagopalan SV, Simpson A, Sammon C: Can real-world data really replace randomised clinical trials? BMC Med 18:13, 2020

4. Blonde L, Khunti K, Harris SB, et al: Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 35:1763-1774, 2018

5. Stewart M, Norden AD, Dreyer N: 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, 2019

6. Baumfeld Andre E, Reynolds R, Caubel P, et al: Trial designs using real-world data: The changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf 29:1201-1212, 2020

7. Cave A, Kurz X, Arlett P: Real-world data for regulatory decision making: Challenges and possible solutions for Europe. Clin Pharmacol Ther 106:36-39, 2019

8. US Food and Drug Administration: Framework for FDA's Real-Word Evidence Program. US Food and Drug Administration, 2018. https://www.fda.gov/media/120060/download.

9. Huang G, Sun X, Liu D: The efficacy and safety of anti-PD-1/PD-L1 antibody therapy versus docetaxel for pretreated advanced NSCLC: A meta-analysis. Oncotarget 9:4239-4248, 2017

10. Henson KE, Elliss-Brookes L, Coupland VH: Data resource profile: National Cancer Registration dataset in England. Int J Epidemiol 49:16-16h, 2019

11. Bright CJ, Lawton S, Benson S, et al: Data resource profile: The Systemic Anti-Cancer Therapy (SACT) dataset. Int J Epidemiol 49:15-15l, 2020

12. Soares M, Antunes L, Redondo P, et al: Real-world treatment patterns and survival outcomes for advanced non-small cell lung cancer in the pre-immunotherapy era in Portugal: A retrospective analysis from the IO optimise initiative. BMC Pulm Med 20:240, 2020

13. Curtis MD, Griffith SD, Tucker M: Development and validation of a high-quality composite real-world mortality endpoint. Health Serv Res 53:4460-4476, 2018

14. Dreyer NA, Hall M, Christian JB: Modernizing regulatory evidence with trials and real-world studies. Ther Innov Regul Sci 54:1112-1115, 2020

15. Velez MA, Burns TF: Is the game over for PD-1 inhibitors in *EGFR* mutant non-small cell lung cancer? Transl Lung Cancer Res 8:S339-S342, 2019 (suppl 4)

16. Dreyer NA: Advancing a framework for regulatory use of real-world evidence: When real is reliable. Ther Innov Regul Sci 52:362-368, 2018

17. Gray CM, Grimson F, Layton D, et al: A framework for methodological choice and evidence assessment for studies using external comparators from real-world data. Drug Saf 43:623-633, 2020

18. New quality and completeness checks on submissions of Systemic Anti-Cancer Therapies (SACT) data-in place June 2017. Public Health England, 2017. www.chemodataset.nhs.uk

19. Franklin JM, Liaw K-L, Iyasu S, et al: Real-world evidence to support regulatory decision making: New or expanded medical product indications. Pharmacoepidemiol Drug Saf 30:685-693, 2021

TABLE A1. Operational Definitions

Term	Definitions
End point	
rwOS	Length of time from the index date that the patient initiates treatment with a PD-(L)1 inhibitor to the date of death or end of follow-up, whichever occurred earliest. For claims data, health plan disenrollment dates are incorporated if deaths are not captured among those who leave health plan coverage. For CAS data, patients were censored at their last vital status date
rwTTD	Length of time from the index date that the patient initiates treatment with a PD-(L)1 inhibitor to the date that the patient discontinues the treatment. The study treatment discontinuation date was defined as the last administration or noncancelled order of a drug contained within the PD-(L)1 regimen. Discontinuation was defined as having a subsequent systemic therapy after the initial PD-(L)1–containing regimen, having a gap of more than 120 days with no systemic therapy after the last administration or having a date of death while on the PD-(L)1–containing regimen. Patients without a discontinuation were censored at their last known PD-(L)1 use. For CAS data, patients were censored at their last vital status date
Other elements	
Structured follow-up time	Length of time from the date that the patient initiates PD-(L)1 therapy or advanced diagnosis date for each patient until the last structured activity (ie, most recent visit or administration), unenrollment when relevant, death, or end of the follow-up period (ie, last structured activity)
LOT	LOT may be available from review of structured medication data, text fields, or other unstructured data from chart review. The first LOT was identified on the basis of the first date of receipt of any anticancer medication for treatment of aNSCLC. A treatment regimen was defined as the combination of anticancer medications that were received within the first 30 days of treatment with the first anticancer drug (in CAS, the first LOT includes all systemic anticancer medications given during the first 28 days). The second LOT was identified after a gap of 120 days or more in infusion or oral anticancer drug therapy or if the combination of drugs being received was changed (in CAS, second LOT was identified after a treatment gap of $\geq$ 70 days or after a change in treatment). Subsequent LOTs were defined similarly

Abbreviations: aNSCLC, advanced non–small-cell lung cancer; CAS, Cancer Analysis System; LOT, line of therapy; PD-(L)1, programmed death-1 or programmed death ligand-1; rwOS, real-world overall survival; rwTTD, real-world time to treatment discontinuation.

TABLE A2. Description of Demographic and Clinical Characteristics of Patients With aNSCLC Treated With PD-(L)1 Checkpoint Inhibitor

Patient Characteristic	CAS (N = 2,035)	Data Set A (n = 2,595)	Data Set B (n = 556)	Data Set C (n = 435)	Data Set D (n = 6,924)	Data Set E (n = 2,860)	Data Set F (n = 269)
Demographic characteristics							
Median age at advanced diagnosis, years (IQR)	67.00 (60.00-73.00)	68 (15)	64 (14)	66 (14)	69 (14)	68 (14)	70 (14)
Median age at PD-(L)1 inhibitor initiation, years (IQR)	68.29 (61.55-73.68)	69 (14)	65 (14)	68 (14)	69 (14)	69 (14)	71 (14)
Age categories at PD-(L)1 inhibitor initiation, years, No. (%)							
≤ 49	87 (4.28)	120 (5)	24 (4)	21 (5)	219 (3)	80 (3)	8 (3)
50-64	567 (27.86)	888 (34)	252 (45)	129 (30)	2,048 (30)	863 (30)	65 (24)
65-74	971 (47.71)	866 (33)	194 (35)	169 (39)	2,504 (36)	1,047 (37)	94 (35)
≥ 75	410 (20.15)	721 (28)	86 (15)	116 (27)	2,153 (31)	870 (30)	102 (38)
Age categories at PD-(L)1 inhibitor initiation, years, No. (%)							
< 75	1,625 (79.85)	1,874 (72)	470 (85)	319 (73)	4,771 (69)	1,990 (70)	167 (62)
≥ 75	410 (20.15)	721 (28)	86 (15)	116 (27)	2,153 (31)	870 (30)	102 (38)
Sex, No. (%)							
Female	931 (45.75)	1,147 (44)	275 (49)	212 (49)	3,172 (46)	1,351 (47)	125 (46)
Male	1,104 (54.25)	1,448 (56)	281 (51)	222 (51)	3,752 (54)	1,509 (53)	143 (53)
Unknown or missing	0	0	0	5	0	0	1
Race or ethnicity, No. (%)							
White	1,865 (91.65)	1,704 (78)	477 (86)	284 (65)	4,969 (79)	676 (87)	160 (87)
Black or African American	—	282 (13)	67 (12)	37 (9)	594 (9)	44 (6)	14 (8)
Asian	27 (1.33)	52 (2)	6 (1)	83 (19)	155 (3)	13 (2)	9 (5)
Others	42 (2.06)	142 (7)	6 (1)	31 (7)	580 (9)	42 (5)	1 (1)
Unknown or missing	67 (3.29)	415	0	0	626	2,085	85
Clinical characteristics							
Stage at initial diagnosis, No. (%)							
0 or occult			0		2 (0)		
I			23 (6)		496 (7)		18 (7)
II			22 (6)		426 (6)		17 (7)
III[a]	373 (18.33)		88 (23)	39 (9)	1,494 (22)		17 (7)
IV	1,662 (81.67)		248 (65)	396 (91)	4,335 (64)		161 (62)
Not reported			175		171		10
Histology, No. (%)							
Nonsquamous NSCLC	1,448 (71.15)		369 (66)	320 (74)	4,679 (70)	1,981 (69)	194 (73)
Squamous cell carcinoma	521 (25.06)		147 (26)	73 (17)	1,983 (30)	659 (23)	61 (23)
NSCLC NOS	24 (1.18)		40 (7)	42 (10)	262 (3)	220 (8)	10 (4)
Others specified	42 (2.06)		—	—	—	—	
Unknown or missing							4
Line of first PD-(L)1 inhibitor, No. (%)							
1 (no prior therapy received)	950 (46.68)	690 (27)	144 (26)	80 (18)	2,074 (30)	777 (27)	77 (29)
2	898 (44.13)	1,440 (56)	272 (49)	205 (47)	3,357 (49)	1,414 (49)	87 (32)
3	129 (6.34)	380 (15)	96 (17)	85 (20)	1,012 (15)	448 (16)	51 (19)

(continued on following page)

TABLE A2. Description of Demographic and Clinical Characteristics of Patients With aNSCLC Treated With PD-(L)1 Checkpoint Inhibitor (continued)

Patient Characteristic	CAS (N = 2,035)	Data Set A (n = 2,595)	Data Set B (n = 556)	Data Set C (n = 435)	Data Set D (n = 6,924)	Data Set E (n = 2,860)	Data Set F (n = 269)
≥ 4	33 (1.62)	85 (3)	44 (8)	65 (15)	481 (7)	221 (8)	54 (20)
Missing or unknown	25 (1.23)						
Follow-up time, months, median (Q1, Q3)							
From advanced diagnosis to PD-(L)1 inhibitor initiation	4.86 (1.61, 11.70)		7 (3, 14)	8 (4, 15)	6 (2, 13)	8 (3, 17)	7 (2, 14)
From advanced diagnosis	16.79 (10.61, 26.05)		18 (10, 28)	18 (10, 31)	14 (8, 25)	18 (10, 30)	18 (10, 28)
From PD-(L)1 inhibitor initiation (any LOT)	10.61 (3.91, 16.00)		8 (3, 16)	9 (3, 16)	6 (2, 12)	8 (3, 14)	8 (4, 13)

Abbreviations: aNSCLC, advanced non–small-cell lung cancer; CAS, Cancer Analysis System; IQR, interquartile range; LOT, line of therapy; NA, not applicable; NOS, not otherwise specified; NSCLC, non–small-cell lung cancer; PD-(L)1, programmed death-1 or programmed death ligand-1; Q1, quartile 1; Q3, quartile 3.

[a]Only stage IIIB patients were included in the analysis of CAS data.

TABLE A3. Median rwOS and 95% CI, Segmented by Treatment Setting and Patient Characteristics

Variable	CAS		Data Set A		Data Set B		Data Set C		Data Set D		Data Set E		Data Set F	
Age, years														
≤ 49	57	6.5 (4.5 to 9.0)	100	18.1 (11.9 to 21.6)	16	9.1 (2.7 to NA)	21	12.0 (4.3 to NA)	219	9.3 (7.8 to 12.1)	80	10.2 (7.9 to 17.4)	8	NA
50-64	347	10.5 (9.3 to 13.7)	723	13.6 (11.8 to 14.6)	164	8.3 (6.8 to 11.4)	129	9.3 (6.7 to 13.3)	2,047	9.3 (8.4 to 10.3)	863	13.8 (11.8 to 15.3)	65	16.9 (9.4 to NA)
65-74	590	12.2 (10.7 to 13.6)	728	13.4 (12.1 to 14.9)	131	8.6 (6.7 to 12.3)	169	8.9 (5.8 to 11.1)	2,504	9.3 (8.8 to 10.3)	1,047	12.2 (10.5 to 14.3)	94	12.3 (7.6 to 21.6)
≥ 75	250	12.2 (9.9 to 14.1)	593	13.2 (11.8 to 14.6)	51	7.0 (5.0 to 15.7)	116	6.8 (4.2 to 9.1)	2,153	8.8 (8.2 to 9.3)	870	13 (10.6 to 14.8)	102	10 (8.7 to 15.6)
Sex														
Female	532	13.5 (11.8 to 15.5)	950	13.7 (12.8 to 15.2)	187	8.9 (6.7 to 12.3)	212	9.3 (7.4 to 13.4)		8.4 (7.3 to 8.9)		14.8 (13.2 to 16.9)		13.2 (8.8 to 23.3)
Male	712	10.0 (8.9 to 11.4)	1,194	13.2 (12.1 to 14.5)	175	7.9 (6.8 to 10.1)	222	7.5 (6.4 to 10.0)	3,172	9.8 (9.4 to 10.7)	1,509	11.1 (10.2 to 12.4)	143	11.9 (8.4 to 20.4)
Stage														
0 or I					9	5.7 (0.5 to 9.7)			498	12.1 (10.7 to 14.0)			18	7 (4.9 to NA)
II					13	6.4 (1.5 to NA)			426	11.8 (10.6 to 13.3)			17	13.9 (4.0 to NA)
III[a]					62	9.6 (6.7 to NA)	39	8.9 (2.7 to 13.4)	1,494	9.8 (9.2 to 10.8)			63	14.9 (9.4 to 22.1)
IV					187	8.8 (6.8 to 10.6)	396	8.7 (6.8 to 10.0)	4,334	8.3 (7.8 to 8.8)			161	12.1 (8.3 to 20.7)
Unknown					91	7.9 (6.2 to 10.6)							10	NA
Histology														
NSQ	867	11.8 (10.4 to 13.2)			239	9.7 (7.7 to 12.6)	320	8.7 (6.7 to 10.2)	4,678	9.6 (9.1 to 10.3)	1,981	14.2 (12.7 to 15.8)	194	11.9 (8.8 to 21.8)
SQ	343	10.5 (8.5 to 12.3)			93	6.8 (4.9 to 8.8)	73	8.4 (4.8 to 12.1)	1,983	8.7 (7.8 to 9.3)	659	10.4 (9.2 to 11.8)	61	13.9 (8.6 to NA)
NOS	14	10.2 (5.0 to NA)			30	10.3 (5.1 to 13.3)	42	7.9 (3.6 to 17.2)			220	11.8 (7.9 to 16.1)	10	20.4 (6.1 to NA)
Others specified	20	21.3 (13.8 to NA)												
LOT number of first PD-(L)1 inhibitor														
1	526	13.4 (11.9 to 15.6)	592	19.8 (17.2 to 22.2)	74	9.4 (6.4 to 12.2)	80	9.2 (5.7 to 17.1)	2,074	10.4 (9.5 to 11.2)	777	20.8 (14.8 to 25.1)	77	15.9 (9.9 to NA)
2	590	10.2 (8.9 to 11.8)	1,174	11.7 (10.9 to 12.8)	191	7.9 (6.5 to 10.5)	205	7.8 (5.8 to 10.3)	3,357	8.7 (8.1 to 9.2)	1,414	10.7 (9.8 to 11.9)	87	9.1 (7.4 to 20.4)
3	95	7.5 (6.0 to 10.6)	304	12.8 (10.7 to 14.7)	63	10.7 (6.9 to 17.5)	85	9.3 (6.0 to 12.0)	1,011	9 (7.8 to 9.9)	448	10.7 (9.0 to 13.9)	51	15.3 (9.6 to NA)
4+	25	4.8 (3.3 to NA)	74	14.2 (10.1 to 17.2)	34	5.1 (2.1 to 11.9)	65	8.7 (5.3 to 13.3)	481	8.5 (6.9 to 10.5)	221	12 (8.3 to 15.6)	54	10.4 (6.9 to 21.8)
Missing	8	NA (14.6 to NA)												

Abbreviations: CAS, Cancer Analysis System; LOT, line of therapy; NA, not applicable; NOS, not otherwise specified; NSQ, nonsquamous; OS, overall survival; PD-(L)1, programmed death-1 or programmed death ligand-1; rwOS, real-world overall survival; SQ, squamous.

[a]Only stage IIIB patients were included in the analysis of CAS data.

TABLE A4. LOT for Age Group Below 50 Years

Variable	CAS, No. (%)
Overall	87 (100)
LOT number of first PD-(L)1 inhibitor (index LOT)	
1	30 (34.48)
2	45 (51.72)
3	—
4+	6 (6.90)
Missing or unknown	—

Abbreviations: CAS, Cancer Analysis System; LOT, line of therapy; PD-(L)1, programmed death-1 or programmed death ligand.

TABLE A5. Correlations Between rwOS and rwTTD in CAS Compared With US Data Sets Using Spearman's Rank Correlation Coefficient

Data Set	Comparison	No.	Correlation (95% CI)
CAS	rwOS *v* rwTTD	920	0.73 (0.69 to 0.76)
A	rwOS *v* rwTTD	254	0.63 (0.55 to 0.70)
B	rwOS *v* rwTTD	254	0.62 (0.54 to 0.69)
C	rwOS *v* rwTTD	295	0.89 (0.86 to 0.91)
D	rwOS *v* rwTTD	4,337	0.80 (0.79 to 0.81)
E	rwOS *v* rwTTD	1,456	0.77 (0.75 to 0.79)
F	rwOS *v* rwTTD	142	0.80 (0.66 to 0.85)

NOTE. Correlation analysis was restricted to those patients who had experienced both death and treatment discontinuation.

Abbreviations: CAS, Cancer Analysis System; rwOS, real-world overall survival; rwTTD, real-world time to treatment discontinuation.

COVID-19 Evidence Accelerator: A parallel analysis to describe the use of Hydroxychloroquine with or without Azithromycin among hospitalized COVID-19 patients

Mark Stewart[1‡], Carla Rodriguez-Watson[2‡], Adem Albayrak[3☺], Julius Asubonteng[4☺], Andrew Belli[5☺], Thomas Brown[6☺], Kelly Cho[7,8☺], Ritankar Das[9☺], Elizabeth Eldridge[3☺], Nicolle Gatto[10☺], Alice Gelman[3☺], Hanna Gerlovin[7☺], Stuart L. Goldberg[11☺], Eric Hansen[5☺], Jonathan Hirsch[6☺], Yuk-Lam Ho[7☺], Andrew Ip[11☺], Monika Izano[6☺], Jason Jones[3☺], Amy C. Justice[12,13☺], Reyna Klesh[14☺], Seth Kuranz[15☺], Carson Lam[9☺], Qingqing Mao[9☺], Samson Mataraso[9☺], Robertino Mera[4☺], Daniel C. Posner[7☺], Jeremy A. Rassen[10☺], Anna Siefkas[9☺], Andrew Schrag[6☺], Georgia Tourassi[16☺], Andrew Weckstein[10☺], Frank Wolf[6☺], Amar Bhat[2☺], Susan Winckler[2☺], Ellen V. Sigal[1,2☺], Jeff Allen[1] *

1 Friends of Cancer Research, Washington, District of Columbia, United States of America, **2** Reagan-Udall Foundation for the FDA, Washington, District of Columbia, United States of America, **3** Health Catalyst, Salt Lake City, Utah, United States of America, **4** Gilead Science, Inc. Foster City, California, United States of America, **5** COTA, Inc., Boston, Massachusetts, United States of America, **6** Syapse, San Francisco, California, United States of America, **7** Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, United States of America, **8** Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **9** Dascena, Oakland, California, United States of America, **10** Aetion, New York, New York, United States of America, **11** Division of Outcomes and Value Research, John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, New Jersey, United States of America, **12** VA Connecticut Healthcare System, West Haven, Connecticut, United States of America, **13** Yale University Schools of Medicine and Public Health, New Haven, Connecticut, United States of America, **14** HealthVerity, Philadelphia, Pennsylvania, United States of America, **15** TriNetX, Cambridge, Massachusetts, United States of America, **16** National Center for Computational Sciences Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America

☺ These authors contributed equally to this work.
‡ TCo-lead authors.
* jallen@focr.org

Abstract

Background

The COVID-19 pandemic remains a significant global threat. However, despite urgent need, there remains uncertainty surrounding best practices for pharmaceutical interventions to treat COVID-19. In particular, conflicting evidence has emerged surrounding the use of hydroxychloroquine and azithromycin, alone or in combination, for COVID-19. The COVID-19 Evidence Accelerator convened by the Reagan-Udall Foundation for the FDA, in collaboration with Friends of Cancer Research, assembled experts from the health systems

Methods

Electronic health record (EHR) and claims data were extracted from seven separate databases. Parallel analyses were undertaken on data extracted from each source. Each analysis examined time to mortality in hospitalized patients treated with hydroxychloroquine, azithromycin, and the two in combination as compared to patients not treated with either drug. Cox proportional hazards models were used, and propensity score methods were undertaken to adjust for confounding. Frequencies of adverse events in each treatment group were also examined.

Results

Neither hydroxychloroquine nor azithromycin, alone or in combination, were significantly associated with time to mortality among hospitalized COVID-19 patients. No treatment groups appeared to have an elevated risk of adverse events.

Conclusion

Administration of hydroxychloroquine, azithromycin, and their combination appeared to have no effect on time to mortality in hospitalized COVID-19 patients. Continued research is needed to clarify best practices surrounding treatment of COVID-19.

Background

Despite a growing body of literature about COVID-19 and its cause, SARS-CoV-2, much remains unclear about which treatment strategies are most effective for the entire clinical course of the disease. Several therapeutic agents have been investigated, including the anti-malarial drug hydroxychloroquine. The evidence regarding its use for the treatment of COVID-19 continues to evolve [1, 2].

The use of hydroxychloroquine for COVID-19 was initially supported by *in vitro* studies showing anti-inflammatory and anti-SARS-CoV-2 activity [3]. Methodological weaknesses have marked subsequent *in vivo* studies; many studies have enrolled small numbers of patients and have not included appropriate control groups or methods to control for confounding variables, making interpretation of findings difficult [4]. While some non-randomized studies have shown a survival benefit for COVID-19 patients receiving hydroxychloroquine [5], others have found no evidence of benefit [6]. Still others have identified safety concerns, including an increased risk of prolonged QT intervals and arrhythmias for patients receiving hydroxychloroquine [7–9]. Additional studies have found that hydroxychloroquine may be more effective when given in combination with azithromycin [5] while emerging experience may indicate otherwise, leading to further uncertainty about the appropriate use of hydroxychloroquine. Differing methodologies to control for potential bias, incomplete capture of the timing of mechanical ventilation in relation to receipt of hydroxychloroquine, may have contributed to these inconsistencies.

Despite the development of treatment guidelines for COVID-19 [10], significant questions remain about best treatment practices. COVID-19 remains a significant global threat, and

epidemiologic models have predicted that transmission will continue through the coming years [11]. Due to the significant morbidity and mortality associated with severe cases of COVID-19, establishing treatment guidelines is an essential step towards improving patient outcomes. It is therefore important to address methodological inconsistencies of existing studies and remaining uncertainties about the efficacy of hydroxychloroquine for the treatment of COVID-19.

Towards this end, the COVID-19 Evidence Accelerator convened by the Reagan-Udall Foundation for the FDA, in collaboration with Friends of Cancer Research, assembles experts from the health systems research, regulatory science, data science, and epidemiology to participate in parallel analyses. Analytic partners align on a common protocol and conduct analyses independently; methods and results are shared side-by-side to evaluate differences and similarities. Results are presented to a larger audience, including experts and leaders from the FDA, to provide informal discussion and review. Several groups, representing distinct populations within the U.S. to conduct parallel analyses of the effect of hydroxychloroquine, azithromycin, and the two drugs in combination on COVID-19 outcomes to compare results and better understand differences in the safety of these treatments for COVID-19.

Methods

Ethical statement

All the data partners received Institutional Review Board (IRB) approval or exemption. The use of VA data was approved by both the Department of Energy (DOE) (Oak Ridge Sitewide IRB00000547 for Protocol ORAU000718) and VA review committees and engages both VA and DOE researchers (VA-DOE Reliance Agreement under the authority of 38 U.S.C. 7303 and 38 U.S.C. 523). In addition to IRB approval, VA R&DC reviewed research proposals for final Institutional approval and ensured that all research in which the facility is engaged is consistent with the VA mission and complies with all applicable statutory and regulatory requirements. A Waiver of HIPAA Authorization and a Waiver of Informed Consent were approved for this study ORAU000718. The Aetion/HealthVerity study was approved under exemption by the New England Institutional Review Board (protocol #1-9757-1) and received a waiver of informed consent. The COTA study received approval by the Hackensack Meridian Health IRB (Pro2020-0342) and received a waiver of informed consent. The Health Catalyst dataset used for this analysis has been de-identified following the expert determination method outlined in 45 CFR 164.514(b)(1). Health Catalyst uses an external vendor to certify that the dataset is de-identified in accordance with 45 CFR 164.514(b)(1). The Dascena study received approval from the Pearl Institutional Review Board (20-DASC-120) and was granted a waiver of informed consent. The TriNetX Platform receives Protected Healthcare Information (PHI) or a Limited Data Set (LDS) from Healthcare Organizations (HCO) strictly under the constraints defined in a Business Associate Agreement (BAA) or a Data Use Agreement (DUA) under the United States (U.S.) Health Insurance Portability and Accountability Act (HIPAA). A fundamental Data Privacy principle is that TriNetX does not expose PHI or LDS to the end users of the TriNetX Platform. The data made available from the TriNetX platform is de-identified based on standard defined in Section §164.514(a) of the HIPAA Privacy Rule. The process by which Data Sets are de-identified is attested to through a formal determination by a qualified expert as defined in Section §164.514(b)(1) of the HIPAA Privacy Rule. Sypase conducted this work through a Research Collaboration Agreement with the FDA to include an IRB exemption through the Office of the Chief Scientist (OCS) Human Subject Protection (HSP) Executive Officer and all of this work involved data from secondary sources. The RCA

work has been performed under an exemption from the Office of the Chief Scientist (OCS) Human Subject Protection (HSP) Executive Officer at FDA.

Data sources

The Evidence Accelerator partnered with seven groups to conduct the parallel analyses: Syapse, COTA/Hackensack Meridian Health (HMH), Dascena, TriNetX, Health Catalyst, Aetion, and Veteran's Health Administration (VA). Each group conducting the parallel analysis collected data from their distinct sources. Syapse, COTA, Dascena, TriNetX, Health Catalyst and VA all utilized electronic health record (EHR) data, while Aetion utilized medical and pharmacy claims, and hospital billing data drawn from the HealthVerity Marketplace. Syapse utilized the EHR and molecular diagnostic lab information from two large Midwestern US health systems. COTA utilized data from the Real-world Evidence COVid RegistrY (RE-COV-RY) database collected at Hackensack Meridian Health System. Dascena utilized data from the EHRs from eight US hospitals. TriNetX drew data from the TriNetX Dataworks USA Network. Health Catalyst drew data from 17 Health Catalyst clients; the group had access to EHR data including medication administration. Aetion drew data from the HealthVerity linked medical and pharmacy claims, labs, and hospital chargemaster dataset. The VA used EHR data from the national VA Healthcare System, with COVID-19 cases adjudicated through a National Surveillance Tool [12]. A graphical description of coverage and overlap is illustrated in Fig 1 and characteristics of participating data sources and populations is described in Table 1.

Patient inclusions

Data were gathered from the 7 EHR datasets for hospital admissions in the U.S. between January 1, 2020 and June 30, 2020 (index hospitalization). Patients were eligible for inclusion if they had tested positive for COVID-19 during or prior to their visit or had an International Classification of Diseases (ICD)-10 code for COVID-19 in the 21 days leading up to admission, during admission or as a discharge diagnosis (Fig 2). All groups using ICD codes considered ICD-10 code U07.1; Syapse, Dascena and COTA additionally considered codes B97.21, B97.29, J12.81, B34.2; for the primary results Health Catalyst required a discharge diagnosis of U07.1 either primary or secondary to a related primary diagnoses (e.g., pneumonia or acute
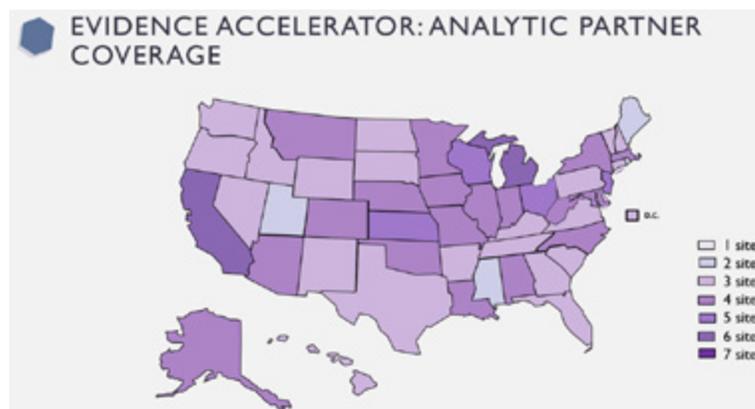
Fig 1. Partner map for HCQ analysis. The Evidence Accelerator partnered with seven groups to conduct the parallel analyses. This is a graphical description of coverage and overlap for each group conducting the parallel analysis and their distinct sources. Republished from https://www.brightcarbon.com/resources/editable-powerpoint-maps/ under a CC BY license, with permission from Bright Carbon, original copyright 2020.

**Table 1. Characteristics of participating data sources and populations.**

Study Variable	Aetion	COTA	Dascena	Health Catalyst	Syapse	TriNetX	VA Dataset
Data source	HealthVerity Medical and Pharmacy Claims and Hospital Chargemaster	Hackensack Meridian Health System; HMH Real-world Evidence COVid-RegistrY (RE-COV-RY)	Electronic health records from eight US hospitals	17 Health Catalyst clients; No claims, pharmacy data available; EMR data: medication administrations, prescriptions, labs, diagnoses	Integrated data from health systems (e.g. electronic medical records, ancillary clinical sources) and molecular diagnostic labs	TriNetX Dataworks USA Network	Veteran's affairs healthcare electronic health records. Covers all VA-hospitals with over 9 million active users.
Region	Coverage across the U.S. Majority (57%) in the Northeast Region.	Northeast Region; New Jersey		Across the U.S. with coverage in 24 states	Midwest	Across the Northeastern, Southern, Midwestern, and Western regions of the US.	Nationwide sample representative of COVID pandemic geographic distribution as of study time period.
Data processing	Open claims data from data clearing houses. Chargemaster data from hospital billing system (with details of hospital stays). Lab data from a major lab provider (Quest) plus several smaller providers. The data is brought together by the HealthVerity de-id engine through a common patient ID.	EHR data abstracted through chart review into REDCap		Data from multiple EMRs standardized into single data model; additional regular expression and curation performed to identify COVID labs	Data from clinical systems at multiple health systems and molecular data from diagnostic labs integrated into a common data model	Various and disparate data is mapped to industry standard terminologies which produces master terminology/ intelligent synonym search.	Corporate data warehouse (CDW) repository of EHR for entire VA. Consists of inpatient and outpatient diagnosis codes, laboratory results, pharmacy prescription fills (inpatient, outpatient, and IV), vitals, health factors, and demographic data.
Data collection period	February 2020-April 2020 with follow up through May 2020	March 2020—May 2020	March 1, 2020-June 20, 2020	Mar 2020—May 2020	February 1 –June 13, 2020	All patients were diagnosed with COVID-19 and hospitalized between January 20, 2020 and April 27th 2020.	Case-positive individuals as of April 30, 2020 followed through EHR data as of July 14, 2020.

respiratory distress syndrome). To ensure accurate assessment of comorbidities, Syapse imposed a one-year minimum enrollment criteria to their study population of patients diagnosed with malignant cancers on or after January 1, 2015.

All patient data was maintained in compliance with the Health Insurance Portability and Accountability Act (HIPAA).

Treatment

Patients were considered to be treated with hydroxychloroquine and/or azithromycin if they received any of those medications at any point during their hospitalization, before discharge or death (the VA only considered treatments that occurred in the first 48 hours following hospitalization; Health Catalyst required treatment initiation within the first 2 days following hospitalization). Three treatment groups were compared to the population of patients that received neither hydroxychloroquine nor azithromycin. These treatment groups were hydroxychloroquine and azithromycin in combination, hydroxychloroquine alone, and azithromycin alone. Groups varied in their approach to determining date of cohort entry and index date of treatment (see S1 File for details). This study was non-interventional and treatment groups were not randomized.
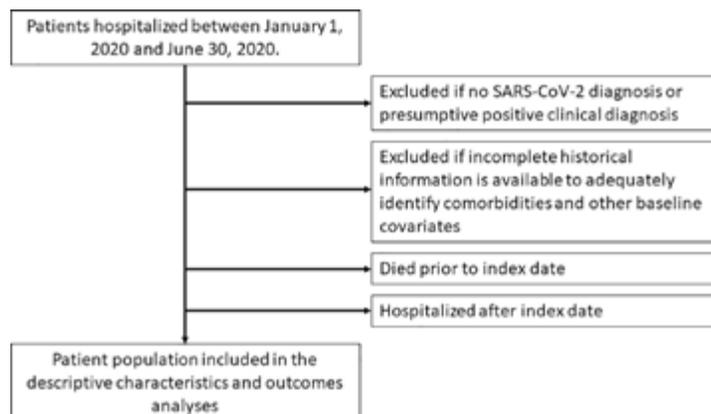
Fig 2. Study diagram. Common exclusion criteria were applied across the datasets to arrive at a final study population to assess patient characteristics and treatment outcomes for patients with COVID-19.

Covariates

For each patient, data on potential demographic, medication and health-related covariates were extracted from structured EHR fields, prescription dates, and ICD-10 codes. In one analysis data were also extracted from structured and unstructured hospital billing data. In two networks, this data was augmented with claims data. A subset of participating groups adjusted for health-related variables and medication use prior to the index hospitalization, sociodemographic factors and comorbid conditions. The ICD codes used to identify comorbidities are presented in the S1 File. Covariates were selected independently by each group. Covariates considered by each group, and the method of their selection, are presented in Table 2.

Outcomes

Three primary outcomes were measured: use of mechanical ventilations (as a potential indicator of overall health status of patients), evidence of benefit of hydroxychloroquine (determined by hospital discharge as an indicator of recovery), and in-hospital mortality (determined by discharge disposition). Measurement of overall health status and evidence of benefit varies across groups, contingent on data availability. A summary of outcome definitions is presented in Table 3.

We also assessed the proportion of patients in each treatment group experiencing any of the following adverse events: diarrhea, hypoglycemia, cardiac arrest, abnormal electrocardiogram (ECG), arrhythmia, or prolonged QT interval. Adverse event data were not provided by Health Catalyst or the VA.

Analytic methods

Study follow-up began at slightly different time points within each group's defined study design (S1 File). To improve baseline balance and minimize immortal time bias, in Aetion's analyses, untreated were matched (on a number of key characteristics including calendar time and time since hospital admission) to treated patients on the day of first administration of HCQ (using risk set sampling). The index date for HCQ treated was the first day of treatment and the index date for the HCQ untreated was defined by the index date of the matched HCQ

Table 2. Covariates reported by each group.

	Health-related Variables	Medication Usage	Sociodemographic variables	Confounder selection method
Aetion	• Baseline health status: chronic comorbidities, lifestyle factors, health resource utilization • Pre-admission confounders related to COVID-19 severity: pre-admission symptoms, pre-admission health resource utilization, no. days since symptom onset • Admitting characteristics: Hospital characteristics (e.g. urban vs rural, no. of beds, teaching status), admitting status, admitting diagnoses • In-hospital confounders: COVID-19 severity at treatment, trajectory in severity, respiratory support and procedural treatments, ICU utilization	• Baseline medication use: chronic medication use, no. of unique medications dispensed, no. prescriptions dispensed • COVID-19-related medications: pre-admission outpatient treatments, inpatient antithrombotics, inpatient antivirals/antibiotics, other experimental COVID-19 therapies administered prior to or concurrent with treatment index date, time from admission to treatment	• Age • Sex • Calendar month of cohort entry • Insurance type • US region	A priori assumptions about confounding structure. and prior literature
COTA Hackensack Meridian Health	• Smoking history • Number of comorbidities at baseline • Comorbid diagnoses • ICU status • Fever • Respiratory rate • C-reactive protein • Oxygenation status • qSOFA	Insulin	• Age • Sex • Race • Nursing home status prior to admission	Lasso regression using 5-fold cross validation, with priority given to variables significant in determining the outcome of interest
Dascena	• Vital signs and lab values at admission (oxygen saturation, D-Dimer, lactate, temperature, white blood cell count, respiratory rate, heart rate, and systolic blood pressure) • Comorbid diagnoses	remdesivir, macrolide antibiotics, angiotensin receptor blockers (ARB), angiotensin-converting enzyme (ACE) inhibitors, non-steroidal anti-inflammatory drugs (NSAID), steroids, tocilizumab, and statins	• Age • Sex • Race/ethnicity • Income	A priori assumptions about confounding structure and prior literature
Health Catalyst	• Chronic comorbid diagnoses • History of supplemental oxygen use • History with health-related behaviors (e.g., smoking)	• remdesivir, macrolide antibiotics, angiotensin receptor blockers (ARB), angiotensin-converting enzyme (ACE) inhibitors, non-steroidal anti-inflammatory drugs (NSAID), steroids, tocilizumab, and statins	• Age • Sex • Race/ethnicity • Income	A priori assumptions about confounding structure and prior literature coupled with empiric interrogation.
Syapse	• Chronic Comorbidities Clinical characteristics at hospital admission	Not Reported	• Age • Sex • Race/ethnicity • Estimated Income	A priori assumptions about confounding structure and prior literature.
TriNetX	• Comorbid diagnoses • Oxygenation status • Baseline invasive	• angiotensin-converting enzyme (ACE) • Angiotensin receptor blockers	• Age • Sex • Race • Ethnicity	A priori assumptions about confounding structure and prior literature
VA	• Lab orders (Lactate dehydrogenase, C-reactive protein. D-dimer, Ferritin) • Height and weight • Smoking status and alcohol use • Concurrent inpatient treatments • Chronic comorbidities • Frailty • Lab results and vital signs	• Chronic medication use • Concurrent inpatient treatments	• Age • Sex • Urbanicity • Region of US • Long-term care status • Calendar week of admission • Station size or number of veterans in care	A priori assumptions about confounding structure and prior literature.

treated patient (for further detail see S1 File). For the VA, the index date was assigned to be 48 hours after hospital admission to avoid immortal time bias. For the other 4 datasets the index date was the date of admission.

Table 3. Summary of outcome definitions across groups.

Group	Mechanical Ventilation	Evidence of Benefit
Aetion	Mechanical ventilation	Not assessed
COTA Hackensack Meridian Health	Mechanical Ventilation	Hospital Discharge
Dascena	Mechanical Ventilation	Hospital Discharge
Health Catalyst	Not assessed	Not assessed
Syapse	Mechanical Ventilation	Not assessed
TriNetX	Not assessed	Improvement from "hospitalized with any oxygen support" to either "hospitalized on room air" or "discharge" following the index date
VA	Mechanical Ventilation within 21 days	Not assessed

Six groups used time-to-event analyses. The mortality outcome was primarily evaluated as in-hospital, for five of the groups, while the VA considered time to all-cause mortality within 30 days of the index date. Syapse examined the cumulative incidence of all-cause mortality during or after the index hospitalization. Health Catalyst conducted a primary analysis using mortality as time-to-event and sensitivity analysis treating mortality as binary.

Statistical analysis

To examine the association between potential confounders and treatment with hydroxychloroquine and azithromycin, hydroxychloroquine alone, and azithromycin alone, we compared the distributions of each covariate in each treatment group and in the group receiving no treatment. In addition, we examined the distribution of adverse events across treatment groups.

Methods to assess the association between treatment and outcomes included logistic regression, competing risk analyses, and propensity score methods. Dascena employed Fine and Gray models for the subdistribution hazard ratio (HR) were used to examine the association between treatment and each of the outcome measures. This method allows for estimation of the incidence of events, despite the presence of a competing event that precludes the observation of the event of interest. Incidence was estimated using Breslow's estimator. All individuals who had not experienced the event were censored at the end of the study period. Additional information about statistical methods used by each group can be found in the S1 File.

To adjust for baseline confounding variables, a subset of groups employed propensity score methods. Logistic regression was used by five groups to predict the probability of treatment with hydroxychloroquine, azithromycin, or both in the study population, conditional on all measured confounders. The VA estimated propensity scores using a gradient boosting machine, implemented using the packages 'gbm' and 'WeightIt' in R [13–15]. Aetion performed all analyses in the Aetion Evidence Platform v4.5. Propensity scores were then used to adjust for confounding either through inverse probability of treatment weighting (IPTW), propensity score matching, or adjustment on the propensity score. Health Catalyst conducted sensitivity analyses using unmatched, 1:1 matched, propensity score matched, propensity score adjusted, and propensity score binned techniques. Details of the adjustment methods used by each group are presented in Table 4 and in the S1 File.

In each dataset, the association between each treatment (HCQ, HCQ+AZ) and each outcome was assessed in comparison to the non HCQ/AZ group (no treatment, "neither"). Therefore, treatment groups were not directly compared to each other in this analysis. For all analyses, a two-sided alpha of 0.05 without adjustment for multiple comparisons was used to determine statistical significance.

Table 4. Key definitions and methodology.

Study Variable	Aetion	COTA	Dascena	Health Catalyst	Syapse	TriNetX	VA Dataset
Inclusion Criteria	Hospitalized patients defined as having any of the suspected COVID-19 criteria occurring in the 21 days prior to (and including) the hospital admission date (cohort entry date) OR in the discharge diagnosis	Hospitalized patients defined having positive SARS-CoV-2 diagnosis by RT-PCR	Hospitalized patients (defined as length of stay > 24 hours) defined as having a positive COVID-19 PCR test or diagnosis within five days of encounter	A discharge diagnosis of COVID 19 (ICD-10:U07.1) either primary or secondary to a specific list of other conditions	Hospitalized patients with malignant cancer (diagnosed in the last 5 years) at two health systems with confirmed COVID-19 diagnosis (via positive lab result and/or ICD code)	Hospitalized patients identified using coronavirus codes used in EMRs for COVID-19. Any Code must be present Jan. 20, 2020 or after to yield patients. Inpatient code required 2 weeks before or anytime after COVID-19	Hospitalized on or after first positive Sars-CoV-2 test. Treatment assignment groups were specified by the intent-to-treat design where HCQ/Azith/ Both was/were initiated in the first 48-hours following admission. Comparison groups (Az, Neither) included individuals from hospitals where at least one person was prescribed HCQ.
COVID-19 Definition	Medical claim or chargemaster event with COVID-19-like diagnosis or Positive or presumptive positive viral lab test result	Positive SARS-CoV-2 diagnosis		ICD-10 diagnosis code		LOINC or ICD-10 diagnosis code	VA cases were based on the National Surveillance Tool classification following NLP and adjudication methodologies described(12)
COVID-19 Diagnosis date	Earliest date of confirmed COVID-19 recorded in 21 days pre-admission (inclusive) or admission date if diagnosis derived from discharge diagnosis	Date of confirmed COVID-19 diagnosis via PCR lab result		Earliest positive PCR lab collection data or clinical diagnosis of COVID-19	Date of confirmed COVID-19 diagnosis via ICD code or positive lab result. For patients with confirmation of COVID-19 diagnosis via ICD and lab test result, date of lab result is used	Minimum of positive PCR lab confirmed date or clinical diagnosis of COVID-19	All individuals included were required to have a case-defined diagnosis date prior to, or on the same day as, hospital admission.
Index Date	Initiation of treatment for HCQ + patients; matched controls (HCQ-) were assigned an index date of their matched HCQ treated patient.	Date of hospital admission		Date of hospital admission		Date of hospital admission	Date of hospitalization + 48 hours
Measure of Overall Health of Patients	Mechanical Ventilation (used as a control outcome to refine comparative approach for future drug evaluation studies)	Mechanical Ventilation	Mechanical Ventilation	NA	Mechanical Ventilation	The rate of recovery, defined as an improvement from hospitalized with any oxygen support to either hospitalized on room air or discharge following the index date	Mechanical Ventilation

(*Continued*)

Table 4. (Continued)

| Study Variable	Aetion	COTA	Dascena	Health Catalyst	Syapse	TriNetX	VA Dataset
Measure of Evidence of Benefit	NA	Hospital discharge	Hospital discharge	NA	NA	The rate of recovery, defined as an improvement from hospitalized with any oxygen support to either hospitalized on room air or discharge following the index date	NA
Adjusted Analysis Approach	RSS+PS; Risk set sampling of HCQ untreated patients at time of HCQ administration in HCQ+, followed by propensity score matching based on key patient demographics and clinical characteristics	Propensity score matching and adjusting based on key patient demographics and clinical characteristics	Inverse probability weight adjusted	5-bin propensity stratified analysis	Not conducted. Only reported crude estimates	Inverse probability weight adjusted	Gradient boosted tree models for estimating weights to use in sIPTW with Cox proportional hazards

Results

Population

In total, 20,371 patient encounters from seven data sources were analyzed. Demographic characteristics of each dataset are laid out in Figs 3–5, with patient comorbidities presented in

Fig 3. Age distribution by treatment group. Categorical age for each treatment group shows similarities and differences across data sets. Note. Numbers represent percent values.
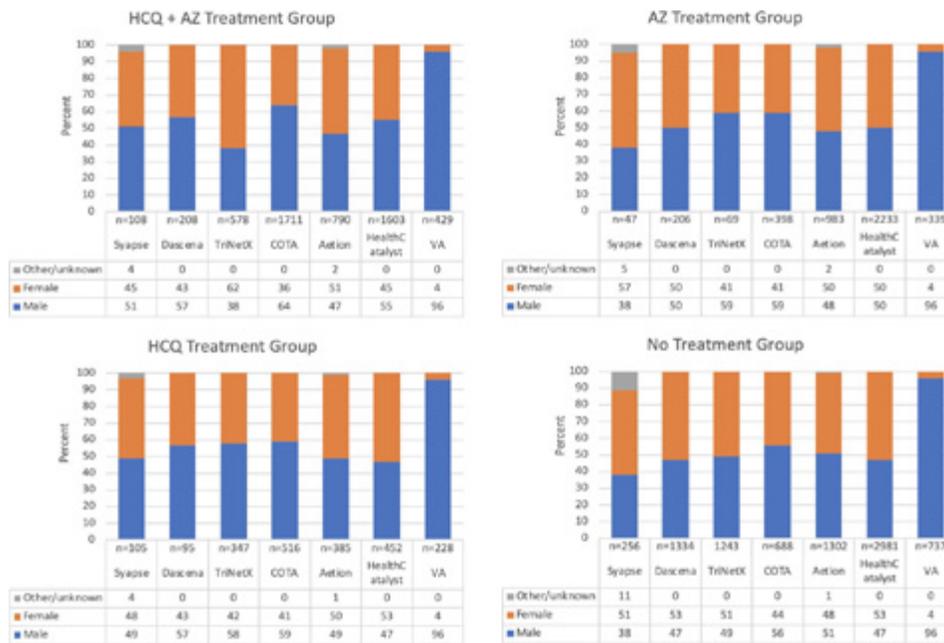
Fig 4. Sex distribution by treatment group. Sex distribution by treatment group shows similarities and differences across data sets. Note. Numbers represent percent values.

Table 5. Patient characteristics were similar across datasets. Patients receiving treatment with hydroxychloroquine were typically older than 45 with a larger proportion of males.

For the adverse event with most complete reporting, any arrhythmia, analysis did not support increased arrhythmia in patients receiving hydroxychloroquine versus those not (Fisher's Exact Test p-value 0.462) (Table 6).

Outcomes

Frequencies of each evaluated outcome are displayed in Table 7. In the Syapse study population of patients with cancer and COVID-19, crude all-cause mortality estimates were greatest among patients receiving hydroxychloroquine plus azithromycin (29.6%), followed by azithromycin alone (21.3%). The crude cumulative incidence of mechanical ventilation was also greatest in the hydroxychloroquine plus azithromycin treatment arm (35.2%), followed by hydroxychloroquine alone (22.9%).

Given constraints of the data, not all groups performed an adjusted analysis. Six groups (Dascena, Health Catalyst, TriNetX, Aetion, COTA/HMH and the VA) conducted adjusted analyses. Among the 3 groups that conducted comparative analyses between hydroxychloroquine plus azithromycin and monotherapy treatment groups (Dascena, Health Catalyst and the VA), after adjusting for confounding, hydroxychloroquine alone was not found to be associated with mortality, overall patient condition, or benefit to the patient. Interrogation of confounding mitigation demonstrated an ability to balance across treatment groups especially on key characteristics such as age and comorbid burden. Hydroxychloroquine plus azithromycin was similarly not associated with any of the outcomes assessed in this study. Adjusted results are presented in Table 8.

Based on these results, hydroxychloroquine and azithromycin, alone or in combination, did not appear to impact outcomes among COVID-19 patients. These results were consistent
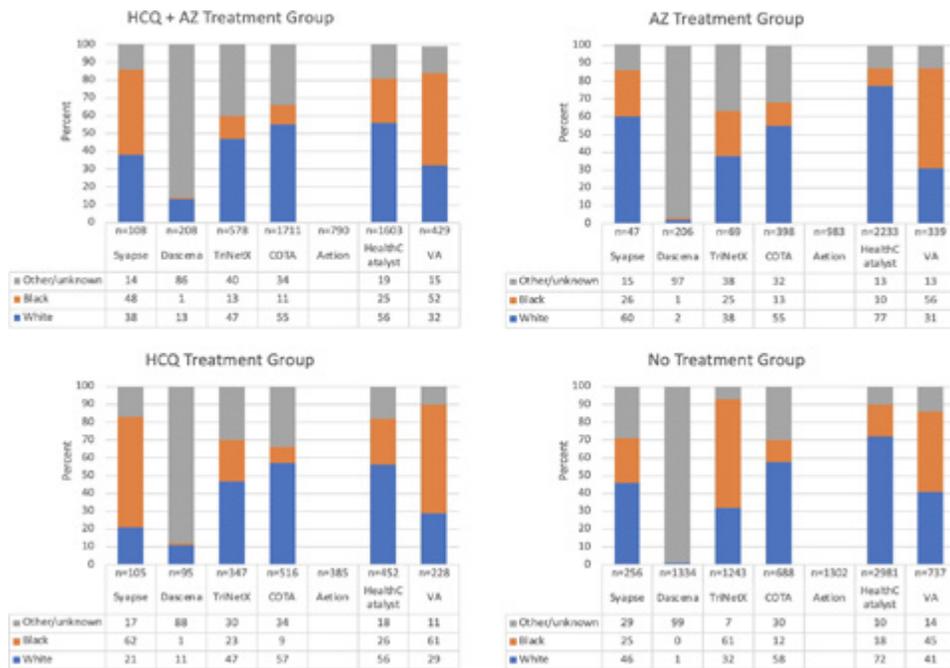
Fig 5. Race distribution by treatment group. Race distribution by treatment group shows similarities and differences across data sets. Note. Numbers represent percent values.

across datasets, with the only notable difference being the evidence of benefit with azithromycin treatment observed by Dascena. Besides the primary analysis, Health Catalyst performed additional sensitivity analyses based upon (a) inclusion criteria (positive lab or ICD-10; require primary discharge ICD-10; allow treatment initiation after the second day), (b) treating mortality as binary, (c) different confounders and confounding mitigation techniques. Results were as expected: (a) broader inclusion criteria resulted in more baseline group differences, (b) treating mortality as binary reduced potential treatment benefit especially when not requiring rapid treatment initiation (e.g., immortal time bias), and (c) more extreme results were observed in the absence of any confounder adjustment technique and different techniques had small impacts of results.

Discussion

In this study, a consortium of groups conducted parallel analyses of the effects of hydroxychloroquine, azithromycin, and their combination on health outcomes of COVID-19 patients. By conducting parallel analyses that aligned on a common protocol, while allowing for flexibility within each group to define covariates, exposure and outcome identification, this study aimed to provide a robust description of outcomes associated with the use of hydroxychloroquine for the treatment of COVID-19.

Among five sites that contributed race data, Syapse, the VA and Health Catalyst reported that Black patients made up a larger distribution of HCQ/HCQ+AZM recipients compared to their distribution without these treatments. Race did not appear to be associated with HCQ administration in the COTA/HMH dataset; and White patients represented a greater proportion of HCQ/HCQ+AZM recipients in the TriNetX dataset as compared to their distribution among those without HCQ treatment.

Table 5. Summary of comorbidities across groups.

Hydroxychloroquine	Population Size, n	Any Cardiovascular Disease	Hypertension	Diabetes	Obesity	Coronary artery disease	Congestive heart failure	Chronic lung disease
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Aetion	385	244 (63)	172 (45)	118 (31)	79 (21)	42 (11)	54 (14)	57 (15)
COTA/HMH	516	334 (65)	313 (61)	185 (36)	186 (36)	94 (18)	Not assessed	37 (7)
Dascena	91	7 (8)	4 (4)	9 (10)	2 (2)	Not assessed	2 (2)	1 (1)
Health Catalyst	335	186 (56)	171 (51)	106 (32)	112 (33)	73 (22)	67 (20)	87 (26)
Syapse	105	49 (47)	29 (28)	37 (35)	79 (75)	9 (9)	8 (9)	0 (0)
TriNetX	347	194 (56)	226 (65)	135 (39)	140 (40)	133 (38)	82 (24)	Not assessed
VA	228	122 (54)	179 (79)	123 (54)	107 (47)	80 (35)	56 (25)	52 (23)

Hydroxychloroquine + Azithromycin	Population Size, n	Any Cardiovascular Disease	Hypertension	Diabetes	Obesity	Coronary artery disease	Congestive heart failure	Chronic lung disease
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Aetion	790	438 (55)	271 (34)	197 (25)	177 (22)	54 (7)	63 (8)	84 (11)
COTA/HMH	1711	985 (58)	917 (54)	558 (33)	629 (37)	232 (14)	Not assessed	114 (7)
Dascena	206	13 (6)	11 (5)	23 (11)	12 (6)	Not assessed	2 (1)	14 (7)
Health Catalyst	1157	582 (50)	531 (46)	321 (28)	324 (28)	185 (16)	132 (11)	237 (21)
Syapse	108	70 (65)	48 (44)	38 (35)	92 (85)	13 (12)	10 (9)	1 (1)
TriNetX	578	225 (39)	280 (48)	202 (35)	198 (34)	129 (22)	53 (9)	Not assessed
VA	429	179 (42)	312 (73)	205 (48)	209 (49)	119 (28)	74 (17)	95 (22)

Azithromycin	Population Size, n	Any Cardiovascular Disease	Hypertension	Diabetes	Obesity	Coronary artery disease	Congestive heart failure	Chronic lung disease
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Aetion	983	538 (55)	347 (35)	259 (26)	210 (21)	90 (9)	100 (10)	133 (14)
COTA/HMH	398	219 (55)	204 (51)	104 (26)	127 (32)	52 (13)	Not assessed	34 (9)
Dascena	201	28 (14)	22 (12)	20 (11)	7 (4)	Not assessed	7 (4)	7 (4)
Health Catalyst	1546	719 (47)	641 (42)	413 (27)	315 (20)	223 (22)	172 (11)	285 (18)
Syapse	47	30 (64)	21 (45)	13 (28)	43 (91)	8 (17)	8 (17)	2 (4)
TriNetX	69	35 (51)	37 (54)	24 (35)	24 (35)	17 (25)	11 (16)	Not assessed
VA	339	154 (45)	244 (72)	151 (45)	149 (44)	106 (31)	71 (21)	72 (21)

Neither	Population Size, n	Any Cardiovascular Disease	Hypertension	Diabetes	Obesity	Coronary artery disease	Congestive heart failure	Chronic lung disease
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Aetion	1302	786 (60)	475 (37)	356 (27)	237 (18)	137 (11)	119 (9)	168 (13)
COTA/HMH	688	342 (50)	319 (46)	159 (23)	191 (28)	98 (14)	Not assessed	50 (7)
Dascena	1284	100 (17)	86 (15)	74 (13)	17 (3)	Not assessed	24 (2)	35 (6)
Health Catalyst	1101	611 (56)	556 (51)	337 (31)	283 (26)	233 (21)	225 (20)	276 (25)

(*Continued*)

Table 5. (Continued)

Syapse	256	142 (55)	88 (34)	68 (27)	207 (81)	21 (8)	36 (14)	1 (0)
TriNetX	1243	646 (52)	764 (61)	453 (36)	378 (30)	248 (20)	245 (20)	Not assessed
VA	737	423 (57)	592 (80)	377 (51)	271 (37)	284 (39)	201 (27)	197 (27)

Across all datasets and treatment groups, the most prominent pre-existing conditions tended to be any cardiovascular disease, hypertension, diabetes and obesity. Overall, obesity was more prevalent among the HCQ treatment groups than in the neither group.

For most data partners, the proportion of patients treated with any of these comorbidities was lower or no different in the HCQ groups than in the neither group–with the exception of Syapse, which was a cancer cohort.

There are several limitations to this study that must be acknowledged. First, despite our goal of carrying out the same set of analyses on multiple datasets, analyses could not be carried

Table 6. Frequency of adverse events across each treatment group.

Hydroxychloroquine	Population Size, n	Any arrhythmia N (%)	Diarrhea N (%)	MI, Stroke, CABG/PCI N (%)	Any conduction disorder N (%)	hypoglycemia N (%)
Aetion*	385	92 (24)	15 (4)	29 (8)	13 (3)	12 (3)
COTA/HMH	516	36 (7)	0 (0)	0 (0)	12 (2)	0 (0)
Dascena	91	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
Syapse	105	7 (7)	1 (1)	4 (4)	0 (0)	1 (1)
TriNetX	347	2 (1)	43 (12)	3 (1)	35 (10)	3 (1)
Hydroxychloroquine + Azithromycin	**Population Size, n**	**Any arrhythmia** N (%)	**Diarrhea** N (%)	**MI, Stroke, CABG/PCI** N (%)	**Any conduction disorder** N (%)	**hypoglycemia** N (%)
Aetion*	790	193 (24)	50 (6)	49 (6)	23 (3)	29 (4)
COTA/HMH	1711	88 (5)	0 (0)	0 (0)	35 (2)	0 (0)
Dascena	206	10 (5)	3 (2)	0 (0)	0 (0)	3 (2)
Syapse	108	15 (14)	4 (4)	4 (4)	0 (0)	0 (0)
TriNetX	578	4 (1)	96 (17)	9 (2)	54 (9)	2 (0)
Azithromycin	**Population Size, n**	**Any arrhythmia** N (%)	**Diarrhea** N (%)	**MI, Stroke, CABG/PCI** N (%)	**Any conduction disorder** N (%)	**hypoglycemia** N (%)
Aetion*	983	232 (24)	51 (5)	53 (5)	36 (4)	36 (4)
COTA/HMH	398	11 (3)	0 (0)	0 (0)	Not assessed	0 (0)
Dascena	201	3 (2)	2 (1)	0 (0)	0 (0)	0 (0)
Syapse	47	6 (13)	3 (6)	3 (6)	1 (2)	0 (0)
TriNetX	69	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Neither	**Population Size, n**	**Any arrhythmia** N (%)	**Diarrhea** N (%)	**MI, Stroke, CABG/PCI** N (%)	**Any conduction disorder** N (%)	**hypoglycemia** N (%)
Aetion*	1302	325 (25)	67 (5)	86 (7)	85 (7)	80 (6)
COTA/HMH	688	21 (3)	0 (0)	0 (0)	Not assessed	0 (0)
Dascena	1284	5 (1)	1 (0)	0 (0)	0 (0)	3 (1)
Syapse	256	30 (12)	3 (1)	8 (3)	3 (1)	2 (1)
TriNetX	1243	24 (2)	35 (3)	9 (1)	149 (12)	2 (0)

*Adverse event data from discharge diagnoses.

Table 7. Frequencies of outcome for each treatment group.

No Treatment	Mortality	Mechanical Ventilation	Evidence of Benefit
Aetion* (n = 1302)	Not assessed	95 (7%)	Not assessed
COTA/HMH (n = 688)	123 (18%)	44 (6%)	492 (72%)
Dascena (n = 1334)	47 (4%)	44 (3%)	539 (40%)
Health Catalyst (n = 1101)	203 (18%)	Not assessed	Not assessed
Syapse (n = 256)	33 (12.9%)	16 (6.2%)	Not assessed
TriNetX (n = 1243)	188 (15%)	Not assessed	728 (59%)
VA (n = 737)	141 (19%)	69 (9%)	Not assessed
Hydroxychloroquine	**Mortality**	**Mechanical Ventilation**	**Evidence of Benefit**
Aetion* (n = 385)	Not assessed	48 (12%)	Not assessed
COTA/HMH (n = 516)	154 (30%)	111 (22%)	270 (52%)
Dascena (n = 95)	20 (21%)	14 (15%)	65 (68%)
Health Catalyst (n = 335)	50 (15%)	Not assessed	Not assessed
Syapse (n = 105)	18 (17%)	24 (23%)	Not assessed
TriNetX (n = 347)	45 (13%)	Not assessed	176 (46%)
VA (n = 228)	49 (21%)	32 (14%)	Not assessed
Hydroxychloroquine + Azithromycin	**Mortality**	**Mechanical Ventilation**	**Evidence of Benefit**
Aetion* (n = 790)	Not assessed	128 (16%)	Not assessed
COTA/HMH (n = 1711)	428 (25%)	479 (29%)	1,089 (64%)
Dascena (n = 208)	46 (29%)	73 (35%)	124 (60%)
Health Catalyst (n = 1157)	212 (18%)	Not assessed	Not assessed
Syapse (n = 108)	32 (30%)	38 (35%)	Not assessed
TriNetX (n = 578)	66 (11%)	Not assessed	316 (55%)
VA (n = 429)	90 (21%)	64 (15%)	Not assessed
Azithromycin	**Mortality**	**Mechanical Ventilation**	**Evidence of Benefit**
Aetion* (n = 983)	Not assessed	144 (15%)	Not assessed
COTA/HMH (n = 398)	97 (24%)	43 (11%)	266 (67%)
Dascena (n = 206)	8 (4%)	28 (14%)	96 (47%)
Health Catalyst (n = 1546)	280 (18%)	Not assessed	Not assessed
Syapse (n = 47)	10 (21%)	3 (6%)	Not assessed
TriNetX (n = 69)	8 (12%)	Not assessed	35 (51%)
VA (n = 339)	56 (17%)	39 (12%)	Not assessed

*To align with other Parallel Analysis partners, Aetion assessed the risk of incident mechanical ventilation among patients in the risk set sampled population, before propensity score matching (used for T1-6). Outcome frequencies in T6 are reported among patients without record of ventilation prior to or concurrent with treatment index.

out identically on all datasets due to differences in and limitations of the data. In particular, not all groups were able to carry out an adjusted analysis that controlled for confounding variables. Second, due to stratification on treatment groups, some analyses were conducted on small sample sizes. Data was also limited to those collected from United States sources. These results may therefore not be generalizable to international settings. Coding of certain outcomes, particularly adverse events, may have been incomplete. Therefore, not all outcomes may have been captured, potentially limiting the accuracy of our results.

These analyses were conducted in parallel among 7 individual groups using their own datasets. Characteristics, definitions, and methodologies used by each of the groups are summarized in Tables 1 and 4. A goal of this project was to develop a common analytical plan for multiple groups to apply to different datasets as they continued to aggregate data on the

Table 8. Adjusted hazard ratios for each treatment group.

Hydroxychloroquine	Mortality	Mechanical Ventilation	Evidence of Benefit
Aetion*	Not assessed	1.29 (0.96, 1.74)	Not assessed
COTA/HMH	1.22 (0.93, 1.60)	Not assessed	Not assessed
Dascena	2.6 (0.82, 8.0)	0.32 (0.04, 2.4)	0.80 (0.37, 1.7)
Health Catalyst	Not assessed	Not assessed	Not assessed
TriNetX	1.4 (0.97, 2.05)	Not assessed	1.03 (0.86, 1.23)
VA	1.18 (0.81, 1.72)	1.31 (0.81, 2.13)	Not assessed
Hydroxychloroquine + Azithromycin	**Mortality**	**Mechanical Ventilation**	**Evidence of Benefit**
Aetion	Not assessed	Not assessed	Not assessed
COTA/HMH	1.16 (0.90, 1.51)	Not assessed	Not assessed
Dascena	1.9 (0.91, 4.1)	2.5 (1.2, 5.2)	0.90 (0.58, 1.4)
Health Catalyst	1.09 (0.76, 1.56)	Not assessed	Not assessed
TriNetX	0.99 (0.73, 1.35)	Not assessed	0.87 (0.75, 1.01)
VA	1.18 (0.88, 1.58)	1.54 (1.07, 2.23)	Not assessed
Azithromycin	**Mortality**	**Mechanical Ventilation**	**Evidence of Benefit**
Aetion	Not assessed	Not assessed	Not assessed
COTA/HMH	1.31 (0.95, 1.82)	Not assessed	Not assessed
Dascena	0.79 (0.31, 1.6)	1.7 (1.0, 3.0)	1.4(1.0, 1.8)
Health Catalyst	Not assessed	Not assessed	Not assessed
TriNetX	Not assessed	Not assessed	Not assessed
VA	0.89 (0.63, 1.25)	1.03 (0.66, 1.61)	Not assessed

*Aetion reports adjusted hazard ratios for incident mechanical ventilation events among hydroxychloroquine initiators and untreated controls matched with both risk set sampling and propensity score models. Hazard ratios are derived from all hydroxychloroquine initiators and all matched controls, regardless of azithromycin treatment.

https://doi.org/10.1371/journal.pone.0248128.t008

experience of diagnosis, treatment, and outcomes associated with COVID-19. Given the novelty of the virus and magnitude of the pandemic, the use of data derived from various sources of healthcare data presents an opportunity to augment clinical trial data with information about patients not enrolled in clinical studies, and provide information about treatment patterns and observations about those experiences in large, diverse populations. By using a common analysis plan, the resulting observations can be more readily compared and if consistent, can further support the findings among individual studies.

In this instance, several of the studies were already underway when the parallel analysis was started. This made it difficult to align on all aspects of the analyses and data parameters. As future parallel analyses are considered, it is important that participants seek to develop as much uniformity to the definitions and methods as possible. However, given the different sources of data, some aspects of the analysis will need to be tailored to the individual dataset and those variations should be clearly described. As future study questions are developed to further characterize COVID-19 treatments, it is important to select the sources of data that are best fit to answer each specific question. In addition, future parallel analysis should consider using a stepwise approach to perform a sample size, demographic, and feasibility assessment and use that initial step to optimally design subsequent comparative analyses.

Conclusion

The Evidence Accelerator successfully brought together seven partners to execute analyses in disparate populations. Representing more than 20,000 patients with COVID-19 across the U.S,

we found similar trends in those getting HCQ treatment–despite minor differences in coding and cohort entry. Across the 5 groups who ran comparative analyses, we observed no association between HCQ treatment and mortality, overall patient condition, or evidence of benefit.

Supporting information

S1 File.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Mark Stewart, Carla Rodriguez-Watson, Adem Albayrak, Andrew Belli, Thomas Brown, Kelly Cho, Ritankar Das, Elizabeth Eldridge, Nicolle Gatto, Stuart L. Goldberg, Eric Hansen, Jonathan Hirsch, Andrew Ip, Jason Jones, Amy C. Justice, Samson Mataraso, Robertino Mera, Andrew Weckstein, Amar Bhat, Susan Winckler, Ellen V. Sigal, Jeff Allen.

Data curation: Adem Albayrak, Thomas Brown, Kelly Cho, Elizabeth Eldridge, Alice Gelman, Hanna Gerlovin, Stuart L. Goldberg, Yuk-Lam Ho, Andrew Ip, Jason Jones, Amy C. Justice, Reyna Klesh, Seth Kuranz, Carson Lam, Daniel C. Posner, Jeremy A. Rassen, Andrew Schrag.

Formal analysis: Julius Asubonteng, Thomas Brown, Nicolle Gatto, Hanna Gerlovin, Stuart L. Goldberg, Eric Hansen, Andrew Ip, Monika Izano, Jason Jones, Amy C. Justice, Reyna Klesh, Carson Lam, Robertino Mera, Daniel C. Posner, Jeremy A. Rassen, Andrew Weckstein.

Funding acquisition: Kelly Cho, Jonathan Hirsch.

Investigation: Adem Albayrak, Julius Asubonteng, Thomas Brown, Kelly Cho, Elizabeth Eldridge, Nicolle Gatto, Alice Gelman, Stuart L. Goldberg, Andrew Ip, Jason Jones, Reyna Klesh, Seth Kuranz, Andrew Schrag, Andrew Weckstein.

Methodology: Adem Albayrak, Julius Asubonteng, Andrew Belli, Thomas Brown, Kelly Cho, Elizabeth Eldridge, Nicolle Gatto, Hanna Gerlovin, Stuart L. Goldberg, Eric Hansen, Jonathan Hirsch, Andrew Ip, Monika Izano, Jason Jones, Amy C. Justice, Carson Lam, Robertino Mera, Daniel C. Posner, Jeremy A. Rassen, Anna Siefkas, Andrew Weckstein, Frank Wolf.

Project administration: Mark Stewart, Carla Rodriguez-Watson, Adem Albayrak, Andrew Belli, Thomas Brown, Kelly Cho, Elizabeth Eldridge, Nicolle Gatto, Andrew Ip, Amy C.

References

1. Cortegiani A, Ingoglia G, Ippolito M, Giarratano A, Einav S. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. J Crit Care. 2020 Jun; 57:279–83. https://doi.org/10.1016/j.jcrc.2020.03.005 PMID: 32173110

2. Fiolet T, Guihur A, Rebeaud ME, Mulot M, Peiffer-Smadja N, Mahamat-Saleh Y. Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2020 Aug;

3. Yao X, Ye F, Zhang M, Cui C, Huang B, Niu P, et al. In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Clin Infect Dis an Off Publ Infect Dis Soc Am. 2020 Jul; 71(15):732–9.

4. Singh AK, Singh A, Singh R, Misra A. "Hydroxychloroquine in patients with COVID-19: A Systematic Review and meta-analysis.". Diabetes Metab Syndr. 2020; 14(4):589–96. https://doi.org/10.1016/j.dsx.2020.05.017 PMID: 32417708

5. Gautret P, Lagier J-C, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents. 2020 Jul; 56(1):105949. https://doi.org/10.1016/j.ijantimicag.2020.105949 PMID: 32205204

6. No clinical benefit from use of hydroxychloroquine in hospitalised patients with COVID-19—RECOVERY Trial [Internet]. 2020. Available from: https://www.recoverytrial.net/news/statement-from-the-chief-investigators-of-the-randomised-evaluation-of-covid-19-therapy-recovery-trial-on-hydroxychloroquine-5-june-2020-no-clinical-benefit-from-use-of-hydroxychloroquine-in-hospitalised-patients-with-co

7. Rosenberg ES, Dufort EM, Udo T, Wilberschied LA, Kumar J, Tesoriero J, et al. Association of Treatment With Hydroxychloroquine or Azithromycin With In-Hospital Mortality in Patients With COVID-19 in New York State. JAMA. 2020 Jun; 323(24):2493–502. https://doi.org/10.1001/jama.2020.8630 PMID: 32392282

8. Ip A, Berry DA, Hansen E, Goy AH, Pecora AL, Sinclaire BA, et al. Hydroxychloroquine and tocilizumab therapy in COVID-19 patients-An observational study. PLoS One. 2020; 15(8):e0237693. https://doi.org/10.1371/journal.pone.0237693 PMID: 32790733

9. Cavalcanti AB, Zampieri FG, Rosa RG, Azevedo LCP, Veiga VC, Avezum A, et al. Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. N Engl J Med [Internet]. 2020 Jul 23; Available from: https://doi.org/10.1056/NEJMoa2019014

10. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines [Internet]. 2020. Available from: Coronavirus Disease 2019 (COVID-19)%0ATreatment Guidelines

11. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020 May; 368(6493):860–8. https://doi.org/10.1126/science.abb5793 PMID: 32291278

12. Chapman A, Peterson K, Turano A, Box T, Wallace K, Jones M. A Natural Language Processing System for National COVID-19 Surveillance in the US Department of Veterans Affairs [Internet]. 2020. Available from: https://openreview.net/pdf?id=ZQ_HvBxcdCv

13. Ridgeway G. Generalized Boosted Models: A guide to the gbm package. Update. 1(1):2007.

14. Greenwell B, Boehmke B, Cunningham J, Developers GBM. gbm: Generalized boosted regression models. R Packag version. 2019; 2(5).

15. Greifer N. WeightIt: Weighting for Covariate Balance in Observational Studies [Internet]. 2020. Available from: https://cran.r-project.org/web/packages/WeightIt/index.html

FRIENDS
of CANCER
RESEARCH

Harmonizing the Definition and Reporting of Cytokine Release Syndrome (CRS) in Immuno-Oncology Clinical Trials

SHAPING THE FUTURE OF EMERGING IMMUNOTHERAPIES AND CELL THERAPIES

Objective

This white paper focuses on establishing a standardized approach for defining and capturing cytokine release syndrome (CRS). It also provides considerations for categorizing the variety of adverse events (AEs) that may accompany CRS, recognizing that presentations of CRS may differ among various immunotherapeutics (e.g., monoclonal antibodies, CAR T-cell therapies, and T-cell engagers, which can include bispecific antibodies and other constructs). The ultimate goal is to ensure accurate and consistent identification of CRS in patients receiving immunotherapies in clinical studies to aid in reporting; enable a more precise evaluation of the therapeutic risk-benefit profile; support evidence-based monitoring and management of novel toxicities; and improve patient care and outcomes. This will be of increasing importance as the number and variety of molecular targets for these therapies expands and immunotherapies with novel mechanisms of action are tested either as a monotherapy or in combinations.

Introduction

The emergence of cancer immunotherapies has resulted in transformational advances across solid and hematological malignancies, bringing new hope to patients with serious, life-threatening diseases. Cancer immunotherapies provide clinically beneficial alternatives and additions to traditional cytotoxic treatments. Recent U.S. Food and Drug Administration (FDA) approvals and the rapid expansion of indications for existing agents are enabling broader availability of immunotherapies to cancer patients.

The immuno-oncology (IO) drug development pipeline continues to grow, and cancer immunotherapies are quickly being integrated into the standard of care for many cancers.[1] Importantly, our increasing clinical experience with these immunotherapeutic agents has brought greater awareness to several unique toxicities when compared to traditional cytotoxic agents. With the success of newer immunotherapies like T-cell engagers and chimeric antigen receptor (CAR) T-cells in several hematologic malignancies, there has been growing recognition of cytokine

Thank You to Our Working Group Members

Carolyn Britten, Amgen

Meredith Chuk, U.S. FDA

Bindu George, U.S. FDA

Nicole Gormley, U.S. FDA

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

Eric Kowack, Xencor

Ke Liu, U.S. FDA

Bruce McCall, Genentech

Candice McCoy, Bristol Myers Squibb

PK Morrow, Amgen

Emmanuel Okoye, Regeneron

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

Rosanna Ricafort, Bristol Myers Squibb

John Rossi, Kite, a Gilead Company

Elad Sharon, National Cancer Institute

Marc Theoret, U.S. FDA

Ferdinando Vegni, Bristol Myers Squibb

Wayne Wallis, Bristol Myers Squibb

Allen Yang, Xencor

Tai Yu, Amgen

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

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

Although CRS is increasingly recognized as an on-target effect associated with CAR T-cells and T-cell engagers, the scope of this syndrome, including effects on end organ function, has not been fully characterized. A standardized approach is needed for diagnosing and reporting CRS and its manifestations in clinical trials, published literature, and in clinical practice. More importantly there is a need to distinguish CRS from other clinical entities, such as acute infusion-related reactions (IRR). Acute IRRs and CRS can have overlapping symptoms and temporality but likely have different pathophysiology and treatments with different prognoses. Inconsistent or inadequate characterization of these toxicities in clinical trials impact how data are presented in publications and prescribing information, potentially resulting in suboptimal description and management of these clinical events. This can put patients at risk if their treatment side effects are not properly managed.

Growing Clinical Experience of Infusion Reactions and CRS

Adverse events (AEs) known broadly as IRRs have been long defined, diagnosed, and reported in an ambiguous and inconsistent manner.[4] This arises, in part, from the fact that the term IRR came into use at a time when few biological therapies were available and acute reactions to an infusion of a biologic agent were reported. Additionally, little was known about the exact mediators involved in these reactions. Since the introduction of therapeutic monoclonal antibodies into clinical practice, IRR continues to be used as a term to describe a variety of symptoms occurring during or shortly after the infusion of the medicinal product. Infusion related reactions following CAR T-cell administration are infrequent and generally mild. However, with the emergence of T-cell-engaging therapeutics, in particular T-cell engagers and fusion proteins, distinguishing CRS from IRR has been a challenge, in that the signs and symptoms may partially overlap.

Infusion-related reaction is a broad term traditionally used to encompass acute findings during or shortly after an infusion that may include hypersensitivity/anaphylaxis, complement activation-related pseudoallergy (CARPA), CRS, or more nonspecific signs and symptoms.[4] During clinical development, IRRs are generally defined as AEs occurring within the first 24 hours after infusion of a therapy, with causality deemed by the investigator as related to the therapy. This operational definition has resulted in this term being used to define a wide array of symptoms with potentially disparate pathophysiology whose main commonality is occurrence within 24 hours of infusion. The majority of IRRs reported with therapeutic monoclonal antibodies are self-limited and treated symptomatically.[5–8] However, a primary clinical concern within the con-

text of IRR is whether the reaction is mediated by immunoglobulin E (IgE) because this specific type of reaction can increase in severity with additional infusions. For that reason, re-challenging is contraindicated with IgE mediated hypersensitivity.[4,9,10]

CRS is a supraphysiologic response driven by the immune system, which is also observed commonly in sepsis and other infections and most recently with COVID-19. CRS is T-cell mediated and can occur within several hours to days after infusion, but rarely presents beyond 14 days after initiation of therapy. CRS can be short lived, but often lasts for several days. Because symptoms of CRS can overlap with other toxicities that have generally been classified as IRRs, and because both CRS and IRR can occur within a day after infusion, careful examination of the signs and symptoms, their attribution, and the response to therapy is important. The presentation of CRS may differ depending on the immunotherapeutic. The timing of the onset of CRS can coincide closely with infusion of T-cell engagers, but for cellular products where T-cell expansion precedes clinical CRS, there may be a significant lag between infusion and CRS symptom onset.[11] CRS typically presents with a fever and may progress to hypotension and/or hypoxia. Flushing and rash may accompany both anaphylactic reactions and CRS, although specific skin and mucosal changes such as hives and mucosal swelling predominate in anaphylactic reactions, occurring in 80% of cases.[12] An underlying factor associated with CRS is the release of cytokines, and this has been identified as a differentiating criterion in the Common Terminology Criteria for Adverse Events (CTCAE) v.5 definition for CRS and IRRs. However, the measurement of cytokines is not yet a routine element in clinical practice, nor are there reliable cytokine thresholds for CRS diagnosis. Thus, this distinction alone may not be helpful to clinicians at the bedside, and emergent clinical interventions are still largely based on the clinical manifestations and severity of CRS as well as response to therapeutic interventions. For example, the role of the IL-6 pathway in CAR T-cell therapy has been characterized and use of IL-6 blocking agents is the primary treatment of CRS.[13,14]

CRS Definition and Severity

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

Fever is a CRS-defining characteristic but does not dictate the severity. Therefore, the MSKCC grading system initially relied on the availability of released cytokine levels measured from patients in real-time to distinguish severe versus non-severe CRS.[21] However, cytokine testing may be limited to specific health care research settings, and there is currently poor correlation between cytokine levels and the intensity of CRS signs and symptoms. The presence and severity of hypotension and hypoxia are most commonly used to assign the grade of severity for CRS, as these two events drive the need for higher level of care (e.g., intensive care). One unique aspect of CRS grading is that the severity is often attributed based on practitioner intervention. For example, the utilization of one or more vasopressor agents to treat hypotension or use of supplemental oxygen or mechanical ventilation for hypoxia would dictate the CRS severity grade.

Table 1: Evolving Definitions and Criteria for Grading and Managing CRS

Lee Criteria[15]

Grade 1: Symptoms are not life-threatening and require symptomatic treatment only (e.g., fever, nausea, fatigue, headache, myalgias, malaise)

Grade 2: Symptoms require and respond to moderate intervention; Oxygen requirement <40% or hypotension responsive to IV fluids or low-dose single vasopressor or grade 2 organ toxicity

Grade 3: Symptoms require and respond to aggressive intervention; Oxygen requirement of ≥40% or hypotension requiring high-dose or multiple vasopressors or grade 3 organ toxicity or grade 4 transaminitis

Grade 4: Life-threatening symptoms; Requirements for ventilator support OR grade 4 toxicity (excluding transaminitis)

CTCAE v5.0[16]

Grade 1: Fever with or without constitutional symptoms

Grade 2: Hypotension responding to fluids; Hypoxia responding to <40% oxygen

Grade 3: Hypotension managed with one vasopressor; Hypoxia requiring ≥40% oxygen

Grade 4: Life-threatening consequences; Urgent intervention indicated

Memorial Sloan Kettering Cancer Center (MSKCC)[17]

Grade 1: Mild symptoms requiring observation or supportive care only (e.g., antipyretics, antiemetics, pain medication)

Grade 2: Hypotension requiring any vasopressors <24 h; Hypoxia or dyspnea requiring supplemental oxygen <40%

Grade 3: Hypotension requiring any vasopressors ≥24 h; Hypoxia or dyspnea requiring supplemental oxygen ≥40%

Grade 4: Life-threatening symptoms; Hypotension refractory to high-dose vasopressors; Hypoxia or dyspnea requiring mechanical ventilation

Chimeric Antigen Receptor Toxicity (CARTOX)[18]

Grade 1: Temperature ≥38ºC Grade 1 organ toxicity

Grade 2: Hypotension responds to intravenous fluids or low-dose vasopressor; Hypoxia requiring oxygen <40%; Grade 2 organ toxicity

Grade 3: Hypotension needing high-dose or multiple vasopressors; Hypoxia requiring oxygen ≥40%; Grade 3 organ toxicity or Grade 4 transaminitis

Grade 4: Life-threatening hypotension; Needing ventilator support; Grade 4 organ toxicity except Grade 4 transaminitis

Penn Criteria[19]
Grade 1: Mild reaction: Treated with supportive care, such as antipyretics, antiemetics
Grade 2: Moderate reaction: Some signs of organ dysfunction (Grade 2 creatinine or Grade 3 liver function tests [LFTs]) related to CRS and not attributable to any other condition; Hospitalization for management of CRS-related symptoms, including neutropenic fever and need for IV therapies (not including fluid resuscitation for hypotension)
Grade 3: More severe reaction: Hospitalization required for management of symptoms related to organ dysfunction, including Grade 4 LFTs or Grade 3 creatinine, related to CRS and not attributable to any other condition; Hypotension treated with multiple fluid boluses or low-dose vasopressors; Coagulopathy requiring fresh frozen plasma, cryoprecipitate, or fibrinogen concentrate; Hypoxia requiring supplemental oxygen (nasal cannula oxygen, high-flow oxygen, CPAP, or BiPAP)
Grade 4: Life-threatening complications, such as hypotension requiring high-dose vasopressors; Hypoxia requiring mechanical ventilation
American Society for Transplantation and Cellular Therapy (ASTCT)[20]
Grade 1: Temperature ≥38°C
Grade 2: Temperature ≥38°C; Hypotension not requiring vasopressor; Hypoxia requiring low-flow nasal cannula or oxygen blow-by
Grade 3: Temperature ≥38°C; Hypotension requiring one vasopressor with or without vasopressin; Hypoxia requiring high-flow nasal cannula, facemask, nonrebreather mask, or Venturi mask
Grade 4: Temperature ≥38°C; Hypotension requiring multiple vasopressors (excluding vasopressin); Hypoxia requiring positive pressure ventilatory support (CPAP, BiPAP, intubation, and mechanical ventilation)

This is important to remember as the use of vasopressors or respiratory support is based on the clinical judgement of the physician, which may vary and lead to individual bias in CRS grading. The presence of other organ function abnormalities is included in some but not all grading systems. These abnormalities of other organs could be reported either as separate AEs with no relationship to CRS or as preferred terms encompassing CRS. Thus, it is important to clarify whether CRS definition would consider these abnormalities to capture the full extent of CRS and minimize the possibilities of under-documenting and/or reporting. Additionally, if a therapeutic modality has the potential to cause clinically severe CRS that requires treatment with fluids, vasopressors, supplemental oxygen, and anti-cytokine therapy, we should assume that low-grade events related to these manifestations that may occur initially are part of that spectrum and define them as CRS. While there are a variety of published manuscripts, descriptions, and adapted grading criteria and management strategies for CRS[20], it is noted that published definitions and grading criteria do not readily articulate the distinctions among CRS and other clinical entities that may have overlapping symptoms and temporality (e.g., IRR, macrophage activation syndrome/hemophagocytic lymphohistiocytosis [MAS/HLH]).

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

1. Alignment on Defining and Grading CRS

The ASTCT defines CRS as "a supraphysiologic response following any immune therapy that results in the activation or engagement of endogenous or infused T cells and/or other immune effector cells. Symptoms can be progressive, must include fever at the onset, and may include hypotension, capillary leak (hypoxia), and end organ dysfunction."[20] ASTCT's definition for CRS represents an opportunity for alignment, prioritization of grading of clinically relevant events, and can be inclusive of currently available and emerging immunotherapies with some considerations noted below.

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

Since these guidelines have been developed based mainly on the clinical experience with CAR T-cell therapy, they may prove, with additional clinical experience, to be incomplete for all cancer immunotherapies and may need to be revised as new data become available from existing and novel therapies.[20] As such, it is important that data collection is aimed at more than meeting the requirements of any one grading system. Therefore, establishing core principles for defining CRS that consider the therapeutic modality, symptom manifestation, timing, and response to intervention will be important to enable flexibility and maximize utility of a harmonized definition for CRS to adequately assess safety profiles of therapeutics being offered to patients (Table 2).

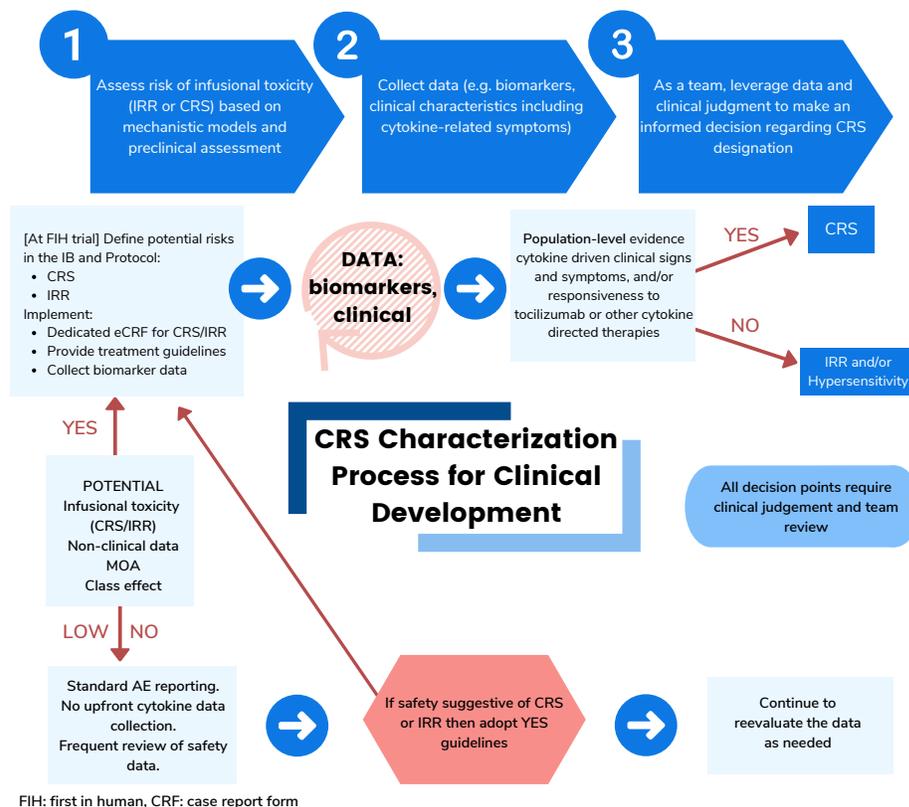## Table 2: Principle Components for Defining Cytokine Release Syndrome

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

2. Strategy for Assessing CRS Over the Course of a Clinical Development Program

The characterization of CRS for a given experimental therapeutic in the course of a clinical development program is crucial to ensure the correct diagnosis and deployment of appropriate toxicity management. This is particularly important since some of the therapies used to manage conditions other than CRS can mitigate the effectiveness of immunotherapy. During the development of protocols for safety data collection and monitoring strategies as it relates to CRS, consideration should be given for how toxicities will be identified and managed in routine clinical care. Recognizing the association between the immunotherapeutic agent and CRS will inform the framework on how best to collect these data.

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

Figure 1. Decision Tree for Assessing a Population-level CRS Risk during Product Development of an Experimental Agent

1 Assess risk of infusional toxicity (IRR or CRS) based on mechanistic models and preclinical assessment

2 Collect data (e.g. biomarkers, clinical characteristics including cytokine-related symptoms)

3 As a team, leverage data and clinical judgment to make an informed decision regarding CRS designation

[At FIH trial] Define potential risks in the IB and Protocol:
• CRS
• IRR
Implement:
• Dedicated eCRF for CRS/IRR
• Provide treatment guidelines
• Collect biomarker data

DATA: biomarkers, clinical

Population-level evidence cytokine driven clinical signs and symptoms, and/or responsiveness to tocilizumab or other cytokine directed therapies

YES → CRS

NO → IRR and/or Hypersensitivity

CRS Characterization Process for Clinical Development

All decision points require clinical judgement and team review

YES

POTENTIAL Infusional toxicity (CRS/IRR) Non-clinical data MOA Class effect

LOW | NO

Standard AE reporting. No upfront cytokine data collection. Frequent review of safety data.

If safety suggestive of CRS or IRR then adopt YES guidelines

Continue to reevaluate the data as needed

FIH: first in human, CRF: case report form

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

For therapeutic classes that are known to be associated with CRS or at particularly high risk for inducing CRS based on mechanism of action or preclinical data, the implementation of a dedicated clinical and safety monitoring plan may be required from the onset. The potential risks of IRR and CRS should be defined in the Investigator Brochure and protocol for the first-in-human trial, with a dedicated case report form (CRF) for IRR/CRS that collects the associated signs and symptoms. In addition, special preparation may be warranted as part of the protocol such as specific site training on CRS and the requirement of certain clinical interventions (e.g., inpatient monitoring, ICU availability, and readily available tocilizumab). In most circumstances, the provision for the physicians to report either IRR or CRS as the preferred term is recommended until human data at the population level are available. If there is evidence at the population level of cytokine-driven clinical signs and symptoms, and/or responsiveness to tocilizumab or other cytokine-directed therapies, it would be concluded that CRS is an identified risk and can be characterized accordingly. Lack of such evidence may suggest that the reaction is a manifestation of IRR or hypersensitivity.

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

3. Harmonized Data Elements for Characterizing CRS

Given the likely evolution of defining and grading CRS in the field, the medical community should ensure that the appropriate data elements are collected to allow derivation with different grading systems. Collection of common data variables using aligned protocols will be important to enable comparison with different therapies in the future. A suspected diagnosis of CRS will most likely be based on clinical signs and symptoms. However, the collection of certain data variables, such as laboratory assessments, cytokine profiles, and biomarkers will be important for future retrospective analyses to assess the relationship of certain signs and symptoms with CRS, the severity of CRS, natural history of the event including response to therapy, or the identification of predictive biomarkers. CRS should be considered as an adverse event of special interest (AESI) whenever there is an association of the IO product and CRS.

CRS is most often characterized by fever, hypotension, hypoxia, and increased release of inflammatory cytokines.[22] The use of proactive premedication (e.g., corticosteroids) may limit or minimize the presence of some symptoms, such as fever, associated with CRS. In addition, the signs and symptoms associated with CRS may represent other adverse events from non-CRS etiologies (i.e., IRR, infection). Early in the clinical development of a novel therapy it is important to collect individual signs and symptoms associated with each case of CRS, since the definition of

CRS has evolved and is likely to continue to evolve as more experience is gained with immunotherapies. A confirmatory diagnosis could be made at a later date and in the context of the evolution of clinical symptoms and cytokine data or response to cytokine-directed interventions (see section "Harmonized Approach for Recording and Reporting CRS Events").

Table 3 outlines key data elements driven largely in part by ASTCT 2019. Review of key data variables from published severity scales should inform the components of a dedicated CRF for CRS. These represent minimal data collection elements for sponsors. Comprehensive data capture will be critical to facilitate new iterations of grading criteria and past criteria to ensure the safe monitoring and administration of T-cell engaging immunotherapies.

Table 3: Harmonized Collection of Discrete Data Elements

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

Vital sign assessment should include body temperature, pulse (heart rate), blood pressure, and oxygen saturation. It is important to note the ASTCT grading depends on the use of supplemental oxygen or positive pressure ventilation and the use of vasopressors. Because the criteria to use these interventions are not standardized, it could introduce some bias into the grading of CRS. Once CRS is further characterized, biomarker testing can be reduced to key time points and biomarkers. Capturing these core data elements may be important for drug label descriptions and management guideline development.

Additional laboratory tests should be considered among patients who experience a more severe manifestation of CRS without initial response to interventional therapy. This can include fibrinogen and complete blood counts, if not already included in the routine hematologic laboratory assessments, triglycerides, and a bone marrow biopsy. The latter being necessary to confirm the diagnosis of MAS/HLH, which likely has a worse prognosis.

In the setting of CAR T-cell therapy, one important determinant associated with CRS and its severity is the in vivo expansion of these cells after infusion. Patients with CRS symptoms with increasing acute phase reactants and expansion of cells might require a different therapeutic approach compared to patients with the same severity of CRS in whom laboratory values are normalizing. While to date treatment guidelines are based on symptoms, it is important to capture the lab value information, including cytokine biomarkers, as lab values and cytokines help improve our understanding of the pathophysiology and may inform future development of management guidelines. Though currently there are no commercially available assays to determine expansion and persistence of CAR T-cells, and real-time cytokine analysis is also not typically available, correlative analyses in the context of clinical trials may allow retrospective analyses to interrogate CRS cases and direct future guidelines for toxicity management.

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

4.   Consistent Method for Recording and Reporting CRS Events

Identification and characterization of CRS can be challenging due the heterogeneity in presentation of signs and symptoms and similarity of these signs and symptoms of CRS to other adverse events, such as IRR or infection. A hypothetical case is shown in **Figure 2**, "Patient Timeline." A patient treated with a T-cell engager experiences several adverse events. Initially, the patient presents with a fever of 40.1 degrees Celsius lasting 6 hours and is accompanied by hypotension that is responsive to a one-liter fluid bolus. The fever and hypotension are CTCAE Grade 3. The next morning liver function test (LFT) increases

are noted (Grade 4), and later that day, the patient has a brief generalized seizure that is self-limited, lasting less than a minute (Grade 2). CRS Grade 2 is diagnosed.[20] While all of these may precede the investigator diagnosing CRS, all these adverse events should be captured into the CRF and independently reported.

Figure 2. Patient Treatment Timeline and Event Reporting Following Immunotherapy Administration
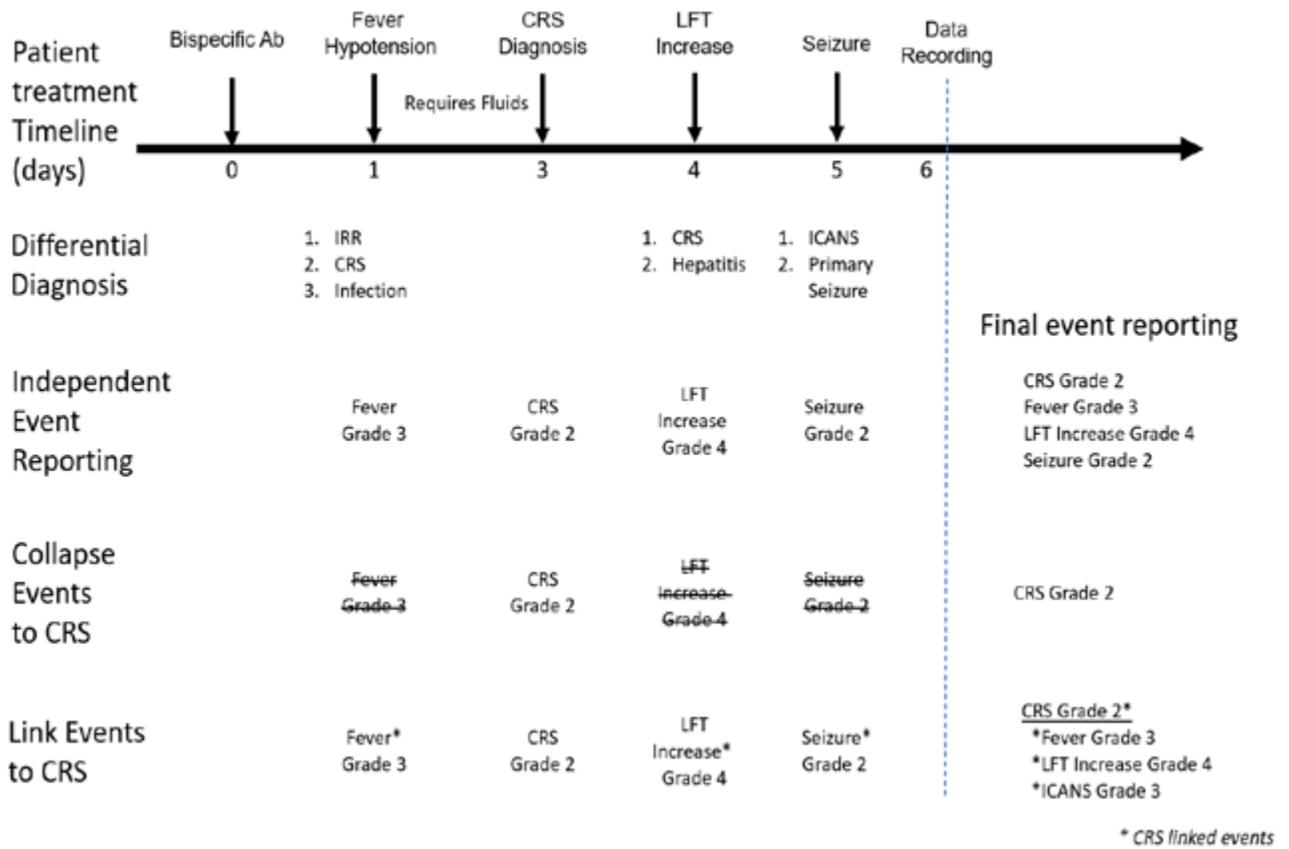
In our example case, the event of "fever" precedes the diagnosis of CRS and would be captured into the AE database independently and graded independently as the differential diagnosis for the fever could include not only CRS but other potential etiologies such as IRR and infection. Additional events such as the increase in LFTs and seizure are attributable to

the CRS but could also be recorded as independent AEs into the database. Once CRS is diagnosed and recorded as an adverse event in the trial database, the symptoms of CRS should be linked together.

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

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

As recommended in **Table 3**, additional information should be captured including use of concomitant medications (i.e., tocilizumab or other cytokine-directed therapy, oxygen, vasopressors, corticosteroids) and specific interventions (i.e., method of oxygen delivery, mechanical intervention, IV fluids). Here in our example case, the use of IV fluids and not vasopressors defines a Grade 2 CRS event. Although these items may be collected in other parts of the electronic data capture record, it is important these events are easily linked to a specific CRS event as CRS grading is dependent on these interventions in most classification systems. In addition, some grading systems can lead to downgrading of events. As an example, liver function laboratory values may increase transiently and meet the criteria for CTCAE Grade 4 CRS based upon these laboratory changes; however, this increase will only meet the definition of a Grade 2 CRS by ASTCT criteria if it is not accompanied by clinically significant changes in blood pressure or oxygen requirement.

Conclusions

Cytokine release syndrome is commonly seen with newer immunotherapies like T-cell engagers and CAR T-cells. All investigators should commit to a harmonized data collection approach using a dedicated CRS eCRF with data elements identified in Table 3 as a guide to ensure consistency in how data is collected and presented. This working group outlined several actionable proposals that can help incentivize more aligned strategies for deployment in early clinical development programs of emerging immunotherapies:

- Alignment on Defining CRS
- Strategy for Assessing CRS Over the Course of a Clinical Development Program
- Harmonized Data Elements for Characterizing CRS
- Consistent Method for Recording and Reporting CRS Events

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

Appendix

An informal survey tool was conducted to provide a landscape assessment of the current approaches and efforts being used to harmonize definitions for CRS and align data collection strategies that can be analyzed retrospectively as definitions change and will maintain its relevance as the field evolves. This survey was circulated amongst participating drug sponsors and organizations and generated three key findings that guided this work:

First, survey responses indicated that, generally, there are not harmonized definitions for IRR and CRS in IO clinical trials. Often, a distinction between IRR and CRS is based on the temporality of the events, but this may be due to the absence of a better parameter. The lack of a standardized definition can be partially explained by the difficulty associated with applying a singular definition to a broad field of diverse agents such as monoclonal antibodies, T-cell engagers, and cell-based therapies. In addition to considering the impact of this context on CRS and IRR definitions, the development of core principles central to any definition of CRS (as opposed to a singular, rigid definition) should be

considered. This approach would allow sufficient flexibility across contexts and as new data emerges.

Next, survey responses indicated that while the ASTCT 2019 CRS severity grading scale is most frequently used, other scales such as CTCAE and the Lee 2014 scales are also used for severity grading. With the evolution of severity grading scales in mind, it will be necessary to collect a core set of raw data elements for CRS events. This would enable retrospective analyses and comparison between therapies developed at different points in time and for which different severity scales were likely used. The collection of a core set of data elements may necessitate a CRS-specific case report form, which, as survey results indicated, is a practice already being implemented by most sponsors for the collection of elements such as grade, associated signs/symptoms, onset, and resolution.

Lastly, survey responses indicated that the uniform collection of data elements will be critical to enabling the mapping of CRS to different severity scales, comparison between drugs, and the future pooling of information. Common data elements such as the timing of events, lab findings, signs/symptoms, severity of events, and management of signs/symptoms should be collected. In thinking through proposed core data elements, it will be important to extend thinking past the current standard of care (SOC) and into the future of SOC for patients.

References

1. Xin Yu, J., Hubbard-Lucey, V. M. & Tang, J. Immuno-oncology drug development goes global. Nature reviews. Drug discovery vol. 18 899–900 (2019).

2. Khadka, R. H., Sakemura, R., Kenderian, S. S. & Johnson, A. J. Management of cytokine re-lease syndrome: an update on emerging antigen-specific T cell engaging immunother-apies. Immunotherapy 11, 851–857 (2019).

3. Murthy, H., Iqbal, M., Chavez, J. C. & Kharfan-Dabaja, M. A. Cytokine Release Syndrome: Current Perspectives. ImmunoTargets Ther. 8, 43–52 (2019).

4. Doessegger, L. & Banholzer, M. L. Clinical development methodology for infusion-related reactions with monoclonal antibodies. Clin. Transl. Immunol. 4, e39 (2015).

5. Roselló, S. et al. Management of infusion reactions to systemic anticancer therapy: ESMO Clinical Practice Guidelines. Ann. Oncol. 28, iv100–iv118 (2017).

6. Bupathi, M. et al. Incidence of infusion reactions to anti-neoplastic agents in early phase clinical trials: The MD Anderson Cancer Center experience. Invest. New Drugs 35, 59–67 (2017).

7. Bernardez, B. et al. Real world of infusion related reactions in an oncology outpatient clin-ic. (2018).

8. Chung, C. H. et al. Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1, 3-galactose. N. Engl. J. Med. 358, 1109–1117 (2008).

9. Joerger, M. Prevention and handling of acute allergic and infusion reactions in oncology. Annals of oncology : official journal of the European Society for Medical Oncology vol. 23 Suppl 1 x313-9 (2012).

10. Merlano, M. C. & Garrone, O. Incidence of cetuximab-related infusion reaction in head and neck cancer patients: may we predict it? (2018).

11. Shimabukuro-Vornhagen, A. et al. Cytokine release syndrome. J. Immunother. cancer 6, 56 (2018).

12. Sampson, H. A. et al. Second symposium on the definition and management of anaphy-laxis: summary report--Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. The Journal of allergy and clinical immunol-ogy vol. 117 391–397 (2006).

13. Maude, S. L., Barrett, D., Teachey, D. T. & Grupp, S. A. Managing cytokine release syndrome associated with novel T cell-engaging therapies. Cancer J. 20, 119 (2014).

14. Le, R. Q. et al. FDA Approval Summary: Tocilizumab for Treatment of Chimeric Antigen Re-ceptor T Cell-Induced Severe or Life-Threatening Cytokine Release Syndrome. Oncologist 23, 943–947 (2018).

15. Lee, D. W. et al. Current concepts in the diagnosis and management of cytokine release syndrome. Blood 124, 188–195 (2014).

16. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) Ver-sion 5.0. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf (2017).

17. Park, J. H. et al. Long-Term Follow-up of CD19 CAR Therapy in Acute Lymphoblastic Leu-kemia. N. Engl. J. Med. 378, 449–459 (2018).

18. Neelapu, S. S. et al. Chimeric antigen receptor T-cell therapy - assessment and manage-ment of toxicities. Nat. Rev. Clin. Oncol. 15, 47–62 (2018).

19. Porter, D., Frey, N., Wood, P. A., Weng, Y. & Grupp, S. A. Grading of cytokine release syndrome associated with the CAR T cell therapy tisagenlecleucel. J. Hematol. Oncol. 11, 35 (2018).

20. Lee, D. W. et al. ASTCT Consensus Grading for Cytokine Release Syndrome and Neurologic Toxicity Associated with Immune Effector Cells. Biol. blood marrow Transplant. J. Am. Soc. Blood Marrow Transplant. 25, 625–638 (2019).

21. Davila, M. L. et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. Sci. Transl. Med. 6, 224ra25 (2014).

22. Wei, J. et al. The model of cytokine release syndrome in CAR T-cell treatment for B-cell non-Hodgkin lymphoma. Signal Transduct. Target. Ther. 5, 134 (2020).

FRIENDS
of CANCER
RESEARCH

Optimizing Dosing in Oncology Drug Development

FRIENDS OF CANCER RESEARCH
ANNUAL MEETING 2021

Introduction: Current dosing paradigm and ongoing challenges

Outside of oncology, most drugs are evaluated in randomized dose-ranging trials that support a broader understanding of the impact of different doses on efficacy and toxicity. In oncology, dose-finding studies are largely performed only in Phase 1 clinical trials and intended to identify the maximum tolerated dose (MTD), a dose initially developed for systemic chemotherapies. This paradigm relies on the notion that an increased dose leads to increased tumor suppression; therefore, the MTD is selected based on safety aspects focused primarily on tolerability.[1,2] With the advent of new molecular targeted agents (MTAs) and immunotherapies, oncology drug dose-finding approaches should be revised. In 2013, Friends of Cancer Research (*Friends*) released an issue brief on "Optimizing Dosing of Oncology Drugs" outlining strategies for optimizing dosing in oncology drug development while acknowledging key challenges and considerations (**Table 1**). Many of the challenges still persist in addition to nonoptimal approaches for dose selection.[3]

The continued focus on identifying and using the MTD may be driven by a desire for speed and misconceptions in the community. There is a notion that it is not worth performing randomized dose-finding clinical trials because they are too time consuming which may delay drug development and keep life-changing therapies from

Objectives

Describe current challenges to the implementation of dose-finding studies in oncology

Discuss opportunities to improve dosing strategies given ongoing challenges

Set expectations for dose-finding studies in the oncology pre-market setting

Identify key consider-ations for selecting appropriate dose optimization strategies in oncology

Thank You to Our Contributors

Gideon Blumenthal
Vice President, Oncology Global Regulatory Affairs, Merck

Lokesh Jain
Senior Director, Quantitative Pharmacology & Pharmacometrics
Oncology Clinical Development, Merck

Anne Loeser
Patient Advocate, Patient Centered Dosing Initiative

Yazdi K. Pithavala
Senior Director, Clinical Pharmacology, Pfizer Inc.

Atiqur Rahman
Division Director, Division of Cancer Pharmacology II,
Office of Clinical Pharmacology, U.S. FDA

Mark Ratain
Leon O. Jacobson Professor of Medicine,
The University of Chicago Medicine

Mirat Shah
Clinical Reviewer, Office of Oncologic Diseases, U.S. FDA

Laurie Strawn
Senior Director, Oncology Global Regulatory Affairs, Pfizer Inc.

Marc Theoret
Deputy Center Director, Oncology Center of Excellence, U.S. FDA

patients.Additionally, there is a misconception that a higher dose leads to higher efficacy and patients often anticipate that cancer treatments come with side effects. However, newer treatments like MTAs and immunotherapies often have target saturation limits below the MTD suggesting drugs can be given at lower doses with similar efficacy and potentially fewer side effects.

Friends convened stakeholders from industry, academia, the U.S. Food and Drug Administration (FDA), and patient advocacy groups to discuss the opportunities for optimizing dosing in oncology. This white paper highlights key findings from the discussions and aims to provide recommendations that precipitate a paradigm shift in oncology drug development to support adequate dose optimization studies. First, we provide strategies for overcoming the challenges outlined above, then highlight expectations for dose-finding studies, and lastly suggest key considerations for improved study design. While improved dosing methods and education are needed in the post-market setting, the recommendations provide focus on the pre-market setting to improve clinical trial design.

Strategies to overcome perceived challenges associated with the execution of appropriate oncology dose-finding studies

Perceived Challenge 1: Dose-finding studies are too time consuming and will prevent patients from quickly getting the drugs they need.

Dose-finding studies are extremely important to understand the therapeutic window of a drug and to ensure patients with cancer are optimally treated. These studies can be completed efficiently, with appropriate planning. FDA expects that sponsors perform dose-finding studies to evaluate exposure-response, efficacy, and safety and inform dose selection for registrational trials.[4]

Performing dose-finding studies in the pre-market setting builds a comprehensive foundation regarding the scientific reason for selecting a dose that not only provides more optimal treatment for patients, but also supports more seamless updates to the drug post approval. Applications for utilizing these data post approval include their use in combinations, adjustments in frequency of administration (e.g., Q3W to Q6W), and changes in the route of administration (e.g., intravenous to subcutaneous). Additionally, adequate dose-finding trials pre-approval may prevent clinical holds, the need for additional studies later in development, or post-marketing requirements if an inadequate dose is selected.

Moving ahead with an ill-optimized dose for the registrational trial can negatively impact the ability to document the true benefit of the drug. Identifying a dose with improved tolerability will lead to more patients missing fewer treatments due to toxicities while enabling them to remain on a working treatment for longer. In addition to individuals staying on treatment longer, a more appropriate dose may also provide an opportunity for additional patients with poor performance status to gain access.

To reduce potential delays in approval, discussions with FDA about dose-finding studies ideally would occur as soon as possible in drug development, as early as in the pre-IND setting. We have provided a list of questions to guide dosing discussions and strategies for different phases of development in **Table 2**. In addition to an early milestone meeting, sponsors could incorporate discussions about dosing in pre-IND meetings and consider additional meeting settings to discuss dose optimization. During the 2021 Beyond Breakthrough meeting hosted by *Friends*, a dosing snapshot was proposed for sponsors to help facilitate the exchange of key considerations and supporting evidence for dosing (**Appendix 1**).[5]

Perceived Challenge 2: Stakeholders believe that lower doses of drug are not as effective as higher doses.

Many involved in decision making for cancer treatment are accustomed to the MTD paradigm. They believe that higher drug doses will be more effective and fear that lower doses will lead to a subtherapeutic or less efficacious treatment regimen. Sponsors are encouraged to consider including an interim assessment and allowing intrapatient dose escalation (e.g., if the primary endpoint is based on early changes in tumor size metrics) into randomized trials of two or more doses to address the potential for underdosing. It is paramount that patient informed consent documentation clearly communicates the reasoning for various doses and explains that the lower dose was chosen based on data and modeling which inform its activity in a clinical trial. In addition, incorporation into clinical trials of tools to protect the patient interest, such as enabling treatment crossover or dose modifications based on interim analyses, is important.

A key solution in updating the approach for dose-finding studies is through stakeholder education. Concerns that lower doses lead to less efficacy can be mitigated through educating patients and providers about the value of using a lower dose, when appropriate, especially for MTAs. Sponsors should understand the utility of appropriate dose-finding trials in the long-term to support further analysis of their products and ensure that the greatest number of patients benefit from their product. An additional avenue of education about dose-finding trial design is enhanced guidance from FDA and how the agency plans to engage with sponsors in selecting the dose for oncology registrational trials.

Educating providers and patients about the outcomes from dose-finding studies could also support an understanding of how lower doses do not necessarily always lead to lower efficacy. Sponsors could be encouraged to publish the results from their dose-finding studies, including certain aspects of the process and rationale for selecting the dose used in the registrational trial. Clinical management guideline developers could consider incorporating information about different doses, including dose reductions, that provides data driven insights about the impact of different doses on safety and efficacy.

Expectations for dose-finding study designs and methodology

The goal of dose-finding studies is to adequately characterize the exposure/safety relationship as well as the exposure/activity relationship to select a dose that will be brought into the registrational trial and ultimately used post approval.[6,7] Establishment of a therapeutic window based on activity and an acceptable level of toxicity, derived from a characterization of pharmacokinetic (PK)/exposure and pharmacodynamic (PD) metrics is integral. Dose-finding trials that effectively and efficiently evaluate at least two doses in a randomized manner are increasingly important for dose selection.

Selecting PK and PD Metrics. Sponsors can use pre-clinical data to define target saturation points and exposure to narrow the range of doses for further clinical evaluation. It is often very helpful to identify biomarkers that translate from animal models or protein modeling into the clinical trial design to characterize PK and PD metrics in addition to those used in patient selection, if appropriate.

When biomarkers are assessed in patients, it is important that the biomarker is well defined. The most appropriate biomarkers are blood-based or imaging biomarkers rather than biopsies to estimate dose-response, especially given the American Society of Clinical Oncology (ASCO) guidelines around such biopsy studies.[8,9] Tumor biopsies are also problematic because the sources of variability are rarely identified and may be influenced by the time of sampling. Blood-based biomarkers can be easily assayed at multiple timepoints but may be less informative than evidence of radiographic improvement.

Well-defined biomarkers can support an understanding of dose-response relationships along with the totality of safety and activity data. Biomarkers that measure activity include those that track the relationship between plasma exposure and change in tumor endpoints such as Response Evaluation Criteria in Solid Tumors (RECIST) measurements. Analysis of tumor dynamics (e.g., depth and duration of tumor change from baseline) as a function of dose using a modeling approach can support an understanding of activity.[10] It is important to consider the nature of the disease when identifying biomarkers, as certain solid tumors (e.g., lobular breast cancer) may not be measurable by RECIST and hematologic cancers will have different measurements than solid tumors. Safety can be tracked through blood biomarkers like neutrophil counts when applicable.

The Dose-Finding Trial Design. Ideally, the pre-registrational dose-finding study would be randomized, compare at least two doses, and confirm the dose selected for the registrational trial, which is the dose that maximizes benefit-risk by measuring efficacy among a sizeable number of patients. The randomized dose-finding trials do not necessarily need to be powered to conduct a rigorous statistical comparison across doses; however, it is important that the trial is sufficiently sized to understand the general shape of the dose/exposure-activity/toxicity relationships, including the minimally active dose. To save time

but provide robust data for multiple doses, sponsors could consider pre-registrational trial protocols that extend monitoring of patients after the registrational trial starts to allow for the long-term characterization of patients treated with different doses.

Important considerations when choosing the doses for comparison in the pre-registrational dose-finding study include selecting doses that are pharmacokinetically distinguishable and do not have overlapping PK exposures (i.e., doses that are 2-3 fold apart). The lowest dose is the minimal dose expected to provide activity based on PK/PD analyses, and the highest dose (chosen within safety allowance) is selected to ascertain whether dose increases result in increased activity with acceptable toxicity.

After the completion of the randomized dose-finding trial, data from this trial can be analyzed to characterize the exposure-response relationships for activity and safety and then integrate with the previous PK/PD analyses results to inform the final dose(s) for the registrational trial. If there is a clear differential benefit with acceptable safety compared to a lower dose, the higher dose can be selected for the registrational trial(s). If the efficacy and safety are similar between the lower and higher doses, the lower dose can be selected as the final dose for the registrational trials.

Key considerations for dose optimization strategies

The study design for determining the optimal dose will differ depending on the product, the target population, and the data that are available. There are key considerations when designing these studies (more details are included in **Appendix 2**):

- *Therapeutic properties.* Differences in the properties of drugs (e.g., small molecule vs. large molecule, agonist vs. antagonist) influence the way drugs interact with the body in terms of safety and efficacy. The selection of the initial doses for the dose-finding studies as well as methods for determining which dose to move into registrational trials are influenced by the therapeutic properties.

- *Patient populations.* There is heterogeneity in patient populations based on tumor type, disease stage, and comorbidities. Especially in the context of expanded clinical trial populations, an understanding of how various factors influence the efficacy of the drug may provide justification for adjusting the dose accordingly.

- *Supplemental vs. original approval.* The differences in disease characteristics and patient populations between tumor types and treatment settings (e.g., monotherapy vs. combination therapy) are important to consider in determining whether any additional dose exploration is necessary for a supplemental application. In instances where further dose exploration may be needed, the study design can incorporate prior understanding of exposure-response from the original approval.

Conclusions and future directions

In conclusion, randomized studies that formally evaluate at least two doses to support dosing decisions are increasingly important in oncology rather than using MTD as the default approach. These studies will improve care in oncology by decreasing toxicities while maintaining efficacy and ultimately allow for more patients to benefit from treatments for a longer period of time. The findings in this white paper provide considerations and expectations for dose-finding studies that offer opportunities for improved patient care.

In the short-term, continued education will support a realization of the value of these studies in the pre-market setting. Patients and providers should understand that treatment with higher doses of oncology therapies is not always better and may, in fact, lead to increased side effects without the added benefit of higher activity. Sponsors should recognize the long-term benefits of these trials and appreciate that FDA has established the expectation to incorporate dose-finding studies in the drug development paradigm sooner.

To complement this, FDA has encouraged sponsors to discuss their dose-finding trial design early in clinical development, as supported by available clinical pharmacology data. FDA's focus has shifted over the past few years to encourage companies to have conversations about their drug development pipeline earlier in development. Additionally, an appreciation for cross-disciplinary discussions has led to an increase in interdisciplinary interactions to address dosing considerations early on. Sponsors should conduct pre-clinical research that supports a basic understanding of pharmacology based on suggestions from MAPPs and guidance documents. After performing pre-clinical work and establishing necessary data, sponsors should engage FDA to refine and build the dose-finding trial. The use of a dosing snapshot (Appendix 1) would likely support more targeted discussions.

How the data from dose-randomized trials are included in drug labels require further discussion among stakeholders (FDA, industry, patients, providers). There may be an opportunity to include data in labeling that may help patients and providers understand the range of efficacy and toxicity as it relates to dose. It may also be helpful to expand on what is included about different doses for different patient populations such as those with altered organ function or pharmacogenetics.

The overarching goal is that dose-finding studies will be a part of standard oncology drug development in the pre-market setting to allow delivery of efficacious and tolerable doses to patients at initial marketing approval of a new drug. Meetings held with sponsors on dose-finding and dose selection as early as possible in development provides an opportunity for the agency to convey their expectations sooner, potentially leading to more efficient studies. Communication of data from dose-ranging trials in drug labels or publications can support shared decision making between patients and providers about dosing choices. Also, rather than patients and providers expecting debilitating side effects, side effects would be regarded as possible but not inevitable. Ultimately, updating dosing regimens should allow patients to be on drugs providing benefit with fewer toxicities for longer and miss fewer treatments due to toxicities.

Table 1: Findings from *Friends'* 2013 White Paper "Optimizing Dosing of Oncology Drugs"	
Proposal	**Suggestion**
Path for study	1. Phase 1 trials should include adequate PK sampling to enable a clear determination of the PK properties of the drug and preliminary characterization of dose-exposure relationships. When feasible and appropriate, PD endpoints should be incorporated to determine the drug exposure that results in inhibition of the drug target. 2. Phase 2 trials should go beyond assessment of drug activity and could include adaptive designs and/or randomized exploration of doses. Continued, sparse PK sampling should be included to gain a sense of relationships between exposure and clinical outcomes. If possible, measurements of PD endpoints should also be continued. 3. Phase 3 trials should incorporate population PK sampling to further evaluate the relationship between covariates influencing exposure and key clinical outcomes. 4. When subjective toxicities are identified in phase 1 trials, patient-reported outcomes (PROs) should be assessed using validated tools if available in phase 2 and phase 3 trials and could be used to guide dose optimization. 5. The PK and PRO dataset collected in phases 1-3 could be used to develop an approach to therapeutic drug monitoring in the post-market setting. This will enable the dose for an individual patient to be adjusted as needed based on observed drug exposure, treatment tolerance and clinical status.
Necessary data elements	1. Sponsors should collect PK and exposure data in oncology phase 2 and 3 clinical trials to estimate a therapeutic index for a defined patient population. Randomized dose comparison studies should be included in phase 2 studies and exposure-response analyses should be performed to better inform the selection of dose for phase 3 registrational trials. 2. PROs should also be collected to understand the patient experience more fully with a drug. PROs can be informative not only of the side-effects of a drug, but also of any beneficial effects a drug may have on symptoms of the cancer itself.
How to integrate data elements	In the proposed approach, the collection of exposure data and data regarding tolerability across a range of doses could enable the definition of a threshold exposure needed for anti-tumor effect as well as the determination of a peak exposure that correlates with excess toxicity. Collection of drug exposure and tolerability data, as well as ongoing evaluation of adverse events and dose modifications, from patients in real-world settings may be useful for post-market evidence generation.
Optimal timing of dose comparison studies	Ideally, randomized dose comparison studies and exposure-response analyses would be performed in the pre-market setting.

Table 2: Dosing Questions by Stage of Drug Development	
	Key Questions
Pre-clinical	• What is the best model to identify the initial dose? • Is there established pharmacological and a dose-pharmacology relationship evidence? • Which biomarkers should be evaluated in the clinical trials to monitor safety? To monitor activity? • What enzymes metabolize the drug? How do polymorphic enzymes influence trial design? • For oral drugs, what is the Biopharmaceutical Classification System (BCS) classification of the drug?
Early phase trial	• How do the PK and PD characteristics justify the dosing interval? • Are there any intrinsic or extrinsic factors that would influence PK? • What is the degree of PK variability, considering both interindividual and intraindividual variability? • Are there any drug interactions that need to be evaluated? • For oral drugs, should the drug be administered with food? • For oral drugs, is there a better time of day to administer (AM vs. PM)?
Prior to conducting trial intended for drug approval*	• What is the relationship between dose/exposure and activity? • What is the relationship between dose/exposure and toxicity? • Are there concerns for chronic or delayed toxicities and have these been considered when evaluating dose/exposure-toxicity? • Is the dose schedule justified based on the Kinetics-PK-PD or modeling approaches? • Is the dosing regimen justified based on dose/exposure-response relationships and other relevant data?
Registrational Trials*	• Does the dosing regimen continue to demonstrate acceptable benefit-risk?
Post-market	• Are there unexpected toxicities which necessitate a re-evaluation of the dosing regimen? • Are there opportunities to optimize the dosing regimen for convenience (e.g., extended dosing interval, new route of administration, etc.)?

*For products with expedited development (for example: Breakthrough Therapy-designated products) these two phases could potentially be combined

References

1. Rogatko A, Babb J, Tighiouart M et al. New Paradigm in Dose-Finding Trials: Patient-Specific Dosing and Beyond Phase I. Clinical Cancer Research 2005;11:5342–5346.

2. Mathijssen R, Sparreboom A, Verweij J. Determining the Optimal Dose in the Development of Anticancer Agents. Nature Reviews Clinical Oncology 2014;11:272–281.

3. Minasian L, Rosen O, Auclair D et al. Optimizing Dosing of Oncology Drugs. Clinical Pharmacology and Therapeutics 2014;96:572–579.

4. Shah M, Rahman A, Theoret MR, Pazdur R. The Drug-Dosing Conundrum in Oncology — When Less Is More. The New England Journal of Medicine 2021;385:1445–1447.

5. Friends of Cancer Research. Beyond Breakthrough: Optimizing the Breakthrough Therapy Designation. 2021. Available at: https://friendsofcancerresearch.org/sites/default/files/2021-09/Beyond-Breakthrough-Optimizing-the-BTD-Draft.pdf

6. Ratain MJ. Redefining the Primary Objective of Phase I Oncology Trials. Nature Reviews Clinical Oncology 2014;11:503–504.

7. Ratain MJ, Tannock IF, Lichter AS. Dose Optimization of Sotorasib: Is the US Food and Drug Administration Sending a Message? Journal of Clinical Oncology 2021;39:3423–3426.

8. Sweis RF, Drazer MW, Ratain MJ. Analysis of Impact of Post-Treatment Biopsies in Phase I Clinical Trials. Journal of Clinical Oncology 2016;34:369–374.

9. Levit LA, Peppercorn JM, Tam AL et al. Ethical Framework for Including Research Biopsies in Oncology Clinical Trials: American Society of Clinical Oncology Research Statement. 2019;37:2368–2377.

10. Bruno R, Mercier F, Claret L. Evaluation of Tumor Size Response Metrics to Predict Survival in Oncology Clinical Trials. Clinical Pharmacology & Therapeutics 2014;95:386–393.

11. Mandrekar S, Cui Y, Sargent D. An Adaptive Phase I Design for Identifying a Biologically Optimal Dose for Dual Agent Drug Combinations. Statistics in Medicine 2007;26:2317–2330.

12. Re M del, Omarini C, Diodati L et al. Drug-Drug Interactions Between Palbociclib and Proton Pump Inhibitors May Significantly Affect Clinical Outcome of Metastatic Breast Cancer Patients. ESMO Open 2021;6:100231.

Appendix 1: Drug Development Snapshot Template—Clinical Pharmacology (Dose & Administration) Snapshot.

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule, and route of administration	• *What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding?* • *When do you anticipate that a R2PD will be selected?* • *Are other routes of administration being investigated?*
Mechanism of action (MOA) and format	• *Is the therapeutic a small or large molecule? Another platform? What is the MOA?*
Translational evidence	• *Is there established pharmacological evidence (e.g., target engagement, outcome based biomarkers, tumor volume) in the relevant pre-clinical species?* • *Is the dose-PK relationship established in the non-clinical species (i.e., is the PK dose proportional)?* • *Are the pharmacological/efficacious target concentrations for patients defined?* • *Is the dose/exposure-response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?*
Clinical Evidence	

Clinical studies	• *List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)*
	• *Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:*
	  • *Dose escalation, expansion cohorts with or without randomization*
	  • *Single arm randomization (i.e., dose and/or control); adaptive design*

PK characteristics	• *Is the dose–PK relationship well established (i.e., is the PK dose proportional)?* • *Do the PK characteristics (accumulation, half-life) justify the dosing interval?* • *Are there any intrinsic or extrinsic factors (e.g., food, body weight, immunogenicity) that would majorly influence PK (i.e., if these warrant dose adjustments in a subset of patients)?* • *Was the PK variability considered when selecting a dose that would achieve target exposure for most patients?*
Safety summary	• *Is the dose–PK relationship well established (i.e., is the PK dose proportional)?* • *Summary of frequencies of key AEs (including chronic low grade AEs, which can affect tolerability) of interest by dose* • *Is there a dose/exposure-safety or PK-PD relationship, upon the adjustment of potential covariates, for safety? If yes, what is the nature of the relationship?* • *Summary of dose interruptions, reductions, and discontinuations by dose/exposures* • *Is there an increased frequency of dose interruptions or reductions or treatment discontinuations with increasing doses/exposures?* • *Are there any late occurrence toxicities beyond the DLT period? Are there early PD biomarkers reflective of the delayed safety endpoints?* • *Are there any overlapping toxicities with the concomitant medications in the patient population (e.g., treatment combinations for NME with SOC and/or treatments for comorbidities/cancer-related symptoms)?* • *Is there an increased frequency of dose interruptions, reductions, or treatment discontinuations with increasing doses/exposures?* • *If acute/transient toxicities were observed, were alternative dosing approaches considered (e.g., step-up dosing)?* • *Do existing data indicate this is a narrow therapeutic window drug with dose limiting toxicity that is monitorable (e.g., biomarkers, BP, HR, neuropathy)?* • *If yes, does this drug provide an opportunity to personalize the dose for an individual patient or a sub-population based on the emerging monitorable toxicity?*
Efficacy summary	• *If yes, does this drug provide an opportunity to personalize the dose for an individual patient or a sub-population based on the emerging monitorable toxicity?* • *Summary of response endpoints by dose (e.g., ORR, PFS)* • *Is there a dose/exposure – efficacy (primary efficacy endpoint) and PK-PD (e.g., mechanism of action/predictive biomarkers) relationship upon the adjustment of potential confounders? If yes, what is the nature of the relationship?* • *Is the dose schedule (e.g., frequency, dose holidays) justified based on the K/PK-PD and/or QSP modeling approaches?* • *Are the relevant exposure metrics for efficacy identified (e.g., AUC, C_{max}, C_{min}, concentration-time, RO)?*

Other considerations	• *Are there any manufacturing considerations (e.g., pill burden, maximal feasible dose, etc.) that need to be considered?* • *Are there any patient factors that need to be considered (e.g., patient convenience/compliance [QD, BID, TID), QW vs Q3W, SC vs IV)?* • *Complimentary M&S approaches (i.e., PK-TGI/QSP/ML, etc.) for dose optimization and/or inform dose adjustments*
Additional Clinical Evidence	
Planned clinical studies	• *Are there additional planned clinical studies that will contribute data to the current D&A plan/rationale or future D&A proposals?*
Other evidence	• *Does additional scientific evidence exist (e.g., from similar class, MOA, or indication) that may support the current D&A plan/rationale (e.g., publications, scientific presentations)?*

Abbreviations: AEs=adverse events; AUC=area under the curve; BP=blood pressure; C_{max}=maximum 'peak' concentration; C_{min}=minimum 'trough' concentration; D&A=dose & administration; DLT=dose-limiting toxicity; HR=heart rate; K=kinetic; MOA=mechanism of action; NME=new molecular entity; ORR=overall response rate; PD=pharmacodynamic; PFS=progression-free survival; PK=pharmacokinetic; QSP=quantitative systems pharmacology; RO=receptor occupancy; SOC=standard of care

Appendix 2: Expanded Key Considerations for Dose-Finding Studies

Therapeutic properties. Differences in the chemical structure of drugs influences the way it interacts with the body in terms of safety and efficacy. The selection of the initial doses for the dose-finding studies as well as methods for determining which dose to move into registrational trials are influenced by the therapeutic properties.

- *Large molecules.* Antibodies have the potential for demonstrating false positive exposure/response relationships with single-dose data due to the impact of confounding factors such as patient health status (i.e., cachexia) on survival. Ascertaining the Target-Mediated Drug Disposition (TMDD) during the dose escalation stage is important as it provides information about target expression and target turnover.
- *Antagonists.* For antagonist monoclonal antibodies, consider a dose that attains target engagement (TE) of >90% in systemic circulation and in tumors (where required). Assess data on target saturation in tumor, which can be informed by approaches like physiologically-based pharmacokinetic (PBPK) models.
- *Agonists.* Unlike antagonist monoclonal antibodies, a high level of receptor occupancy may not be necessary for agonists to elicit a maximum pharmacological effect. PK/PD analysis on biomarker data can help in determining the level of receptor occupancy needed for therapeutic effect.
- *Non-traditional therapies.* Specific consideration may be required for antibody drug conjugates (ADCs), bispecific antibodies, and cell therapies. ADCs have relatively narrow therapeutic indices and require optimization of both dose and dosing frequency to reduce toxicity. For bispecific antibodies, it may become challenging to optimize target engagement

for two targets. Efficacy and on-target toxicity of bispecific antibodies may be driven by trimer formation (ternary complex) between bispecific antibody, T cell, and tumor cell. These ternary complexes usually have a bell-shaped exposure-response relationship, and it is important to determine optimal concentrations of bispecific antibodies that maximize formation of trimer formation for maximal pharmacological activity. For cell therapies, cellular kinetics models are used to describe the relationship between the number of cells infused and expansion of modified T-cells in vivo. The understanding of cell kinetics along with measurable PD response informs the selection of dose of cell therapy.

- *Combination regimens.* In combination trials, the dose of each drug in the combination is often based on the MTD of each drug, rather than considering their additive toxicities and efficacies.[11] Doses should be selected based on maximum pharmacology (and not MTD) with special consideration when treatments have overlapping toxicities.
- *Drug-drug interactions.* Specific consideration may be required based on drug-drug interactions and effect of renal and hepatic impairment on PK. Sometimes drugs depend on pH for solubility so differences in body chemistry or use of proton pump inhibitors may impact efficacy.[12]

Patient populations. There is heterogeneity in patient populations based on tumor type, disease stage, and comorbidities. Especially in the context of expanded clinical trial populations, an understanding of how various factors influence the efficacy of the drug may provide justification for adjusting the dose accordingly.

- *Small molecules.* Drugs that are taken orally may have different bioavailability in a fed versus fasted states. If there is a food effect, then those prandial conditions that reduce bioavailability should be avoided (since that often increases GI toxicity).
- *Heavy pre-treatment.* Patients in Phase 1 dosing trials tend to be heavily pre-treated and have strict inclusion/exclusion criteria, which often differs from the average patient who will use the drug in registrational trials or in the real world and thus may impact metabolism or tolerability of the drug.
- *Altered organ function.* Some patients with cancer have altered end-organ function either due to their disease or previous treatments. Others may have differences in pharmacogenetics, specifically genetic polymorphisms in drug transporters or metabolizing enzymes which may ultimately impact drug clearance.
- *Changes in tolerability.* Some patients have changes in tolerability over time. Dosing efficacy may be impacted by age and bodyweight. Clearance may also change over time, particularly for monoclonal antibodies.
- *Tumor stage.* Clearance of monoclonal antibodies may be different between patients with metastatic disease and without metastatic disease. The latter patients will have lower clearance, and thus a lower dose may be effective than for patients with advanced disease.
- *Long-term treatment.* In the metastatic setting, patients can be treated regularly for years, so identifying the optimal dosing regimen is important since longer term safety is an issue. Special consideration should be taken for chronic treatment use to avoid buildup of toxicities.

Expedited Development Programs at the Food and Drug Administration: Insights and Opportunities

Grace Collins[1] · Mark Stewart, PhD[1,2] · Ellen Sigal, PhD[1] · Jeff Allen, PhD[1]

Timely access to therapies that treat serious illnesses is critical for patients, particularly for those with rare or serious disease types that have no current treatments. Congress and the Food and Drug Administration (FDA) have addressed this by periodically establishing programs designed to expedite different steps associated with the development and review of drugs and biologics. Over time, these efforts have been effective in getting new treatments to patients faster than traditional approval processes [1]. Here, we present an analysis of original therapeutic oncology agents approved between January 1, 2013, and September 4, 2020, to understand how expedited programs are utilized in oncology, a disease area where these pathways have been utilized the most.

We analyzed development timelines using publicly available FDA review documents through the online database Drugs@FDA. We compiled Investigational New Drug (IND) Application submission dates, New Drug Application (NDA) and Biologics Licensing Application (BLA) receipt dates, approval dates, and noted use of Fast Track, Breakthrough Therapy Designation, Priority Review, and Accelerated Approval. During this 7.5-year period, the FDA approved 98 original oncology treatments [2]. To gain a comprehensive view of how expedited programs are utilized, we examined a time period during which all pathways were active, beginning in 2013 when the most recently established pathway, Breakthrough Therapy Designation, became available for use. We note that the number of oncology drug approvals over the past 4 years (2017–2020 | $n = 61$) increased 65% compared to the 4 years before that (2013–2016 | $n = 37$). The two most used pathways were Priority Review (86%, $n = 84$) and Breakthrough Therapy

✉ Mark Stewart
    mstewart@focr.org

1   Friends of Cancer Research, Washington, DC, USA

2   1800 M Street NW, Suite 1050 South, Washington, DC 20036, USA

Designation (54%, $n = 52$), and 92% ($n = 90$) of all approvals used at least one expedited pathway (Table 1). Our analysis shows expedited pathways were rarely used alone. 76% of expedited approvals were approved using a combination of two or more expedited pathways (Fig. 1).

When comparing the median development time (IND submission to NDA/BLA submission) of novel agents using expedited programs (2013–2020 | $n = 90$) to novel approvals that used traditional approaches (2013–2020 | $n = 8$), we found the use of expedited programs reduced the median development time by 3.4 years and shortened median review time by 4 months (Fig. 2). For approvals that used only one expedited program ($n = 16$), the median time to development was 9.62 years, compared to 5.76 years for those approved using two or more expedited approaches ($n = 74$).

By evaluating the processes associated with these programs based on the wealth of experience gained over the past decade in oncology, insights into their effectiveness and opportunities for improvement emerge. The shifting utilization and utility of these pathways must be considered in the context of current scientific capabilities and cutting-edge drug development procedures. While Priority Review has consistently been used since 2013, the percent of approvals using Fast Track significantly decreased overtime. More than half of expedited approvals in our analysis used a combination of expedited programs as opposed to one alone. For example, of the 52 drugs that used Breakthrough Therapy Designation and Priority Review, 19 used Fast Track, and 34 used Accelerated Approval. While there are overlapping benefits when using these programs in combination, there are also duplicative application and administrative processes for those with similar requirements. Despite these redundancies, sponsors continue to use programs in combination. To ensure optimal use of all programs—alone or in combination—FDA's resources must be allocated most efficiently and developers' processes optimized. To that end, it may be worth creating a more streamlined process to avoid redundancy in the administrative processes associated
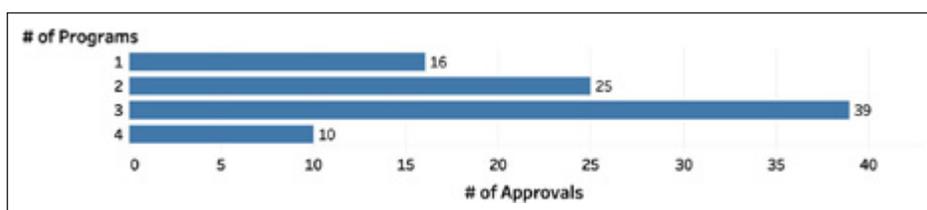
Table 1 Utilization of expedited programs by year for oncology drug approvals

	Year of approval (n=# of approvals)							
	2013 (n=8)	2014 (n=8)	2015 (n=16)	2016 (n=5)	2017 (n=16)	2018 (n=17)	2019 (n=10)	2020 (n=18)
Fast Track	7 (87.5%)	4 (50.0%)	7 (43.8%)	2 (40.0%)	7 (43.8%)	7 (41.2%)	3 (30.0%)	6 (33.3%)
Breakthrough Therapy Designation	2 (25.0%)	5 (62.5%)	5 (31.3%)	5 (100.0%)	12 (75.0%)	6 (35.3%)	6 (60.0%)	11 (61.1%)
Priority Review	6 (75.0%)	8 (100.0%)	12 (75.0%)	5 (100.0%)	14 (87.5%)	15 (88.2%)	9 (90.0%)	15 (83.3%)
Accelerated Approval	2 (25.0%)	7 (87.5%)	5 (31.3%)	4 (80.0%)	6 (37.5%)	3 (17.6%)	7 (70.0%)	10 (55.6%)

Percentages total greater than 100% because multiple expedited programs can be used for a single drug. Expedited pathways were established overtime to reflect the evolving and modernizing landscape of regulatory science. Priority Review and Accelerated Approval were established in 1992, Fast Track in 1997, and, most recently, Breakthrough Therapy Designation in 2012

Fig. 1 Utilization of expedited programs alone or in combination for oncology drugs. Expedited programs are rarely used in isolation and are often combined with one or more other expedited programs
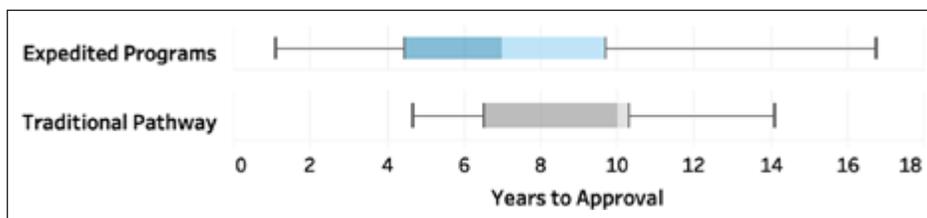
Fig. 2 Median years to approval for oncology drugs utilizing expedited programs versus the traditional approval pathway. Use of expedited programs shortened median time to approval for qualifying drugs (Expedited Development=6.58 years) compared to drugs that do not qualify for an expedited program (Traditional Pathway=10 years)

with pathways frequently used together. A more streamlined process would maximize resources and time for both FDA and Sponsors to continue driving innovation and prioritize development of promising drugs.

In summary, expedited mechanisms effectively facilitate development and review processes and shorten time to approval for original therapeutic approvals in oncology. They collectively help provide effective new treatments to patients faster than the traditional approval processes. Many of these new therapies have since demonstrated long term population-level benefits, such as a significant reduction in overall lung cancer mortality [3]. A delay for these therapies would result in a lag in such benefits for potentially thousands of patients. The frequent use of expedited programs in oncology provides a wealth of experience and learning

which can be applied to optimize the use of expedited programs in other serious disease spaces where there are few or no available treatments.

To ensure the benefit of these pathways evolves to reflect current science and technological capabilities, it is essential that FDA has adequate resources to optimize these expedited programs. As steps for the seventh reauthorization of the Prescription Drug User Fee Act begin, there is an opportunity to modernize approaches to reflect the current state of drug development and regulation. A periodic review of these programs will allow planning for future capacity and ensure processes associated with their use are optimized. In an era where we will likely see an expansion of emerging new therapies [4, 5] aimed at treating serious and life-threatening diseases, it is critical to ensure expedited programs

continue to facilitate the science, provide appropriate access for patients, and are sustainable.

Compliance with Ethical Standards

Conflict of interest
The Authors declare that they have no conflicts of interest.

Open Access

References

1. US Food and Drug Administration. Expedited programs for serious conditions—drugs and biologics 2014. https://www.fda.gov/media/86377/download. Accessed 17 Dec 2020.
2. Friends of Cancer Research. Drug Development Dashboard 2020. https://friendsofcancerresearch.org/drug-development-dashboard. Accessed 17 Dec 2020.
3. Howlader N, Forjaz G, Mooradian MJ, et al. The effect of advances in lung-cancer treatment on population mortality. N Engl J Med. 2020;383(7):640–9.
4. Yu JX, Hubbard-Lucey VM, Tang J. Immuno-oncology drug development goes global. Nat Rev Drug Discov. 2019;18(12):899–900.
5. Yu JX, Upadhaya S, Tatake R, et al. Cancer cell therapies: the clinical trial landscape. Nat Rev Drug Discov. 2020;19(9):583–4.

Beyond Breakthrough: Optimizing the Breakthrough Therapy Designation

Introduction:

Advances in our understanding of disease processes, genetics, manufacturing technologies, and innovative clinical trial design have enabled the development of novel therapeutic agents for the treatment of patients with cancer. In oncology, the ability to target a novel agent against a driver oncogene or protective immune checkpoint has led to several therapeutic breakthroughs in diseases with limited or no systemic treatment options. These breakthroughs have established new classes of therapeutics leading to, in some instances, unprecedented improvements in clinical outcomes for patients with cancer.

Regulatory review processes are time and resource intensive for drug sponsors and the United States Food and Drug Administration (FDA). The FDA leverages several tools to safely and efficiently facilitate development and review of agents intended for treatment of patients with life-threatening conditions without compromising the rigorous standards established for their approval.

Breakthrough Therapy Designation (BTD) facilitates the efficient development of both drugs and biologics (hereafter referred to as "drugs") intended to treat serious or life-threatening illnesses for which there is preliminary clinical evidence demonstrating that the investigational therapy may offer substantial improvement on a clinically significant endpoint(s) over available therapies.[1] BTD provides

Objectives

Characterize the rate-limiting steps and challenges encountered throughout the development of oncology products with Breakthrough Therapy Designation (BTD).

Propose recommendations that address commonly encountered challenges and identify best practices to ensure development programs maximize the benefits of BTD.

Delineate key topics and optimal timing for interactions with the FDA when requesting BTD (pre-BTD) and after receiving BTD (post-BTD).

Acknowledgements

sponsors with early opportunities for FDA interaction and enhanced guidance, including proactive organizational commitment and coordination involving senior FDA managers and experienced regulatory project management staff. Additionally, BTD often offers a pathway to eligibility for rolling or priority review.[2]

Since 2012, the number of BTD requests each year has increased. To date, the FDA has received over one thousand requests for BTD and granted more than four hundred requests.[3] Both FDA and commercial sponsors prioritize internal resources to help ensure that the most promising products receive BTD and undergo clinical development as efficiently as possible without compromising safety, efficacy, or quality. As a result of this work, BTD has facilitated timely development and approval of 205 products, 61% of which are oncology products. The program has been particularly successful in getting safe and effective novel treatments approved for patients with cancer, particularly new treatments that have not been previously approved for other indications. It is estimated that BTD has shortened the time from IND submission to approval for 60 original applications (not previously approved) for oncology products by a median of 2.3 years compared to oncology products without BTD.[4]

Historically, much of the focus on BTD has been on the qualifying criteria and processes leading to receipt of BTD; however, given the breadth of experience with BTD in oncology over the past eight years, stakeholders now have the opportunity to identify pressure points and best practices to help optimize its implementation in order to better support efficient, successful development of new safe and effective cancer treatments. To this end, Friends of Cancer Research (*Friends*) convened a multistakeholder working group to conduct a landscape analysis to identify opportunities for improved implementation of the BTD program in oncology. This effort focused on evaluation of the successful use of the BTD program and challenges associated with optimal use of BTD in oncology in order to help inform strategies for sustaining its impact on drug development.

Friends conducted a survey soliciting input from over 20 commercial sponsors that varied in company size and the range of experiences in terms of number of therapies that have received BTD and approved therapies with BTD to identify challenges and formulate recommendations for optimizing the use of BTD. All sponsors noted the positive impact of BTD in oncology drug development. Several key areas with the potential to further optimize use of BTD emerged.

1. **Clarify expectations for necessary evidence to receive BTD:** Sponsors may find it challenging to anticipate when and what to submit with preliminary and formal BTD requests. There is also a need for additional clarity about the types and quantity of evidence needed to support BTD, particularly with respect to early preliminary clinical evidence in oncology.

2. **Enhance communication between sponsors and FDA:** The opportunity for enhanced interactions with FDA provided by BTD support proactive identification and resolution of issues; however, at certain stages of development, such as pre-BTD or between milestone meetings, there may be additional opportunities to streamline clinical development through sponsor–FDA interaction.

3. **Support inter-disciplinary coordination and improve transparency:** In oncology, many drug development programs are increasingly complex and require close coordination between multiple disciplines, and often between Offices and Centers at the FDA (e.g., applications with BTD involving complimentary or companion diagnostics, employing a novel platform or end-point, targeting rare diseases, and/or incorporating innovative trial designs). Additional guidance on how best to strengthen coordination between Centers, when relevant, at the time of BTD may be beneficial to ensure complexity does not lead to delays.

4. **Provide additional support to address rate-limiting steps in drug development:** BTD facilitates both clinical and CMC aspects of development; however, challenges can occur during drug development, particularly for novel or innovative technology platforms. CMC development issues can be the rate-limiting step in drug development and approval, and flexibility and timely interactions between FDA and sponsors can be crucial to identify and resolve these issues to mitigate delays. Dose selection and justification is also an important component of drug development and can be a challenge with an expedited drug development timeline.

After conducting the survey, *Friends* convened multiple focus groups with key stakeholders, including the FDA, to identify potential practical solutions in these key areas to support optimal use of BTD. A summary of the outcomes of these focus group discussions is provided below.

Opportunity 1. Optimize the timing of BTD and improve communication on expectations for data necessary to receive and maintain BTD.

Provide additional clarity regarding the criteria for BTD to optimize the timing for submission of a BTD request.

BTD has the most potential to positively impact development of drugs that have not previously received FDA approval for another indication because it confers opportunities for enhanced interactions between sponsors and a multidisciplinary FDA team including senior FDA staff. These interactions can help address critical aspects of drug development such as dose optimization and manufacturing, which can be rate-limiting. Similarly, the timing of BTD is critical; for example, when BTD is granted based on top-level results of a pivotal trial that will provide the primary evidence to support a marketing application, there may be limited potential for enhanced interactions to result in meaningful improvements to drug development, as opposed to when BTD is granted prior to initiation of a registrational trial.

It is important for sponsors to apply for BTD at a time when there is sufficient data to meet the qualifying criteria for BTD but early enough to fully leverage the enhanced interactions provided by BTD ideally no later than the time of completion of Phase 2 development. Among sponsors, there can be uncertainty regarding the level of clinical evidence required to meet the qualifying criteria for BTD. Routine use of preliminary BTD advice teleconferences to discuss eligibility of requests for BTD and gain a better understanding regarding the

appropriate timeline and data package necessary to support a BTD request facilitates timely submission and review of BTD applications. Inclusion of additional annotation in the preliminary BTD teleconference template to describe the type and scope of preliminary evidence that are generally needed to support a BTD request submission for an oncology product could help facilitate preparation of documents and meaningful preliminary BTD teleconference discussions, while also reducing the number of preliminary BTD teleconferences that clearly lack sufficient data. An oncology-specific guidance describing general guidelines for preparing for a preliminary BTD discussion, content of a BTD request submission, and efficacy considerations for meeting the criteria for BTD may be beneficial.

Clarify procedures and decision-making regarding withdrawing or rescinding BTD and better understand its downstream impact on development/approval to support integrity of the program.

If a program no longer qualifies for BTD, the sponsor can voluntarily withdraw their BTD, or FDA can rescind it. It is sometimes unclear what the timepoints are for re-evaluating the status of BTD or common reasons for withdrawal. Transparent communication regarding the considerations and procedures used by FDA to evaluate whether BTD should be rescinded or withdrawn could ensure a designation is robust and fair. There are context-dependent considerations, such as the timing of the receipt of BTD, stage of development of the program when BTD is withdrawn, and the status of other available therapies, which may impact the public messaging by a BTD sponsor around why withdrawal occurred. Prior to the voluntary withdrawal or rescinding of BTD, the sponsor and the FDA could engage in discussions surrounding the rationale for a planned withdrawal or rescinding of BTD and the implications of these actions on the drug development program. Sponsors may also benefit from an explanation of how withdrawal impacts a program's ability to participate in other Oncology Center for Excellence (OCE) pilots (e.g., Real Time Oncology Review, Project Orbis).

Opportunity 2. Improve mechanisms to enable meaningful discussions between FDA and sponsors that clearly align with key decision points.

Ensure productive and timely discussions between the FDA and sponsor.

It is beneficial to have post-BTD discussions as early as possible so the drug development program can take full advantage of BTD's enhanced opportunities for interaction and advice. Ideally, sponsors should prepare for and request the comprehensive post-BTD multidisciplinary meeting so that it occurs in a timely fashion (within the first six months of receiving BTD).[5] In this meeting, the sponsor may propose a high-level communication plan and estimated timeline for future interactions aimed at accelerating development of their BTD drug.

Meetings following the comprehensive post-BTD multidisciplinary meeting could be focused to align on the specific needs of a drug development program at each stage of development. The benefit of collaborative discussions with the FDA and sponsors may be more fully realized when questions are focused on a single discrete issue. This could also shorten the lead time for the

meeting and reduce the burden on both the FDA and the sponsor to prepare for the meeting. If questions arise between formal meetings, meeting requests targeted to address specific topics could be considered. Proposed PDUFA VII goals include a proposed Type D meeting that would be suitable to address a narrow set of issues[6] which will be further described in a revised draft of the draft guidance entitled "Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products." Specifically, Type D meetings are intended to address a follow up issue after a formal meeting, a narrow issue the sponsor would like FDA input on that requires input from no more than 3 disciplines, or a general question that does not require detailed advice. During focus group discussions, participants suggested topics such as CMC/product quality-related hurdles, trial design-related issues, and timing of dose optimization studies for meetings like Type D meetings. Timelines for drug development may need to be coordinated when the program is expedited, and key questions that may be rate-limiting may not necessarily arise in alignment with the timing of traditional milestone meetings for aspects such as CMC and clinical considerations. The development of a mechanism to update FDA on key components of drug development could help both sponsors and the FDA identify when meetings might be valuable to head off potential rate-limiting obstacles to oncology drug development.

Provide additional clarity to ensure a better understanding of which types of meetings are optimal for specific aspects of drug development for products with BTD.

Formal meetings such as Type B meetings are generally held within 60 days of their request (70 days for end-of-phase meetings) and require extensive preparation on the part of sponsors and FDA staff. As such, these meetings may not always be amenable to post-BTD drug development timeframes, therefore additional meeting strategies are likely to be helpful for BTD product development. Table 1 outlines available and proposed meeting types to support development of drugs with BTD. Strategies to formally integrate and operationalize issue specific meetings for products with BTD could be informed through pilot projects such as the Complex Innovative Trial Design Pilot Meeting Program (CID) and Model-Informed Drug Development (MIDD) Pilot Program. Given the fast pace of development post-BTD, decisions are made decisively and quickly, and timely interactions are extremely important.

In addition to timely discussions, external stakeholders expressed that it may help to have enhanced interactions and feedback from the FDA. Proactive and thoughtful interactions support expedited development and review of products with BTD. Optimizing meeting structure and approach could enable discussions between FDA and sponsors to occur more frequently, help address issues earlier in development, and provide opportunities for proactive planning of manufacturing and testing strategies and clinical development relative to traditional drug development approaches. It may be important to promote timely dialogue between the sponsors and FDA review divisions to enable proactive management of potential issues, which could become major issues either for submission or review of the premarket application. Additionally, the development of guidance that outlines best practices or novel approaches for avoiding commonly encountered issues may be of value.
Rather than a lengthy briefing document, a proposed Type D meeting (See Table 1 below)

may be supported by a more focused briefing document containing the information needed to address the drug development issue(s) at hand to inform the discussion and feedback provided for focused meetings. Additionally, templates that provide high-level summaries of specific aspects of drug development such as the status of CMC development and dose optimization might be beneficial; such high-level summaries could potentially facilitate early FDA identification of potential drug development issues that could be addressed in future meetings with the sponsor post-BTD might be beneficial (see discussion on use of "Drug Development Snapshots" as a potential communication tool).

During the focus group meetings, there was discussion of informal meeting requests but ultimately the value of a more "informal request" may be limited when there are questions requiring input from multiple review disciplines or limited background is provided by the sponsor.

Table 1: Available FDA Meeting Types and Applicability for Products with BTD[7,8,9]			
Meeting Type	**General Purpose**	**Timeframe**	**Application for BTD Products**
Type A Meetings	Meetings that are necessary for an otherwise stalled product development program to proceed or to address an important safety issue.	Within 30 days of request	Strengthening communication between the FDA and sponsor could help avoid issues that would require a Type A meeting for products with BTD.
Type B Meetings	Routine meetings occurring at pre-defined endpoints between FDA and a sponsor. Meetings typically occur right after or right before the submission of clinical data or a new drug filing.	Within 60 to 70 days of request	Formal meetings for products with BTD, including the initial comprehensive BTD meeting, are granted as Type B, unless they qualify as Type A meetings.
Type C Meetings	A meeting that is not a Type A or Type B meeting regarding the development and review of a product.	Within 75 days of request	Very rarely used for BTD products, as BTD default would be to Type B meetings unless reclassified by FDA and have a potentially shorter turnaround. With the addition of a Type D meeting (see below), Type C meetings could be reserved for development issues that would not warrant as quick of a turnaround.
Type F Meetings	Early advice meetings to discuss pediatric development plans.	Within 30 days of request	Meetings encouraged for BTD drugs to ensure that an agreed initial pediatric study plan (iPSP) is in place prior to marketing application. This may be particularly important with respect to mechanism of action based pediatric requirements (Section 504 of FDARA 2017) for original application of oncology drugs.
Proposed Type D Meetings*	Focused on a narrow set of issues (no more than 2 focused topics) which could include: • a follow-up question that raises a new issue after a formal meeting • a narrow issue with few associated questions • a general question about an innovative development approach	Within 50 days of request	This meeting type could promote proactive, collaborative discussion to help identify and mitigate/troubleshoot potential rate-limiting steps to development articulated by the sponsor in a more focused briefing document. Type D meetings could be reserved for issues that require more rapid feedback versus issues that may be best routed through a Type B meeting request.

*Meeting type described in the PDUFA VII Commitment Letter
(https://www.fda.gov/media/151712/download)

The ability to connect informally with review staff, Division leaders, and Office Directors, while potentially valuable for timely decision making, may be counterproductive unless it includes all relevant members of the FDA review team and is incorporated as official FDA correspondence to the sponsor.

Opportunity 3. Develop communication tools and optimize processes to support interdisciplinary coordination within and between the FDA and sponsors.

Create a program for voluntary submission of "Drug Development Snapshots" for earlier identification of issues that could result in delays in development.

A strength of BTD is the product-specific dialogue provided through cross-functional structured interactions between the FDA and sponsors, which involves senior leadership at FDA. However, challenges may be encountered at different times and in different areas of drug development depending on the drug development program. Identification of challenges may not align with milestone meetings that are typically attended by cross-functional groups within the FDA and the sponsor drug development team, which can result in delays addressing these issues. Voluntary sponsor submission of high-level product development information in the form of periodic Drug Development Snapshots (see mock template for dosing snapshot in Appendix) could help prompt earlier identification of rate-limiting aspects of development (e.g., dosing, CMC/Product quality aspects, diagnostic co-development, plans for confirmatory trials if accelerated approval pathway is anticipated, etc.), serve as a vehicle to support information exchange, and help determine when meetings outside of the normal milestone cadence would be most beneficial. This increased transparency throughout development outside of normal milestone meetings can also promote FDA cross-discipline and inter-center communication on the development program, plans for upcoming milestones, and necessary interventions. While these snapshots could be particularly useful for enabling improved real-time communication, they could also be leveraged beyond BTD products. The FDA could consider a pilot project to explore the utility of Drug Development Snapshots, including their optimal timing with respect to drug development, frequency of submission, and content.

Refine best practices for communicating with RPMs to help facilitate efficient collaboration.

FDA Regulatory Project Managers (RPMs) play a vital role in triaging and prioritizing sponsors' requests, as well as in identifying the appropriate person(s) for FDA-sponsor and internal meetings, and in coordinating responses for such meetings when necessary. While sponsor interactions with RPMs are extremely helpful, there is opportunity to improve these interactions. Defining the best communication practices for sponsors and RPMs may help sponsors understand expectations specifically in the context of a program with BTD. One opportunity to optimize efficient communication is the ability for sponsors to "flag" requests for feedback that are time-sensitive, and clearly identify they are requesting a reviewer's feedback on a specific topic, to help RPMs appropriately prioritize requests. RPMs could also provide a time estimate for how long it will take to provide feedback for the request, at the point of acknowl-

edgement of the request. Updating CDER's Manual of Policies and Procedures (MAPP) 6030.9 Good Review Practice: Good Review Management Principles and Practices for Effective IND Development and Review, which describes review management principles and practices, may provide additional clarity on best communication practices. A one-on-one meeting between the FDA and Sponsor RPMs could also help set communication expectations. Further, CDER's 2017 Best Practices for Communication Between IND Sponsors and FDA During Drug Development Good Review Practice[10] outlines appropriate communication strategies between sponsors and FDA, and additional awareness and following of the best practices may increase efficiencies.

Communicate the sponsor's role in preparing materials for cross-discipline meetings.

For cross-discipline meetings in BTD drug development to be productive, it may be helpful to outline the role of both sponsors and the FDA in identifying when general or more specific feed-back is warranted and the key disciplines needed for each interaction. Sponsors could provide concise, focused information necessary for FDA to answer the questions at hand. RPMs could then distribute the materials to the review team, including to reviewers or consultants outside of the Division or Center.

Opportunity 4. Provide a roadmap for addressing key pressure points for products with BTD.

Encourage early collaboration, alignment, and prioritization between pharmaceutical and device sponsors and CDER and CDRH.

Challenges to efficient development of a companion diagnostic can arise particularly for drugs developed for rare patient populations or in the setting of a product with BTD. Early identifica-tion of the need for a companion diagnostic and plans for parallel development with the goal of contemporaneous approval of a BTD drug and companion diagnostic (if needed) could be a key component of the comprehensive interdisciplinary post-BTD meeting. As noted earlier, out-lining specific meeting types (Type B meetings or otherwise) and timelines to focus on incorpo-rating companion diagnostic co-development can increase collaboration between CDER, CDRH, the pharmaceutical sponsor, and the diagnostic sponsor and enable early preparation for pos-sible bridging studies, as well as strategies for saving patient samples and adequate patient ascertainment. Additionally, notification to CDRH upon designation of a BTD could allow for effi-cient mobilization of appropriate resources. There is also a need for additional clarity around the level of evidence and data elements needed prior to approval for a companion diagnostic. It may be helpful if diagnostic tests developed to direct the use of therapies with BTD were also considered for breakthrough device designation to assist in alignment, prioritization, and collab-oration between senior leadership within and across medical product Centers responsible for each breakthrough product.

Facilitate timely discussion and agreement on dose selection, exposure-response analyses, and study design.

Rapid development programs associated with BTD can give the impression that there is insufficient time for robust dose finding approaches; however, identifying the optimal dosage to support safe and effective use of oncology drugs, and accumulating sufficient information to support this dosage, is extremely important; the selection of the recommended dose without adequate investigation is unacceptable. FDA's OCE has highlighted dose optimization, including for BTD drugs, as a priority by introducing Project Optimus, calling for dose selection justification and earlier discussion of dose selection during the IND phase. BTD may be granted prior to identification of the optimal dose; however, the approach planned to support the dosage(s) intended for further development could be discussed with the FDA prior to embarking on a clinical trial intended to provide evidence of safety and effectiveness to support a marketing application. These approaches could integrate PK, PD, efficacy, safety, and tolerability data to adequately support dose selection and may result in the selection of 2 or more doses for further exploration. Sponsors could seek out these discussions, which may occur before or after receipt of BTD, as early in the development process as possible. FDA could outline opportunities to discuss strategies for dose optimization and selection of the pivotal dose(s) in pre-BTD meetings. FDA could also clarify and provide feedback on the appropriate use of systems and model-based approaches to support dose selection, study design, and exposure-response analyses and provide feedback on proposals to leverage relevant markers of activity to inform the dosing decisions early in development. Further, learnings from FDA's work through its PDUFA VI commitment on model-informed drug development (MIDD) could be leveraged to facilitate appropriate dose selection. Clarity is needed on the appropriate use of MIDD to support dose selection, including the evidence needed to show that a model is credible and the role of the model in supporting or supplementing clinical data.

Identify processes to support early Office of Pharmaceutical Quality (OPQ) discussions and facilitate timely submission of CMC information to address rate limiting-steps in the commercialization process.

Sponsors are encouraged to provide available CMC/manufacturing information and commercialization plans, which may be in the form of a "Drug Development Snapshot," early to FDA to potentially maximize the ability to address rate-limiting steps in the development and marketing of a breakthrough product. Currently, discussions with CMC/OPQ generally occur later in development, as CMC development often lags behind clinical development for expedited programs. Sponsors may initiate these conversations earlier in the development program, in a more proactive manner, to aid in planning and development of manufacturing and product quality strategies. An opportunity to engage in a discussion specific to late phase/commercial manufacturing and testing approaches as well as to troubleshoot Quality-related development challenges may expedite the commercialization process. As described in the PDUFA VII Commitment Letter, FDA plans to issue a new MAPP on approaches to address CMC challenges for products with accelerated clinical development timelines and will describe early engagement with sponsors of such products. Further, early identification of the regulatory business project manager (RBPM) for the OPQ related inquiries would be helpful. Processes for rolling submission and review of information before all the necessary stability data are avail-

able could also be explored; this could lead to submission of all Module 3 content except for all or part of the 3.2.S.7 and 3.2.P.8 sections months ahead of the submission of the final components of the NDA or BLA, which could include these remaining sections. The sponsor should propose and reach agreement with FDA on plans for early rolling submission of segments of Module 3 at the EOP 2 or a Pre-NDA/BLA meeting. For example, early submission of detailed manufacturing site information (e.g., list of manufacturing facilities with addresses and FEI numbers, current CGMP status, facilities' prior experience with similar manufacturing processes, manufacturing area and filling line or equipment used) can allow earlier coordination and planning if a pre-licensure or preapproval inspection is necessary. Discussions with OPQ to help clarify the level of CMC information needed for approval and types of plans that could be implemented in the post-approval setting can also be helpful.

Consider FDA's available strategies to assess facility risks and enable more efficient inspections.

Certain flexibilities were allowed during the COVID-19 pandemic, including operational processes (both for FDA and sponsors), and clarifications on regulatory approaches toward application components. The FDA has used alternative tools to inform facility assessments including   examination of a firm's compliance history, inspection reports from trusted foreign regulatory partners, records requests, and the use of remote interactive evaluations.[11] Proposed PDUFA VII goals include that some of these flexibilities, or the principles behind them, will be explored to ensure that facility assessments for BTD products are timely and focused on critical areas for coverage, thus alleviating delays in approval and enabling sponsors and FDA to allocate resources efficiently.

Explore decoupling drug substance and drug product process performance qualification (PPQ).

When drug substance and drug product PPQ occur sequentially (for those programs requiring PPQ data in the initial application), the PPQ timeline may delay submission of the application and product approval. Inclusion of CMC into Real-Time Oncology Review (RTOR) and the development of the CMC Assessment Aid have brought flexibility and enhanced CMC review efficiency to oncology application reviews. Concurrent execution and completion of the drug substance and drug product PPQs could build on experiences in small molecule development and may result in expedited CMC readiness to meet clinical timelines. Feasibility of this approach has been demonstrated in the development of small molecules over the last few decades as well as for the BTD product pembrolizumab (a monoclonal antibody).[12] Exploring the conditions where it might be possible to successfully decouple drug substance PPQ and drug product PPQ, could help to expedite the timeline safely and efficiently. Concurrent validation approaches could be useful for a BTD product to market the PPQ batches. Circumstances and rationale for concurrent release could be fully described in a PPQ protocol, which for BLAs should be submitted in the application. More details can be found in the process validation guidance.[13]

References

1. FDA. Guidance for Industry: Expedited Programs for Serious Conditions––Drugs and Biologics. (2014). Available at: https://www.fda.gov/media/86377/download. (Accessed: 7th July 2021)

2. FDA. Guidance for Industry: Expedited Programs for Serious Conditions––Drugs and Biologics. (2014). Available at: https://www.fda.gov/media/86377/download. (Accessed: 7th July 2021)

3. FDA. CDER Breakthrough Therapy Designation Requests Received by Fiscal Year. (2020). Available at: https://www.fda.gov/media/95292/download. (Accessed: 12th August 2021).

4. Friends of Cancer Research. Drug Development Dashboard: Original Approvals in Oncology, 2013-2021. (2021). Available at: https://friendsofcancerresearch.org/drug-development-dashboard

5. FDA. Manual of Policies and Procedures for Good Review Practice management of BTD drugs and biologics (2014). Available at: https://www.fda.gov/files/about%20fda/published/Good-Review-Practice--Management-of-Breakthrough-Therapy-Designated-Drugs-and-Biologics.pdf (Accessed 29 July 2021).

6. FDA. PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2023 Through 2027 (2022). Available at: https://www.fda.gov/media/151712/download (Accessed 9 September 2021).

7. FDA. Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products Guidance for Industry (2017). Available at: https://www.fda.gov/media/109951/download (Accessed 29 July 2021).

8. FDA. Manual of Policies and Procedures for Good Review Practice management of BTD drugs and biologics (2014). Available at: https://www.fda.gov/files/about%20fda/published/Good-Review-Practice--Management-of-Breakthrough-Therapy-Designated-Drugs-and-Biologics.pdf (Accessed 29 July 2021).

9. FDA. Pediatric Oncology Product Development Early Advice Meeting (Type F)1. Available at: https://www.fda.gov/about-fda/oncology-center-excellence/pediatric-oncology-product-development-early-advice-meeting-type-f1 (Accessed 29 July 2021).

10. FDA. Best Practices for Communication Between IND Sponsors and FDA During Drug Development Guidance for Industry and Review Staff Good Review Practice (2017). Available at: https://www.fda.gov/media/94850/download (Accessed 8 September 2021).

11. FDA. Manufacturing, Supply Chain, and Drug and Biological Product Inspections During COVID-19 Public Health Emergency Questions and Answers. (2021) Available at: https://www.fda.gov/media/141312/download (Accessed on 13 December, 2021).

12. Dye, E., Sturgess, A., Maheshwari, G. et al. Examining Manufacturing Readiness for Breakthrough Drug Development. AAPS PharmSciTech: 17, 529–538. (2016). https://doi.org/10.1208/s12249-015-0455-1

13. FDA. Guidance for Industry Process Validation: General Principles and Practices (2011). Available at: https://www.fda.gov/files/drugs/published/Process-Validation--General-Principles-and-Practices.pdf (Accessed 9 September 2021).

Appendix 1. Drug Development Snapshot Template – Clinical Pharmacology (Dose & Administration) Snapshot

Please note: The table below describes the supportive evidence for the proposed dose and schedule. The target length of the completed snapshot is 2–5 pages.

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule and route of administration	• What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding? • When do you anticipate that a R2PD will be selected? • Are other routes of administration being investigated?
Mechanism of action (MOA) and format	• Is the therapeutic a small or large molecule? Another platform? What is the MOA?
Translational evidence	• Is there established pharmacological evidence (e.g., target engagement, MOA, outcome-based bio-markers, tumor volume) in the relevant preclinical species? • Is the dose–PK relationship established in the non-clinical species (i.e., is the PK dose proportional)? • Are the pharmacological/efficacious target concentrations for patients defined? • Is the dose/exposure-response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?
	Clinical Evidence

Clinical studies	• List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)
	• Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:
	  ○ Dose escalation, expansion cohorts with or without randomization
	  ○ Single arm randomization (i.e., dose and/or control); adaptive design

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule and route of administration	• What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding? • When do you anticipate that a R2PD will be selected? • Are other routes of administration being investigated?
Mechanism of action (MOA) and format	• Is the therapeutic a small or large molecule? Another platform? What is the MOA?
Translational evidence	• Is there established pharmacological evidence (e.g., target engagement, MOA, outcome-based biomarkers, tumor volume) in the relevant preclinical species? • Is the dose–PK relationship established in the non–clinical species (i.e., is the PK dose proportional)? • Are the pharmacological/efficacious target concentrations for patients defined? • Is the dose/exposure–response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?
Clinical Evidence	

Clinical studies	• List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)
	• Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:
	  ○ Dose escalation, expansion cohorts with or without randomization
	  ○ Single arm randomization (i.e., dose and/or control); adaptive design

PK characteristics	• Is the dose–PK relationship established (i.e., is the PK dose proportional)? • Do the PK characteristics (accumulation, half–life) justify the dosing interval? • Are there any intrinsic or extrinsic factors (e.g., food, body weight, immunogenicity) that would majorly influence PK? (i.e., if these warrants dose adjustments in a subset of patients) • Was the PK variability considered when selecting a dose that would achieve target exposure for the majority of patients?

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule and route of administration	• What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding? • When do you anticipate that a R2PD will be selected? • Are other routes of administration being investigated?
Mechanism of action (MOA) and format	• Is the therapeutic a small or large molecule? Another platform? What is the MOA?
Translational evidence	• Is there established pharmacological evidence (e.g., target engagement, MOA, outcome-based biomarkers, tumor volume) in the relevant preclinical species? • Is the dose-PK relationship established in the non-clinical species (i.e., is the PK dose proportional)? • Are the pharmacological/efficacious target concentrations for patients defined? • Is the dose/exposure-response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?

Clinical Evidence	

Clinical studies	• List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)
	
	• Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:
	
	  o Dose escalation, expansion cohorts with or without randomization
	
	  o Single arm randomization (i.e., dose and/or control); adaptive design

PK characteristics	• Is the dose-PK relationship established (i.e., is the PK dose proportional)? • Do the PK characteristics (accumulation, half-life) justify the dosing interval? • Are there any intrinsic or extrinsic factors (e.g., food, body weight, immunogenicity) that would majorly influence PK? (i.e., if these warrants dose adjustments in a subset of patients) • Was the PK variability considered when selecting a dose that would achieve target exposure for the majority of patients?

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule and route of administration	• What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding? • When do you anticipate that a R2PD will be selected? • Are other routes of administration being investigated?
Mechanism of action (MOA) and format	• Is the therapeutic a small or large molecule? Another platform? What is the MOA?
Translational evidence	• Is there established pharmacological evidence (e.g., target engagement, MOA, outcome-based biomarkers, tumor volume) in the relevant preclinical species? • Is the dose–PK relationship established in the non-clinical species (i.e., is the PK dose proportional)? • Are the pharmacological/efficacious target concentrations for patients defined? • Is the dose/exposure–response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?
Clinical Evidence	

Clinical studies	• List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)
	• Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:
	  ○ Dose escalation, expansion cohorts with or without randomization
	  ○ Single arm randomization (i.e., dose and/or control); adaptive design

PK characteristics	• Is the dose–PK relationship established (i.e., is the PK dose proportional)? • Do the PK characteristics (accumulation, half-life) justify the dosing interval? • Are there any intrinsic or extrinsic factors (e.g, food, body weight, immunogenicity) that would majorly influence PK? (i.e., if these warrants dose adjustments in a subset of patients) • Was the PK variability considered when selecting a dose that would achieve target exposure for the majority of patients?

Key Area of Consideration	Supporting Evidence
Recommended dose, schedule and route of administration	• What is the current dose(s), schedule(s) and route of administration that are currently being evaluated in clinical trials? Has the RP2D been selected? If the RP2D has not been selected, what key questions are outstanding? • When do you anticipate that a R2PD will be selected? • Are other routes of administration being investigated?
Mechanism of action (MOA) and format	• Is the therapeutic a small or large molecule? Another platform? What is the MOA?
Translational evidence	• Is there established pharmacological evidence (e.g., target engagement, MOA, outcome-based biomarkers, tumor volume) in the relevant preclinical species? • Is the dose–PK relationship established in the non-clinical species (i.e., is the PK dose proportional)? • Are the pharmacological/efficacious target concentrations for patients defined? • Is the dose/exposure–response (i.e., biomarkers, tumor size, etc.) relationship identified from the in vitro cellular systems or the in vivo animal models?
Clinical Evidence	

Clinical studies	• List of ongoing and completed studies (i.e., single agent and/or combination studies, indication, etc.)
	• Brief description of study design including patient population/cancer type(s) under study, line of therapy, and doses and schedules evaluated, sample size. For example, the following elements can be considered:
	  ○ Dose escalation, expansion cohorts with or without randomization
	  ○ Single arm randomization (i.e., dose and/or control); adaptive design

PK characteristics	• Is the dose–PK relationship established (i.e., is the PK dose proportional)? • Do the PK characteristics (accumulation, half-life) justify the dosing interval? • Are there any intrinsic or extrinsic factors (e.g., food, body weight, immunogenicity) that would majorly influence PK? (i.e., if these warrants dose adjustments in a subset of patients) • Was the PK variability considered when selecting a dose that would achieve target exposure for the majority of patients?

Safety summary	• Summary of frequencies of key AEs (including chronic low grade AEs which can affect tolerability) of interest by dose • Is there a dose/exposure–safety or PK-PD relationship, upon the adjustment of potential covariates, for safety? If yes, what is the nature of the relationship? • Summary of dose interruptions, reductions, and discontinuations by dose/exposures ○ Is there an increased frequency of dose interruptions or reductions or treatment discontinuations with increasing doses/exposures? • Are there any late occurrence toxicities beyond the DLT period? Are there early PD biomarkers reflective of the delayed safety endpoints? • Are there any overlapping toxicities with the concomitant medications in the patient population (e.g., treatment combinations for NME with SOC and/or treatments for comorbidities/cancer related symptoms)? • Is there an increased frequency of dose interruptions, reductions, or treatment discontinuations with increasing doses/exposures? • If acute/transient toxicities were observed, were alternative dosing approaches considered (e.g., step-up dosing)? • Does existing data indicate this is a narrow therapeutic window drug with dose limiting toxicity that is monitorable (e.g., biomarkers, BP, HR, neuropathy)? • If yes, does this drug provide an opportunity to personalize the dose for an individual patient or a sub-population based on the emerging monitorable toxicity?
Efficacy summary	• Summary of response endpoints by dose (e.g., ORR, PFS) • Is there a dose/exposure – efficacy (primary efficacy endpoint) and PK-PD (e.g., mechanism of action/predictive biomarkers) relationship upon the adjustment of potential confounders? If yes, what is the nature of the relationship? • Is the dose schedule (e.g., frequency, dose holidays) justified based on the K/PK-PD and/or QSP modeling approaches? • Are the relevant exposure metrics for efficacy identified (e.g., AUC, C_{max}, C_{min}, concentration-time, RO)?

Other considerations	• Are there any manufacturing considerations (e.g., pill burden, maximal feasible dose, etc.) that needs to be taken into account? • Is there any patient factors that need to be considered (e.g., patient convenience/compliance [QD, BID, TID), QW vs Q3W, SC vs IV) • Complimentary M&S approaches (i.e., PK-TGI/QSP/ML etc.) for dose optimization and/or inform dose adjustments
Additional Clinical Evidence	
Planned clinical studies	• Are there additional planned clinical studies that will contribute data to the current D&A plan/rationale or future D&A proposals?
Other Evidence	• Does additional scientific evidence exist (e.g., from similar class, MOA or indication) that may support the current D&A plan/rationale (e.g., publications, scientific presentations)?

Abbreviations: AEs=adverse events; AUC=area under the curve; BP=blood pressure; C_{max}=maximum 'peak' concentration; C_{min}=minimum 'trough' concentration; D&A=dose & administration; DLT=dose-limiting toxicity; HR=heart rate; K=kinetic; MOA=mechanism of action; NME=new molecular entity; ORR=overall response rate; PD=pharmacodynamic; PFS=progression-free survival; PK=pharmacokinetic; QSP=quantitative systems pharmacology; RO=receptor occupancy; SOC=standard of care

Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: phase II of the Friends of Cancer Research TMB Harmonization Project

D. M. Vega[1], L. M. Yee[2], L. M. McShane[2], P. M. Williams[3], L. Chen[3], T. Vilimas[3], D. Fabrizio[4], V. Funari[5], J. Newberg[4], L. K. Bruce[5], S.-J. Chen[6], J. Baden[7], J. Carl Barrett[8], P. Beer[9], M. Butler[10], J.-H. Cheng[6], J. Conroy[11], D. Cyanam[12], K. Eyring[13], E. Garcia[14], G. Green[7], V. R. Gregersen[15], M. D. Hellmann[16], L. A. Keefer[17], L. Lasiter[1], A. J. Lazar[18], M.-C. Li[2], L. E. MacConaill[14], K. Meier[19], H. Mellert[20], S. Pabla[11], A. Pallavajjalla[21], G. Pestano[20], R. Salgado[9], R. Samara[15], E. S. Sokol[4], P. Stafford[22], J. Budczies[23], A. Stenzinger[23], W. Tom[12], K. C. Valkenburg[17], X. Z. Wang[24], V. Weigman[25], M. Xie[8], Q. Xie[26], A. Zehir[16], C. Zhao[19], Y. Zhao[2], M. D. Stewart[1*] & J. Allen[1], on behalf of the TMB Consortium

[1]Friends of Cancer Research, Washington; [2]National Cancer Institute, Bethesda; [3]Molecular Characterization Laboratory, Frederick National Lab for Cancer Research, Leidos Biomedical Research Inc., Frederick; [4]Foundation Medicine Inc., Cambridge; [5]NeoGenomics Laboratories, Aliso Viejo, USA; [6]ACT Genomics, Taipei, Taiwan; [7]Bristol Myers Squibb Co., Princeton; [8]AstraZeneca Pharmaceuticals LP, Waltham, USA; [9]European Organisation for Research and Treatment of Cancer, Brussels, Belgium; [10]LGC Clinical Diagnostics, Gaithersburg; [11]OmniSeq Inc., Buffalo; [12]Clinical Sequencing Division, Thermo Fisher Scientific, Ann Arbor; [13]Intermountain Precision Genomics, St. George; [14]Brigham and Women's Hospital, Boston, USA; [15]QIAGEN Inc, Aarhus, Denmark; [16]Memorial Sloan Kettering Cancer Center, New York; [17]Personal Genome Diagnostics, Baltimore; [18]The University of Texas MD Anderson Cancer Center, Houston; [19]Illumina Inc, Clinical Genomics, San Diego; [20]Biodesix, Inc., Boulder; [21]Johns Hopkins University, Baltimore; [22]Caris Life Sciences Inc, Phoenix, Arizona, USA; [23]Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany; [24]EMD Serono Research and Development Institute, Inc., Billerica; [25]Q Squared Solutions, Durham; [26]General Dynamics Information Technology, Inc., Columbia, USA

Background: Tumor mutational burden (TMB) measurements aid in identifying patients who are likely to benefit from immunotherapy; however, there is empirical variability across panel assays and factors contributing to this variability have not been comprehensively investigated. Identifying sources of variability can help facilitate comparability across different panel assays, which may aid in broader adoption of panel assays and development of clinical applications.

Materials and methods: Twenty-nine tumor samples and 10 human-derived cell lines were processed and distributed to 16 laboratories; each used their own bioinformatics pipelines to calculate TMB and compare to whole exome results. Additionally, theoretical positive percent agreement (PPA) and negative percent agreement (NPA) of TMB were estimated. The impact of filtering pathogenic and germline variants on TMB estimates was assessed. Calibration curves specific to each panel assay were developed to facilitate translation of panel TMB values to whole exome sequencing (WES) TMB values.

Results: Panel sizes >667 Kb are necessary to maintain adequate PPA and NPA for calling TMB high versus TMB low across the range of cut-offs used in practice. Failure to filter out pathogenic variants when estimating panel TMB resulted in overestimating TMB relative to WES for all assays. Filtering out potential germline variants at >0% population minor allele frequency resulted in the strongest correlation to WES TMB. Application of a calibration approach derived from The Cancer Genome Atlas data, tailored to each panel assay, reduced the spread of panel TMB values around the WES TMB as reflected in lower root mean squared error (RMSE) for 26/29 (90%) of the clinical samples.

Conclusions: Estimation of TMB varies across different panels, with panel size, gene content, and bioinformatics pipelines contributing to empirical variability. Statistical calibration can achieve more consistent results across panels and allows for comparison of TMB values across various panel assays. To promote reproducibility and comparability across assays, a software tool was developed and made publicly available.

Key words: precision medicine, biomarker, tumor mutational burden, immunotherapy, cancer

Correspondence to: Dr Mark D. Stewart, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC 20036, USA. Tel: +1-202-944-6706
E-mail: mstewart@focr.org (M. D. Stewart).

INTRODUCTION

The use of anti-programmed death-ligand 1 (PD-L1)/anti-programmed cell death protein 1 (PD-1) therapies has risen dramatically over the last few years, with an increasing number of regulatory approvals in several cancer types.[1]

Despite these successes, only a fraction of cancer patients benefit from immune checkpoint blockade, which has led to broader exploration of biomarkers to identify patient populations more likely to respond to immunotherapy.[2-4]

Tumor mutational burden (TMB) is defined as the number of somatic mutations per megabase of interrogated genomic sequence. There has been early success in using TMB to predict responses to immune checkpoint inhibitors for patients with melanoma and lung cancer, among others.[5-7] Importantly, the use of TMB as a biomarker is tumor agnostic. Recently, data from KEYNOTE-158 (NCT02628067) supported the use of pembrolizumab for the treatment of TMB-high adult and pediatric patients with unresectable or metastatic solid tumors that had progressed after previous treatment. TMB high was set at TMB $\geq$10 mut/Mb for patients' formalin-fixed paraffin-embedded (FFPE) tumor tissue samples tested with the Foundation Medicine (Cambridge, MA) FoundationOne CDx assay.[8] The findings of this study led to the first United States Food and Drug Administration (FDA) approval of pembrolizumab using TMB high as a positive predictive biomarker for patient selection in a tissue-agnostic setting. The FoundationOne CDx assay is the first FDA-approved companion diagnostic to measure TMB and to help identify patients who may be appropriate for treatment with pembrolizumab, regardless of solid tumor type.

In clinical practice, next generation sequencing (NGS) targeted gene panel assays are preferred over whole exome sequencing (WES) approaches for TMB estimation due to already relevant clinical use for identification of targetable oncogenes, broader availability, quicker turnaround time, and cost. To date, FoundationOne CDx is the only FDA-approved panel assay that reports TMB, while the Memorial Sloan Kettering Cancer Center MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets), NantHealth's Omics Core test, and the PGDx elio™ tissue complete assay have received FDA 510(k) clearance.[9] However, access to these regulated panel assays is not ubiquitous, leading to the development of additional NGS targeted gene panel assays by local and commercial providers.

Several factors impact variation among panel assays including sample input, tumor content, panel size, gene content, quality control (QC), NGS platform, and bioinformatics pipeline, which may influence TMB estimates and lead to inconsistent TMB calculation and reporting.[10] Although 10 mut/Mb is the cut-off for TMB-high designation with the FoundationOne CDx assay, other panels may have different clinical cut-offs. Because of these inherent differences, the standardization of clinical validation practices, harmonization of TMB assessment, and alignment across TMB panel assays are critical steps to improve consistency of results and comparability across panel assays, and to promote confidence in the use of this biomarker. This is a crucial time to seek harmonization in TMB measurement and assess comparability across TMB assays to prevent the inconsistencies seen in past biomarkers. For example, the lack of alignment across PD-L1 immunohistochemistry assays, lack of comparability of panel assay results, and different cut-offs defined for each drug have posed a significant challenge for the implementation of PD-L1 expression testing.[11-17]

With the aim of facilitating harmonization and alignment across tissue TMB assays, the Friends of Cancer Research (Friends) TMB Harmonization Consortium was formed. The TMB Consortium, which consists of several diagnostic manufacturers, academics, pharmaceutical companies, the National Cancer Institute (NCI), Frederick National Laboratory for Cancer Research, and the FDA, previously reported results from the first phase of the project where the theoretical variability across 11 commercial and academic panel assays was described and consortium-endorsed recommendations were proposed for the analytical validation of TMB assays.[18] Moreover, the TMB Consortium partnered with Quality in Pathology (QuIP) in Germany to complement its approach and enrich its perspective on the variability in TMB estimates across laboratories through a technical comparability study.[19]

In this study, we set out to characterize the empirical variability in TMB measurements across platforms using a common set of cell lines and clinical samples tested across 16 panel assays from 16 participating laboratories. Further, we aimed to elucidate how certain factors such as panel size, gene content, and bioinformatics pipelines impact TMB estimates, and to investigate the use of a calibration tool based on The Cancer Genome Atlas (TCGA) data and human tumor-derived cell lines that will facilitate comparability across different panel assays. Based on these results, we aim to provide data and guidance that will help improve the consistency and reliability of panel tissue TMB estimation across platforms and facilitate the use of this complex biomarker in clinical decision making.

MATERIALS AND METHODS

Samples (clinical samples and cell lines)

Thirty-six FFPE clinical tumor samples and matched buffy coat were acquired from iSpecimen (Lexington, MA) and processed at a reference laboratory (MoCha Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD), where tumor tissue specimens were enriched by histological macrodissection to the extent possible, with the estimated tumor cell content in macrodissected specimens ranging from 30% to 95% (Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2021.09.016). All samples were categorized into the following broad tumor types: bladder, colon, gastric, gastrointestinal stromal tumor (GIST), and lung. Specific histologic diagnoses and demographic data can be found in Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2021.09.016. Genomic DNA from tumor specimens was extracted using AllPrep FFPE Nucleic acid Extraction kit and the QIAcube automated platform (QIAGEN, Germantown, MD). Genomic DNA from buffy coat specimens was extracted using the QIAsymphony automated platform (QIAGEN). DNA was quantitated using Qubit dsDNA BR assay kit (Thermo Fisher

Scientific, Waltham, MA). After performing QC, seven GIST samples were excluded from further analyses mostly due to low DNA yield, poor DNA quality, and low depth of coverage (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2021.09.016); thus, only 29 clinical samples were evaluated in this study.

Ten (two breast, eight lung cancer) human-derived matched tumor-normal cell lines were selected and obtained from the American Type Culture Collection (ATCC) (Supplementary Table S3, available at https://doi.org/10.1016/j.annonc.2021.09.016) and processed at a reference laboratory, SeraCare (now LGC Clinical Diagnostics Division). Cell lines were grown in accordance with ATCC specifications with no more than five passages. DNA was extracted from frozen cell pellets (80-100M cells) using the QIAGEN Gentra Puregene Kit. Purified genomic DNA concentrations were normalized to 50 ng/μL in 0.1x Tris-EDTA buffer as measured by the Qubit dsDNA BR assay kit. Integrity of purified genomic DNA was assessed by agarose gel electrophoresis. All 10 matched cell line samples passed QC, and thus were evaluated in this study.

Whole exome sequencing and TMB estimation

The reference laboratory carried out WES, where 50 ng of genomic DNA was sheared to 150-180 bp using Covaris LE220 sonicator (Covaris, Woburn, MA). Library preparation was automated on a SciClone G3 liquid handling workstation using custom scripts (Supplementary Material, available at https://doi.org/10.1016/j.annonc.2021.09.016). A NovaSeq 6000 (Illumina, San Diego, CA) was used with 2 × 150 bp paired-end (PE) sequencing mode. WES TMB was calculated using the previously described uniform method using two Novaseq S4 flowcells generating ∼400M PE 150-bp reads on tumor and ∼135M reads on normal samples to generate a median target coverage of >400× in tumor and >200× in normal tissue.[20] GATK-based Sentieon pipeline (version v201808) was used to call somatic variants (https://github.com/FNL-MoCha/nextgenseq_pipeline).

Gene panel assay sequencing and TMB estimation

Aliquoted DNA samples extracted from clinical samples and cell lines were distributed to all 16 participating laboratories, and each used their own sequencing and bioinformatics pipelines to estimate TMB from the genes represented in their respective panel assays. Some of these pipelines have been previously published (Table 1). Clinical samples were run as singletons and cell lines were run in duplicate or triplicate as available.

Panel assay size analysis

The simulated positive percent agreement (PPA) and negative percent agreement (NPA) of each of the panel assays (in silico) were calculated as a function of both the size of the panel assay used for its calculation as well as the respective TMB cut-off (Supplementary Material, available at https://doi.org/10.1016/j.annonc.2021.09.016).

Panel gene content analysis

Ten laboratories volunteered their BED file formats to anonymously evaluate the gene content of their panel assays. All panel data were lifted over to hg19 coordinates if they were not already. The intervals in these panel assays were intersected with the xgen-exome-research-panel-v2-targets exome reference panel assay. TCGA mutations from WES (in MAF format) were then overlayed on to the panel assays. We explored the removal of variants flagged as pathogenic as per the Catalogue of Somatic Mutations in Cancer (COSMIC) version 88, as well as synonymous variants, to determine the impact of including or excluding certain variants. TMB estimates per sample and per gene were tabulated.

Germline analysis

Three laboratories that use a tumor-only approach for the removal of germline variants volunteered to estimate the TMB value of the 29 clinical samples using three specific population minor allele frequency (pMAF) thresholds (0%, 0.5%, and 1%) to assess the impact that different population pMAF thresholds have on TMB estimates. Each laboratory used their own combination of population allele databases, including some custom databases, but no additional methods for the removal of germline variants were used (i.e. custom copy number-based germline prediction methods). (Supplementary Material, available at https://doi.org/10.1016/j.annonc.2021.09.016).

Calibration analysis

Statistical analyses were conducted to develop calibration curves specific to each panel assay that would facilitate translation of panel TMB values to WES TMB values. For each panel assay, two potential calibration curves were constructed. One curve modeled the association between panel TMB and WES TMB based on in silico analysis of the TCGA validation data as previously described.[18] These WES TMB values, which were previously calculated, are available on Precision FDA (https://precision.fda.gov/). The second curve modeled the association based on the "wet lab" results obtained on 10 human tumor cell lines newly generated and reported on in the current article. Parameter estimates were then used to compute 95% prediction limits. WES-calibrated TMB estimates and 95% intervals of uncertainty were obtained by inverting the fitted regression line and prediction limits. Supplementary Figure S1, available at https://doi.org/10.1016/j.annonc.2021.09.016, provides a pictorial representation of the calibration process. Additional documentation describing details of the model fit, calculation of the prediction limits, and method of obtaining the WES-calibrated TMB estimates and intervals of uncertainty can be found in Supplementary Material, available at https://doi.org/10.1016/j.annonc.2021.09.016.

After fitting the calibration curves for each panel assay according to the TCGA and cell line methods (training sets), the calibration curves were applied to TMB measurements generated by the panel assay on a completely independent

Table 1. Characteristics of 16 participating diagnostic NGS panel assays

Factor type	ACT	AZ	BWH	Caris	FMI	ILLUM	IPG	JHU	MSKCC	NeoGenomics	OmniSeq	PGDx	Q2	QIAGEN	Thermo_OCA	Thermo_OTMLA
Panel assay characteristics																
Name of panel assay	ACTOnco+	AZ650	OncoPanel v3.1	SureSelectXT	F1 CDx	TSO500	TheraMap Solid Tumor (TSO500)	JHOP2	MSK-IMPACT	NeoTYPE Discovery Profile for Solid Tumors	Ion AmpliSeq Comprehensive Cancer Panel	PGDx elio tissue complete	TSO500	QIAseq TMB panel	Oncomine Comprehensive Assay Plus (OCA Plus)	Oncomine Tumor Mutation Load Assay (OTMLA)
Number of genes	440	649	447	592	324	523	523	432	468	372	409	505	523	486	517	409
TMB region covered	1.1 Mb	1.65 Mb	1.94 Mb	1.40 Mb	0.8 Mb	1.33 Mb	1.27 Mb	1.14 Mb	1.14 Mb	0.935 Mb	1.17 Mb	1.3 Mb	1.2 Mb	1.33 Mb	1.06 Mb	1.2 Mb
Processing																
Minimum DNA input	40 ng	100 ng	50 ng	50 ng	50 ng	40 ng	40 ng	50 ng	150 ng	20 ng	30 ng	50 ng	40 ng	40 ng	20 ng	20 ng
Quantification method	Fluorescence	Fluorescence	Fluorescence	Electrophoresis	Fluorescence	Fluorescence	Fluorescence	Electrophoresis	Fluorescence	Fluorescence	Fluorescence	Fluorescence	Fluorescence	Fluorescence	Fluorescence	Fluorescence
Technology uses UMIs	No	Yes	No	No	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	No	No
Deduplication	No	Yes, UMI based	Yes, not UMI based	Yes, not UMI based	Yes, UMI based	Yes, UMI based	Yes, UMI based	Yes, not UMI based	Yes, not UMI based	Yes, UMI based	No	Yes, UMI based	Yes, UMI based	Yes, UMI based	No	No
Sequencing																
Seq platform Library prep/target enrichment	Ion Torrent Amplicon	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Hybrid	Illumina Amplicon	Ion Torrent Amplicon	Illumina Hybrid	Illumina Hybrid	Illumina Amplicon (single primer extension)	Ion Torrent Amplicon	Ion Torrent Amplicon
Sample-level minimum coverage threshold	800×	NA	30×	300×	250×	150×	300×	300×	50×	500×	125×	100×	50×	100×	500×	500×
Sample-level avg coverage for cell line exercise	1400×	1060.5×	394×	750×	982×	549×	800×	>400×	753×	>500×	314×	1517×	100×	500×	2000×	1300×
Variant-level minimum coverage	20×	50×	50×	100×	100×	50×	50×	50×	20×	100×	20×	Position-specific threshold. Determined by ML	150×	100×	60×	60×
Variant-level minimum read (ALT depth)	20	5	5	10	5	2[a]	3	3	8	10	4	6	2	4	10	10
Variant calling																
Type of variant	Non-synonymous and synonymous	Non-synonymous and synonymous	Non-synonymous only	Non-synonymous only	Non-synonymous and synonymous	Non-synonymous and synonymous	Non-synonymous and synonymous	Non-synonymous and synonymous	Non-synonymous only	Non-synonymous and synonymous	Non-synonymous only	Non-synonymous and synonymous	Non-synonymous and synonymous	Non-synonymous only	Non-synonymous only	Non-synonymous only
Germline variant filtration approach	Tumor only	Normal tissue	Tumor only	Tumor only	Tumor only	Tumor only	Tumor only	Tumor only	Normal tissue	Tumor only	Tumor only	Tumor only	Tumor only	Tumor only	Tumor only	Tumor only
Removes variants from known cancer genes	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	No
Published performance characteristics			30,31	32	9,20,33,34				9,35-37			38				39-42

ACT, ACT Genomics; AZ, AstraZeneca; BWH, Brigham and Women's Hospital; Caris, Caris Life Sciences; FMI, Foundation Medicine; ILLUM, Illumina; IPG, Intermountain Precision genomics; JHU, Johns Hopkins University; Mb, megabase; MSKCC, Memorial Sloan Kettering Cancer Center; NGS, next generation sequencing; PGDx, Personal Genome Diagnostics; Q2, Q Squared Solutions; Thermo_OCA, Thermo Fisher Scientific Oncomine Comprehensive Assay; Thermo_OTMLA, Thermo Fisher Scientific Oncomine Tumor Mutation Load Assay; UMI, Unique Molecular Identifier.

[a] Plus sample- and position-specific cut-off by likelihood ratio model.

set of 29 clinical samples that were not used in any way to develop the calibration curves (testing set). For each clinical sample, the uncalibrated, TCGA-calibrated, and cell line-calibrated panel TMB values were visually compared by boxplots. Root mean squared error (RMSE), relative to the observed WES value for each sample, was calculated on the sample and on the panel assay level.

The calibration tool, tmbLab, is an open-source software package written in the publicly available statistical software R that was created as part of this study.[21] This package, vignettes, documentation, and associated source code have been made freely available for public use at https://brb.nci.nih.gov/tmbLab/. The 'tmbLab' package was applied for the calibration analyses (Supplementary Methods, available at https://doi.org/10.1016/j.annonc.2021.09.016). Output produced by the package includes calibration plots as well as intercept, slope, and variance parameters associated with the fitted calibration curves relating panel TMB to WES TMB.

Compliance with ethics guidelines

Institutional review board (IRB) approval of the study protocol was obtained by each laboratory before study conduct. In all cases, the IRB determined this study is exempt from IRB review because it does not meet the definition of human subject research as defined in 45 CFR 46.102. Specifically, the investigators did not obtain information or biospecimens through intervention or interaction with individuals, and the DNA samples utilized by the participating laboratories was de-identified.

RESULTS

Variability across panel assays of participating laboratories

Sixteen targeted gene panel assays from academic and diagnostics laboratories participated in this study (Table 1). Each panel assay had a unique combination of characteristics that encompassed different sample processing requirements and sequencing platforms and chemistries. Each laboratory used their own analytical and bioinformatics methodologies to estimate TMB, which were optimized to their own panel assay specifications. If available, published panel assay performance characteristics are reported (Table 1). Size of the coding regions used to estimate TMB ranged from 0.8 to 1.94 Mb; minimum DNA input ranged from 20 to 150 ng and sample-level depth of coverage ranged between 30 and 800× for the participating laboratories. Seventy-five percent (12/16) of panel assays used an Illumina sequencing platform, while the others used the Thermo Fisher Scientific Ion Torrent platform. Sixty-three percent (10/16) used hybridization as a target enrichment approach, while the remaining panel assays used an amplicon-based approach.

The locally developed bioinformatics pipelines used in this phase II study also varied. All 16 panel assays included non-synonymous variants for TMB estimation, while 9 panel assays (56%) also included synonymous variants. Two panel assays used paired normal tissue to remove germline variants for TMB estimation, and the remaining 14 used their own tumor-only approach that utilized a combination of population frequency databases and proprietary methods for germline variant removal (Table 1). The variability in panel TMB values is described with boxplots in Figure 1 for the 25 clinical samples with WES TMB values <20 mut/Mb (Figure 1A and Supplementary Table S4), the 4 clinical samples with WES TMB values >20 mut/Mb (Figure 1B), and the 10 cell line samples (Figure 1C and Supplementary Table S5). Overall, the empirical variability across panel assays increased with increasing TMB value, which is consistent with findings of our previous study.[18] This trend in variance is evidenced by the wider (vertically stretched) boxplots proceeding from left to right within each figure and by comparing Figure 1A to Figure 1B. We noted that in clinical samples, WES TMB was occasionally lower than many of the reported panel TMB values (e.g. TMB-38, TMB-51, TMB-36), whereas in the cell lines, WES TMB was sometimes higher than many of the reported panel TMB values (e.g. NCI-H1437, NCI-H2009). Patient demographic and clinical variables as well as some specimen characteristics are also described via heatmaps below the boxplot figures (Figure 1).

Impact of panel assay size on panel TMB estimates

We used an *in silico* approach to estimate the impact of panel size on the PPA and NPA of TMB calling. At a TMB cut-off of 10, all 16 panel assays evaluated have a theoretical NPA of at least 95%, with a theoretical NPA falling <95% for panel sizes under 667 Kb (Figure 2A). The theoretical PPA at a TMB cut-off of 10 ranged from 87% to 92%, with a theoretical PPA falling <85% for panel sizes under 577 Kb. At a TMB cut-off of 5, theoretical NPAs ranged from 87% to 91%, while theoretical PPAs ranged from 86% to 92%, with larger panel assays having higher theoretical PPA and NPA. At TMB cut-offs of 15 and 20, theoretical NPAs ranged from 98% to 99%, while theoretical PPAs ranged from 88% to 92%.

While actual panel performance reflects many factors, including depth of sequencing and accuracy of mutation calling, we observed a substantial acceleration of decrease in PPA of panels at critical intersections of small panel sizes and low TMB cut-offs (Figure 2A). These findings support the hypothesis that small panels are insufficient to maintain adequate PPA and NPA for calling TMB high versus TMB low across the range of cut-offs for positivity likely to be used in practice.

Impact of panel assay gene content on panel TMB estimates

Failure to filter out pathogenic variants in panel TMB estimates results in overestimation of TMB relative to WES for all panel assays investigated (Figure 2B). In this *in silico* analysis, removing known pathogenic cancer gene mutations, as identified in COSMIC, showed a closer approximation to WES TMB. When synonymous variants are additionally filtered, thereby keeping only non-synonymous
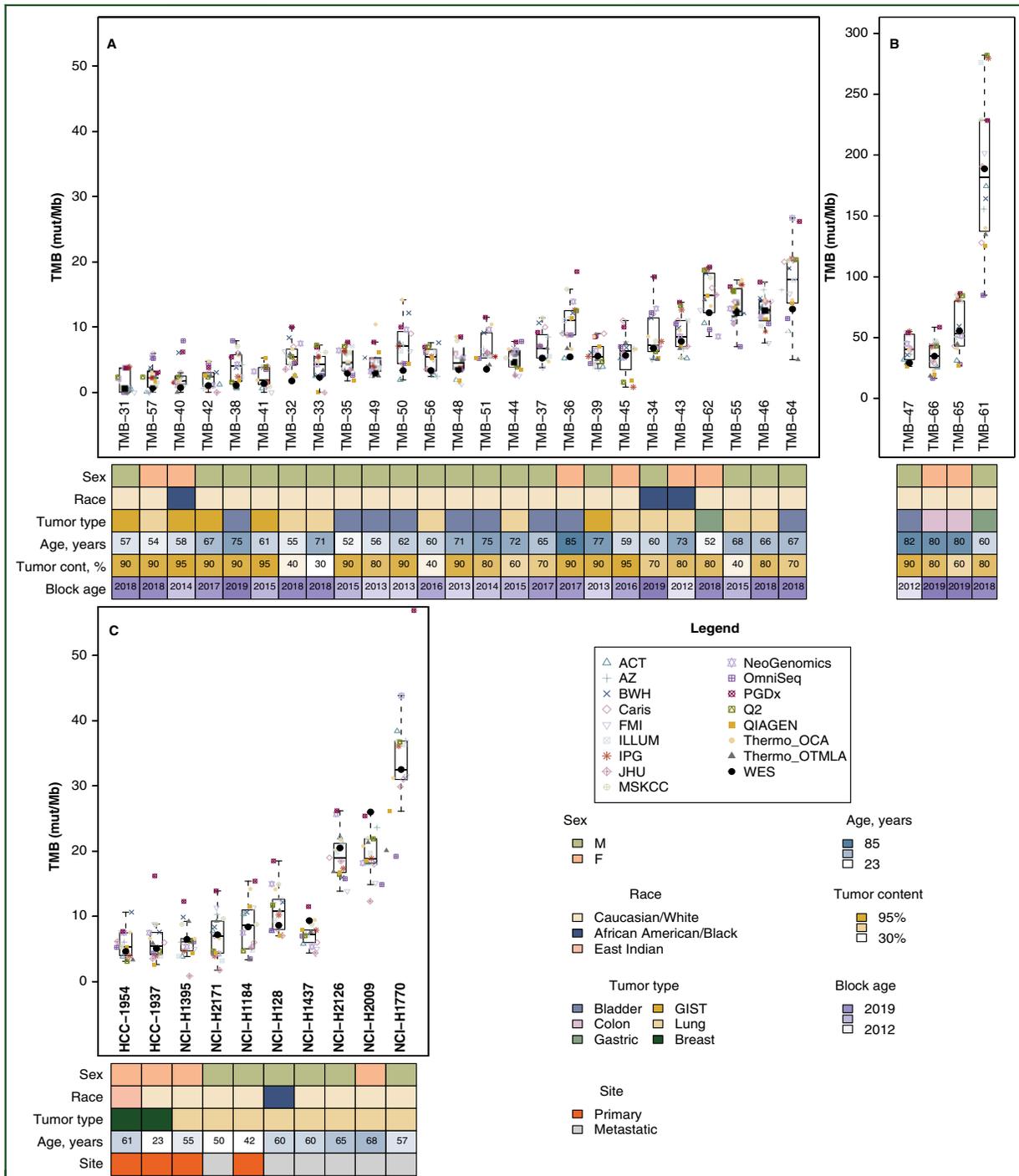
Figure 1. Variability of reported TMB values across panel assays participating in the experiment as depicted by boxplots.
Sample-level boxplots are ordered by observed WES TMB value (low to high). Heatmaps describe demographic and clinical characteristics of the sample.
(A) Clinical samples with WES TMB values <20. (B) Clinical samples with WES TMB values >20. (C) Cell lines.
ACT, ACT Genomics; AZ, AstraZeneca; BWH, Brigham and Women's Hospital; Caris, Caris Life Sciences; FMI, Foundation Medicine; GIST, gastrointestinal stromal tumor; ILLUM, Illumina; IPG, Intermountain Precision genomics; JHU, Johns Hopkins University; MSKCC, Memorial Sloan Kettering Cancer Center; PGDx, Personal Genome Diagnostics; Q2, Q squared Solutions; Thermo_OCA, Thermo Fisher Scientific Oncomine Comprehensive Assay; Thermo_OTMLA, Thermo Fisher Scientific Oncomine Tumor Mutation Load Assay; TMB, tumor mutational burden; WES, whole exome sequencing.
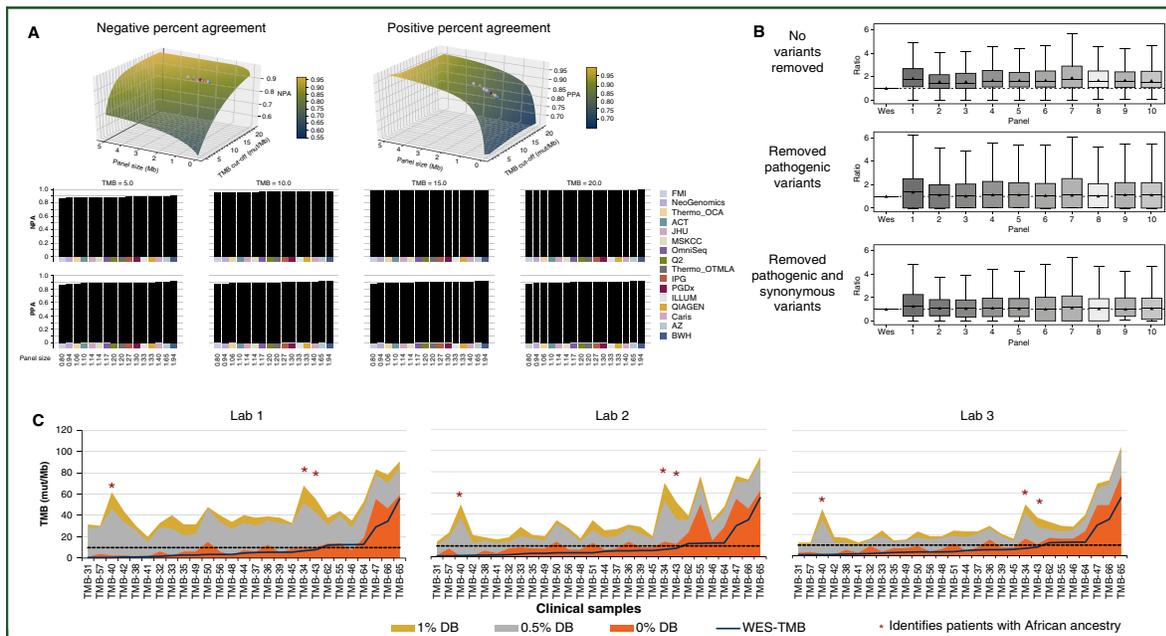
Figure 2. Panel assay design and bioinformatics factors affecting panel TMB estimates.
(A) Impact of panel assay size on NPA and PPA of panel TMB estimate, (B) impact of gene content (including pathogenic variants and synonymous variants) on panel TMB estimate, and (C) impact of population allele thresholds on germline variant filtering. *Identifies patients with African ancestry.

ACT, ACT Genomics; AZ, AstraZeneca; BWH, Brigham and Women's Hospital; Caris, Caris Life Sciences; FMI, Foundation Medicine; ILLUM, Illumina; IPG, Intermountain Precision genomics; JHU, Johns Hopkins University; MSKCC, Memorial Sloan Kettering Cancer Center; NPA, negative percent agreement; PGDx, Personal Genome Diagnostics; Q2, Q Squared Solutions; PPA, positive percent agreement; Thermo_OCA, Thermo Fisher Scientific Oncomine Comprehensive Assay; Thermo_OTMLA, Thermo Fisher Scientific Oncomine Tumor Mutation Load Assay; TMB, tumor mutational burden; WES, whole exome sequencing.

variants to estimate TMB, only a minimal effect is observed on panel TMB estimates as approximations to WES TMB. However, it was evident that removing synonymous variants also widened the boxplot, thus signaling greater variability across panel TMB estimates when the number of variants was reduced. Variability in this context was also associated with panel assay size. Boxplot width was the smallest for panel 2, which also corresponded to the largest panel assay (1.5 Mb). In contrast, panel 7, with the smallest panel assay (0.8 Mb), exhibited the greatest boxplot width.

Impact of germline variant filtering on panel TMB estimates

The tumor-only approach utilized by 14 out of 16 panel assays included the identification of common variants in a single or a combination of population-based genotyping databases (Supplementary Table S6, available at https://doi.org/10.1016/j.annonc.2021.09.016). Filtering out potential germline variant calls, defined as >0% of the pMAF, provides the strongest correlation to WES TMB independent of the panel assay utilized (Figure 2C). In some instances, use of 0% pMAF could even lead to underestimation of panel TMB. Conversely, setting the germline variant allele frequency filter to >0.5% pMAF significantly overestimates panel TMB compared to WES TMB and this effect is even more pronounced when the filter is raised to >1% pMAF. Notably, three of the clinical samples evaluated were from

patients of African descent (TMB-34, TMB-40, TMB-43) and were observed to have panel TMB values that were grossly overestimated by the majority of platforms, especially if 0.5% or 1% pMAF thresholds were used for the removal of germline variants.

Calibration tool

The range of fitted calibration curve slopes across the panel assays was 0.868-1.647 when TCGA data were used as the calibration reference, and 0.551-1.142 when the cell line data were used as the reference (Supplementary Tables S7 and S8, available at https://doi.org/10.1016/j.annonc. 2021.09.016). The TCGA- and cell line-derived calibration results are depicted in Figure 3 for samples with WES TMB values between 5 and 15. The boxplots of all 29 clinical samples are included in Supplementary Figure S2, available at https://doi.org/10.1016/j.annonc.2021.09.016. In general, the TCGA calibration approach tends to yield boxplots that are compressed and/or closer to the WES TMB value, when compared to uncalibrated TMB values. Numerically this is demonstrated by the lower RMSE (Supplementary Table S9, available at https://doi.org/10.1016/j.annonc. 2021.09.016); in particular, the RMSE for the TCGA-calibrated TMB values as compared to the uncalibrated TMB values is equal or lower in 26/29 (90%) clinical samples (Supplementary Table S9, available at https://doi.org/10. 1016/j.annonc.2021.09.016). In contrast, the cell line
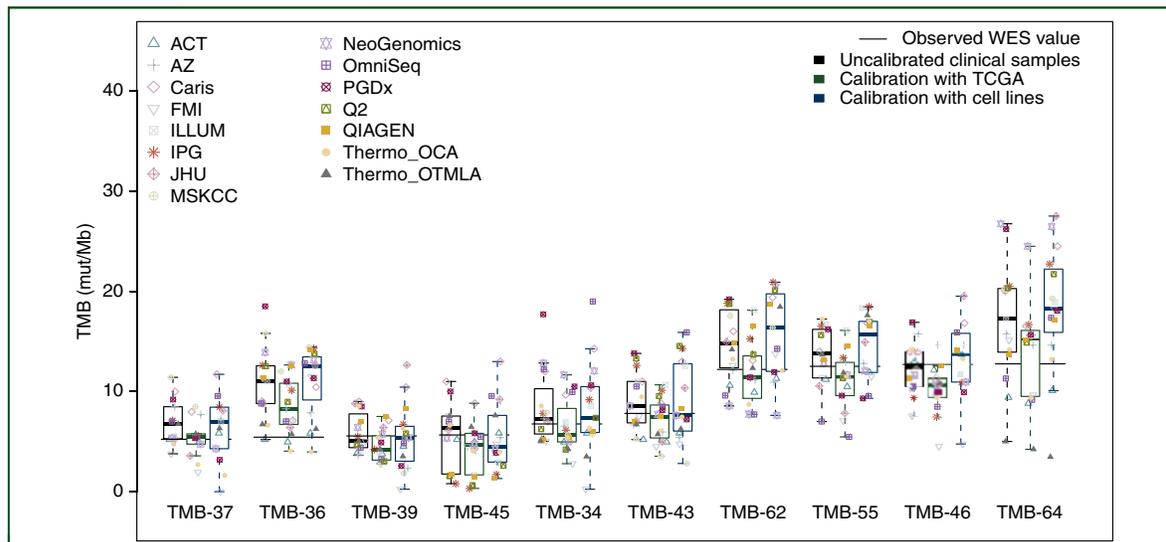
Figure 3. Application of two calibration approaches (TCGA and cell line) to the clinical samples.
Sets of boxplots are depicted for samples with WES TMB values between 5 and 15 (full sample set depicted in Supplementary Figure S2, available at https://doi.org/10.1016/j.annonc.2021.09.016). For each sample, the three boxplots correspond to (left to right) the uncalibrated TMB values, the TCGA-calibrated TMB values, and the cell line-calibrated TMB values. The black horizontal line across the three boxplots corresponds to the observed WES TMB value.
ACT, ACT Genomics; AZ, AstraZeneca; BWH, Brigham and Women's Hospital; Caris, Caris Life Sciences; FMI, Foundation Medicine; ILLUM, Illumina; IPG, Intermountain Precision genomics; JHU, Johns Hopkins University; MSKCC, Memorial Sloan Kettering Cancer Center; PGDx, Personal Genome Diagnostics; Q2, Q Squared Solutions; TCGA, The Cancer Genome Atlas; Thermo_OCA, Thermo Fisher Scientific Oncomine Comprehensive Assay; Thermo_OTMLA, Thermo Fisher Scientific Oncomine Tumor Mutation Load Assay; TMB, tumor mutational burden; WES, whole exome sequencing.

calibration approach does not yield less variable/less biased boxplots or lower RMSE than uncalibrated TMB values (Figure 3, Supplementary Figure S2 and Table S9, available at https://doi.org/10.1016/j.annonc.2021.09.016). RMSE was also calculated on the panel assay level and shows that RMSE may increase or decrease on the panel assay level (Supplementary Table S10, available at https://doi.org/10.1016/j.annonc.2021.09.016).

DISCUSSION

In an environment where diverse NGS assays will be available, to confidently use TMB estimation in clinical decision making, sources of measurement variability must be understood and controlled for when interpreting results. In this large collaboration-driven study, we describe the empirical variability in TMB estimation across 16 different panel assays applied to a common set of FFPE clinical tumor samples and to human tumor-derived cell lines. Additionally, we developed a publicly available calibration tool to align TMB estimates using different panel assays.

A certain degree of variability in the estimation of TMB on clinical samples across panel assays was expected, similar to our *in silico* assessment.[18] Factors such as panel assay content, sequencing platforms, and bioinformatics pipelines were expected to contribute to variability. Since standardization of these variables is impractical, we utilized publicly available samples to quantitatively characterize the empirical variability in panel TMB estimation and provide the opportunity to achieve more consistent results through calibration.

Our results agree with previous reports showing that a sufficiently sized panel is required to maintain reasonable PPA of panel TMB measurements.[22-26] There is a small but consistent association between panel assay size and the PPA and NPA, regardless of the TMB cut-off. However, we also found relatively marginal gains in assay performance above a certain threshold of panel size.

In addition to size alone, gene content is also a key factor. We show that filtering out known cancer gene mutations, as identified in COSMIC, significantly improved the accuracy of panel TMB estimates relative to WES TMB for all of the panel assays. Another approach is to remove synonymous alterations and count only non-synonymous variants when estimating TMB; seven participating laboratories did so in our study. However, this did not significantly affect accuracy of TMB estimates in the clinical samples (perhaps related to few silent alterations in the gene regions tested by each panel assay).[23,27]

Another issue in TMB estimation is the impact of tumor-only sequencing, which can lead to inadvertent inclusion of germline variants. Inclusion of germline variants within 1% pMAF as part of a tumor-only germline variant removal approach resulted in significant overestimation of panel TMB, which has also been observed by Parikh et al.[28] Population databases are commonly used by various panel assays, and here we showed that the most stringent filtering approach, using a filter of >0% pMAF, offers the closest approximation to WES TMB compared to other pMAF values (0.5% or 1%). Additionally, TMB values for patients of African descent within our clinical samples were overestimated. Analysis using more than one population

database may help to reduce biases, especially as databases may vary with their representation of different racial ancestries.[29] Overall, it is important to accurately filter germline variants using available bioinformatics methods. Additionally, the use of FFPE specimens may have an impact on TMB estimation by generating false positives due to artifacts created during the fixation process. These factors must be considered and assessed during assay development, including development of the bioinformatics pipeline to reduce potential false positives. Clinically, if the TMB value of a cancer is close to a predetermined threshold that would make it eligible for treatment, the variability added by the suboptimal removal of germline variants could translate to potential overtreatment of patients and unnecessary exposure to immune-related adverse events. Patient-matched normal samples are not always available to identify a patient's germline variants for filtration. Thus, it is important to accurately filter germline variants using available bioinformatics methods.

Beyond characterizing and quantifying factors that can impact variability in panel TMB estimates, we also built a tool to promote alignment and optimize the functionality of TMB as a clinical biomarker.

Our calibration tool aims to improve clinical consistency and interpretability and is a free and open-source software. Ideally, a calibration tool could be used for regulatory purposes to permit different tests to align to common treatment recommendations, resulting in expanded patient access and reduced variability in oncology care. Application of the calibration tool using TCGA data as a reference does not account for differences in wet-lab procedures across panel assays. We attempted to use human tumor-derived cell lines as a reference material; however, there were insufficient cell lines with matched normal cell lines and calibration using the 10 cell lines in this study did not meaningfully reduce variability (Figure 3). See Supplementary Table S11 for considerations for the use of different sources as reference material. More generally, the calibration tool reduced the overall variability across laboratories, but calibration did not improve concordance between panel TMB and WES TMB for every lab. Further work is needed to optimize the calibration tool for this purpose. Our findings should be interpreted considering several limitations, including the heterogeneity of tumor specimens acquired and inclusion of a few tumors for which immunotherapies are less relevant (i.e. GIST cancers) as well as use of samples with high tumor purity (≥30%) which may not represent all samples acquired in the clinical setting.

Despite these limitations, our tool effectively demonstrates that calibration of panel TMB values can be achieved to an extent that supports development and utilization of TMB applications across platforms. While use of each TMB platform will likely be optimized to specific drug indications, there is value in considering the harmonization and standardization principles we present here. Based on our findings, we strongly encourage diagnostics developers to conduct their own calibration analyses and compare their panel assays to others in order to achieve optimal reproducibility and improve assay utility in the clinic. Clinicians can use findings from this study to contextualize a single TMB output. Clinically, if the TMB value of a cancer is close to a predetermined threshold that would make a patient eligible for treatment, being able to recognize variability of individual panel-level TMB values could help avoid potential over- or undertreatment of patients or unnecessary exposure to immune-related adverse events. In addition to direct clinical care, calibration may facilitate synthesizing panel TMB data across studies for translational research and enable increased scale and power of studies to examine TMB along with other predictors of response to immunotherapy.

Conclusion

The TMB Harmonization Project leveraged the expertise and insight of 16 different diagnostic laboratories to objectively evaluate the empirical variability across panel TMB values and to propose best practices for panel TMB alignment. Our work demonstrates that the utilization of a calibration tool based on a universal reference standard derived from TCGA data can enhance comparability of TMB across different panel assays. Use of different NGS platforms for TMB testing will necessitate a combinatorial approach, including consensus guidelines and availability of a universal reference standard, in order to maintain a satisfactory level of consistency in the measurement and clinical application of this complex biomarker. Availability of reference material annotated with analytical and associated clinical truth would be of value to assay development efforts. Our results provide proof of principle that this level of alignment is achievable and will support the consistent assessment, adoption, and application of TMB to optimally guide immunotherapy decisions. We hope that this process can serve as a model for future biomarker technologies and alignment efforts.

REFERENCES

1. Vaddepally RK, Kharel P, Pandey R, Garje R, Chandra AB. Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. *Cancers*. 2020;12:738.
2. Adam T, Becker TM, Chua W, Bray V, Roberts TL. The multiple potential biomarkers for predicting immunotherapy response-finding the needle in the haystack. *Cancers*. 2021;13:277.
3. Klempner SJ, Fabrizio D, Bane S, et al. Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence. *Oncologist*. 2020;25:e147-e159.
4. Wang C, Yu X, Wang W. A meta-analysis of efficacy and safety of antibodies targeting PD-1/PD-L1 in treatment of advanced nonsmall cell lung cancer. *Medicine (United States)*. 2016;95:e5539.
5. Rizvi H, Sanchez-Vega F, La K, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol*. 2018;36:633-641.
6. Hellmann MD, Ciuleanu T-E, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378:2093-2104.
7. Hellmann MD, Callahan MK, Awad MM, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. *Cancer Cell*. 2018;33:853-861.e4.
8. Marabelle A, Fakih M, Lopez J, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol*. 2020;21:1353-1365.
9. FDA unveils a streamlined path for the authorization of tumor profiling tests alongside its latest product action. Available at https://www.fda.gov/news-events/press-announcements/fda-unveils-streamlined-path-authorization-tumor-profiling-tests-alongside-its-latest-product-action. Accessed April 30, 2021.
10. Stenzinger A, Allen JD, Maas J, et al. Tumor mutational burden standardization initiatives: recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions. *Genes Chromosomes Cancer*. 2019;58:578-588.
11. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. *J Thoracic Oncology*. 2017;12:208-222.
12. Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non—small cell lung cancer. *JAMA Oncology*. 2017;3:1051-1058.
13. Kerr KM. The PD-L1 Immunohistochemistry biomarker: two steps forward, one step back? *J Thorac Oncol*. 2018;13:291-294.
14. Diggs LP, Hsueh EC. Utility of PD-L1 immunohistochemistry assays for predicting PD-1/PD-L1 inhibitor response. *Biomark Res*. 2017;5:12.
15. Abdul Karim L, Wang P, Chahine J, Kallakury B. Harmonization of PD-L1 immunohistochemistry assays for lung cancer: a working progress. *J Thorac Oncol*. 2017;12:e45.
16. Thunnissen E, de Langen AJ, Smit EF. PD-L1 IHC in NSCLC with a global and methodological perspective. *Lung Cancer*. 2017;113:102-105.
17. Salgado R, Bellizzi AM, Rimm D, et al. How current assay approval policies are leading to unintended imprecision medicine. *Lancet Oncol*. 2020;21:1399-1401.
18. Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J Immunother Cancer*. 2020;8:e000147.
19. Stenzinger A, Endris V, Budczies J, et al. Harmonization and standardization of panel-based tumor mutational burden measurement: real-world results and recommendations of the quality in pathology study. *J Thorac Oncol*. 2020;15:1177-1189.
20. Fabrizio DA, Milbury C, Yip W-K, et al. Analytic validation of tumor mutational burden as a companion diagnostic for combination immunotherapy in non-small cell lung cancer. *Ann Oncol*. 2018;29:VIII16.
21. Team R Development Core. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2018. Available at https://www.R-project.org/. Accessed April 30, 2021.
22. Wang Z, Duan J, Cai S, et al. Assessment of blood tumor mutational burden as a potential biomarker for immunotherapy in patients with non-small cell lung cancer with use of a next-generation sequencing cancer gene panel. *JAMA Oncol*. 2019;5:696-702.
23. Allgäuer M, Budczies J, Christopoulos P, et al. Implementing tumor mutational burden (TMB) analysis in routine diagnostics—a primer for molecular pathologists and clinicians. *Transl Lung Cancer Res*. 2018;7:703-715.
24. Meléndez B, van Campenhout C, Rorive S, Remmelink M, Salmon I, D'Haene N. Methods of measurement for tumor mutational burden in tumor tissue. *Transl Lung Cancer Res*. 2018;7:661-667.
25. Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. 2018;24:1441-1448.

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

27. Fancello L, Gandini S, Pelicci PG, Mazzarella L. Tumor mutational burden quantification from targeted gene panels: major advancements and challenges. *J Immunother Cancer*. 2019;7:183.

28. Parikh K, Huether R, White K, et al. Tumor mutational burden from tumor-only sequencing compared with germline subtraction from paired tumor and normal specimens. *JAMA Netw Open*. 2020;3:e200202.

29. Green ED, Gunter C, Biesecker LG, et al. Strategic vision for improving human health at The Forefront of Genomics. *Nature*. 2020;586:683-692.

30. Garcia EP, Minkovsky A, Jia Y, et al. Validation of oncopanel a targeted next-generation sequencing assay for the detection of somatic variants in cancer. *Arch Pathol Lab Med*. 2017;141:751-758.

31. Hanna GJ, Lizotte P, Cavanaugh M, et al. Frameshift events predict anti-PD-1/L1 response in head and neck cancer. *JCI Insight*. 2018;3:e98811.

32. Vanderwalde A, Spetzler D, Xiao N, Gatalica Z, Marshall J. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Med*. 2018;7:746-756.

33. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013;31:1023-1031.

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

35. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn*. 2015;17:251-264.

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

37. Zehir A, Benayed R, Shah RH, et al. Erratum: mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23:1004.

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

39. Heeke S, Benzaquen J, Hofman V, et al. Comparison of three sequencing panels used for the assessment of tumor mutational burden in NSCLC reveals low comparability. *J Thorac Oncol*. 2020;15:1535-1540.

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

41. Alborelli I, Bratic Hench I, Chijioke O, et al. Robust assessment of tumor mutational burden in cytological specimens from lung cancer patients. *Lung Cancer*. 2020;149:84-89.

42. Wong CN, Fessas P, Dominy K, et al. Qualification of tumour mutational burden by targeted next-generation sequencing as a biomarker in hepatocellular carcinoma. *Liver Int*. 2021;41:192-203.

Assessing the Use of ctDNA as an Early Endpoint in Early-Stage Disease

FRIENDS OF CANCER RESEARCH
ANNUAL MEETING 2021

Introduction

Circulating tumor DNA (ctDNA) is a dynamic biomarker with potentially broad clinical and regulatory applicability in oncology. To date, the use of ctDNA has been studied to the greatest extent in the metastatic solid tumor setting for molecular profiling at diagnosis, targeted therapy selection, treatment response monitoring, and long-term post-treatment tumor surveillance.[1] However, there is great opportunity and potential value to patients to further explore the use of ctDNA in early-stage solid tumors including:

- Determining the need for adjuvant therapy after definitive surgery, radiation, or chemoradiation by indicating the presence of minimal (or molecular) residual disease (MRD) or optimizing neoadjuvant therapy regimens,
- Monitoring for disease recurrence in a simpler and less invasive way compared to existing tools (e.g., clinical imaging, biopsies),
- Enabling the identification of patients at the highest risk of recurrence for enrollment in clinical studies (prognostic enrichment strategies), reducing patient numbers as well as the time and cost of studies, and
- Serving as a potential predictive biomarker for a patient's response to therapy as an early endpoint to predict long-term survival outcomes, allowing for faster identification of drugs that may be most efficacious and support regulatory decision-making.[2]

Objectives

Detail the opportunities and challenges in using ctDNA in the early-stage disease setting.

Identify and prioritize clinical questions supporting its use as an early endpoint to support regulatory approval.

Define data elements where alignment is needed across datasets for easier contextualization and analysis to answer these questions.

Thank You to Our Contributors

AstraZeneca
Carmen Lee
Chris Abbosh
Darren Hodgson
Jian Wang
Rachel Hodge

Bristol Myers Squibb
George Green
Joseph Fiore
Jonathan Baden

Genentech, a Member of the Roche Group
Eric Peters
Josina Reddy
Kathleen Winson
Silke Maier

Johns Hopkins School of Medicine
Valsamo Anagnostou

Natera, Inc.
Alexey Aleshin
Stephen Willingham

Personal Genome Diagnostics
Mark Sausen

Predicine
Shidong Jia

U.S. Food and Drug Administration
Anand Pathak
Erik Bloomquist
Julia Beaver
Paz Vellanki
Reena Philip
Shenghui Tang

Ultimately, the hope is that use of ctDNA in early-stage disease will improve the approach to drug development in this setting, enabling effective therapies to get to patients faster.
To explore the opportunities and unique challenges for use of ctDNA in early-stage solid tumors, Friends of Cancer Research (*Friends*) convened a multi-stakeholder group of experts in ctDNA and early-stage disease including the U.S. Food and Drug Administration (FDA), drug sponsors, ctDNA assay developers, and academic clinicians. The working group focused on the use of ctDNA as an early endpoint to predict long-term survival outcomes to support regulatory approval, noting the need for collaboration across sponsors. The working group strongly endorsed this collaboration for multiple reasons. There is recognition that validating the use of ctDNA as an early endpoint in early-stage disease will require large amounts of data from multiple prospective clinical trials. These data will need to represent robust clinical outcomes and come from multiple sources.[3] Also, a more coordinated and collaborative approach will help to accelerate the understanding of ctDNA in this setting and to establish ctDNA as a potential early endpoint earlier. Lastly, previous collaborative efforts in this space have laid the foundation for this effort. Friends established a multi-phased collaborative research initiative to harmonize the use of ctDNA to monitor treatment response (ctMoniTR) to determine if changes in ctDNA levels accurately reflect the therapeutic effect of immunotherapies in advanced lung cancer.[4,5] The ctMoniTR Project affirmed that multiple sponsors can work collaboratively to effectively combine data from multiple clinical trials to demonstrate a correlation between ctDNA and response, and has expanded efforts to a second phase that includes additional cancer types and treatments. With this foundation for a collaborative framework, the working group discussed the investigation of ctDNA as an early endpoint in early-stage disease to support its use in regulatory decision-making. Through these findings and leveraging previous ctMoniTR work, we propose a collaborative effort to align data from multiple trials for the investigation of the use of ctDNA as an early endpoint in early-stage disease.

Opportunities for Use of ctDNA in Early-Stage Disease

There are numerous opportunities to utilize ctDNA in early-stage disease that rely on the potential to detect disease burden, such as MRD or molecular relapse, earlier and in a less invasive manner than standard of care imaging technologies or tissue biopsy. Opportunities exist at various stages of established use and validity in oncology, summarized in **Figure 1**. Within the use cases of ctDNA in early-stage disease, one large category pertains to informing and assessing efficacy of therapies. We detail these use cases below.

Figure 1: Potential Use Cases of ctDNA in Oncology. Depicted is a time course through a patient's cancer treatment journey and the opportunities for use of ctDNA to guide treatment. (Adapted from Natera)
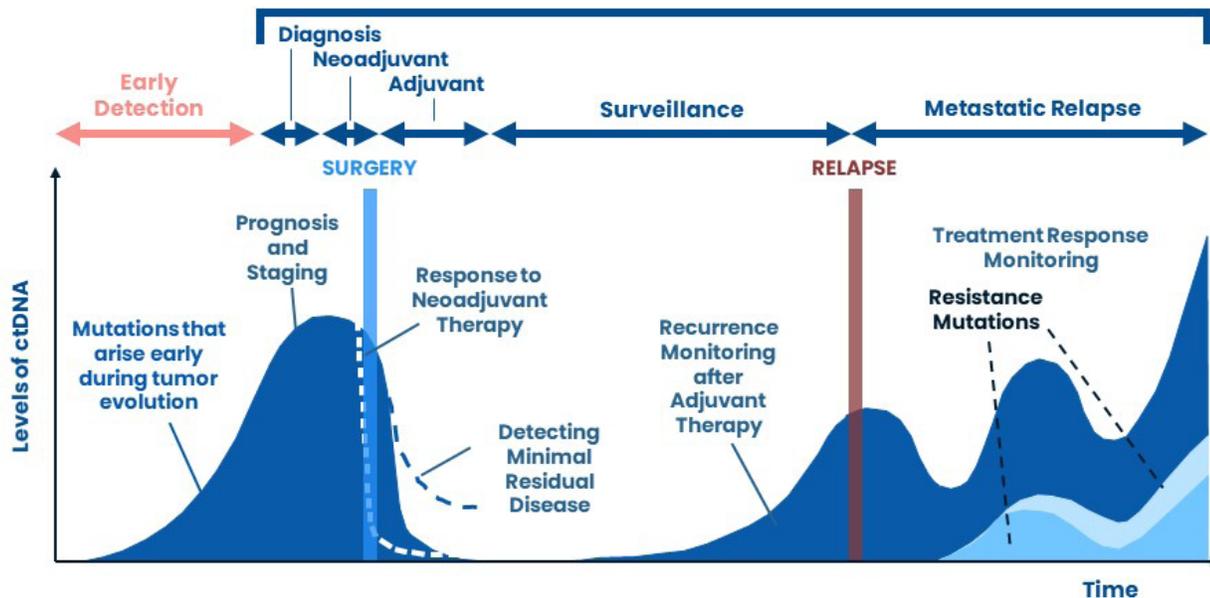
ctDNA for Risk Stratification and Treatment Selection

Evidence is emerging on the potential to detect MRD by ctDNA assessment post-surgery to guide decisions on adjuvant therapy. The prognostic value has been demonstrated across multiple tumor types, demonstrating that the detection of ctDNA post-definitive intervention could be utilized to direct patients to appropriate adjuvant therapy in early-stage disease or potentially spare them of unneeded treatment. A study of patients with operable urothelial cancer found that the presence of ctDNA after surgery was significantly associated with poor prognosis and those with detectable ctDNA appeared to derive the most relative benefit with adjuvant immunotherapy.[6] Additionally, multiple studies in early-stage colorectal cancer found the presence of ctDNA after surgery strongly correlated with recurrence[7,8] and inferior disease-free survival (DFS), after adjusting for clinicopathological risk factors.[9]

ctDNA for Patient Selection

Utilizing the prognostic value of ctDNA in the early-disease setting, MRD-selected adjuvant trials can help define a more homogenous patient population with higher relapse-event rates, leading to smaller, higher-risk patient populations and reduced time to reach endpoints. In the early-stage disease setting, adjuvant trial patient populations are heterogeneous with low relapse-event rates leading to the need for large numbers of patients to adequately power studies to analyze outcomes and reach their endpoints, which can also take a significant amount of time and expose some patients to treatments which they ultimately may not need. The potential value of ctDNA in this setting is highlighted by recently launched Phase III trials,

including the MERMAID-1 and MERMAID-2 trials to assess adjuvant treatment in patients with resected stage II and III NSCLC with MRD by ctDNA measurement[10] and the IMvigor011 trial to assess adjuvant treatment in patients with muscle-invasive bladder cancer who are ctDNA positive after cystectomy.[11]

ctDNA to Monitor and Predict Treatment Response in the Neoadjuvant and Adjuvant Settings
Measuring serial ctDNA prior to and throughout treatment may be useful to monitor response to treatment as an early endpoint to potentially predict long-term outcomes. This has been illustrated in a retrospective analysis of prospectively collected samples in early-stage breast cancer where clearance of ctDNA was a predictor of pathological complete response (pCR) to neoadjuvant treatment and was associated with a lower risk of recurrence.[12] In the adjuvant setting, ctDNA can also potentially be used to monitor and predict treatment response as an early endpoint. Early work by a pair of small prospectively designed studies at academic institutions investigating serial ctDNA collection during adjuvant chemotherapy treatment from patients with locally resected colon cancer found that increases in ctDNA levels during treatment was an early indicator of radiologic recurrence[7] and could be an early predictor of relapse.[9] Evidence from these early phase studies supports the association of ctDNA changes as an early predictor of treatment outcome and suggests there is an opportunity in both the neoadjuvant and adjuvant settings to further generate robust evidence.

Defining a Specific Use Case: ctDNA Changes as an Early Endpoint

While there are many possible use cases for ctDNA in early-stage disease, for the purposes of this white paper, the group decided to focus on ctDNA changes (e.g., clearance, reductions, kinetics) in response to therapy as an early endpoint to predict long-term survival outcomes to support regulatory approval. We use the term "early endpoint" for the purposes of this white paper to distinguish the potential to measure ctDNA changes earlier than other endpoints (e.g., disease-free survival, event-free survival, and overall survival) rather than defining the timeframe of when the endpoint is measured (i.e., not insinuating ctDNA measurement occurs early in a clinical trial, as this may vary based on the context of different cancer types or treatment settings). In order for ctDNA to support an Accelerated Approval as a primary efficacy endpoint, ctDNA changes would need to be proven to be reasonably likely to predict clinical benefit. This utility may have therapeutic class specific (e.g., chemotherapy, immunotherapy, targeted therapy, etc.) and tumor type specific considerations, however more data and evidence are needed to delineate these factors. Many clinical and technical questions exist regarding use of ctDNA as an early endpoint and robust evidence generation will be necessary to support its use for regulatory decision-making.

Challenges and Variability in ctDNA Detection

In early-stage disease, there are low amounts of ctDNA due to the small, localized nature of these tumors, and detecting the levels may be limited by current technologies. Furthermore,

ctDNA levels vary due to differences in tumor growth rate (e.g., indolent vs. fast-progressing), tumor ctDNA shedding rates, and other biological factors, which vary significantly between different tumor types and metastatic sites (e.g., intracranial metastases).[3] Additionally, both personalized (tumor informed) and non-personalized (plasma only) approaches integrating varying single-omic or multi-omic approaches (e.g., sequence mutations, structural alterations, methylation, fragmentomics, etc.) and platforms are currently being utilized with many others in development. These approaches, coupled with clinical variables and trial methodology, result in significant sources of variability in early-stage disease ctDNA clinical studies **(Table 1)**.

Table 1: Sources of Variability in Early-Stage Disease ctDNA Clinical Studies

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

Key Questions for the Use of ctDNA in Early-Stage Disease

Amidst this variability, there are many questions regarding the ability to use ctDNA in early-stage disease as an early endpoint that will be critical to address. These include both technical and clinical questions.

Key Technical Questions to Be Addressed to Enable the Use of ctDNA in Early-Stage Disease
Due to the multiple approaches and platforms for ctDNA detection in the early-stage disease setting, there are several technical questions regarding the feasibility and best approach for aligning the various methodologies to generate meaningful data on the use of ctDNA as an early endpoint. Questions fall into two categories:

Multi-Use Case Considerations
- Are there different minimum analytical performance requirements for different early-stage disease applications (e. g., neoadjuvant vs. adjuvant, tumor type, stage of disease, etc.)?
- Assuming there are minimum diagnostic analytical performance requirements, are there mechanisms to baseline/compare analytical performance (e.g., LOD, LOB, etc.) across different platforms from both a qualitative (ctDNA detection) and quantitative (ctDNA levels) perspective, and is there a common unit of measurement across assays?

Early Endpoint Considerations
- Given similar analytical performance, are different ctDNA features equally informative to reflect long-term outcomes after surgical or therapeutic intervention (neoadjuvant or adjuvant)?
- Do differences in sample collection (e.g., timing) and pre-analytical processing (e.g., whole blood collection and plasma preparation) affect the ability of ctDNA changes to reflect long-term outcomes?
- Given various lower LOD for different platforms, how can data be pooled and stratified based on the absence or presence of ctDNA to correlate with long-term outcomes?

Future work is needed by multi-stakeholder groups to prioritize the questions and further expand on the necessary evidence to answer these technical questions. There are few harmonized definitions across assays suggesting a need to define common assay metrics and standards to align across datasets.

Key Clinical Questions to be Addressed to Support ctDNA as an Early Endpoint
There are multiple clinical questions regarding the use of ctDNA changes as an early endpoint in early-stage disease. The prioritized questions center around whether changes in ctDNA following treatment reflect long-term outcomes (DFS/EFS and/or OS) at the patient

and trial level, as well as whether the ability to use ctDNA as an early endpoint varies by the therapy setting, therapeutic class, or tumor type. Additional considerations explore the nuances of these questions, focusing on the appropriate timing of ctDNA measurement to predict long-term outcomes, including further delineating the predictive value of a drug on reduction, increase, or clearance of ctDNA as reflected in long-term outcomes. These questions will also likely have different answers depending on the therapeutic setting and tumor type. In the adjuvant setting, it is important to look at ctDNA clearance, while percent change of ctDNA levels may be more relevant in the neoadjuvant setting where the tumor has not been removed. Key questions include:

Do ctDNA changes in response to a drug reflect long-term outcomes (DFS/EFS and/or OS)?
- For example, are certain categorical changes (reduction or rise) in ctDNA more predictive of long-term survival outcomes?

Does the predictive value of ctDNA vary by:
- early-stage disease therapy setting (e.g., neoadjuvant vs. adjuvant)?
- therapeutic class (e.g., immunotherapy, chemotherapy, targeted therapy)?
- tumor type?

When should ctDNA be measured (i.e., should there be set time points for measurement throughout treatment for all trials)?

What is the optimal threshold, in terms of percent change in ctDNA levels (or clearance), that should be used to define ctDNA response?

At what time point does ctDNA response (e.g., early response from pre-treatment to on-treatment, maintaining ctDNA response at a landmark on-treatment timepoint) correlate with long-term survival benefit?

Aligning on a Core Set of Data Elements for Assessing the Use of ctDNA in Early-Stage Disease

There are multiple pragmatic challenges with early-stage disease ctDNA studies including the size and time needed to reach clinical endpoints. Therefore, proactive planning of data elements and analysis methodology is important. To generate sufficient evidence to begin to answer these key clinical questions, collaboration across groups and clinical trials to aggregate data is necessary. If alignment on a core set of data elements occurs before prospective clinical trials are designed and executed, validating ctDNA as an early endpoint can be achieved more efficiently.

Technical Considerations
Due to the significant variability in the analytical approaches and platforms used to

Table 2: Examples of Analytical Data Elements to Align Across Clinical Trials

Assay Data Element Type	Data Element	Description
Approach	Assay Approach	Tumor informed, plasma only; Personalized or non-personalized
	Genomic Features Assessed	Sequence mutations, structural alterations, methylation, fragmentation, etc.
	Assay Performance Metrics	Reference range/interval, LOB, LOD, accuracy, repeatability, reproducibility
	Platform for ctDNA Assessment	NGS, ddPCR, etc.
	ctDNA Positive Definition	Threshold for calling samples positive
	ctDNA Detection and Quantification Approach	Requirement for ctDNA+ result, how ctDNA levels are calculated, and unit of measurement (e.g., mean tumor molecules (MTM)/ml plasma, and/or Variant Allele Fraction (VAF))
	Non-Genomics Features Assessed	Protein biomarkers, lipid biomarkers, etc.
Measurements	cfDNA Assay Input and Method of Measurement	
	ctDNA Level at Baseline	
	ctDNA Level at Detection	
	ctDNA Fraction at Detection	VAF
	Plasma Volume	
	Per Sample LOD	Estimated LOD for an individual sample

measure ctDNA levels, an important first step will be to align on key definitions and metrics for measuring ctDNA to begin to address the technical questions. Studies of ctDNA changes should include sufficient detail regarding the specific approach and measurements of the assay **(Table 2)**, such that the data can be optimally understood and appropriately analyzed across multiple studies to answer the key technical questions.

Clinical Considerations

To generate large datasets with robust clinical outcomes and ctDNA data, it is important to align on a core set of data elements that should be captured in randomized controlled clinical trials to allow for better data contextualization and to optimally assess the use of ctDNA as an early endpoint. In the previous ctMoniTR efforts, challenges arose when harmonizing the data across data sources to answer key clinical questions due to the fact that this was a retrospective analysis and key data elements varied, making it

Table 3: Examples of Clinical Data Elements to Align Across Clinical Trials	
Data Elements	**Considerations for Alignment**
Baseline Disease Characteristics and History	Timing of adjuvant therapy after definitive therapy (adjuvant) History and timing of prior therapy (neoadjuvant and adjuvant)
Timing and Frequency of Plasma Collection	Time of day sampling
	Timing of collection relative to definitive therapy (adjuvant)
	Timing of collection with therapy (e.g., cycle of administration)
	Timing of collection at baseline measurement
Frequency of Radiographic Tumor Surveillance and Imaging Modality	Frequency of imaging
	Timing of imaging with plasma collection

challenging to answer some clinical questions of interest. For example, in order to evaluate how early changes in ctDNA can predict response, there must be appropriate timepoint measurements of ctDNA levels at baseline and prior to the first imaging assessment. However, these timepoints were not routinely collected in all clinical trials, making it challenging to effectively answer the clinical question of how early changes in ctDNA levels can predict response. Therefore, pre-specifying the necessary elements to embed in clinical trial protocols can help maximize the types of clinical questions that can be answered and prevent later analysis issues due to discordant clinical trial methodology.

For each of these core data elements there are multiple considerations, each with the opportunity to align on a standard methodology for the data elements to maximize learnings later. For example, there is varying methodology for the timing and frequency of plasma collection, and further work is needed to define the frequency so that analysis of ctDNA changes over time can be more effectively analyzed across datasets. The frequency of collection may depend on the therapeutic class, cycle of administration, treatment setting (neoadjuvant or adjuvant), or tumor type. These clinical variables will also inform the clinical endpoints that are measured in a clinical trial (e.g., DFS, EFS, OS, and pCR), and will therefore affect the conclusions that can be drawn regarding the use of ctDNA as an early endpoint. Future work is needed to align on these core data components and set forth recommendations to be followed in future clinical trials to allow for effective assessment of ctDNA.

Conclusions

There is great opportunity in early-stage solid tumors to assess the use of ctDNA changes as a potential early endpoint to predict long-term patient outcomes for regulatory approval. The focus on establishing ctDNA as an early endpoint has the potential to expedite and improve confidence in the efficacy of novel therapies, bringing beneficial treatments to patients sooner. To rigorously evaluate the use of ctDNA as an early endpoint for regulatory decision-making, aggregating data across studies will be necessary. There are many critical clinical and technical questions to address to establish ctDNA as an early endpoint, and alignment on key data elements in clinical studies will help to accelerate the answers to these questions. As the group continues to develop a roadmap for assessing the use of ctDNA as an early endpoint in early-stage disease, ongoing conversations and collaboration between stakeholders is crucial.

As a first step, a landscape assessment of the current data available from previously conducted randomized controlled trials in early-stage disease is needed. The group aims to establish an inventory of data availability, categorizing the data available by clinical variables such as tumor type, treatment setting, and therapeutic class. An analysis will need to be conducted to understand the methodology for obtaining the core clinical data elements, such as the frequency of plasma collection in the studies. This insight into current practice, with an understanding and justification of the clinical context supporting the practice, will inform future recommendations for data capture in clinical studies. Prospectively designed studies, following the specified recommendations for data capture, will then lessen the variability seen in retrospective datasets allowing for more effective analysis.

There are additional significant technical and analytical questions about the assays that measure ctDNA to be addressed. Activities must be coordinated with other relevant stakeholders, as efforts to set pre-analytical and analytical standards for assays measuring ctDNA will be important. Further, the technical and statistical considerations for effectively conducting pooled meta-analyses from multiple trials, given the variability, will need to be discussed to determine the optimal statistical approaches and potential limitations of meta-analyses.

As demonstrated by the previous ctMoniTR work in late-stage disease, collaboration across sponsors for data analysis from multiple clinical trials is possible. Collaboration will be necessary to generate large datasets with robust and aligned clinical data to evaluate the use of ctDNA as an early endpoint in early-stage disease. Initial considerations are presented in this white paper, and work will continue to build a roadmap for assessment of this early endpoint that has the potential to transform drug development and benefit patients.

Abbreviations

cfDNA – Cell Free DNA

cfRNA – Cell Free RNA

ctDNA – Circulating Tumor DNA

ctMoniTR – ctDNA to Monitor Treatment Response (*Friends'* collaboration)

ddPCR – Digital Droplet Polymerase Chain Reaction

DFS – Disease-Free Survival

EFS – Event-Free Survival

IO – Immuno-Oncology

LOB – Limit of Blank

LOD – Limit of Detection

MRD – Minimal (or Molecular) Residual Disease

MTM - Mean Tumor Molecules

NSCLC – Non-Small Cell Lung Cancer

NGS – Next Generation Sequencing

OS – Overall Survival

pCR – Pathological Complete Response

VAF – Variant Allele Fraction

References

1. Chin R-I, Chen K, Usmani A et al. Detection of Solid Tumor Molecular Residual Disease (MRD) Tumor DNA (ctDNA). Mol Diagn Ther 2019;23:311.

2. Chakrabarti S, Xie H, Urrutia R et al. The Promise of Circulating Tumor DNA (ctDNA) in the Management of Early-Stage Colon Cancer: A Critical Review. Cancers (Basel) 2020;12:1–18.

3. Siravegna G, Mussolin B, Venesio T et al. How liquid biopsies can change clinical practice in oncology. Ann Oncol 2019;30:1580–1590.

4. ctMoniTR Step 1 Results: Do Changes in ctDNA reflect response to treatment? Available at https://friendsofcancerresearch.org/blog/engaging-innovation/ctmonitr-step-1-results-do-changes-ctdna-reflect-response-treatment. Accessed September 15, 2021.

5. Friends of Cancer Research Launches Step 2 of the ctMoniTR Project. Available at https://friendsofcancerresearch.org/news/friends-cancer-research-launches-step-2-ctmonitr-project. Accessed September 15, 2021.

6. Powles T, Assaf ZJ, Davarpanah N et al. ctDNA guiding adjuvant immunotherapy in urothelial carcinoma. Nat 2021;595:432–437.

7. Tie J, Wang Y, Tomasetti C et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med 2016;8:346ra92.

8. Reinert T, Henriksen TV, Christensen E et al. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. JAMA Oncol 2019;5:1124–1131.

9. Tarazona N, Gimeno-Valiente F, Gambardella V et al. Targeted next-generation sequencing of circulating-tumor DNA for tracking minimal residual disease in localized colon cancer. Ann Oncol 2019;30:1804–1812.

10. Phase III Study to Determine the Efficacy of Durvalumab in Combination With Chemotherapy in Completely Resected Stage II-III Non-small Cell Lung Cancer (NSCLC) - Full Text View - ClinicalTrials.gov. Available at https://www.clinicaltrials.gov/ct2/show/NCT04385368. Accessed September 16, 2021.

11. A Study of Atezolizumab Versus Placebo as Adjuvant Therapy in Patients With High-Risk Muscle-Invasive Bladder Cancer Who Are ctDNA Positive Following Cystectomy - Full Text View - ClinicalTrials.gov. Available at https://clinicaltrials.gov/ct2/show/NCT04660344. Accessed October 1, 2021.

12. Magbanua MJM, Swigart LB, Wu H-T et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Ann Oncol 2021;32:229–239.

FRIENDS
of CANCER
RESEARCH