Regulatory Advancements for Patients

FRIENDS
of CANCER
RESEARCH

INTRODUCTION

For more than two decades, Friends of Cancer Research (*Friends*) has been instrumental in the creation and implementation of policies ensuring patients receive the best treatments in the fastest and safest way possible. *Friends* has been successful due to convening the right people at the right time and putting forth revolutionary, yet realistic ideas. Through collaborative and meaningful initiatives, *Friends*' programs foster solutions to issues encountered by researchers and regulators as they strive to translate discoveries into safe and effective new treatments.

Each year, *Friends* convenes working groups, hosts scientific conferences, and conducts research on a range of topics to inform regulatory policy, oncology drug development, and clinical practice. These venues are vital to facilitating the creative partnerships and dialogue that ultimately yield many whitepapers, scientific abstracts, and peer-reviewed manuscripts led by *Friends* that infuse innovative ideas and strategies into the collective science and regulatory landscape. In 2019, we expanded our research portfolio with several large scale pilot projects that are represented in this book. These pilot projects are designed to expand our science understanding as well as inform policy.

The 2019 Scientific Report represents *Friends*' ongoing mission to drive collaboration among partners from every healthcare sector to power advances in science, policy, and regulation that speed life-saving treatments to patients. This journal is intended to be a resource for those in the drug development and regulatory space and informative for those interested in science and regulatory issues in oncology. The scientific report contains the full text of the *Friends* 2019 publications and whitepapers, which focused on several key themes:

❶ **Real-world evidence:** Characterizing the use and power of real-world data
❷ **Patient-focused drug development:** Aligning patient needs with oncology drug development
❸ **Complex biomarkers:** Informing development and standards for diagnostic tests
❹ **Optimal drug development:** Addressing emerging opportunities

> Only **35%** of cancer trials include sicker, higher risk patients. Over **60%** of cancers occur in people 65 and older but only **22%** of clinical trial patients are over 65.

> Specific patient populations and **characteristics** can be collected from several RWD sources in a similar manner to produce similar results— this will **increase** confidence in results from studies with real-world data.

REAL-WORLD EVIDENCE: CHARACTERIZING THE USE AND POWER OF REAL-WORLD DATA

Real-world evidence (RWE) is generated from data collected from patients receiving routine care. This information is recorded in the patient's medical record, medical billing and claims documents, and patient registries. Although the type of RWE gathered varies depending on the source, this data can help researchers gather insights into patient characteristics, treatment patterns, and outcomes of patients treated outside of clinical trials.

Randomized controlled trials (RCTs) are the gold standard for demonstrating evidence of safety and efficacy in the development of new therapies. While this will always be the case, there may be opportunities for real-world data (RWD) to provide important supplemental information to inform the use of therapies. For example, clinical trials use eligibility criteria to select for specific patient populations that help maximize the ability to measure treatment efficacy. While this patient selection can improve the statistical quality of the clinical data, it may not reflect the broader patient population that will likely receive the drug in clinical practice. The use of RWD can help assess the use and effectiveness of a therapy in "real-world" patients.

Increasingly, the healthcare community is grasping the capabilities of RWE and its potential role in facilitating drug development. Congress has also recognized this potential and has included mandates in the 21st Century Cures Act and the Prescription Drug User Fee of 2017 for FDA to explore the use of RWE in prescription drug regulation. The FDA's *Framework for Evaluating RWD/RWE for Use in Regulatory Decisions* will "evaluate the potential use of RWD to generate RWE of product effectiveness to help support approval of new indications for drugs… or to help to support or satisfy post approval study requirements."

In response to this growing need, the *Friends* RWE Pilots 1.0 and 2.0 were first of their kind efforts that brought together a dozen key stakeholders to study RWE for use in drug development and evalutation over time. The collective results of this work are helping inform the use of RWD/RWE.

- Specific patient populations and characteristics can be collected from several RWD sources in a similar manner to produce similar results—this will increase confidence in results from studies with real-world data
- Some real-world endpoints correlate to endpoints commonly used in clinical trials—this suggests that real-world endpoints could be used as proxies for clinical trial endpoints in real-world data and warrants further investigation

PATIENT-FOCUSED DRUG DEVELOPMENT: ALIGNING PATIENT NEEDS WITH ONCOLOGY DRUG DEVELOPMENT

Patient engagement in research and clinical trials has evolved over time. Patients are being actively sought as partners to help design, implement, and disseminate clinical trial findings. It is estimated that less than 5% of adult cancer patients enroll in a clinical trial despite many indicating a desire to participate. Engaging patients early and often throughout the entire research and drug development process can help inform appropriate trial designs, answer patient relevant questions, and encourage participation.

Designing clinical trials that answer questions important to patients and maximize the information gained are critical. Listening to patients and capturing their experience can help better characterize the patient's health, quality of life, and functional status while on a cancer treatment. Patient reported outcome (PRO) tools are helping better define the value of a treatment and enabling more informed patient decision-making.

Throughout 2019, *Friends* helped advance the understanding of PRO use in treatment decision making as well as identify key recommendations to improve safety monitoring in trials.

RWE PARTNERS

Aetion
ASCO CancerLinQ & Concerto
HealthAI
Cancer Research Network
COTA
FDA
Flatiron Health
IQVIA™
Mayo Clinic
McKesson
NCI SEER-Medicare Linked Database
OptumLabs®
PCORnet
Syapse
Tempus

Some real-world endpoints correlate to **endpoints** commonly used in clinical trials— but are difficult to measure. This suggests that real-world endpoints could be used as **proxies** for clinical endpoints in real-world data.

> Improved character-ization of treatment toxicities and how **patients feel** and function while on treatment can ena-ble a more thorough assessment of a therapy's **benefit** and can help prioritize future clinical trials

- Improved characterization of treatment toxicities and how patients feel and function while on treatment can enable a more thorough assessment of a therapy's benefit and can help prioritize future clinical trials
- Mechanisms for more rapid sharing of adverse events information during trials and improved consistency in reporting can help more precisely describe potential treatment-related adverse events
- Use of PRO data by physicians in decision-making is variable; however, enhancing the quality and availability of data can improve its usefulness

PUBLICATIONS RELATED TO THIS TOPIC

- *The Promise of Immuno-oncology: Implications for Defining the Value of Cancer Treatment (Page 45)*
- *How Oncologists Perceive the Availability and Quality of Information Generated from Patient-Reported Outcomes (PROs) (Page 56)*
- *Improving Attribution of Adverse Events in Oncology Trials (Page 64)*

> Mechanisms for more **rapid sharing** of adverse events information during trials and **improved consistency** in reporting can help more precisely describe potential treatment-related **adverse events**

COMPLEX BIOMARKERS: INFORMING DEVELOPMENT AND STANDARDS FOR DIAGNOSTIC TESTS

Targeted therapies, which are drugs that target specific molecular pathways, and their associated diagnostic tests allow physicians and researchers to identify the patients most likely to respond to a specific treatment. When patients are matched to the right drug at the right time, they may experience an improved outcome since these targeted therapies can provide substantial improvement over currently available treatment. Diagnostic tests are used to identify specific mutations or biomarkers that can then identify the therapy to which a patient is most likely to respond.

Characterization of genetic signatures, such as tumor mutational burden (TMB), are emerging as important tools in treatment decision-making. A snapshot of clinicaltrials.gov demonstrates that the number of clinical trials incorporating TMB is increasing rapidly and have a total patient accrual goal of more than 20,000 patients **(Figure 1)**. To maximize the utility and benefit of these emerging complex biomarkers, it is important that there

is consistency and accuracy across tests to optimally inform treatment decisions.

Friends convened a consortium of key stakeholders, including diagnostic manufacturers, academics, pharmaceutical companies, the National Cancer Institute (NCI), and the FDA, to recommend best practices and approaches for TMB measurement, validation, alignment, and reporting well ahead of the adoption of this powerful biomarker for clinical decision-making.

Key findings and recommendations from this work help provide important development and regulatory considerations for complex biomarkers.

- Preliminary analyses from the TMB harmonization effort highlight the importance of assay characteristics and bioinformatic pipeline for reliable TMB estimation
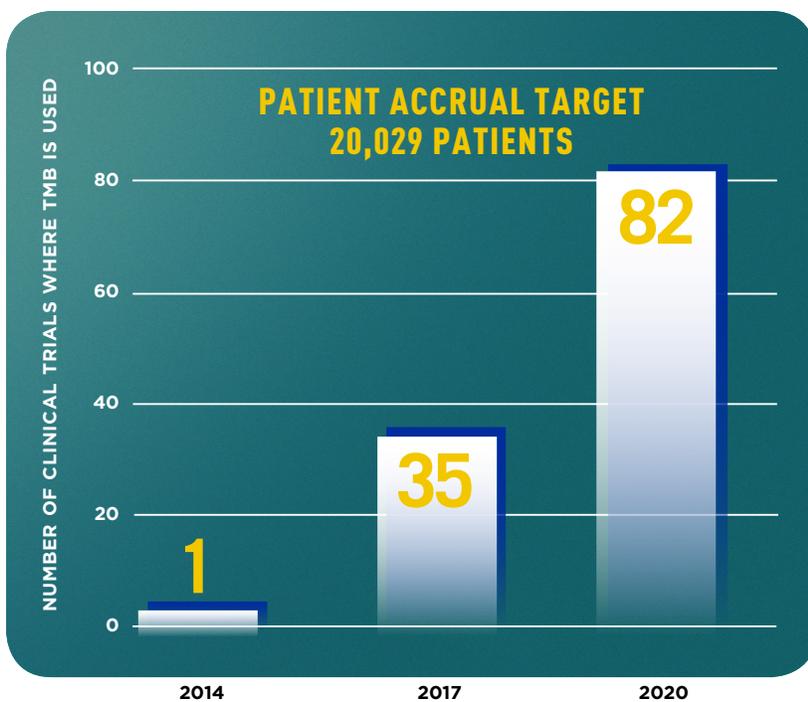
FIGURE 1: Number of clinical trials where TMB is used during the past 6 years according to clinicaltrials.gov

A framework for **evidentiary standards** could help establish confidence in the safe and effective use of **diagnostic tests** and inform test and drug labels

> **Preliminary analyses from the TMB harmonization effort highlight the importance of assay characteristics and bioinformatic pipeline for reliable TMB estimation**

- In addition, variation due to technical aspects of diagnostic tests and biological variation can also play an important role in determining TMB
- A framework for evidentiary standards could help establish confidence in the safe and effective use of diagnostic tests and inform test and drug labels

PUBLICATIONS RELATED TO THIS TOPIC

- *Tumor Mutational Burden Standardization Initiatives: Recommendations for Consistent Tumor Mutational Burden Assessment in Clinical Samples to Guide Immunotherapy Treatment Decisions (Page 73)*
- *Spatial and Temporal Heterogeneity of Panel–Based Tumor Mutational Burden in Pulmonary Adenocarcinoma: Separating Biology from Technical Artifacts (Page 84)*
- *TMB standardization by alignment to reference standards: Phase II of the Friends of Cancer Research TMB Harmonization Project. (Page 97)*
- *Data Generation (and Review Considerations) for Use of a Companion Diagnostic for a Group of Oncology Therapeutic Products (Page 99)*

OPTIMAL DRUG DEVELOPMENT: ADDRESSING EMERGING OPPORTUNITIES

During the last two decades, the field of oncology has undergone rapid change as novel mechanisms for treating cancer have been discovered, promising new therapies have emerged, and innovations in regulatory science have been made. Although these advances present tremendous opportunities for the field, they are also accompanied by new challenges related to the optimization of drug development processes. In oncology, there are clinical settings and scenarios where randomization may be difficult or not feasible (e.g., rare disease, small patient population, loss of equipoise, availability of the investigational agent outside of the clinical trial), which requires innovative approaches for studying drugs in these situations.

> **Use of PRO data by physicians in decision-making is variable; however, enhancing the quality and availability of data can improve its usefulness**

Advancements in cancer immunology and recent clinical experience with emerging cellular therapies, such as CAR-T, are generating huge interest and activity both academically and industrially. Additionally, combination therapies are continuing to demonstrate benefit for some patients and the number of trials

using immunotherapy drugs is growing. These emerging new therapies and combinations have the potential to rapidly change cancer treatment, positively impacting patients. New science- and risk-based approaches to optimize development may need to be considered.

- Regulatory strategies and adaptive manufacturing processes could be optimized to expedite CAR-T therapies into early phase studies and ensure that CAR-T are impactful for the greatest number of patients
- Disease areas with unmet medical needs (e.g., rare cancers, specific cancer subtypes) often represent clinical settings and scenarios that are difficult to study and randomization in clinical trials may be difficult or not feasible. Trial designs that use external data (e.g., RWD, other clinical trial data) could help expedite development in these challenging areas
- As the number of combination therapies and codeveloped new investigational drugs increases in the era of immunother-apies, rational combination development and thoughtful trial designs are needed

> In addition, variation due to technical aspects of **diagnostic tests** and **biological variation** can also play an important role in determining TMB

PUBLICATIONS RELATED TO THIS TOPIC

- *Immuno-Oncology Combination Drug Development for Patients with Disease Progression After Initial Anti-PD-(L)1 Therapy (Page 117)*
- *Characterizing the Use of External Controls for Augmenting Randomized Control Arms and Confirming Benefit. (Page 133)*
- *Non-small cell lung cancer (NSCLC) case study examining whether results in a randomized control arm are replicated by a synthetic control arm. (SCA) (Page 170)*
- *Opportunities for Combination Drug Development: Data Sources and Innovative Strategies to Assess Contribution of Components (Page 171)*
- *Designing the Future of Cell Therapies (Page 186)*

CONCLUSION

We thank the numerous contributors, partners, and collaborators who have contributed their time, expertise, and data. We look forward to continued collaborations in 2020.

An Exploratory Analysis of Real-World End Points for Assessing Outcomes Among Immunotherapy-Treated Patients With Advanced Non–Small-Cell Lung Cancer

Mark Stewart, PhD[1]; Andrew D. Norden, MD, MPH, MBA[2]; Nancy Dreyer, MPH, PhD[3]; Henry Joe Henk, PhD[4]; Amy P. Abernethy, MD, PhD[5]; Elizabeth Chrischilles, PhD[6,7]; Lawrence Kushi, ScD[8,9]; Aaron S. Mansfield, MD[10]; Sean Khozin, MD, MPH[11]; Elad Sharon, MD, MPH[12]; Srikesh Arunajadai, PhD[2]; Ryan Carnahan, PharmD, MS[6,7]; Jennifer B. Christian, PharmD, MPH, PhD[3]; Rebecca A. Miksad, MD, MPH[5]; Lori C. Sakoda, PhD, MPH[8]; Aracelis Z. Torres, MPH, PhD[5]; Emily Valice, MPH[8]; and Jeff Allen, PhD[1]

abstract

PURPOSE This pilot study examined the ability to operationalize the collection of real-world data to explore the potential use of real-world end points extracted from data from diverse health care data organizations and to assess how these relate to similar end points in clinical trials for immunotherapy-treated advanced non–small-cell lung cancer.

PATIENTS AND METHODS Researchers from six organizations followed a common protocol using data from administrative claims and electronic health records to assess real-world end points, including overall survival (rwOS), time to next treatment, time to treatment discontinuation (rwTTD), time to progression, and progression-free survival, among patients with advanced non–small-cell lung cancer treated with programmed death 1/programmed death-ligand 1 inhibitors in real-world settings. Data sets included from 269 to 6,924 patients who were treated between January 2011 and October 2017. Results from contributors were anonymized.

RESULTS Correlations between real-world intermediate end points (rwTTD and time to next treatment) and rwOS were moderate to high (range, 0.6 to 0.9). rwTTD was the most consistent end points as treatment detail was available in all data sets. rwOS at 1 year post–programmed death-ligand 1 initiation ranged from 40% to 57%. In addition, rwOS as assessed via electronic health records and claims data fell within the range of median OS values observed in relevant clinical trials. Data sources had been used extensively for research with ongoing data curation to assure accuracy and practical completeness before the initiation of this research.

CONCLUSION These findings demonstrate that real-world end points are generally consistent with each other and with outcomes observed in randomized clinical trials, which substantiates the potential validity of real-world data to support regulatory and payer decision making. Differences observed likely reflect true differences between real-world and protocol-driven practices.

INTRODUCTION

Randomized clinical trials (RCTs) are the optimal method by which to demonstrate causal effects between treatments and outcomes, but are often slow to accrue and expensive[1] or are difficult to conduct because of practical or ethical reasons.[2] Moreover, their results may not generalize to patients who are treated in the real-world setting.[3] The unprecedented availability of real-world data (RWD), emergence of new RWD sources, improved analytic methods, and the accelerating need for clinical evidence in the face of constrained RCT resources has increased the demand for real-world evidence (RWE). Study of routinely collected health care data is increasingly important for

various stakeholders who are interested in better understanding particular patient populations, evaluating drug safety in the postmarketing setting, measuring health care use and clinical outcomes, performing comparative effectiveness research, and optimizing drug pricing models.[4] However, before RWD finds widespread use as an adjunct to—or in unique settings, an alternative for—RCTs, the validity of readily extractable clinical outcomes measures—real-world end points—must be established. A fundamental step is to characterize and contrast the patient populations and methods used for aggregation and curation of RWD across various sources to understand the natural variability of key parameters in real-world

settings and the extent to which they differ from that observed under highly controlled settings.[5]

The US Congress and the US Food and Drug Administration (FDA) recognize the importance of further developing the use of RWD for regulatory decision making as evidenced by recent publications by the FDA[6-8] and passage of the 21st Century Cures Act (Cures Act),[9] and the Prescription Drug User Fee Act[10] VI reauthorization.[7] The Cures Act, passed in December 2016, requires the FDA to develop a framework for and issue guidance on the use of RWE for a new indication for an already-approved drug or for postmarket study as a requirement for regulatory approval. In addition, RWD comparator or benchmark data have been used in recent approvals of new cancer treatments on the basis of phase II trials.[11,12]

Academia, public and private companies, health policy organizations, and the FDA are working to establish best practices for the generation and evaluation of RWD in regulatory settings.[8,13] To support these efforts, Friends of Cancer Research convened six organizations with oncology-focused health care data to conduct a pilot RWD project. The primary collective goals of the study were to agree on and execute a common protocol using diverse RWD and to explore how real-world end points could be used to rapidly address clinically relevant questions about treatment effectiveness.

A framework was established for data collection, end point definitions, and planned analyses, with flexibility incorporated to allow for differences in data elements across multiple RWD sources. The project examined patients with advanced non–small-cell lung cancer (aNSCLC) who were treated with programmed death 1/programmed death-ligand 1 [PD-(L)1] inhibitors in the real-world setting. Initial results and potential implications were presented publicly at The Future Use of Real-World Evidence meeting hosted by Friends of Cancer Research in Washington, DC, on July 10, 2018.

PATIENTS AND METHODS

Study Design and Objectives

Data sets generated for this study included relevant and accessible patient-level RWD for eligible individuals. The project had three key objectives:

1. Identify, describe, and compare the demographic and clinical characteristics of eligible patients in each data source.
2. Assess the ability to operationalize a common protocol and generate real-world (rw) end points [overall survival (rwOS), progression-free survival (rwPFS), time to progression (rwTTP), time to next treatment (rwTTNT), and time to treatment discontinuation (rwTTD)].
3. Assess how rwOS compares with clinical end points as measured in RCTs.

General approaches to identifying analytic populations, defining specific variables, and conducting analyses were discussed and agreed on by all participating organizations or networks. Given the variability of the types of data and data sources available, these general approaches were tailored to each of the six contexts, as data sources differed, and may have included data from health claims, electronic health records (EHRs), data that had been extracted from text or other unstructured fields in medical charts, or some combination of these data sources. We conducted and completed database analyses within approximately 3 months from the completion of the broad study protocol.

Study Populations and Inclusion and Exclusion Criteria

Eligible patients included those who were diagnosed with aNSCLC on or after January 1, 2011. Patients were identified as having aNSCLC if they were diagnosed initially with American Joint Committee on Cancer stage IIIB or IV NSCLC, or with early-stage NSCLC with evidence of recurrence or progression described or documented in available data. Treatment with a PD-(L)1 inhibitor was identified from each organization's data sources, which may have included a medication order, a claim, or infusion databases in EHRs. To limit analyses to patients who could have been observed by health care providers who were represented in each participating organization's databases, patients had to have at least two documented clinical visits during the calendar period of interest as defined above or, alternatively, in integrated health care systems, evidence of continuous enrollment in the health plan, defined as no gap in insurance coverage greater than 90 days. For claims data sources, in which stage and progression data were not typically available, patients were included if they received treatment with a PD-(L)1 inhibitor after a diagnosis of lung cancer. During the project timeframe, insurer coverage for these agents required evidence of advanced disease as defined above. Data were sought for patients with lung cancer who were diagnosed as early as January 2011 and who initiated treatment with a PD-(L)1 inhibitor between January 2014 and October 2017, which allowed for at least 6 months of potential follow-up and identification of prior lines of therapy. End of follow-up varied by participating organization on the basis of the most recent date of reasonably complete information on outcomes of interest, with some data sets having documentation of outcomes as recent as April 30, 2018.

Patients were excluded if they had a date of diagnosis more than 90 days before the first activity date—visit or treatment administration—on the assumption that this reflected missing data on historical treatment.

Participating Organizations and Data Sources

Data partners represent a range of care models in the United States, from community oncology centers, health systems, academic medical centers, and integrated delivery system networks to mixtures of these care settings. Data curation included different approaches that were unique to each participant, including natural language

processing, artificial intelligence tools, and technology-enabled abstraction and general chart review. Key characteristics of the data sources are listed in Table 1.

All data partners in this pilot project have been using their data extensively for research over many years. Consequently, each has used a variety of curation processes designed to evaluate the quality and completeness and to strengthen data management processes to assure reliable data integration and transformations, as needed for research conduct. Thus, these research-ready networks are not typical of EHR data in general, nor of health insurance claims data that have not been subjected to such ongoing data curation.

End Point Definitions

Each data provider used the agreed upon definitions to calculate end points (Table 2). Study treatment refers to treatment with a PD-(L)1 inhibitor, defined here as treatment with nivolumab, pembrolizumab, or atezolizumab.

Statistical Analysis

Each data provider analyzed their own data, as is common in many federated research networks and, hence, there may have been specific nuances to each data source that are not captured directly in the above definitions. Results were shared with Friends of Cancer Research, who, as a neutral third party, summarized the findings in an anonymous fashion. Patient, tumor, and treatment characteristics were analyzed using descriptive statistics. Continuous variables were summarized using medians and interquartile ranges. Categorical variables were calculated as frequencies. We used Kaplan-Meier methods to estimate time-to-event end points with 95% CIs estimating median times to event. Correlations between rwOS and each time-to-event end point were calculated using Spearman's rank-order correlation. Correlation analysis was restricted to those patients who had experienced both death and the event of interest.

RESULTS

Patient Identification and Characteristics

Table 3 lists the demographic and clinical characteristics of each patient population. The six data sets included 269 to 6,924 patients with lung cancer who were diagnosed as early as January 1, 2011, and who initiated treatment with a PD-(L)1 inhibitor between January 1, 2014, and October 30, 2017. Median age at diagnosis of lung cancer ranged from 64 to 70 years. Data sets were composed of 50% to 56% male patients and a large majority of patients (65% to 87%) were white with 6% to 13% Black or African American. Source and missing data of information on race/ethnicity varied significantly by data set. In data set C, 19% of patients were identified as Asian, otherwise, Asians represented 1% to 5% of the cohort. Median household income information was available for only two of the six data sets. Information on tobacco use was available in four datasets and, as expected, most patients (78% to 92%) were documented as having a history of tobacco smoking.

In the four data sets with stage at initial diagnosis, 69% to 100% of patients were diagnosed with stage III or IV disease. [In addition, for data set A, evidence of advanced disease, defined as either stage IIIB or IV NSCLC at initial diagnosis or early-stage (I, II, and IIIA) NSCLC with a recurrence or progression is required by the health plan for coverage of a PD-(L)1 during this study period.] Tumor histology was available in five data sets, among which 66% to 74% of patients had non–squamous-cell carcinoma and 17% to 30% had squamous-cell carcinoma. In some data sets, PD-L1 expression testing was available in a subset of patients. Where results were available from *ALK* or *EGFR* testing, few patients had *ALK* translocations or *EGFR* mutations. In the largest proportion of cases (32% to 56%), PD-(L)1 inhibitor therapy represented the second line of treatment in the advanced disease setting.

Line of therapy information was derived in five of six data sets. During the study period, most patients received a PD-(L)1 inhibitor after first-line treatment. Few patients received a second PD-(L)1 inhibitor. Overall, median follow-up time from the initiation of PD-(L)1 inhibitor therapy was 6 to 9 months, with an interquartile range of 9 to 13 months.

Patient Identification and Characteristics: What We Learned

Whereas the depth of information varied by data source, with EHR and cancer registry data providing access to richer clinical information, all data providers were able to expeditiously identify a cohort of patients with aNSCLC who received PD-(L)1 and observe them for at least 6 months. In addition, the option of identifying a patient's diagnosis date or simply identifying patients at the time of first PD-(L)1 inhibitor treatment provides the flexibility to study all treatment regimens or only immunotherapy.

Real-World End Points

Table 4 lists median times for real-world end points in each data set. The range of median rwOS times was 8.58 months to 13.50 months, presumably driven in large part by the sources of death information, which varied from the confirmation of death being reported in the EHR to linkage with the Social Security Administration (SSA) Death Master File. Data set B demonstrates the variability that can exist in determining death and the potential effect on survival end points by calculating rwOS in all sites and when only including sites with SSA or state death data available. rwTTD was the most consistent end point as treatment detail was available in all data sets. Excluding data set A, which seems to be an outlier, median rwTTD for the remaining data sets ranged between 3.2 months and 4.7 months. Similarly, rwTTNT was consistent, ranging from 11.6 months to 14.0 months. rwTTP and rwPFS were calculated with data sets D and F in which information was extracted from the text or other unstructured fields in the EHR. Other data providers used primarily structured data from claims or EHRs for this analysis, which precluded capture of these end points.

Stewart et al

TABLE 1. Characteristics of Participating Data Sources

Variable	Cancer Research Network	Cota Healthcare	Flatiron Health	IQVIA	OptumLabs Data Warehouse	PCORnet
Data Source	Virtual data warehouse*	De-identified, longitudinal data source comprised of abstracted patient-level EHR data from contributing provider sites, including academic medical centers and community practices and hospital systems	De-identified patient-level clinical data from OncoEMR, Flatiron's oncology-specific EHR, and integrations with academic EHRs (eg, Epic); data include a complete copy of the medical record as well as patient-level linkages to other data sets to fill gaps (eg, mortality)	Diversified oncology EHRs, including EHRs from TransMed, that are de-identified at the patient level	Medical claims data and enrollment information for commercial and Medicare Advantage enrollees in a large US health plan	Pooled data set from 11 medical centers participating in a PCORnet Rapid Cycle Project—PCORnet Common Data Model linked with tumor registries
Setting	Community-based health care systems	Predominantly community practices (≥ 90%) in this project	Predominately community oncology practices (80% of patients) and academic medical centers (20% of patients)	Predominantly community oncology practices (≥ 90%)	Paid claims from provider, which include enrollees treated in either community practice or academic settings or both	Ten academic medical centers and one multihospital health care system
Region	Across the United States	Northeast and Mid-Atlantic regions in this project	>800 sites of care at >280 clinics across the United States	Sources from sites across the United States	Geographically diverse across the United States	Eleven states: Great Plains/Midwest, Mid-South, and Florida
Single source or linked	Linked	Multisourced EHR data linked for supplementation (eg, mortality data sources)	Linked to resolve data gaps (eg, SSDI and commercial death data)	EHRs from several sources, including multiple EHR software systems and networks integrated at a patient level	Linked data; for this project, race and SES information and death information from the SSA DMF are linked to health plan enrollees (NOTE. SSA DMF results are not reported)	Linked
Data processing	Structured EHR and other clinical and administrative data only	Structured and unstructured EHR data collected through Cota's software platform and subject to a rigorous quality control process	Structured EHR data and unstructured EHR and document data curated by clinical experts using a single software interface and PHI controls	Structured EHR elements included; custom abstractions of unstructured elements were excluded for this analysis	Structured files included in this analysis	Structured EHR and other clinical and administrative data only

Abbreviations: DMF, Death Master File; EHR, electronic health record; PHI, protected health information; SES, socioeconomic status; SSA, Social Security Administration; SSDI, Social Security Disability.

*Common data model used by participating organizations to systematically structure and aggregate data from medical records dating back to 1996 or earlier on multiple domains, including but not limited to, health plan enrollment periods, cancer registries, mortality and cause-of-death files, encounters including associated diagnoses and procedures, prescribed oral and infusion medications, laboratory results, and demographics.[14]

© 2019 by American Society of Clinical Oncology

TABLE 2. Common Definitions Used in the Pilot Project

Term	Definition
End point	
Real-world overall survival	Length of time from the date the patient initiates treatment with a PD-(L)1 inhibitor to the date of death or end of follow-up, whichever occurred earliest, and for claims data, health plan disenrollment, if deaths are not captured among those who leave health plan coverage
Real-world time to next treatment	Length of time from the date the patient initiates study treatment to the date the patient initiates his or her next systemic treatment. When subsequent treatment is not received (eg, continuing current treatment or unenrollment not because of confirmed death), patients were censored at their last known activity
Real-world time to treatment discontinuation	Length of time from the date the patient initiates treatment with a PD-(L)1 inhibitor to the date 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
Real-world progression event	Distinct episode in which the treating clinician concludes that there has been growth or worsening in the cancer, as determined by review of the patient chart. As this is typically determined by review of the patient chart and progression events are not documented in structured fields, this was readily available only for participating organizations in which chart review was performed
Real-world progression-free survival	Length of time from the date the patient initiates treatment with a PD-(L)1 inhibitor to the date of a real-world progression event or death, at least 14 days after study treatment initiation. Patients without a real-world progression event or date of death were censored at the most recent visit with the treating oncologist or end of follow-up
Real-world time to progression	Length of time from the date the patient initiates the study treatment to the date that a real-world progression event is documented in the patient's EHR, at least 14 days after study treatment initiation. Death is excluded as an event. Patients without a real-world progression event were censored as in real-world progression-free survival above
Other elements	
Structured follow-up time	Length of time from the date 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. 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. Subsequent LOTs were defined similarly

Abbreviations: aNSCLC, advanced non–small-cell lung cancer; EHR, electronic health record; LOT, line of therapy; PD-(L)1, programmed death 1/programmed death-ligand 1.

rwOS proportions were calculated for each data set at 12 months (Table 4). The proportion of patients who were alive at 12 months after initiation of PD-(L)1 therapy ranged from 40% to 57%. rwTTD and rwOS median times and 95% CIs segmented by treatment setting and demographic characteristics were also calculated (Table 5). This further illustrates the ability of RWD to assess treatment effectiveness in patient populations that may not be routinely

TABLE 3. Description of Demographic and Clinical Characteristics of Patients With Advanced Non–Small-Cell Lung Cancer Treated With PD-(L)1 Checkpoint Inhibitors

Demographic	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)
Median age at advanced diagnosis, years (IQR)	68 (15)	64 (14)	66 (14)	69 (14)	68 (14)	70 (14)
Median age at PD-(L)1 inhibitor initiation, years (IQR)	69 (14)	65 (14)	68 (14)	69 (14)	69 (14)	71 (14)
Age categories at PD-(L)1 inhibitor initiation (categorical), years, No. (%)						
≤ 49	120 (5)	24 (4)	21 (5)	219 (3)	80 (3)	8 (3)
50-64	888 (34)	252 (45)	129 (30)	2,048 (30)	863 (30)	65 (24)
65-74	866 (33)	194 (35)	169 (39)	2,504 (36)	1,047 (37)	94 (35)
≥ 75	721 (28)	86 (15)	116 (27)	2,153 (31)	870 (30)	102 (38)
Age categories at PD-(L)1 inhibitor initiation (binary), years, No. (%)						
< 75	1,874 (72)	470 (85)	319 (73)	4,771 (69)	1,990 (70)	167 (62)
≥ 75	721 (28)	86 (15)	116 (27)	2,153 (31)	870 (30)	102 (38)
Sex, No. (%)						
Female	1,147 (44)	275 (49)	212 (49)	3,172 (46)	1,351 (47)	125 (46)
Male	1,448 (56)	281 (51)	222 (51)	3,752 (54)	1,509 (53)	143 (53)
Unknown/missing	0	0	5	0	0	1
Race/ethnicity, No. (%)						
White	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	52 (2)	6 (1)	83 (19)	155 (3)	13 (2)	9 (5)
Other	142 (7)	6 (1)	31 (7)	580 (9)	42 (5)	1 (1)
Unknown/missing	415	0	0	626	2,085	85
Group stage at initial diagnosis, No. (%)						
0/occult		0		2 (0)		
I		23 (6)		496 (7)		18 (7)
II		22 (6)		426 (6)		17 (7)
III		88 (23)	39 (9)	1,494 (22)		17 (7)
IV		248 (65)	396 (91)	4,335 (64)		161 (62)
Group stage not reported		175		171		10
Histology, No. (%)						
Non–squamous-cell carcinoma		369 (66)	320 (74)	4,679 (70)	1,981 (69)	194 (73)
Squamous-cell carcinoma		147 (26)	73 (17)	1,983 (30)	659 (23)	61 (23)
NSCLC histology, not otherwise specified		40 (7)	42 (10)	262 (3)	220 (8)	10 (4)
Missing						4
Smoking status, No. (%)						
History of smoking			340 (78)	6,185 (90)	448 (92)	182 (87)
No history of smoking			94 (22)	717 (10)	38 (8)	28 (13)
Unknown/not documented			5	22	2,374	210
PD-L1 tested on or before PD-(L)1 inhibitor start	326 (13)			2,384 (34)	96	80/96 (83)
PD-L1 expression status (among those tested), No. (%)						
PD-L1 positive				512 (22)	45 (50)	65 (68)
PD-L1 negative/not detected				691 (29)	45 (50)	29 (30)
Unsuccessful/indeterminate test				1,012 (42)	0	2 (2)
Results pending/unknown				169 (7)	6	173

(Continued on following page)

TABLE 3. Description of Demographic and Clinical Characteristics of Patients With Advanced Non–Small-Cell Lung Cancer Treated With PD-(L)1 Checkpoint Inhibitors (Continued)

Demographic	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)
ALK tested on or before PD-(L)1 inhibitor start	258 (10)			4,513 (65)	582	143/173 (83)
ALK status (among those tested), No. (%)						
Rearrangement present				57 (1)	8 (1)	1 (1)
Rearrangement not present				4,145 (92)	570 (99)	170 (98)
Results pending/unknown				68 (2)	0	2 (1)
Unsuccessful/indeterminate test				243 (5)	4	96
EGFR tested on or before PD-(L)1 inhibitor start	543 (21)		171 (39)	4,684 (68)	953	115/142 (81)
EGFR status (among those tested), No. (%)						
Mutation positive				305 (7)	68 (11)	6/142 (4)
Mutation negative				4,161 (89)	525 (89)	135/142 (95)
Results pending/unknown				60 (1)	358	1/142 (1)
Unsuccessful/indeterminate test				158 (3)	2	127
Line of first PD-(L)1 inhibitor in advanced setting, No. (%)						
1 (no prior therapy received)	690 (27)	144 (26)	80 (18)	2,074 (30)	777 (27)	77 (29)
2	1,440 (56)	272 (49)	205 (47)	3,357 (49)	1,414 (49)	87 (32)
3	380 (15)	96 (17)	85 (20)	1,012 (15)	448 (16)	51 (19)
≥ 4	85 (3)	44 (8)	65 (15)	481 (7)	221 (8)	54 (20)
Patients receiving a second PD-(L)1 inhibitor in a subsequent line, No. (%)						
No		167 (30)	402 (92)	1,740 (25)		
No subsequent therapy received		375 (67)		4,879 (71)		
Yes	93 (4)	14 (3)	33 (8)	305 (4)	112	14
Line of second PD-(L)1 inhibitor in advanced setting, No. (%)						
2	28 (30)	1 (7)	11 (33)	99 (33)	9 (8)	5 (36)
3	45 (48)	3 (21)	10 (30)	134 (44)	51 (46)	4 (29)
≥ 4	20 (22)	10 (71)	12 (36)	72 (24)	52 (46)	5 (36)
N/A		541	402			
Median time from advanced diagnosis to first PD-(L)1 inhibitor initiation, months (Q1, Q3)		7 (3, 14)	8 (4, 15)	6 (2, 13)	8 (3, 17)	7 (2, 14)
Structured follow-up time						
Structured follow-up time from advanced diagnosis, months, median (Q1, Q3)		18 (10, 28)	18 (10, 31)	14 (8, 25)	18 (10, 30)	18 (10, 28)
Structured follow-up time from PD-(L)1 inhibitor initiation, months, median (Q1, Q3)		8 (3, 16)	9 (3, 16)	6 (2, 12)	8 (3, 14)	8 (4, 13)

NOTE. Empty data fields indicate variables that were not collected.

Abbreviations: N/A, not applicable; PD-(L)1, programmed death 1/programmed death-ligand 1; Q, quarter.

represented in clinical trials. For each of the available real-world end points, correlation with rwOS was assessed (Table 6). With few exceptions, correlation was in the range of 0.60 to 0.89.

Real-World End Points: What We Learned

All data partners were able to collect information on treatment and mortality data and rapidly assemble this information to quantify real-world treatment duration and rwOS; however, information that confirmed death, including date and cause of death, varied by data source and proved to be challenging. Assessing rwTTP and rwPFS requires extracting information from text and other unstructured fields within EHRs as a result of the specific information needed. Whereas health plans that implement prior authorization systems may collect and

retain progression information, it will be limited to those seeking a next line of treatment. Therefore, we believe EHRs will be a key data source for evaluating the impact of treatment on rwTTP and rwPFS compared with claims data.

Although clinical trials often have rigid inclusion and exclusion criteria to assess the safety and efficacy of a therapy, this study assessed real-world end points in a much broader patient population. Overall survival in five RCTs that assessed PD-(L)1 therapies in patients with aNSCLC had a median OS of 12.6 months (POP-LAR clinical trial), 13.8 months (OAK clinical trial), 12.2 months (CheckMate 057), 9.2 months (CheckMate 017), and 10.4 months or 12.7 months, depending on dosage (KEYNOTE-010).[15] Of interest, rwOS from this study falls within the range observed in these clinical trials. Additional work to understand how real-world end points relate to more traditional measures of clinical benefit used in clinical trials is needed.

Agreement on and monitoring of statistical analyses through a research project plan to assure similarity in execution is fundamental because of the important differences between data sources. Given the timeliness and real-world nature of the data, methods to assess the impact of and account for censoring are important. Some patients continue to receive PD-(L)1 inhibitor treatment for many months and others may be lost to follow-up or unenroll from the health plan. These facts must be considered when estimating real-world end points.

DISCUSSION

This pilot project represents an effort to bring together diverse providers of established RWD drawn from EHRs, cancer registries, and administrative claims sources to assess the feasibility of using RWD to address questions that are relevant to clinical development (eg, identification of unmet needs and contextualization of clinical trial results for new therapies, or expanded indication for existing therapies) and use (eg, adverse event and dosing considerations). This pilot project successfully brought together experienced data providers who created a common framework to address a singular question to assess whether real-world end points could be extracted from RWD of patients with aNSCLC who were treated with a PD-(L)1. Recognizing that these data partners were selected because of their research-ready data, the data protocol was executed by each group within approximately 3 months, using staff members who were already experienced in the databases, data management, and data curation practices.

Key findings of this preliminary validation exercise demonstrate that clinical questions can be addressed in a relatively short timeframe as a result of the ability to access contemporaneous cohorts, and RWD can produce findings that are directionally similar to those from RCTs,

particularly with regard to OS. In fact, some data sets had outcome data as recent as 3 months before the analysis readout. rwOS as assessed through EHRs and claims data fell within the median OS values observed in several PD-(L)1 clinical trials.[15] Variations in the rwOS signal are likely a result of challenges with accessing mortality data, as death would not by itself trigger an entry in most EHRs. Clinical workflows and their documentation in EHRs and claims data are not designed to routinely capture information about death, including the date or cause of death. In a recent EHR-based study using a single EHR, sensitivity of the structured mortality variable was only 66% and the publicly available SSA Death Master File was even lower at 35%.[16] To address known gaps in death data, some of the participating data providers rely on proprietary data sources that harvest published obituary data.[17] Such data linkages, which leverage the scale and breadth of multiple data types and sources, were highlighted as a critical mechanism to address missing data and create more robust data sources. In addition, a recent study helped to elucidate at what threshold does incompleteness begin to affect findings. It has been observed that the impact of missing death data on survival analyses and estimates of OS is small when mortality capture sensitivity is high (eg, approximately 90% or more)[18]; however, this was not analyzed in this study.

Directional patterns observed in the project data provide useful information about the utility of real-world end points and important information on patient populations that are often excluded from clinical trials. Moreover, recognizing that these data reflect contemporary treatment of cancer, the data offer important signals about how treatments are administered and how patients' disease responds outside of rigidly controlled clinical trials. In addition, levels of correlation to rwOS between several of the other real-world intermediate end points that were assessed ranged from 0.6 to 0.9, which indicates that real-world end points could have utility in supporting regulatory and payer decision making. Moreover, several characteristics are shared among the analyzed cohorts, despite varying sample sizes, data capture/curation processes, case identification, and data sources.

The implication of these findings is that RWD can provide useful and timely evidence to quantify the benefits and risks of new cancer treatments used in real-world settings. These results further demonstrate the utility of RWE and the need for additional investigations to assess readily extractable end points from RWD sources. Although there is a great deal of discussion about what constitutes regulatory-grade RWD, concordance between RWD and RCT data shown here demonstrates the basic principles needed to support RWD, namely that enough patients who meet the criteria of interest can be aggregated and selected without bias, and consistent follow-up data and validated end points are available for the same population.

TABLE 4. Median Time and 95% CI For Real-World Extracted End Points

Data Set	rwOS	rwTTNT	rwTTD	rwTTP	rwPFS	1-Year rwOS Landmark Analysis
A	13.50 (12.80 to 14.50)*	22.50 (N/A)	7.03 (6.27 to 9.97)			0.57 (0.52 to 0.57)
B	15.78 (12.2 to 24.59); 8.58 (7.56 to 10.36)†	12.95 (10.29 to 14.73)	3.25 (2.76 to 3.75)			0.54 (0.48 to 0.59); 0.41 (0.32 to 0.49)†
C	8.67 (6.83 to 10.02)	11.60 (8.80 to 16.10)	4.70 (3.68 to 5.52)			0.40 (0.35 to 0.46)
D	9.15 (8.82 to 9.51)	14.03 (12.89 to 15.15)	3.21 (3.21 to 3.44)	5.41 (5.18 to 5.67)	3.28 (3.18 to 3.41)	0.42 (0.41 to 0.43)
E	12.69 (11.7 to 13.87)	12.07 (11.24 to 13.48)	3.63 (3.40 to 3.87)			0.51 (0.49 to 0.53)
F	12.30 (9.61 to 16.94)	12.50 (9.29 to N/A)	4.60 (3.71 to 6.32)	9.37 (7.42 to 11.93)	9.37 (7.42 to 11.93)	0.40 (0.34 to 0.48)

NOTE. Empty data fields indicate variables that were not collected.

Abbreviations: N/A, not applicable; rwOS, real-world overall survival; rwPFS, real-world progression-free survival; rwTTD, time to treatment discontinuation; rwTTNT, time to next treatment; rwTTP, real-world time to progression.

*OS was calculated as months between PD-(L)1 initiation and disenrollment.

†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.

Once those criteria have been satisfied, the next level of examination is to determine whether the must-have data for a given study are available and are of sufficient quality and completeness,[5] and FDA guidance on end points is available to support oncology drug approvals.[19] Ongoing data curation processes used by these data sources allow for quick utilization of these data as many data partners assess the quality and completeness of data through ongoing quality assurance activities. Whereas this exercise demonstrated that not all data sources readily contain the same information, nor the same depth of clinical data of interest, they all contributed value toward estimating real-world treatment benefits and risks. That said, the completeness and accuracy of mortality information have not yet been assessed in results presented here. It would be helpful to establish best practices for managing missing data as well as curation efforts used to link and combine data sets, ensuring analytical consistency and establishing optimal effectiveness and other end points in a fashion similar to the clinical trials community.

RWD may differ from protocol-driven data collected through RCTs for various reasons. For example, RCTs have specified timing and frequency of follow-up assessment, per study protocol, and non–standard of care molecular or biomarker testing for inclusion eligibility and/or follow-up. Thus, generalizability can be enhanced using RWE as a supplement to RCTs or as external comparators.[20] A related challenge involves the lack of standardized biomarker assays in the real world, which may affect comparisons among results of studies conducted in different settings. Moreover, some variables of interest, such as the date of progression, are not typically available in structured EHRs, unavailable in claims data, and are not captured in tumor registries[21]; however, proxy measures, such as time to change in treatment, are often useful and have been used with both EHR and claims data. When available, the date of progression is generally found only in EHR text or other unstructured fields—for example, clinician notes, radiology and pathology reports, and documents from outside the institute scanned into the EHR. Even then, comparability of rwPFS with PFS is uncertain with less systematic follow-up for progression and requires additional research. Other important elements, such as the date of diagnosis, stage of disease, intended and received chemotherapy treatments, and various clinical and socioeconomic factors, may be missing for some patients, and the magnitude of missing data may vary across data sources.

In the case of claims-based data, some or all of the care that patients receive in a clinical trial setting may not generate insurance claims as the costs of these tests are borne by the trial sponsor and not the health care insurer, thus creating a systematic data gap. Because patients with advanced cancer are more likely to participate in clinical trials, this may be a meaningful issue in the current set of analyses. Similarly, in the case of EHR-based data, coverage may be limited when patients receive their care from multiple providers across different settings—for example, patients who receive portions of their care at an academic medical center and portions in the community practice setting. If a data provider sources information exclusively from the academic center or community practice, important clinical data gaps may exist. Furthermore, it is challenging to assemble and contrast the experience of patients who are treated at the same point in the course of their disease, as patients do not always present for care at prescribed intervals as they would in an RCT.

RWD may have some missing common data elements of interest, will almost always have nonstandard timepoints at which data from clinical encounters are documented, and reflect variability in types of diagnostic tests and data quality. Nonetheless, RWE is not posed here as a solution to every problem, but rather as a cost-effective and relatively reliable tool for understanding cancer treatment

TABLE 5. Median Times and 95% CI [indexed to initial PD-(L)1 inhibitor line start in advanced setting] Segmented by Treatment Setting and Demographic Characteristics as Described in Table 1

Variable	Data Set A No.	Data Set A Median rwTTD, Months (95% CI)*	Data Set A Median rwOS, Months (95% CI)†	Data Set B No.	Data Set B Median rwTTD, Median (95% CI)	Data Set B No.	Data Set B Median rwOS, Months (95% CI)†	Data Set C No.	Data Set C Median rwTTD, Months (95% CI)	Data Set C Median rwOS, Months (95% CI)	Data Set D No.	Data Set D Median rwTTD, Months (95% CI)	Data Set D Median rwOS, Months (95% CI)	Data Set E No.	Data Set E Median rwTTD, Months (95% CI)	Data Set E Median rwOS, Months (95% CI)	Data Set F No.	Data Set F Median rwTTD, Months (95% CI)	Data Set F Median rwOS, Months (95% CI)
Demographic																			
Age categories at PD-(L)1 inhibitor initiation, years																			
≤ 49	100	5.87 (3.97 to 9.80)	18.1 (11.87 to 21.63)	24	3.98 (1.84 to 17.03)	16	9.07 (2.66 to N/A)	21	6.44 (1.28 to 16.29)	12.02 (4.27 to N/A)	219	2.89 (2.30 to 3.64)	9.28 (7.77 to 12.07)	80	2.33 (1.43 to 4.40)	10.20 (7.96 to 17.36)	8	8.52 (2.77 to N/A)	N/A
50-64	723	6.23 (5.20 to 7.27)	13.60 (11.83 to 14.6)	252	3.62 (2.96 to 4.14)	164	8.25 (6.81 to 11.41)	129	5.35 (3.35 to 9.23)	9.33 (6.73 to 13.27)	2,047	3.21 (2.92 to 3.44)	9.34 (8.43 to 10.26)	863	3.53 (3.17 to 4.03)	13.84 (11.80 to 15.32)	65	6.16 (3.71 to 12.26)	16.94 (9.37 to N/A)
65-74	728	7.40 (6.10 to 9.60)	13.40 (12.07 to 14.93)	194	2.79 (2.27 to 3.75)	131	8.58 (6.67 to 12.33)	169	4.63 (3.45 to 6.44)	8.97 (5.78 to 11.14)	2,504	3.41 (3.21 to 3.67)	9.34 (8.79 to 10.26)	1,047	3.77 (3.30 to 4.23)	12.16 (10.49 to 14.33)	94	5.38 (3.42 to 6.90)	12.30 (7.55 to 21.58)
≥ 75	593	7.70 (6.37 to 9.37)	13.22 (11.83 to 14.61)	86	2.47 (1.45 to 4.87)	51	7.00 (5.03 to 15.68)	116	3.57 (1.84 to 5.26)	6.83 (4.24 to 9.13)	2,153	3.25 (3.18 to 3.61)	8.79 (8.23 to 9.28)	870	3.77 (3.30 to 4.23)	13.02 (10.62 to 14.79)	102	3.67 (2.82 to 5.74)	10.00 (8.71 to 15.55)
Sex																			
Female	950	6.80 (5.90 to 8.23)	13.70 (12.8 to 15.23)	275	3.35 (2.76 to 3.98)	187	8.94 (6.67 to 12.33)	212	4.76 (3.45 to 7.72)	9.33 (7.42 to 13.44)	3,751	2.98 (2.75 to 3.21)	8.43 (7.93 to 8.98)	1,351	3.83 (3.53 to 4.27)	14.79 (13.15 to 16.87)	125	5.03 (3.77 to 7.03)	13.23 (8.77 to 23.33)
Male	1,194	7.23 (6.10 to 8.43)	13.20 (12.13 to 14.5)	281	3.16 (2.37 to 3.81)	175	7.92 (6.77 to 10.06)	222	4.62 (3.35 to 5.52)	7.49 (6.34 to 10.02)	3,172	3.51 (3.21 to 3.70)	9.84 (9.38 to 10.72)	1,509	3.30 (2.87 to 3.77)	11.08 (10.16 to 12.39)	143	3.87 (2.93 to 6.40)	11.87 (8.40 to 20.40)
Clinical characteristic																			
Group stage at initial diagnosis																			
0/I				23	3.45 (0.92 to 5.69)	9	5.69 (0.49 to 9.70)				498	4.36 (3.67 to 5.28)	12.07 (10.69 to 14.03)	18	5.18 (3.43 to 15.65)	7.03 (4.87 to N/A)			
II				22	3.68 (1.41 to 6.38)	13	6.38 (1.48 to N/A)				426	3.90 (3.28 to 4.95)	11.84 (10.59 to 13.28)	17	4.10 (2.80 to N/A)	13.90 (4.03 to N/A)			
III				88	4.14 (2.76 to 3.88)	62	9.60 (6.67 to N/A)				1,494	3.67 (3.44 to 4.13)	9.84 (9.18 to 10.79)	63	5.73 (2.90 to 9.67)	14.87 (9.37 to 22.10)			
IV				248	3.35 (2.76 to 3.88)	187	8.77 (6.77 to 10.55)				4,334	2.89 (2.75 to 3.18)	8.26 (7.80 to 8.79)	161	4.32 (3.19 to 6.47)	12.10 (8.30 to 20.73)			
Unknown				175	2.47 (1.87 to 3.85)	91	7.92 (6.15 to 10.55)							10	8.83 (3.74 to N/A)	N/A			

(Continued on following page)

TABLE 5. Median Times and 95% CI [indexed to initial PD-(L)1 inhibitor line start in advanced setting] Segmented by Treatment Setting and Demographic Characteristics as Described in Table 1 (Continued)

Variable	Data Set A No.	Data Set A Median rwTTD (95% CI)* / rwOS (95% CI)†	Data Set B No.	Data Set B Median rwTTD (95% CI) / rwOS (95% CI)	Data Set C No.	Data Set C Median rwTTD (95% CI) / rwOS (95% CI)	Data Set D No.	Data Set D Median rwTTD (95% CI) / rwOS (95% CI)	Data Set E No.	Data Set E Median rwTTD (95% CI) / rwOS (95% CI)	Data Set F No.	Data Set F Median rwTTD (95% CI) / rwOS (95% CI)
Histology												
Non–squamous-cell carcinoma			369 / 239	3.19 (2.56 to 3.85) / 9.70 (7.73 to 12.62)	320	4.76 (3.81 to 5.81) / 8.67 (6.7 to 10.18)	4,678	3.34 (3.21 to 3.51) / 9.61 (9.11 to 10.30)	1,981	3.60 (3.30 to 3.90) / 14.14 (12.69 to 15.81)	194	4.15 (3.23 to 5.97) / 11.87 (8.77 to 21.80)
Squamous-cell carcinoma			147 / 93	3.25 (2.47 to 4.01) / 6.81 (4.87 to 8.78)	73	4.14 (1.68 to 7.33) / 8.38 (4.8 to 12.06)	1,983	3.21 (2.98 to 3.54) / 8.66 (7.84 to 9.25)	659	3.77 (3.30 to 4.30) / 10.36 (9.24 to 11.77)	61	7.55 (3.67 to 12.97) / 13.90 (8.57 to N/A)
NSCLC histology, not otherwise specified			40 / 30	3.88 (1.94 to 5.10) / 10.26 (5.13 to 13.32)	42	3.80 (2.07 to 9.13) / 7.92 (3.58 to 17.15)			220	3.57 (2.83 to 5.17) / 11.84 (7.89 to 16.08)	10	5.07 (2.29 to N/A) / 20.40 (6.13 to N/A)
Smoking status												
History of smoking					340	4.53 (3.45 to 5.35) / 8.67 (6.7 to 10.18)	6,185	3.28 (3.21 to 3.48) / 9.21 (8.85 to 9.64)	448	5.17 (3.90 to 6.53) / 19.17 (14.30 to 24.46)	182	4.27 (3.43 to 6.19) / 11.43 (8.40 to 20.73)
No history of smoking					94	5.52 (3.22 to 9.20) / 8.34 (6.04 to 12.88)	716	2.75 (2.49 to 3.11) / 8.69 (8.03 to 9.77)	38	3.30 (2.37 to 9.07) / 14.50 (4.57 to N/A)	28	6.08 (3.23 to 8.71) / 12.10 (8.71 to N/A)
Unknown/not documented									2,374	3.53 (3.30 to 3.77) / 12.00 (10.82 to 13.08)	59	5.97 (2.77 to 10.00) / 13.21 (9.37 to N/A)
PD-L1 expression status (among those tested)												
PD-L1 positive							512	4.10 (3.38 to 4.82) / 10.79 (9.05 to 13.28)	45	5.63 (2.83 to 18.23) / N/A	65	3.68 (2.82 to 5.97) / 9.63 (7.55 to N/A)
PD-L1 negative/not detected							690	2.75 (2.49 to 3.11) / 8.69 (7.48 to 9.84)	45	9.63 (N/A) / N/A	29	6.32 (4.19 to N/A) / 20.80 (11.43 to N/A)
Line of first PD-(L)1 inhibitor in advanced setting												
1	592	9.10 (7.97 to 12.40) / 19.83 (17.23 to 22.23)	144 / 74	2.76 (2.01 to 3.88) / 9.40 (6.38 to 12.23)	80	5.26 (2.63 to 6.64) / 9.17 (5.68 to 17.05)	2,074	3.90 (3.67 to 4.23) / 10.36 (9.48 to 11.18)	777	5.90 (4.93 to 6.80) / 20.78 (14.79 to 25.12)	77	5.03 (3.67 to 8.83) / 15.87 (9.87 to N/A)
2	1,174	6.57 (6.03 to 7.50) / 11.7 (10.97 to 12.83)	272 / 191	3.45 (2.60 to 4.11) / 7.92 (6.54 to 10.45)	205	4.53 (3.22 to 5.81) / 7.75 (5.75 to 10.28)	3,357	3.21 (2.82 to 3.21) / 8.66 (8.13 to 9.15)	1,414	3.00 (2.83 to 3.30) / 10.68 (9.83 to 11.93)	87	4.81 (2.74 to 7.00) / 9.07 (7.39 to 20.40)

(Continued on following page)

Stewart et al

TABLE 5. Median Times and 95% CI [indexed to initial PD-(L)1 inhibitor line start in advanced setting] Segmented by Treatment Setting and Demographic Characteristics as Described in Table 1 (Continued)

Variable	Data Set A No.	Data Set A — Median rwTTD, Months (95% CI) / Median rwOS, Months (95% CI)*	Data Set B No.	Data Set B — Median rwTTD, Median (95% CI)	Data Set B No.	Data Set B — Median rwOS, Months (95% CI)†	Data Set C No.	Data Set C — Median rwTTD, Months (95% CI) / Median rwOS, Months (95% CI)	Data Set D No.	Data Set D — Median rwTTD, Months (95% CI) / Median rwOS, Months (95% CI)	Data Set E No.	Data Set E — Median rwTTD, Months (95% CI) / Median rwOS, Months (95% CI)	Data Set F No.	Data Set F — Median rwTTD, Months (95% CI) / Median rwOS, Months (95% CI)
3	304	4.47 (3.8 to 6.1) / 12.8 (10.7 to 14.67)	96	3.16 (2.56 to 4.50)	63	10.65 (6.87 to 17.49)	85	4.63 (2.30 to 8.05) / 9.33 (6.04 to 12.02)	1,011	2.98 (2.75 to 3.44) / 9.02 (7.80 to 9.97)	448	3.53 (3.00 to 4.23) / 10.72 (9.04 to 13.87)	51	5.67 (3.43 to 8.71) / 15.29 (9.63 to N/A)
≥ 4	74	3.83 (2.83 to 5.47) / 14.2 (10.1 to 17.17)	44	3.35 (1.97 to 4.37)	34	5.10 (2.14 to 11.90)	65	4.76 (3.94 to 11.10) / 8.67 (5.26 to 13.27)	481	2.59 (2.30 to 3.18) / 8.52 (6.89 to 10.46)	221	3.30 (2.60 to 4.23) / 12.00 (8.25 to 15.58)	54	3.47 (2.77 to 7.16) / 10.43 (6.90 to 21.80)

NOTE. Empty data fields indicate variables that were not collected.

Abbreviations: N/A, not applicable; NSCLC, non–small-cell lung cancer; PD-(L)1, programmed death 1/programmed death-ligand 1; rwOS, real-world overall survival; rwTTD, time to treatment discontinuation.

*rwOS was calculated as the time between PD-(L)1 initiation and disenrollment.

†rwOS estimates include sites with Social Security or state death data available; excluded are sites with only local/electronic health record death data available.

TABLE 6. Correlation Between Real-World Overall Survival and Real-World Extracted End Points Using Spearman's Rank Correlation Coefficient

Data Set	Comparison	No.	Correlation (95% CI)
A	rwOS v rwTTNT	83	0.36 (0.15 to 0.53)
	rwOS v rwTTD	254	0.63 (0.55 to 0.70)
B	rwOS v rwTTNT	86	0.77 (0.67 to 0.85)
	rwOS v rwTTD	254	0.62 (0.54 to 0.69)
C	rwOS v rwTTNT	96	0.70 (0.58 to 0.79)
	rwOS v rwTTD	295	0.89 (0.86 to 0.91)
D	rwOS v rwTTNT	1,203	0.61 (0.57 to 0.64)
	rwOS v rwTTD	4,337	0.80 (0.79 to 0.81)
	rwOS v rwPFS	4,337	0.75 (0.74 to 0.76)
	rwOS v rwTTP	2,286	0.60 (0.57 to 0.63)
E	rwOS v rwTTNT	358	0.62 (0.54 to 0.68)
	rwOS v rwTTD	1,456	0.77 (0.75 to 0.79)
F	rwOS v rwTTNT	39	0.46 (0.33 to 0.81)
	rwOS v rwTTD	142	0.80 (0.66 to 0.85)
	rwOS v rwPFS	142	0.84 (0.62 to 0.86)
	rwOS v rwTTP	55	0.56 (0.21 to 0.71)

Abbreviations: 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; rwTTP, real-world time to progression.

heterogeneity and effectiveness. RWD provides an opportunity to rapidly address clinically relevant questions. RWE also provides an opportunity to investigate the effectiveness of therapies in patient populations and in combinations and treatment sequences that have not been studied in a clinical trial and can supplement drug development programs in meaningful ways. For example, effectiveness of therapies and long-term surveillance after initial FDA approval of medications is an area in which RWD can provide important insights. In addition, the scarcity of patients or loss of clinical equipoise may make random assignment difficult or impossible. By creating a well-reasoned approach for assessing the quality of RWD and how to apply these data to the development of RWE, we can ensure that the massive amounts of data that are generated in the course of routine health care provision and transactions can be useful for advancing drug development and efforts at generating knowledge.

This project demonstrates that acceptable data can be aggregated from research-ready RWD with short lag time and that outcomes can be measured from these data sources. Additional studies are needed to further support the use of RWE and inform the development of regulatory guidance. Standardizing definitions for real-world end points and determining appropriate analytic methodologies for RWD will be critical for broader adoption of real-world studies and will provide greater confidence in associated findings. As more refined and standardized approaches are developed that incorporate deep clinical and bioinformatics expertise, the greater the utility of RWD will be for detecting even small, but important, differences in treatment effects.

AFFILIATIONS

[1]Friends of Cancer Research, Washington, DC
[2]Cota Healthcare, New York, NY
[3]IQVIA, Durham, NC
[4]OptumLabs, Cambridge, MA
[5]Flatiron Health, New York, NY
[6]University of Iowa College of Public Health, Iowa City, IA
[7]PCORnet, Washington, DC
[8]Kaiser Permanente, Oakland, CA
[9]Cancer Research Network, Oakland, CA
[10]Mayo Clinic, Rochester, MN
[11]US Food and Drug Administration, Bethesda, MD
[12]National Cancer Institute, Bethesda, MD

A.P.A. participated in this work before joining the US Food and Drug Administration (FDA). This work and its related conclusions reflect the independent work of study authors and do not necessarily represent the views of the FDA or United States.

CORRESPONDING AUTHOR

Jeff Allen, PhD, Friends of Cancer Research, 1800 M St NW, Suite 1050 South, Washington, DC 20036; Twitter: @canceresrch; e-mail: jallen@focr.org.

PRIOR PRESENTATION

Presented at the Friends of Cancer Research public meeting, The Future Use of Real-World Evidence, Washington, DC, July 10, 2018.

Stewart et al

REFERENCES

1. Martin L, Hutchens M, Hawkins C, et al: How much do clinical trials cost? Nat Rev Drug Discov 16:381-382, 2017
2. Frieden TR: Evidence for health decision making: Beyond randomized, controlled trials. N Engl J Med 377:465-475, 2017
3. Franklin JM, Schneeweiss S: When and how can real world data analyses substitute for randomized controlled trials? Clin Pharmacol Ther 102:924-933, 2017
4. Khosla S, White R, Medina J, et al: Real world evidence (RWE): A disruptive innovation or the quiet evolution of medical evidence generation? F1000 Res 7:111, 2018
5. Ali AK, Hartzema AG (eds): Enriched Studies in Post-Authorization Safety Studies of Medicinal Products: The PASS Book (ed 1). London, United Kingdom, Elsevier, 2018, pp 150-158
6. Corrigan-Curay J, Sacks L, Woodcock J: Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 320:867-868, 2018
7. US Food and Drug Administration: PDUFA reauthorization performance goals and procedures fiscal years 2018 through 2022. https://www.fda.gov/media/99140/download
8. US Food and Drug Administration: Use of Electronic Health Record Data in Clinical Investigations: Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry
9. US Congress: 21st Century Cures Act. P.L. 114-255, 2016.
10. US Congress: Prescription Drug User Fee Act. P.L. 102-571, 1992.
11. US Food and Drug Administration. BLA multidisciplinary review and evaluation: BLA 761049. Bavencio (Avelumab). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761049Orig1s000MultidisciplineR.pdf
12. US Food and Drug Administration. Summary Review for Regulatory Action: Blincyto (Blinatumomab). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000MedR.pdf
13. Berger M, Daniel G, Frank K, et al: A Framework For Regulatory Use of Real-World Evidence. Durham, NC, Duke-Margolis Center for Health Policy, 2017. https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf
14. Ross TR, Ng D, Brown JS, et al: The HMO Research Network virtual data warehouse: A public data model to support collaboration. EGEMS (Wash DC) 2:1049, 2014
15. Huang G, Sun X, Liu D, et al: The efficacy and safety of anti-PD-1/PD-L1 antibody therapy versus docetaxel for pretreated advanced NSCLC: A meta-analysis. Oncotarget 9:4239-4248, 2017
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. Curtis MD, Griffith SD, Tucker M, et al: Development and validation of a high-quality composite real-world mortality endpoint. Health Serv Res 53:4460-4476, 2018
18. Carrigan G, Whipple S, Taylor MD, et al: An evaluation of the impact of missing deaths on overall survival analyses of advanced non-small cell lung cancer patients conducted in an electronic health records database. Pharmacoepidemiol Drug Saf 28:572-581, 2019

19. Johnson JR, Williams G, Pazdur R: End points and United States Food and Drug Administration approval of oncology drugs. J Clin Oncol 21:1404-1411, 2003

20. Gökbuget N, Kelsh M, Chia V, et al: Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood Cancer J 6: e473, 2016

21. In H, Bilimoria KY, Stewart AK, et al: Cancer recurrence: An important but missing variable in national cancer registries. Ann Surg Oncol 21:1520-1529, 2014

■ ■ ■

FRIENDS
of CANCER
RESEARCH

ALEXANDRIA®
Building the Future of Life-Changing Innovation™

8th Annual Blueprint for Breakthrough Forum
Validating Real-World Endpoints for an Evolving Regulatory Landscape
September 18, 2019
Washington, DC

Introduction

Advances in data analytics and data capture through electronic health records (EHRs) and medical/pharmacy claims have brought the opportunities and challenges associated with using real-world evidence (RWE) to the forefront of the US healthcare industry. Increasingly, the promise of RWE to contribute to a more complete picture of the benefits and risks associated with therapies, when paired with results from randomized, controlled clinical trials, is being realized. RWE provides an opportunity to collect data rapidly on a broader patient population outside of a strict clinical trial protocol to help provide evidence for new indications or describe rare safety events, provide information that is more generalizable than clinical trial results, and confirm clinical benefit in the post-market setting.

Applications for RWE extend across the spectrum of therapeutics development from regulatory decision-making, to clinical use, to coverage and payment decisions. In the regulatory space, RWE has been utilized most frequently to evaluate drug safety through pharmacovigilance and adverse event monitoring in pre- and post-approval settings. However, RWE has increasingly been used to support effectiveness claims. Beyond regulatory decisions, RWE is frequently used to support clinical trial design, development of clinical practice guidelines, confirmation of population/subgroup size, and payment decisions including formulary placement.

Significant progress has been made in data collection efforts to support use of RWE in regulatory settings, however challenges remain, chiefly with developing methodologies and standard definitions when organizing and analyzing data from different sources that ensure appropriate translation of real-world data (RWD) into "fit-for-use" RWE. Friends of Cancer Research initially proposed a pilot project, comprised of six leading healthcare organizations with oncology data, to develop a data set curation process and framework to operationalize RWD collection and explore potential real-world endpoints that may be fit for regulatory purposes as well as assessing long-term benefits of a product. The results of this pilot were presented in July 2018[1].

A result of the initial RWE collaborative oncology research pilot showed that several different data sets were able to extract real-world time to treatment discontinuation (rwTTD) in a relatively consistent manner. In addition, rwTTD correlated well to real-world overall survival (rwOS) in the context of anti-PD-(L)1 use for treating advanced non-small cell lung cancer (aNSCLC).

---

[1] Friends of Cancer Research. The Future Use of Real-World Evidence Meeting. July 10, 2018. Washington D.C.

Establishing a New Pilot Project

Informed by the recent release of FDA's Real-world Evidence Program framework[2], and building off the successes from our 2018 RWE pilot project, "Establishing a Framework to Evaluate Real-World Endpoints", *Friends* initiated a new pilot project to further characterize how RWD can fill evidence gaps about the performance of approved agents used in a real-world setting. Additional insights can also be gained about populations that may not have been included in clinical trials for various reasons, such as feasibility or ethical concerns, rarity of the cancer, etc.

The RWE Pilot Project 2.0, which is ongoing, has been designed to provide insight into the opportunities and limitations of real-world endpoints and the ability to compare differences in effectiveness between therapies in terms of patient characteristics and observed effectiveness. The pilot will also allow us to understand the extent to which similar conclusions may be observed in real-world patient populations using established clinical trial patient populations (defined by applying agreed upon inclusion and exclusion criteria) as a relative benchmark. In addition to evaluating real-world endpoints in a case study, this pilot project will also provide an opportunity to start to a) align on how to evaluate data quality, b) define data standards, and c) determine essential elements of a potential analytic framework to evaluate real world endpoints.

This pilot project was initiated to help determine whether RWD can be used to develop an early perspective on real-world outcomes, as defined by real-world endpoints from EHR and claims data. Additionally, we sought insight into the generalizability of clinical trial results to patients treated in real-world settings. The pilot project evaluates the performance of real-world endpoints across multiple data sets by focusing on a common question: ***What are the real-world outcomes for aNSCLC patients treated with frontline therapies in usual care settings?***

Participating organizations began by agreeing upon necessary data elements to define demographic and clinical characteristics and internal processes to define real-world endpoints in the context of clinical trial definitions, taking into account the FDA regulatory framework and the variation of available data within EHR and claims-based datasets. While the project is in the preliminary stages, later phases of the pilot project will ultimately help evaluate whether the various data sets included in this study can reach similar conclusions upon application of uniform critical inclusion/exclusion criteria and appropriate analytic methodologies.

Pilot Project Study Design and Objectives

The on-going RWE Pilot Project 2.0 leverages parallel analyses from common data elements across multiple data sources to assess three frontline treatment approaches in real-world patients with aNSCLC. It is a retrospective observational analysis derived from EHR and claims data. The data sets generated for the study include all relevant, retrospective patient-level HIPAA-compliant de-identified data available for eligible individuals up to a single specific data cutoff date of March 31, 2018.

---

[2] US Food and Drug Administration. FRAMEWORK FOR FDA'S REAL-WORLD EVIDENCE PROGRAM. December 2018. Retrieved from https://www.fda.gov/media/120060/download.

It is important to note that this pilot is not intended to replicate results observed in RCTs nor draw formal conclusions regarding the performance of any product in real-world settings.

The study design includes two objectives that are being carried out in a phased manner:

Objective 1: Description of demographic and clinical characteristics of patients with aNSCLC receiving frontline chemotherapy doublet, PD-(L)1 monotherapy, or PD-(L)1 + doublet chemotherapy.

Purpose: Provide baseline understanding of the similarities/differences among the datasets to better understand what confounding factors may need to be considered when interpreting the data.

Objective 2: Evaluate treatment effect size in frontline therapy regimens using real-world endpoints.

Purpose: Agree on data source specific definitions and measurement of endpoints assessed through real-world data, in order to ensure reliability, consistency, and conservation of clinical meaning.

Methods

PROJECT DETAILS

BROAD COHORT AND INCLUSION / EXCLUSION CRITERIA	**Inclusion**: • <u>EHR-based data sets</u>: Physically present at a practice or having an encounter (defined as a physician visit, intravenous medication administration, or vitals documentation) in the real-world database on at least two separate occasions on or after January 1, 2011 until data cutoff date (March 31, 2018). • <u>Claims-based data sets</u>: Continuous enrollment in the health plan beginning on or after January 1, 2011 and before data cutoff date (March 31, 2018). <u>All Data Sets</u> • Diagnosis of (identified by ICD-9 code of 162.x or ICD-10 code of C33.x or C34.x) or pathology consistent with NSCLC • Evidence of advanced disease on or after January 1,2011 with advanced disease defined as either stage IIIB, IIIC or IV NSCLC at initial diagnosis or early -stage (stages I, II, and IIIA) NSCLC with a recurrence or progression to advanced or metastatic status. • Date of recurrence or progression defined based on physician assessment through curation or as last radiology date prior to use of chemotherapy agents of interest. • Regimen given to NSCLC patients subsequent to the patient's date of advanced diagnosis including all agents received within 30 days following the day of first infusion:

○ Platinum doublet chemotherapy (cisplatin, carboplatin, oxaliplatin, or nedaplatin with pemetrexed, paclitaxel, nab-paclitaxel, gemcitabine)
○ PD-(L)1 monotherapy (pembrolizumab, nivolumab, atezolizumab)
○ Any PD-(L)1 + doublet chemotherapy combination (pembrolizumab, pemetrexed and platinum or pembrolizumab, platinum and paclitaxel or nab-paclitaxel)

Exclusion:
- <u>EHR-based data sets</u>: Greater than 120 days from time of advanced diagnosis to evidence of clinical encounter
- <u>Claims-based data sets</u>: Less than 180 days baseline before the date of diagnosis

<u>All Data Sets</u>
- Incomplete historical treatment data available within the real-world database
- Treatment at sites without consistent historical reporting such that confidence of identification of frontline therapy is diminished.
- Received other therapies during frontline.

EHR AND CLAIMS-DERIVED ENDPOINTS DEFINITION AND ANALYTICAL GUIDANCE	Index Date ● *Definition*: Earliest drug episode (e.g., first administration or non-cancelled order) of the frontline therapy for advanced disease. Real-world Overall Survival (rwOS) ● *Definition*: Length of time from the index date to the date of death, or disenrollment (need to define gap in enrollment). For claims data, health plan disenrollment date is incorporated if deaths are not captured among those who leave health plan coverage. ● *Censor date*: Last structured recorded clinical activity within the real-world database including prescription, office or institutional billing claims data, or end of follow-up period, whichever occurs earliest. Real-world Time to Next Treatment (rwTTNT) ● *Definition*: Length of time from the index date to the date the patient received an administration of their next systemic treatment regimen or to their date of death if there is a death prior to having another systemic treatment regimen. ● *Censor date*: Last known activity or end of follow-up. Real-world Time to Treatment Discontinuation (rwTTD) ● *Data*: Length of time from the index date to the date the patient discontinues frontline treatment (i.e., the last administration or non-cancelled order of a drug contained within the same frontline regimen). ○ Discontinuation is defined as: ■ having a subsequent systemic therapy regimen after the frontline treatment;

- 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 frontline regimen.
- *Censor date*: Last known usage (i.e., administration or non-cancelled order) of frontline treatment.

Real-world Progression Free Survival (rwPFS)
- *Definition*: Length of time from the index date to the date of a real-world progression (rwP) event (i.e., distinct episode in which the treating clinician concludes that there has been growth or worsening in the aNSCLC based on review of the patient chart) at least 14 days after frontline treatment initiation, or death.
- *Censor date*: Date of rwTTNT. For patients without a rwP event or a rwTTNT event and at least 180 days follow-up from last frontline treatment, censor date will be rwTTD event date.

ANALYSES	Graphs 1-13:
	• Description of demographic and clinical characteristics of aNSCLC patients treated with one of the above frontline treatment categories, example characteristics include:
	o Demographic: age, gender, and race
	o Clinical: smoking status, histology, group stage at time of initial diagnosis, PD-L1 expression status and staining, performance (ECOG) status, and presence/absence of brain metastasis.
	o Treatment description of population by treatment category including time from advanced diagnosis to index date (not shown), year of index date, structured follow-up time from advanced diagnosis (not shown), and structured follow-up time from index date (not shown).
	Graphs 14-21:
	• Real-world endpoints (rwOS, rwPFS, rwTTNT, rwTTD) for aNSCLC patients treated with frontline therapies of interest in the advanced setting (Kaplan-Meier curves for each endpoint and median time to event estimates), stratified by treatment category.

Contributing Organizations for Pilot Project Study

Aetion

Aetion is a health care technology company that delivers real-world evidence for biopharma, payers, and regulatory agencies. The Aetion Evidence Platform™ analyzes data from the real world to produce transparent, rapid, and scientifically validated answers to guide treatment development, commercialization, and payment innovation. For this engagement, Aetion is supporting data aggregation and analysis across participant data sources.

ASCO CancerLinQ/Concerto HealthAI

CancerLinQ®, an initiative of the American Society of Clinical Oncology (ASCO), is a web-based platform that collects and analyzes structured and unstructured real-world cancer data from multiple electronic health record systems (EHRs) to improve care and drive new research. Concerto HealthAI, a technology leader in AI solutions for real-world oncology data, also aggregates structured and unstructured data from multiple EHRs, to revolutionize clinical and outcomes research that will enhance patient care and improve outcomes. Concerto HealthAI curates data from both sources, which together hold two million patient records from more than 100 practices, generating de-identified datasets that support high-quality research by non-profit organizations, academia, government agencies, and industry.

COTA

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

Flatiron Health

The Flatiron Health database is a nationwide longitudinal, demographically and geographically diverse database derived from de-identified electronic health record (EHR) data from over 280 cancer clinics (~800 sites of care) representing more than 2.2 million US cancer patients available for analysis. The de-identified patient-level data in the EHRs includes structured data (e.g., laboratory values, and prescribed drugs) in addition to unstructured data collected via technology-enabled chart abstraction from physician's notes and other unstructured documents (e.g., biomarker reports).

IQVIA™

IQVIA™ is a leading global provider of information, innovative technology solutions and contract research services focused on using data and science to help healthcare clients find better solutions for their patients. For this engagement, IQVIA's Real World team analyzed data from structured Oncology Electronic Medical Records (EMR) fields, combined with medical and pharmacy claims, and supplemented where possible with NLP and chart abstraction. IQVIA's data is sourced through multiple partners, including Inteliquet and IntrinsiQ Specialty Solutions. IntrinsiQ's affiliate Xcenda supported the project through the use of both their NLP technology and their chart abstraction team. The data are comprised of all payer types, all practice sizes and both community practices and hospital centers across the United States. The IQVIA Integrated EMR platform includes linkage to medical and pharmacy claims to capture activity outside of the oncology site and to apply a mortality index algorithm.

Kaiser Permanente Analysis Using Cancer Research Network

The Cancer Research Network originated as an NCI-funded consortium of research groups affiliated with integrated health care systems across the US; the participating health systems are a subset of those participating in the Health Care Systems Research Network. In the early 2000's, the CRN created the Virtual Data Warehouse (VDW), a living common data model to facilitate collaborative research across these health care systems. Data in the VDW are extracted from multiple source databases, including, but

not limited to, electronic health records, legacy databases, and databases for specific applications such as prescription medication orders and fills. The VDW is maintained by each research group with the possibility of pooling data under IRB-approved research protocols. For most participating institutions, the VDW has essentially complete information on care dating back to 1996 or earlier for most data domains. Domains include health plan enrollment periods, cancer registries, encounters including diagnoses and procedures, prescription and infusion medications, laboratory results, and other areas. The data provided are results from one of the participating CRN organizations.

Mayo Clinic Analysis using OptumLabs® Data Warehouse

OptumLabs® is an open, collaborative research and innovation center founded in 2013 as a partnership between Optum and Mayo Clinic with its core linked data assets in the OptumLabs Data Warehouse (OLDW). The database contains de-identified, longitudinal health information on enrollees and patients, representing a diverse mixture of ages, ethnicities and geographical regions across the United States. The claims data in OLDW includes medical and pharmacy claims, laboratory results and enrollment records for commercial and Medicare Advantage enrollees. The EHR-derived data includes a subset of EHR data that has been normalized and standardized into a single database. For this pilot project, clinical information from the health plan's cancer registries and prior authorization systems were linked to the health plan's administrative claims data for privately insured enrollees and death records.

McKesson

McKesson Data, Evidence & Insights uses robust regulatory-grade data to deliver meaningful, timely insights so informed clinical, regulatory, commercial and payer strategy decisions can be made. McKesson leverages iKnowMed[SM], its oncology practice electronic health record (EHR) system, as well as reimbursement data from integrated structured retrospective and prospective databases. Using this innovative model, biopharma and life sciences companies are able to bring life-saving drugs to market faster and support rapid label expansion, as well as create commercialization plans and outreach strategies to support appropriate utilization of commercial products.

SEER-Medicare

The SEER-Medicare data reflect the linkage of two large population-based sources of data that provide detailed information about Medicare beneficiaries with cancer. SEER is supported by the Surveillance Research Program (SRP) within the Division of Cancer Control and Population Sciences (DCCPS) at the National Cancer Institute, which provides national leadership in the science of cancer surveillance as well as analytical tools and methodological expertise in collecting, analyzing, interpreting, and disseminating population-based cancer statistics. SEER collects demographic, tumor characteristic, treatment, and survival data as a part of legally mandated reporting requirements for cancer surveillance from registries within 19 geographic areas representing 34% of the US population. The SEER data provide information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. Medicare is a federally funded insurance program in the US administered by CMS, insuring beneficiaries over 65 years old or those meeting other requirements. The Medicare data include administrative claims for health services submitted for reimbursement across various care settings. The publicly available linked dataset provides the ability to study population-based epidemiologic questions related to screening, treatment, costs, and outcomes among elderly cancer patients.

Syapse

Syapse is on a mission to improve outcomes for every cancer patient through precision medicine. By bringing together leading healthcare providers into a unified ecosystem, we have built one of the world's largest Learning Health Networks of provider-driven precision medicine, comprising over 10% of cancer care at over 400 hospitals in the United States and South Korea. Our real-world evidence platform integrates clinical, molecular, treatment, and outcomes data from multiple structured and unstructured sources including EHRs, registries, radiology systems, and molecular testing labs, building a complete, longitudinal, and continuously updated picture of each cancer patient's journey, including non-oncology care. In collaboration with our partners — including Advocate Aurora Health, CommonSpirit Health, Henry Ford Health System, Providence St. Joseph Health, and Seoul National University Hospital — we are working toward a future in which all cancer patients have access to the best personalized care.

Tempus

The Tempus real-world oncology database comprises longitudinal oncology care data from a variety of stakeholders across the healthcare ecosystem (e.g. community practices, integrated delivery networks, and academic institutions) including more than 50 NCI cancer centers. Our data assets include structured patient-level data from various healthcare sources and formats (e.g. electronic medical records, enterprise data warehouses, tumor and death registries), integrated with abstracted clinical information from unstructured documents (e.g. physician notes, pathology, radiology, laboratory, and genomic sequencing/biomarker reports) and corresponding molecular data produced by our lab. This is acquired through purpose-built, semi-automated pipelines and harmonized to standard terminologies (eg. MedDRA, NCBI, NCIt, NCIm, RxNorm, SNOMED, etc.). Data captured include demographic, diagnostic, biomarker and genomic testing, laboratory values, treatment, outcome, and adverse event data.

FRIENDS
of CANCER
RESEARCH

ALEXANDRIA.
Building the Future of Life-Changing Innovation™

Results[3,4]

Graph 1. Percentage of Patients with Advanced vs Early Stage NSCLC at Initial Diagnosis

---
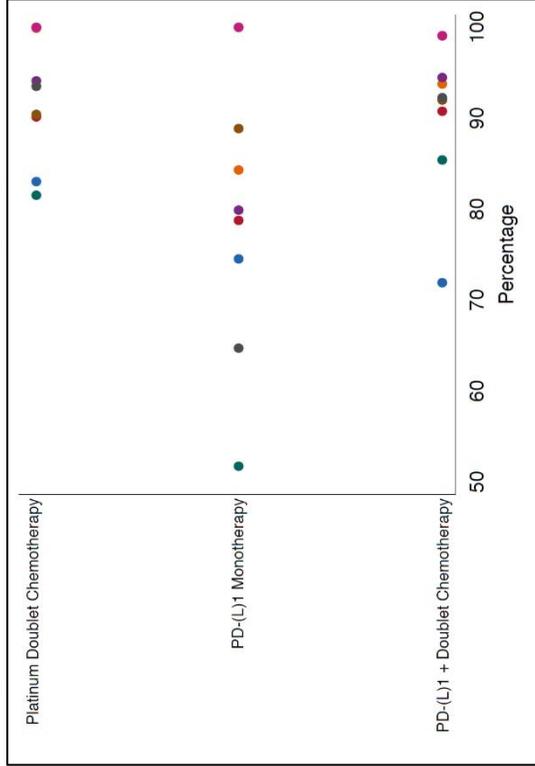
[3] * Graphs are based on structured or unstructured information depending on the data source
[4] Graphs represent data of patients with values reported. Missing/unknown data are not represented in these graphs. See Assumptions and Limitations section for more explanation.

Graph 2. Percentage of aNSCLC Patients Age 75 or Older at Index
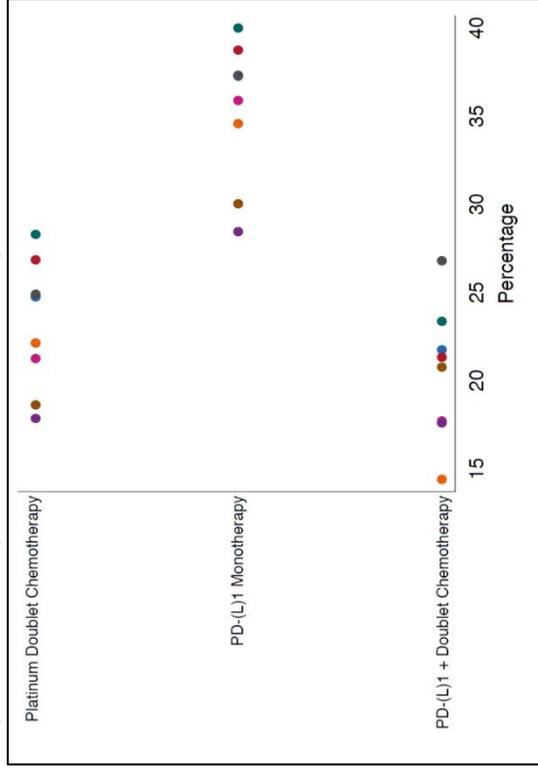
Graph 3. Median and Lower/Upper Quartiles of Age at Index
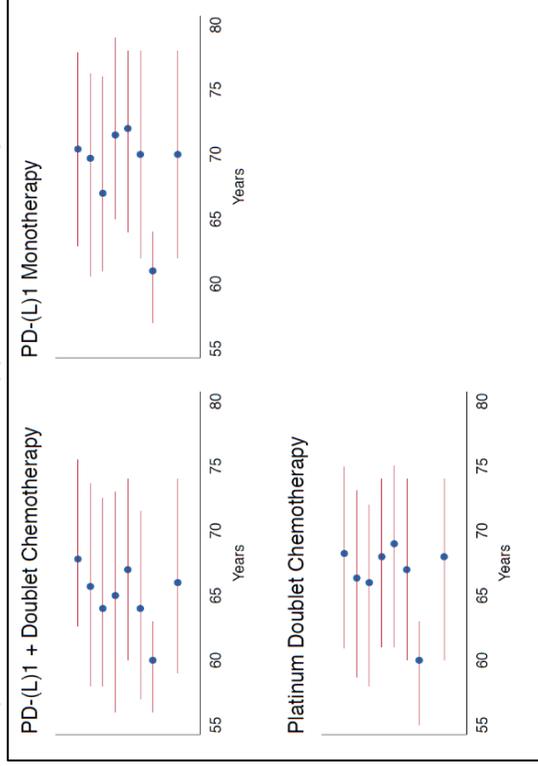
Graph 4. Percentage of Male aNSCLC Patients
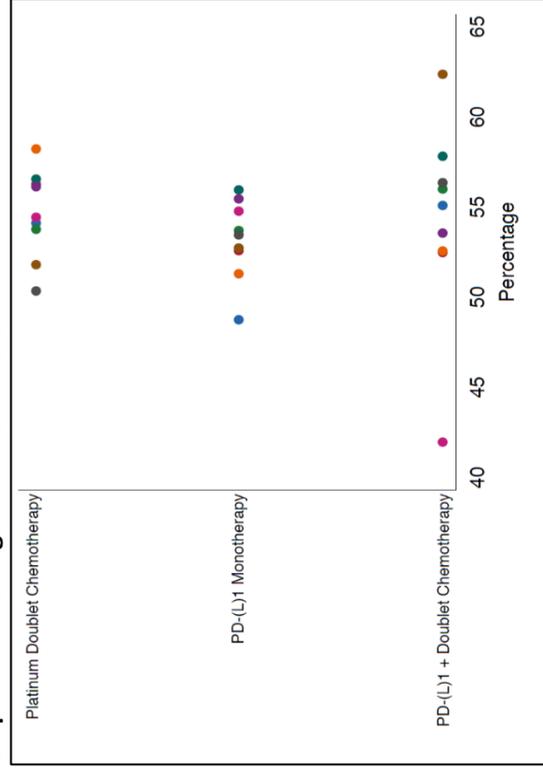
Graph 5. Percentage of aNSCLC Patients Whose Race is White

35
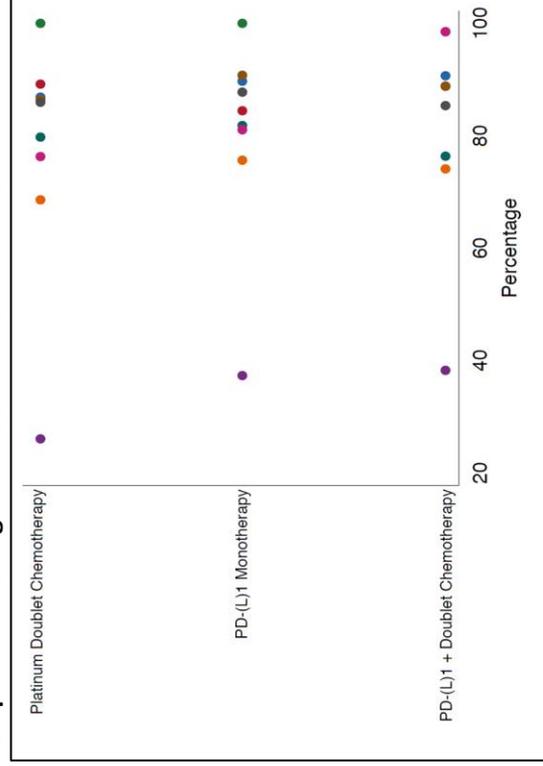
Graph 6. Percentage of aNSCLC Patients with a History of Smoking*
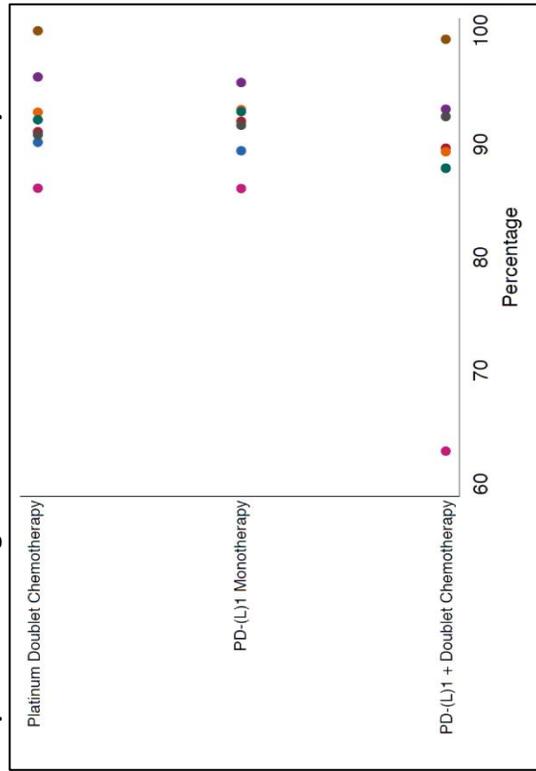
Graph 7. Histology of Patients with aNSCLC by Treatment Category

NOS: Histology not otherwise specified
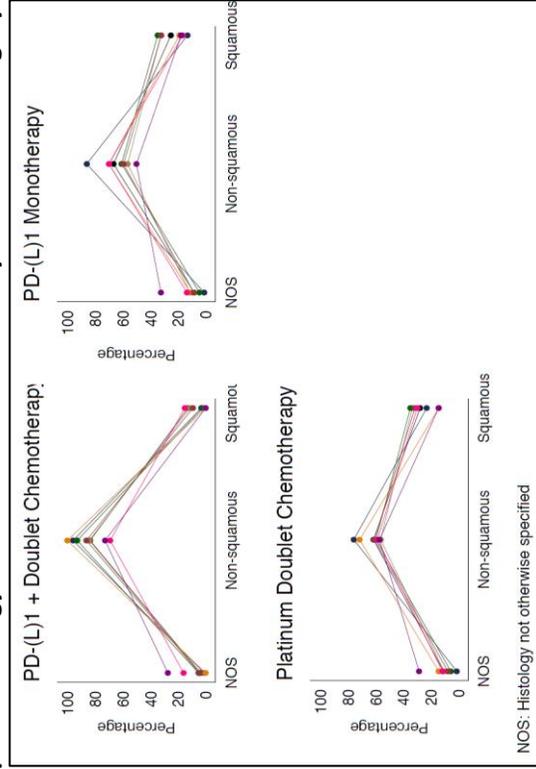
Graph 8. Group Stage of Patients with aNSCLC by Treatment Category[5]
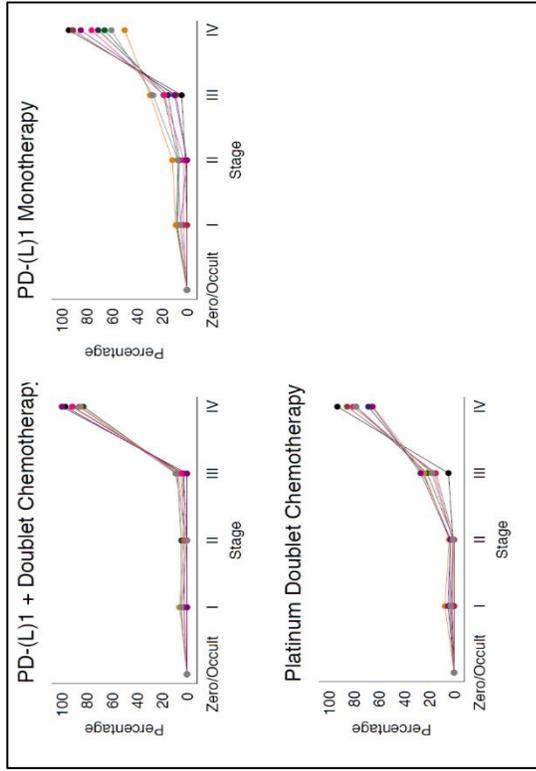
Graph 9. Percentage of PD-L1 Positive* aNSCLC Patients
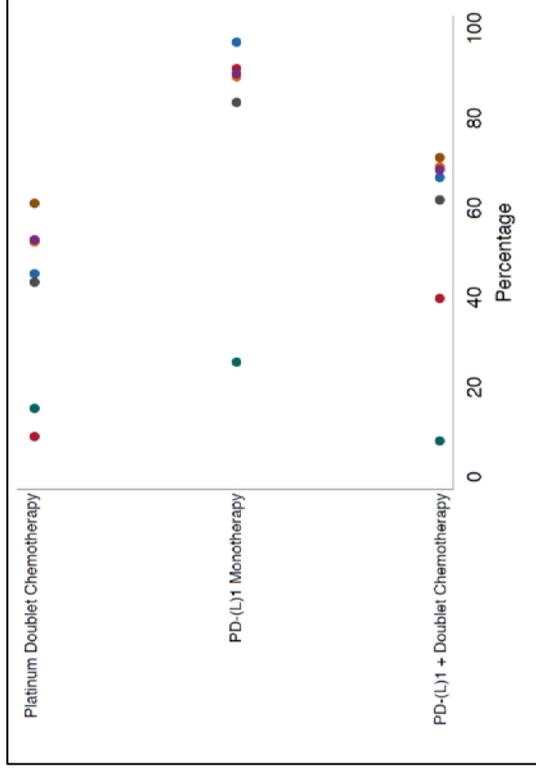
Graph 10. PD-(L)1 Staining of Patients with aNSCLC by Treatment Category[6]*

Graph 11. aNSCLC Patient Performance Status (ECOG) at Index[7]
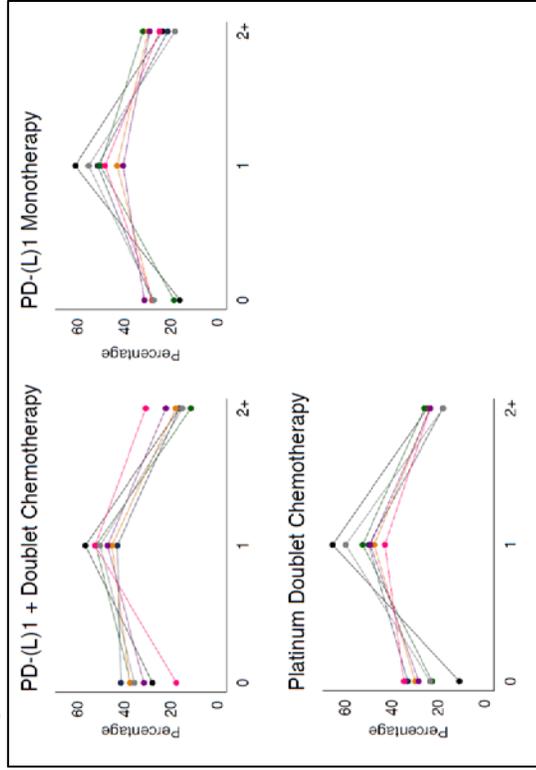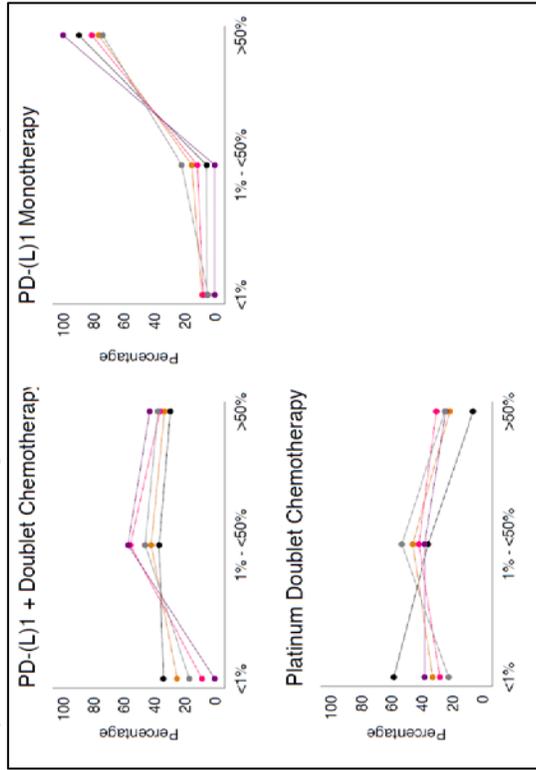
[6] Among patients tested for PD-(L)1. Five (out of ten) data partners reported PD-(L)1 staining.
[7] For patients where performance status was reported, patients with unkown performance status not included in percent calculation.

Graph 12. Percentage of aNSCLC Patients with Brain Metastasis at Index[8]
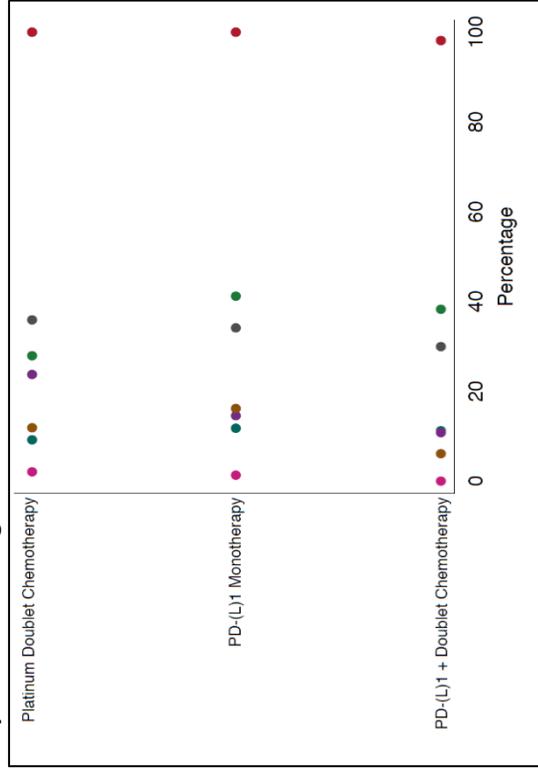
Graph 13. Year of Index Date by Treatment Category[9]*

ALEXANDRIA.
Building the Future of Life-Changing Innovation™

FRIENDS
of CANCER
RESEARCH
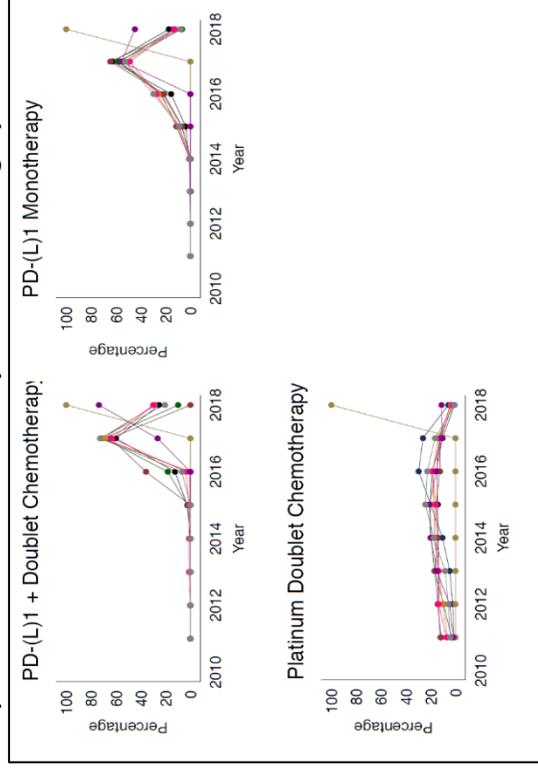
Graph 14. Kaplan-Meier Curve of rwOS by Treatment Category

Graph 15. Kaplan-Meier Curve of rwPFS by Treatment Category

Graph 16. Kaplan-Meier Curve of rwTTD by Treatment Category
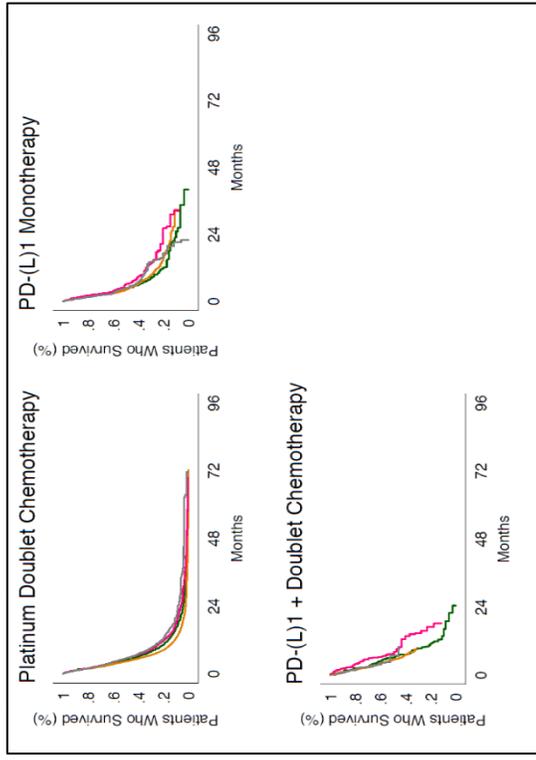
Graph 17. Kaplan-Meier Curve of rwTTNT by Treatment Category
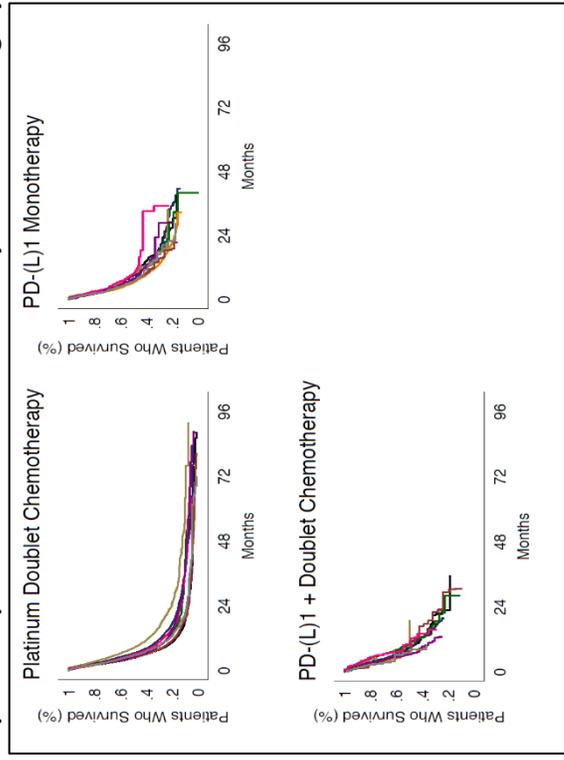
Graph 18. Estimates of Median (95% CI) rwOS by Treatment Category

95% confidence limits shown. Open circles represent unbounded limits. Undefined estimates were excluded.
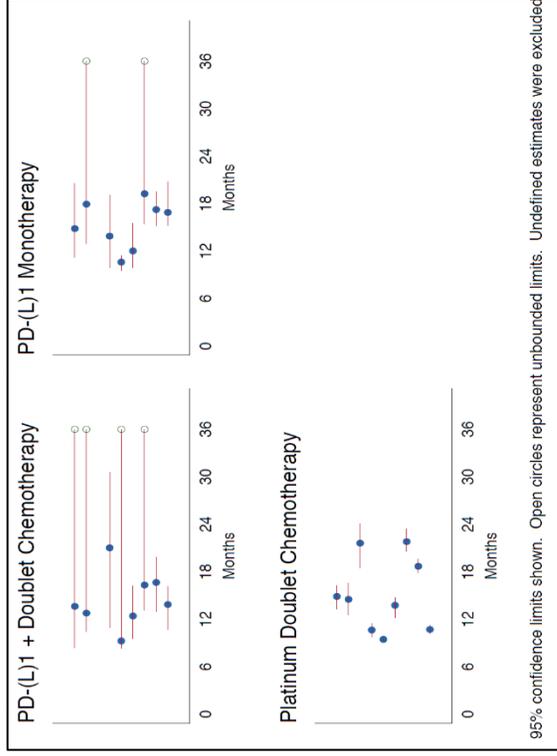
Graph 19. Estimates of Median (95% CI) rwPFS by Treatment Category

95% confidence limits shown. Open circles represent unbounded limits. Undefined estimates were excluded.
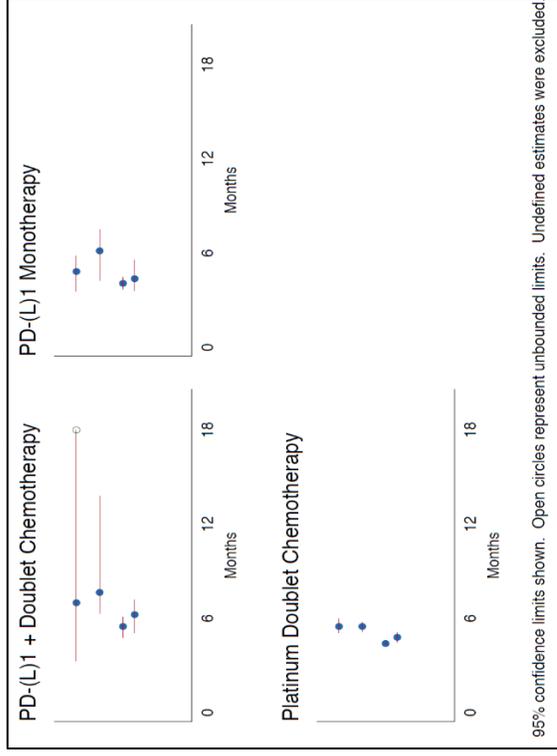
Graph 20. Estimates of Median (95% CI) rwTTD by Treatment Category

95% confidence limits shown. Open circles represent unbounded limits. Undefined estimates were excluded.
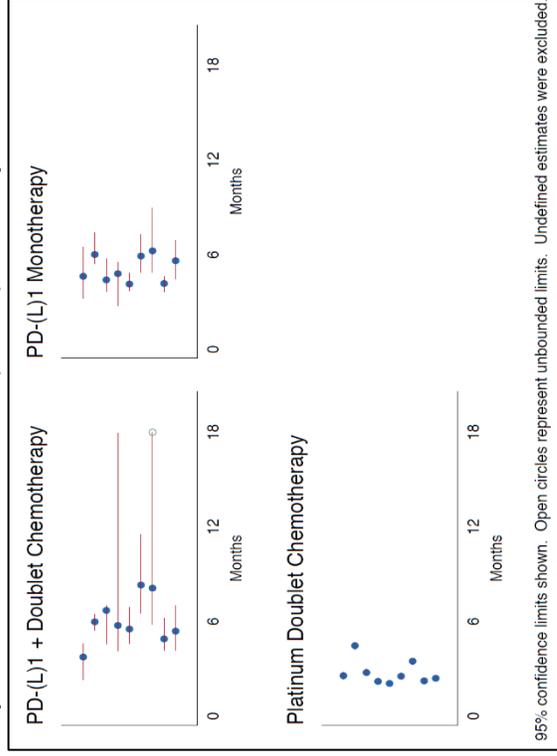
Graph 21. Estimates of Median (95% CI) rwTTNT by Treatment Category

95% confidence limits shown. Open circles represent unbounded limits. Undefined estimates were excluded.
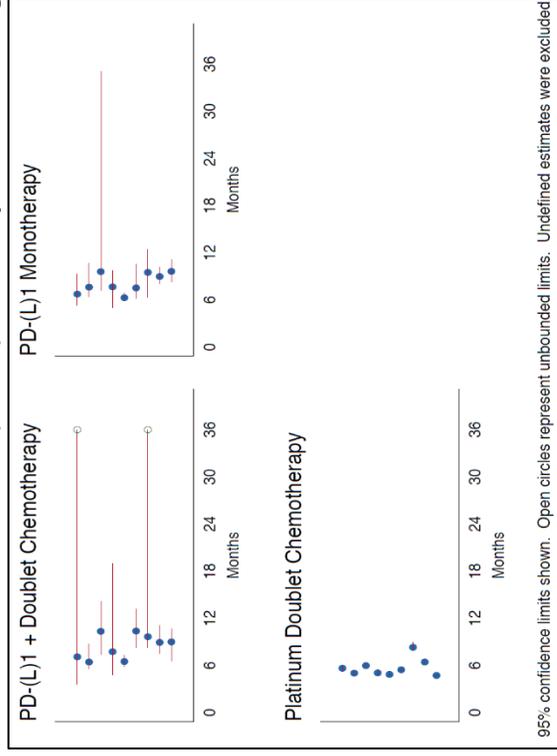
Discussion

Conclusions from Pilot Project Study

1. It is possible to coordinate the efforts across numerous real-world oncology data organizations to reach high-level alignment on important data elements and definitions for real-world endpoints in the context of a focused research question. As part of this collaboration, there was a shared understanding of the important considerations to take into account when identifying aNSCLC patients treated with frontline therapy across diverse RWD sources.

2. The depth of data varied across data providers and distinct characteristics were identified among the cohorts provided by each organization, likely attributable to the characteristics of the data source and the underlying population it is capturing. These differences may influence the measurable outcomes observed.

3. The results of this phase of the pilot project highlighted the ability to show differences in important prognostic demographic as well as clinical characteristics between trial patients and heterogenous real-world patient populations (e.g., median age, histology). It also demonstrated the ability to provide insight into recent trends in clinical care.

Assumptions and Limitations of Pilot Project Data Sets

- Preliminary findings are being presented today and subsequent analyses are planned.
- The observed unadjusted outcomes were evaluated in a broad set of patients with aNSCLC. In subsequent phases of the Pilot Project 2.0, it is important to apply relevant inclusion/exclusion criteria along with appropriate analytic methodologies to account for imbalances across critical prognostic variables.
- Discussions at the public meeting will also help identify additional action items.
- Interpretation of variable definitions may vary based on assumptions made in the conduct of analyses, even when using a common protocol and statistical analysis plan; a careful review and collaboration is needed to align on a consistent and reliable approach to be able to distinguish differences due to differences in the population characteristics, data source, and/or subtle differences in methodological assumptions made during the analyses.
- Granularity of certain variables within RWD may be limited because it is not always possible to distinguish between data that has not been captured in the data source versus data that is missing because the event never occurred.
- Cells with $\leq$N patients (N ranges from 5 to 11 depending upon data source) were masked to maintain patient privacy in compliance with each data source internal policies. Certain data reported in this document reflect this masking (Graphs 5 and 13).
- Verifying and determining date of death may also prove challenging. Although discharge status and some diagnosis codes may be a source of mortality information, it is often incomplete. Some data partners rely on external linkages, such as to the public Social Security Administration death master file (DMF), but the public DMF has been shown to under identify deaths

highlighting the need to understand the underlying quality of specific data elements[10]. Other data partners linked to additional data sources (linking EMR with billing and pharmacy claims) to apply a mortality algorithm and reduce potential loss to follow-up and confirm mortality/survival status.

- For claims-based data, some patients with advanced disease may enroll in clinical trials and some or all the care received in a clinical trial setting may not generate insurance claims, thus, data for these patients may not be fully captured or captured at all.
- Provider data (EHR) may not identify all chemotherapy as patients may seek care inside and outside a provider group that contributes to the EHR data (e.g., chemotherapy at an academic center then move to a community setting). This may or may not be a source of missing information in the aNSCLC setting. Some data partners linked EHR data with billing data to minimize this risk and improve capture of care outside of the clinic setting.
- Ability to distinguish proportion of different therapies used within each treatment group will impact outcomes observed in RWD.

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

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

2. How to overcome difficulties associated with inherent biases within RWE?

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

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

5. Can extractable endpoints from clinical trial "eligible" patients within EHR and claims databases be used to inform an internal assessment or sensitivity analysis of RWE?

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

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

---

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

Journal for ImmunoTherapy of Cancer

POSITION ARTICLE AND GUIDELINES   Open Access

The promise of Immuno-oncology: implications for defining the value of cancer treatment

Howard L. Kaufman[1], Michael B. Atkins[2], Prasun Subedi[3], James Wu[4], James Chambers[5], T. Joseph Mattingly II[6], Jonathan D. Campbell[7], Jeff Allen[8], Andrea E. Ferris[9], Richard L. Schilsky[10], Daniel Danielson[11], J. Leonard Lichtenfeld[12], Linda House[13] and Wendy K. D. Selig[14]*

Abstract

The rapid development of immuno-oncology (I-O) therapies for multiple types of cancer has transformed the cancer treatment landscape and brightened the long-term outlook for many patients with advanced cancer. Responding to ongoing efforts to generate value assessments for novel therapies, multiple stakeholders have been considering the question of "What makes I-O transformative?" Evaluating the distinct features and attributes of these therapies, and better characterizing how patients experience them, will inform such assessments. This paper defines ways in which treatment with I-O is different from other therapies. It also proposes key aspects and attributes of I-O therapies that should be considered in any assessment of their value and seeks to address evidence gaps in existing value frameworks given the unique properties of patient outcomes with I-O therapy. The paper concludes with a "data needs catalogue" (DNC) predicated on the belief that multiple key, unique elements that are necessary to fully characterize the value of I-O therapies are not routinely or robustly measured in current clinical practice or reimbursement databases and are infrequently captured in existing research studies. A better characterization of the benefit of I-O treatment will allow a more thorough assessment of its benefits and provide a template for the design and prioritization of future clinical trials and a roadmap for healthcare insurers to optimize coverage for patients with cancers eligible for I-O therapy.

Keywords: Immunotherapy, Immuno-oncology, Value, Patient experience, Patient reported outcomes (PROs)

Introduction: current clinical landscape

Compared with traditional cancer therapies, the approach described as Immuno-Oncology (I-O) therapy offers a more effective treatment alternative for some patients with cancer [1]. Rather than aiming treatments directly at the tumor, I-O therapies generally engage the immune system to recognize and eradicate tumor cells. Key features of immune-mediated therapy include specificity, breadth of response, and memory. These can contribute to complete tumor regressions, often providing more durable clinical outcomes and improved quality of life relative to cytotoxic chemotherapy, molecularly targeted therapeutics, and radiation, particularly in metastatic settings. The unique kinetics and properties of immunotherapy also result in different incidence and types of side effects, treatment length, and durability of response, as we describe in detail below. These differences need to be considered in studies of cost-effectiveness and value-based outcomes research, since I-O therapies are now approved by the U.S. Food and Drug Administration (FDA) in a variety of solid and hematologic malignancies, including melanoma, lung, kidney, bladder, head and neck, Merkel Cell, hepatocellular, certain gastrointestinal cancers, Hodgkin lymphoma, non-Hodgkin lymphoma, certain forms of leukemia, as well as in primary, site-agnostic tumors with Micro-Satellite-Instability High (MSI-Hi).

Most of the advances in I-O therapy to date have been demonstrated in patients with late-stage and metastatic cancer, but early results of adjuvant clinical trials using I-O therapies in patients with melanoma and lung cancer are promising. In addition, innovative approaches

* Correspondence: wendy@wscollaborative.com
[14]WSCollaborative, 934 Bellview Rd, McLean, VA, USA
Full list of author information is available at the end of the article

to patient selection, use of combinations, and sequencing of therapies lead to more patients benefitting from I-O therapy, expanding its potential impact. Typically, assessing impact of cancer therapeutics requires a minimum of five years follow-up to identify the benefit in overall survival. In melanoma, where I-O therapy has been available for the longest time, durable survival after I-O treatment has been confirmed [2].

There is urgent need to engage all stakeholders in maximizing I-O therapy's impact for current patients and those diagnosed in the near term. Optimal I-O therapy utilization will require clinically appropriate quality benchmarks and an understanding of its true clinical and economic value.

Immuno-oncology in the context of Cancer treatment

Until recently, the basic arsenal for treating cancer included surgery, radiation therapy, chemotherapy, and more recently, targeted therapy, sometimes in combination and often in sequence, to remove, reduce, eliminate or alleviate tumors. While these modalities often proved effective in producing durable remissions in patients with early, non-metastatic cancers, they generally failed to produce lasting benefit in patients with late-stage disease, except in certain leukemia, lymphomas, germ cell tumors and testicular carcinoma. Moreover, this multifaceted approach was often associated with serious negative consequences for patients, including disfigurement and a variety of treatment-related side effects caused by the total dose of radiation and the indiscriminate impact of cytotoxic agents on normal cells and physiologic functions.

Genomic studies conducted in the past two decades identified the molecular drivers of certain cancers and led to the advent of targeted therapies as an important additional pillar of the cancer therapy armamentarium. The current strategy generally follows a "one gene-one target" paradigm and is based on an assessment of specific gene mutations within an individual patient's cancer. This approach, however, has been associated with high rates of acquired drug resistance largely through cancer cell upregulation of bypass pathways to circumvent the block in driver pathways. Many driver mutations have proven challenging for drug targeting [3].

Immunologists have evaluated a variety of approaches designed to stimulate and enhance the immune system's response to tumors. The characterization of immune checkpoint pathways that can be targeted with immune-modulating antibodies energized a raft of drug development programs focused on inhibiting the effects of these immune checkpoints. The first of these immune checkpoint inhibitors, the anti-CTLA-4 antibody ipilimumab, was shown to produce durable survival in as many as 22% of patients with advanced melanoma,

leading to FDA approval in 2011. Subsequent studies with a variety of PD1/PDL1 antibodies led to regulatory approvals as single agents and in combination with either anti-CTLA-4 or other agents in more than a dozen cancer indications [4].

The most recent frontier has leveraged chimeric antigen receptor (CAR) T cells as a successful treatment modality for patients with hematologic malignancies [5] and is a modality under clinical investigation for use in patients with some solid tumors. Oncolytic virus therapy, in which a virus can be used to infect and kill cancer cells, received approval in 2015 by the FDA of the treatment of patients with unresectable recurrent melanoma [6].

Emboldened by this progress, many pre-clinical and clinical activities are underway to advance and exploit therapeutic options through either exploiting means of activating antitumor immunity or crippling other mechanisms for evading immune destruction either alone or in combination with checkpoint inhibitors.

What makes I-O therapy different? Scientific and clinical perspectives
Unique mechanisms of action

Cancer is basically a process of the patient's own cells dividing rapidly and failing to die normally. For a cancer to become established in a host, the transformed cells must also develop mechanisms to avoid eradication by the immune system. Therapeutic manipulation can activate the innate immune system leading to cell death and, under appropriate conditions, innate and adaptive immunity leading to oncolysis while promoting long-term memory responses.

I-O therapy involves a fundamentally different approach from conventional chemotherapy, which unleashes an indiscriminate, static, and toxic direct attack on all cells – malignant and normal -- in hopes of damaging the cancer cells more than the host cells. Recent studies have suggested that cytotoxic chemotherapy and targeted therapy may also target stromal cells and immune cells within the tumor microenvironment [7, 8]. These observation suggest the potential for combining chemotherapy with IO agents with the goal not of killing as many tumor cells as possible, but rather to optimize immunologic clearance, which may allow for lower chemotherapy dosing. Because immunotherapy for cancer primarily relies on an indirect approach rather than a direct attack on cancer cells, the observed kinetics of response related to I-O therapies can be delayed [9] and, at times, the tumor may appear to be growing in the near term, when in fact the observed increase in volume is instead related to an inflammatory immune response that is working to eliminate the cancer [4].

Significant and increased durability of response

An adaptive immune response is characterized by the ability to persist, creating "immune memory" that, once effectively triggered by an immunotherapy, can enable the body to maintain an ongoing defense against a threat like a virus or a cancer cell expressing specific antigens, even after therapy is discontinued and perhaps for the lifetime of the patient. I-O therapies can also evolve over time, broadening and deepening anti-tumor immunity, preventing the cancer's ability to escape through the selective growth of variants that can evade immune detection. The rapid co-evolution of tumor cells and immune responses may also result in immunoediting resulting in loss of antigen-specific immunity explaining, in part, IO drug resistance and the need to reconsider the pharmacologic drug class, dosing, schedule and combination to optimize anti-tumor activity [10].

Evidence of an effective and durable immune response against cancer dates back more than three decades, as high-dose interleukin 2 (IL-2) therapy produced durable responses with few relapses among approximately 10% of patients with advanced renal cell carcinoma (RCC) and melanoma [11]. These experiences demonstrated a unique hallmark of immunotherapy for the treatment of cancer: the flattening of the Kaplan Meier survival curve, in which a long, plateau of the curve represents durable responses that, for some patients, may extend throughout their lives.

With the advent of checkpoint inhibitors as single agents and in combination, dramatic results were first seen in patients with melanoma. The proportion of patients with metastatic melanoma experiencing objective responses increased to 20–22% with ipilimumab (anti-CTLA 4) treatment and 35–40% with anti-PD-1 agents, and above 50% with a combination approach [1].

Similarly, significant results with checkpoint inhibition approaches have yielded regulatory approvals of novel drugs and combination regimens, leading to new standards of care for patients with RCC, non-small cell lung cancer (NSCLC) [12], small cell lung cancer (SCLC) [13], bladder cancer [14], Merkel cell cancer [15], head and neck cancer [16], gastrointestinal cancer [17]and certain lymphomas [18]. Investigators are motivated by early success in identifying potential predictive biomarkers to select patients most likely to benefit (including programmed death ligand-1 or PDL1, and micro-satellite instability high or MSI-Hi), as checkpoint inhibition strategies are yielding even higher response rates in some tumors [19, 20]. Durable responses have also led to FDA approval of two CAR-T cell approaches for the treatment of acute lymphoblastic leukemia in children and young adults, and in certain forms of non-Hodgkin lymphoma in adults [21, 22].

Distinct side effect profiles

In general, immune checkpoint inhibitors have been associated with immune-related adverse events while CAR T cell treatment has been associated with cytokine release syndrome and neurologic toxicities. While serious adverse events are rare, mortality has been reported for patients receiving immune checkpoint inhibitors [23]. Nevertheless, I-O treatment has been suggested to have less impact on patients' quality of life than conventional therapies [24], especially when adverse events are expeditiously managed early with corticosteroids and other immunosuppressive agents [25]. This parallels the experience with CAR-T research, in which cytokine release syndrome (CRS) was identified as an early potentially lethal clinical syndrome [26, 27], but an effective clinical management strategy was quickly identified [26], which did not appear to interfere with efficacy [28, 29], and actually led to a concomitant FDA-approved indication expansion for the anti-IL-6 monoclonal antibody, tocilizumab, since IL-6 is believed to be a major cytokine released in patients experiencing IO-induced cytokine release syndrome [30]. Research is ongoing to better define the most serious immune-related adverse events and identify patient characteristics most likely associated with them (recognizing that patient cohorts in most pre-approval studies did not fully reflect the general population) [31, 32].

What makes I-O therapy different? Patient experience perspective

Many thousands of patients have been treated with immunotherapies in clinical trials and more recently, as standard of care. A holistic narrative is emerging about the patient experience with these novel therapies, providing important insights about how patients and caregivers perceive the value of these treatments. Patients often describe their experience with I-O agents in broader terms than the clinical outcome measures usually used in a trial. In addition to considering traditional effectiveness and safety measures like response rates, overall survival, and side effects, patients focus on the potential for limited treatment period duration, durability of response, the possibility of being "cured," a more manageable side effect profile, and a better overall quality of life. Evaluating these aspects can provide important context and completeness for assessing the value of these therapies.

Limited treatment period duration: treatment free survival

Because I-O therapies act on the immune system, they may be effective if administered for a shorter period. As a result, many I-O treated patients experience significant "Treatment-free survival" (TFS), the period that occurs after treatment ends, and while the impact of the

therapy endures, patients may not require other treatment(s) [33]. TFS provides an important opportunity for patients and their families to resume routine activities, travel, and generally approach their daily lives free from ongoing cancer treatment [34].

Effectiveness of therapy: return to productivity

There may also be financial benefits to individual patients, their families, and society that result from patients being able to return to work earlier and for longer periods of time while also reducing the need for additional or subsequent cancer treatments and perhaps less frequent medical tests and interventions. When effective, I-O treatment should boost productivity for many patients and may save individuals, families, and society considerable expenditures throughout the rest of their lives.

Impact of Treatment & Possibility of "cure"

While there is risk of serious toxicities associated with current I-O regimens, I-O therapies generally do not lead to the side effects commonly associated with cytotoxic chemotherapy such as nausea/vomiting, hair loss, and risks to fertility. In fact, the knowledge about 1) what side effects are likely to occur from I-O therapies and 2) that most can be managed in the near-term (by experiences providers) without impacting the effect of the cancer treatment, adds to patients' current willingness to try them – especially when faced with few other potentially curative treatment options.

Late-stage patients facing the possibility of dying from their cancer often value the opportunity to pursue a hopeful gamble and receive a novel therapy that offers the potential for long-term disease control in a small percentage of patients rather than a treatment that offers potential benefit to a higher proportion of patients but for a shorter duration. Further, patients often will place a higher value overall on survival than their clinicians, who typically focus more on progression-free survival and managing patient's treatment and disease related symptoms [35]. Of course, there are other factors at play in determining whether a patient has access to such hopeful gamble therapies, e.g. geographic access to healthcare provider expertise in IO delivery, drug availability, negative reimbursement incentives, high out-of-pocket expenses and others, raising important issues for society that are beyond the scope of this paper.

Reports of significant positive outcomes with I-O therapy for an increasing number of tumor types have fueled hope among patients for long-term survivorship and even cure in some cancers. This type of hope – especially for patients with dismal prognoses -- has been recognized to provide positive benefits to the patient's quality of life [36] and is a powerful incentive for patients to seek access to these therapies, even while recognizing the longer odds of success. There is active debate within the oncology community about if and when to try immunotherapies when patients have few other valid options, even though the evidence is not yet conclusive about the potential benefit [37]. This may be especially important for patients with orphan cancers where clinical trials are lacking and where few approved agents are available.

Assessing the value of I-O therapies

Economists frequently use the Incremental Cost Effectiveness Ratio (ICER) to assess and compare value in healthcare among available treatment options. ICERs are calculated by measuring or estimating the incremental costs and improvements in patient outcomes versus a therapeutic comparator through cost-effectiveness and cost utility models. The ICER measure is designed to be standardized across diseases. Health care payers often use the ICER to assess whether the improvements in patient health are worth the extra costs for one treatment versus another. For some, the ICER addresses an efficiency question, which can be helpful in a constrained resource environment. There are divergent views about the utility of the ICER measure in capturing value, especially given limitations in its ability to assess patient perspectives.

Currently, economic models are based on the metrics reported in the medical literature and are complicated by statistical uncertainty. These metrics generally describe treatment effects and adverse events reported in pivotal trials necessary to gain marketing approval by various national regulatory bodies, such as the FDA. While these metrics have rarely included patient-centered outcomes, the FDA has recently implemented a Patient-Focused Drug Development (PFDD) program to attempt to incorporate patient experience metrics into the regulatory pathway [38]. In the meantime, such outcomes are generally compiled during late stage development, especially for products that have gone through an accelerated approval.

The ability of current economic models to estimate ICERs is tied to the robustness of the data that are used to create the model itself. In oncology, economic modeling is challenging, in part because:

- Disease mechanisms vary by tumor type, genetic alteration, and location, that suggest heterogeneity of effect;
- Trial data are limited due to small study populations and relatively short follow-up; and
- Therapeutic effects of the therapy under investigation may be impacted by previous therapies a patient may have received.

These factors increase the uncertainty of economic model outputs and therefore negatively impact their

capacity to precisely measure value in oncology. Various health technology assessment (HTA) bodies attempt to compensate for special cases such as disease severity, rare diseases, or end of life therapies, by adopting a lower ICER threshold by which 'value' is judged [39]. Others maintain the ICER threshold, evaluating all drugs against a common standard.

The definition of 'value' varies among stakeholders. For instance, patients and caregivers mostly overlap in how they define value, but subtle differences often exist between how patients differentially value returning to work or the impact of regaining their activities of daily living. Similarly, subtle but meaningful differences exist among how physicians, researchers, payers and employer groups define 'value.' In addition, the views of other stakeholders, such as drug developers, patients' employers and family members are often not considered in the value assessment.

Within oncology, and specifically I-O, the assessment of value is made that much more difficult due to the principal impact of the therapy on landmark OS and the height of the plateau on the OS curve, rather than median PFS or OS, small numbers of patients assessed, and lack of long-term follow-up. These elements compound the uncertainty normally found within economic models [40].

Cost-effectiveness analysis (CEA) is an important tool when weighing the value of certain treatments using a common measure of health benefit. However, CEA is limited when accounting for other important aspects of 'value' to patients and may be misleading when long-term follow-up data on critical endpoints, such as overall survival, are not available. While these other aspects of value are arguably less important to decision makers allocating resources from a fixed budget, they should be accounted for when assessing value to patients and making decisions that may affect patient access.

Existing value frameworks and tools

Traditional clinical outcome measures, or clinical outcome assessments (COAs), in trials include overall survival (OS), progression-free survival (PFS) and objective response rate (ORR). These have long proved to be useful measures for assessment of cytotoxic chemotherapy, but a more complete assessment of the value of I-O requires identifying and measuring the impact of I-O therapy on patient's lives. Some I-O therapy studies have shown significant improvement in overall survival without any impact on PFS, making the use of OS surrogates problematic in value frameworks that are not accounting for the potential differences in endpoint analyses.

A recent review by the ISPOR (International Society for Pharmacoeconomics and Outcomes Research) Special Task Force on US Value Frameworks has identified multiple value frameworks in the U.S. [41] In Europe, where HTA bodies are much more prevalent, there is less need for discrete value frameworks, but the European Society for Medical Oncology (ESMO) has created one based on "magnitude of clinical benefit." [42] Others strive to be more patient-centered, emphasizing the patient experience [43]. In addition to understanding how each framework defines 'value', it is also important to consider that those designed by clinically-oriented bodies are meant to inform clinician-patient decisions, while those geared for payers are meant to inform payer and pharmacy benefit manager decision-making around coverage or formulary tiering.

The report of the Second Panel on Cost-Effectiveness in Health and Medicine (Second Panel) has defined four normative perspectives for consideration in evaluating value: 1) the payer perspective; 2) the health care sector perspective; 3) the health care sector with time cost perspective; and 4) societal perspective [44]. While each is scientifically valid and informative for their respective decision makers, the Second Panel recommended that analyses should include "reference cases" from the health care sector perspective and the societal perspective, which could be helpful in understanding how the value assessment informs a comparison within the therapeutic class or across therapeutic classes. Some stakeholders have noted a shortcoming in the Second Panel's work, noting that it did not specifically call out patient perspectives in its report [45].

While some observers have criticized the recent value frameworks [46, 47], those meant to inform clinician-patient decisions do have elements of patient preference included in them, which may make the 'value' resulting from them reflective of an individualized assessment, and possibly then fit for informing individualized clinician-patient decisions. More payer-centric value frameworks also include elements of patient preferences but given the goal of informing population-level decision making, such value estimates are conducted at the average of a population. Thus, heterogeneity in individual patient preferences are often lost in these population-geared exercises.

Identifying shortcomings of traditional metrics in assessing I-O value
Clinical efficacy measures for I-O
Because of the mechanistic differences between I-O therapies and traditional chemotherapy, conventional trial designs and endpoints generally do not fully capture the novel patterns of treatment response. This unique aspect of I-O suggests that longer-term assessment at multiple timepoints is needed to adequately evaluate outcomes [48]. Traditional parametric survival models

used commonly to estimate long-term survival cannot adequately represent complex hazard functions and may not be appropriate for modelling the underlying mechanism of action associated with I-O treatments [49].

Recent work reported by the ISPOR Special Task Force in rare pediatric diseases presents some of the unique challenges in selection of clinical outcome assessments (COAs), and highlight the importance of developing uniform methods and metrics to capture relevant outcomes of interest for the I-O setting [50]. Additionally, recent work from the ISPOR Rare Disease Special Interest Group has identified several key challenges to research in rare diseases, which may be particularly relevant for I-O, and result in a lack of tailored health technology methods for rare disease treatments, as well as significant uncertainty for HTA authorities [51]. Many of the factors result from the evolving evidence base, including difficulties in establishing specific and sensitive diagnostic criteria, and evaluating the treatment effect (or heterogeneity of treatment effect). Combined with ethical challenges in designing appropriate clinical trials, insufficient knowledge of the natural history of the disease, and often poor patient recruitment for trials, the result is high levels of uncertainty in assessing value for these therapies. These uncertainties are factored into health technology assessments by global authorities, as comprising the level of certainty that is generally attributed to the value of a product. In addition, the model structure may not reflect the full patient experience, often failing to assess the value of treatment-free survival.

Safety assessments for I-O

While the long-term clinical and economic impact of safety monitoring with I-O therapy is not defined, current practice suggests that limited baseline screening and on-going laboratory monitoring with detailed clinical surveillance and patient education can identify adverse events early, allowing rapid intervention [52]. Whether this results in better compliance with planned treatment duration or prevents chronic toxicity is unknown. The optimal duration of treatment with I-O has also recently undergone considerable debate and discussion with some clinicians suggesting that early drug discontinuation may be possible without increasing rates of tumor progression [53].

An improved understanding of tumor immunology has led to new combination treatments, although it is unclear whether concurrent or sequential administration impacts outcomes. Further studies will focus on better defining effective combination regimens, treatment schedules and duration of therapy while refining safety monitoring measures that will allow appropriate patient management while limiting unnecessary diagnostic work-ups. These advances

in limiting and mitigating toxicities should provide additional I-O relevant evidence to support better value assessments. There is a need for long-term follow up via accurate registries, capturing patient outcomes in community settings as well as academic medical centers.

PRO measures for I-O

One of limitations of reliance on the QALY within certain value frameworks is its primary dependence on survival endpoints (or improvements in OS and/or PFS) in determining the incremental cost per QALY gained for interventions that have OS and/or PFS primary endpoints in clinical trials [54]. Indeed, the ISPOR Special Task Force on Value Frameworks echoed the recommendation that cost-effectiveness analysis "as measured by cost per QALY [should serve] as a *starting point* to inform payer and policy maker deliberations" [55]. A natural question arises as to whether or not the QALY can be a comprehensive estimate of health outcome for the purposes of characterizing I-O therapies. Some cases of incremental cost-per-QALYs for I-O therapies suggest good value for money [56]. However, the question remains as to whether QALYs are sufficiently comprehensive to address the unique long-term outcomes for I-O, especially when compared to more traditional chemotherapy and targeted therapy regimens.

There is increasing interest in 'going beyond QALYs', to measure and systematically incorporate patient reported outcomes (PRO) in oncology [57–59], as there are signals (from markets outside the U.S.) that surrogate endpoints like PFS may not be closely associated with improvements in health-related quality of life in oncology clinical trials [60], or that current health-related quality of life instruments lack uniformity when applied across therapeutic areas [61]. While various work has suggested how to set standards for PRO use for cancer clinical trials with international standards [62], or in clinical trial protocols [63], there is more to be done before this work is ready for inclusion in value assessments. In fact, a recent FDA analysis has noted that health-related quality of life components most impacted by anti-PD-1/PD-L1 therapies (including disease symptoms, symptomatic toxicity and physical function) have been 'variable,' but that "these data, along with other important clinical data such as hospitalizations, ER visits and supportive care medications can help inform the benefit risk assessment for regulatory purposes." [64]

In the U.S., the Centers for Medicare and Medicaid Services (CMS) has recently opened a National Coverage Determination (NCD) for Chimeric Antigen Receptor T-cell (CAR-T) Therapy for Cancers [65] and has focused on the PRO instruments themselves, and whether sufficient scientific evidence exists to support application of PROs to health outcomes research [66]. Presentations by the FDA and PRO experts provided optimism for

several of the PRO instruments [67], and a final recommendation from the MEDCAC in the form of a proposed Decision Memo is expected in 2019 [68].

There is increasing interest in incorporating more patient centric elements in value assessments, especially as recent evidence appears to suggest an OS improvement among metastatic cancer patients who had PROs integrated into their routine care, compared to usual care [69]. While Basch had previously pointed out the lack of PRO data in existing value frameworks [70], he also argues for greater uniformity in how the PROs are incorporated into the value assessment for CAR-T cell therapies and to include patient representatives in consensus processes. While there seems to be increasing use of validated PRO instruments in oncology clinical trials, there are challenges to incorporating the PRO measures into existing value frameworks [71].

It is also challenging to weigh the different trade-offs between therapies in a class and the added layer of complexity associated with evaluating combination therapies. Likewise, there is the challenge of distinguishing between novel I-O therapies and their chemotherapy comparators, with the concept of treatment-free survival raising additional questions for researchers to address. An emphasis on integrating data collection regarding both PRO and quality of life (QOL) into modern I-O clinical trials will be important to developing benchmark metrics for understanding the impact of these measures related to specific drug agents and tumor types. The development of benchmark data will also provide a basis for comparisons to patient outcome data with more traditional cancer therapeutics.

Recommendations for framework to develop value metrics for I-O: Data Needs Catalogue

This paper recommends the generation and synthesis [72] of evidence that will enable patients, health care providers, payers, and other stakeholders to make informed value-based decisions about I-O therapies (see Table 1). In addition to the clinical trials used for regulatory approval, more studies performed in real-world settings, e.g., pragmatic clinical trials, patient registries, health surveys, and administrative claims studies [73], would provide decision makers with a better understanding of the cost and benefits of treatments in the real world. As new data are generated, researchers must simultaneously work to incorporate them into value assessments.

Develop better evidence, especially post-market

Post-market research is important to our understanding of the costs and real-world effectiveness of novel therapy approaches post launch. An important aspect of measuring real-world effectiveness is comparison of available treatment options in real world populations (i.e. comparative effectiveness). Thus, careful consideration for study design is needed not only to collect important elements of value but also to ensure that observed signals can be attributable to the I-O therapy.

Incorporate additional evidence into value assessments /modeling considerations

While evidence to support costs and real-world effectiveness estimates improves, researchers should advance models that support informed decisions. This may include, but is not limited to, increased modeling transparency [74], clearly outlining data and underlying assumptions used for calculations [75, 76], consensus on value elements [77] to incorporate into individual assessments, and continuous patient engagement [78, 79] throughout the process to ensure a patient-centric approach.

We recommend a concerted effort to develop models for looking beyond the median and conducting appropriate pre-planned sub-group analysis of the

Table 1 Assessment of conventional value metrics in evaluating I-O therapies

Conventional value metric (examples)	Why insufficent For I-O	Areas where new I-O value measures are needed (beyond QALY)
Clinical Efficacy Assessment OS, PFS, ORR	I-O therapies offer potential for durable response and due to delayed kinetics may not demonstrate early ORR or improvements in PFS	Milestone Survival; Treatment Free Survival
Safety Assessment	Late-stage cancer patients may be more willing to accept high risk of toxicity for possible benefit (durable response); long-term impact of adverse events not fully known	More nuanced evaluation of patient preferences based on their risk tolerance and profile; longer follow up studies post treatment
Patient Reported Outcome	Current measures fall short in measuring the value to patients of Treatment-free Survival; (extended time off treatment)	Treatment Free Survival impact on patient's QoL; Hope for durable response
Economic Measures, e.g. Cost of ongoing treatment; Cost of treatment for side effects; cost of lost productivity	Typically focuses on patient-related expenses or drug cost during active treatment	Return to productivity; Economic benefit of Treatment-Free Survival, including reduced expenditures on ongoing treatment, scans and other follow up; Amortize costs over the longer horizon of benefit in a "cure-rate" model; consider other stakeholder fiscal impact

patients who see long-term benefit (e.g. "the Tail of the Curve" phenomenon, which within oncology, is seen by clinicians and patients as a defining hallmark of I-O). Table 2 describes considerations for such I-O specific elements to enhance a traditional ICER calculation.

Future strategies for I-O analyses

While the field of I-O has advanced significantly in the past several decades, much more knowledge is needed to achieve a future where the potential benefit of these therapies can be maximized for the greatest number of patients. Key questions remain about how to select those patients who are most likely to respond to I-O therapy, how to combine I-O therapies with one another and with other treatment modalities, how to predict limit and mitigate I-O treatment related toxicities, how to reduce resistance to I-O therapies, how to use these therapies in newly defined standards of care and when to stop treatment.

Answers to these important questions – and addressing the important questions surrounding access to these therapies -- will help define and realize a promising vision for the future of cancer treatment, one that maximizes the potential of I-O therapy and further enhances its value to patients, their families, and society.

We envision a time when:

- Many more cancer patients will receive some form of I-O therapy during their treatment journey;
- We leverage patient reported outcomes, real world evidence and other tools to expand the knowledge base and continuously improve patient outcomes from I-O therapies;
- Careful patient selection ensures that treatments are provided only to those patients most likely to benefit;
- The numbers and cancer profiles of patients who are likely to benefit has expanded;
- Potential resistance to I-O therapy is reduced and we succeed in turning previously non-immunogenic cancers into ones that can respond to I-O therapy;
- The benefits are established for I-O therapy in the adjuvant and neo-adjuvant settings, thereby reducing the incidence of late-stage cancers; and
- Cancer can become a treatable and even curable set of diseases [80] with combination approaches that include I-O leading to maximized therapeutic equations for every cancer and a resulting favorable economic impact for patients, their families and society.

Table 2 I-O specific elements to enhance traditional value calculations

Costs (numerator considerations)	Net Prices vs. List Prices	Wholesale acquisition costs may significantly overestimate the true cost of a drug. We recommend accounting for discounts and rebates where appropriate to reflect the true price paid for the new therapy.
	Consider alternate stakeholder perspectives	More research emphasis on a societal perspective – While many payers require a focus on the health sector specific costs, to fully understand the costs and benefits of a drug to society taking a societal perspective (accounting for caregiver costs, productivity gains/losses, etc.) in costeffectiveness analysis is warranted.
Effects (denominator considerations)	QALY	Many economic models are sensitive to the variations of the utility value used for each health state. We recommend engaging current or former patients as advisors to validate the assumptions made with the base case QALY inputs as well as the sensitivity analysis.
	Life Years	Conduct the same analysis with no QALY adjustment so that absolute mortality reductions can be easily reported for the decision-maker.
	Patient Specific	Identify other potential outcomes as denominators by engaging current and former patients. Addressing the outcomes that "matter" to patients can help decision-makers compare drugs within the same disease state for the specific population that it is impacting. Consider stratifying analyses based on risk tolerance of patient subpopulations.
Other factors (beyond the incremental cost effectiveness ratio)	Value of Hope	The ISPOR Special Task Force identifies this as an area needing more research to quantify, but it is conceptually intuitive and very relevant to IO. A cancer patient facing a terminal diagnosis may be willing to risk taking a more novel therapy if his or her chances include the possibility of durable response and even functional cure.
	Real Option Value	For a cancer patient, any innovation that can extend life (even at the same or worse quality of life) may give a patient a chance to live long enough for a new treatment to develop, possibly even a cure.
	Scientific Spillovers	New mechanisms of action may or may not benefit current patients, but we often fail to consider the steps in the path to future discovery. Without learning from the research in the 1950s, would we be here today with ~ 26 IO regimens benefitting thousands of patients?

Availability of data and materials
Not applicable.

Authors' contributions
WKDS developed the overall concept for this paper, developed the initial manuscript draft, incorporated all author comments and edits throughout multiple versions, and completed the final draft for submission. HLK and MBA helped develop the overall concept for this paper, contributed to the initial manuscript draft and provided additional edits and final approval. PS and JW developed the concept for the Data Needs Catalogue section, provided significant expert input into the construct of the entire manuscript and reviewed edits for inclusion and provided final approval. JM, JC, JC, and DD provided expert input and content for the Data Needs Catalogue, contributed to the initial manuscript draft and provided additional edits and final approval. RS, JLL, JA, LH, and AEF contributed to the initial manuscript draft and provided additional edits and final approval.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
H.L. Kaufman is an employee of Replimune, Inc.
W.K.D. Selig is owner of WSCollaborative, LLLC.
R. Schilsky receives research funding from Astra-Zeneca, Bayer, Bristol Myers Squibb, Genentech, Lilly, Merck and Pfizer.
P. Subedi is an employee of Pfizer.
J. Wu is an employee of Amgen.
D. Danielson was an employee of Premera Blue Cross Blue Shield.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Society for Immunotherapy of Cancer (SITC) Policy Committee, Replimune, Inc, 18 Commerce Way, Woburn, MA, USA. [2]Georgetown University, 3970 Reservoir Road NW, Washington, D.C, USA. [3]Pfizer, Inc, 235 E 42nd St, New York, NY, USA. [4]Amgen, Inc, One Amgen Center Drive, Thousand Oaks, CA, USA. [5]Tufts University, 145 Harrison Ave, Boston, MA, USA. [6]University of Maryland, 20 N. Pine St, Baltimore, MD, USA. [7]University of Colorado, 12850 E. Montview Blvd, Aurora, CO, USA. [8]Friends of Cancer Research, 1800 M St. NW, Washington, DC, USA. [9]LUNGevity Foundation, 228 S Wabash Ave, Chicago, IL, USA. [10]ASCO, 2318 Mill Rd, Alexandria, VA, USA. [11]Premera Blue Cross, 7001 220th St. SW, Mountlake Terrace, WA, USA. [12]American Cancer Society, 250 Williams St, NW, Atlanta, GA, USA. [13]Cancer Support Community, 734 15th St, NW, Washington, DC, USA. [14]WSCollaborative, 934 Bellview Rd, McLean, VA, USA.

References
1. Kaufman HL, Atkins MB, Dicker AP, Jim HS, Garrison LP, Herbst RS, et al. The Value of Cancer Immunotherapy Summit at the 2016 Society for Immunotherapy of Cancer 31st anniversary annual meeting. Journal for ImmunoTherapy of Cancer. 2017;5(1).
2. Ascierto PA, Long GV, Robert C, Brady B, Dutriaux C, Giacomo AMD, et al. Survival outcomes in patients with previously untreated BRAF wild-type advanced melanoma treated with Nivolumab therapy: three-year follow-up of a randomized phase 3 trial. JAMA. 2018.
3. Kakadia S, Yarlagadda N, Awad R, Kundranda M, Niu J, Naraev B, et al. Mechanisms of resistance to BRAF and MEK inhibitors and clinical update of US Food and Drug Administration-approved targeted therapy in advanced melanoma. OncoTargets and Therapy. 2018;Volume 11:7095–107.
4. Ventola CL. Cancer Immunotherapy, Part 2: Efficacy, Safety, and Other Clinical Considerations. Pharmacy and Therapeutics. 2017;42(7):452–63.
5. June CH, Sadelain M. Chimeric Antigen Receptor Therapy. New England Journal of Meicine. 2018;379(1):64–73.
6. Andtbacka RH, Kaufman HL, Collichio F, Amatruda T, Senzer N, Chesney J, et al. Talimogene Laherparepvec improves durable response rate in patients with advanced melanoma. Journal of Clinical Oncology: Current neurology and neuroscience reports; 2015.
7. Fabbri R, Macciocca M, Vicenti R, et al. Doxorubicin and cisplatin induce apoptosis in ovarian stromal cells obtained from cryopreserved human ovarian tissue. Future Oncol. 2016;12 (14):1699–1711.
8. Okkenhaug K, Graupera M and Vanhaeesbroceck B. Targeting Pi#K in cancer: Impact on tumor cells, their protective stroma, angiogenesis, and immunotherapy. Cancer Discov. 2016;6(10):1090–1105.
9. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with Ipilimumab in patients with metastatic melanoma | NEJM. N Engl J Med. 2010.
10. Dharmaraj N, Piotrowski SL, Huang C, Newton JM, Golfman LS, Hanoteau A, Koshy ST, Li AW, Pulikkathara MX, Zhang B, Burks JK, Mooney DJ, Lei YL, Sikora AG, Young S. Anti-tumor immunity induced by ectopic expression of viral antigens is transient and limited by immune escape. OncoImmunology 2019;8(4):e1568809.
11. Dutcher JP. Current status of interleukin-2 therapy for metastatic renal cell carcinoma and metastatic melanoma. Current neurology and neuroscience reports. U.S. National Library of Medicine; 2002 [cited 2019 Jan 13].
12. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2015;373(2):123–35.
13. Yu D-P, Cheng X, Liu Z-D, Xu S-F. Comparative beneficiary effects of immunotherapy against chemotherapy in patients with advanced NSCLC: meta-analysis and systematic review. Oncol Lett. 2017;14(2):1568–80.
14. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet Oncology. 2016;387(10031).
15. Kaufman HL, Russell J, Hamid O, Bhatia S, Terheyden P, Dangelo SP, et al. Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. Lancet Oncol. 2016;17(10):1374–85.
16. Seiwert TY, Burtness B, Mehra R, Weiss J, Berger R, Eder JP, et al. Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. Lancet Oncol. 2016; 17(7):956–65.
17. Moehler M, Delic M, Goepfert K, Aust D, Grabsch HI, Halama N, et al. Immunotherapy in gastrointestinal cancer: recent results, current studies and future perspectives. Eur J Cancer. 2016 May;59:160–70.
18. Savage KJ, Steidl C. Immune checkpoint inhibitors in Hodgkin and non-Hodgkin lymphoma: how they work and when to use them. Expert Rev Hematol. 2016;9(11):1007–9.
19. Hellmann MD, Gettinger SN, Goldman JW, Brahmer JR, Borghaei H, Chow LQ, et al. CheckMate 012: Safety and efficacy of first-line (1L) nivolumab (nivo; N) and ipilimumab (ipi; I) in advanced (adv) NSCLC. Journal of Clinical Oncology. 2016;34(15_suppl):3001.
20. Diaz LA, Marabelle A, Delord J-P, Shapira-Frommer R, Geva R, Peled N, Kim TW, Andre T, Cutsem EV, Guimbaud R, Jaeger D, Elez E, Yoshino T, Joe AK, Lam B, Gause CK, Pruitt SK, Kang SP, Le DT. Pembrolizumab therapy for microsatellite instability high (MSI-H) colorectal cancer (CRC) and non-CRC. J. Clin. Oncol. 2017;35(15_suppl):3071.
21. Neelapu SS, Locke FL, Bartlett NL, Lekakis LJ, Miklos DB, Jacobson CA, et al. Axicabtagene Ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. New England Journal of Medicine. 2017;377(26):2531–44.
22. Schuster SJ, Svoboda J, Chong EA, Nasta SD, Mato AR, Anak O, Brogdon JL, Pruteanu-Malinici I, Bhoj V, Landsburg D, Wasik M, Levine BL, Lacey SF,

Melenhorst JJ, Porter DL, June CH. Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas. N. Engl. J. Med. 2017;377(26):2545–54.

23. Wang DY, Salem JE, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. JAMA. 2018;4(12):1721–8.

24. Ciren B, Wang X, Long Z. The evaluation of immunotherapy and chemotherapy treatment on melanoma: a network meta-analysis. Oncotarget. 2016;7(49).

25. Kroschinsky F, Stölzel F, Bonin SV, Beutel G, Kochanek M, Kiehl M, et al. New drugs, new toxicities: severe side effects of modern targeted and immunotherapy of cancer and their management. Critical Care. 2017;21(1).

26. Davila ML, Riviere I, Wang X, Bartido S, Park J, Curran K, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. Sci Transl Med. 2014;6(224):224ra25.

27. Lee DW, Gardner R, Porter DL, Louis CU, Ahmed N, Jensen M, Grupp SA, Mackall CL. Current concepts in the diagnosis and management of cytokine release syndrome. Blood 2014;124(2):188–195.

28. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor–modified T cells for acute lymphoid leukemia. New England J Med. 2013;368:1509–18.

29. Park JH, Rivière I, Gonen M, Wang X, Sénéchal B, Curran KJ, Sauter C, Wang Y, Santomasso B, Mead E, Roshal M, Maslak P, Davila M, Brentjens RJ, Sadelain M. Long-Term Follow-up of CD19 CAR Therapy in Acute Lymphoblastic Leukemia. N. Engl. J. Med. 2018;378(5):449–459.

30. Center for Drug Evaluation and Research. Approved Drugs - FDA approves tisagenlecleucel for B-cell ALL and tocilizumab for cytokine release syndrome [Internet]. U S Food and Drug Administration Home Page. Center for Drug Evaluation and Research; 2017 [cited 2018 Nov 17]. Available from: https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm574154.htm.

31. Adverse effects from certain immunotherapies may be more common than initially reported. Healio. [cited 2019 Jan 7]. https://www.healio.com/hematology-oncology/lung-cancer/news/online/45aa720b-8739-4ada-bf35-9c7e9e75553c/adverse-effects-from-certain-immunotherapies-may-be-more-common-than-initially-reported .

32. Johnson DB, Sullivan RJ, Menzies AM. Immune checkpoint inhibitors in challenging populations. Cancer. 2017;123(11):1904–11.

33. Melanoma Research Foundation. Advances in Melanoma Oncology - Mike Atkins, MD LinkedIn SlideShare. 2017 [cited 2019 Jan13.

34. Howard L. Kaufman, Michael B. Atkins, Adam P. Dicker, Heather S. Jim, Louis P. Garrison, Roy S. Herbst, et al. The Value of Cancer Immunotherapy Summit at the 2016 Society for Immunotherapy of Cancer 31st Anniversary Annual Meeting. Journal for ImmunoTherapy of Cancer. BioMed Central; 2017 [cited 2019Jan7]. https://jitc.biomedcentral.com/articles/10.1186/s40425-017-0241-6

35. Addario BJ, Fadich A, Fox J, Krebs L, Maskens D, Oliver K, et al. Patient value: Perspectives from the advocacy community. Health Expect. 2018;21(1):57–63.

36. Azizi M, Elyasi F. Effective strategies for increasing the hope in patients with cancer: a review. Journal of Cellular Immunotherapy. 2017;3(1):14–5.

37. Kolata G. 'Desperation Oncology': When Patients Are Dying, Some Cancer Doctors Turn to Immunotherapy. The New York Times. The New York Times; 2018 [cited 2019Jan7]. https://www.nytimes.com/2018/04/26/health/doctors-cancer-immunotherapy.html?rref=collection/sectioncollection/health&action=click&contentCollection=health%ion=rank&module=package&version=highlights&contentPlacement=5&pgtype=sectionfront

38. Office of the Commissioner. Oncology Center of Excellence - Patient-Focused Drug Development. U S Food and Drug Administration Home Page. Center for Drug Evaluation and Research; 2018 [cited 2018Nov17]. https://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/OCE/ucm544143.htm

39. Detiček A, Locatelli I, Kos M. Patient access to medicines for rare diseases in European countries. Value Health. 2018;21:553–60.

40. Diaby V, Ali AA, Adunlin G, Kohn CG, Montero AJ. Parameterization of a disease progression simulation model for sequentially treated metastatic human epidermal growth factor receptor 2 positive breast cancer patients. Current Medical Research and Opinion. 2016;32(6):991–6.

41. Willke RJ, Neumann PJ, Garrison LP Jr, Ramsey SD. Review of recent US value frameworks – a health economics approach: an ISPOR special task force report. Value Health. 2018;21(2):155–60.

42. Cherny NI, Sullivan R, Dafni U. A standardized, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society of Medical Oncology Magnitude of clinical benefit scale (ESMO-MCBS). Ann Onc. 2015;26:1547–73.

43. Patient Pe rspective Value Framework. FasterCures. [2018Nov17]. https://www.fastercures.org/programs/patients-count/patient-perspective-value-framework/Open-Source Value Project. Innovation and Value Initiative. [cited 2018Nov17]. https://www.thevalueinitiative.org/open-source-value-project/

44. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost effectiveness in health and medicine. New York: Oxford University Press; 2016.

45. Perfetto EM. ISPOR's Initiative on US Value Assessment Frameworks: A Missed Opportunity for ISPOR and Patients. Value in Health. 2018;21(2):169–70.

46. Cohen JT, Anderson JE, Neumann PH. Three sets of case studies suggest logic and consistency challenges with value frameworks. Value in Health. 2017;20:193–9.

47. Bentley TGK, Cohen JT, Elkin EB, Huynh J, Mukherjea A, Neville TH, Mei M, Copher R, Knoth R, Popescu I, Lee J, Zambrano JM, Broder MS. Validity and Reliability of Value Assessment Frameworks for New Cancer Drugs. Value in Health. 2017;20(2):200–205.

48. Gemmen E, Parmenter L. Special considerations for the analysis of patient-level Immuno-oncology data. Value & Outcomes Spotlight. 2018:23–4.

49. Huang M, Latimer N, Zhang Y, et al. Estimating the Long-term outcomes associated with Immuno-oncology therapies: challenges and approaches for overall survival extrapolations. Value & Outcomes Spotlight. 2018:28–30.

50. Phillips D, Leiro B. Clinical outcome assessments: use of normative data in a pediatric rare disease. Value Health. 2018;21:S08–34.

51. Nestler-Parr S, Korchagina D, Toumi M, et al. Challenges in research and health technology assessment of rare disease technologies: report of the ISPOR rare disease special interest group. Value Health. 2018;21:493–500.

52. ASCO, NCCN Provide Guidance on Understanding and Managing Immunotherapy Side Effects. ASCO Connection. 2018 [cited 2019Jan13].

53. Medscape Log In. [cited 2019Jan13]. https://www.medscape.com/viewarticle/889689

54. Lipscomb J, Drummond M, Fryback D, et al. Retaining, and enhancing, the QALY. Value Health. 2009;12(Supp 1):S18–26.

55. Garrison L, Neumann PJ, Willke RJ, et al. A health economics approach to US value assessment frameworks – summary and recommendations of the ISPOR special task force report. Value Health. 2018;21:161–5.

56. Whittington MD, Mcqueen RB, Ollendorf DA, Kumar VM, Chapman RH, Tice JA, et al. Long-term Survival and Value of Chimeric Antigen Receptor T-Cell Therapy for Pediatric Patients with Relapsed or Refractory Leukemia. JAMA Pediatrics. 2018;172(12):1161.

57. Basch E. Toward a Patient-Centered Value Framework in Oncology. JAMA. 2016;315(9):2073–4.

58. Mott FE. Patient Reported Outcomes (PROs) as Part of Value-Based Care Can Shape Therapy Guidelines: Impact on Emerging Targeted Agents and Immunotherapy Protocols in Resource-Limited Regions. Oncology and Therapy. 2017;5(1):69–74.

59. Devlin NJ, Lorgelly PK. QALYs as a measure of value in cancer. J. Cancer Policy 2017;11:19–25.

60. Kovic B, Jin X, Kennedy SA, Hylands M, Pedziwiatr M, Kuriyama A, et al. Evaluating Progression-Free Survival as a Surrogate Outcome for Health-Related Quality of Life in Oncology: A Systematic Review and Quantitative Analysis. JAMA Internal Medicine. 2018;178(12):1586–96.

61. Efthymiadou O, Mossman J, Kanavos P. Differentiation of Health-Related Quality of Life Outcomes Between Five Disease Areas: Results from An International Survey Of Patients. International Journal of Technology Assessment in Health Care. 2018;34(05):498–506.

62. Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. Lancet Oncol. 2016;17(11):e510–4.

63. Calvert M, Kyte D, Mercieca-Bebber R. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols, The SPIRIT-PRO Extension. JAMA. 2018;319(5):483–94.

64. Howie LJ, Singh H, King-Kallimanis B, Roydhouse J, Theoret MR, Blumenthal GM, et al. Patient-reported outcomes in PD-1/PD-L1 inhibitor registration trials: FDA analysis of data submitted and future directions. Journal of Clinical Oncology. 2018;36(5_suppl):134.

65. National Coverage Analysis (NCA) for Chimeric Antigen Receptor (CAR) T-cell Therapy for Cancers (CAG-00451N). CMS.gov Centers for Medicare & Medicaid Services. [cited 2018Sep27]. https://www.cms.gov/medicare-coverage-database/details/nca-tracking-sheet.aspx?NCAId=291

66. MEDCAC Meeting 8/22/2018 - Chimeric Antigen Receptor (CAR) T-Cell Therapy and Patient Reported Outcomes. CMS.gov Centers for Medicare &

Medicaid Services. [cited 2018Sep27]. https://www.cms.gov/medicare-coverage-database/details/medcac-meeting-details.aspx?MEDCACId=76&NCAId=291&bc=gIAAAAAQAAA&

67. CMS MEDCAC Voting Scores: https://www.cms.gov/Regulations-and-Guidance/Guidance/FACA/downloads/id76c.pdf, Accessed 20 Oct 2018.

68. CMS NCD website: https://www.cms.gov/medicare-coverage-database/details/nca-tracking-sheet.aspx?NCAId=291&bc=gIAAAAAQAAA&. Accessed 2 Oct 2018.

69. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine Cancer treatment. JAMA. 2017;318(2):197–8.

70. Basch E. Toward a patient-centered value framework in oncology. JAMA. 2016;315(9):2073–4.

71. Ben-Aharon O, Magnezi R, Leshno M, Goldstein DA. Association of Immunotherapy with durable survival as defined by value frameworks for Cancer care. JAMA. 2018;4(3):326–32.

72. Harold Sox's paper defining "comparative effectiveness research" from the Institute of Medicine to kick off this section. Sox HC. Med Care. 2010; 48: S7-S8.

73. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR real-world data task force report. Value Health. 2007;10(5):326–35.

74. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices--overview: a report of the ISPOR-SMDM modeling good research practices task Force-1. Med Decis Mak. 2012;32(5):667–77.

75. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahm M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316(10):1093–103.

76. Neumann PJ, Kim DD, Trikalinos TA, Sculpher MJ, Salomon JA, Prosser LA, Owens DK, Meltzer DO, Kuntz KM, Krahn M, Feeny D, Basu A, Russell LB, Siegel JE, Ganiats TG, Sanders GD. Future Directions for Cost-effectiveness Analyses in Health and Medicine. Medical. Med. Decis. Making 2018;38(7):767–777.

77. Lakdawalla DN, Doshi JA, Garrison LP, Phelps CE, Basu A, Danzon PM. Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. Value in Health. 2018;21(2):131–9.

78. Mullins CD, Abdulhalim AM, Lavallee DC. Continuous Patient Engagement in Comparative Effectiveness Research. JAMA. 2012;307(15):1587–8.

79. Perfetto EM, Oehrlein EM, Boutin M, Reid S, Gascho E. Value to Whom? The Patient Voice in the Value Discussion. Value in Health 2017;20(2):286–291.

80. Emens LA, Ascierto PA, Darcy PK, Demaria S, Eggermont AM, Redmond WL, et al. Cancer immunotherapy: Opportunities and challenges in the rapidly evolving clinical landscape. European Journal of Cancer. 2017;81:116–29.

Research Article

How Oncologists Perceive the Availability and Quality of Information Generated From Patient-Reported Outcomes (PROs)

Michael Shea, PhD[1], Céline Audibert, PhD[2], Mark Stewart, PhD[1], Brittany Gentile, PhD[3], Diana Merino, PhD[1], Agnes Hong, PharmD[3], Laura Lassiter, PhD[1], Alexis Caze, PharmD[4], Jonathan Leff, MBA[4], Jeff Allen, PhD[1], and Ellen Sigal, PhD[1]

Abstract

Background: Despite increased incorporation of patient-reported outcome (PRO) measures into clinical trials, information generated from PROs remains largely absent from drug labeling and electronic health records, giving rise to concerns that such information is not adequately informing clinical practice. **Objective:** To evaluate oncologists' perceptions concerning the availability and quality of information generated from PRO measures. Additionally, to identify whether an association exists between perceptions of availability and attitudes concerning quality. **Method:** An online, 11-item questionnaire was developed to capture clinician perspectives on the availability and use of PRO data to inform practice. The survey also asked respondents to rate information on the basis of 4 quality metrics: "usefulness," "interpretability," "accessibility," and "scientific rigor." **Results:** Responses were received from 298 of 1301 invitations sent (22.9% response rate). Perceptions regarding the availability of PRO information differed widely among respondents and did not appear to be linked to practice setting. Ratings of PRO quality were generally consistent, with average ratings for the 4 quality metrics between "satisfactory" and "good." A relationship was observed between ratings of PRO data quality and perceptions of the availability. **Conclusion:** Oncologists' attitudes toward the quality of information generated from PRO measures are favorable but not enthusiastic. These attitudes may improve as the availability of PRO data increases, given the association we observed between oncologists' ratings of the quality of PRO information and their perceptions of its availability.

Keywords

cancer, health information technology, medical decision-making, survey data

Introduction

Cancer drugs often carry substantial treatment-related toxicities that may negatively impact patients' physical functioning and overall health-related quality of life (HRQoL) (1). While measures of treatment activity have provided the primary support for drug approval and payment decisions in oncology, they do not necessarily reflect patient perceptions of treatment benefit. Patient-reported outcomes (PROs) and clinical outcome assessments more broadly are important for characterizing clinical benefit, or "the impact of a treatment on how a patient feels, functions or survives," and can contribute meaningfully to efficacy and safety evaluations of a new treatment (2–4).

Much of the recent excitement around PROs stems from the recognition that these tools can be meaningful and reproducible and in many cases more accurate than clinician assessments (5). Historically, PRO tools were used primarily in oncology as research tools or in the measurement of palliative care interventions. However, PROs are now also used to measure HRQoL, disease-related symptoms, functional

[1] Friends of Cancer Research, Washington, DC, USA
[2] Deerfield Institute, Épalinges, Switzerland
[3] Genentech, South San Francisco, CA, USA
[4] Deerfield Management, New York, NY, USA

Corresponding Author:
Céline Audibert, Deerfield Management, Route de la Corniche 3a, Epalinges 1066, Switzerland.
Email: caudibert@deerfield.com

Table 1. Questionnaire Items.[a]

Item	Question	Response Choices
1-6	Items 1-6 of the questionnaire did not address the use of patient-reported outcomes	
7	What is your level of agreement with the following statement? *Patient-reported Outcome (PRO) data are widely available to prescribers in my field.*	a. Strongly agree b. Somewhat agree c. Neither agree or disagree d. Somewhat disagree e. Strongly disagree
8	To what extent have you considered patient-reported outcome (PRO) data when making prescribing decisions?	a. Always b. Very Often c. Sometimes d. Rarely e. Never f. Not applicable
9	Please rate the PRO information you have consulted in your practice on the following metrics: *Accessibility* *Usefulness* *Scientific rigor* *Interpretability*	a. Excellent b. Very good c. Good d. Satisfactory e. Poor f. Unsure
10	In the past, what sources have you used to access PRO data on a specific drug? (select all that apply)	a. Journal articles b. Conference abstracts/posters c. Sponsor company resources d. Patient forums e. Product labels f. Clinical guidelines g. Other h. None
11	What is your level of agreement with the following statement? *Adding a new section to the FDA product label that would contain PRO data would have a positive and meaningful impact on my prescribing decisions.*	a. Strongly agree b. Somewhat agree c. Neither agree or disagree d. Somewhat disagree e. Strongly disagree

[a]An online, 11-item questionnaire was developed to collect anonymized information on oncologists' attitudes toward different sources of prescribing information. The final 5 items of the questionnaire were the focus of this analysis. The questionnaire was comprised of Likert/Likert-type scale and multi-response questions. Respondents were asked to indicate the extent to which they have consulted information from patient-reported outcomes and then to rate that information using Likert-type scales.

impacts, treatment-related toxicities, treatment satisfaction, and in some cases the anticancer activity of drug interventions (6,7). Particular focus in recent years has centered around the measurement of adverse events reported by the patient (eg, PRO-CTCAE) (8).

A recent review of ClinicalTrials.gov found that between 2007 and 2013, the number of oncology trials that included at least one PRO measure has increased to approximately one-third of registered trials (9). Accordingly, regulatory agencies in the United States and Europe have taken steps to establish guidance for the use of PROs in clinical trials (10–12).

Despite a growing consensus regarding the importance of PROs and their regular incorporation into trial designs, there are concerns that the information generated from PROs is not reaching clinicians and patients (13). Much of this concern centers around the limited inclusion of patient-reported information in US Food and Drug Administration (FDA) product labeling. A recent analysis found that out of 160 approved hematology and oncology drugs between 2010 and 2014, only 3 included information generated from PROs in

labeling (14), although there have been additional label claims in the years since. More broadly, the literature on clinical trials that has included PROs has suffered from heterogeneity in the way data are analyzed, presented, and interpreted, hindering the incorporation of information into clinical guidelines and health policy (15).

We developed a survey to find out the degree to which clinicians felt that PRO information was available to them, where they typically find such information, and their opinion on the quality of that information. Insights from this survey are intended to help policymakers and others discover how to disseminate PRO data more effectively.

Methods

An 11-item, online physician survey was developed to collect anonymized information from physicians on their use of different sources of prescribing information (Table 1). Five items (items 7-11) specifically asked about physicians' use of and attitudes about PRO information and are the subject of this

analysis. The survey was piloted by 4 physicians prior to being distributed via e-mail to 1301 oncologists, who were recruited from a panel of medical professionals in the United States by a commercial research organization specializing in online physician surveys. Physicians were eligible if they reported being a board-certified oncologist or neurologist and had treated at least 10 patients in the past 12 months. The survey company, M3, verified the credentials of physicians opting in for survey research. Demographic and professional information was collected from each physician, including gender, type of practice (private, academic or community), and number of years in practice. Physicians were informed of the sponsors of the survey. The survey was open from December 2017 to February 2018. Each participating physician was given a small honorarium as compensation for their time.

By electing to complete the survey, respondents provided consent to use their anonymous responses. This study qualified as market research, as it did not involve patients or data on patient characteristics. As such, institutional review board and ethics committee approval and informed consent were not required, per current US regulations.

The survey was comprised of Likert/Likert-type scale and multiresponse questions. Data were pooled across participants and analyzed at the item-level using the R software package. Respondents were excluded from the analysis who answered "never" or "not applicable" to item 8 (15 respondents) and "unsure" to item 9 (20 respondents for "interpretability," 21 for "accessibility," 23 each for "usefulness" and "scientific rigor"). The rational for excluding those who answered "unsure" at item 9 was that they represented a small fraction of the total number of respondents and it was unclear why they were not sure how to rate the PRO information on the suggested metrics. Hypothesis testing was performed to assess whether there is strong evidence that the majority of oncologists (more than 50%) report that they "somewhat agree" or "strongly agree" that PRO information are widely available to them (item 7); "always" or "very often" consider PRO information when making prescribing decisions (item 8); and "somewhat agree" or "strongly agree" with the utility of adding a new section to the FDA product label that would contain PRO information (item 11). Hypothesis testing was also conducted to determine which resources are used by the majority of oncologists to access PRO information.

An oncologist's overall view of PRO data was quantified using a composite score based on their ratings of accessibility, interpretability, usefulness, and scientific rigor (item 9). The score was computed and validated using weights from a factor analysis (Supplemental Methods). Hypothesis testing compared the scores derived from the factor analysis across different populations of oncologists.

Results

Surveys were distributed to 1301 oncologists across the United States and responses were received from 298

(22.9% response rate). Of these respondents, 73% were male, 46% practiced in a private setting, and 41% had been in practice for 10 to 24 years (Supplemental Table 1). Information about the respondents' main area of focus was captured. One hundred seventy-four (58%) respondents focus on general oncology, 142 (48%) on hematology, and 98 (33%) mentioned some specific areas of specialization, with breast, lung, and gastrointestinal cancers being named most often. Given the high proportion of respondents mentioning general oncology, as well as the high proportion of respondents who mentioned more than one area of specialization, no analysis at the specialty level was performed. Geographic information was captured for 59% of respondents.

Perceptions regarding the availability of PRO information and frequency of use in making prescribing decisions differed widely among respondents (Figure 1). For example, 43% of respondents agreed (either "strongly" or "somewhat") that PROs are widely available and 34% disagreed (either "strongly" or "somewhat"). Additionally, 22% of respondents reported they "always" or "very often" consider PRO information when making prescribing decisions, whereas 27% reported they "rarely" or "never" consider PRO information. The most commonly cited sources of PRO information were "journal articles" (62%) and "clinical guidelines" (45%).

Respondents rated the quality of PRO data between "satisfactory" and "good" on average (Figure 2). No major differences in ratings of the 4 quality metrics "usefulness," "accessibility," "interpretability," and "Scientific rigor" were observed; however, respondents gave slightly higher scores to PRO data on the basis of the "usefulness," with 54% of respondents providing a rating of "good," "very good," or "excellent."

Hypothesis testing was used to investigate the impact of specific criteria on ratings of PRO quality (Table 2). As theorized, there was evidence that oncologists who believe that PRO data are widely available and those who use PRO data to prescribe medications rated it higher on average. Results also showed that the majority (63%) of oncologists "somewhat to strongly agree" that adding a new section to the FDA product label with PRO data would have a meaningful impact on their prescribing decisions.

Discussion

We surveyed oncologists regarding their perceptions of available PRO information and the extent to which they use PRO data to inform treatment decision-making. Overall, we found that oncologists hold heterogeneous views on the extent to which PRO data are available and the quality of the information they have access to. Oncologists currently hold favorable but not enthusiastic opinions regarding the quality of PRO information they have considered. On average, respondents rated PRO information between "satisfactory" and "good" on the

7. What is your level of agreement with the following statement?
"Patient-reported Outcome (PRO) data are widely available to prescribers in my field."

	Number	Total	Percent
Strongly agree	12	298	4%
Somewhat agree	115	298	39%
Neither agree or disagree	70	298	23%
Somewhat disagree	82	298	28%
Strongly disagree	19	298	6%

8. To what extent have you considered patient-reported outcome (PRO) data when making prescribing decisions?

	Number	Total	Percent
Always	6	298	2%
Very often	61	298	20%
Sometimes	150	298	50%
Rarely	66	298	22%
Never	15	298	5%
Not applicable	0	298	0%

10. In the past, what sources have you used to access PRO data on a specific drug? (select all that apply)

	Number	Total	Percent
Journal articles	164	265	62%
Conference abstracts/posters	105	265	40%
Sponsor company resources	59	265	22%
Patient forums	40	265	15%
Product labels	99	265	37%
Clinical guidelines	119	265	45%
Other	4	265	2%
None	21	265	8%

11. What is your level of agreement with the following statement?
"Adding a new section to the FDA product label that would contain PRO data would have a positive and meaningful impact on my prescribing decisions."

	Number	Total	Percent
Strongly agree	53	265	20%
Somewhat agree	113	265	43%
Neither agree or disagree	65	265	25%
Somewhat disagree	27	265	10%
Strongly disagree	7	265	3%

Figure 1. Oncologists' perceptions of availability of patient-reported outcome (PRO) data. Items 7 and 8 asked respondents to report their perceptions of the availability of PRO information as a prescribing resource. Item 10 asked respondents to select sources they have used to access PRO information in the past. Item 11 gauged respondents' level of agreement with the utility of adding a new section to product labeling that would contain PRO data. Responses for 33 respondents were not considered for items 10 and 11 due to a response of "never" to item 8 or "unsure" to item 9.

basis of 4 quality metrics: usefulness, interpretability, accessibility, and scientific rigor.

We also found that the majority of oncologists do not frequently use PRO data when making prescribing decisions. Given that PRO data have not traditionally been well represented in product information and the lack of standardization with regard to how such information is presented in the clinical trial literature (5), this is not entirely

9. Please rate the PRO information you have consulted in your practice on the following metrics:

Usefulness, Interpretability, Accessibility, Scientific rigor

	1 (Poor)		2 (Satisfactory)		3 (Good)		4 (Very good)		5 (Excellent)	
Usefulness	27	10%	77	27%	72	25%	58	20%	26	9%
Interpretability	35	12%	85	30%	74	26%	53	19%	16	6%
Accessibility	56	20%	62	22%	76	27%	50	18%	18	6%
Scientific rigor	62	22%	70	25%	64	23%	47	17%	17	6%

Figure 2. Oncologists' ratings of PRO information. Item 9 asked respondents to rate the PRO information they have consulted on the basis of 4 quality metrics: "usefulness," "interpretability," "accessibility," and "scientific rigor." Twenty-three respondents selected "unsure" when asked to provide ratings, and their responses were eliminated from the analysis.

Table 2. Hypothesis Test Results.[a]

Item(s)	Research Question	Hypothesis Test	Results, *P* [95% CI]
7	Do the majority of oncologists "somewhat agree" or "strongly agree" that PRO data are widely available to prescribers in their field?	One-sample proportion $H_0: P = .5$ $H_A: P > .5$	.9946 [.37-.48]
8	Do the majority of oncologists consider PRO data when making prescribing decisions "always' or "very often"?	One-sample proportion $H_0: P = .5$ $H_A: P > .5$	1 [.18-.27]
9	Do the majority of oncologists consider PRO information "excellent" or "very good" on the basis of the following metrics? "Accessibility" "Interpretability" "Usefulness" "Scientific rigor"	One-sample proportion $H_0: P = .5$ $H_A: P > .5$	1 [.21-.31] 1 [.21-.32] 1 [.27-.38] 1 [.20-.30]
10	Do the majority of oncologists use the following sources to access PRO data on a specific drug? Journal articles Clinical guidelines Product labels Sponsor company resources	One-sample proportion $H_0: P = .5$ $H_A: P > .5$	**.0005** [.54-.65] .98 [.38-.50] 1 [.28-.41] 1 [.16-.26]
11	Do the majority of oncologists "somewhat agree" or "strongly agree" with adding a new section to the FDA product label that would contain PRO data would have a positive and meaningful impact on their prescribing decisions?	One-sample proportion $H_0: P = .5$ $H_A: P > .5$	**<.0001** [.57-.69]
7, 9	Are "an oncologist's opinion that PRO data is widely available" and "an oncologist's rating of PRO data" related?	Chi-square Null hypothesis: No relationship	**<.0001** n/a
7, 9	Are oncologists who believe that PRO data are widely available more likely to rate it higher than those who do not believe it is widely available?	Two-sample *t* test $H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 > 0$	**<.0001** [3.8-5.8]

Abbreviations: CI, confidence interval; FDA, Food and Drug Administration; N/A, not applicable.

[a]Hypothesis testing was performed to assess whether there is strong evidence that the majority of oncologists (more than 50%) report that they: "somewhat agree" or "strongly agree" that PRO information are widely available to them (item 7); "always" or "very often" consider PRO information when making prescribing decisions (item 8); consider PRO information "excellent" or "very good" on the basis of 4 quality metrics; and "somewhat agree" or "strongly agree" with the utility of adding a new section to the FDA product label that would contain PRO information (item 11). Hypothesis testing was also conducted to determine which resources are used by the majority of oncologists to access PRO information.

Boldface values indicate statistically significant differences.

surprising. However, we found a clear link between those who consider PRO data as generally available and more positive attitudes about data quality. This suggests that familiarity with PRO data and scientific acceptance are associated with integration into practice. Although it is not possible to make statements about causality, increasing

access to PRO data and improving the quality of the data may encourage integration into clinical practice and are worthy goals in the move toward patient-focused drug development.

Given the relationship between perceptions of PRO availability and ratings of PRO data quality, our research suggests that increased exposure to PRO information may improve physician regard for such data. Therefore, we lay out the following recommendations for how to increase utilization and uptake of PRO data for treatment decision-making by physicians.

First, continued efforts should be directed toward conveying PRO information through drug labeling. Information found on drug labels is used by a range of other prescribing resources and may thus increase prescriber exposure to such information in a range of venues. Moreover, some have suggested that market forces will encourage manufacturers to invest more in PRO labeling if they observe more success cases (16). However, given the many barriers to the inclusion of PRO data in labels, especially for cancer products, the FDA may need to consider additional opportunities for disseminating PRO data, such as through the development of a separate section of product labels specifically devoted to such information. As stated in a May 2017 public meeting, FDA officials are actively considering such an approach, either through the creation of a new section on printed package inserts or as online labeling appendices (17).

Second, in the absence of widespread access to PRO information on labels in the short term, clinical investigators will need to consider more digestible formats for the information in peer-reviewed publications. Peer-reviewed literature was identified in this study as the most relied upon source for accessing PRO information. As previously noted, the peer-reviewed literature has suffered from heterogeneity in the presentation of PRO data, hindering its accessibility.

Finally, utilization and uptake of PRO data will continue to increase if sustained support for patient-focused drug development continues. The 21st Century Cures Act and the most recent reauthorization of the Prescription Drug User Fee Act both contained important provisions related to the dissemination of PRO data and signaled policymakers' support for a more patient-focused drug development process (18,19). Careful implementation of these statutes, as well as the timely development of new regulatory guidance, will further advance understanding of and support for patient-focused drug development.

Conclusion

This research summarizes the current acceptance and usage of PRO data for treatment decision-making among a sample of oncologists. Current attitudes toward PROs, though favorable, may improve as availability is increased, given the link between perceptions of PRO availability and

oncologists' rating of PRO information. Regulators should continue to evaluate new methods of conveying data from PROs to prescribers, such as through expansions of physician package inserts.

ORCID iD

Céline Audibert ⓘ https://orcid.org/0000-0002-2215-1462
Mark Stewart ⓘ https://orcid.org/0000-0002-4847-0736

Supplemental Material

Supplemental material for this article is available online.

References

1. Basch E, Geoghegan C, Coons SJ, Gnanasakthy A, Slagle AF, Papadopoulos EJ, et al. Patient-reported outcomes in cancer drug development and US regulatory review: perspectives from industry, the food and drug administration, and the patient. JAMA Oncol. 2015;1:375-9.

2. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, eds. Clinical Measurement in Drug Evaluation. New York, NY: John Wiley; 1995.

3. Levit LA, Balogh EP, Nass SJ, Ganz PA. Delivering High-Quality Cancer Care: Charting a New Course for a System in Crisis. Washington, DC: National Academies Press; 2013.

4. Kluetz PG, O'Connor DJ, Soltys K.Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol. 2018;19:e267-74. doi:10.1016/S1470-2045(18)30097-4.

5. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care—hearing the patient voice at greater volume. Nat Rev Clin Onc. 2017;14:763-72.

6. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med. 2010;363:733-42.

7. Mesa RA, Gotlib J, Gupta V, Catalano JV, Deininger MW, Shields AL, et al. Effect of ruxolitinib therapy on myelofibrosis-related symptoms and other patient-reported outcomes in COMFORT-I: a randomized, double-blind, placebo-controlled trial. J Clin Oncol. 2013;31:1285-92.

8. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014;106:dju244.

9. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials. 2015;43:1-9.

10. US Food and Drug Administration. The 21st Century Cures Act. 2016. Retrieved July 6, 2018, from: https://www.fda.gov/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/21stCenturyCuresAct/default.htm.

11. US Food and Drug Administration. Prescription Drug User Fee Act VI. 2017. Retrieved July 6, 2018, from: https://www.fda.gov/forindustry/userfees/prescriptiondruguserfee/ucm446608.htm.

12. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: the use of patient-reported outcome (PRO) measures in oncology studies. 2016. Retrieved July 6, 2018, from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf.

13. Gensheimer SG, Wu AW, Snyder CF; PRO-EHR Users' Guide Steering Group; PRO-EHR Users' Guide Working Group. Oh, the places we'll go: patient-reported outcomes and electronic health records. Patient. 2018;11:591-8. doi:10.1007/s40271-018-0321-9.

14. Gnanasakthy A, DeMuro C, Clark M, Haydysch E, Ma E, Bonthapally V. Patient-reported outcomes labeling for products approved by the office of hematology and oncology products of the US Food and Drug Administration (2011–2014). J Clin Oncol. 2016;34:1928-34.

15. Bottomley A, Pe M, Sloan J, Basch E, Bonnetain F, Calvert M. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. Lancet Oncol. 2016;17:e510-4.

16. Gnanasakthy A, Mordin M, Evans E, Doward L, DeMuro C. A review of patient-reported outcome labeling in the United States (2011-2015). Value Health. 2017;20:420-9.

17. Sutter S. Will patient-reported data require new drug labeling framework? The Pink Sheet. May 05, 2017.

18. 21st Century Cures Act, Section 3001, H.R. 34, 114th Cong. 2015.

19. FDA Reauthorization Act of 2017, Section 605, 115th Cong. 2017.

Author Biographies

Michael Shea, PhD, analyzes how federal and state health care policies impact biomedical innovation and the public health. His main research areas are: prescription drug labeling, FDA expedited programs, and clinical diagnostics policy.

Céline Audibert, PhD, has over 15 years experience in market research and business intelligence in the healthcare sector. She designs and executes primary market research projects in EU and the US to better understand current healthcare practices, evaluate markets trends, product potential and competitive dynamics for a broad range of indication.

Mark Stewart, PhD, serves as a Vice President, Science Policy at Friends of Cancer Research (Friends). Mark leads the development and implementation of the organization's research and policy agenda as well as overseeing the conduct of research projects to help develop innovative policy proposals and consensus-driven solutions for cancer drug development.

Brittany Gentile, PhD, is a senior scientist within the Patient Centered Outcomes Research (PCOR) group at Genentech. She supports clinical outcome assessment strategy development for clinical trials in hematology, ophthalmology, and respiratory conditions. She holds a PhD in social psychology from the University of Georgia.

Diana Merino, PhD, serves as Science Policy Analyst at Friends of Cancer Research. Diana is passionate about disseminating scientific research findings and implementing solutions to accelerate progress in cancer care. Her own experience with the disease drives her desire to improve awareness of the needs cancer patients and survivors face, and the importance of cancer research. She has accumulated a wealth of experience as a spokesperson, cancer advocate, committee member, and 2016-2017 Chair of the American Association for Cancer Research (AACR) Associate Member Council, whose goal is to promote the professional development of early career scientists.

Agnes Hong, PharmD, is an outcome research scientist in oncology. Agnes Hong received her Doctor of Pharmacy from Rutgers University and completed a 2-year post-doctoral PharmD Fellowship with Pfizer Inc in the Medical Affairs and Health Economics Outcomes Research groups. She joined Genentech Inc as part of the Oncology Patient-Centered Outcomes Research group, where she leads the Skin Franchise and supports lung, head & neck and breast cancer programs. Agnes is responsible for generating data on the patients' experience with disease and treatment burden.

Laura Lassiter, PhD, serves as a Science Policy Analyst at Friends of Cancer Research. Prior to joining Friends, Laura worked for the American Association for the Advancement of Science as a Congressional Science Fellow. As a Congressional Science Fellow, Laura handled Senator Al Franken's health portfolio. Laura primarily focused on issues relating to the FDA and prescription drug prices. Additionally, during grad school Laura served as the Director of the Mid-South Academic Alliance, the workforce development arm of Life Science Tennessee, a nonprofit that advocates for the life science industry in the state. At Friends, Laura works on the development of evidence-based policies that will improve care for cancer patients and survivors, facilitate dialogue between stakeholders through the organization of conferences and symposia, and continue advocating on behalf of cancer patients and survivors.

Alexis Caze, PharmD, heads Deerfield Management's research and consulting group that he established in 2006 when he joined Deerfield. The Institute provides insight and expertise in areas such as market research, marketing, market access, medical and intellectual property that is used to better understand commercial and regulatory dynamics of the healthcare industry and for the development and testing of investment theses. Prior to joining Deerfield, Alexis spent three years at Gerson Lehrman Group, a leading provider of systems to manage expert networks where he was managing and serving key accounts contributing to the rapid expansion of the healthcare practice. Before that, Alexis was Strategic Marketing Manager at Sanofi, analyzing research and data needed to develop

the launch strategy for two of the main portfolio products. His responsibilities included product positioning, segmentation, KOL mapping, managed care strategy and market analysis. Alexis began his career as a consultant at PricewaterhouseCoopers serving clients in the healthcare space. He earned his PharmD from the University of Pharmacy in Paris and his MSc in Business from E.M. Lyon Business School.

Jonathan Leff, MBA, is a partner on the Private Transactions team at Deerfield Management and Chairman of the Deerfield Institute. He joined Deerfield in 2013, and focuses on venture capital and structured investments in biotechnology and pharmaceuticals. Prior to joining Deerfield, for more than sixteen years, Mr. Leff was with Warburg Pincus, where he led the firm's investment efforts in biotechnology and pharmaceuticals. He is a member of the Boards of several public and private healthcare companies as well as several not-for-profit Boards, including the Spinal Muscular Atrophy Foundation, Friends of Cancer Research, the Reagan-Udall Foundation for the Food and Drug Administration and the Columbia University Medical Center Board of Advisors. Mr. Leff has also been active in public policy discussions related to healthcare and medical innovation. He previously served as a member of the Executive Committee of the Board of the National Venture Capital Association (NVCA), where he led NVCA's life sciences industry efforts as Chair of NVCA's Medical Innovation and Competitiveness Coalition (NVCA-MedIC), and also previously served on the Board of the Biotechnology Innovation Organization. Mr. Leff received his A.B. from Harvard University, and earned his M.B.A. from The Stanford Graduate School of Business.

Jeff Allen, Ph.D. serves as the President and CEO of Friends of Cancer Research (Friends). During the past 20 years, Friends has been instrumental in the creation and implementation of policies ensuring patients receive the best treatments in the fastest and safest way possible. As a thought leader on many issues related to Food and Drug Administration, regulatory strategy and healthcare policy, he is regularly published in prestigious medical journals and policy publications, and has contributed his expertise to the legislative process on multiple occasions. Recent Friends initiatives include the establishment of the Breakthrough Therapies designation and the development of the Lung Cancer Master Protocol, a unique partnership that will accelerate and optimize clinical trial conduct for new drugs. Dr. Allen received his Ph.D. in cell and molecular biology from Georgetown University, and holds a Bachelors of Science in Biology from Bowling Green State University.

Ellen Sigal, PhD, is chairperson and Founder of Friends of Cancer Research, a think tank and advocacy organization based in Washington that drives collaboration among partners from every healthcare sector to power advances in science, policy and regulation that speed life-saving treatments to patients. Dr. Sigal is Chair of the Board of Directors of the Reagan-Udall Foundation, serves on the Board of the Foundation for the National Institutes of Health, and on the Board of Governors of the Patient Centered Outcomes Research Institute. In 2016, Dr. Sigal was named to Vice President Biden's Cancer Moonshot Blue Ribbon Panel, the Parker Institute for Immunotherapy Advisory Group and joined the board of advisors for the George Washington University's Milken Institute of Public Health. She also holds leadership positions with a broad range of advocacy and policy organizations and academic health centers including; MD Anderson, Duke Cancer Center, Sidney Kimmel Comprehensive Cancer Center and the Sylvester Comprehensive Cancer Center.

Controversy

Improving attribution of adverse events in oncology clinical trials

Goldy C. George[a,1], Pedro C. Barata[b,1], Alicyn Campbell[c], Alice Chen[d], Jorge E. Cortes[e],
David M. Hyman[f], Lee Jones[3], Thomas Karagiannis[g], Sigrid Klaar[h], Jennifer G. Le-Rademacher[i],
Patricia LoRusso[j], Sumithra J. Mandrekar[i], Diana M. Merino[k], Lori M. Minasian[l],
Sandra A. Mitchell[m], Sandra Montez[e], Daniel J. O'Connor[n], Syril Pettit[o], Elaine Silk[3],
Jeff A. Sloan[i], Mark Stewart[l], Chris H. Takimoto[p], Gilbert Y. Wong[q], Timothy A. Yap[e],
Charles S. Cleeland[e,2,*], David S. Hong[e,2,*]

[a] The University of Texas MD Anderson Cancer Center (MD Anderson), Houston, TX, United States
[b] Tulane University, New Orleans, LA, United States
[c] Genentech, South San Francisco CA
[d] National Cancer Institute (NCI), Bethesda, MD, United States
[e] MD Anderson, Houston, TX, United States
[f] Memorial Sloan Kettering Cancer Center, New York, NY, United States
[g] Genentech, Chicago, IL, United States
[h] Swedish Medical Products Agency, Uppsala, Sweden
[i] Mayo Clinic, Rochester, MN, United States
[j] Yale University Cancer Center, New Haven, CT, United States
[k] Friends of Cancer Research, Washington, DC, United States
[l] NCI, Bethesda, MD, United States
[m] NCI, Rockville, MD, United States
[n] Medicines and Healthcare Products Regulatory Agency, London, United Kingdom
[o] Health and Environmental Sciences Institute, Washington DC, United States
[p] Forty Seven, Inc., Menlo Park, CA, United States
[q] Pfizer, New York NY, United States

ARTICLE INFO

ABSTRACT

Attribution of adverse events (AEs) is critical to oncology drug development and the regulatory process. However, processes for determining the causality of AEs are often sub-optimal, unreliable, and inefficient. Thus, we conducted a toxicity-attribution workshop in Silver Springs MD to develop guidance for improving attribution of AEs in oncology clinical trials. Attribution stakeholder experts from regulatory agencies, sponsors and contract research organizations, clinical trial principal investigators, pre-clinical translational scientists, and research staff involved in capturing attribution information participated. We also included patients treated in oncology clinical trials and academic researchers with expertise in attribution. We identified numerous challenges with AE attribution, including the non-informative nature of and burdens associated with the 5-tier system of attribution, increased complexity of trial logistics, costs and time associated with AE attribution data collection, lack of training in attribution for early-career investigators, insufficient baseline assessments, and lack of consistency in the reporting of treatment-related and treatment-emergent AEs in publications and clinical scientific reports. We developed recommendations to improve attribution: we propose transitioning from the present 5-tier system to a 2–3 tier system for attribution, more complete baseline information on patients' clinical status at trial entry, and mechanisms for more rapid sharing of AE information during trials. Oncology societies should develop recommendations and training in attribution of toxicities. We call for further

* Corresponding authors at: Department of Investigational Cancer Therapeutics, Professor and Deputy Chair, Associate Vice President of Clinical Research, The University of Texas MD Anderson Cancer Center, 1400 Holcombe Boulevard, Unit 455, Houston, TX 77030, United States (D.S. Hong). Department of Symptom Research, McCullough Professor of Cancer Research, The University of Texas MD Anderson Cancer Center, 1400 Pressler, Unit 1450, Houston, Texas 77030, United States (C.S. Cleeland).
E-mail addresses: ccleeland@mdanderson.org (C.S. Cleeland), dshong@mdanderson.org (D.S. Hong).
[1] First authors contributed equally.
[2] Senior authors contributed equally.
[3] Patient (no affiliation)

harmonization and synchronization of recommendations regarding causality safety reporting between FDA, EMA and other regulatory agencies. Finally, we suggest that journals maintain or develop standardized requirements for reporting attribution in oncology clinical trials.

Introduction

The reporting of adverse events (AEs) is an essential aspect of oncology drug development and the regulatory process. AE reporting is key to determining a new drug's toxicity profile [1], which will ultimately contribute to the benefit–risk assessment and will be included in the label [2]. However, the process for determining the origin of the AE is challenging [3–5], sub-optimal and inefficient and produces information that may be of limited or uncertain value for regulatory decision-making and for informing clinical practice and future research steps. These inefficiencies can result in added burden upon resources (eg, cost, time, and effort) for investigators, industry, ethics

committees, and regulatory bodies. Ultimately, inefficient processes affect patients by limiting the number of new medicines that can advance through the clinical trial trajectory and reducing the collection of actionable information that could guide optimal care delivery and support.

The American Society of Clinical Oncology (ASCO) and other groups have recognized these concerns and made recommendations for streamlining the reporting of serious adverse events [3]. Although the published ASCO guidelines address the issue of attribution of the AE (initial assessment of whether or not the event is caused by the agent being tested), they focus primarily on other parts of the process, such as expedited Investigational New Drug safety reporting. The ASCO panel

Early drug development
When little is known about a drug, the focus of attribution is to assess safety during an early-phase trial, so as to protect study participants. The attribution of AEs contributes towards defining dose-limiting toxicities and thereby the dosing strategy. Concepts such as "maximum tolerated dose" and "dose-limiting toxicity" must be carefully defined to best capture relevant toxicities across the entire treatment period. Anticipated toxicities or AEs are based on knowledge of the pharmacological class as well as on mechanistic investigations conducted in early non-clinical animal or in vitro studies.

Late and confirmatory stage
Characterizing the safety profile is key in order to inform on the benefit risk assessment and the content of the product information/ label. As the safety database is compiled, mild-to-moderate and expected AEs sometimes receive less attention; consequently, chronic low-grade toxicities or delayed-onset toxicities may be underreported, whereas toxicities of special interest may be over-reported. Stakeholders are increasingly interested in integrating traditional methods of safety reporting with data collection from other sources (eg, patient-reported outcomes) that may reflect important aspects of tolerability and effects on functioning.[1, 2] [3-5]

Post marketing safety aspects
Smaller pre-market exposure equates with increased importance of and need for post-market monitoring and data collection. Postmarket assessment is important for identifying safety signals that reflect diverse patient populations, varied therapeutic delivery patterns, and expanded treatment timescales.
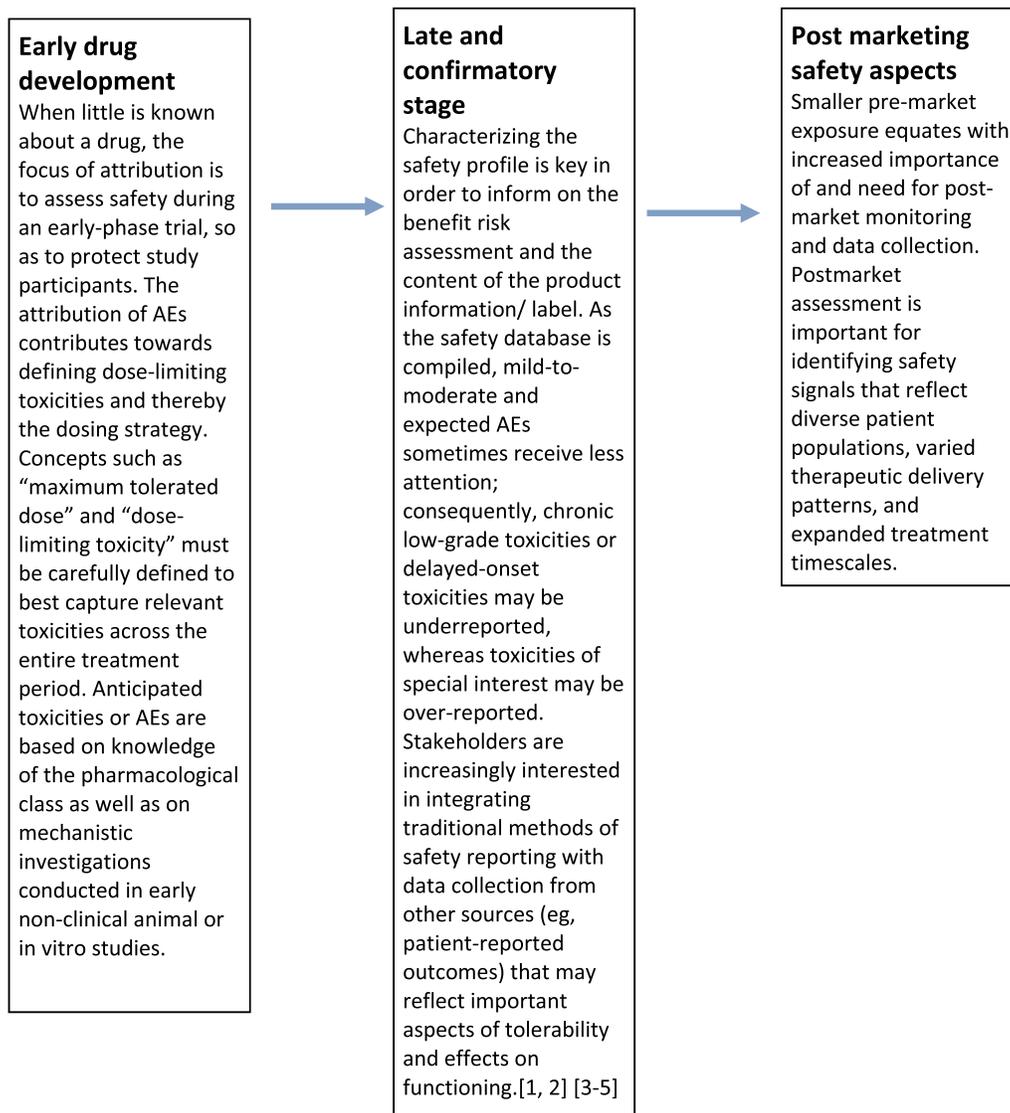
Fig. 1. The Focus of Attribution through the Drug Development Program.

acknowledged that it was often difficult to distinguish AEs that result from an intervention from those with other causes.

In order to review the challenges with attribution and develop recommendations for improving attribution, a consensus-building workshop on toxicity attribution was convened in Silver Springs, Maryland in September 2017. Discussions were held with multiple stakeholders, including representatives from regulatory agencies from both the United States and Europe, the US National Cancer Institute (NCI), pharmaceutical sponsors and contract research organizations, academic clinical trial principal investigators, non-clinical translational safety scientists, and research staff involved in capturing AE attribution information. The working group also included patients who had been treated in early-phase clinical trials and academic researchers with expertise in attribution. The working group attribution stakeholder experts were selected based on nominations from regulatory agencies, experienced clinical trialists, and the NCI. The working group focused on reviewing current practice, including the rigor and utility of AE attribution, identifying major concerns with the current process of attribution, and developing consensus on how AEs might be more precisely attributed to treatment-related versus non–treatment-related causes.

Preparation for the toxicity-attribution workshop included a series of teleconferences with multiple expert stakeholders and a pre-workshop in Houston, Texas in February 2017 to review the current state of AE attribution, identify key issues for further work, and create the agenda topics and key issues and questions for the September workshop. The workshops were co-sponsored by the Friends of Cancer Research, The University of Texas MD Anderson Cancer Center, and The Health and Environmental Sciences Institute®.

Defining attribution and its importance

Attribution is defined as "the act of saying or thinking that something is the result or work of a particular person or thing."[6] In cancer research, attribution is the determination of whether or not an untoward clinical event that occurred during (or after) the administration of a treatment is related to the treatment. The term is primarily used in relation to the study drug or intervention in a clinical trial. In certain regulatory contexts (eg, the European Union), the terms "causality" or "relatedness to study treatment" are used when referring to attribution. Attribution is referred to as "relatedness to study treatment" in the scientific literature [7]. The purpose of attribution varies during the different phases of the development program for a new medicine (Fig. 1). For the purposes of this paper, we will focus on attribution in the context of its relation to oncology clinical trials.

Implications of misattribution

Accurate attribution of AEs to an experimental drug versus other potential causes, such as other concomitant therapies, symptoms of the underlying disease, or comorbidities, is not always straightforward. An event may be incorrectly attributed to other causes when it is in fact related to the experimental therapy (Type A error), or it may be attributed to the experimental therapy when in fact it is associated with other causes (Type B error) [7,8]. Type A errors can result in more patients being exposed to potentially toxic levels of the drug, with a negative impact on safety, whereas Type B errors may lead to premature study termination [8]. These errors are known to negatively affect estimates of the "true" maximum tolerated dose (MTD) and, consequently, the accuracy, safety, sample size, and/or treatment dose or duration of a future confirmatory trial [8].

The magnitude of impact of these errors is related to the trial design used. Data suggest that the standard "3 + 3" dose-escalation schema (wherein patients are enrolled at increasing dose levels based on the presence or absence of dose-limiting toxicities in a pre-specified proportion of patients until the MTD is determined) [9] is particularly sensitive to Type B errors [8,10]. Misattribution of dose-limiting toxicities also has the potential to lead to underestimation of the MTD. Even with biologic agents (for which a minimally effective dose, rather than an MTD, is frequently preferred), proper understanding of the expected toxicities and therapeutic window of a given agent has important implications for further development and clinical use.

Downstream consequences of misattribution include the potential for evaluating sub-therapeutic doses of the drug and inaccurate safety profiling of the drug in the label. As a worst-case scenario, poor attribution can lead to a faulty final causality assessment, affecting the

Table 1
Principles for Assessing Causality of an Adverse Event.

Broad Factor/Category	Principle	Explanatory wording
Patient-level factors	Timing of the adverse event (AE) relative to drug exposure; plausible temporal relationship	Is the timing of the AE compatible with its being caused by the drug? Did it occur, or increase in severity, during or after exposure to the drug?
	Relation of AE to baseline symptoms, including severity	Is the AE an existing comorbidity, disease symptom, or residual toxicity from previous therapy, as illustrated by existence at baseline? Did it increase in severity after administration of the drug?
	Response to interruption of administration ("de-challenge") or to readmission of the agent after recovery from the AE ("re-challenge")	Did the AE resolve with drug interruption? Did it recur if or when the drug was restarted?
	Dose-response patterns in the individual patient that indicate a causal relationship	Is the event increasing and decreasing with dose reductions and increases? Note: Not relevant at first occurrence
	Likelihood of alternative causes, such as disease symptoms, other medication, other disease	Does the patient have comorbidities or concomitant medication likely to cause the AE, or is the AE expected in the patient population (eg, due to age)?
Agent-level factors	Preclinical and clinical knowledge of the drug, its pharmacology and toxicology; AE identified as a drug reaction in the reference safety information ("expectedness")	Is the AE something that the drug is expected to cause? It the AE biologically plausible if related to the drug?
Trial or program level/ aggregate data level factors	Dose-response patterns across patients that indicate a causal relationship	Is the event increasing and decreasing with dose/exposure?
	Incidence of the AE in the intervention group versus placebo or active comparator groups	Is there a relevantly higher frequency in the experimental arm (which usually indicates that the event has a causal relationship with the drug)? Note: Lower frequencies in the experimental arm versus an active comparator arm still may be compatible with a causal relationship and must undergo further biological-pharmacological plausibility assessment.

benefit–risk assessment. Patients may be taken off active therapy unnecessarily, and the product label may incorrectly identify an event as being causally linked to the drug.

Current status of attribution

Framework for AE reporting

The lexicon for AE identification and grading in the context of oncology clinical trial reporting is the Common Terminology Criteria for Adverse Events (CTCAE), first developed in 1983 by the NCI's Cancer Therapy Evaluation Program (CTEP) [11]. The newest CTCAE version, v5.0, was released in 2018 and includes new AE terms, clarified definitions, and updated grading [12]. In 2016, the NCI released its patient-reported outcomes version of the CTCAE, a new standardized method for assessing symptomatic AEs from the patient's perspective. In 1998, the CTEP introduced the idea of collecting and reporting AE attribution data in clinical trials, based on a set of 5 nominal categories of attribution to study drug: "definitely related," "probably related," "possibly related," "unlikely related," and "unrelated" [13].

Attribution is assigned at the patient level and then summarized at the trial level. However, even in those instances in which attribution appears to be well defined, the selection of an AE term and its grading are highly user dependent and are a potential source of variation in the reporting of trial data [14].

Methods for making attribution

Sponsors are charged with identifying all AEs that are attributable to an agent being tested in a clinical trial, especially in early-phase studies that evaluate the safety of new medicines. This information is collected in a standardized format by site investigators and summarized by the sponsor, who is ultimately responsible for reporting serious and unexpected suspected adverse reactions to the FDA and other regulatory authorities that have jurisdiction over the study locale. A product of these efforts in the United States was a guideline for streamlining AE reporting produced by the Clinical Trials Transformation Initiative, a publicprivate partnership to develop and drive adoption of practices that will increase the quality and efficiency of clinical trials [3,15]. To further this process, the present working group discussed and summarized the main principles for assessing causality in order to provide a hands-on tool to guide study investigators (Table 1).

Challenges in the current state of attribution

The current system for attribution is not optimal. In a retrospective

Table 2
Challenges Associated with Attribution.

Issue	Description
Trial logistics, costs, and time	• Site investigator time to gather sufficient clinical data to review and determine adverse event attribution Extensive amounts of AE attribution data are collected and reported, and this represents a burden for researchers, research site personnel, and regulators [31,32]. Cancer patients may have multiple different AEs over the course of a trial. AE attribution reporting may be particularly burdensome for patients with diseases associated with multiple events as part of the nature of the disease, such as patients with hematologic disease.
Lack of education about assigning attribution	• Specific training and concrete guidelines on how to reliably attribute AEs in clinical research either are often insufficient or lacking altogether [33,33].
Advanced disease, multiple previous therapies, and multiple comorbidities	• Attribution of AEs is particularly challenging in patients with advanced disease who may have undergone multiple treatments (including chemotherapy, radiation, hormonal therapy, targeted agents, and/or hospitalizations) [3]. Some patients with advanced disease may also be older, with higher levels of disease-related symptoms, multiple comorbidities, and concomitant medications [3]. With the broadening of clinical trial eligibility criteria [34], there will likely be more patients with pre-existing conditions and, subsequently, a greater potential for drugdrug interactions and comorbidities as possible causes of AEs.
Insufficient baseline information about health status at trial entry	• Baseline data are often inadequate. There is great variability across institutions in the symptoms or clinical abnormalities assessed at baseline and the degree of comprehensiveness with which baseline assessments are performed. Inadequate baseline examination and documentation and insufficient washout periods can increase the potential for confounding toxicities that are actually late effects of a previous drug. This may cause AEs to be incorrectly attributed to the new therapy being tested in a clinical trial. Also, patients may underreport the use of over-the-counter medications and supplements that could have potential drugdrug interactions, which could affect expected adverse events.
Multiagent studies	• Combinations of cancer drugs are often tested in early-phase clinical trials, yet the safety profile for each of the novel drugs being tested, as well as their combination, may not be fully known. Although some of the individual drugs being tested may not have direct pharmacokinetic or pharmacodynamic interactions, they may have overlapping toxicities, making it very difficult to determine whether the different oncology compounds are synergistic or additive with each other in terms of AEs.
Inconsistencies in reporting of treatment-related AEs	• A lack of consistency in how treatment-related and treatment-emergent AEs are reported in publications and clinical scientific reports can add to confusion around the attribution of AEs [14]. Moreover, results from an individual trial are now being reported in multiple forms, including publications, regulatory documents, clinical study reports, registries, medical meetings and presentations, and patient-level data portals and other databases [35], creating the possibility of greater variation and heterogeneity in AE reporting due to these differing formats [35]. It is also common to default to reporting only serious adverse events (SAEs). This may be confusing, as the trial definition of an SAE may differ from what a reader may understand as "serious" as it relates to an AE. Thus, despite somewhat uniform definitions of what defines an SAE, there is wide heterogeneity on what is reported as an SAE, at least in hematological malignancies.
Attribution process	• Classification that has insufficient sensitivity and reliability Reluctance of investigators to rule out possibility of causal relationship based on the wording of the present 5-tier scale for attribution (i.e., unrelated, unlikely to be related, possibly related, probably related, definitely related)
Utility of attribution	• Lack of clarity to trial sponsors and regulators on the utility of attribution of non-serious AEs. (Expedited reporting of serious AEs and suspected unexpected serious adverse reactions (SUSARs) does require attribution, however, to promote safety of patients on clinical trials.)

analysis of data from Phase III, randomized, double-blind, placebo-controlled trials, Hillman et al. [16]. found a relatively high frequency (up to 50%) of AE reports in placebo arms being reported as at least possibly related to treatment. It was also found that attribution of repeated AEs within the same patient (defined as a second or subsequent occurrence of the same event) changed over time. A subsequent pooled analysis now using data from nine randomized, multicenter, placebo-controlled trials, found that 7585% of all AEs that were deemed related to treatment were classified as only possibly related [7]. They concluded that AE causality determinations were often complex, unreliable, and subjective, suggesting that attribution should be eliminated in randomized double-blind placebo-controlled trials [7]. The working group recognizes that many of the challenges that contribute to the burden of attribution are inherent to the field and cannot easily be overcome, such as frequent overlap of drug toxicity and disease symptoms [5]. The major challenges identified by the working group are summarized in Table 2.

Recommendations for improving attribution

The working group identified actionable issues and proposes the following recommendations, summarized in Table 3.

Recommendation 1: improve attribution efficiencies

Recommendation 1A: collapse the current 5-tier AE attribution categories into a 2-tier or 3-tier system

There was strong consensus that the 5-tier system needs to be changed. In a study based on early-phase trials, Eaton et al. [17] found that toxicities rated as "possibly," "probably," or "definitely" related were associated with dose of study drug, whereas "unlikely" or "unrelated" toxicities were not. Based on this study, Eaton et al. [17]. recommended collapsing attribution categories. Also, multiple attribution stakeholders at the workshop indicated that the differences between "possibly related" and "probably related" categories are difficult to delineate, and that many investigators tend to avoid specifying an AE as unrelated, given that only rarely can one completely rule out the possibility that the drug contributed to the event.

We propose migrating from the 5-tier system to a 2-tier (related or unrelated) or 3-tier (related, unrelated, or unknown) system of attribution. This would simplify the attribution process without losing valuable information. Arguments in favor of a 2-tier system are that it forces investigators to commit to whether or not an AE is related to study treatment and that it may be consistent with what some investigators do on an instinctive basis. Arguments in favor of 3-tier system are that sometimes there may be insufficient information to guide the attribution and that an option to reflect this may be needed, particularly in early-phase development, when the knowledge of the agent's safety profile is limited. Thus, a 3-tier system may provide greater transparency. Also, any uncertainty that may be related to an individual physician or center, i.e., "center effects," will become more transparent and can then be documented and further analyzed. A 3-tier system may allow study designs to prescribe different trial consequences for AEs that are attributed to the unknown middle tier versus AEs that are likely related to study drug. It may further facilitate the principal investigator's assessment of sub-investigator attributions by allowing him or her to focus on the difficult ones, which would not be identified in a 2-tier system. For a 2-tiered system, more time and investigative effort may be needed to identify the true cause. On the other hand, a potential shortfall of the 3-tiered system is that the middle option could become a default, non-committal selection in some instances (for example, for interventions testing less-known drugs).

Regardless of whether a 3-tiered or 2-tiered system is selected, we propose the use of likelihood-based wording that clearly communicates that the attribution to be made is a probability assessment based on the investigator's current knowledge, and not the ultimate "true"

attribution. This approach would likely increase the quality of attribution by reducing the proportion of attributions that lean toward the "safe side." Wording for the 2-tier system could be: "more likely related to study drug (than other causes)" or "more likely related to other causes (than study drug)." Wording for the 3-tier system could be: "more likely related to study drug (than other causes)," "equally likely related to study drug and other causes," or "more likely related to other causes (than study drug)." The choice of the middle tier, "equally likely," could be made when there is insufficient information to make the call toward either side; it could be used when it is impossible to differentiate between drug toxicity that overlaps with symptoms of the disease under treatment (eg, fatigue, nausea, or myelosuppression).

Studies would be needed to compare the two- vs. three-tier system of toxicity attribution, and to also test different phrasings of the likelihood-based wording used to describe each tier.

Recommendation 1B: remove attribution of non-SAEs from randomized placebo-controlled trials

Attribution of AEs may be of less value for the regulatory benefitrisk evaluation of randomized placebo-controlled trials. For the protection of trial participants, SAEs still need to be attributed for potential expedited reporting. The working group proposes that AE attribution for non-SAEs should be eliminated in randomized double-blind placebo-controlled trials where objective data are available to determine the relatedness of AEs [7].

Recommendation 1C: for combination regimens, consider attribution in terms of the entire regimen

It is frequently not possible to attribute AEs to individual drugs in combination regimens, but it should always be possible to attribute toxicity to the combination regimen as a whole. This is in line with guidance in the EMA's revision 5 of the "Guideline on the evaluation of anti-cancer medicinal products in man," which encourages defining causality of AEs in relation to the overall combination treatment regimen being evaluated when definition of causality in relation to individual drugs may not be possible [18].

Recommendation 1D: define a process for removing an AE from the list of expected events

In early-phase trials, the inclination is to assume that all toxicities are from the agent under investigation, as it is difficult to rule out its lack of contribution. Conversely, in later-phase trials, AE causality should be reviewed and refined. A standard operating procedure for refining AE causality could be envisaged. An example is provided by the

Table 3
Recommendations for Improving Attribution.

1. Improving attribution efficiencies
    1. a Collapse the current 5-tier AE attribution categories into a 2-tier or 3-tier system and use likelihood-based wording to increase the sensitivity of attribution
    1. b Remove attribution of non-SAEs from randomized placebo-controlled trials
    1. c For combination regimens, consider attribution in terms of the entire regimen
    1. d Define a process for removing an AE from the list of expected events
2. Improve processes and tools for attribution
    2. a Require more clinically actionable information about expected toxicities
    2. b Establish online access to updated safety profile information
    2. c Establish standard operating procedures to facilitate and improve the quality of clinical research, especially baseline assessments
    2. d Promote the importance of experienced clinical trial investigators and trial sites
    2. e Explore the utility of patient-reported outcomes for capturing AEs at baseline and longitudinally
3. Improving the consistency of attribution
    3. a Educate those who conduct clinical trials on best practices for attribution
    3. b Standardize reporting of attribution in publications
    3. c Pursue harmonization of AE reporting recommendations among regulatory agencies

NCI's Cancer Therapy Evaluation Program (CTEP)'s Comprehensive Adverse Events and Potential Risks (CAEPR) source documents [19]. CAEPRs list expected toxicities for each drug as a guide to help the investigator with AE reporting; they are developed on the basis of the number of patients treated and allow for a more realistic assessment of AEs from an agent [19].

CTEP examines various items to develop a CAEPR, including the investigator's brochure, available animal data, safety communications, its own sponsored trial database, and publications [19]. Defining a process for when to remove an AE from the list of expected events would make attribution a fluid process as knowledge accumulates. This should be continued from first-in-human through Phase IV trials and incorporate prescribing data of drugs once approved, to avoid unfairly tagging drugs in perpetuity with events of dubious significance or association.

Additional considerations on improving attribution efficiencies

In addition, the working groups discussed the possibility to limit the reporting of lower grade AEs in order to facilitate the attribution process. However, although lower-grade AEs have often been regarded as clinically less important in the past, and therefore of less interest to collect, this may not hold true for newer agents, such as targeted and immunotherapies and oral agents that are meant for long-term chronic administration, when low-grade adverse drug reactions (such as fatigue and diarrhea) can have a major impact on overall tolerability and the possibility to maintain an efficacious dose-intensity.

Recommendation 2: Improve processes and tools for attribution

Recommendation 2A: require more clinically actionable information about expected toxicities

As part of best practices, investigators need either more clinically relevant information or greater understanding of how to interpret the available pre-clinical data in order to make robust calls on attribution. Although the protocol and the investigator's brochure are required to have preclinical data and information on anticipated AEs, toxicities, and symptoms based on drugs of a similar action, the working group suggests that these resources could be augmented with enhanced discussion on the interpretation of such data in the context of the anticipated presentation of symptoms/toxicities in the patient. Easier access to such information could be provided by sponsors to investigators by digitizing the investigator's brochure to make it searchable. Additionally, the provision in a standardized format of comprehensive lists of reported and potential AEs associated with an investigational agentsimilar to the CAEPR list required by the NCI for CTEP-sponsored clinical trialscould be of value [19]. Better and more standardized, digitalized databases of drugs with potential drugdrug interactions (such as strong inhibitors and inducers of CYP3 and others) also could be included.

Recommendation 2B: establish online access to updated safety profile information

The use of integrated electronic systems with possibilities for a bidirectional flow of safety information could be very useful. For example, investigators and clinical staff in early-phase trials could use a software system to collect and record, as appropriate, patient characteristics, safety and accrual data, patient-reported outcomes (PROs), and laboratory data essential for AE determination, and to report to sponsors safety and accrual data in a more efficient and accurate manner [20]. At the same time, sponsors should be able to give investigators real-time access to cumulative summary data for AEs associated with an experimental therapy from all sites and/or all clinical trials using that therapy. The availability of an online, searchable drug-safety database with expected AEs and observed toxicities that is continuously updated and always accessible may improve the quality of attribution decisions and enhance patient safety during a trial.

Recommendation 2C: establish standard operating procedures to facilitate and improve the quality of clinical research, especially baseline assessments

Stakeholders expressed a clear need for a more standardized collection of baseline data and measurements across institutions to improve attribution, to accurately establish whether a patient's clinical status is stable, worsening, or improving. The standardized baseline assessment and documentation of clinical data can include PROs, clinical laboratory values (such as hepatic enzymes, hemoglobin, fasting blood sugar, blood lipid levels, and blood cell counts), comorbidities, AEs, previous cancer treatments, and current medications, and may use tools such as CAEPRs to help with attribution assessments. A mechanism would ideally include collecting both solicited and unsolicited AEs. Factors to be measured at baseline could be based on preclinical animal toxicity data, expected AEs for the drug class (for example, rash, diarrhea, and pulmonary symptoms may be expected for immune checkpoint inhibitors), and symptoms and AEs that are most commonly seen in early-phase clinical trials. Baseline standardization could include a core set of factors that are measured at baseline in all early-phase trials, the number of baseline assessments, definition of baseline days, and the grading criteria to be used.

Recommendation 2D: promote the importance of experienced clinical trial investigators and trial sites

The quality and experience of a clinical research center and its investigators are recognized as important components of successful AE reporting, including high-quality attribution, particularly in early-phase trials. Changes in the goals, populations, and conduct of early-phase trials have resulted in a shift towards multi-institutional trials and centralized study management by contract research organizations instead of research centers. A disadvantage of this shift is that if too many sites are involved in a single clinical trial, each participating site may contribute only a limited number of patients, thus resulting in investigators at each site having limited experience with the experimental agent and consequently little sense of the toxicities that may be associated with it.

It has been shown that the ability of Phase I trials to predict clinically relevant toxicities in later-phase trials increases as the number of patients on the initial Phase I trial increases (up to 60 patients) [21]. Phase I trials would therefore need sufficient numbers of patients and clear expansion numbers to accurately define and attribute toxicities before Phase II trials were commenced. Limiting the number of sites and ensuring sufficient numbers of patients at each site, whenever feasible, would increase site investigators' experience with the agent and with observing toxicities. When this is not possible, a study management committee can serve as an advisory group to help with attributions and discussion of options.

Recommendation 2E: explore the utility of PROs for capturing AEs at baseline and longitudinally

Scheduled systematic assessment of symptomatic AEs by PROsfor example, at baseline and over time (longitudinal symptom trajectories) - may aid investigators in assigning attribution and grading severity [22,23]. PROs would be very helpful for identifying important symptomatic toxicities that are best reported by the patient (such as fatigue, pain, nausea, and neuropathy). The feasibility of real-time PRO data collection in trial settings has been demonstrated [24,25]. Nonetheless, although PRO data can enhance the detection of AEs, the working group was quite clear that the attribution of symptomatic toxicities should remain the responsibility of the investigator. Patients can report symptoms, but the group agreed that patients should not normally make an attribution for these symptoms. The subjective attribution of AEs by patients themselves might be a topic of interest for academic research, but patient attribution of AEs should not be required for drug development.

Recommendation 3: improve the consistency of attribution

Recommendation 3A: educate those who conduct clinical trials on best practices for attribution

Educating investigators, sponsors, and other professionals involved in clinical trials on best practices and regulations related to attribution will improve the consistency of this process. Including attribution in educational activities, such as the American Association for Cancer Research (AACR)/ASCO Methods in Clinical Cancer Research workshop, and developing content (for example, by using case studies from previous trials) that will better prepare physician-researchers to make accurate and consistent attribution. Mandatory periodic courses that focus on the regulatory framework and Good Clinical Practice guidelines could also be included to ensure that investigators and sponsors comply with relevant regulatory requirements and are up-to-date with changes in best practices [26].

This could be supported by developing an attribution webinar by the FDA, NCI, and possibly even ASCO or AACR, that could be updated periodically. The webinar would include information on how to assess for attribution, how to conduct attribution assessments, and how to report to regulatory agencies. Physician investigators could be required to attend the webinar and to repeat it on a reasonable basis as part of 1572 certification or Good Clinical Practice requirements. For research conducted in other parts of the world (e.g., as in multicenter international trials), sponsors could play the role of ensuring that investigators are educated by making this information a condition for participation in conducting the trial.

For investigators involved in clinical research, engagement with sponsors by participating in safety conference calls is highly recommended, as knowledge and expertise accumulate over the course of a clinical trial. Moreover, a minimum requirement to join safety calls should be employed, as the information shared during these calls can have major implications for the attribution process and MTD definition and can improve patients' safety and experience during the clinical trial.

Recommendation 3B: standardize reporting of attribution in publications

Understanding of the relatedness (attribution) of an AE to a drug will be greatly enhanced by an insistence on standardized reporting of AEs in publications and clinical scientific reports. Reporting of all-cause and treatment-emergent AEs above a certain threshold (for example, for Phase I trials, treatment-emergent AEs exceeding 10% incidence), in addition to AEs of special interest, should be required in publications and clinical scientific reports to help build consistency and standardization in AE reporting. Reporting of all-cause AE frequencies (rather than, or in addition to, treatment-related frequencies) are suggested as all-cause AE frequencies are the measure least likely to be biased by pre-existing understanding [18]. The type and frequency of AEs leading to dose reduction, dose interruption, or permanent treatment discontinuation should also be reported in papers and clinical scientific reports, as should SAEs and deaths. Also, for phase I trials, AEs > 10% should be reported not only for first cycle but for all cycles. Others have also called for more transparent reporting of the seriousness of adverse events in oncology clinical trials [27]. Some have suggested that it would be a great benefit if trialists could provide evidence users with some level of certainty about whether an AE was caused by the intervention. The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) criteria have been used in certain settings to provide stakeholders and decision makers (for example, patients, physicians and policy makers) with levels of certainty regarding evidence presented [28,29]. However, early-phase clinical trials may represent too early a setting to gauge certainty of evidence for whether or not an AE was caused by a drug, as the number of patients involved in an early-phase trial are typically relatively small. The rating of certainty of evidence may be more feasible and practical at the regulatory level when larger amounts of data are available, for example, by combining

data from early-phase and later-phase clinical trials, as well as post-marketing safety reports. On a related note, there have been recent calls in the literature that the FDA require reports of SAEs associated with black-box warnings to include descriptions of certainty of evidence as a guide to support decision making and implementation [30].

Recommendation 3C: pursue harmonization of AE reporting recommendations among regulatory agencies

Increased harmonization and synchronization of attribution/causality-based safety reporting requirements between international regulatory agencies, such as the FDA, EMA, and others, will greatly facilitate consistent reporting of causality to these various agencies and will increase efficiency, improve compliance [3], and lower costs related to different reporting requirements.

Summary and conclusions

The working group has identified a number of challenges with current safety attribution processes and present 3 broad recommendations (summarized in Table 3) that may streamline and facilitate the attribution process and increase its value. We propose a move from the present 5-tier system of AE reporting to a 2–3-tier system. We also recommend that oncologic societies such as ASCO and AACR move toward developing recommendations and training to improve attribution in clinical trials, baseline assessments, and the logistical ways in which protocols are designed. Finally, we suggest that journals incorporate consistent, standardized requirements for reporting attribution in oncology clinical trials.

Funding

Conflicts of interest (COI)

Alicyn Campbell: Employee of Genentech, a member of the Roche group while conducting this work. **Jorge E. Cortes:** Grants (Institution): BMS, Novartis, Pfizer, Astellas, Daiichi, Takeda, Immunogen, Arog, Amphivena, BergenBio, Merus. Consulting (Personal): BMS, Novartis, Pfizer, Astellas, Daiichi, Takeda. **David M. Hyman:** Consulting/Advisory Role: Chugai Pharma, CytomX Therapeutics, Boehringer Ingelheim, AstraZeneca, Pfizer, Bayer Pharmaceuticals, Genentech / Roche; Research Funding: Loxo Oncology, PUMA Biotechnology, AstraZeneca, Bayer Pharmaceuticals (does not include industry-sponsored clinical trials). **Thomas Karagiannis:** Employee and Stock Ownership in Genentech. **Patricia LoRusso:** Data Safety Monitoring Board/Committee: Agios, FivePrime; Advisory Board Member: Alexion, Ariad, GenMab, Glenmark, Halozyme: Menarini, Novartis, Genentech, CytomX, Omniox, Ignyta, Takeda; Consultant: SOTIO, Cybrexa, Agenus. **Gilbert Y. Wong:** Employee and stock shareholder of Pfizer, Inc. **Chris H. Takimoto:** Stock holder and current employee of Forty Seven, Inc; Stock holder and former employee of Johnson & Johnson. **Timothy A. Yap:** Employment: Medical Director of the Institute for Applied Cancer Science and Associate Director for Translational Research of the Institute for Personalized Cancer Therapy at the University of Texas MD Anderson Cancer Center; Previous employee of the Institute of Cancer Research, London, England; Research support: AstraZeneca, Bayer, Pfizer, Tesaro, Jounce, Eli Lilly, Seattle Genetics, Kyowa, Constellation, and Vertex Pharmaceuticals Consultancies: Aduro, Almac, AstraZeneca, Atrin, Bayer, Bristol-Myers Squibb, Calithera, Clovis, Cybrexa, EMD Serono, Ignyta, Jansen, Merck, Pfizer, Roche, Seattle Genetics, and Vertex Pharmaceuticals; Speaker bureau: AstraZeneca, Merck, Pfizer,

References

[1] Cleeland CS, Allen JD, Roberts SA, Brell JM, Giralt SA, Khakoo AY, et al. Reducing the toxicity of cancer therapy: recognizing needs, taking action. Nat Rev Clin Oncol 2012;9:471–8.

[2] Kluetz PG, Kanapuru B, Lemery S, Johnson LL, Fiero MH, Arscott K, et al. Informing the tolerability of cancer treatments using patient-reported outcome measures: summary of an FDA and critical path institute workshop. Value Health J. Int. Soc. Pharm. Outcomes Res. 2018;21:742–7.

[3] Levit LA, Perez RP, Smith DC, Schilsky RL, Hayes DF, Vose JM. Streamlining adverse events reporting in oncology: an american society of clinical oncology research statement. J Clin Oncol 2018;36:617–23.

[4] FDA. Questions and answers on FDA's Adverse Event Reporting System (FAERS). 2018.

[5] Thanarajasingam G, Minasian LM, Baron F, Cavalli F, De Claro RA, Dueck AC, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. Lancet Haematol 2018.

[6] Cambridge English Dictionary 4ed. Cambridge: Cambridge University Press; 2013.

[7] Le-Rademacher J, Hillman SL, Meyers J, Loprinzi CL, Limburg PJ, Mandrekar SJ. Statistical controversies in clinical research: Value of adverse events relatedness to study treatment: analyses of data from randomized double-blind placebo-controlled clinical trials. Ann Oncol 2017;28:1183–90.

[8] Iasonos A, Gounder M, Spriggs DR, Gerecitano JF, Hyman DM, Zohar S, et al. The impact of non-drug-related toxicities on the estimation of the maximum tolerated dose in phase I trials. Clin Cancer Res 2012;18:5179–87.

[9] Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 2009;101:708–20.

[10] Zohar S, O'Quigley J. Sensitivity of dose-finding studies to observation errors. Contemp Clin Trials 2009;30:523–30.

[11] Trotti A, Colevas AD, Setser A, Rusch V, Jaques D, Budach V, et al. 0: development of a comprehensive grading system for the adverse effects of cancer treatment. Semin Radiat Oncol 2003;13:176–81.

[12] NCI. Common Terminology Criteria for Adverse Events (CTCAE). 2018.

[13] Committee IS. ICH harmonised tripartite guideline. Clinical safety data management: definitions and standards for expedited reporting E2A. Step 4 version. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use1994.

[14] Sivendran S, Latif A, McBride RB, Stensland KD, Wisnivesky J, Haines L, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol. 2014;32:83–9.

[15] CTTI. Project: Safety Reporting. 2018.

[16] Hillman SL, Mandrekar SJ, Bot B, DeMatteo RP, Perez EA, Ballman KV, et al. Evaluation of the value of attribution in the interpretation of adverse event data: a North Central cancer treatment group and american college of surgeons oncology group investigation. J Clin Oncol 2010;28:3002–7.

[17] Eaton A, Iasonos A, Gounder MM, Pamer EG, Drilon A, Vulih D, et al. Toxicity attribution in phase I trials: evaluating the effect of dose on the frequency of related and unrelated toxicities. Clin Cancer Res 2016;22:553–9.

[18] EMA. Guideline on the evaluation of anticancer medicinal products in man. London: Committee for Medicinal Products for Human Use (CHMP). 2017.

[19] National Cancer Institute. NCI Guidelines For Investigators: Adverse Event Reporting Requirements for DCTD (CTEP and CIP) and DCP INDs and IDEs. 2013.

[20] Pietanza MC, Basch EM, Lash A, Schwartz LH, Ginsberg MS, Zhao B, et al. Harnessing technology to improve clinical trials: study of real-time informatics to collect data, toxicities, image response assessments, and patient-reported outcomes in a phase II clinical trial. J Clin Oncol 2013;31:2004–9.

[21] Jardim DL, Hess KR, Lorusso P, Kurzrock R, Hong DS. Predictive value of phase I trials for safety in later trials and final approved dose: analysis of 61 approved cancer drugs. Clin Cancer Res 2014;20:281–8.

[22] Banerjee AK, Okun S, Edwards IR, Wicks P, Smith MY, Mayall SJ, et al. Patient-reported outcome measures in safety event reporting: PROSPER consortium guidance. Drug Saf 2013;36:1129–49.

[23] Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and reliability of the US National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). JAMA Oncol 2015;1:1051–9.

[24] Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. J Clin Oncol 2016;34:557–65.

[25] Basch E, Dueck AC. Patient-reported outcome measurement in drug discovery: a tool to improve accuracy and completeness of efficacy and safety data. Expert Opin Drug Discov 2016;11:753–8.

[26] Food and Drug Administration. Best Practices for Communication Between IND Sponsors and FDA During Drug Development Guidance for Industry and Review Staff 2018.

[27] Gyawali B, Shimokata T, Honda K, Ando Y. Reporting harms more transparently in trials of cancer drugs. BMJ 2018;363:k4383.

[28] Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017;87:4–13.

[29] Murad MH, Mustafa RA, Schunemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. Evidence-Based Med 2017;22:85–7.

[30] Elraiyah T, Gionfriddo MR, Montori VM, Murad MH. Content, consistency, and quality of black box warnings: time for a change. Ann Intern Med 2015;163:875–6.

[31] Perez R, Archdeacon P, Roach N, Goodwin R, Jarow J, Stuccio N, et al. Sponsors' and investigative staffs' perceptions of the current investigational new drug safety reporting process in oncology trials. Clin Trials (London, England) 2017;14:225–33.

[32] Health UDo, Services H, Food, Administration D. Guidance for Industry and Investigators Safety Reporting Requirements for INDs and BA/BE Studies. 2010

[33] Crepin S, Villeneuve C, Merle L. Quality of serious adverse events reporting to academic sponsors of clinical trials: far from optimal. Pharmacoepidemiol Drug Saf 2016;25:719–24.

[34] 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] Lineberry N, Berlin JA, Mansi B, Glasser S, Berkwits M, Klem C, et al. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. BMJ 2016;355.

WILEY

RESEARCH ARTICLE

Tumor mutational burden standardization initiatives: Recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions

Albrecht Stenzinger[1] | Jeffrey D. Allen[2] | Jörg Maas[3] | Mark D. Stewart[2] |
Diana M. Merino[2] | Madison M. Wempe[2] | Manfred Dietel[4]

[1]Institute of Pathology, University Hospital Heidelberg (on behalf of QuIP® GmbH, Berlin, Germany), Heidelberg, Germany

[2]Science Policy, Friends of Cancer Research, Washington, District of Columbia

[3]Management Department, QuIP® GmbH, Berlin, Germany

[4]Institute of Pathology, Charité Berlin (on behalf of QuIP® GmbH, Berlin, Germany), Berlin, Germany

Correspondence
Albrecht Stenzinger, Institute of Pathology, University Hospital Heidelberg, Im Neuenheimer Feld 672, 69120 Heidelberg, Germany.
Email: albrecht.stenzinger@med.uni-heidelberg.de
and
Jeffrey D. Allen, Friends of Cancer Research, 1800 M Street NW, Suite 1050 South, Washington, DC 20036, USA.
Email: jallen@focr.org

Abstract

Characterization of tumors utilizing next-generation sequencing methods, including assessment of the number of somatic mutations (tumor mutational burden [TMB]), is currently at the forefront of the field of personalized medicine. Recent clinical studies have associated high TMB with improved patient response rates and survival benefit from immune checkpoint inhibitors; hence, TMB is emerging as a biomarker of response for these immunotherapy agents. However, variability in current methods for TMB estimation and reporting is evident, demonstrating a need for standardization and harmonization of TMB assessment methodology across assays and centers. Two uniquely placed organizations, Friends of Cancer Research (Friends) and the Quality Assurance Initiative Pathology (QuIP), have collaborated to coordinate efforts for international multistakeholder initiatives to address this need. Friends and QuIP, who have partnered with several academic centers, pharmaceutical organizations, and diagnostic companies, have adopted complementary, multidisciplinary approaches toward the goal of proposing evidence-based recommendations for achieving consistent TMB estimation and reporting in clinical samples across assays and centers. Many factors influence TMB assessment, including preanalytical factors, choice of assay, and methods of reporting. Preliminary analyses highlight the importance of targeted gene panel size and composition, and bioinformatic parameters for reliable TMB estimation. Herein, Friends and QuIP propose recommendations toward consistent TMB estimation and reporting methods in clinical samples across assays and centers. These recommendations should be followed to minimize variability in TMB estimation and reporting, which will ensure reliable and reproducible identification of patients who are likely to benefit from immune checkpoint inhibitors.

KEYWORDS

biomarkers, immune checkpoint inhibitors, neoantigens, next-generation sequencing, tumor mutational burden/load

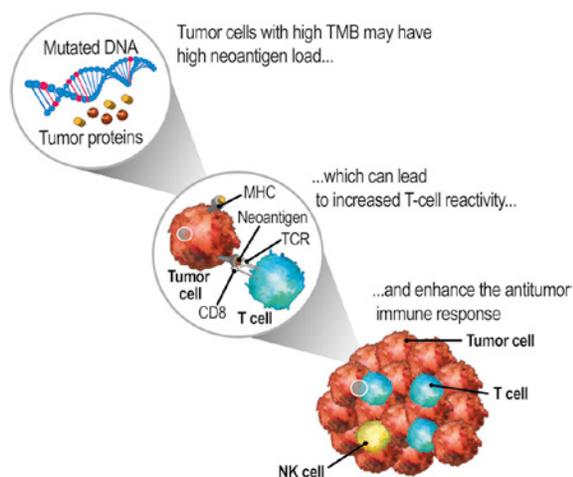Albrecht Stenzinger and Jeffrey D. Allen are co-first authors

1 | TUMOR MUTATIONAL BURDEN AS A BIOMARKER OF RESPONSE TO IMMUNE CHECKPOINT INHIBITORS

Tumor mutational burden (TMB) is the total number of somatic mutations in a defined region of a tumor genome and varies according to tumor type as well as among patients.[1–4] For some tumors, particularly those with high TMB, such as melanoma and lung cancers, evidence is emerging for the association of TMB with neoantigen load.[2–5] Neoantigens are novel tumor cell surface epitopes, some of which can be recognized as foreign to the body by the immune system, resulting in increased T-cell reactivity and thereby leading to an antitumor immune response (Figure 1).[1,4,6–9] Immune checkpoint inhibitors enhance antitumor T-cell activity via inhibition of immune checkpoint molecules, such as programmed death-1/programmed death ligand-1 (PD-1/PD-L1) and cytotoxic T lymphocyte antigen-4 (CTLA-4), which negatively regulate T-cell activation and contribute to tumor immune response evasion.[10–12] Therefore, for some tumor types, neoantigen load or TMB may be a suitable clinical biomarker to guide treatment decisions for immune checkpoint inhibitors. While not all mutations result in immunogenic neoantigens and determining which mutations are likely to induce immunogenic neoantigens remains a challenge, TMB represents a quantifiable measure of the number of mutations in a tumor that can be used to inform treatment selection.[4] Clinical data demonstrating that patients with tumors that have high neoantigen load or high TMB are more likely to achieve clinical benefit from treatment with immune checkpoint inhibitors are accumulating.[1,13–15]

Investigation of TMB as a biomarker of response to immune checkpoint inhibitors has increased over recent years. These studies have identified an association between elevated TMB and improved patient outcomes in response to anti-PD-1/PD-L1 and anti-CTLA-4 therapies in multiple tumor types.[16–25] Most studies to date have investigated the association of patient outcomes and TMB in patients with non-small cell lung cancer (NSCLC). Other studies have assessed this association in patients with melanoma, squamous cell carcinoma of the head and neck, small cell lung cancer, and urothelial carcinoma. Data from retrospective or exploratory analyses indicate that TMB may be an independent biomarker for clinical efficacy of PD-1/PD-L1 and CTLA-4 inhibitors.[16–20,24,26–29] These observations were recently corroborated in clinical studies in patients with NSCLC treated with nivolumab in combination with ipilimumab and with atezolizumab, where high TMB (defined as ≥10 mutations per megabase [mut/Mb] and ≥14 mut/Mb, respectively) was prospectively assessed as clinically predictive for increased progression-free survival.[21,23] The escalation of published studies in 2017 and 2018 compared with previous years demonstrates the increased awareness of assessing TMB as a predictive marker for response to immune checkpoint inhibitors, a trend that is set to continue.

2 | THE FUTURE CLINICAL LANDSCAPE OF TMB

Alongside data from published studies demonstrating the association of TMB and response to immune checkpoint inhibitors, additional ongoing and planned clinical trials with a key TMB component in their design are emerging.[16–25,30] A search of the United States-focused ClinicalTrials.gov database (search terms "tumor mutation burden", "tumor mutational burden", "tumor mutation load", "tumor mutational load" [performed July 26, 2018]) demonstrates that the number of trials with key TMB components (defined as TMB assessment listed under study description, study design, outcome measures, or eligibility criteria) has greatly increased from 1 in 2014 to 35 in 2017, and the data for the first half of 2018 continue to follow the trend (14 trials from January 1 to July 26, 2018). Fifty-four trials were identified in the search, which have a total estimated enrollment of over 11 000 patients, and their projected primary completion dates suggest that patient TMB data will continue to accumulate through 2019 and beyond. Of the 54 trials, 37 investigate immune checkpoint inhibitors and TMB, and findings show that integration of TMB as a biomarker for response to immune checkpoint inhibitors in clinical trials is diversifying from mostly melanoma and NSCLC trials into a range of other tumor types, including endometrial, colorectal, urothelial, and breast cancers.[25] These findings underlie expectations that diagnostic assessment of TMB could provide benefit across many tumor types. Furthermore, as is common in the field of precision medicine, TMB assessment can be included in clinical trials as part of multiparameter assessments encompassing potential protein, DNA, and RNA biomarkers. While TMB represents one aspect of the genomic landscape, whole genome or exome and RNA sequencing may reveal functional aspects of the tumor profile, such as targetable gene mutations and/or fusions, and assessment of protein markers in the tumor microenvironment may provide additional information. Therefore, multiomic analyses may provide a more complete patient biomarker profile for guiding treatment decisions. Indeed, other biomarkers commonly investigated alongside TMB include PD-L1, microsatellite instability (MSI) or deficient mismatch repair, and immune signatures.[31]

FIGURE 1 TMB association with the antitumor response. Abbreviations: CD8, cluster of differentiation 8; MHC, major histocompatibility complex; NK, natural killer; TCR, T-cell receptor

The increase in integration of TMB assessment in ongoing and upcoming clinical trials investigating immune checkpoint inhibitors demonstrates increased awareness of TMB as a potential clinical biomarker for guiding patient treatment decisions and identifying patients likely to benefit from these therapies. It also brings to the forefront the crucial need for clinicians to be aware of different TMB methodologies and reporting so that they may make informed clinical decisions.

3 | THE NEED FOR STANDARDIZATION AND HARMONIZATION OF TMB ASSESSMENT IN CLINICAL SAMPLES

TMB is most commonly measured by assessing formalin-fixed, paraffin-embedded (FFPE) tissue samples using next-generation sequencing (NGS) methods, whole genome sequencing (WGS), whole exome sequencing (WES), and various targeted gene panels. With advances in technology enabling targeted gene panel assays to be performed more affordably, with quick turnaround times, and with increased assay sensitivity that enables analyses of small biopsy samples or those with low tumor cellularity, such assays are increasingly being used to assess TMB, MSI status, and other genomic biomarkers.[4,32,33] The FDA recently granted the genomic profiling assays FoundationOne CDx and MSK-IMPACT approval and authorization, respectively, as tests for actionable mutations, copy number alterations, and fusions in solid tumors.[33-36] Although not yet approved for such use in the clinical setting, these assays can be used for TMB assessment. Furthermore, several targeted gene panel assays are currently being developed and validated by diagnostic companies and academic institutes, including some specific for TMB assessment in blood.

The increase in TMB assessment by various methods has brought with it a confusing array of information that documents how TMB has been determined and reported. The wide variation in TMB estimation and reporting methods across studies that have already been published demonstrates an evident lack of standardization and harmonization of current TMB assessment methods (Table 1).[16-24,26-29,37-39] These extensive differences may arise from the theoretic framework, technical methods applied, and the way that TMB data are reported, and will be described in more detail in the later sections of this article.

Together, the increased interest in using TMB to select patients who will most likely benefit from immune checkpoint inhibitors, increased integration of TMB assessment in ongoing clinical trials, and variability in current TMB assessment methods can create confusion for physicians and may influence critical treatment decisions. Further investigation is warranted to assess how these methods compare with one another and highlights the need for standardization and harmonization efforts for TMB estimation and reporting across assays and centers.[4,40-42] Standardization of TMB assessment methodology will ensure consistency of TMB estimation and reporting across assays and centers, and harmonization will enable TMB score to be more accurately compared across assays and centers. It has been recognized that the need for standardized and harmonized methodology for clinical assays can be addressed by the collaborative efforts of accredited agencies, pathologists, and oncologists. Here, we describe the

Friends of Cancer Research (Friends) and Quality Assurance Initiative Pathology (QuIP), which have collaborated to coordinate efforts to address this need for standardized and harmonized TMB estimation and reporting in clinical samples. Friends and QuIP are well placed to coordinate the international multistakeholder initiatives to understand the differences in TMB assessment methodology and propose approaches that will standardize and harmonize TMB assessment across assays and centers globally.

4 | FRIENDS AND QuIP TMB STANDARDIZATION AND HARMONIZATION INITIATIVES

The international collaboration between Friends and QuIP has been initiated to propose recommendations for achieving consistency in TMB estimation and reporting in clinical tissue samples across different assays, platforms, and centers (Figure 2A). Using multidisciplinary approaches, Friends and QuIP review the current methods of TMB assessment in FFPE samples and propose recommendations on how to standardize them (Figure 2B). These recommendations will inform the oncology community (including diagnostic companies, pathologists, clinicians, and the pharmaceutical industry) of best practices for TMB assessment in FFPE samples and ultimately will improve patient care by guiding treatment decisions and enabling maximum clinical benefit for patients.

Friends is a nonprofit and patient advocacy organization, founded in 1996 and based in Washington, DC, that drives collaboration between partners across diverse healthcare sectors to drive advances in science, policy, and regulation that advance treatments in patients. The organization has been instrumental in the development and implementation of policies that ensure patients quickly receive the best treatments in the safest way possible. QuIP, founded in 2004, is a joint venture between the German Society of Pathology and the German Pathologists' Association, encompassing specialists from the fields of pathology, quality management, administration, and marketing/public relations, that provides and studies pathological testing services. The organization values continued education and training for pathologists and the highest standards of quality assurance to ensure that patients receive optimal personalized treatment. Friends and QuIP are therefore uniquely positioned to perform these collaborative initiatives and provide evidence-based recommendations for reliable and reproducible TMB assessment in clinical samples across assays and centers.

Friends and QuIP have partnered with a number of academic institutes and diagnostic and pharmaceutical companies, bringing together key experts from diverse backgrounds from around the world, including pathologists, bioinformaticians, physicians, drug sponsors and regulators, diagnostic assay developers, patient advocates, and healthcare policy advisors, to achieve the coordinated goal of proposing recommendations for TMB assessment in FFPE clinical samples (Figure 2A).[43,44] Complementary analytical and clinical approaches have been adopted by the two organizations to provide a breadth of data to ensure that the recommendations proposed by Friends and
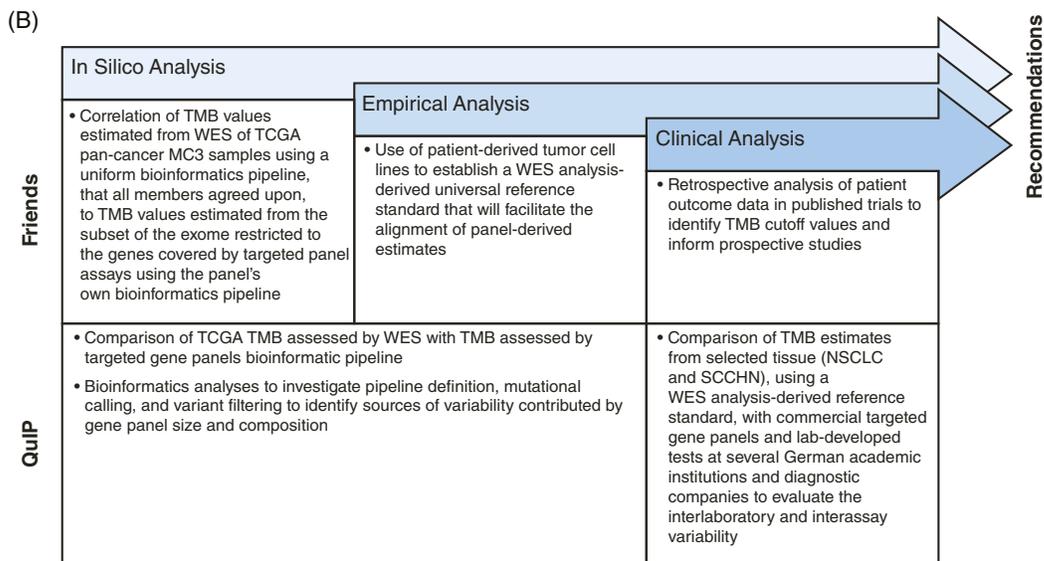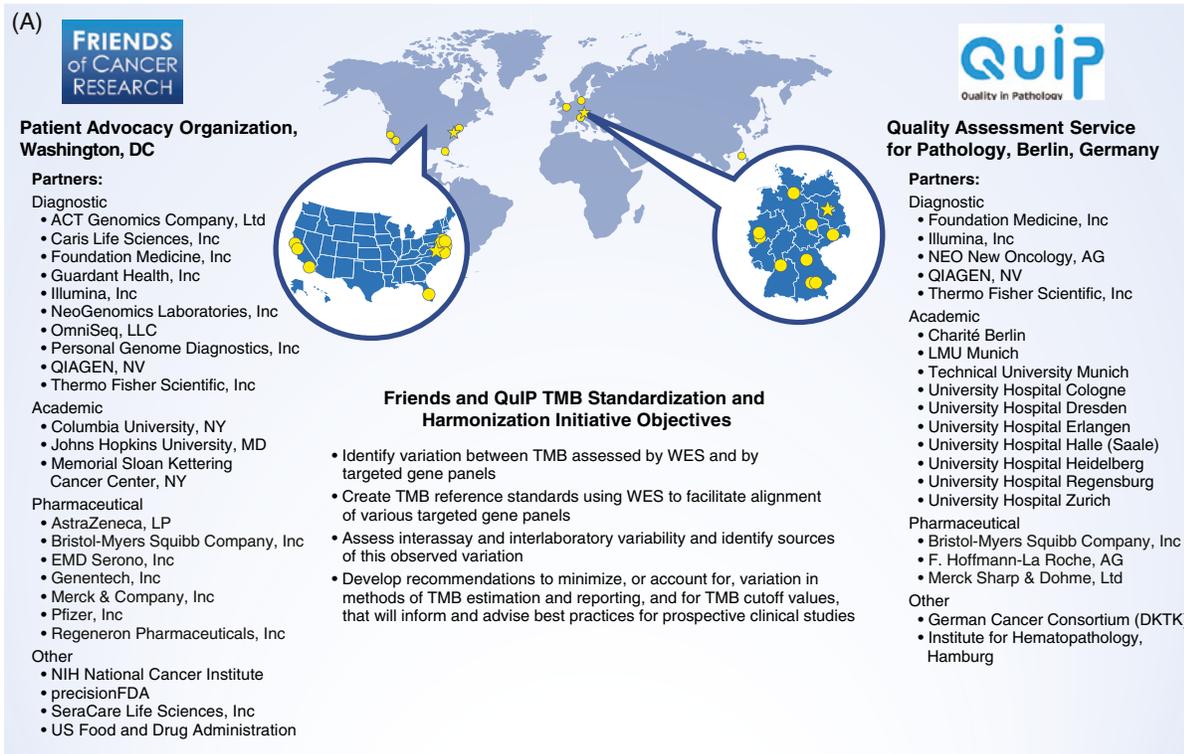
TABLE 1 Methodology for published key trials demonstrating TMB as a biomarker of clinical response to immune checkpoint inhibitors

Study name (NCT number)	Tumor type and therapy agent	Methodology	Reporting	Cutoff for high TMB
KEYNOTE-001[16] (NCT01295827)	NSCLC Pembrolizumab	WES • SureSelect All Exon v2 • Illumina HiSeq 2000 • VAF = 10%	Somatic coding nonsynonymous mutations per exome	≥178 mutations
POPLAR, FIR, and BIRCH[17] (NCT02031458 NCT01846416 NCT01903993)	NSCLC Atezolizumab	FoundationOne assay[4] • 315 genes assessed • 1.1 Mb coverage	Somatic coding SNVs (synonymous and nonsynonymous) and indels per megabase	≥75th percentile (≥13.5 mut/Mb for first line and ≥17.1 mut/Mb or ≥15.8 mut/Mb for second line populations)
CheckMate 026[18] (NCT02041533)	NSCLC Nivolumab	WES • AllPrep DNA isolation (tumor tissue)/QIAamp DNA isolation (blood) • SureSelect All Exon v5 • Illumina HiSeq 2500	Total somatic missense mutations per sample (tumor and blood)	Upper tertile (≥243 mutations)
KEYNOTE-012 and KEYNOTE-028[19,29] (NCT01848834 NCT02054806)	Solid tumors Pembrolizumab	WES Details not specified	Somatic coding nonsynonymous mutations per exome	≥102 mutations
IMvigor 210[26,27] (NCT02108652)	UC Atezolizumab	FoundationOne assay-based panel • 315 genes assessed	Somatic coding SNVs (synonymous and nonsynonymous) and indels per megabase	>16 mut/Mb
POPLAR and OAK[20,24] (NCT01903993 NCT02008227)	NSCLC Atezolizumab	bTMB assay (based on the FoundationOne assay)[24] • 394 genes assessed • 1.1 Mb coverage • Illumina HiSeq 4000 • VAF ≥0.5%	Total somatic SNVs (synonymous and nonsynonymous) per assay	≥14 mut/Mb
CheckMate 032[38] (NCT01928394)	SCLC Nivolumab ± ipilimumab	WES • AllPrep DNA isolation (tumor tissue)/QIAamp DNA isolation (blood) • SureSelect All Exon v5 • Illumina HiSeq 2500	Somatic missense mutations per exome	Upper tertile (≥248 mutations)
CheckMate 012[28] (NCT01454102)	NSCLC Nivolumab + ipilimumab	WES • SureSelect All Exon v2, v4, or Nextera Rapid Capture Exome kit • Illumina HiSeq 2000, 2500, or 4000 • VAF = 5%	Nonsynonymous mutations (SNVs or indels) per exome	Upper tertile (not specified), median (>158 mutations), or upper quartile (≥307 mutations)
CheckMate 038[39] (NCT01621490)	Melanoma Nivolumab ± ipilimumab	WES • SureSelect All Exon v2 • Illumina HiSeq 2000 or 2500 • Allele read count ≥5	Nonsynonymous mutations (SNVs or indels) per exome	100 mutations
CheckMate 275[37] (NCT02387996)	UC Nivolumab	WES • Details not specified	Somatic missense mutations per tumor	Upper tertile (≥167 mutations)
CheckMate 227 and CheckMate 568[21,22] (NCT02477826 NCT02659059)	NSCLC Nivolumab and ipilimumab	FoundationOne CDx assay[34] • 324 genes assessed • 0.8 Mb coverage • Illumina HiSeq 4000 • VAF = 5%	Somatic SNVs (synonymous and nonsynonymous) and indels per megabase	≥10 mut/Mb
B-F1RST[23,24] (NCT02848651)	NSCLC Atezolizumab	bTMB assay (based on FoundationOne) • 394 genes assessed • 1.1 Mb coverage • Illumina HiSeq 4000 • VAF ≥0.5%	Total somatic SNVs (synonymous and nonsynonymous) per assay	≥14 mut/Mb

Abbreviations: bTMB, blood tumor mutational burden; indels, short insertions and deletions; mut/Mb, mutations per megabase; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; SNV, single nucleotide variant; TMB, tumor mutational burden; UC, urothelial carcinoma; VAF, variant allele frequency; WES, whole exome sequencing

both Friends and QuIP have utilized publicly available data from The Cancer Genome Atlas to compare TMB values derived using WES, which is currently considered as the gold standard for calculating TMB, with those calculated using targeted gene panels. Both these approaches have focused on identifying factors that contribute to variation in TMB calculation and harmonizing bioinformatic pipelines.

(A)

FRIENDS of CANCER RESEARCH

Patient Advocacy Organization, Washington, DC

Partners:

Diagnostic
• ACT Genomics Company, Ltd
• Caris Life Sciences, Inc
• Foundation Medicine, Inc
• Guardant Health, Inc
• Illumina, Inc
• NeoGenomics Laboratories, Inc
• OmniSeq, LLC
• Personal Genome Diagnostics, Inc
• QIAGEN, NV
• Thermo Fisher Scientific, Inc

Academic
• Columbia University, NY
• Johns Hopkins University, MD
• Memorial Sloan Kettering Cancer Center, NY

Pharmaceutical
• AstraZeneca, LP
• Bristol-Myers Squibb Company, Inc
• EMD Serono, Inc
• Genentech, Inc
• Merck & Company, Inc
• Pfizer, Inc
• Regeneron Pharmaceuticals, Inc

Other
• NIH National Cancer Institute
• precisionFDA
• SeraCare Life Sciences, Inc
• US Food and Drug Administration

QuIP Quality in Pathology

Quality Assessment Service for Pathology, Berlin, Germany

Partners:

Diagnostic
• Foundation Medicine, Inc
• Illumina, Inc
• NEO New Oncology, AG
• QIAGEN, NV
• Thermo Fisher Scientific, Inc

Academic
• Charité Berlin
• LMU Munich
• Technical University Munich
• University Hospital Cologne
• University Hospital Dresden
• University Hospital Erlangen
• University Hospital Halle (Saale)
• University Hospital Heidelberg
• University Hospital Regensburg
• University Hospital Zurich

Pharmaceutical
• Bristol-Myers Squibb Company, Inc
• F. Hoffmann-La Roche, AG
• Merck Sharp & Dohme, Ltd

Other
• German Cancer Consortium (DKTK)
• Institute for Hematopathology, Hamburg

Friends and QuIP TMB Standardization and Harmonization Initiative Objectives

• Identify variation between TMB assessed by WES and by targeted gene panels
• Create TMB reference standards using WES to facilitate alignment of various targeted gene panels
• Assess interassay and interlaboratory variability and identify sources of this observed variation
• Develop recommendations to minimize, or account for, variation in methods of TMB estimation and reporting, and for TMB cutoff values, that will inform and advise best practices for prospective clinical studies

(B)

FIGURE 2 A, Objectives and partners, and B, methodological approaches adopted for the collaborative Friends and QuIP TMB standardization and harmonization initiatives. Abbreviations: Friends, Friends of Cancer Research; FDA, Food and Drug Administration; MC3, Multi-Center Mutation Calling in Multiple Cancers; NSCLC, non-small cell lung cancer; QuIP, Quality Assurance Initiative Pathology; SCCHN, squamous cell carcinoma of the head and neck; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden; WES, whole exome sequencing

Both organizations will then develop TMB reference standards, with Friends using commercially available tumor cell lines and QuIP using clinical samples, that will facilitate the alignment of TMB assessed by WES and by targeted gene panels. As part of clinical analyses, QuIP recommendations to minimize interassay and interlaboratory variation in TMB estimation and reporting. The targeted panel assays being tested by QuIP include ThermoFisher Oncomine Tumor Mutation Load Assay, QIAGEN QIAseq Targeted DNA IO Panel, QIAGEN QIA-
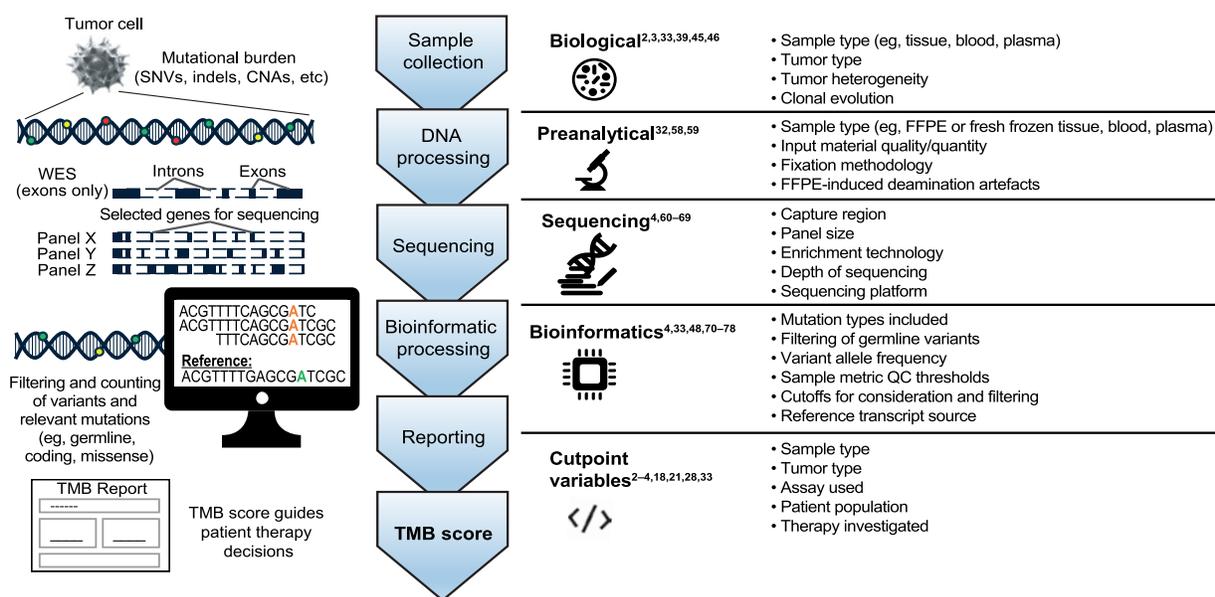
FIGURE 3 Factors that impact TMB or TMB estimation and reporting throughout the TMB assessment process. Abbreviations: CNA, copy number alteration; FFPE, formalin-fixed, paraffin-embedded; indels, short insertions and deletions; QC, quality control; SNV, single nucleotide variant; TMB, tumor mutational burden; WES, whole exome sequencing

Foundation Medicine FoundationOne Panel, Illumina TruSight Oncology 500, and several laboratory-derived assay panels developed in German academic institutes. The Friends initiative evaluates 11 TMB platforms and assays with different TMB assessment parameters, including the FoundationOne CDx and MSK-IMPACT assays, to provide an overview of how these panels compare with one another and to highlight how different factors can influence TMB estimation and reporting. As part of clinical analyses, Friends will evaluate TMB cutoff values in published studies to propose recommendations to inform prospective clinical studies.

Together, data from these multidisciplinary TMB standardization and harmonization approaches cover a wide spectrum of critical aspects of TMB assessment to propose recommendations for consistent TMB estimation, assay comparability, and TMB cutoff values for potential clinical use.[43]

5 | VARIATION IN TMB ASSESSMENT AND FACTORS THAT IMPACT TMB OUTPUT

Review of the published literature indicates that several factors influence TMB assessment, and results of preliminary analyses from the Friends and QuIP initiatives indicate that certain factors have greater impact than others on TMB estimation and reporting; as summarized in Figure 3 and Table 2, and discussed below.

Biological parameters result in differences between overall tumor mutational frequency, with the most basic being tumor type and sample type.[2,3,32] Alexandrov et al. observed that TMB varies according to tumor type, with some tumors intrinsically having higher TMB than others.[2] For example, melanoma and lung tumors have higher TMB than renal carcinomas, brain-related tumors, and hematological

cancers. TMB can also be affected by tumor cellularity and heterogeneity, with subclonal events having a higher impact on mutational burden than clonal evolution.[3,39,45,46] Additionally, tumor transcriptional and/or splicing profiles may differ from reference profiles, resulting in miscounts and impacting the TMB score reported.[47,48]

To date, most of the published studies have assessed TMB in solid tumor samples; however, blood TMB assessment assays are increasingly being used to assess TMB association with response to immune checkpoint inhibitors (Table 1). Because TMB is most commonly assessed using FFPE tumor tissue samples, the initiative by Friends and QuIP proposes recommendations for standardized TMB assessment in these samples; however, TMB assessment using liquid samples is being evaluated by many other groups. Currently, there are several limitations to using other samples for TMB assessment, including that due to low levels of circulating tumor DNA (ctDNA), liquid samples may not yield sufficient quantity for NGS analysis.[24,49–53] Reports show that the sensitivity and accuracy of TMB assay results from liquid samples depend on, among other factors, variability in tumor DNA in the blood. ctDNA can have heterogeneous origins and can be altered by treatment, thereby leading to variation in the final TMB score.[24,49–53] Several ongoing studies are evaluating reliability of TMB assessment from blood samples and harmonizing tissue and blood-derived TMB, including use of the bTMB assay developed by Foundation Medicine.[20,24,53,54] The potential limitations of specificity, sensitivity, and robustness of TMB assessment using blood samples should be further investigated and appropriate guidance should be given on how to address such limitations. Similarly, genome profiling in cytology samples requires a minimum level of cellularity and tumor content, and use for TMB assessment should be further investigated.[55]

TABLE 2 How factors impact TMB score

Factor	Select parameter/technical consideration	Impact on TMB score
Biological	Tumor type	Alternative splicing patterns are dependent on tumor types, and some tumor types have higher TMB than others[2,47]
Preanalytical	Sample type	FFPE samples may harbor artefactual deamination alterations that may impact mutation calling and TMB calculation[56,57]
	Tumor purity	Infiltration of tumor with immune or TME cells may impact TMB score (lower tumor purity is associated with reduced sensitivity)[32]
Sequencing parameters	Genomic region covered	TMB score will depend on panel size and genomic region covered. Greater panel sizes are associated with more precise TMB estimated values[4,62–69]
	Genes included in panel	Gene selection in panels is biased toward frequently mutated cancer-associated genes, and mutation patterns of these genes are often nonrandom.[33] TMB scores may depend on whether the panel contains specific genes that harbor frequent mutations in specific tumor types
	Depth of coverage	Reduced depth of coverage is associated with reduced sensitivity[33,61]
Bioinformatics	Germline variant removal/filtration	Major germline genomic databases have different population race distribution and allele frequency spectrum of variants. TMB score will depend on selection of population allele frequency database when matched tumor-normal tissue is not available[4]
	Reference transcript source	The choice of reference transcript source may impact TMB score depending on the variants considered and counted[48]
	Variants counted in TMB calculation	Panels may consider all variant types or only some of them during their TMB calculations.[1,4,33] TMB score will depend on how comprehensive the variant counting rules are
	Mutation callers	Mutation callers will count variants differently, with some being more comprehensive than others.[71] There is no optimal mutation caller, so a combination of different callers may be most optimal
	Allele frequency/fraction	Reduced variant allele fraction is associated with reduced sensitivity[74]
	Minimum variant count	Reduced variant counts are associated with reduced sensitivity[62]
Cutoff variables	Tumor type	TMB differs widely across tumor types. The cutoff chosen must be appropriate for the tumor type being tested for a reliable and clinically meaningful TMB score to define high TMB[2–5]

Abbreviations: FFPE, formalin-fixed, paraffin-embedded; TMB, tumor mutational burden; TME, tumor microenvironment

TABLE 3 Proposed recommendations for consistent TMB assessment

Factor	Parameter	Recommendations
Preanalytical	Sample processing	• Standardize sample processing protocols • Minimize interlaboratory variability
Sequencing parameters	Genomic region covered	• Select gene panels that screen for actionable mutations or biomarkers • Select panels with larger genome coverage (ideally ~1 megabase or greater)
Bioinformatics	Standardization of workflow	• Align panel-derived TMB values to a WES analysis-derived reference standard to ensure consistency regardless of the assay • Standardize bioinformatic algorithms used for mutation calling and filtering
Comparison of results	Calibration of outputs	• Ensure reporting consistency by developing templates for clinically meaningful reporting (eg, report TMB as mutations per megabase) • Allow calibration of results from different studies

Abbreviations: TMB, tumor mutational burden; WES, whole exome sequencing

TMB estimation and reporting can be heavily influenced by differing working processes across clinical and research laboratories; primarily, the choice of assay, platform, and how the assay is implemented.[32] Preanalytical factors can also have significant effects on TMB estimation, including those that apply to all genomic profiling assays, such as sample collection and processing, input material quality and quantity, sample fixation methodology, FFPE-induced deamination artefacts, and NGS library preparation.[56,57] These factors affect the quantity and quality of DNA extracted for TMB assessment by either WES or targeted gene panel assays, and therefore, TMB estimation output. For example, low tumor purity, which can result from infiltration of immune or tumor microenvironment cells, can lead to reduced TMB assay sensitivity. Also, fixation time is a preanalytical factor that influences the introduction of FFPE-induced deamination artefacts, which also impacts TMB estimation at the stage of bioinformatic analysis.[58,59]

For sequencing, genome coverage differs between WGS, WES, and targeted gene panel assays. WGS covers the whole genome, WES covers the entire exome coding region, and targeted gene panels cover specified areas that may or may not include tumor suppressor genes, driver genes, or intronic regions.[4,60,61] Moreover, the size and location of the capture region differs between targeted gene panel assays. It is important to carefully consider the panel size and composition for accurate TMB assessment. Supporting this concept, it has been observed that confidence intervals for TMB estimation increase with the use of gene panels that assess a smaller area of the genome compared with those that assess a larger area, which suggests that using smaller coverage gene panels could lead to the overestimation or underestimation of TMB.[4,62–69] Depth of sequencing also differs between WES and targeted gene panel assays; sequencing depth is greater for targeted gene panels (~500×) than for WES (~100×).[4,61,64] Genome coverage and sequencing depth together determine assay sensitivity and specificity, and therefore, influence TMB estimation output.

Bioinformatic algorithms can differ widely across targeted gene panels and although these factors heavily influence TMB estimation and reporting, the specifics are often not reported (Table 1). The mutation types considered for TMB assessment can vary from one assay to another. These may include or exclude short insertions and deletions (indels) and/or synonymous and nonsynonymous base substitutions/single nucleotide variants.[4,33,70] For example, from retrospective analyses, it has been observed that TMB assessed by WES often includes missense mutations only, leaving out indels and other mutations, whereas some targeted panels include these variant types.[4,18,20,21,24,26,27] This is an important consideration due to the impact of indels and frameshift mutations on neoantigen formation.[4] However, calling indels can be challenging and their inclusion may depend on the sensitivity of the methods used to detect them.[71] Other bioinformatic parameters that impact TMB estimation and reporting include quality control metrics and various data-filtering procedures for inclusion/exclusion of a variant in the TMB estimation.[4,18,33,38,48] Filtering algorithms and cutoffs for putative germline variants, variant allele frequency (VAF), and FFPE-induced deamination artefacts vary between assays and can be affected by biological and preanalytical factors. For example, VAF cutoffs can vary from 0.5 to 10%, with lower thresholds increasing the risk of including false-positives arising from contamination or sequencing artefacts.[4,24,36,48,72–74]

For calculation of the TMB denominator, genome coverage and bioinformatic parameters must be considered. Most studies have reported a TMB value in mut/Mb, whereas others have reported total mutations per tumor (WES studies); this makes it difficult to compare TMB values across patients and studies. Alongside the way in which TMB is reported, a key factor that must be aligned to ensure consistent identification of patients who are likely to benefit from immune checkpoint inhibitors, and which is currently variable among assays and centers, is the cutoff threshold that defines tumor TMB as high or low. Cutoffs may differ depending on sample type, tumor type, patient subgroup, therapy investigated, and assay used, and the recommendations proposed by Friends and QuIP aim to facilitate the identification of such cutoffs to inform prospective clinical studies.[2–4,18,21,28,32]

6 | RECOMMENDATIONS FOR RELIABLE TMB ESTIMATION AND REPORTING IN CLINICAL SAMPLES

The Friends and QuIP initiatives have proposed recommendations for the standardization of TMB assessment to improve reproducibility and reliability, and best practices for how to minimize and account for variability among assays (Table 3). From results of preliminary analyses, we recommend that NGS assays provide as much patient-relevant genetic/molecular information as possible to avoid the need for rebiopsy and retesting of quality samples at baseline. This will be critical to guide immediate therapy selection with targeted therapies. For example, testing of actionable driver mutations (eg, EGFR inhibitor therapies for EGFR-mutated lung cancers), genes associated with mutagenesis (eg, POLE), and potential negative predictors of response (eg, mutated β2M, JAK1/2, PTEN, STK11).[75–78] We recommend that targeted gene panel assays that have larger genome coverage (ideally with ~1 megabase being the lower limit) are used because they yield more reliable TMB estimation than smaller panels.[67–69] Of note, panels that cover less than 1 megabase are useful; however, accuracy may be reduced.[67–69] We also recommend the use of external reference sequence data, generated using agreed standard methodology such as WES, as this may enable and facilitate TMB assessment interpretability across panel assays.

Ongoing empirical and clinical analyses to generate reference standards, compare TMB measured by WES with TMB measured by various targeted gene panels, and evaluate and minimize interlaboratory and interassay variability are underway. These data will investigate additional aspects of TMB measurement to ensure consistency between assays and laboratories, based on the expectation that many laboratories may develop their own tests for TMB assessment.

7 | CONCLUSIONS

Standardization and harmonization of TMB assessment across assays and centers are essential for reliable and reproducible use of TMB as a clinical biomarker of response to immune checkpoint inhibitors. There is a recent increase in the integration of TMB as a biomarker to select patients who will most likely benefit from immune checkpoint

inhibitors in clinical trials. Increased use of TMB, as well as the current variations in methods of TMB estimation and reporting, highlights the need for standardized and harmonized methods for TMB assessment.

Results of preliminary analyses from Friends and QuIP highlight the importance of targeted gene panel size and composition, and bioinformatic pipeline for reliable TMB estimation in FFPE samples. Following the critical and timely recommendations proposed by Friends and QuIP will help minimize variability in TMB estimation and reporting, which will ensure consistency of TMB assessment in clinical samples across assays and centers. This will improve interpretability of TMB data across assays and studies and lead to the more reliable and accurate use of TMB as a biomarker to identify patients likely to benefit from immune checkpoint inhibitors and to effectively guide patient treatment decisions.

CONFLICT OF INTEREST

MD reports honoraria from Bristol-Myers Squibb. JM reports honoraria for talks from Bristol-Myers Squibb. AS reports advisory board fees from AstraZeneca, Bayer, Bristol-Myers Squibb, Illumina, Novartis, and Thermo Fisher Scientific, and honoraria for talks from AstraZeneca, Bristol-Myers Squibb, MSD, Illumina, Novartis, Roche, Bayer, and Thermo Fisher Scientific. JDA, DMM, MDS, and MMW declare no conflicts of interest.

ORCID

Albrecht Stenzinger https://orcid.org/0000-0003-1001-103X

REFERENCES

1. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69-74.
2. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-421.
3. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-218.
4. 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(1):34.
5. Chen YP, Zhang Y, Lv JW, et al. Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. *Theranostics*. 2017;7(14):3585-3594.
6. Kim JM, Chen DS. Immune escape to PD-L1/PD-1 blockade: seven steps to success (or failure). *Ann Oncol*. 2016;27(8):1492-1504.
7. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458:719-724.
8. Chabanon RM, Pedrero M, Lefebvre C, Marabelle A, Soria JC, Postel-Vinay S. Mutational landscape and sensitivity to immune checkpoint blockers. *Clin Cancer Res*. 2016;22(17):4309-4321.
9. Giannakis M, Mu Xinmeng J, Shukla Sachet A, et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep*. 2016; 15(4):857-865.
10. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252-264.
11. Iwai Y, Ishida M, Tanaka Y, Okazaki T, Honjo T, Minato N. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc Natl Acad Sci U S A*. 2002;99(19):12293-12297.
12. Nobel Media AB. The Nobel Prize in Physiology or Medicine 2018. https://www.nobelprize.org/prizes/medicine/2018/press-release/. Accessed October 2, 2018.
13. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med*. 2017;377(25):2500-2501.
14. Liontos M, Anastasiou I, Bamias A, Dimopoulos M-A. DNA damage, tumor mutational load and their impact on immune responses against cancer. *Ann Transl Med*. 2016;4(14):264.
15. Sharma P, Allison JP. The future of immune checkpoint therapy. *Science*. 2015;348(6230):56-61.
16. Rizvi NA, Hellmann MD, Snyder A, et al. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science*. 2015;348(6230):124-128.
17. Kowanetz M, Zou W, Shames D, et al. Tumor mutation burden (TMB) is associated with improved efficacy of atezolizumab in 1L and 2L+ NSCLC patients. *J Thorac Oncol*. 2017;12(1):S321-S322.
18. Carbone DP, Reck M, Paz-Ares L, et al. First-line nivolumab in stage IV or recurrent non–small-cell lung cancer. *N Engl J Med*. 2017;376(25):2415-2426.
19. Cristescu R, Mogg R, Ayers M, et al. Mutational load (ML) and T-cell-inflamed microenvironment as predictors of response to pembrolizumab. *J Clin Oncol*. 2017;35(suppl 7S): Abstract 1. Available at http://ascopubs.org/action/showCitFormats?doi=10.1200/JCO.2017.35.7_suppl.1.
20. Fabrizio D, Lieber D, Malboeuf C, et al. A blood-based next-generation sequencing assay to determine tumor mutational burden (bTMB) is associated with benefit to an anti-PD-L1 inhibitor, atezolizumab. Presented at: AACR Annual Meeting; April 14–18, 2018; Chicago, IL. *Cancer Res*. 2018;78(13 suppl). Abstract 5706.
21. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378(22):2093-2104.
22. Ramalingam SS, Hellmann MD, Awad MM, et al. Tumor mutational burden (TMB) as a biomarker for clinical benefit from dual immune checkpoint blockade with nivolumab (nivo) + ipilimumab (ipi) in first-line (1L) non-small cell lung cancer (NSCLC): identification of TMB cutoff from CheckMate 568. *Cancer Res*. 2018;78(suppl 13): CT078. Available at http://cancerres.aacrjournals.org/content/78/13_Supplement/CT078.
23. Velcheti V, Kim ES, Mekhail T, et al. Prospective clinical evaluation of blood-based tumor mutational burden (bTMB) as a predictive biomarker for atezolizumab (atezo) in 1L non-small cell lung cancer (NSCLC): interim B-F1RST results. *J Clin Oncol*. 2018;36(suppl): 12001. Available at http://ascopubs.org/action/showCitFormats?doi=10.1200/JCO.2018.36.15_suppl.12001.
24. 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(9):1441-1448.

67. Buchhalter I, Rempel E, Endris V, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden (TMB) analysis. *Int J Cancer*. 2018;144:848-858. https://doi.org/10.1002/ijc.31878.

68. 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(6):703-715.

69. Endris V, Buchhalter I, Allgauer M, et al. Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in-silico and real-life analysis of three larger gene panels. *Int J Cancer*. 2018. https://doi.org/10.1002/ijc.32002.

70. Singh RR, Patel KP, Routbort MJ, et al. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn*. 2013;15(5):607-622.

71. Kroigard AB, Thomassen M, Laenkholm AV, Kruse TA, Larsen MJ. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLoS One*. 2016;11(3):e0151664.

72. Koeppel F, Blanchard S, Jovelet C, et al. Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients. *PLoS One*. 2017;12(11):e0188174.

73. Lauss M, Donia M, Harbst K, et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat Commun*. 2017;8(1):1738.

74. Gerstung M, Beisel C, Rechsteiner M, et al. Reliable detection of sub-clonal single-nucleotide variants in tumour cell populations. *Nat Commun*. 2012;3:811.

75. Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171(5):1042-1056.

76. Rajendran BK, Deng CX. Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget*. 2017;8(30):50252-50272.

77. 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(11):1023-1031.

78. Thomas SJ, Snowden JA, Zeidler MP, Danson SJ. The role of JAK/-STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br J Cancer*. 2015;113(3):365-371.

Spatial and Temporal Heterogeneity of Panel-Based Tumor Mutational Burden in Pulmonary Adenocarcinoma: Separating Biology From Technical Artifacts

Check for updates

Daniel Kazdal, PhD,[a,b] Volker Endris, PhD,[a] Michael Allgäuer, PhD,[a]
Mark Kriegsmann, MD,[a] Jonas Leichsenring, MD,[a] Anna-Lena Volckmar, PhD,[a]
Alexander Harms, MD,[a,b] Martina Kirchner, PhD,[a] Katharina Kriegsmann, MD,[c]
Olaf Neumann, PhD,[a] Regine Brandt, PhD,[a] Suranand B. Talla, PhD,[a]
Eugen Rempel, PhD,[a] Carolin Ploeger, PhD,[a] Moritz von Winterfeld, PhD,[a]
Petros Christopoulos, MD,[b,d] Diana M. Merino, PhD,[e] Mark Stewart, PhD,[e]
Jeff Allen, PhD,[e] Helge Bischoff, MD,[b,d] Michael Meister, PhD,[b,f]
Thomas Muley, PhD,[f] Felix Herth, MD,[b,g] Roland Penzel, PhD,[a] Arne Warth, MD,[h]
Hauke Winter, MD,[b,i] Stefan Fröhling, MD,[j,k] Solange Peters, MD, PhD,[l]
Charles Swanton, MD, PhD,[m] Michael Thomas, MD,[b,d] Peter Schirmacher, MD,[a,k]
Jan Budczies, PhD,[a] Albrecht Stenzinger, MD[b,k,*]

[a]Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[b]Translational Lung Research Center (TLRC) Heidelberg, German Center for Lung Research (DZL), Heidelberg, Germany
[c]Department of Hematology, Oncology and Rheumatology, University Hospital Heidelberg, Heidelberg, Germany
[d]Department of Thoracic Oncology, Thoraxklinik at the University Hospital Heidelberg, Heidelberg, Germany
[e]Friends of Cancer Research, Washington, DC
[f]Translational Research Unit, Thoraxklinik at University Hospital Heidelberg, Heidelberg, Germany
[g]Department of Pneumonology and Critical Care Medicine, Thoraxklinik at the University Hospital Heidelberg, Germany
[h]Institute of Pathology, Cytopathology, and Molecular Pathology UEGP MVZ Giessen/ Wetzlar/Limburg, Germany
[i]Department of Thoracic Surgery, Thoraxklinik at the University Hospital Heidelberg, Heidelberg, Germany

[j]*Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT), Heidelberg, Germany*
[k]*German Cancer Consortium (DKTK), Partner Site, Heidelberg, Germany*
[l]*Department of Oncology, Centre Hospitalier Universitaire Vaudois (CHUV) and Lausanne University, Lausanne, Switzerland*
[m]*Cancer Evolution and Genome Instability Translational Cancer Therapeutics Laboratory, Francis Crick Institute, London, United Kingdom*

ABSTRACT

Background: Tumor mutational burden (TMB) is an emerging biomarker used to identify patients who are more likely to benefit from immuno-oncology therapy. Aside from various unsettled technical aspects, biological variables such as tumor cell content and intratumor heterogeneity may play an important role in determining TMB.

Methods: TMB estimates were determined applying the TruSight Oncology 500 targeted sequencing panel. Spatial and temporal heterogeneity was analyzed by multiregion sequencing (two to six samples) of 24 pulmonary adeno-carcinomas and by sequencing a set of matched primary tumors, locoregional lymph node metastases, and distant metastases in five patients.

Results: On average, a coding region of 1.28 Mbp was covered with a mean read depth of 609x. Manual validation of the mutation-calls confirmed a good performance, but revealed noticeable misclassification during germline filtering. Different regions within a tumor showed considerable spatial TMB variance in 30% (7 of 24) of the cases (maximum difference, 14.13 mut/Mbp). Lymph node–derived TMB was significantly lower ($p = 0.016$). In 13 cases, distinct mutational profiles were exclusive to different regions of a tumor, leading to higher values for simulated aggregated TMB. Combined, intratumor heterogeneity and the aggregated TMB could result in divergent TMB designation in 17% of the analyzed patients. TMB variation between primary tumor and distant metastases existed but was not profound.

Conclusions: Our data show that, in addition to technical aspects such as germline filtering, the tumor content and spatially divergent mutational profiles within a tumor are relevant factors influencing TMB estimation, revealing limitations of single-sample–based TMB estimations in a clinical context.

Introduction

Tumor mutational burden (TMB) has emerged as a novel biomarker to identify patients more likely to respond to immune checkpoint inhibitor therapy targeting the programmed cell death protein 1 axis or cytotoxic T-lymphocyte associated protein 4 (CTLA-4).[1-4] For many tumor entities, assessment of programmed death ligand 1 (PD-L1) expression on tumor and/or immune cells by immunohistochemistry (IHC) is the approved companion diagnostic.[5] TMB can potentially identify — independently from PD-L1 expression status — different patient cohorts likely to respond and in conjunction with PD-L1 status help to predict non-responders and exceptional responders.[6-11]

Immuno-oncology (IO) therapy has the potential to overcome tumor-mediated immune suppression. However, such therapies are not only associated with significant costs, but also potentially severe adverse side effects.[12] Because only a minority of patients benefit from this strategy, it is of utmost importance to establish biomarkers to guide therapy decisions. However, the cancer immune system interaction is complex and multilayered.[13] As a basis for immunogenicity, it has been hypothesized that tumors with a high number of coding mutations are more likely to generate tumor-specific neoantigens that will be recognized by the immune system.[14] Recent data support a predictive potential of TMB for checkpoint inhibitor therapy (single substance and combinations) in various cancer types.[8,15,16]

Whole-exome sequencing (WES)–based TMB assessment was initially used in clinical trials. Many clinical sites and recent clinical trials have implemented targeted next-generation sequencing (NGS) of widely accessible formalin-fixed, paraffin-embedded (FFPE) tissue samples to approximate TMB as an alternative to WES-based analysis with its higher costs, longer turn-around time, and limited availability of suitable specimens. Definition of cutpoints, critical panel size, and various technical and bioinformatical aspects are important to consider.[17-19,21]

Here, we investigate intratumoral heterogeneity (ITH) as a biological factor affecting TMB determination using panel-based NGS. ITH is a well-described phenomenon in lung adenocarcinoma (ADC) on a radiologic, histopathologic, genetic, epigenetic, and tumor-microenvironmental level.[22-31] According to the most widely accepted theory, ITH is mainly the result of subclonal evolution during natural tumor progression and therapeutic interventions.[32,33] The great molecular variability associated with high levels of ITH is seen as

COMPLEX BIOMARKERS: INFORMING DEVELOPMENT AND STANDARDS FOR DIAGNOSTIC TESTS

one of the leading causes for lack of response or development of resistance under therapy.[34,35] In addition, ITH can interfere with molecular diagnostic testing for prognostication or selection of optimal systemic therapy. Although some actionable targets (e.g., truncal sensitizing mutations in *EGFR*) are generally present across all tumor sites, other alterations such as tumor protein p53 (*TP53*) mutations or the expression of PD-L1 were described to exhibit a heterogeneous distribution.[36-39] Given this molecular heterogeneity, TMB counts might be influenced by ITH.[40]

First, adding to previous in silico data from our group, we validated the TruSight Oncology 500 (TSO500) targeted sequencing panel (Illumina Inc., San Diego, California) for TMB estimation using a cohort of patients with known WES-based TMB counts.[19,21] Then we used this panel to perform multiregion sequencing of a well-characterized cohort of pulmonary ADC and further compared a set of primary tumors to their locoregional and distant metastases.

Our data show that regional variability of TMB is significant in lung ADC; it can alter TMB classification of individual patients and thus influence therapeutic decisions.

Materials and Methods
Samples

All patient ADC specimens analyzed were obtained from surgical procedures at the Thoraxklinik at University Hospital Heidelberg and diagnosed according to the criteria of the 2015 WHO classification of lung tumors at the Institute of Pathology, at University Hospital Heidelberg.[41] FFPE tissue sections were supplied by the tissue bank of the National Center for Tumor Diseases (NCT; project: # 1746, # 2015) in accordance with its ethical regulations approved by the local ethics committee.

To validate panel sequencing–based bTMB (psTMB) estimation with the TSO500 panel against the gold standard of WES-based TMB calculations, FFPE samples of 16 NSCLC specimens (biopsy and resection specimens) were obtained from the Heidelberg Lung Biobank, member of the BioMaterialbank Heidelberg and the Biobank Platform of the German Center for Lung Research (ethical approval S-270/2001, S-206/2011) of which corresponding WES data were available, derived from the DKFZ HIPO and the NCT MASTER programs.[21,42]

For the evaluation of ITH, a cohort of 24 patients with ADC, each consisting of two to four multiregional samples (see Table 1 for clinicopathologic details) was constructed as described before, but excluding tumors with clinically targetable driver mutations.[43] In short, a central section of each tumor was fixed in formalin and subsequently cut into 5 × 5 mm segments according to a

Cartesian grid. Ink marks maintained the original orientation of each segment during histologic processing. Tumor regions considered for sequencing were selected in accordance with the tumor size (the larger the tumor the more regions), different histologic growth patterns, as well as sufficient tumor cell content (≥10%) and DNA concentration (≥4 ng/μL). The predominant histologic growth pattern in each segment (defined as the pattern with the highest percentage) was determined by an experienced pathologist. Additionally, FFPE samples of locoregional lymph node metastases were analyzed if present.

For the assessment of TMB over time, a cohort of five patients with ADC and local as well as distant metastases was investigated. For each patient, one sample of the primary tumor plus one locoregional lymph node site (available in four cases) and one to two distant metastatic sites were tested (see Table 1 for clinicopathologic details).

In Silico TMB Computation

TMB was defined as the total number of missense mutations from WES data generated in the DKFZ HIPO and the NCT MASTER programs. The mean sequencing depth of the WES data set ranged from 180 to 200×. Additionally, a simulated panel-sequencing–based TMB (sim-psTMB) was calculated as the number of missense mutations detected by WES within the coding region covered by the TSO500 panel divided by the size of this region (1.34 Mbp). The TSO500 panel has a total size of 1.95 Mbp and covers 1.34 Mbp of coding region. psTMB and sim-psTMB levels and cutpoints were calibrated against WES-based TMB using linear regression fits.

DNA Extraction and Quantification

For DNA extraction, six consecutive 10-μm thick FFPE sections of each sample were pooled, deparaffinized, and digested with proteinase K overnight. Subsequently, DNA was extracted automatically using a Maxwell 16 Research system and the Maxwell 16 FFPE Tissue LEV DNA Purification Kit (both Promega, Madison, Wisconsin). DNA concentrations were determined with the Qubit HS DNA assay (Thermo Fisher Scientific, Waltham, Massachusetts). All assays for DNA extraction and quantification were performed according to the manufacturers' protocols.

Library Preparation and Massive Parallel Sequencing

In the initial step of the library preparation for the capture-based TruSight Oncology 500 panel (Illumina), the grade of DNA integrity of a sample was assessed

Table 1. Clinicopathologic Data of Spatial Heterogeneity (1-24) and Temporal Heterogeneity (I-V) Cohorts

Case	Sex	Age, years	Smoking status	pTNM Classification	Histologic Pattern[a]	Driver Mutation	Tumor Area, cm^2
1	f	68	NA	pT4, pN2 (17/31), pMX	A, MP, (S)	n.d.	3.75
2	m	84	former	pT2b, pN2 (6/24), pMX	S	KRAS:p.Gly12Cys	11.50
3	f	75	NA	pT2a, pN0 (0/23), pMX	A, S	n.d.	4.00
4	m	63	NA	pT2a, pN0 (0/39), pMX	A, P	KRAS:p.Gly12Cys	8.75
5	m	77	active	pT4, pN2 (4/37), pMX	A, P, (MP)	n.d.	7.50
6	f	66	NA	pT2a, pN1 (6/30), pMX	A, S, (L)	KRAS:p.Gly12Cys	7.25
7	m	74	former	pT3, pN0 (0/32), pMX	A, S, (MP)	KRAS:p.Gly12Cys	6.50
8	m	70	active	pT1a (mi), pN0 (0/31), pMX	L, A	KRAS:p.Gly12Cys	6.25
9	m	54	active	pT2a, pN1 (1/10), pMX	A, P, S, L	KRAS:p.Gly13Cys	6.25
10	m	57	former	pT3, pN0 (0/25), pMX	A, L, P	n.d.	5.50
11	f	54	NA	pT3, pN2 (1/17), pMX	S	KRAS:p.Gly12Cys	8.00
12	m	60	active	pT2b, pN1 (4/38), pMX	A, S, (MP)	KRAS:p.Gly12Asp	3.25
13	m	63	NA	pT3, pN2 (7/35), pMX	S, (A)	KRAS:p.Gly12Val	3.00
14	m	60	former	pT2a, pN1 (3/51), pMX	L, A	KRAS:p.Gly13Cys	3.50
15	m	51	never	pT2a, pN0 (0/21), pMX	P	KRAS:p.Gly12Asp	4.75
16	m	66	never	pT3, pN1 (1/48), pMX	A, S	ERBB2:p.Ser310Phe	4.50
17	m	60	former	pT2a, pN0 (0/37), pMX	A	n.d.	2.00
18	m	72	active	pT3, pN2 (5/40), pMX	A, (L, M, P)	KRAS:p.Gly12Asp	14.50
19	m	66	active	pT1a, pN0 (0/22), pMX	L, (A)	KRAS:p.Gly12Val	1.5
20	f	74	NA	pT1a, pN0 (0/28), pMX	L, P	n.d.	2.25
21	f	66	NA	pT1b, pN0 (0/22), pMX	M, S, (P, L)	KRAS:p.Gly12Ser	3.75
22	f	65	NA	pT3, pN2 (4/52), pMX	S	n.d.	13.50
23	f	80	NA	pT1b, pN1 (3/32), pMX	M	n.d.	3.75
24	f	59	NA	pT2a(m), pN2 (12/18), pMX	S, M (A)	KRAS:p.Gly12Cys	12.25
I	f	74	NA	pT2a, pN2 (3/34), pM(ADR, BRA)	A	BRAF:p.Gly466Val	NA
II	m	61	NA	pT3, pN2 (14/35), pM(ADR)	A	n.d.	NA
III	m	52	NA	pT4, pN2 (9/24), pM (OTH[b], HEP)	A	KRAS:p.Gly12Cys	NA
IV	m	60	NA	pT3, pN2 (12/29), pM(ADR)	S	n.d	NA
V	m	58	NA	pT3, pN2 (15/28), pM(ADR)	M	n.d.	NA

[a]Minor component shown in parentheses.
[b]Case III with metastases to pancreas.
f, female; m, male; A, acinar; L, lepidic, MP, micropapillary; P, papillary; S, solid; ADR, adrenal; BRA, brain; HEP, liver; OTH, other; NA, not available; n.d., not detected.

using the Genomic DNA ScreenTape Analysis on a 4150 TapeStation System (both Agilent, Santa Clara, California). To fragment the DNA strands to a length of 90 to 250 bp, 80 ng DNA of each sample were sheared according to their degradation level for 50 to 78 seconds using a focused ultrasonicator ME220 (Covaris, Woburn, Massachusetts). Following two-target capture and purification steps, the enriched libraries were amplified (15 cycles polymerase chain reaction [PCR]) and subsequently quality controlled using the KAPA SYBR Library Quantification Kit on a StepOnePlus quantitative PCR system (both Thermo Fisher Scientific). Up to eight libraries were sequenced simultaneously on a NextSeq 500 (Illumina) using high-output cartridge and v2 chemistry. All assays were performed according to the manufacturers' protocols.

NGS Data Analysis and TMB Determination

Procession of raw sequencing data and variant calling was carried out using the TruSight Oncology 500 Local App (Illumina, pipeline version 1.3.0.39). All variants considered for TMB estimation were manually validated by visual inspection in the integrative genome viewer.[44] Further, the presence of a variation called in one sample of a respective patient was checked in all associated samples. Polymorphisms/germline mutations identified by the TSO500 germline filter were evaluated by the comparison of multiple samples of the same tumor with varying tumor cell content and in nine cases by sequencing matched adjacent non-neoplastic lung tissue for nine specific cases in addition. TMB counts were calculated as the number of synonymous and non-synonymous mutations divided by the covered coding region. For the multiregion sequencing approach, additionally an aggregated TMB was calculated to simulate a pooling of samples. To this end, the number of individual mutations considering all samples of a patient was divided by the average covered coding region of a patient. For this study, considerable ITH was defined as a variation of 5 mut/Mbp as it represents 50% of the most

widely used cutpoint (10 mut/Mbp) applied in clinical trials that use tissue-based TMB testing.[8]

IHC

For the determination of tumor cell content, immune cell composition, and PD-L1 status, IHC stainings for thyroid transcription factor 1 (TTF-1), cluster of differentiation (CD) 45, CD8, and PD-L1 (see Supplementary Material 1 for antibody details) were prepared using an autostainer (BenchMark ULTRA, Ventana Medical Systems, Tucson, Arizona) according to the manufacturer's instructions. The IHC sections were digitalized with a slide scanner (Aperio CS2, Leica Biosystems, Wetzlar, Germany) and evaluated with QuPath (v.0.1.2; Queen's University, Belfast, United Kingdom) applying standard settings for cell detection. For automated cell categorization, specific classifiers were trained and verified by an experienced pathologist.[45] The tumor cell content as the histologic tumor purity (ratio of tumor cells to total cell number) was determined based on digital evaluation of the TTF1-IHC. PD-L1 positivity was defined as linear membranous PD-L1 staining greater than or equal to 1% of tumor cells.

Statistical Data Analysis and Plot Generation

For statistical analyses, the R software (v.3.3.0; R Core Team, 2016) was used with the following functions of the "stats" package (v.3.3.0): chisq.test() for chi-squared contingency table tests; cor.test() to test for association between paired samples using Pearson's product moment correlation coefficient; fisher.test() for

Fisher's exact test; lm() to perform linear regression; and wilcox.test() for the Mann-Whitney U test.

For plot generation the "ggplot2" (v2.1.0) and the "waffle" (v.0.7) package or Microsoft Excel 2013 (Microsoft, Redmond, Washington) and the "Daniel's XL Toolbox NG" (7.1.4, https://www.xltoolbox.net) add-in were used.

Results

Study Outline

In the present study (Fig. 1), we investigated the spatial distribution of TMB estimates in a multiregional sample set of ADC addressing ITH and subsequently the temporal impact on TMB counts in a sample set of primary tumor and local as well as distant metastases. Considering both cohorts, 109 samples were sequenced with a mean read depth of 609×, covering an average coding region of 1.28 Mbp. Initially, the correlation of psTMB with WES data was examined and the mutations called for TMB measurement were manually validated.

Agreement of TMB Measurement Based on WES and the TSO500 Panel

psTMB estimated by sequencing with the TSO500 panel was compared to TMB determined by WES sequencing in a cohort of 16 NSCLC cases (Fig. 2). A strong Pearson correlation of R = 0.9 ($p < 0.01$) was observed between the two approaches, which is in line with in silico data from our group.[19] WES TMB cutpoints of 158 (from clinical trial CheckMate 012), 199 (CM227), and 243 (CM026) somatic mutations were converted to
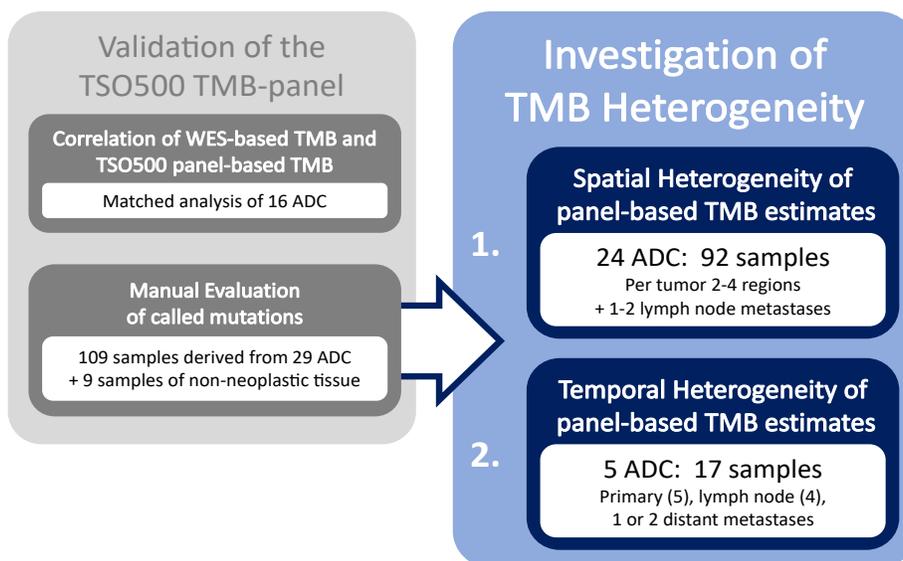
Figure 1. Study outline. Following an initial validation of the TSO500 TMB panel, we investigated the spatial distribution of TMB estimates in a multiregional ADC sample set and subsequently the temporal impact on TMB counts in a sample set of primary tumor and local as well as distant metastases. ADC, adenocarcinoma; TMB, tumor mutational burden.
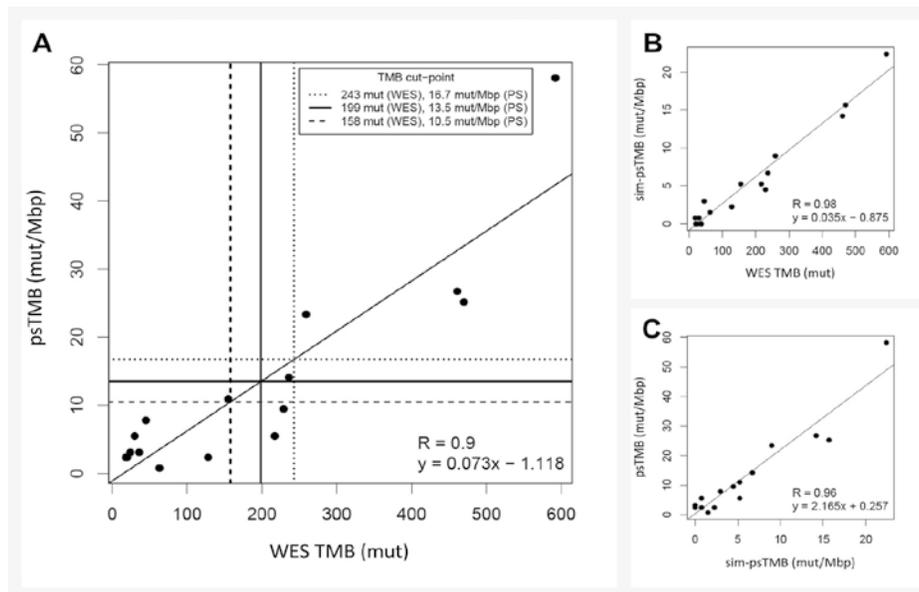
Figure 2. Comparison of the performance of panel sequencing and whole exome sequencing (WES) for the estimation of tumor mutational burden (TMB) in NSCLC. *A*, Strong correlation of panel sequencing–based TMB (psTMB) and WES TMB (R = 0.9). For the WES TMB cutpoints of 158, 199, and 243 mutations corresponding to the psTMB cutpoints of 10.5, 13.5, and 16.7 mut/Mbp classification as TMB-high versus TMB-low was in agreement for 81%, 88%, and 100% of the tumors, respectively. Correlations between a simulated panel sequencing–based TMB (sim-psTMB) and WES TMB (*B*) as well as between psTMB and sim-psTMB (*C*) were higher than the correlation between psTMB and WES TMB. R = Pearson correlation.

psTMB cupoints of 10.5, 13.5, and 16.7 mut/Mbp using a linear regression curve, which served for subsequent individual calibration of the TSO500 panel.[6,8,9] Using these predefined thresholds, classification as TMB-high versus TMB-low was in agreement for 13 (81%), 14 (88%), and 16 (100%) of the 16 investigated tumors, respectively. To decompose different sources of the deviations of psTMB from WES TMB, we investigated the impact of the limited panel size in more detail. To this end, we simulated psTMB by counting the number of mutations detected by WES in the regions captured by the panel, resulting in a sim-psTMB estimate. Correlations between sim-psTMB and WES TMB as well as between psTMB and sim-psTMB were significant (both $p <$ 0.01) and higher than the correlation between psTMB and WES TMB (Figs. 2*B* and *C*). Differences between psTMB and WES TMB were typically (cohort median) 2.7 times higher than differences between sim-psTMB and WES TMB (Supplementary Material 2).

Validation of Mutations Called for TMB Measurement

All mutations identified by the TSO500 mutation calling pipeline were manually validated, in nine cases also including a comparison to matched non-neoplastic tissue. In 21% (23 of 109) of the analyzed samples, mutation calling could be confirmed, whereas one to two and even more than two mutations were either missed (false-negative) or unjustified (false-positive) in 38% or 41% of samples, respectively (Fig. 3, *left*). The main reasons for the miss or nonconsideration of a mutation were the assumption of a single-nucleotide polymorphism/germline mutation (72%) and borderline allele frequencies (21%). An incorrect mutation call resulted mostly from a misclassification as somatic (84%) or the annotation of a complex mutation as two distinct mutations (14%). There was no association between the numbers of missed or unjustified mutations to the estimated TMB (Fig. 3, *left gray line*). In total, 583 somatic mutations (Supplementary Material 3; of which 581 were case specific) could be validated and were considered for TMB estimation (by design excluding recurrently mutated genes such as *KRAS* to avoid overestimation of TMB). Thus, the discordant number of mutations affected pipeline-automated TMB calculations. For the great majority of samples (72%) the absolute difference to the manually validated TMB was smaller than 1 mut/Mbp (Fig. 3, *right*). However, a switch of the TMB designation from high to low (n = 2) or vice versa (n = 2) was observed in 3.6% of the samples.

Spatial TMB Heterogeneity: Multiregional Analysis

The spatial distribution of TMB counts was investigated assessing central tumor sections in a multiregional
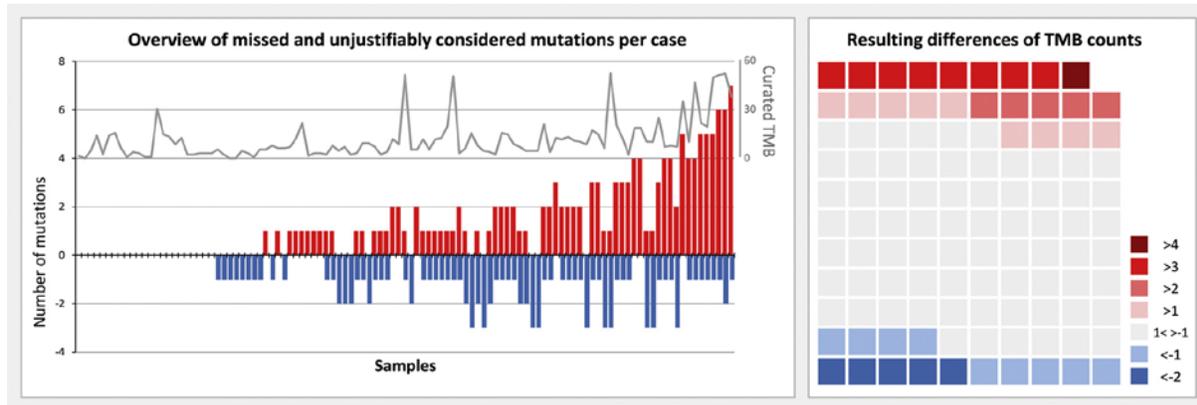
Figure 3. *Left*, Manual validation of mutations considered for tumor mutational burden (TMB) estimation; *red*, missed mutations; *blue*, unjustified mutations; *gray line* gives the respective curated TMB estimation of each sample. *Right*, Resulting differences of the TMB estimation following manual validation of the called mutations.

approach as well as locoregional lymph node metastases (Fig. 4A). Following segmentation, two to four samples of each tumor and up to two lymph node metastases were selected based on sufficient tumor cell content (>10%; Supplementary Material 4), and DNA yield (>4 ng/μL), with differing histologic growth patterns and distances to each other, if applicable. In total, TMB was determined with the TSO500 panel in 69 tumor segments and 23 locoregional lymph node metastases derived from 24 patients. TMB counts of the analyzed samples ranged from 0 to 52.55 mut/Mbp (Fig. 4B) and had a median value of 7.04 mut/Mbp. Considerable ITH of TMB counts, defined by us as a variation of at least 5 mut/Mbp between different regions of a given tumor, was detected in a third (7 of 24) of the analyzed cases (#3, #4, #6, #9, #13, #16, and #20). The highest TMB estimates in these tumors were 5.5, 7.8, 17.2, 14.1, 6.3, 52.6, and 14.8 mut/Mbp, respectively. Mean absolute deviations ranged from 2.43 to 6.11 with a maximum difference of 14.13 mut/Mbp in case #16. The variation of TMB counts within individual cases was even greater when lymph node metastases were included in the analysis (12 cases with ±5 mut/Mbp; maximum difference: 14.21 mut/Mbp; mean absolute deviations: 1.57 to 5.67).

Besides intratumoral variation of the mutation numbers, in 13 cases (#6 through #9, #11, #14, #16 through #20, #22, and #24) distinct mutations were exclusive to different regions within a tumor, also indicating branched tumor evolution. We also simulated pooling of DNA from a patient's various tissue samples by calculating an aggregated TMB count of all detected mutations (Figure 4B, red horizontal bars) which resulted in 0.79 to 7.03 mut/Mbp higher TMB values for these tumors.

Because universally accepted cutpoints regarding the classification of psTMB counts are not yet established

and highly controversial, we applied various cutpoints from recent clinical trials. In Figure 4B, the TMB status of tumor segments, lymph node metastases, or the aggregated TMB counts was determined applying a cutpoint of 10 mut/Mbp, a clinically prospectively validated threshold. Respective corresponding analyses for additional cutpoints (10.5, 13.5, and 16.7 mut/Mbp) derived from WES data (158, 199, and 243 mutations) as referred to above, are given in Supplementary Material 5.

Despite substantial intratumoral variability, in 71% (17 of 24) of the analyzed tumors, estimated TMB status was consistent in different tumor regions and lymph node metastases, with 12 cases (#1 through #5, #8, #10, #13 through 15, #18, and #23) found to be TMB-low and 5 cases (#11, #16, #17, #19, #21) to be TMB-high. Cases #12, #22, and #24 were TMB-high in all tumor segments and consequently was their aggregated TMB, but had at least one lymph node metastasis classified as TMB-low. In three cases (#6, #9, and #20) inconsistent TMB approximations were observed in different segments of the same tumor. Here, the analyzed lymph node metastases were TMB-low. In case #7, only the aggregated TMB would justify a TMB-high classification, whereas all individual tumor segments were found to be TMB-low.

The occurrence of lower TMB values in lymph node metastases compared to corresponding primary tumors was significant ($p = 0.016$) in a paired analysis of the average TMB values of tumor segments and lymph nodes of respective cases (Fig. 4B, upper left inset). Only 6 of 23 analyzed lymph node metastases (#9-N1, #14-N1, #22-N2, #23-N1 + N2, and #24-N1) had a private mutation that was not detectable in the corresponding tumor. Except for one, these mutations had allele frequencies below 10%.
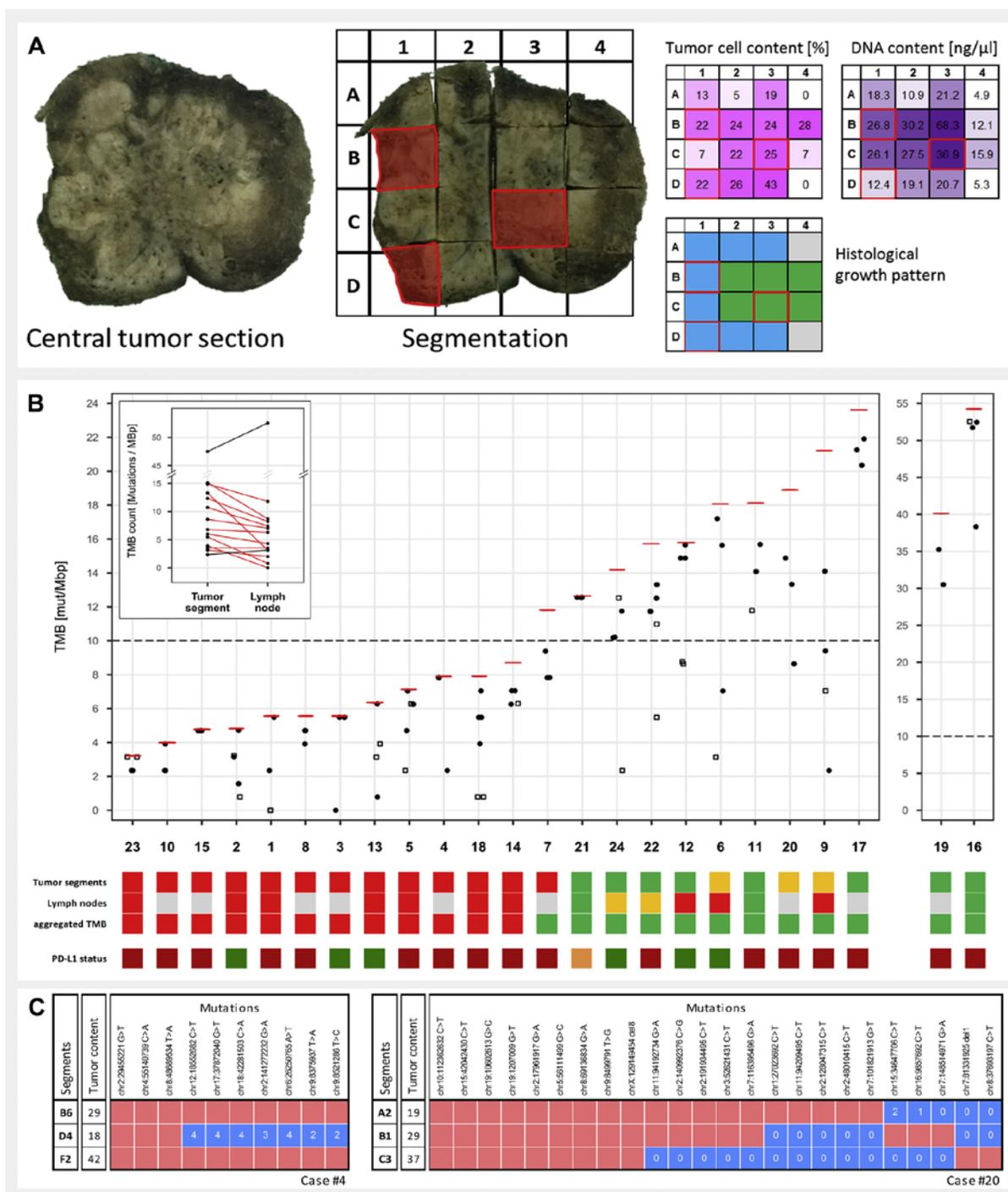
In an IHC analysis, six tumors (25%) were found to be PD-L1–positive in all analyzed regions, whereas one case (#21) showed intratumoral variation of PD-L1 status. TMB status or TMB count did not correlate with PD-L1 status nor did it correlate with immune cell infiltrates (CD8 and CD45). There was merely a significant correlation ($p < 0.01$; R = 0.49) of tumor-infiltrating cytotoxic T cells (CD8) to the total number of present leucocytes

(CD45), and both levels were significantly ($p < 0.05$) higher in PD-L1–positive tumors (Supplementary Material 6).

An in-depth analysis of the seven cases with substantial intratumoral differences of TMB estimates revealed varying tumor cell content (4 of 7) and the development of distinct mutational profiles (3 of 7) as the two primary contributing factors (Fig. 4C). For example, in case #4, the tumor cell content of segment D4 was considerably lower (18%) when compared to regions B6 (29%) and F2 (42%). Here, 7 of 10 somatic mutations that were present in the two other segments could not be called due to allele frequencies (2% to 4%) below the assay's detection limit of 5%. The three analyzed tumor segments of case #20 shared nine somatic alterations. Additionally, two to five private alterations were detected in each region as well as five alterations shared only between segments A2 and B1.

TMB Heterogeneity in Tumor Progression: Comparison of Primary Tumor and Metachronous Distant Metastasis

Next, we investigated the potential variation of TMB during tumor progression. Therefore, we assessed TMB in a smaller cohort (n = 5) of matched primary tumors, locoregional lymph node metastases (n = 4) resected at the time of surgery, and one to two distant metastases per case resected (n = 5) or biopsied (n = 3) several months (median = 11 months; range = 4 – 58 months) after initial surgery (Fig. 5). Supporting our previous observation, two of four lymph node metastases had lower TMB values compared to the primary tumor. TMB estimates for the other two lymph node metastases and for the distant metastases were generally in a similar range as for the respective primary tumor. Despite similar TMB estimates, we detected distinct private mutations in different samples. In case IV (Fig. 5, *bottom*), the great majority of mutations (56) were detectable in all samples, but up to four mutations were exclusively found in the primary tumor (n = 4), the lymph node metastasis (n = 1), and in the first (n = 4) or in the second distant metastasis (n = 1), respectively. Additionally, two mutations were shared between the tumor and the lymph node metastasis, one between lymph node and second metastases, two between tumor and both distant metastasis, and four between both distant metastases.

Discussion

In this study, we comprehensively analyzed the applicability of a 523-gene–spanning targeted sequencing panel for estimation of TMB. Following recent in silico and now wet-lab assay validation, we investigated TMB estimates in multiregional and in temporally separated sample sets using two different ADC cohorts.[19] Our data show considerable ITH of TMB estimates which could impact clinical decision-making. We uncovered critical technical and biological aspects that significantly influence diagnostic TMB assessment.

We observed a strong correlation of TMB levels estimated with the TSO500 panel and TMB levels determined by WES (R = 0.9). Agreement of the two methods in classification of TMB as high or low increased when using higher TMB cutpoints. In a recent comprehensive theoretical analysis of psTMB estimates, we showed that the relative error of psTMB levels decreased proportionally to both the square root of the panel size and the square root of the TMB level.[20] Thus, relative errors are lower for high TMB levels, which is in line with a better classification performance of psTMB for high cutpoints. In terms of correlation with WES, the TSO500 panel performed similar to a panel of comparable size, but better than panels of smaller size, again in line with the theoretical analysis and with the observation that size matters.[19,21] Finally, we showed that a substantial part of the deviation of psTMB from WES TMB is connected to the evaluation of mutations in only a restricted region of the exome, underscoring the relevance of our earlier work on simulations of psTMB. Although differences between sim-psTMB and WES-based TMB can be explained with the smaller sequence covered by the panel, the comparison of sim-psTMB and psTMB indicated that additional factors (e.g., different

---

Figure 4. *A*, Representative sample (case #14): central tumor section, before (*left*) and after segmentation (*middle*), as well as the tumor cell content, DNA content, and histologic growth pattern (*blue* indicates lepidic; *green* indicates acinar; and *gray* indicates non-neoplastic) determined for each tumor segment separately. Segments selected for tumor mutational burden (TMB) measurement are circled in *red*. *B*, Overview of the multiregional TMB analysis of 24 adenocarcinoma (ADC) samples (two to six samples per tumor) using the TSO500 panel. *Black dot* indicates tumor segment, *white square* indicates lymph node metastasis, *red line* indicates aggregated TMB value, considering the mutations detected in all samples of a tumor. *Bottom panel*, TMB status considering 10 mut/Mbp as cutpoint for tumor segments/lymph nodes/aggregated TMB; *green* indicates positive, *red* indicates negative, *yellow* indicates positive and negative results, and *gray* indicates not available. Programmed death ligand 1 (PD-L1) status of tumor cells is shown in homologous darker color code. *Upper left inset*, Paired analyses of the average TMB values of tumor segments and lymph node metastases of respective cases. *Red* indicates decrease; *black* indicates increase. *C*, Showcases illustrating different factors influencing intratumor heterogeneity (ITH) of TMB counts. *Left*, case #4 low tumor cell content; #20 ITH, subclonal development; *red* indicates mutation present; *blue* indicates mutation not present/detection not valid; numbers indicate allele frequencies.

sequencing technologies, FFPE versus fresh frozen tissue, sampling bias, ITH, tumor purity, and the analysis of matched normal tissue) may influence TMB assessment.

Upon manual evaluation of mutations called by the TSO500 local app, we could confirm these as confident and reliable mutational calls with an excellent reduction of artifacts from formalin fixation or misalignment. With using an increasing panel size for TMB estimation, correct germline filtering becomes eminently important. The algorithm applied here appeared sufficient in most instances, but leaves room for improvement considering a switch in the TMB designation in 3.6% of the analyzed samples after manual validation. Our matched germline analysis of non-neoplastic tissue revealed that several somatic mutations were considered germline or vice versa and that filtering was inconsistent between different samples of the same tumor based on, for example, varying allele frequencies. Although optimized algorithms for germline variant filtering without matched non-neoplastic tissue are described and may have their strengths, our data indicate that neither querying polymorphism databases nor approximations based on allele frequencies are sufficient for this task.[46] Future studies are warranted to investigate whether concurrent germline sequencing for panel-based TMB estimation is as essential as it is for WES approaches.[47] In this regard, legal restrictions of germline analysis as well as available laboratory capacity or economic feasibility might impede rapid implementation in routine diagnostics.

Legend: ■ Tumor   ■ Lymph node   ■ Metastasis (t1)   ■ Metastasis (t2)

Top chart — TMB [mut/Mbp] vs cases I–V (synchronous annotation near case IV). NA indicates not available.

Case IV — Mutations

	56 mutations	chr1:120465306 ins6	chr18:42456611 G>A	chr18:736906 C>A	chr19:1621885 C>T	chr17:385730313 A>C	chr16:3639092 C>A	chr17:7577536 T>A	chr6:151694768 T>C	chr6:157100573 G>T	chr2:16082360 G>C	chr18:42530295 G>C	chrX:70586342 G>T	chr16:72828424 G>C	chr5:44305191 A>T	chr9:8341873 T>A	chr17:74732458 A>T	chr20:57415248 C>A	chr3:41274916 C>T	chr8:38315006 T>A
Tumor						0	0	0	0	0	0						0	0	0	0
Lymph node		0	0	0	0		0	0	0	0	0	4	2				0	0	0	0
Metastasis (t1a)		0	0	0	0	0				0			0	0	0					
Metastasis (t1b)		0	0	0	0	0	4	0	0	4			0	0						

Figure 5. *Top*, Tumor mutatioinal burden (TMB) estimation in matched samples of the primary tumor, locoregional lymph node metastases and distant metastases: t1 indicates time point 1, t2 indicates time point 2; NA indicates not available. *Bottom*, Detailed illustration of the detected mutations found in the samples of case IV, revealing shared and private mutations between the samples. *Red* indicates mutation present; *blue* indicates mutation not present/detection not valid. Numbers indicate allele frequencies. Metastasis: *t1a* indicates adrenal right, *t1b* indicates adrenal left.

The multiregional assessment of psTMB estimates revealed that ITH was considerable and would have a critical impact on IO therapy decision making in 12.5% (3 of 24) of the analyzed cases. In these, a designation as TMB-high versus TMB-low was dependent on the tumor area sampled. Our findings are consistent with recently reported findings based on the TRACERx cohort, where 21% of the tumors had inconsistent TMB designations using WES data and a cutpoint of 10 mut/Mbp.[40] In this study, variations in tumor purity between sampled regions were found to be an important factor influencing TMB estimates.

Additionally, the multiregional approach revealed distinct mutational profiles in different regions of the same tumor. Incorporating all TMB estimates of a tumor into an aggregated TMB, a post hoc pooling of samples, led to an increase of TMB estimates posing the intriguing question of whether a single sample approach is valid for estimation of a tumor's TMB. Assuming that an aggregated TMB above the cutpoint would predict IO therapy response or treatment decisions similar to a single sample TMB, this would have even led to a different classification in one case. Tumors with high TMB were more likely to have multiple mutation sets, as seen in a recent study by Zhang et al.[48] However, the predictive value of an aggregated TMB remains to be evaluated given that so far clinical trials have used tissue-based TMB estimates by sequencing single samples. ITH of the TMB status and TMB-high designation for the aggregated TMB, but not the individual tumor samples were seen for all applied cutpoints.

Obtaining sufficient tissue for molecular testing in advanced-stage lung cancer patients can be challenging. Mostly, only little biopsy material is available precluding multiregional analysis. Despite its known limitations (e.g., variable DNA shedding), analyzing cell-free DNA derived from a liquid biopsy might have the potential to provide a more holistic picture of the present mutations, hence blood derived TMB (bTMB) estimation is in the focus of current clinical trials.[49,50] A recent exploratory analysis in a phase III trial was able to prove the feasibility of large-scale bTMB assessment and suggests that bTMB is a predictive biomarker for the combination of anti–CTLA-4 and anti–PD-L1 drugs in naive, *EGFR and ALK* wild type, advanced NSCLC.[16] In this consideration, investigating the correlation between bTMB estimates and aggregated tumor TMB (multiple regions) estimates with therapy response would be of great importance.

In lymph node metastases, we observed mostly lower TMB estimates. This might be due to lower tumor cell content because of surrounding lymphocytes or more likely, reflecting the oligoclonal nature of metastases, with less subclonal diversity when compared to a heterogeneous primary tumor which harbors multiple intermixed distinct subclones. This finding questions the use of lymph node metastases for TMB estimation of the tumor, or at least suggests that different cutpoints might need to be further evaluated.

Interestingly, we also detected ITH in distant metastases, but this did not affect TMB estimates considerably. Although the relatively small sample size limits definite conclusions and futher studies are required to fully understand the impact on clinical decision-making, our results indicate that biopsy samples of (potentially easily accessible) metastatic sites can provide similar TMB estimates as the primary tumor.

In conclusion, we show crucial factors influencing TMB estimation. Besides technical aspects such as tumor cell content, sufficient coverage, and germline filtering the subclonal development of tumors and subsequent ITH also affect TMB estimation. From a clinical point of view, ITH of the TMB status or a switch to a TMB-high designation when considering an aggregated TMB for the whole tumor could lead to misclassification of TMB greater than or equal to 17% of the cases, indicating limitations of single-sample based TMB estimations. In this regard, our findings point to open questions considering the definition of TMB as a predictive biomarker when used in clinical trials or therapeutic workflows.

Acknowledgments

Supplementary Data

References

1. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124-128.
2. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet*. 2016;387:1909-1920.
3. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014;371:2189-2199.

4. Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350:207-211.

5. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014;515:563-567.

6. Carbone DP, Reck M, Paz-Ares L, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med*. 2017;376:2415-2426.

7. 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.

8. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378:2093-2104.

9. Hellmann MD, Nathanson T, Rizvi H, et al. Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell*. 2018;33:843-852 e844.

10. Goodman AM, Kato S, Bazhenova L, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther*. 2017;16:2598-2608.

11. Budczies J, Seidel A, Christopoulos P, et al. Integrated analysis of the immunological and genetic status in and across cancer types: impact of mutational signatures beyond tumor mutational burden. *Oncoimmunology*. 2018;7:e1526613.

12. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*. 2012;366:2455-2465.

13. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2018;48:812-830 e814.

14. Gubin MM, Artyomov MN, Mardis ER, et al. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J Clin Invest*. 2015;125:3413-3421.

15. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389:255-265.

16. Peters S, Cho BC, Reinmuth N, et al. CT074 Tumor mutational burden (TMB) as a biomarker of survival in metastatic non-small cell lung cancer (mNSCLC): Blood and tissue TMB analysis from MYSTIC, a phase III study of first-line durvalumab ± tremelimumab vs chemotherapy. Paper presented at American Association for Cancer Research (AACR) Annual Meeting. March 29-April 3, 2019; Atlanta, Georgia.

17. Allgauer 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.

18. Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol*. 2018;30:44-56.

19. Buchhalter I, Rempel E, Endris V, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int J Cancer*. 2019;144:848-858.

20. Budczies J, Allgauer M, Litchfield K, et al. Optimizing panel-based tumor mutational burden (TMB) measurement [e-pub ahead of print]. *Ann Oncol*. 2019;30:1496-1506.

21. Endris V, Buchhalter I, Allgauer M, et al. Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels. *Int J Cancer*. 2019;144:2303-2312.

22. Ma Y, Feng W, Wu Z, et al. Intra-tumoural heterogeneity characterization through texture and colour analysis for differentiation of non-small cell lung carcinoma subtypes. *Phys Med Biol*. 2018;63:165018.

23. Park S, Ha S, Lee SH, et al. Intratumoral heterogeneity characterized by pretreatment PET in non-small cell lung cancer patients predicts progression-free survival on EGFR tyrosine kinase inhibitor. *PLoS One*. 2018;13:e0189766.

24. Cadioli A, Rossi G, Costantini M, et al. Lung cancer histologic and immunohistochemical heterogeneity in the era of molecular therapies: analysis of 172 consecutive surgically resected, entirely sampled pulmonary carcinomas. *Am J Surg Pathol*. 2014;38:502-509.

25. Warth A, Muley T, Meister M, et al. The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. *J Clin Oncol*. 2012;30:1438-1446.

26. Jamal-Hanjani M, Wilson GA, McGranahan N, et al. tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. 2017;376:2109-2121.

27. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168:613-628.

28. Kazdal D, Harms A, Endris V, et al. Subclonal evolution of pulmonary adenocarcinomas delineated by spatially distributed somatic mitochondrial mutations. *Lung Cancer*. 2018;126:80-88.

29. Dietz S, Lifshitz A, Kazdal D, et al. Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. *Int J Cancer*. 2019;144:1061-1072.

30. Alizadeh AA, Aranda V, Bardelli A, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med*. 2015;21:846-853.

31. Mony JT, Schuchert MJ. prognostic implications of heterogeneity in intra-tumoral immune composition for recurrence in early stage lung cancer. *Front Immunol*. 2018;9:2298.

32. de Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014;346:251-256.

33. Warth A, Endris V, Stenzinger A, et al. Genetic changes of non-small cell lung cancer under neoadjuvant therapy. *Oncotarget*. 2016;7:29761-29769.

34. Lovly CM, Salama AK, Salgia R. Tumor heterogeneity and therapeutic resistance. *Am Soc Clin Oncol Educ Book*. 2016;35:e585-e593.

35. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*. 2018;15:81-94.

36. Dietz S, Harms A, Endris V, et al. Spatial distribution of EGFR and KRAS mutation frequencies correlates with histological growth patterns of lung adenocarcinomas. *Int J Cancer*. 2017;141:1841-1848.

37. Zhang LL, Kan M, Zhang MM, et al. Multiregion sequencing reveals the intratumor heterogeneity of driver mutations in TP53-driven non-small cell lung cancer. *Int J Cancer*. 2017;140:103-108.

38. Casadevall D, Clave S, Taus A, et al. Heterogeneity of tumor and immune cell PD-L1 expression and lymphocyte counts in surgical NSCLC samples. *Clin Lung Cancer*. 2017;18:682-691 e685.

39. Liu Y, Dong Z, Jiang T, et al. Heterogeneity of PD-L1 expression among the different histological components and metastatic lymph nodes in patients with resected lung adenosquamous carcinoma. *Clin Lung Cancer*. 2018;19:e421-e430.

40. Rosenthal R, Cadieux EL, Salgado R, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*. 2019;567:479-485.

41. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol*. 2015;10:1243-1260.

42. Horak P, Klink B, Heining C, et al. Precision oncology based on omics data: the NCT Heidelberg experience. *Int J Cancer*. 2017;141:877-886.

43. Kazdal D, Harms A, Endris V, et al. Prevalence of somatic mitochondrial mutations and spatial distribution of mitochondria in non-small cell lung cancer. *Br J Cancer*. 2017;117:220-226.

44. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.

45. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7:16878.

46. Sun JX, He Y, Sanford E, et al. A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal. *PLoS Comput Biol*. 2018;14:e1005965.

47. 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.

48. Zhang Y, Chang L, Yang Y, et al. The correlations of tumor mutational burden among single-region tissue, multi-region tissues and blood in non-small cell lung cancer. *J Immunother Cancer*. 2019;7:98.

49. Volckmar AL, Sultmann H, Riediger A, et al. A field guide for cancer diagnostics using cell-free DNA: from principles to practice and clinical applications. *Genes Chromosomes Cancer*. 2018;57:123-139.

50. 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.

TMB standardization by alignment to reference standards: Phase II of the Friends of Cancer Research TMB Harmonization Project.

DEVELOPMENTAL IMMUNOTHERAPY AND TUMOR IMMUNOBIOLOGY

Diana M. Merino, Lisa McShane, Matt Butler, Vincent Anthony Funari, Matthew David Hellmann, Ruchi Chaudhary, Shu-Jen Chen, Wangjuh Sting Chen, Jeffrey M. Conroy, David Fabrizio, Laura E MacConaill, Aparna Pallavajjala, Arnaud Papin, Mark Sausen, Victor J Weigman, Mingchao Xie, Ahmet Zehir, Chen Zhao, Paul M. Williams

Friends of Cancer Research, Washington, DC; Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD; SeraCare, Gaithersburg, MD; NeoGenomics Laboratories, Aliso Viejo, CA; Memorial Sloan Kettering Cancer Center, New York, NY; Thermo Fisher Scientific, San Francisco, CA; ACT Genomics, Taipei, Taiwan; Caris Life Sciences, Phoenix, AZ; OmniSeq, Inc., Buffalo, NY; Foundation Medicine, Inc., Boston, MA; Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA; Johns Hopkins University, Baltimore, MD; QIAGEN, Waltham, MA; Personal Genome Diagnostics, Inc., Baltimore, MD; Q Squared Solutions, EA Genomics, Morrisville, NC; AstraZeneca, Waltham, MA; Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY; Illumina, Inc., San Diego, CA; Molecular Characterization Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD

2624

Background: Tumor mutational burden (TMB) is a predictive biomarker of response to immune checkpoint inhibitors across multiple cancers. In Phase 1 of the Friends of Cancer Research Harmonization Project, we demonstrated a robust correlation between TMB estimated using targeted next-generation sequencing (NGS) gene panels and whole exome sequencing (WES) applied to MC3-TCGA data. These findings demonstrated variability in TMB estimates across different panels. Phase 2 evaluates sustainable TMB reference standard materials for TMB alignment to assess this variability. The goal of this effort is to establish best practices for estimating TMB in order to improve consistency across panels, for the sake of optimizing clinical application and facilitating integration of datasets generated from multiple assays. **Methods:** Fifteen laboratories with targeted panels at different stages of development participated. We identified a set of reference standards consisting of 10 well-characterized human-derived lung and breast tumor-normal matched cell lines. WES was performed using a uniform bioinformatics pipeline agreed upon by all team members (WES-TMB). Each laboratory used their own sequencing and bioinformatics pipelines (tumor-only and tumor-normal) to estimate TMB according to genes represented in their respective panels (panel-TMB). The association between WES-TMB and each panel-TMB was investigated using regression analyses. Bias (relative to WES-TMB) and variability in TMB estimates across panels were rigorously assessed. All analyses were blinded. **Results:** The set of reference standards spanned a clinically meaningful TMB range (4.3 to 31.4 mut/Mb). Preliminary data from 12 laboratories shows a good correlation between panel-TMB and WES-TMB in this empirical analysis. Across panels, regression R^2 values range 0.77-0.96 with slopes ranging 0.60-1.26. Calibration analyses that seek to minimize variability of TMB estimates across panels using the established set of reference standards are ongoing, as well as investigating cancer type dependence on the relationship between panel-TMB vs. WES-TMB, which will be available at the time of presentation.
Conclusions: Preliminary findings demonstrate feasibility of using sustainable reference control cell lines to standardize and align estimation of TMB across different targeted NGS assays. Future studies aim to validate reference standard material as a reliable alignment tool by using formalin-fixed paraffin-embedded human tumor samples.

FRIENDS
of CANCER
RESEARCH

A FRIENDS OF CANCER RESEARCH WHITEPAPER

DATA GENERATION (AND REVIEW CONSIDERATIONS) FOR USE OF A COMPANION DIAGNOSTIC FOR A GROUP OF ONCOLOGY THERAPEUTIC PRODUCTS

CONTRIBUTORS

CLAUDIA DOLLINS
EMD SERONO

ANDREA FERRIS
LUNGEVITY FOUNDATION

STEFFAN HO
PFIZER, INC.

DAVID HYMAN
MEMORIAL SLOAN KETTERING CANCER CENTER

RITESH JAIN
EMD SERONO

JOCHEN LENNERZ
MASSACHUSETTS GENERAL HOSPITAL

LYNNE MCBRIDE
THERMO FISHER SCIENTIFIC

LAKSHMAN RAMAMURTHY
GLAXOSMITHKLINE

OBJECTIVE

Friends of Cancer Research (*Friends*) convened a working group to explore evidentiary standards that could be useful in supporting a determination that an IVD companion diagnostic (CDx) device is appropriate for use with a class of therapeutic products, rather than with one or more specific products within the class.[1] This whitepaper constructs a framework within which the evidentiary standards necessary to establish confidence in the safe and effective use of a CDx to direct treatment of a specific group or class of therapeutic products, rather than specific individual products, may be considered for the benefit of patients to provide increased information regarding therapeutic options from a single test. This framework is intended to define categories, informed by technical and biologic considerations, where a approval of or expansion of a CDx label to include use in directing treatment with a specific group or class of oncology therapeutic products may be associated with different evidentiary requirements for class/group labeling considerations.

BACKGROUND

An increasingly detailed understanding of the genetic basis and molecular heterogeneity of cancer has driven the development of targeted therapies and associated companion diagnostic tests that have provided significant benefit to patients. These advances in precision medicine have given rise to approvals of subsequent same-in-class therapeutic products each of which are, most often, paired with a different companion diagnostic test. The benefits of multiple therapeutic options offered by approvals of same-in-class therapeutic products, such as the EGFR and PARP inhibitors, may in part be compounded by added complexity in CDx development as well as clinical testing workflows and practice, inadvertently introducing obstacles to access.

A unique characteristic of certain targeted therapies is their reliance on the detection of a biomarker using a specific companion diagnostic test as an aid to identify the patient population most likely to benefit from that therapeutic. A CDx is the regulatory title given to tests approved that are essential to the use of a specific drug or biologic based on the detection of a biomarker. When a diagnostic test is approved as a CDx, its intended use in identifying patients who are appropriate for treatment with a specific therapeutic agent (including the name of the therapeutic), which is typically supported by results demonstrating an acceptable benefit-risk profile of the therapeutic agent used to treat patients identified using the CDx, is described in the CDx test label. Conversely, the indication statement of the corresponding drug or biologic label describes the requirement to test for the relevant biomarker using an approved test without naming the specific test. The Center for Devices and Radiologic Health (CDRH) typically considers companion diagnostics to be high-risk devices (Class 3) requiring pre-market approval, due to the potential for life-altering adverse events associated with incorrect test results. The clinical utility of a companion diagnostic is most often determined in the context of the test informing use of a single targeted therapy. However, more recently there have been approvals of multi-marker, panel-based CDx tests that can inform the use of multiple therapeutic products across multiple tumor types.[2,3,4]

In the Guidance for Industry and Food and Drug Administration (FDA) Staff on In Vitro Companion Diagnostic Devices, the concept of more broadly labeling an IVD companion diagnostic device that would enable use with a class of therapeutic products was introduced:

> The labeling for an in vitro diagnostic device is required to specify the intended use of the diagnostic device (21 CFR 809.10(a)(2)). Therefore, an IVD companion diagnostic device that is intended for use with a therapeutic product must specify the therapeutic product(s) for which it has been approved or cleared for use. In some cases, if evidence is sufficient to conclude that the IVD companion diagnostic device is appropriate for use with a class of therapeutic products, the intended use/indications for use should name the therapeutic class, rather than each specific product within the class.

Additional guidance pertaining to the definition of a class of therapeutic products or elaboration on the evidence that would be sufficient to support expanding a CDx label to reference a class of therapeutic products was not provided. Therefore, a draft Guidance was issued in December 2018 on Developing and Labeling In Vitro Companion Diagnostic Devices for a Specific Group or Class of Oncology Therapeutic Products – Guidance for Industry. Included among the important considerations regarding broader labeling was further definition of a class or group of therapeutic products, which would be "Approved for the same indication, including the same mutation(s) and the same disease for which clinical evidence has been developed with at least one device for the same specimen type for each therapeutic product."[5]

Development of a new targeted therapeutic agent within a potential drug class requires the identification and treatment of patients within the same indication using a test for the same or a biologically

highly related biomarker. However, the new or next generation "same-in-class" (e.g. targeting the same enzyme) therapeutic agent may be intentionally designed to overcome limitations (e.g. resistance mechanisms) associated with the previously approved same-in-class drug that has become established as the standard of care. Use of the new same-in-class drug subsequent to treatment with the approved same-in-class drug eliminates the need to utilize a companion diagnostic test to identify patients for treatment with the new drug given that patients have already been identified to direct the earlier line of treatment with the approved same-in-class drug. Rather, patients are enrolled based on their prior treatment. For example, five drugs are currently approved for the treatment of patients with metastatic non-small cell lung cancer (NSCLC) that is determined to be anaplastic lymphoma kinase (*ALK*)-positive, but due to differences in the line of therapy for which the drug was approved, differences in the requirement for an FDA-approved test differ across these drugs. Drugs that are used subsequent to treatment with a prior (e.g. first-line) same-in-class therapeutic agent do not directly rely on an approved test "as an aid in identifying patients eligible for treatment" but rather take advantage of the existing standard of care established by same-in-class agents, with associated companion diagnostic tests, previously approved as an earlier line of treatment.

Further, given the limitations of tumor tissue availability, testing with multiple CDx for the same biomarker in order to enable treatment with specific or different *ALK* inhibitors may not be feasible. Similarly, serial or parallel application of the multiple single-analyte CDx tests now relevant for the optimal management of NSCLC is challenging and, in some cases, impractical to implement or unfeasible due to tissue availability. In addition, subsequent testing companies that come to market with a test for *ALK* could encounter problems, for example with accessing clinical trial tissue samples, with expansion of their label indication to include all drugs.

The FDA published draft guidance in December 2018 to inform the development and labeling of companion diagnostics for indication with multiple therapeutic products across a group or class of therapeutic products and final guidance is pending.[5] The draft guidance provides an important first step to advancing the use of group labeling for companion diagnostics, but further discussions are needed in order to address the issues outlined in this whitepaper. For example, the draft guidance refers to diagnostic devices for the identification of specific EGFR mutations in tumors of patients with NSCLC. Five different FDA-approved therapeutic products are indicated for patients with NSCLC whose tumors have EGFR mutations – deletions in exon 19 or base-substitution mutations in exon 21 (excluding the T780M and other resistance mutations). In many of these cases, the CDx may only have been clinically validated with one of the therapeutic products in the class. Prior FDA guidance documents[5,6] address how a CDx may seek approval for additional drugs in the same class beyond the agent for which it was originally approved. The guidance suggests how thorough analytical validation of the biomarker including cut-offs for the specific indication, and potentially clinical experience of the diagnostic with at least two therapeutic products can help broaden the labeling of the companion diagnostic for multiple therapeutic products that are in the same class.

The current whitepaper will consider case studies for three biomarkers, EGFR, *ALK*, and *BRCA*/HRD, to 1) define categories of biomarkers based upon biological and technical complexity, 2) explore how FDA's draft guidance could be implemented for simple or moderately technical biomarkers, and 3) begin to develop a common solution on how to establish a shared definition and evidentiary standard for high complexity biomarkers.

The formulation of a scientific evidentiary standard will be helpful to stakeholders as follows:

a. **Industry** – make for efficient diagnostic development by providing a clear, consistent understanding of the types of validation studies required.
b. **FDA** – help align various definitions of the "same" biomarker CDx and help FDA evaluate the safety and efficacy of the drug and diagnostic more efficiently.
c. **Physicians** – communicate information about new and exciting targeted therapies to physicians using 'simplicity in labeling'. This will be of enormous help to them as they manage their patients.
d. **Patients** – who seek streamlined and efficient access to both innovative life-changing therapies and to high-quality diagnostic tests that are critical in directing their safe and effective use.

A Framework for Companion Diagnostic Group Labeling

A framework to inform group labeling for companion diagnostics requires accurate classification (**Table 1**). Diagnostic tiers should be stratified by complexity of the principle of operation/technology, the biology of the drug target and diagnostic biomarker, including an understanding of mechanism of action, and the test's clinical application. This framework is predicated on the assumption that a group of therapeutic products can be appropriately defined, as described in the draft guidance (a specific group or class of oncology therapeutic products are those approved for the same indications, including the same mutation(s) and the same disease for which clinical evidence has been developed with at least one device for the same specimen type for each therapeutic product).

Classification Schema

Tier A companion diagnostics would include tests designed to identify biomarkers that are technically or biologically "simple", such as SNVs or indels associated with dominant driver oncogenes, where measurement of the biomarker in the intent-to-test population demonstrates a distribution that is largely bimodal, supporting a binary (positive vs negative) readout in which classification is not highly sensitive to the cutpoint. In this Tier, group labeling would be based upon other tests targeting the same analyte (for example, a nucleic acid change), using the same technology, and from the same matrix. We propose the creation of a regulatory pathway for review and approval of Tier A biomarkers primarily on the basis of analytical and clinical validation, including assessment of clinical concordance, and demonstration of non-inferiority, with at least one other approved assay measuring the same analyte. **Tier B** companion diagnostics would represent a slightly more complicated or "moderate" biological and technical complexity and/or require a higher level of evidence to support group labeling and may include tests using the same or different platform technologies. An example of a Tier B CDx may be detection of a gene fusion event that defines a biologically distinct subgroup within a given indication, which can involve a variety of upstream partners, or assays that require a high degree of clinical interpretation (for example, a test that involves pathogenicity assessment of a germline variant). Lastly, **Tier C** companion diagnostics would represent the most technical tests, such as algorithmically determined biomarkers and/or require a high level of evidence to support group labeling where different platform technologies or matrices are used, or the algorithms are so unique to each test that a "group" labeling may not be feasible for Tier C biomarkers. Examples here include assignment of homologous recombination deficiency (HRD) scores or tumor mutation burden as a continuous variable each based on next-generation sequencing.

Table 1 outlines a rough framework of how a test might qualify for each tier based upon a general pattern of characteristics and provides examples of those characteristics. Placement in a tier is dependent upon the biologic and technical considerations of the test itself but also the diversity that exists between tests within a group label. A test would not have to meet all the listed characteristics to be placed in a tier. For example, the currently FDA approved CDx for EGFR are FFPE tumor tissue specific and are placed under Tier A here. However, if an NGS-based EGFR CDx were developed for cell-free DNA (cfDNA) isolated from plasma, this difference in matrix used by the test would merit placement in Tier C for the type of evidence needed to support a label expansion to a group label where other tests within the group use FFPE. Further, a detailed understanding of the mechanism of action of the indicated class of therapeutic products and the interaction between the therapeutic product and the biomarker would contribute to consideration of whether tests evaluating different matrices or utilizing distinctly different platform technologies would warrant placing the test in a different tier.[5,6]

Table 1: Categories Relevant to Consideration of Evidentiary Requirements for Class-Based Companion Diagnostic Test Labeling

Tier	Complexity[1]	Example					Type of evidence to support Dx label expansion to a specific group or class of therapeutic products[2]
		Rx	Alteration Type	Dx Analyte	Dx Platform	Dx Matrix	
A	Low	EGFR TKI	Simple: SNVs, indels	Single gene (e.g. EGFR) or small # of genes (e.g. BRCA1 & BRCA2)	NGS	FFPE tumor tissue	Concordance[4] with approved test(s)
B	Moderate	ALK TKI	Moderate: Gene fusion, low # fusion partners	ALK gene fusions	IHC; FISH; NGS	FFPE tumor tissue	Concordance[4] with approved test(s); additional supporting evidence and rationale
C	High	PARPi	Complex: Many variants, some may be unknown (e.g. tumor suppressor mutants; fusions with many partners)	HRD[3]	NGS	FFPE tumor tissue; plasma	Concordance[4] with approved test(s) along with additional supporting evidence and rationale; prospective design of the test/test algorithm (e.g. cut-off) to optimize concordance with established tests may be necessary; if acceptance criteria cannot be met, test-specific association with clinical outcome is needed and a class-specific label would not be feasible

Footnotes:
[1] Referring to the Dx target (specific analyte), the test platform technology and the biospecimen matrix (see Table 2), all considered within the context of the biology of the drug and Dx target and associated understanding of the drug class mechanism of action
[2] Predicated on appropriate consideration of the key factors relevant to drug class Dx labeling (see below; group or class definition, MOA and biomarker-drug interaction, clinical experience, analytical validity, clinical validity)
[3] HRD, homologous repair deficient; this term is intended to refer to a biologically defined state (i.e. "BRCAness") that can be variably measured using different types of assays that may employ different platform technologies, bioinformatic pipelines, biospecimen matrices, etc.
[4] Acceptance criteria for PPA and NPA are defined on a case-by-case basis informed by prevalence of biomarker-positive subgroup within the intent-to-test population and the benefit-risk profile of the class

Note: In all cases, as per draft FDA guidance, sponsors are encouraged to engage in discussions with FDA (CBER, CDRH, or CDER), in coordination with the Oncology Center of Excellence, early in the development or consideration of possible class-based labeling of a companion diagnostic test

Figure 1: Decision Tree for Tier Placement
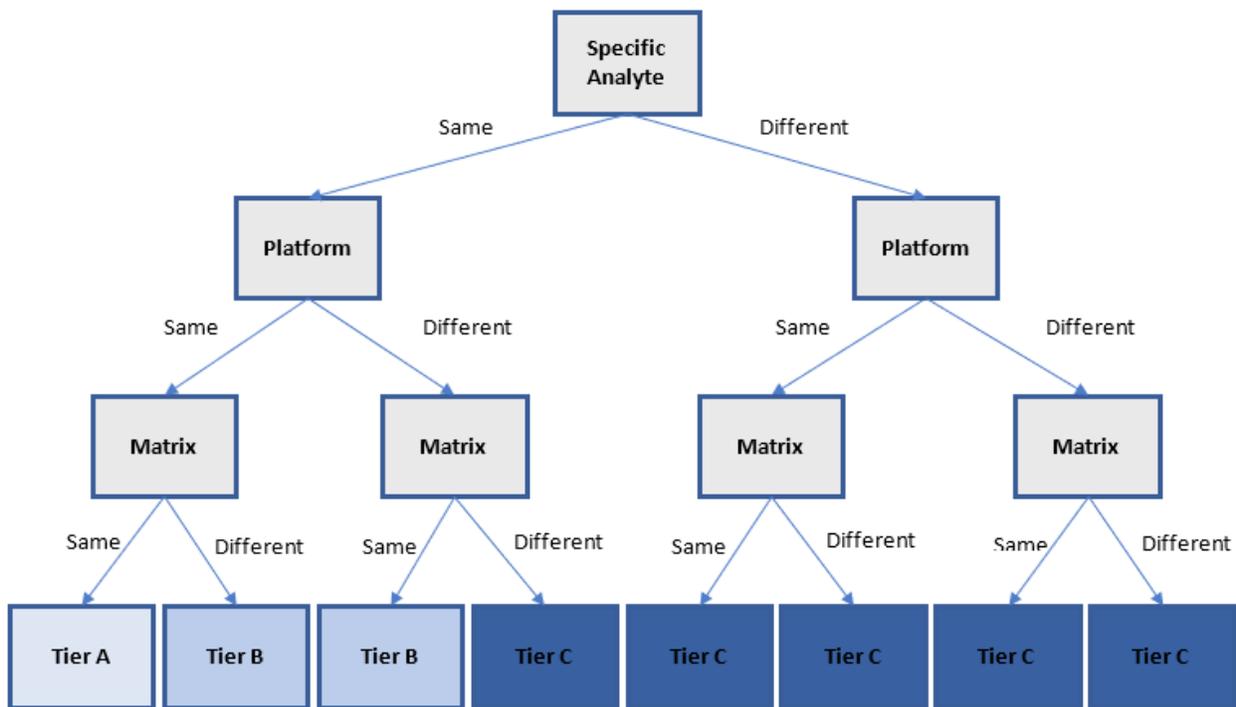
Challenges to Address and Evidence to Support Label Expansion by Category

In its draft guidance, FDA outlines five specific factors companion diagnostic developers should consider when deciding to pursue a broader labeling claim:

1. **Group or class definition.** Whether there is a specific group or class of oncology therapeutic products that can be defined (according to the indication, mutation(s), or disease listed in the therapeutic product's label) for which a companion diagnostic will identify an appropriate patient population for potential treatment.
2. **Understanding of MOA and biomarker-therapeutic interaction.** Whether there is a detailed understanding of a) the mechanism of action of the specific group or class of oncology therapeutic products being considered for use with the companion diagnostic and b) the interaction between the therapeutic products and the biomarker(s), at the mutation level, detected by the companion diagnostic.
3. **Sufficient clinical experience.** Whether there is sufficient clinical experience with at least two therapeutic products for the same biomarker-informed indications.
4. **Demonstration of analytical validity.** Whether analytical validity of the companion diagnostic has been demonstrated across the range of biomarkers that inform the indication.
5. **Demonstration of clinical validity.** Whether clinical validity of the companion diagnostic has been demonstrated with the therapeutic products in the disease of interest.

The below case studies seek to apply the framework outlined above with this guidance to development of a companion diagnostic test where a group label is pursued. By application to examples of companion diagnostic tests used to detect biomarkers representing each tier in **Table 1**, this whitepaper will outline the variables that could be used to provide assurance of drug efficacy across a drug class when indicated by a CDx with a group label across increasingly technical and biological complexity of biomarkers.

CASE STUDY 1: APPLICATION OF TIER A TO EGFR MUTATIONS

According to the categorization schema in Table 1, CDx currently used to identify patients with EGFR-positive NSCLC as an aid in directing treatment with specific members of the class of therapeutic products that inhibit the EGFR receptor tyrosine kinase fit the characteristics designed for Tier A CDx. The biomarker measured by EGFR CDx tests is a specific nucleotide deletion in exon 19 and specific SNVs in exon 21 of the EGFR gene, and the tests utilized to identify these alterations all evaluate the same analyte derived from the same biospecimen matrix. In addition, the alterations represent reasonably well-understood oncogenic driver mutations. The FDA's draft guidance identified EGFR as a case study to illustrate the thought process that would identify appropriate companion diagnostics for group labeling and demonstration of evidence to support a group label.[4]

Group or class definition
As noted in FDA's draft guidance:

> In this example, the oncology community would be better served by a companion diagnostic that detects EGFR exon 19 deletions or exon 21 (L858R) substitution mutations indicated for "identifying patients with NSCLC whose tumors have EGFR exon 19 deletions or exon 21 (L858R) substitution mutations and are suitable for treatment with a tyrosine kinase inhibitor approved by FDA for that indication." This could enable greater flexibility for clinicians in choosing the most appropriate therapeutic product based on a patient's biomarker status.[5]

Understanding of MOA and biomarker-therapeutic interaction
As noted in FDA's draft guidance:

> EGFR exon 19 deletions and exon 21 (L858R) substitution mutations are known to upregulate EGFR phosphorylation and respond to treatment with tyrosine kinase inhibitors of EGFR based on functional studies. Many mutations in EGFR exon 20 are tyrosine kinase inhibitor resistant, so these mutations would be excluded from this group or class.[5]

Sufficient clinical experience
As noted in FDA's draft guidance:

> Afatinib, erlotinib, gefitinib, osimertinib, and dacomitinib are all indicated for the treatment of patients with NSCLC whose tumors have EGFR exon 19 deletions or exon 21 (L858R) substitution mutations, so they will fall under one specific group or class (tyrosine kinase inhibitor indicated for the treatment of patients with NSCLC whose tumors have EGFR mutation exon 19 deletions or exon 21 (L858R) substitution mutations). Also it would not be appropriate to include therapeutic products in this specific group or class that only target resistant mutations, such as EGFR T790M and C797S, for which there may not be sufficient or consistent clinical experience.[5]

Demonstration of Analytical and Clinical Validity

The FDA guidance discusses considerations for demonstration of analytical and clinical validity as it applies to group labeling, although it does not provide an EGFR example for demonstration of analytical and clinical validity. For the discussion in this whitepaper, CDx tests that evaluate these nucleotide mutations are appropriately considered within Tier A because, unlike Tier B/C CDx tests, Tier A CDx tests do not require the identification of a complex rearrangement, do not evaluate different analytes that are directly or indirectly linked to the specific gene alterations, do not require a complex algorithm, and are often analytically

validated by comparing to bi-directional sequencing as the gold standard. Although the guidance does mention use of a reference test to detect false results and consideration of discordance between technologies, examples are needed to address the many outstanding questions, some of which are outlined in the **Discussion Questions** section below.

CASE STUDY 2: APPLICATION OF TIER B TO CDX ESTABLISHING *ALK* STATUS

Group or Class Definition and Understanding of MOA and Biomarker-therapeutic Interaction

Anaplastic Lymphoma Kinase (*ALK*) inhibitors belong to a class of compounds called Tyrosine Kinase inhibitors (TKI). These therapeutic products have proven effective in patients with metastatic NCSLC that is determined to be ALK-positive, reflecting the presence of a rearrangement in the *ALK* gene that functions as an oncogenic driver. During the past eight years, there have been five *ALK* inhibitors developed and approved, representing three generations of therapeutic products - crizotinib (first-generation), ceritinib, alectinib and brigatinib (second generation), and lorlatinib (third generation) - with additional drugs in development.

In general, subsequent generations of *ALK* inhibitors are designed to overcome limitations in potency, selectivity, brain penetrance, and mechanisms of resistance involving mutations within the *ALK* catalytic domain[6]. Studies have shown that patients can develop resistance to *ALK* inhibitors over time and that these mutations can represent a biomarker of response in previously treated patients.[7] Studies continue to shed light on the extent to which the various *ALK* inhibitors differ from each other in terms of mechanisms of resistance.

Sufficient Clinical Experience

The sequential development and approval, over a period of time, of next-generation *ALK* inhibitors using different CDx tests establishes increasing clinical experience and evidence demonstrating the clinical utility of an established class of therapeutic products in patient populations identified by different CDx tests. As mentioned previously, two of the five *ALK* inhibitors currently approved by the FDA for the treatment of patients with *ALK*-positive metastatic NSCLC do not further specify "as detected by an FDA approved test" because these two drugs were approved as 2nd line or 3rd line and greater treatments for patients who progressed on or may be intolerant to another *ALK* inhibitor that was previously approved as a first line treatment. While the safe and effective use of all of these drugs ultimately requires identifying patients with metastatic NSCLC whose tumors are *ALK*-positive as detected using an approved test, those drugs that are used subsequent to treatment with a prior (e.g. first-line) same-in-class therapeutic agent take advantage of the existing standard of care established by these previously approved same-in-class agents and their companion diagnostics.

There are presently three CDx tests approved by the FDA to identify patients with *ALK*-positive NSCLC appropriate for treatment with specific *ALK* inhibitors. Of particular note, each of the three approved tests measures distinctly different analytes (chromosomal DNA, protein, DNA sequence) using completely different platform technologies (FISH, IHC, NGS). These CDx tests therefore are best considered as Tier B tests, as additional evidence and supporting rationale would be necessary to support the expansion of any one test

with a class label.

Demonstration of Analytical and Clinical Validity

There are several technologies that have been developed to detect *ALK* rearrangements including IHC, fluorescent in-situ hybridization (FISH), and NGS. Because the analytical validity of each test and test platform is reviewed in the context of a single trial, the level of cross-platform divergence is unknown. The sensitivity of each of these tests varies, and interpretation of clinical data derived from the use of these different methods should be performed carefully.[8] Inconsistent results have been observed in the analysis of *ALK* rearrangements in NSCLC.[9] Core datasets and/or standard assays should be developed to facilitate harmonization of test sensitivity and analytical validity across tests within a test group. A recent study on *ALK* testing trends and patterns using Flatiron Health electronic health record-derived database reviewed results over 6 years for patients diagnosed with Stage IIIB or IV NSCLC. Average *ALK* testing rates increased over time from 32.4% in 2011 to 62.1% in 2016 and showed that FISH was the most common *ALK* testing method and may help understand relative performance of the various testing methods.[10] Harmonization efforts have been undertaken by comparing IHC testing methods across multiple centers and laboratories leading to standardized methods and interpretation criteria.[11]

In addition to each *ALK* diagnostic being required to demonstrate clinical validity in the context of the therapeutic for which it is a companion, for NGS based testing the FDA has required that the NGS panel test, for example, FoundationOneCDx, demonstrate clinical concordance to previously FDA-approved IHC/FISH tests.[12] While it is helpful to compare the performance of the NGS test with the IHC/FISH tests, if the various *ALK* therapies slightly vary in their mechanisms of action, one wonders if there should be an expectation of clinical concordance between the various diagnostics. In such cases, the same-in-class diagnostics category may have to be considered more carefully.

CASE STUDY 3: APPLICATION OF TIER C TO HOMOLOGOUS REPAIR DEFICIENCY AND SIMILAR TESTS

Poly (ADP-ribose) polymerases (PARP) have shown true promise in early clinical studies due to reported activity in *BRCA*-associated cancers. As a drug class, PARP inhibitors have had their greatest impact on the treatment of women with epithelial ovarian cancers (EOC). PARP inhibition exploits this cancer vulnerability by further disrupting DNA repair, thus leading to genomic catastrophe. Early clinical data demonstrated the effectiveness of PARP inhibition in women with recurrent EOC harboring *BRCA*1/2 mutations and those with platinum-sensitive recurrences. Three PARP inhibitors (olaparib, niraparib, and rucaparib) are now approved for use in women with recurrent EOC.[13]

These new therapeutics have demonstrated clinical use variously in treatment and maintenance settings, and more clinical trials are underway to expand use of this new generation of medicines.[14] Olaparib, Rucaparib, and Niraparib have all been approved with the requirement of a companion diagnostic, for certain indications. They are summarized in a recent FDA presentation.[15] These drugs have shown differential activity in patients with *BRCA* mutations or whose cancers demonstrate *BRCA* mutations or genomic scar-

ring resulting from homologous repair deficiency (HRD) of a variety of origins, including mutations, dele-tions, loss of heterozygosity (LOH), miRNA and DNA methylation.

Various diagnostic tests to detect *BRCA* or HRD have been approved: Myriad BRACAnalysisDx, Myriad myChoice, FoundationFocusCDxbrca and FoundationOneCDx. It is important to consider here that some of the tests only interrogate germline mutations in *BRCA* while others also detect tumor-derived mutations. Even with the approved diagnostics, there may be potential variation with the way homologous repair deficiency is defined (also referred to as genomic instability). In the case of one NGS panel, the HRD is represented by BRCA mutations and genomic score-based alteration called loss of heterozygosity (LOH).[16] This contrasts with another NGS-based diagnostic where *BRCA* mutations are supplemented by three algo-rithmic-score based alterations, namely telomere allelic imbalance (TAI), large-scale state transitions (LST) in addition to LOH.[17] Furthermore, recently a direct-to-consumer testing device has also secured FDA approval for detecting *BRCA* mutations, albeit not as a companion diagnostic to prescribe therapeutic, leading a prominent researcher in the field to worry that there may be insufficient testing of the *BRCA* pathological mutation with this test.[18]

BRCA certainly is gaining importance as a window into the tumorigenic process due to its role as a tumor suppressor, but there are even more genes implicated in the repair pathway that also seem to play a role. In addition to *BRCA*1/2, there are variously 15 or 17 other genes referred to as HRR pathway (homologous recombination and repair), where alterations in those genes are also being studied for response to PARPi therapies. The next iteration of Myriad's diagnostic named myChoicePlus will have an additional 90 genes compared to the original version.[19]

While these are exciting advances, the community will have to come together to define, classify, and har-monize these diagnostic devices as they are all likely going to apply to the same class of therapies, namely PARP inhibitor therapies. HRD or PARPi diagnostic devices, for lack of a better term, are sufficiently com-plex in their differences and nuances, that the average community physician may not be commensurate in understanding how each of them may detect slightly different tumor genotypes resulting in differences in clinical outcomes for the therapeutic.

Potential Implications for Clinical Trial Design

As FDA and industry consider these questions and other concepts to facilitate CDx development, the result-ing policies and their implications need to be considered in the broader clinical context. For example, cur-rent CDx development pathways and regulations can impact the flow of patients onto clinical trials because current regulatory guidance may favor enrollment strategies that utilize prospective patient selection on the basis of an investigational device exemption (IDE) that will eventually form the basis of the CDx. By compar-ison, enrollment strategies that utilize locally obtained testing performed outside of the auspices of the clin-ical trial for the purpose of enrollment, with storage of samples tested by the local lab test for retrospective bridging testing is currently permitted, primarily where the biomarker is very rare, but this approach may not be favored. This "retrospective approach" can also be associated with challenging-to-meet down-stream requirements such as collection of negative samples to be tested by the most prevalent local lab test used for eligibility determination.

Enrollment strategies that utilize prospective central confirmation via an IDE (if needed) for eligibility determination may result in duplication of testing if patients known to harbor the relevant biomarker are re-tested. This could create several concerns including duplication of testing, exhaustion of tissue sometimes requiring repeat invasive procedures to obtain more material necessary for central testing, and delays in patient enrollment during which the tissue is sent, accessioned, tested and results returned. These potential barriers to clinical trial participation, and the evidence they generate, should be carefully weighed against the potential benefits of this approach from an assay validation perspective.

Given these considerations, both regulators and sponsors may need to consider whether retrospective confirmation and enrollment of patients based on local testing could be sufficient to reduce the potential of duplicative testing and how to clearly articulate retrospective pathways that might be used for patient enrollment.

WHITEPAPER DISCUSSION QUESTIONS

- How should a same-in-class drug be defined in the context of a CDx group label?
  - □ What is the minimum number of drugs needed for creation of a group label?
  - □ How should variability in efficacy between drugs within a class be addressed?
- How can parity in measurement between tests within a test group be maintained?
  - □ Can tests be awarded a group label based upon comparison to a reference test?
  - □ How should harmonization of measurement between technologies be achieved?
  - □ What if harmonization cannot be obtained as newer technology is more accurate and provides for more efficient use of tissue (NGS)?
- When demonstrating analytical and clinical validity with reference to a comparator test, what characteristics should be considered when choosing the comparator test?
  - □ Should the first-in-kind or first approved test be the de facto comparator for all tests within a group label?
  - □ For example, in the case of EGFR, the Cobas[20] may be the reference diagnostic to harmonize to, but for BRAF[21], Biomerieux test may be the better reference diagnostic to harmonize to instead of Cobas. Examples of successful harmonization efforts exist, such as for validation of blood glucose monitors in which a standardized enzyme-based assay was used to establish a set range of performance values that all tests are required to meet. Further, the *Friends* TMB Harmonization project is an example of a molecular biomarker harmonization effort in which the use of NCI's The Cancer Genome Atlast (TCGA) data, cell lines, and clinical samples were used to help define and establish analytical performance thresholds.[22,23]
  - □ Should clinical trial data demonstrating validity be required for the comparator test?
  - □ Due to limited access to quality banked samples, are there alternate approaches that can be used?
- How should concordance be demonstrated for a pan-tumor indication?
  - □ Does concordance have to be demonstrated within each separate tumor type or is across a number of tumor types acceptable?
  - □ How many tumor types would be necessary?
- Are there situations when "unacceptable concordance" is acceptable?

CONCLUSION

The purpose of considering these proposed clinical trial strategies is to accelerate development of combination therapies that include an unapproved PD-(L)1 through regulatory flexibility, to accelerate the potential utilization of combination therapies across a more diverse range of tissue types, and to potentially alleviate noted challenges by some drug developers.

Additional considerations may also need to be explored to further facilitate the development of combination therapies containing immuno-oncology agents.

- Obtaining sufficient data on safety and efficacy will be important to consider both in the context of regulatory decision-making and in providing adequate data for patients and physicians who may be considering several therapeutic options.
- Improving the understanding of how preclinical analytical data or animal models can inform the toxicity profiles between an approved PD-(L)1 and an unapproved PD-(L)1 should be further defined.
- Creating incentives or policies to encourage greater collaboration between sponsors of approved PD-(L)1s and sponsors seeking to conduct combination studies with a PD-(L)1 backbone could be explored.

FUTURE CONSIDERATIONS

Areas that may require additional guidance:

- **Interactions Between FDA and Drug Sponsors.**
  - » Define parameters and timing for conversations between FDA and sponsors evaluating two or more drugs for use in combination.
  - » Define parameters for FDA input on adaptations or for the pre-specification of adaptations
- **Class Definition.**
  - » Define process for determining a drug class
  - » Demonstration of early activity
  - » Define how preclinical analytical data or animal models can inform the toxicity profiles between an approved PD-(L)1 and an unapproved PD-(L)1
  - » Suggest strategies for demonstrating early activity for drugs being developed in combination
  - » Suggest strategies for demonstrating the biological rationale for use of a combination
  - » Suggest strategies for demonstrating a combination has a significant therapeutic advance over existing therapeutic options
  - » Establish general criteria for when factorial clinical trial designs are not needed and data

References:

1. FDA's Guidance for Industry and FDA Staff: In Vitro Companion Diagnostic Devices, August 2014, page 11; FDA's Draft Guidance for Industry and FDA Staff: Developing and Labeling In vitro Companion Diagnostic Devices for a Specific Group or Class of Oncology Therapeutic Products Guidance for Industry, December 2018

2. United States Food and Drug Administration. (2017, June 22). FDA grants regular approval to darafenib and trametinib combination for metastatic NSCLC with BRAF V600E mutation [Press release]. Retrieved from https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-regular-approv-al-dabrafenib-and-trametinib-combination-metastatic-nsclc-braf-v600e.

3. United States Food and Drug Administration. (2017, November 30). FDA announces approval, CMS proposes coverage of first breakthrough-designated test to detect extensive number of cancer bio-markers [Press release]. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-an-nounces-approval-cms-proposes-coverage-first-breakthrough-designated-test-detect-extensive.

4. United States Food and Drug Administration. (2014). In Vitro Companion Diagnostic Devices: Guidance for Industry and Food and Drug Administration Staff. Retrieved from https://www.fda.gov/index.php/media/81309/download.

5. United States Food and Drug Administration. (2018). Developing and Labeling In vitro Companion Diagnostic Devices for a Specific Group or Class of Oncology Therapeutic Products Guidance for Industry: Draft Guidance. Retrieved from https://www.fda.gov/media/120340/download

6. Shaw, Alice (2018) Newer Targeted Agents in NSCLC: ALK Inhibitors https://www.cancernetwork.com/asco-lung-cancer/newer-targeted-agents-nsclc-alk-inhibitors

7. Shaw, Alice T et al. (2019) ALK Resistance Mutations and Efficacy of Lorlatinib in Advanced Anaplastic Lymphoma Kinase-Positive Non–Small-Cell Lung Cancer https://ascopubs.org/doi/full/10.1200/JCO.18.02236

8. Zhang X et al. (2017) Oncotarget 8(43):75400 Diagnostic accuracy of PCR for detecting ALK gene rear-rangement in NSCLC patients: A systematic review and meta-analysis [PMID:29088875]

9. Mattsson JSM et al (2016) BMC Cancer 16: 603 Inconsistent results in the analysis of ALK rearrange-ments in non-small cell lung cancer

10. Illei PB et al (2018) JCO Precision Oncology – ALK Testing Trends and Patterns Among Community Practices in the United States

11. von Laffert M et al (2014) J Thoracic Oncology 9 (11): 1685 – Multicenter Immunochemical ALK-Testing of Non-Small-Cell Lung Cancer shows high concordance after harmonization of techniques and inter-pretation criteria

12. Summary of Safety & Effectiveness Document https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019B.pdf (Pg 11)

13. Cook, SA and Tinker AV (2019) PARP Inhibitors and the Evolving Landscape of Ovarian Cancer Management: A Review https://link.springer.com/article/10.1007%2Fs40259-019-00347-4

14. Franzese, E et al. (2019) PARP inhibitors in ovarian cancer. Cancer Treatment Reviews 73:1-9 https://www.cancertreatmentreviews.com/article/S0305-7372(18)30200-7/pdf

15. Ison, Gwynn (2018) FDA Perspective: Evolving Development of PARP inhibitors https://www.fda.gov/media/114655/download

16. https://www.fda.gov/medical-devices/recently-approved-devices/foundationfocus-cdx-

brca-loh-p160018s001

17. https://myriad.com/products-services/precision-medicine/mychoice-hrd/

18. Murphy H (2019), Don't Count on 23andMe to Detect Most Breast Cancer Risks, Study Warns. New York Times (April 16, 2019) https://www.nytimes.com/2019/04/16/health/23andme-brca-gene-testing.html

19. AZ to Use Myriad's myChoice HRD Plus Test in Phase III Trial https://www.clinicalomics.com/topics/precision-medicine-topic/cancer/az-to-use-myriads-mychoice-hrd-plus-test-in-phase-iii-trial/

20. Summary of Safety & Effectiveness Document https://www.accessdata.fda.gov/cdrh_docs/pdf12/P120019B.pdf

21. Summary of Safety & Effectiveness Document https://www.accessdata.fda.gov/cdrh_docs/pdf12/P120014B.pdf

22. www.focr.org/tmbhttps://www.ncbi.nlm.nih.gov/pubmed/30664300

23. www.focr.org/tmb and/or https://www.ncbi.nlm.nih.gov/pubmed/30664300

FRIENDS of CANCER RESEARCH

A FRIENDS OF CANCER RESEARCH WHITEPAPER

IMMUNO-ONCOLOGY COMBINATION DRUG DEVELOPMENT FOR PATIENTS WITH DISEASE PROGRESSION AFTER INITIAL ANTI-PD-(L)1 THERAPY

BACKGROUND

During the past five years, one CTLA-4 and six PD-(L)1 inhibitors have gained approval by the Food and Drug Administration (FDA) for a variety of malignancies including melanoma, non-small cell lung cancer (NSCLC), head and neck squamous cell cancer (HNSCC), lymphoma, urothelial carcinoma, and microsatellite instability-high (MSI-H) cancers.[1,2] The use of PD-(L)1 inhibitors as single agent therapies in first- and second-line settings is becoming the standard of care for several indications, such as NSCLC, increasing the number of patients being exposed to these IO therapies earlier in the course of their disease.[3] However, durable benefit from these PD-(L)1 monotherapies is only observed in a small fraction of patients as many of these patients appear to develop primary resistance.

Novel combination immunotherapy regimens using PD-(L)1 inhibitors as a backbone that modulate different immune pathways simultaneously or in tandem and override the risk of acquired resistance to a single immunotherapy agent are being developed and studied in different indications.[4–8]

Given the potential for overcoming anti-PD-(L)1 resistance using a combination drug approach, many patients who could benefit from these combination therapies are those whose disease has progressed during or after anti-PD-(L)1 monotherapy. However, it is not fully understood how these previously treated patients will respond to re-exposure to additional anti-PD-(L)1 therapies given in combination with additional agents. In some cases, the PD-(L)1 inhibitors or their combination agents may be already FDA-approved, but there may be cases in which they are not. Several scenarios exist, including the combination of two or more investigational drugs, an investigational drug with a previously approved drug for a different indication, or two (or more) previously approved drugs for a different indication as

CONTRIBUTORS

JULIE BRAHMER
JOHNS HOPKINS UNIVERSITY

ILLARIA CONTI
EMD SERONO

BHARANI DHARAN
NOVARTIS PHARMACEUTICAL
CORPORATION

ERIC RUBIN
MERCK INC.

T.J. SHARPE
PATIENT ADVOCATE

RYAN SULLIVAN
MASSACHUSETTS GENERAL HOSPITAL

MARC THEORET
FDA

MAURIZIO VOI
NOVARTIS PHARMACEUTICAL
CORPORATION

a novel combination therapy. These scenarios have been previously explored and innovative strategies that properly assess the contribution of components of the combination drug regimen have been discussed.[9]

The mechanisms of resistance to PD-(L)1 inhibitors are not well understood as some patients may not respond to these inhibitors at all and develop progressive disease right away, while others may respond to treatment initially or partially, and eventually develop progressive disease. A better understanding of the mechanisms by which patients develop refractory or relapsed disease will help guide subsequent drug alternatives for patients whose disease progressed during or after PD-(L)1 inhibitors. Moreover, refining the definition of disease that has relapsed or has become refractory to treatment will also further elucidate the population being studied, which will help guide the interpretation of the study findings.

Exploring the development of promising combination therapies using PD-(L)1 therapy as a backbone is imperative and a rational next-step to overcome resistance to monotherapies. However, knowing that the study population will most likely be composed of both PD-(L)1 inhibitor-pre-treated and PD-(L)1 inhibitor-naïve patients, it is crucial to discuss any additional considerations that the pre-treated population may require to closely monitor the safety and efficacy of the novel combination drugs while maintaining proper equipoise. For instance, would there be a lack of equipoise if a patient whose tumor progressed after anti-PD-(L)1 therapy is randomized to the single-agent PD-(L)1 inhibitor control arm in a late-stage randomized controlled clinical trial? And would this be dependent on disease type? Because there are not enough data to guide treatment decisions in this rapidly-growing pre-treated population, there is great uncertainty as to whether a patient's tumor would respond when re-exposed to the same agent in combination, to monotherapy with another same-in-class agent, or even if the patient would respond to an inhibitor that targets PD-(L)1 if they have received a PD-1 inhibitor (or vice-versa). It is not fully understood whether the patient's immune system will behave similarly to an immunotherapy-naïve patient, or if further considerations, such as timing or a specific washout period from anti-PD-(L)1 therapies, will impact subsequent response to re-exposure to additional anti-PD-(L)1 therapy.

Friends of Cancer Research (*Friends*) convened a group of experts from various healthcare sectors to discuss important considerations to keep in mind that patients whose disease has progressed after anti-PD-(L)1 therapies face when seeking to enroll in clinical trials testing combination therapies including a PD-(L)1 inhibitor. The objectives of the working group and this whitepaper encompass the development of a framework that will help harmonize the definition of a population whose disease has progressed after PD-(L)1 inhibitors, and the identification of flexible trial design strategies and innovative approaches that allow for earlier exploration and modifications based on interim analyses, and the characterization of roles that external data may have to support immuno-oncology combination trials. The primary goal of these discussions is to propose actionable, practical, and rational solutions for the unique needs of patients whose tumors have progressed after anti-PD-(L)1 therapies, which will promote the development of drug combinations and increase accessibility to better treatment options for this growing population.

Framework for the Harmonization of a Definition for a Population Whose Disease has Progressed After Initial Anti-PD-(L)1 Therapies

Disease that has progressed past treatment can be referred to as (1) relapsed disease when the disease has initially responded positively to treatment but later reappeared or grew after having been in remission for a time, or (2) refractory disease when the disease has not responded positively to treatment or even progressed during treatment. However, relapsed disease can become refractory to the treatment it once responded to, so it is not surprising that these two terms are often confused, or at times used interchangeably. Actually, various publications have repeatedly combined both relapsed and refractory (r/r) diseases into a single category. As the use of this combined term to define solid tumors that ultimately fail to respond to treatment increases, and as the community learns more about the unique patterns of response to immunotherapies, it is important to accurately define what is meant by r/r disease and refine these terms within the context of immunotherapies, more specifically, after PD-(L)1 inhibitors.

Assessing response to PD-(L)1 inhibitors is complex because clinical response to immunotherapies is unique and does not follow the established patterns observed with cytotoxic therapies. Various reports have shown delayed clinical responses in studies with immunotherapies where patients have shown an increase in total tumor burden, either by growth of existing lesions or appearance of new lesions, followed by decreased tumor burden[10]. This atypical response pattern is known as pseudoprogression and seems to be unique to immunotherapy. If such a response was evaluated using the conventional Response Evaluation Criteria In Solid Tumors (RECIST) criteria established to assess whether a solid tumor responded, stayed the same, or progressed, patients receiving immunotherapies would be classified as having progressive disease even if their tumors actually responded to treatment.[11–14] Several efforts subsequently addressed this challenge,[10,15–17] which led to the development of response criteria that incorporated RECIST 1.1 recommendations, but is better able to address the atypical patterns of response associated with immunotherapies: iRECIST.[18] Use of iRECIST would ensure consistency in the way the trials were designed and the way data was collected, which would enable the comparison of results across trials. It is important to note, however, that to date, no drug has been approved based on immune-related response criteria only.

The complexity of identifying clinical efficacy, or lack thereof, in patients receiving PD-(L)1 inhibitors is one of the remaining challenges that confounds the definition of a population of patients whose disease has truly progressed past PD-(L)1 inhibitors. These remaining challenges have been acknowledged by the research community, launching several initiatives that further investigate, discuss and develop strategies to align definitions to better characterize patients with r/r disease after initial anti-PD-(L)1 therapy, such as the Society for Immunotherapy of Cancer (SITC) PD-(L)1 Resistance Definition Task Force. Open discussion among experts will drive research that investigates mechanisms of resistance to PD-(L)1 inhibitors, and thus promote a greater understanding on how patients who progress past these therapies should be treated.

Friends conducted a survey with six pharmaceutical companies that have a marketed FDA-approved PD-(L)1 inhibitor to better assess the variability in definitions for r/r disease being utilized in current clinical trials of PD-(L)1 inhibitors, and to learn whether the definition is harmonized across each pharmaceutical company. All six companies surveyed expressed interest in the idea of a harmonized definition of r/r disease and commented this is an area where further guidance is necessary. Three of the six companies (50%) surveyed had

a company-wide harmonized definition of r/r disease, and those who did not mentioned they are working on incorporating a more consistent definition of disease progression into their clinical trials (Table 1). The survey also asked sponsors to share their definition of r/r disease (if available) in order to compare the variability across company definitions.

When analyzing the definitions provided by the different sponsors, three main principles emerged. These revolved around 1) identifying adequate exposure to anti-PD-(L)1 therapies by specifying dose or length of anti-PD-(L)1 therapy that was used before disease progression; 2) identifying and confirming progression of disease, including the type of scan, or the timing at which this scan would be done; and 3) identifying the likelihood of responding to re-exposure of anti-PD-(L)1 therapies (Table 2).

Some pharmaceutical companies raised concerns about a harmonized r/r definition as they acknowledge there are considerations that need to be taken into account when defining r/r disease in different populations, as there are various factors that may influence the evaluation of disease progression. Seeing as how the assessment of disease progression in patients treated with PD-(L)1 inhibitors is so nascent, the influence of factors such as cancer type, the natural history of disease, the biology of the drug assessed, and the timing of scans need to be further investigated within this unique context.

Table 1: State of Harmonization of Relapsed/Refractory Disease Definitions of Six Major Pharmaceutical Companies with a Marketed FDA-approved PD-(L)1 Inhibitor

Question	Sponsor A	Sponsor B	Sponsor C	Sponsor D	Sponsor E	Sponsor F
Harmonized r/r disease definition within company	Yes	Yes	No	No	No	Yes
Comments	Variations can occur related to specific tumor types	Definitions are specific for primary, secondary, and adjuvant resistance	Due to differences in disease setting, prior therapy, and stage of drug development	Company does not have enough studies in this space yet	Company has proposed criteria but no harmonized language yet	Terminology for "progressed on or recurred after an anti-PD-1 agent"
Definitions	Patients must have progressed on treatment with an anti-PD1/(L)1 monoclonal antibody (mAb) administered either as monotherapy, or in combination with other checkpoint inhibitors or other therapies. PD-1 treatment progression is defined by meeting all of the following criteria: • Has received at least 2 doses of an approved anti-PD-1/(L)1 mAb. • Has demonstrated disease progression after PD-1/(L)1 as defined by RECIST v1.1. The initial evidence of disease progression (PD) is to be confirmed by a second assessment no less than four weeks from the date of the first documented PD, in the absence of rapid clinical progression.1,2 • Progressive disease has been documented within 12 weeks from the last dose of anti-PD-1/(L)1 mAb.	The recommendation is to generally include all categories of resistance, but use "primary," "secondary," and "adjuvant" resistance for patient stratification purposes. • Primary Resistance: patients must have experienced progressive disease (PD) within 12 weeks of initiation of PD-(L)1 inhibitor-based treatment. Radiographic confirmation of the PD must be documented*, after a minimum of 4 weeks after the initial identification of progression, unless: i) investigator confirms clinical progression/deterioration attributed to PD, or ii) the first radiographic assessment indicated critical tumor growth by imaging (size or location). *The purpose of radiographic confirmation is to minimize inclusion of patients with pseudoprogression; however, study teams have the option to waive this requirement.	Generally, we have incrementally modified our definition of progression to be more consistent across clinical trials that include anti-PD-(L)1 therapies in solid and hematological diseases, to account for tumor flare, and to incorporate feedback received from Health Authorities in our programs. » For proof-of-concept studies only: - CPI refractory: best response by RECIST is PD - CPI-responsive: best response by RECIST is PR or SDx 6 months followed by PD	At present definitions are still being discussed	We will be happy to share the definition, when harmonized	We believe there are at least 3 distinct patient populations: • Patients who do not respond & progress on anti-PD-(L)1 (or within 6 months of treatment) • Patients who progress after initial response while on anti-PD-(L)1 • Patients who progress after initial response to anti-PD-(L)1 off drug

Question	Sponsor A	Sponsor B	Sponsor C	Sponsor D	Sponsor E	Sponsor F
Definitions (con't)	1.Seymour et al; iRECIST: Guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 18: e143-52 2.This determination is made by the investigator. Once PD is confirmed, the initial date of PD documentation will be considered the date of disease progression.	• Secondary Resistance: patients must have experienced progressive disease (PD), either during or within 3 months of discontinuing treatment with anti-PD-(L)1-based therapy, occurring after previous clear benefit (any complete (CR) or partial response (PR)), or after previous stable disease (SD). No requirement for radiographic confirmation of progression. • Adjuvant Resistance: patients with documented loco-regionally and/or systemic relapse of their disease occurring <6 months after the last dose of anti-PD-(L)1-based systemic adjuvant treatment. In an effort to ensure that Phase 1/2 signal-finding studies are conducted in a population that is not potentially still sensitive to their prior anti-PD-(L)1 therapy, we are not including patients that experience progression > 3 months after cessation of anti-PD-(L)1-based therapy in the metastatic setting or > 6 months in the adjuvant setting, regardless of the rationale for cessation of treatment. In addition, we are generally recommending exclusion of patients that have received intervening systemic therapy following prior anti-PD-(L)1-based therapy.				We believe it is important to study these patient populations separately. An additional patient population to consider in this context (and not covered by "relapsed/refractory" definitions) is patients with stable disease while on PD-(L)1 inhibitors Additional criteria: • Patients must have confirmed disease progression on anti-PD-(L)1 therapy as defined by RECISTv1.1 and further defined as: • Previous exposure to anti-PD-1 containing regimen for at least 12 consecutive weeks, and • Progression must be while on treatment with anti-PD-1 or within 6 months of discontinuing anti-PD-1, and regardless of any intervening therapy

Table 2: Principles and Considerations for the Definition of Relapsed/Refractory Disease

Principle	Considerations	Example
1. Identifying adequate exposure to anti-PD-(L)1 therapies	• Dose of anti-PD-(L)1 therapies • Length of anti-PD-(L)1 therapies	Has received at least 2 doses of an approved anti PD-(L)1 therapy.
2. Identifying progression of disease		
Evaluation of progression	• Tumor-specific criteria • Adequacy of measurement method	• Different cancer types may require different approaches to evaluate progression. • Progression in prostate cancer is evaluated using the PCWG3 criteria,[19] and in glioblastoma, the modified RANO criteria.[20]
Confirmation of progression	• Ability to address pseudo-progression • Timing • Equipoise: patients with immediate life-altering disease & timing	• Radiographic confirmation of disease. • Documented after a minimum of 4 weeks of initial identification of progression.
3. Identifying likelihood of responding to re-exposure of anti-PD-(L)1 therapies	• Time from anti PD-(L)1 therapy initiation • Time from last anti-PD-(L)1 therapy administration • Refractory – disease progressed because it did not respond to drug • Relapsed – disease initially responded to drug & then progressed • Adjuvant vs. metastatic setting • Intervening treatment • Does resistance to one drug within class mean resistance to all drugs within class?	

Considerations for the Assessment of Combination Drugs Using a PD-(L)1 Inhibitor Backbone in Patients Whose Disease Progressed After PD-(L)1 Inhibitors

Combination drug trial design strategies: maintaining a fine balance between efficiency and equipoise in patients who have been previously treated with PD-(L)1 inhibitors

The development of innovative combination drug clinical trial designs, such as master protocol platform designs and seamless adaptive designs that allow for modifications based on interim analyses while achieving the appropriate statistical rigor, would greatly benefit patients and enable the collection of data to support clinical decision making in this unique population of previously-treated patients.

In addition to striking the right balance between providing potentially life-saving therapies to advanced cancer patients with very few therapeutic options, and minimizing a patient's exposure to ineffective and harmful therapies by rapidly identifying patients who do not derive any benefit from their assigned therapy (via early efficacy or futility evaluation), combination drug trials must also determine the contribution of each of the investigational drugs assessed in combination.

Several combination drug trial designs and approaches have been previously explored to help isolate the treatment effects of the agents used in combination.

- **2x2 factorial design.** Several reports have comprehensively reviewed the benefits and challenges of using the most optimal 2x2 factorial clinical trial design (e.g. SOC vs. A vs. B vs. A+B) to understand the attribution of effects for the single agents and their combination; however, this approach may generate duplicative data and reduce the lack of equipoise created when patients are assigned to the control arm knowing they are predicted to receive no benefit from it.
- **Randomized early-stage clinical trials.** Assessing efficacy and safety through randomized early-stage clinical trials, such as randomized, open label, phase 2 trials that incorporate a "master" protocol framework (such as umbrella, basket, or platform trials) would enable sponsors to identify a treatment arm that shows the best activity in a smaller number of patients and would signal the need to increase development efforts.
- **Single-arm trials.** Another alternative method involves supportive single-arm trials, when randomized trials may not be feasible. In such cases a single-arm trial may be the next best approach to translate preliminary results into predictions of Phase 3 benefit and risk. In the absence of randomized trials, however, a comprehensive evaluation of the contribution of each individual component in both preclinical and clinical data would be needed, given that time-to-event endpoints, such as OS, will likely not be informative.
- **Common controls.** The use of a common control may incorporate the flexibility needed to better assess efficacy and safety when there is a desire to minimize the number of patients randomized to a control arm. The i-SPY2 trial used this method to more rapidly accrue patients and minimize the number of patients assigned to a standard of care (SOC) control arm that may be lacking equipoise as in the case of previously-treated patients enrolling in a combination drug trial using a PD-(L)1 backbone.[21] In the i-SPY2 trial, the FDA supported the use of a common control arm, but additional guidance and further work to better characterize this type of design is necessary given that this is not a common method to assess clinical benefit.

Additionally, the FDA has generated guidance on the Codevelopment of Two or More New Investigational Drugs Used in Combination, which describes criteria for knowing when codevelopment is appropriate, and identifies various development strategies as well as regulatory considerations.[22]

All these strategies seek to address one of the main concerns about investigating the efficacy and safety of a combination regimen that has a PD-(L)1 inhibitor backbone in patients whose disease has progressed after an initial PD-(L)1 inhibitor: **Will the patient's disease be able to respond to the challenge by either the same PD-(L)1 inhibitor or a similar in-class inhibitor when used in combination with another drug or biologic?** This is not a particularly novel question, given that there have been several studies where patients treated with earlier-generation therapies have been subsequently re-challenged with a same-in-class novel agents and demonstrated clinical benefit (e.g., retreatment of advanced NSCLC patients with later-generation ALK inhibitors after becoming resistant to a first-generation ALK inhibitor[23,24]). However, if focusing on immunotherapy, much can be learnt from the first trials assessing the use of PD-1 inhibitors (nivolumab and pembrolizumab) in patients who developed melanoma that is refractory to CTLA-4 inhibitor (ipilimumab), another immune checkpoint inhibitor (Table 3).

Table 3: Characteristics of KEYNOTE-001, KEYNOTE-002 and CheckMate 037, Clinical Trials Investigating the Efficacy of PD-1 Inhibitors in Patients with Advanced Melanoma who Progressed After Anti-CTLA-4 Therapy

	KEYNOTE-001 (Robert et al., 2014)	KEYNOTE-002 (Ribas et al., 2015 & Hamid et al., 2017)	CheckMate 037 (Weber et al., 2015 & Larkin et al., 2017)
Clinical trial type	Randomized, dose-comparison, open label, expansion cohort of a phase 1 international trial	Randomized, controlled, phase 2 international trial	Randomized, controlled, open label, phase 3 international trial
Number of patients	173 given pembrolizumab, 89 at 2mg/kg, and 84 at 10mg/kg	357 given pembrolizumab (178 at 2mg/kg, 179 at 10mg/kg) and 171 given investigators choice chemotherapy (ICC)	268 given nivolumab and 102 given ICC
Definition of anti-CTLA-4 refractory melanoma population	• Progressive, measurable, unresectable melanoma • Previously treated with at least 2 doses of ipilimumab 3 mg/kg or higher administered every 3 weeks • Confirmed disease progression using immune related response criteria within 24 weeks of the last dose of ipilimumab • Previous BRAF or MEK inhibitor therapy or both (if BRAF V600 mutant-positive) and no limitations on the number of previous treatments	• Histologically or cytologically confirmed unresectable stage III or stage IV melanoma not amenable to local therapy • Confirmed disease progression within 24 weeks of the last ipilimumab dose • Minimum two doses, 3 mg/kg once every 3 weeks; • Previous BRAF or MEK inhibitor therapy or both (if BRAF V600 mutant-positive)	• Histologically confirmed, unresectable stage IIIC or IV metastatic melanoma • Patients with BRAF wild-type tumors must have had progression after anti-CTLA-4 treatment, such as ipilimumab, and patients with a BRAF V600 mutation-positive tumor mutation must have had progression on anti-CTLA-4 treatment and a BRAF inhibitor
Crossover	• N/A	• Allowed • Effective crossover rate= 58%	• Prohibited until the interim analyses • High percentage of patients in the ICC arm withdrawing consent (17%)

KEYNOTE-001 started as a phase 1 adaptive clinical trial that sought to define the safety and tolerability of pembrolizumab in patients with advanced solid tumors (reviewed in Kang et al.).[25] Although these initial study cohorts were not powered for efficacy, a substantial antitumor activity was observed, which provided the necessary rationale for an expansion randomized dose-comparison cohort of a phase 1 trial investigating pembrolizumab in patients with advanced and ipilimumab-refractory melanoma.[26] The definition of their study cohort used the recently developed immunotherapy-related response criteria guidelines[15] to ensure they were studying patients who had truly progressed after their initial immunotherapy (ipilimumab). Moreover, the adaptive design used in this trial was key in the early identification of substantial antitumor activity that led to the accelerated approval of pembrolizumab for unresectable or metastatic melanoma with disease progression after ipilimumab and, if BRAF V600 mutation positive, a BRAF inhibitor. Following KEYNOTE-001, KEYNOTE-002, a randomized, controlled, phase 2 trial, conducted a safety and efficacy study between patients treated with two different doses of pembrolizumab and investigator's choice of chemotherapy (ICC) in an equally defined population of ipilimumab-refractory melanoma.[27] This trial had planned two interim analyses that would allow for the identification of early response outcomes.

CheckMate 037 was a randomized, controlled, open-label, phase 3 trial that compared nivolumab with ICC in a population of ipilimumab-refractory melanoma patients. The trial design included an interim analysis assessing objective response as the primary analysis in a predefined population.[10] Moreover, a descriptive interim progression-free survival (PFS) analysis was also conducted in the intention-to-treat population at the same timepoint as the first analysis.

These trials initially demonstrated clinically meaningful improvements in objective response and PFS respectively, as well as fewer toxic effects compared to patients treated with ICC. Final analyses for KEYNOTE-002 and CheckMate 037 trials showed improvement in overall survival as well as durable response with the PD-1 inhibitors; however, these were not statistically significant.[28,29] Various factors could have contributed to the lack of significance in overall survival between the treatment and control arms, including allowing crossover between treatment groups. In KEYNOTE-002, the effective crossover rate was 58%, while in CheckMate 037, prohibiting crossover until the interim analysis could have been the reason why a high fraction of patients in the ICC arm withdrew consent.

Trial design determinations, such as whether crossover would be allowed or not, hinge on a fine balancing act between a trial's ability to detect significant drug efficacy and maintaining proper equipoise. All data derived from all stages of drug development (e.g. preclinical, early clinical trial, late and confirmatory trial, etc.) should be considered to make these determinations, and trialists are required to make trial design and statistical determinations that provide patients the care most likely to benefit them. This is the impetus behind the need for more flexible clinical trial designs that are able to meet the necessary statistical rigor for approval while placing the patient's safety and interests first and providing them a choice when preliminary findings reveal a potential lack of equipoise.

Currently, a few active trials are assessing the clinical utility of combination drugs using PD-(L)1 inhibitors as a backbone in patients whose disease progressed after anti-PD-(L)1 therapy, while adopting a flexible trial design, which allows for greater adaptability to changes driven by earlier assessment of patient safety and efficacy outcomes. As trial data becomes available, it will be important to assess how the added flexibility

of the platform trials contributes to a finer balance between trial efficiency and equipoise. Examples of such trials include:

- The HUDSON study is a phase 2 study that assesses novel biomarker-directed drug combinations that include durvalumab, an approved PD-(L)1 inhibitor, as a backbone in patients with NSCLC who progressed on an anti-PD-(L)1 containing therapy (NCT03334617).[30] This ongoing trial is an umbrella study with a modular design, which is able to conduct initial assessments of efficacy, safety, and tolerability in multiple treatment arms. This flexible design also allows trialists to add future treatment arms as needed via protocol amendment.
- The PLATforM study is a randomized phase 2 study of the novel PD-1 inhibitor Spartalizumab in combination with novel drugs and biologics in patients with unselected, unresectable, or metastatic melanoma previously-treated with PD-(L)1 ± CTLA-4 inhibitors, and a BRAF inhibitor, alone or in combination with a MEK inhibitor, if BRAF mutation positive (NCT 03484923).[31] In addition, based on an extensive tumor biopsy and blood sampling at baseline and on treatment, a key secondary endpoint of the study is to assess the percentage of patients with a favorable biomarker profile, as defined by favorable changes in number of cells expressing T-cell markers.

Challenges for the Assessment of Combination Immuno-oncology Therapies in the Adjuvant Setting

Combination IO trials are not only assessing response in patients with advanced disease who no longer have treatment options. Immune checkpoint inhibition is also being used earlier in the disease course, more specifically, in the adjuvant setting. A couple of scenarios that are becoming increasingly common in the clinic include the use of adjuvant anti-PD-(L)1 therapy in patients with Stage III/IV resected melanoma and anti-PD-(L)1 therapy after definitive chemoradiation therapy in patients with Stage IIIB NSCLC.

- Scenario A: Patient with Stage III melanoma treated with PD-(L)1 inhibitor monotherapy recurs while on adjuvant anti-PD1 therapy. This recurrence represents resistance to therapy.
- Scenario B: Patient with Stage III melanoma treated with PD-(L)1 inhibitor monotherapy develops recurrent disease after completing the planned treatment cycles or sooner (e.g., in case of toxicity). Recurrence may represent resistance to therapy, but this determination is less clear.

It is well described that patients with Stage IV melanoma who have a complete response and then discontinue therapy may again respond when re-challenged with the same or similar therapy. It stands to reason that patients who discontinue therapy after completing a planned year of adjuvant therapy may respond to re-treatment in the setting of disease recurrence. However, the magnitude of the effect may be smaller if there is ongoing target engagement of the PD-1 antibody with T-cells. As this occurs for at least 12 weeks and perhaps up to 6 months, by convention, it has been accepted generally to consider a patient resistant to anti-PD-(L)1 therapy if the last dose was within 3 months, and in some definitions, 6 months. This convention is reflected in Table 1. It is imperative to develop and implement a consistent framework for the proper documentation of response in patients who are re-challenged with immune checkpoint inhibitors. Whether these are given as monotherapies or in combination, either on clinical trials or off study, it will be critical to determine what the true rate of "resistance" is in patients whose disease progress after adjuvant

PD-(L)1 inhibitor therapy and whether there are predictive factors, including timing of last adjuvant dose to time of recurrence or specific biomarkers that may be useful in patient risk stratification.

Advantages to the Use of External Data for the Assessment of Combination Immuno-oncology Therapies

It is important to explore the use of external data to complement clinical trial data and further confirm the benefit of the combination regimen. Several efforts are being carried out to better understand the use of synthetic control arms derived from historical clinical trial data to augment clinical trial data, especially in instances where assigning a randomized control arm lacks equipoise or is not possible due to scarcity of patients, or when elevated crossover rates may compromise control arm data and make it unusable (2018 and *Friends* 2019 Annual Meeting whitepaper on external controls).[32]

Assessing the safety and efficacy of combination drugs with an anti-PD-(L)1 therapy backbone in patients whose disease has progressed after an initial PD-(L)1 inhibitor is not straight forward and will require out-of-the-box thinking. There are several remaining questions that need to be further discussed and potentially several areas that require further evidence development to better inform treatment alternatives for this unique and growing population of patients previously treated with an immunotherapy.

Remaining Questions or Areas that Warrant Evidence Development and Continued Discussion

- What type of data needs to be collected to enable a better understanding of potential patient response to re-challenge by either the same PD-(L)1 inhibitor or a similar in-class inhibitor when used in combination with another drug or biologic?

- Consistency in collecting data to determine timing of progression—will a harmonized method for data collection help investigate the association with likelihood of response to re-challenge?

- What preclinical models or clinical translational data would be helpful to identify combinations most likely to be effective in patients who have progressed on PD-(L)1 therapies?

- What is the role of biomarkers in better understanding the drug combinations most likely to be effective in patients who have progressed on PD-(L)1 therapies?

- Randomization approaches that allow for earlier examination of effect via interim analyses
  - □ Earlier identification of patients who may not be deriving benefit from monotherapy arm using, for example, response adaptive randomization

- What are some statistical considerations or approaches to evaluate early efficacy or early futility in these trials?

- Statistical considerations for addressing crossover
  - □ Knowing that crossover is a common issue when preclinical and early phase data for a novel

agent demonstrates significant antitumor activity, what are some innovative statistical strategies to properly deal with crossover?

◻ An example may include crossover-adjusted overall survival using rank preserved structural failure time (RPSFT). Under certain assumptions, the RPSFT model can be used to identify what survival difference would have been observed had all patients remained on the original assigned treatment

◻ Not all statistical approaches apply to all cases. Several approaches may be needed.

• Is there a role for non-invasive monitoring of treatment response in the adjuvant setting (i.e. ctDNA monitoring)? Would this enhance the identification of patients who respond to treatment vs. patients who never achieved a benefit?

References:

1. Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. Science (80-.). 359, 1350–1355 (2018).
2. Hargadon, K. M., Johnson, C. E. & Williams, C. J. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. Int. Immunopharmacol. 62, 29–39 (2018).
3. Zimmermann, S., Peters, S., Owinokoko, T. & Gadgeel, S. M. Immune Checkpoint Inhibitors in the Management of Lung Cancer. Am. Soc. Clin. Oncol. Educ. B. 682–695 (2018). doi:10.1200/edbk_201319
4. O'Donnell, J. S., Smyth, M. J. & Teng, M. W. L. Acquired resistance to anti-PD1 therapy: Checkmate to checkpoint blockade? Genome Med. 8, 1–3 (2016).
5. Ott, P. A., Hodi, F. S., Kaufman, H. L., Wigginton, J. M. & Wolchok, J. D. Combination immunotherapy: A road map. J. Immunother. Cancer 5, 1–15 (2017).
6. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: Toward combination strategies with curative potential. Cell 161, 205–214 (2015).
7. Schmidt, E. V. Developing combination strategies using PD-1 checkpoint inhibitors to treat cancer. Semin. Immunopathol. 41, 21–30 (2019).
8. Jenkins, R. W., Barbie, D. A. & Flaherty, K. T. Mechanisms of resistance to immune checkpoint inhibitors. Br. J. Cancer 118, 9–16 (2018).
9. Opportunities for Novel-Novel Therapy Combination Development: Data Sources and Strategies to Assess Contribution of Components. (2019).
10. Hodi, F. S. et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with Pembrolizumab. J. Clin. Oncol. 34, 1510–1517 (2016).
11. Chiou, V. L. & Burotto, M. Pseudoprogression and Immune-Related Response in Solid Tumors. JCO 33, (2015).
12. Nishino, M. Immune-related response evaluations during immune-checkpoint inhibitor therapy: Establishing a 'common language' for the new arena of cancer treatment. J. Immunother. Cancer 4, 4–7 (2016).
13. Nishino, M. et al. Tumor response dynamics of advanced non–small cell lung cancer patients treated with PD-1 inhibitors: Imaging markers for treatment outcome. Clin. Cancer Res. 23, 5737–5744 (2017).
14. Nishino, M. et al. Immune-related tumor response dynamics in melanoma patients treated with pembrolizumab: Identifying markers for clinical outcome and treatment decisions. Clin. Cancer Res. 23, 4671–4679 (2017).
15. Wolchok, J. D. et al. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. Clin. Cancer Res. 15, 7412–7420 (2009).
16. Ades, F. & Yamaguchi, N. WHO, RECIST, and immune-related response criteria: Is it time to revisit pembrolizumab results? Ecancermedicalscience 9, 9–12 (2015).
17. Nishino, M. et al. Developing a common language for tumor response to immunotherapy: Immune-related response criteria using unidimensional measurements. Clin. Cancer Res. 19, 3936–3943 (2013).
18. Seymour, L. et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 18, 143–152 (2017).
19. Scher, H. I. et al. Trial design and objectives for castration-resistant prostate cancer: Updated recommendations from the prostate cancer clinical trials working group 3. J. Clin. Oncol. 34, 1402–1418

(2016).

20. Ellingson, B. M., Wen, P. Y. & Cloughesy, T. F. Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. Neurotherapeutics 14, 307–320 (2017).

21. The I-SPY Trials. ispytrials.org.

22. US Food and Drug Administration. Guidance for Industry on Codevelopment of Two or More New Investigational Drugs for Use in Combination. Federal Register 35940–35942 (2013).

23. Lin, J. J. et al. Brigatinib in Patients With Alectinib-Refractory ALK-Positive NSCLC. J. Thorac. Oncol. 13, 1530–1538 (2018).

24. Kim, D. W. et al. Brigatinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small-cell lung cancer: A randomized, multicenter phase II trial. J. Clin. Oncol. 35, 2490–2498 (2017).

25. Kang, S. et al. Pembrolizumab KEYNOTE-001: An adaptive study leading to accelerated approval for two indications and a companion diagnostic. Ann. Oncol. 28, 1388–1398 (2017).

26. Robert, C. et al. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: A randomised dose-comparison cohort of a phase 1 trial. Lancet 384, 1109–1117 (2014).

27. Ribas, A. et al. Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): A randomised, controlled, phase 2 trial. Lancet Oncol. 16, 908–918 (2015).

28. Larkin, J. et al. Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: A Randomized, Controlled, Open-Label Phase III Trial. J. Clin. Oncol. 36, 383–390 (2018).

29. Hamid, O. et al. Final analysis of a randomised trial comparing pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory advanced melanoma. Eur. J. Cancer 86, 37–45 (2017).

30. Phase II Umbrella Study of Novel Anti-cancer Agents in Patients With NSCLC Who Progressed on an Anti-PD-1/PD-L1 Containing Therapy. (HUDSON).

31. Study of Efficacy and Safety of Novel Spartalizumab Combinations in Patients With Previously Treated Unresectable or Metastatic Melanoma (PLATforM).

32. Davi, R. et al. Augmenting Randomized Confirmatory Trials for Breakthrough Therapies with Historical Clinical Trials Data. (2018).

FRIENDS
of CANCER
RESEARCH

A F R I E N D S O F C A N C E R R E S E A R C H W H I T E P A P E R

CHARACTERIZING THE USE OF EXTERNAL CONTROLS FOR AUGMENTING RANDOMIZED CONTROL ARMS AND CONFIRMING BENEFIT

OBJECTIVE

Friends of Cancer Research (*Friends*) convened a working group to characterize methodological processes and to discuss the implementation and opportunities for formal regulatory use of external controls. This whitepaper describes several approaches to constructing an external control and also considers the use of hybrid designs that supplement or augment the control group in the randomized control trials (RCT) with data from an external population. This whitepaper further discusses statistical methodology to help address potential biases and improve the usefulness of the data as well as other adjustment methods that rely on patient summary data. In addition, we describe several scenarios where the use of external controls may be advantageous and practices that can help guide the implementation within a clinical study. A use case was prepared that characterizes the construction of an external control using clinical trial data in multiple myeloma to compare the treatment effect with a randomized control versus an external control and assesses the potential impact of unmeasured confounders.

INTRODUCTION

In drug development, RCT are the gold standard for evaluating the safety and efficacy of medical treatments. However, oncology drug development increasingly relies on the use of single-arm clinical trials especially in certain settings where there are ethical or feasibility challenges with deploying a concurrent control arm. While single-arm trials alone may yield important safety and efficacy signals and can be relied on for regulatory decision making in certain clinical and regulatory contexts, external controls (sometimes referred to as synthetic controls) may provide additional context and supplementary evidence. Expanding the use of external controls to other difficult-to-study indications may reduce patient burden where research may be slowed or unin-

ABOUT FRIENDS OF CANCER RESEARCH

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

WORKING GROUP

Francois Beckers
EMD Serono

William Capra
Genentech

Adrian Cassidy
Novartis

Frank Cihon
Bayer U.S.

Ruthie Davi
Acorn AI, A Medidata Company

Ritesh Jain
EMD Serono

Alaknanda Joshi
Novartis

Bindu Kanapuru
FDA

Laura Koontz
Flatiron Health

Dominic Labriola
Bristol-Myers Squibb

Michael LeBlanc
University of Washington

Nicole Mahoney
Flatiron Health

Michael Menefee
FDA

Pallavi Mishra-Kalyani
FDA

Reena Nadpara
Novartis

Jim Omel
Cancer Research Advocate

Erik Pulkstenis
AbbVie

Jeremy Rassen
Aetion

Dirk Reitsma
PPD

Gary Rosner
Johns Hopkins University

Meghna Samant
Flatiron Health

Chenguang Wang
Johns Hopkins University

Xiang Yin
Acorn AI, A Medidata Company

We are grateful for the data, expertise, and/or review each working group member has provided.

terpretable due to the use of a concurrent randomized control. The latter may be the case with some confirmatory trials of medical products made available through the accelerated approval pathway where the control arm may be compromised by early discontinuation or treatment cross-over to the investigational therapy made available by an accelerated approval.[1]

Study designs that deviate from the traditional RCT, such as single arm or externally controlled trials, are considered in guidance for regulatory approval when their use is justified.[2,3] These types of trial designs can be warranted for scenarios where randomization may be difficult or infeasible due to the rarity of the disease, scarcity of patients, scientific concerns about treatment switching/crossover, or ethical considerations. For example, challenges introduced by treatment crossover can be observed in the double-blind, randomized study comparing sunitinib to placebo. An inter-im analysis demonstrated a large effect on progression free survival (PFS) and patients on the placebo arm were offered sunitinib. During the final analysis, the treatment effect size for overall survival (OS) was diminished, which was likely due to treatment crossover.[4]

As described in regulation, external controls have generally been allowed only in, "special circum-stances;" for example, "diseases with high and predictable mortality" and when, "effect of the drug is self-evident." This restricted use is due, in part, to the perceived inability for external con-trols to be "well assessed with respect to pertinent variables as can concurrent control popula-tions," as stated in FDA guidance and regulation.[2] However, our ability to electronically store and manage continually aggregating real-world data (RWD) from electronic medical records, claims data, prior clinical trials data, and other sources is opening opportunities that were not possible before. Moreover, higher quality external controls are more available today than in the past due to the availability of patient level data and statistical methods for achieving balance in baseline characteristics between the clinical trial and external controls.

There are several examples of the use of external controls for regulatory applications evaluating effectiveness, but most have been used for informal, rather than direct, statistical comparison. The use of external controls is most common in orphan disease settings where it can be difficult to accrue patients, especially for a randomized clinical trial. There are some notable examples of the use of external controls in oncology drug development:

1. Blinatumomab (Blincyto)[5,6]: Historical clinical trial site patient data and propensity score meth-ods were used to construct complete remission and OS reference rates for comparison to the single-arm study of blinatumomab for Ph-negative B-precursor cell relapsed/refractory acute lymphoblastic leukemia. As per the sponsor, the historical clinical trial data of 1,139 patients from the EU and the US was used to support the FDA's breakthrough therapy designation and accelerated approval in December 2014.

2. Bavencio (Avelumab)[7]: In 2017, Bavencio received accelerated approval for Merkel Cell Carcinoma on the basis of an 88-patient single arm Phase II trial. Real-world evidence (RWE),

contributed by external data from a registry, was used as supportive evidence, but the regulatory approval was based primarily on data from the Phase II trial.

Additional efforts and case studies have helped inform methodology for constructing external controls and describe limitations and opportunities with these types of analyses. For instance, a case study in non-small cell lung cancer demonstrated that it is possible to produce a "matched" cohort to a randomized control arm.[1] More experience and understanding of the circumstances where external data may serve as an external control are needed to characterize the full utility and potential of external controls. This paper explores the design and analyses of studies leveraging an external control built from external historical or contemporaneous patient-level data selected to be similar in important prognostic (or clinical) characteristics to patients treated with the experimental regimen.

METHODOLOGICAL APPROACHES AND CONSIDERATIONS FOR CONSTRUCTING AN EXTERNAL CONTROL

Multiple sources of data exist to populate an external control cohort. These data sources include clinical trial data, published clinical data, and real-world data derived from electronic health records (EHRs) and other sources. As external data sources are considered, the advantages and limitations associated with the various sources and whether patient-level data is available will need to be evaluated when designing a clinical study. Methods discussed in this whitepaper focus on the use of individual patient-level data rather than aggregate-level data.

COHORT SELECTION AND ADJUSTMENT METHODS

Careful cohort selection is critical to developing a robust external control to control for potential biases that can be encountered in clinical research (Table 1). Lack of randomization can result in several potential biases. In particular, selection bias and confounding bias need to be considered when selecting patients in the external control cohort. Selection bias occurs when the observed patients are not representative of the broader population of interest and thereby can challenge the external validity of the results.

Some examples include selecting patients from a specific geographic region or with certain clinical characteristics such as age, comorbidities, prognostic indices or prior/concurrent therapies that are not representative of the clinical trial population. It is also important to select a cohort in which we can account for confounding that may arise due to lack of randomization. Confounding bias occurs when there is an imbalance in the distribution of key baseline characteristics that are associated with both the outcome and exposure to treatment. Such characteristics are called confounders and are typically characterized as "measured" and "unmeasured." The presence of confounders is particularly important to consider when using controls from RWD sources, like electronic medical record data, since certain patient characteristics that are likely to impact outcomes

Table 1: Select biases encountered in clinical research

Bias	Explanation	Methods to Reduce Bias
Confounding Bias	Selection of experimental and control patients completed in such a way that the patient characteristics are systematically different across treatment groups, perhaps with those with better prognoses preferentially receiving one therapy over another.	Randomization
Selection Bias	Occurs when the observed patients are not representative of the broader population of interest and thereby can challenge the external validity of the results.	Randomization; Improved sampling
Performance Bias	Follow-up differs by treatment. Differential care according to treatment beyond the treatment itself. Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	Standardization of treatment and follow-up plans for all patients
Detection Bias	Outcome assessment differs by treatment leading to systematic differences in outcome determination.	Masking
Attrition Bias	Systematic differences between groups in withdrawals from a study or treatment exist.	Analysis by intention to treat
Time-trend Bias	Prognostic characteristics of available patients change during the course of the trial especially for trials with long recruitment periods.	Maintain randomization

(e.g., age, access to clinical care, socioeconomic status) are also likely to influence treatment exposure. For example, we may see a distribution of patients in the clinical trial skewed more toward younger and fitter patients, while the population of real-world patients may comprise a much broader patient population including the elderly and patients with more comorbidities. Even within the real-world population, confounding by indication may occur as certain types of patients may be more likely to be prescribed certain treatments because of their characteristics. Finally, differences in the characteristics of sites participating in clinical trials (e.g., site effect, site volume, clinical care protocols, access to multimodality care, academic vs. community centers, etc.) can also confound outcomes and bias results. The myriad of considerations discussed above make it challenging to isolate the treatment effect in externally controlled studies, and analytical approaches need to be considered to mitigate such biases, where possible.

A fundamental step when considering external controls is thoughtful and rigorous planning in the design phase. This involves careful identification of key baseline prognostics and confounding factors through tools such as directed acyclic graphs (DAGs), and accordingly pre-specifying the key inclusion/exclusion criteria for external cohort selection.[8,9] The identification and prioritization of key criteria for selection is critical because not all criteria typically applied in clinical trials may be available or possible to collect using completed historical trials or retrospective real-world datasets. It is thus important to align, as much as possible, the criteria between the clinical trial and the external control. Once a prioritized list of criteria is identified, all efforts must be made to collect the relevant data to a high degree of completeness and accuracy, noting that in some cases prospective approaches may be needed to intentionally collect the required data element. Sponsors should clearly and transparently document these efforts (e.g., through patient attrition diagrams, data deficiencies), hypothesize the impact of missing data elements on overall outcomes, and have plans to address this impact.

Despite careful selection of the external cohort in alignment with the trial eligibility criteria, imbalances in key confounding factors may still exist that need to be further mitigated through thoughtful consideration and pre-specification of appropriate statistical methodologies. There is no one adjustment method universally preferred over others; Table 2 below outlines methods that are commonly used to drive greater balance in patient distributions among measured covariates. This is not intended to be a comprehensive list and variations of these methods are common. The choice of the statistical method in a particular context ultimately depends on a variety of factors including available external cohort size, number of key variables to consider, tolerance for complexity, etc. Propensity score methods are especially important in the creation of external controls and are further discussed in the next section.[10,11] A propensity score (PS) is the probability of being treated with one drug versus another, based on the measured factors known about the patient. In a single number, the PS captures much of the nuance about treatment choice and allows us to control for a substantial amount of confounding using a single variable (a detailed description of propensity scores, propensity score matching, and propensity score weighting is in Appendix 1).

For hybrid designs (randomized controls augmented with external controls), some statistical methods determine the degree to which external information enters the analysis of a clinical trial in a data-dependent way. If the external data, particularly outcomes or covariate-adjusted outcomes, seem consistent with the outcomes of the current trial's control group, the algorithms will give relatively more weight to the external data than when there appear to be heterogeneities. Some example methods are commensurate priors, power priors, and meta-analytic predictive priors and are commonly used in trial designs with hybrid controls.[12–16]

Table 2. Commonly used statistical methods to balance baseline factors[8]

Method	Description	Key Benefits	Key Limitations
Exact matching	Trial patients are matched 1:1 or 1:many to external controls on a set of important baseline characteristics	- Simple and intuitive	- Need large external cohort sample size to find matched controls for all patients and some trial patients may remain unmatched - Often very limited number of baseline factors can be used for matching - Inefficient use of data from unmatched trial and control patients

Propensity score matching[17]	Trial patients are matched with fixed or various ratios to external controls on propensity scores (probability a patient is in the trial cohort vs external control conditional on baseline covariates). Since scores are continuous, calipers/intervals are commonly used	- Can be simple and intuitive - Large number of baseline factors can be captured and balanced through propensity score - Matching is based on one single score rather than on the full multivariate set of baseline factors - Calipers provide flexibility to relax matching requirements and enable more efficient use of external controls	- Some matching algorithms need a large external cohort sample size to find matched controls for all patients and some trial patients may remain unmatched - Inefficient use of data from unmatched trials and control patients when insufficient number of matches are found - Requires correct specification of the propensity score model - Pre-specifying width of the caliper may be challenging depending on the context and sample size
Propensity score weighting – **Inverse probability of treatment weights (IPTW)**[18]	Propensity scores are typically used to weight patients in the trial and external cohorts in a way that achieves balance in the baseline characteristics	- Efficient use of all trial and external control patients	- Distorts the original distribution of the trial patients since they are also weighted along with the external controls, thereby changing the target population for which treatment efficacy is being assessed - Requires correct specification of the propensity score model - May require more complex analytic decisions, e.g. trimming, in case of extreme propensity scores
Propensity score weighting – **Weighting by odds**[19]	Patients in the trial arm are given a weight of 1 (i.e. all information is included) while odds of propensity scores are used to weight patients in the external cohorts	- Distribution of trial patients remains intact and full information from all trial patients is utilized - Efficient use of all trial and external control patients	- Requires correct specification of the propensity score model

Outcome regression models	Association between treatment and outcome is modeled adjusting for baseline covariates Doubly robust regression models are sometimes considered whereby a function of propensity scores is used as weights in the model, making it more robust to model misspecification[20]	- Generally easy to understand as familiarity with regression models is high among research community - Efficient use of all trial and external control patients - Doubly robust models provide insurance against model misspecification (i.e the results are unbiased so long as either the outcome model or propensity model are correctly specified)	- The outcome model must be correctly specified if no propensity score weighting is used - Either the outcome or propensity score model must be correctly specified if a weighted model is used - No separation of design for balancing baseline factors from the outcome analysis

NOTES ON EFFECT ESTIMATES

Matching and weighting are on their surface very similar, but there is a subtle difference in the values one estimates from each approach. The matching approach will estimate the average treatment effect in the treated, which can be more tangibly thought of as the treatment effect among those patients who were reasonable candidates for either treatment choice: this is a notion of clinical equipoise. On the other hand, IPTW weighting estimates the average treatment effect in the entire population and considers what would happen if all patients were moved from control to treatment.[21]

For the questions considered here, we would expect there to be little difference between the average treatment effect in the treated and the average treatment effect in the entire population, as all patients would have met stringent inclusion/exclusion criteria, and thus would in all likelihood be eligible for either treatment pathway. As such, considerations of the feasibility of matching should outweigh considerations of the estimated treatment effect.

CONTROL OF CONFOUNDING BIAS

Confounding bias results from not accounting for factors that are associated with both the treatment choice and the outcome, independent of any effect via treatment.[22] In studies of medications, some of the strongest confounding comes from confounding by indication, in which patients' level of illness drives treatment choice (sicker patients may get "stronger" treatments) as well as outcome (sicker patients may experience worse outcomes).[23] This can be particularly difficult to address, though design approaches such as fit-for-purpose data, RCT-like study design,[24] new user cohorts,[25] and principled process,[26] as well as analytic approaches, such as multivariable regression, propensity scores, and high-dimensional propensity scores,[27]

can eliminate the effect of measured (or measureable) confounding.

However, unmeasured confounding may yet remain, and control of a factor that is ultimately unmeasurable is a substantial challenge. Approaches that can control for this unmeasurable confounding, such as instrumental variable analysis, are not frequently seen in the medical literature but can be effective.[28] Separately, high dimensional propensity scores can "uncover" previously-unmeasured confounders and reduce bias. There are powerful techniques that allow us to assess unmeasured confounding.

Causal diagrams can help elucidate potential sources of bias.[8] More quantitatively, sensitivity analyses allow us to ask ourselves questions like, "If we had an unmeasured confounder (or group of confounders) of strength x, how much would our results be affected?" and "How powerful would an unmeasured confounder (or group of confounders) have to be to meaningfully alter our interpretation of the situation we've observed?"[29] E-values and tipping point analyses may be potential solutions for assessing the impact of unmeasured confounders on the overall treatment effect. The use case included in Appendix 2 illustrates a tipping point analysis, which shows the strength a confounder would need in order to change the statistical significance or numerical direction of the original estimates of the treatment effect. By addressing these questions, we can better reason about the robustness of our results to issues like unmeasured confounding; presenting such results can strengthen readers' and reviewers' confidence in the evidence.

While the potential for unmeasured confounding is a key issue in any non-randomized study, in the single-arm study with external controls scenario, a more important issue is whether the experience of the controls truly represents the counterfactual experience of the treated patients. That is, would standard of care patients have been treated with the single-arm treatment had the single-arm treatment been available to them, and vice-versa? To ensure this, we implement strong inclusion/exclusion criteria, draw controls from populations similar to that of treated patients, and apply other key design approaches.

OUTCOMES AND ENDPOINT CONSIDERATIONS

Even when a set of patients comparable to the experimentally treated patients can be identified for the external control, to create valid inference regarding the treatment effect, one must also ensure comparable ascertainment and measurement of the outcomes of interest for the external control and experimentally treated patients. Differences between arms in endpoint collection methods and endpoint definitions can bias the treatment effect estimates. But the way the endpoints are captured for the external control patients generally is not within the researcher's control and may not be completely consistent with the experimentally treated patients. Additionally, assessment of response or progression free survival endpoints may be performed

locally or centrally, and those assessment differences should be a consideration when external control data are utilized. In addition, some response criteria, especially in hematologic malignancies, are complex, which may result in differences in implementation from study to study.

These inconsistencies may be more or less challenging based on the source of data. For example, external controls built from historical clinical trial data enjoy the benefit of similar collection and definition of efficacy and safety endpoints while endpoints representing similar clinical concepts may be captured differently in external controls built from real-world data. Endpoints that are objective may be less affected by different measurement techniques, timing, or settings and may be preferred when using an external control. For instance, progression free survival may have more complex considerations than an endpoint related to tumor shrinkage when considering options for external controls.[30,31]

OPERATIONAL CONSIDERATIONS

Situations that may support the use of an external control include those where randomization may not be feasible due to ethical, scientific, or operational considerations (Table 3).[32] For example, for certain orphan diseases, rare diseases, or rare biomarker-defined cohorts, it may not be possible to enroll a sufficient number of patients to have a concurrent control, meriting consideration of external data sources. Hybrid designs could also be considered to reduce the number of patients assigned to the control by augmenting with external data. In some cases, it may be unethical to randomize patients to the control arm. All patients could receive a promising experimental drug in an externally controlled trial, making this type of study more attractive to patients, and lessening the risk of trials closing due to poor accrual. Externally controlled studies may also be valuable when treatment crossover from a deployed control arm to the experimental arm of an RCT, or to off-study treatments including new treatments approved during the course of the study, compromises the interpretability of treatment effects. In some respects, externally controlled data may be preferable to single-arm studies that are often employed to address the limitations noted in the situations above and given that some time to event endpoints may be difficult to interpret in a single-arm study.

Table 3. Select scenarios that may benefit from the use of an external control

Scenario	Challenge	Role of External Controls
Uncontrolled studies (e.g., single-arm trial, expanded access)	Outcomes of the experimentally treated patients are difficult to interpret without an understanding of expected outcomes for patients who did not receive the experimental treatment	-To provide context needed to interpret outcomes of experimentally treated patients by comparing to a group of patients who did not receive experimental treatment
Studies of orphan diseases, rare diseases or rare biomarker-defined cohorts	Recruitment of patients is very difficult due to rarity of defined disease so that a concurrent control may not be possible and resulting single arm data is difficult to interpret	-To improve patient recruitment and allow a design where all patients can be treated with the experimental product -To provide context needed to interpret outcomes of experimentally treated patients by comparing to a group of patients who did not receive experimental treatment -To function as a natural history cohort to describe patient characteristics and outcomes in these settings
Post-marketing confirmatory study following accelerated approval	Recruitment and/or retention to a randomized controlled trial when the experimental product is available on the market is very difficult and sometimes impossible	-To augment or replace the randomized control of the confirmatory trial so that an external control may be constructed and confirmatory studies could be completed. An additional benefit would be that patients enrolling in the trial have a higher probability or even assurance of receiving the experimental therapy
High rate of treatment cross-over	Patients assigned to the control arm of a randomized controlled trial may use the experimental product or a similar product in the same class when the experimental product or a similar product in the same class is available on the market thereby diluting the ability of the study to demonstrate a difference between arms	-To augment or replace a randomized control with patients who did not receive the experimental product (since perhaps they were studied at a time when the experimental product was not available) so that the difference between arms is a more accurate estimate of the actual treatment effect

To date, from a regulatory perspective, external controls have been used to provide a benchmark or context for interpreting single arm effectiveness studies. With careful planning and scientifically rigorous approaches, external controls may be compared through formal statistical methods and support regulatory decisions. The clinical questions and regulatory decisions sought should drive the selection of data source, study design, and analytic approaches.

Bias inherent in externally controlled studies may be difficult to account for, but certain approaches may increase the credibility of such studies and reduce concerns (see above). Once a decision has been made to use an external comparator, there are specific considerations that may strengthen or limit the credibility of resulting data. These considerations, described in current FDA guidance, include ensuring similarity between the external populations and those receiving the experimental drug with respect to critical baseline characteristics such as disease severity, duration of illness, prior treatments, and other critical prognostic factors.

Another important consideration is the comparability of endpoint assessments regarding both definitions and ascertainment (timing, measurement). Historical clinical trial data may have more applicable data than data derived from EHRs or registries, which may not collect the sorts of endpoints used in clinical trials or collect them at consistent time points. For example, an endpoint like overall survival is less likely to suffer from ascertainment bias than is expected from more complex endpoints like response rate or progression free survival, which may differ in definition and ascertainment as well as analytical approaches across different datasets or physician assessments.

Patient management also matters, especially for cancer types in which the standard of care is not agreed upon or has rapidly changed over time. For example, as toxicity management improves over time, this may in turn impact patient outcomes. It would be beneficial if management of patients from historical data sources was similar enough to the current clinical trial to limit any resulting bias. This may be assessed by looking at the constancy of treatment outcomes historically for the control regimen. Consistency would lead to a higher level of confidence that if a randomized control had been deployed in the current trial, then it would have behaved similarly. To the extent that patient management in clinical trials differs from patient management in clinical practice, this also may result in differences between using historical clinical trial data vs. RWD. It should also be noted that the patient population(s) are rapidly changing in many areas. The immunotherapy revolution has dramatically changed many patient populations available for clinical trials relative to data that may be available historically. A complete and transparent assessment of these issues will help researchers and reviewers understand the scientific strength of the evidence of safety and effectiveness resulting from the study.

REGULATORY CONSIDERATIONS

FDA regulations explicitly recognize the use of external controls, including a hybrid approach where a clinical trial control group is augmented with external data, to support regulatory decision-making in limited circumstances.[2,3] While the use of external control data matures to the point where it may support regulatory approval more broadly, careful consideration should be given to near-term uses in appropriate regulatory and clinical contexts. Rather than replacing RCTs in situations where randomization is feasible, new methodological approaches and data sources may allow the use of external comparators, in situations where randomization would be unethical or infeasible. For example, external patient level data may be used to augment randomized control arms as part of a hybrid approach that could reduce the number of patients that are randomized to the control arm within a study. Such data may come from completed RCTs or from real-world sources such as electronic medical records.

Given a solid rationale for an external control, and a careful assessment of whether an external control would be scientifically feasible based on the considerations just outlined, the actual implementation of the external control requires care and planning. Several procedural best practices are advised as part of the regulatory process to increase the credibility of externally controlled studies. Pre-specification of protocols and statistical analysis plans provide confidence that the external control group selection process follows a prospective methodology and plan that could be independently performed or duplicated. This should include a detailed protocol with clear objectives and description of the study population, as well as details regarding data sources and critical features of the study design and analysis plan. The approach should be specified in the statistical analysis plan or other companion document and should not be biased by actual analysis of candidate external control group data that may be perceived to introduce selection bias. This may happen for example if historical data/trials with superior results are preferentially omitted. As a result, it is important that the entire selection process of a dataset and patient-level data be prespecified independent of outcome data.[33,34] The final statistical analysis and any sensitivity analyses should also be clearly pre-specified consistent with good statistical practice.

Early discussions with regulators and review of key planning documents is likely to result in valuable feedback for sponsors using external controls. Sponsors should consider soliciting FDA feedback by means of protocol submissions or formal product meetings. Sponsors may also explore opportunities for participation in the Agency's Complex Innovative Trial Designs pilot program, which exists to further the use of new trial designs.[35] When external data comes from real-world data sources, sponsors may request input on study designs from the FDA's RWE Subcommittee and should note any submission of RWD to the agency for tracking purposes.

CONCLUSIONS AND RECOMMENDATIONS

In oncology, there are clinical settings and scenarios where randomization may be difficult or not feasible (e.g., rare disease, small patient population, loss of equipoise, availability of the investigational agent outside of the clinical trial). Additionally, patients with serious, life threatening diseases may often seek trials where the likelihood of receiving the investigational agent is high (e.g., single arm studies, designs that allow treatment crossover). However, these scenarios (described in Table 3) may make interpreting the clinical trial results difficult or could introduce uncertainty in the results. The use of external controls in clinical studies represents an opportunity to potentially reduce the number of patients in the control arm, enhance data obtained from clinical trials, and improve the interpretability of results.

This whitepaper describes methodological approaches for constructing an external control cohort and reducing or managing potential biases that can be introduced in these types of analyses as well as operational and regulatory considerations to help guide their successful use. The case study developed for this whitepaper (Appendix 2) also helps demonstrate how to operationalize several of the concepts described in this whitepaper and inform the design of future clinical studies.

Additional considerations may also need to be explored to further facilitate the use of external control cohorts more formally in oncology drug development and regulatory discussions:

- Identify methods and mechanisms to share patient-level data to facilitate robust analyses
- Clarify how sponsors and investigators can incorporate external controls for formal analyses to support regulatory decisions
- Establish best practices for the use of specific data sources and appropriate methodologies to help develop and promote standards
- Characterize appropriate uses of specific endpoints in external controls and the ability to compare across studies

Glossary

Control Arm – In a clinical trial, the group of participants that is not given the experimental intervention being studied is the control arm. A control arm is used to establish the expected outcome without the effect of the new experimental therapy, and the result in the experimentally treated patients is judged relative to this. The control arm may receive an intervention that is considered effective (the standard), a placebo, or no intervention.

Randomized Control Arm – In a randomized controlled clinical trial, the group of participants who are randomly selected to not receive the experimental intervention is a randomized control arm. Random selection of patients and concurrent study of the randomized control arm with the study of the experimental intervention group provides high levels of assurance that differences between the randomized control arm and the experimental intervention arm are attributable to the intervention, not imbalances in baseline characteristics or differences in time, place, or circumstances of treatment.

External Control Arm – An umbrella term referring to any control that is not a randomized control. Can be used as a reference for interpretation of a set of experimental data especially when randomization is unethical or unfeasible.

Concurrent Control Arm – A type of external control. A group chosen from the same or similar population as the experimental intervention group and treated over the same period of time as the experimentally treated patients. Ideally, the experimental intervention and control groups should be similar with regard to all baseline and on-treatment variables that could influence the outcome, except for the study treatment. May be patient-level data or summary information gained from medical literature or other sources.

Historical Control – A type of external control. A non-concurrent comparator group of patients who received treatment (placebo or active treatments) in the past or for whom data are available through records. May be patient-level data or summary information gained from medical literature or other sources.

Synthetic Control Arm – A type of external control consisting of patient level data from patients external to the trial and selected with statistical methods such as propensity scores to provide confidence that the baseline characteristics of the selected external patients are balanced and comparable with the baseline characteristics of the experimentally treated patients. Can be formed from external clinical trials data, real-world data, or other data sources.

References

1. Friends of Cancer Research. Exploring Whether a Synthetic Control Arm can be Deriving from Historical Clinical Trials that Match Baseline Characteristics and Overall Survival Outcome of a Randomized Control Arm. (2018). Available at: https://www.focr.org/sites/default/files/SCA White Paper.pdf.
2. 21 CFR 314.126.
3. FDA. Guidance for Industry: E10 Choice of Control Group and Related Issues in Clinical Trials. (2001). Available at: https://www.fda.gov/media/71349/download.
4. Faivre, S. et al. Sunitinib in pancreatic neuroendocrine tumors: updated progression-free survival and final overall survival from a phase III randomized study. Ann. Oncol. 28, 339–343 (2016).
5. Barlev, A. et al. Estimating Long-Term Survival of Adults with Philadelphia Chromosome-Negative Relapsed/Refractory B-Precursor Acute Lymphoblastic Leukemia Treated with Blinatumomab Using Historical Data. Adv. Ther. 34, 148–155 (2017).
6. Gökbuget, N. et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood Cancer J. 6, e473–e473 (2016).
7. Cowey, C. L. et al. Real-world treatment outcomes in patients with metastatic Merkel cell carcinoma treated with chemotherapy in the USA. Futur. Oncol. 13, 1699–1710 (2017).
8. Greenland, S., Pearl, J. & Robins, J. M. Causal Diagrams for Epidemiologic Research. Epidemiology 10, (1999).
9. Miguel A. Hernán, J. M. R. Causal Inference: What If. (Chapman & Hall/CRC, 2019).
10. Ho, D. E., Imai, K., King, G. & Stuart, E. A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Polit. Anal. 15, 199–236 (2007).
11. Stuart, E. A. Matching methods for causal inference: A review and a look forward. Stat. Sci. 25, 1–21 (2010).
12. Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. & Sargent, D. J. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. Biometrics 67, 1047–1056 (2011).
13. Murray, T. A., Hobbs, B. P. & Carlin, B. P. Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. Ann. Appl. Stat. 9, 1549–1570 (2015).
14. Ibrahim, J. G., Chen, M.-H., Gwon, Y. & Chen, F. The power prior: theory and applications. Stat. Med. 34, 3724–3749 (2015).
15. Schmidli, H. et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics 70, 1023–1032 (2014).
16. Viele, K. et al. Use of historical control data for assessing treatment effects in clinical trials. Pharm. Stat. 13, 41–54 (2014).
17. Rubin, D. B. & Thomas, N. Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics 52, 249–264 (1996).
18. Lunceford, J. K. & Davidian, M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat. Med. 23, 2937–2960 (2004).

19. Li, F., Morgan, K. L. & Zaslavsky, A. M. Balancing Covariates via Propensity Score Weighting. J. Am. Stat. Assoc. 113, 390–400 (2018).

20. Bang, H. & Robins, J. M. Doubly Robust Estimation in Missing Data and Causal Inference Models. Biometrics 61, 962–973 (2005).

21. Austin, P. C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav. Res. 46, 399–424 (2011).

22. K Rothman, S Greenland, T. L. Modern Epidemiology, 3rd Edition. (RTI International, 2007).

23. Walker, A. M. Confounding by Indication. Epidemiology 7, (1996).

24. Hernán, M. A. & Robins, J. M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am. J. Epidemiol. 183, 758–764 (2016).

25. Ray, W. A. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. Am. J. Epidemiol. 158, 915–920 (2003).

26. Schneeweiss, S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol. Drug Saf. 19, 858–868 (2010).

27. Schneeweiss, S. et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 20, 512–522 (2009).

28. Rassen, J. A., Brookhart, M. A., Glynn, R. J., Mittleman, M. A. & Schneeweiss, S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. J. Clin. Epidemiol. 62, 1226–1232 (2009).

29. Schneeweiss, S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol. Drug Saf. 15, 291–303 (2006).

30. Seymour, L. et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. Clin. Cancer Res. 16, 1764–1769 (2010).

31. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2018 Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics. (Accessed: 1st November 2019)

32. FDA. FRAMEWORK FOR FDA'S REAL-WORLD EVIDENCE PROGRAM. (2018). Available at: https://www.fda.gov/media/120060/download.

33. FDA. Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products. (2018). Available at: https://www.fda.gov/media/117976/download.

34. FDA. Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics Guidance for Industry. (2019). Available at: https://www.fda.gov/media/124795/download.

35. FDA. Complex Innovative Trial Designs Pilot Program. (2019). Available at: https://www.fda.gov/drugs/development-resources/complex-innovative-trial-designs-pilot-program.

36. ROSENBAUM, P. R. & RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55 (1983).

37. Stürmer, T., Wyss, R., Glynn, R. J. & Brookhart, M. A. Propensity scores for confounder

adjustment when assessing the effects of medical interventions using nonexperimental study designs. J. Intern. Med. 275, 570–580 (2014).

38. Cepeda, M. S., Boston, R., Farrar, J. T. & Strom, B. L. Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders. Am. J. Epidemiol. 158, 280–287 (2003).

39. Seeger, J. D., Bykov, K., Bartels, D. B., Huybrechts, K. & Schneeweiss, S. Propensity Score Weighting Compared to Matching in a Study of Dabigatran and Warfarin. Drug Saf. 40, 169–181 (2017).

40. Stürmer, T., Rothman, K. J., Avorn, J. & Glynn, R. J. Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution—A Simulation Study. Am. J. Epidemiol. 172, 843–854 (2010).

41. Hernán, M. Á., Brumback, B. & Robins, J. M. Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 11, (2000).

Appendix 1: Detailed Description of Propensity Scores, Propensity Score Matching, and Propensity Score Weighting

Propensity Scores

The propensity score is a method developed in the early 1980's, and further developed substantially over the past decades, to reduce bias due to confounding in observational (non-randomized) studies.[21,36,37] A more novel application of propensity scores is to create balance between a clinical trial treatment arm and an external control group (see Table 2).[6] While it does not control for unmeasured confounding, there are several advantages:
- A propensity score makes it possible to create balance across many factors simultaneously, avoiding issues of cutting data "too thin" when exact matching on many factors.
- In scenarios where patient n is limited but confounding is strong, the single propensity score value allows us to capture a large amount of confounding using substantially fewer degrees of freedom in an outcome model.[38]
- The propensity score can be effectively used in a variety of ways, including matching, weighting, or regression.

Propensity Score Matching

The most common use of the propensity score is in matching. The idea is straightforward: if we are able to estimate the probability of a patient being treated with the investigational treatment, as compared to the standard of care (SOC) measured in external controls, then if we match patients who had similar probabilities of being treated with the investigational treatment and treated with the SOC, then the choice between investigational treatment and SOC for that patient can be thought of as essentially random. That is, if we can (1) estimate a propensity score using all relevant confounders, and then (2) take each patient treated with the investigational treatment and find a similar patient treated with the SOC, we will (3) create a cohort of patients in which each confounder tends to be balanced between the investigational and SOC treatment groups, with no need to further adjust for confounding. As a result, the data can be analyzed like that of an RCT, even though we created the balance by construction rather than design.

The advantages of matching are substantial, and include:
- Clear methodology that is easily understood by readers, reviewers, and others.
- A table of baseline characteristics that can be verified for balance, building confidence in results.
- Simple analytics that do not require, for example, bootstrapped variances or other statistical nuances.

However, matching can also introduce a challenge: if we seek to match all patients treated

with the investigational treatment but fail to find a match among those treated with the SOC, then we will lose one or more patients that received the investigational treatment in the analysis. In cases where there is a substantial n, this is often manageable, but in small trials where each patient's clinical experience is of extraordinary value, losing patients is highly unfavorable.

Propensity Score Weighting

An alternative to matching in which no data are lost is propensity score weighting. Weighting is a propensity score-based approach to standardization; while there are a wide variety of weighting techniques that can be used, the one most commonly seen (and the one used in the Blincyto example) is inverse probability of treatment (IPTW) weighting. Another technique, weighting by propensity odds, is discussed in Table 2.

As a propensity score estimates a patient's probability of receiving a given treatment, the inverse probability of treatment weight is the inverse of the propensity score (that is, 1/PS) for patients receiving the investigational treatment and 1/(1-PS) for SOC patients. (We use 1-PS because that is the probability of being treated with the SOC.) When using IPTW weights, we model a population in which both treated patients and control patients are "standardized" to resemble the entire study population, such that treated patients may be standardized to more resemble controls and vice-versa.

The clear advantage of this technique is that no patient data are lost; we are able to use all data and achieve confounding control. However, there are several disadvantages:
• The method appears somewhat opaque and may not create confidence by readers, reviewers, and other stakeholders.
• Because this method counts certain patients more than others (those with high weights versus those with low weights), it is possible that it may overweight the experience of one or more patients.[39] Control of the maximum assigned weight is often necessary.[40]
• This method will change the weight of both patients receiving the investigational agent and SOC patients, and as such, the data as weighted will not represent patients' actual experience in the single-arm trial.
• Technical adjustments are often needed to stabilize weights and to accurately report variance.[41]

Appendix 2: Developing a Synthetic Control Arm Derived from Historical Multiple Myeloma Clinical Trials and Assessing Unobserved Confounders

1. CASE STUDY OBJECTIVES

This case study builds on previous work (Friends of Cancer Research whitepaper, 2018, case study in non-small cell lung cancer) and continues exploration of whether a synthetic control arm (SCA) can be useful for assessment of medical product efficacy and safety in indications where a randomized control presents ethical or practical challenges. This case study has two primary objectives.

- Objective 1: To explore whether the treatment effect based on a SCA (i.e., investigational arm vs. SCA) can mimic the treatment effect based on the randomized control (i.e., investigational arm vs. randomized control).
- Objective 2: To develop and illustrate statistical methods (e.g., tipping point analyses) useful for assessing the impact of unobserved confounders on the demonstration of efficacy in the setting of a SCA.

This case study will also address some of the concerns regarding incomplete matching in the previous work by utilizing matching methods that do not require exclusion of a large proportion of investigational product (IP) treated patients and by extending SCA exploration to additional indications.

2. DATA SOURCES

This case study is based on patient-level data from multiple historical clinical trials in relapsed/refractory multiple myeloma. These trials have been conducted by the pharmaceutical industry for the purposes of drug development and are available through the Medidata Enterprise Data Store (MEDS). MEDS is a collection of thousands of previous clinical trials with patient-level data recorded through the Medidata electronic data capture system, Rave. Per the legal agreements with the sponsors of these historical clinical trials and Medidata, these data are available for use in deidentified (e.g., patients and original sponsor of the trial cannot be identified) and aggregated (e.g., every analysis must include data from two or more sponsors) form.

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

As shown in Table 1, the historical data originated from open label phase 3 multinational trials that were conducted between 2010 and 2017. At baseline, all patients had:

- Relapsed or refractory multiple myeloma
- Received at least 2 prior lines of treatment
- Received prior treatment with lenalidomide and bortezomib
- Age ≥ 18 years

Including both investigational and control arms from the historical trials, there were 946 historical patients available for this case study.

Table 1: Features of Historical Data							
	Design	Region	Start/End of Trial(s)	Baseline Characteristics	Endpoints	Number of Patients in All Arms	Control Regimen
Historical Data (from multiple trials)	Open label, phase 3	Multi-national	Trial conducted between 2010 and 2017	- Relapsed or refractory multiple myeloma - Received at least 2 prior lines of treatment - Received prior treatment with lenalidomide and bortezomib - Age ≥ 18 years	Overall survival	946	Dexamethasone

Because the historical data in this case study came from trials that had been conducted as part of clinical development programs, the populations, study design, data collection methods, and endpoints utilized in these trials are fairly consistent across trials. Nevertheless, differences across studies in some variable definitions were present but have been reconciled as part of the data standardization process. Clinically important baseline covariates available across studies and to be used in the creation of the SCA are shown in Table 2. Overall survival is the endpoint of interest for this case study and was measured as a key outcome in all historical trials that had similar study designs, such as the disease population and follow-up time.

Table 2: Clinically Important Baseline Covariates Available Across Historical Trials
1. Race (White vs. Others/unknown)
2. Region (Europe vs. Others/unknown)
3. ECOG=0 vs 1 vs 2 or 3
4. Number of drug classes refractory (≥4 vs. <4)
5. Cytogenetic risk (High vs. Standard/unknown)
6. Prior stem cell transplant (Yes vs. No/unknown)
7. Age (continuous)
8. Days since last PD/relapse to first study dose (continuous)
9. Sex (F vs. M)
10. Bone lesion (Yes vs. No/unknown)
11. Best response to last therapy (≥PR vs. <PR/unknown)
12. Number of prior lines of therapy (continuous)
13. Years since diagnosis (continuous)
14. Weight (continuous)

3. RATIONALE AND METHODS

3.1 For objective 1, we explored whether the treatment effect based on a SCA can mimic the treatment effect based on a randomized control using a historical randomized controlled trial in multiple myeloma. This trial, the 'Target Randomized Trial', had a 2:1 treatment assignment ratio and included 294 patients assigned to investigational treatment and 149 patients assigned to dexamethasone as a control. An SCA was selected from the remaining 201 patients assigned to dexamethasone control in all other studies available within this project. Patients assigned to investigational therapies in all trials except the target trial made up the remainder of the total 946 patients referenced above (table 1) and were not utilized in this case study. Historical patients were selected for inclusion in the SCA to balance the baseline characteristics of the IP treated patients in the Target Randomized Trial and the SCA using propensity score methods. Selection of the historical patients for the SCA was completed using only baseline characteristics without knowledge of any post-randomization data.

While appealing in its simplicity and similarity to a randomized design, the commonly used approach to propensity score matching, Greedy 1-1 matching, was not possible for this case study due to the limited number of historical control patients available. Rather, we used a matching method called optimal full matching (often referred to as full matching), which was introduced by Rosenbaum (Rosenbaum 1991) and recommended recently (Hansen 2004, Austin and Stuart 2015a). Full matching subdivides the subjects into strata of different sizes, consisting of either one IP treated subject and at least one control subject or one control subject and at least one IP treated subject. The algorithm of full matching is to minimize the average differences within a matched set in the propensity score between IP treated and control subjects. An attractive feature of this approach is that it can use most or even the entire set of all IP treated subjects in the analysis. This contrasts with conventional matching approaches such as Greedy matching where a portion of treated subjects cannot be matched and therefore are excluded from the final analysis. As a result, full matching might avoid potential bias due to incomplete matching, which can occur when some treated subjects are excluded from the matched sample.

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

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

where T denotes the investigational product in the target trial (T=1)/historical control (T=0) and X is a vector representing the covariates to be included in the propensity score model. The predictors included in the propensity score model are all available baseline characteristics described in Table 2. These baseline covariates will be utilized without further variable selection or trimming to obtain optimal balance between the matched subjects. Using a large set of covariates is recommended, even if some of the covariates are only related to self-selection and other covariates, and not necessarily to the outcome of interest (Stuart & Rubin 2008, Harris 2016). Some researchers recommend using all available baseline covariates in the analysis (Lim 2018) if the sample size permits.

Step 2: Create SCA by selecting historical patients to match investigational patients in the Target Randomized Trial using full matching. SAS PROC PSMATCH (SAS/STAT® 15.1) will be used for matching, and the maximum number of historical controls to be matched with each IP treated patients and the maximum number of IP treated patients to each historical control will be determined based on the ratio of the number of subjects between IP treated patients and historical controls (Hansen 2004) as well as the performance of balancing baseline characteristics listed in table 2.

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

For a continuous covariate, the standardized difference is:

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

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

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

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

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

To account for the difference in the number of treated and control subjects within each matched set in full matching, a weighted standardized difference will be used and weights will be derived from the strata imposed by the full matching and constructed as follows: IP treated patients are assigned a weight of one, while each historical control patient has a weight calculated as the number of IP treated patients in its matched set divided by the number of controls

in the matched set. (Ho 2007) The weights of controls are scaled such that the sum of the weights from matched controls across all the matched sets is equal to the number of uniquely matched treated subjects.

Each sample estimate (sample means, variances, and prevalences) in the above formulas will be replaced by its weighted equivalent. The weighted mean $\bar{x}_{WT} = \frac{\sum w_i x_i}{\sum w_i}$ and weighted sample variance

$s_{WT}^2 = \frac{\sum w_i}{(\sum w_i)^2 - \sum w_i^2} \sum w_i (x_i - \bar{x}_{WT})^2$ will be used, where w_i is the weight assigned to the ith subject (Austin and Stuart 2015b).

The absolute standardized differences should generally be less than 0.25 (Stuart et al., 2008). An absolute standardized difference of less than 0.10 has been taken to indicate a negligible difference in the mean or prevalence of a covariate between treatment groups (Normand et al., 2001). In addition, the matching process will be evaluated by examining the distribution of propensity scores as well as individual baseline characteristics, including prognostic factors between the Target Randomized Trial investigational arm and SCA using graphical methods such as cloud plots.

The treatment effect on overall survival based on the SCA will be described alongside the treatment effect from the Target Randomized Trial using a Kaplan Meier curve, log rank test, hazard ratio, and 95% confidence interval for the hazard ratio. Weighted estimates incorporating the weights induced by the full matching will be examined.

3.2 Objective 2 is undertaken to illustrate an approach for testing the robustness of the treatment effect to an unobserved or unknown covariate, a potential confounder. While methods such as propensity score matching can adjust for observed confounding, unobserved confounding or unavailable measurement is often a concern compared to the gold standard randomized clinical trial where both observed and unobserved confounders can be balanced. When a key variable is not available for historical patients used to build the SCA, balance between groups in this factor cannot be assured or even described. For example, there may be situations where a key biomarker discovered to have prognostic value only in recent years is available in today's investigational patients, but was not measured or is otherwise unavailable in historical trials. Imbalance in this known or unknown factor could bias the comparison between groups. Under this objective, we illustrate a special type of sensitivity analyses (i.e., tipping point analyses) designed to assess how strong the association of an unobserved confounder with the treatment assignment and the outcome would have to be to change the study inference. If the effects of the investigational product (efficacy or safety) is insensitive over a wide range of plausible assumptions regarding the confounding, then the qualitative effects can be concluded to be secure despite the possibility of unobserved confounders.

Utilizing methods proposed by Lin (Lin 1998), we will adjust the observed treatment effect (HR and 95% confidence intervals) for overall survival to reflect the impact of a theoretical unobserved confounder. Let β and β* denote the true and apparent regression parameters for the treatment effects, respectively. The β is the parameter of interest adjusting for the potential unobserved confounder; while β*, obtained from the observed analysis and necessarily produced by a reduced model due to the unavailability of the unobserved confounder will be adjusted by specifying the distributions of the unobserved confounder among the treatment arms as well as the effects of the unobserved confounder on outcome as

$$\beta \approx \beta^* - \log \frac{e^{\gamma_1}P_1 + (1 - P_1)}{e^{\gamma_0}P_0 + (1 - P_0)}$$

where P_0 and P_1 are the assumed prevalence of the unmeasured confounder among the investigational group and SCA respectively, and the assumed hazard ratio of the unmeasured confounder on the event of interest among the investigational group and SCA is $\Gamma_0 = e^{\gamma_0}$ and $\Gamma_1 = e^{\gamma_1}$, respectively. Without loss of generalizability, we can assume $\Gamma = e^{\gamma_0} = e^{\gamma_1}$. The strength of these assumed relationships between the potential confounder and treatment arm imbalance (ie, prevalences P_0 and P_1) and the potential confounder and overall survival (ie, hazard ratio Γ) will be varied over a range of relevant values so that the point where the conclusion regarding the effect of the drug is changed can be identified.

The assumptions that result in a loss of statistical significance of the treatment effect with the SCA will be highlighted as the 'statistical tipping point'. Assumptions at which the numerical direction of the treatment effect is changed will be highlighted as the 'clinical tipping point'. These tipping points allow an understanding of how imbalanced and influential an unobserved confounder would have to be in order to change the qualitative conclusion. The reader may then make a judgement regarding whether a confounder with this degree of imbalance and impact is likely to exist in the clinical setting and therefore whether the efficacy conclusion is robust against unobserved confounders.

4 RESULTS

4.1 SCA CREATION AND BASELINE BALANCE ACHIEVED

As described in the methods section, full matching was used to select, match, and weight the appropriate patients from the historical pool for inclusion in the SCA to balance the distribution of baseline characteristics between the SCA and the investigational arm from the Target Randomized Trial. Propensity scores were calculated as described in the methods section and utilizing the covariates listed in Table 2. The Cloud Plot in Figure 1 shows the distribution of propensity scores for the investigational arm from the Target Randomized Trial (top) and all available control patients from other trials (bottom). The figure illustrates the degree to which these distributions overlap. The investigational arm from the Target Randomized Trial included 294 patients. Overlap in the distribution of propensity scores for the investigational arm in the Target Randomized Trial and the historical controls was nearly complete. Green dots represent patients who are successfully matched with a patient in the opposite group with a similar propensity score. Red circles and blue x's represent patients for whom a match is not available. These are generally in the tails of the distributions and visually we can see that there are no analogous patients available in this region in the opposite group. Two hundred ninety (99%) in the investigational arm in the Target Randomized Trial were successfully matched. The remaining 4 patients (1%) were not matched and were removed from further analysis. A larger number of control patients are not matched and are excluded from further analysis, but this is of no consequence since our interest is inference regarding the investigational treatment, not the controls themselves.

Excluding unmatched target trial patients from further analysis is a common practice when utilizing match-

ing methods. To many accustomed to analyzing clinical trials, this practice may seem concerning and in direct contradiction to the intent-to-treat principle normally relied upon in clinical trials to preserve the balance between treatment groups afforded by random treatment assignment. However, in this setting, randomization is not utilized and removing patients from the target improves balance between groups rather than threatens it (in essence, prioritizing internal validity over external validity). This practice of removing patients from the target could restrict the matched patients to a set of patients with baseline characteristics that are not as wide ranging as is present in the target or overall disease setting and so the appropriateness of extrapolating the analysis of this precise set and applying it to a more varied population should be considered. But with only 4 patients excluded in this case, there is likely to be very little impact on extrapolation and may illustrate a possible advantage of full matching over greedy 1-1 matching, which may result in more patient exclusions in certain cases.

Figure 1. Cloud Plot: Distribution of Propensity Scores for Investigational Arm of Target Randomized Trial Versus All Available Historical Control Patients
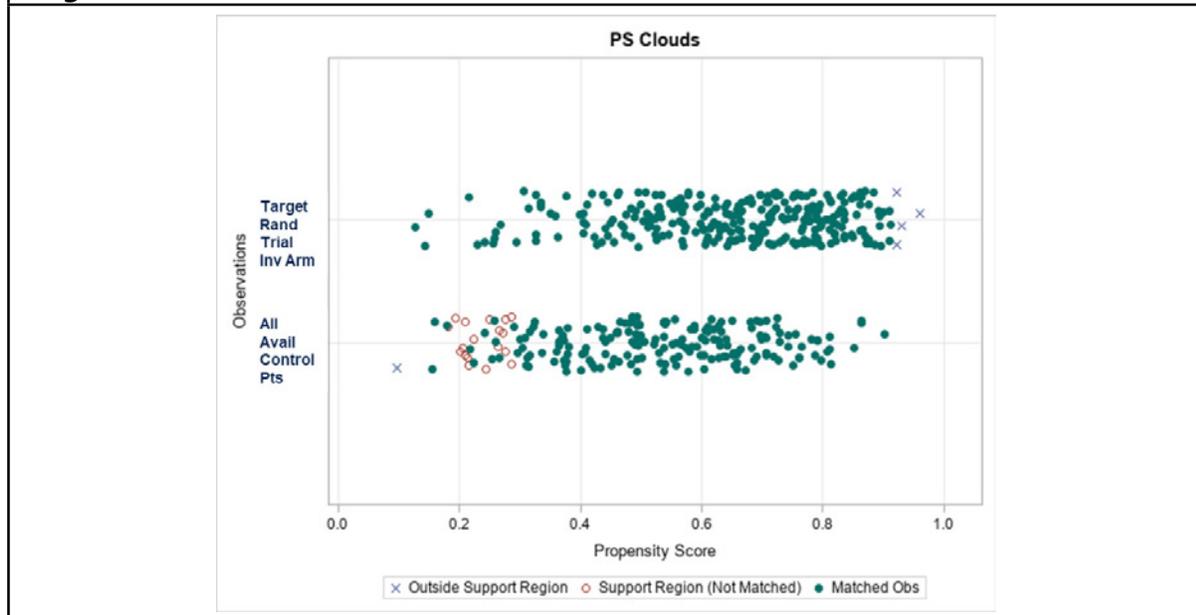
We now consider the degree of balance that has been achieved by the propensity score full matching. The propensity score can be considered a summarization of all baseline characteristics and so we begin by examining the balance achieved in the propensity score.

The distributions of the propensity score for the investigational arm of the Target Randomized Trial and all available historical control patients before matching are shown in the lower set of boxplots in Figure 2. The analogous distributions after matching are shown in the upper region of these figures. There is considerable discordance between the investigational arm of the Target Randomized Trial and all available historical controls before matching. For example, the median for the investigational arm is higher than that of the historical pool. However, after matching, the medians of the groups are very similar.

Figure 2. Distribution of Propensity Scores Before and After Matching
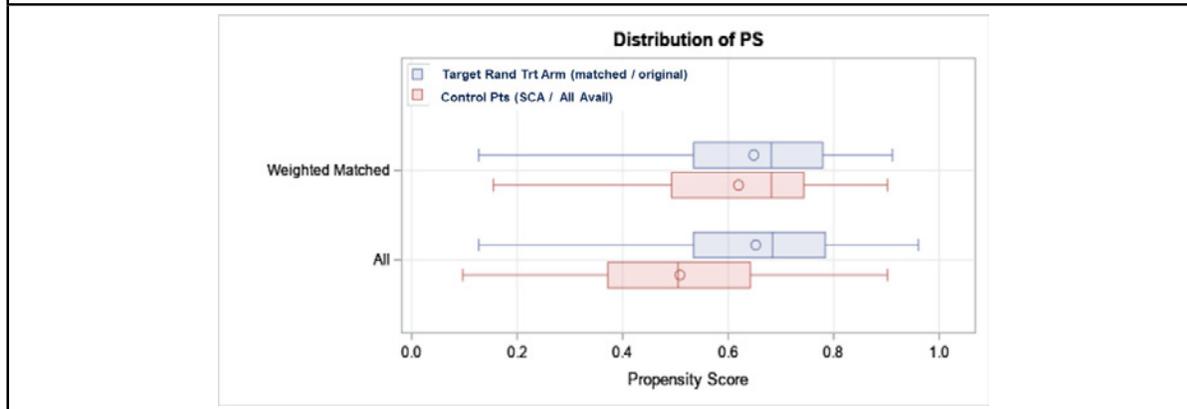
Assessment of balance in terms of individual baseline covariates yields observations consistent with the conclusions afforded above by examination of the propensity scores and indicates very good balance between groups after matching. Figure 3 illustrates the standardized difference between the investigational arm of the Target Randomized Trial and historical controls (before matching) on the left and the same between the matched investigational arm of the Target Randomized Trial and SCA (after matching) for each baseline characteristic examined in this case study. In all cases, reductions in the absolute standardized difference between groups for each variable are observed and the absolute standardized differences after matching are equal to or below 0.10, a commonly used threshold for designating a negligible difference in the mean or prevalence of a covariate between groups, for all but two instances.

Figure 3. Plot of Standardized Difference of Important Baseline Covariates Before and After Matching
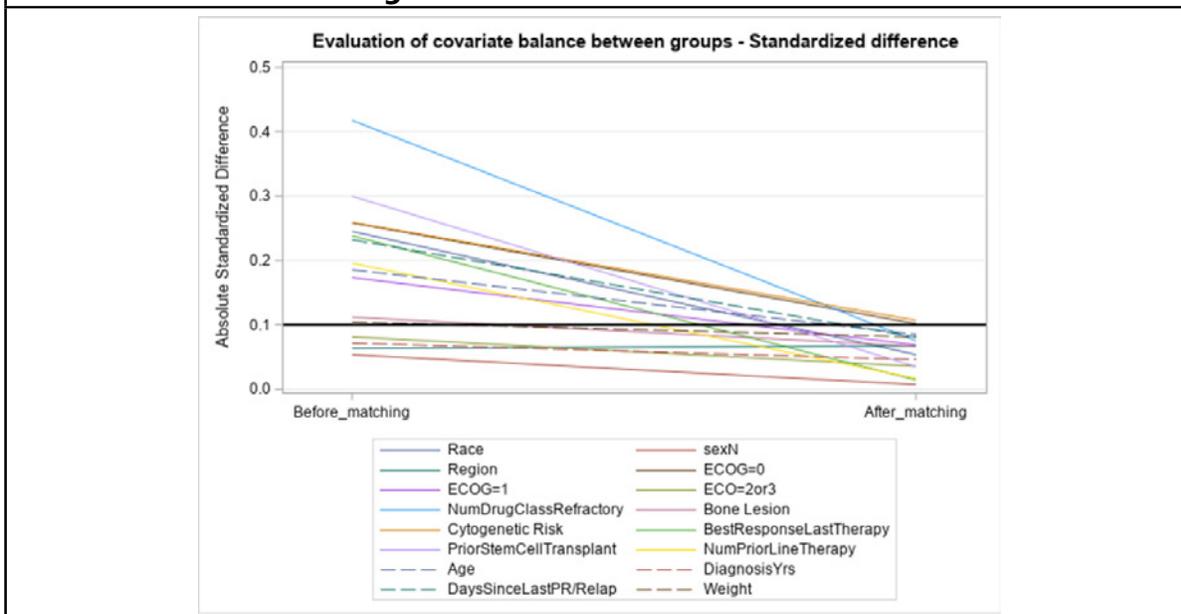
Similarly, examination of the baseline characteristics (on their original scales) for the matched investigational arm of the Target Randomized Trial and the SCA reveals good balance between groups.

The matched investigational arm of the Target Randomized Trial includes 290 patients. As shown in Table 3, most patients were white males from the Europe region with an average age of 63.5 years. Many patients were refractory to 4 or more drug classes (72.1%) and/or had prior stem cell transplant (61%) at baseline. The SCA is quite well balanced with the investigational arm and is weighted to represent 290 patients. Similar to the investigational arm, white males from the Europe region were common in the baseline estimates for the SCA and the average age for the SCA was 64.3 years. Also like the investigational arm, the SCA includes many patients who were refractory to 4 or more drug classes (68.6%) and/or had prior stem cell transplant (59.3%) at baseline. Overall, very good balance in baseline characteristics is achieved between the investigational arm and SCA.

Table 3. Baseline Characteristics - SCA vs. Matched Investigational Arm in Target Randomized Trial		
Baseline Characteristic	**Matched Investigational Arm in Target Randomized Trial (N=290)**	**SCA Weighted Summary (N=290)**
Race (White)	235 (81.0)	241 (83.1)
Region (Europe)	232 (80.0)	224 (77.2)
ECOG=0 ECOG=1 ECOG=2 or 3	105 (36.2) 134 (46.2) 51 (17.6)	91 (31.4) 144 (49.7) 55 (19.0)
Number Drug Classes Refractory (>=4)	209 (72.1)	199 (68.6)
Cytogenetic Risk (high)	29 (10.0)	39 (13.4)
Prior Stem Cell Transplant	177 (61.0)	172 (59.3)
Age (continuous)	63.5 (9.4)	64.3 (9.6)
Days since last PD/relapse to first study dose (continuous)	64.6 (80.1)	71.2 (104.5)
Sex (Male)	174 (60.0)	175 (60.3)
Bone lesion	204 (70.3)	195 (67.2)
Best response to last therapy (≥PR vs. <PR/unknown)	106 (36.6)	104 (35.9)
Number of prior lines of therapy (<4)	64 (22.1)	62 (21.4)
Years since diagnosis (continuous)	6.3 (4.1)	6.1 (4.4)
Weight (continuous)	74.5 (15.3)	73.3 (18.4)

4.2 REPLICATION OF TREATMENT EFFECT ON OVERALL SURVIVAL WITH SCA (OBJECTIVE 1)

In previous sections, we have demonstrated that the propensity score full matching successfully balanced the distribution of baseline characteristics between the SCA and the investigational arm of the Target Randomized Trial. We now move to the first primary objective of this case study, to explore whether the treatment effect based on a SCA (i.e., matched investigational arm from Target Randomized Trial vs. SCA) can mimic the treatment effect based on the randomized control (i.e., investigational arm vs. randomized control in Target Randomized Trial).

Figure 4 provides a description of OS for four groups:
• Investigational arm of the Target Randomized Trial (red)
• Randomized control arm of the Target Randomized Trial (blue)
• Matched investigational arm of the Target Randomized Trial (brown)
• SCA (teal)

The Target Randomized Trial demonstrated a positive treatment effect on overall survival, as evidenced by a separation of the Kaplan Meier curves representing the investigational and randomized control arms of a Target Randomized Trial. The hazard ratio for the investigational arm versus the randomized control is 0.743 with a confidence interval that excludes 1 (95% CI: (0.60, 0.92)). This difference between groups is also supported by the log rank test (p=0.0061).

The treatment effect utilizing SCA is very similar. The Kaplan Meier curve for the SCA visually overlaps and crosses with that of the randomized control and the quantified differences between SCA and the matched investigational arm of the Target Randomized Trial are very similar to the original trial. The hazard ratio for the matched investigational arm versus the SCA is 0.758 with a confidence interval that excludes 1 (95% CI: (0.63, 0.91)). This difference between groups is also supported by the log rank test (p=0.0158).

Figure 4 Overall Survival Treatment Effect Using the Randomized Control Versus Using SCA

Investigational vs. Control from Target Randomized Trial	Matched Investigational from Target Rand Trial vs. SCA
Log-Rank 2-sided P-value 0.0063	Log-Rank 2-sided P-value 0.0158
Hazard Ratio (95% CI) 0.74 (0.60, 0.92)	Hazard Ratio (95% CI) 0.76 (0.63, 0.91)
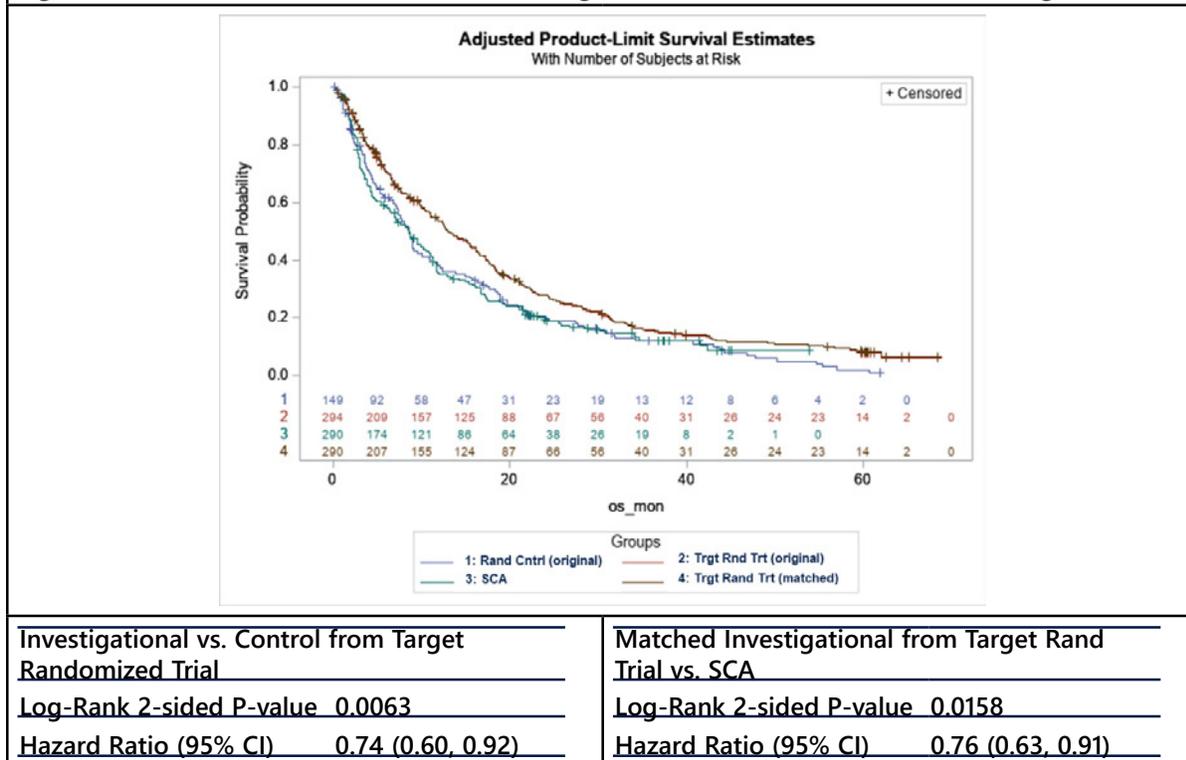
5.0 TIPPING POINT ANALYSES FOR UNOBSERVED CONFOUNDERS – OBJECTIVE 2

This section illustrates an approach for testing the robustness of the treatment effect to an unobserved or unknown covariate. While propensity score matching can be used to balance observed covariates, it cannot guarantee to balance or describe balance for unobserved covariates. The HR and 95% confidence interval for the effect of treatment in the investigational arm of the Target Randomized Trial relative to SCA was estimated to be 0.76 (0.63, 0.91) but one may question whether this is due to the investigational product or due to an imbalance in an unknown or unmeasured confounder.

Using the methods of Lin (Lin, 1998), as described in section 3.2, the observed treatment effect can be adjusted to reflect the possibility of an unknown confounder when the prevalence of the confounder in each treatment arm is known (or assumed) and the influence the confounder has on outcomes is known (or assumed). For example, suppose an unknown confounder is present for only 10% of the investigational arm in this case study while it is present for 30% of the SCA and that the confounder is moderately predictive of overall survival with a hazard ratio for overall survival for those with versus without the confounder of 1.5. Then the adjusted treatment effect separate from the effect of this confounder is estimated to be HR=0.83 with 95% CI (0.69, 0.99). This leads to a conclusion that is qualitatively consistent with that of the original unadjusted treatment effect, that the investigational product is providing a statistically significant benefit. If, however, we had assumed a little stronger imbalance between groups and set the prevalence of the confounder in the SCA slightly higher, say 35%, while all other assumptions remained the same, the adjusted treatment effect separate from the effect of this confounder is estimated to be HR=0.85

with 95% CI (0.71,1.02). These results indicate no statistically significant difference between the investigational arm and the SCA and is qualitatively inconsistent with the original unadjusted analysis. That is the assumption of a 35% prevalence in the SCA rather than 30% is the 'statistical tipping point' where statistical significance of the treatment effect is changed from the original unadjusted analysis. A similar threshold, a 'clinical tipping point', exists where the numerical estimate of the HR exceeds 1 and the numerical trend for the treatment effect is no longer consistent with the original unadjusted analysis.

The example provided above represents just a few possible sets of assumptions regarding the unobserved confounder. To fully understand the possible impact of an unobserved confounder, many sets of assumptions, a grid across all possible or plausible assumptions should be considered. Tables 4 and 5 provide estimates of the treatment effect (HR and 95% confidence interval) adjusted for a theoretical unobserved confounder. The prevalence of this unobserved confounder in the investigational group and SCA are assigned all possibilities, between 0 and 0.8 in increments of 0.05 and are included in the rows and columns of Tables 4 and 5. The relationship between the theoretical unobserved confounder and overall survival is assumed moderate (hazard ratio for those with and without the confounder set to 1.5) in Table 4 and strong (hazard ratio for those with and without the confounder set to 2.0) in Table 5. Entries in each of the cells are the adjusted treatment effects (HR and 95% CI) under these sets of conditions.

The diagonal entries indicated in red text are under the assumption that the unobserved confounder is balanced between the investigational arm and the SCA and therefore the adjusted treatment effect is identical to the original analysis. Moving to the right of the diagonal, as the prevalence of the confounder is assumed to be higher in the SCA than the investigational arm, the HR and 95% confidence intervals initially provide the same conclusion as the original analysis, that there is a statistically significant benefit of the investigational product. Eventually though the imbalance in the theoretical confounder becomes enough to lead to the conclusion that the treatment effect is not statistically significant. This is the 'statistical tipping point' and is represented in Tables 4 and 5 by yellow shading. Moving even further to the right and increasing the discrepancy in prevalence of the confounder between arms even further eventually leads to a numerical estimate of the HR that is bigger than 1 and is no longer directionally consistent with the original analysis. This is the 'clinical tipping point' and is represented in Tables 4 and 5 by green shading.

These tipping points allow an understanding of how imbalanced and influential an unobserved confounder would have to be in order to change the qualitative conclusion regarding the statistical significance or numerical direction of the original unadjusted treatment effect. With this information, the reader may make a judgement regarding whether a confounder with this degree of imbalance and impact is likely to exist in the clinical setting and therefore whether the efficacy conclusion is robust against unobserved confounders.

6.0 CONCLUSION

In this case study in relapsed/refractory multiple myeloma, we have demonstrated that it is possible to produce an SCA from historical clinical trial data using propensity score methods that is well balanced with the investigational arm at baseline. This case study further illustrated that this is possible even when the historical data size is limited and without excessive exclusion of nonmatched patients from the investigational arm, both benefits possibly attributable to the full matching approach.

Importantly, this case study also demonstrated the treatment effect on OS estimated in comparison to the randomized control was very closely matched by that of the SCA, suggesting that SCA could be used to augment or replace a randomized control in future trials in indications where a randomized control is ethically or practically challenging.

Tipping point analyses illustrated in this case study are an effective way of understanding the possible impact of unobserved confounders on the treatment effect estimates and whether the statistical and numerical direction of those effects are reliable despite a reasonable degree of confounding expected in the particular clinical setting.

Table 4: Statistical and Clinical Tipping Points for Overall Survival Analysis (when HR for overall survival of those with and without confounder set to 1.5)

Trt Effect HR (95% CI)

Assumed prevalence of unobserved confounder in investigational arm (rows) × Assumed prevalence of unobserved confounder in SCA (columns)

Inv \ SCA	0.0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
0.0	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.80 (0.66,0.95)	0.81 (0.68,0.98)	0.83 (0.70,1.00)	0.85 (0.71,1.02)	0.87 (0.73,1.04)	0.89 (0.74,1.07)	0.91 (0.76,1.09)	0.93 (0.77,1.11)	0.95 (0.79,1.14)	0.97 (0.81,1.16)	0.99 (0.82,1.18)	1.00 (0.84,1.20)	1.02 (0.85,1.23)	1.04 (0.87,1.25)	1.06 (0.88,1.27)
0.05	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.68,0.97)	0.83 (0.69,1.00)	0.85 (0.71,1.02)	0.87 (0.72,1.04)	0.89 (0.74,1.06)	0.91 (0.76,1.09)	0.92 (0.77,1.11)	0.94 (0.79,1.13)	0.96 (0.80,1.15)	0.98 (0.82,1.17)	1.00 (0.83,1.20)	1.02 (0.85,1.22)	1.04 (0.86,1.24)
0.1	0.72 (0.60,0.86)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.68,0.97)	0.83 (0.69,0.99)	0.85 (0.71,1.02)	0.87 (0.72,1.04)	0.88 (0.74,1.06)	0.90 (0.75,1.08)	0.92 (0.77,1.10)	0.94 (0.78,1.12)	0.96 (0.80,1.15)	0.97 (0.81,1.17)	0.99 (0.83,1.19)	1.01 (0.84,1.21)
0.15	0.71 (0.59,0.84)	0.72 (0.60,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.68,0.97)	0.83 (0.69,0.99)	0.85 (0.71,1.01)	0.86 (0.72,1.03)	0.88 (0.73,1.06)	0.90 (0.75,1.08)	0.92 (0.76,1.10)	0.93 (0.78,1.12)	0.95 (0.79,1.14)	0.97 (0.81,1.16)	0.99 (0.82,1.18)
0.2	0.69 (0.57,0.83)	0.71 (0.59,0.85)	0.72 (0.60,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.68,0.97)	0.83 (0.69,0.99)	0.84 (0.70,1.01)	0.86 (0.72,1.03)	0.88 (0.73,1.05)	0.90 (0.75,1.07)	0.91 (0.76,1.09)	0.93 (0.78,1.11)	0.95 (0.79,1.14)	0.96 (0.80,1.16)
0.25	0.67 (0.56,0.81)	0.69 (0.58,0.83)	0.71 (0.59,0.85)	0.72 (0.60,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.67,0.97)	0.83 (0.69,0.99)	0.84 (0.70,1.01)	0.86 (0.72,1.03)	0.88 (0.73,1.05)	0.89 (0.74,1.07)	0.91 (0.76,1.09)	0.93 (0.77,1.11)	0.94 (0.79,1.13)
0.3	0.66 (0.55,0.79)	0.68 (0.56,0.81)	0.69 (0.58,0.83)	0.71 (0.59,0.85)	0.73 (0.60,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.67,0.97)	0.82 (0.69,0.99)	0.84 (0.70,1.01)	0.86 (0.71,1.03)	0.87 (0.73,1.05)	0.89 (0.74,1.07)	0.91 (0.76,1.09)	0.92 (0.77,1.11)
0.35	0.65 (0.54,0.77)	0.66 (0.55,0.79)	0.68 (0.56,0.81)	0.69 (0.58,0.83)	0.71 (0.59,0.85)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.67,0.97)	0.82 (0.69,0.99)	0.84 (0.70,1.00)	0.85 (0.71,1.02)	0.87 (0.73,1.04)	0.89 (0.74,1.06)	0.90 (0.75,1.08)
0.4	0.63 (0.53,0.76)	0.65 (0.54,0.78)	0.66 (0.55,0.79)	0.68 (0.57,0.81)	0.69 (0.58,0.83)	0.71 (0.59,0.85)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.65,0.93)	0.79 (0.66,0.95)	0.81 (0.67,0.96)	0.82 (0.68,0.98)	0.84 (0.70,1.00)	0.85 (0.71,1.02)	0.87 (0.72,1.04)	0.88 (0.74,1.06)
0.45	0.62 (0.52,0.74)	0.63 (0.53,0.76)	0.65 (0.54,0.78)	0.67 (0.55,0.80)	0.68 (0.57,0.82)	0.70 (0.58,0.83)	0.71 (0.59,0.85)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.93)	0.79 (0.66,0.95)	0.80 (0.67,0.96)	0.82 (0.68,0.98)	0.84 (0.70,1.00)	0.85 (0.71,1.02)	0.87 (0.72,1.04)
0.5	0.61 (0.51,0.73)	0.62 (0.52,0.74)	0.64 (0.53,0.76)	0.65 (0.54,0.78)	0.67 (0.56,0.80)	0.68 (0.57,0.82)	0.70 (0.58,0.84)	0.71 (0.59,0.85)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.93)	0.79 (0.66,0.94)	0.80 (0.67,0.96)	0.82 (0.68,0.98)	0.83 (0.70,1.00)	0.85 (0.71,1.02)
0.55	0.59 (0.50,0.71)	0.61 (0.51,0.73)	0.62 (0.52,0.75)	0.64 (0.53,0.77)	0.65 (0.55,0.78)	0.67 (0.56,0.80)	0.68 (0.57,0.82)	0.70 (0.58,0.84)	0.71 (0.59,0.85)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.93)	0.79 (0.66,0.94)	0.80 (0.67,0.96)	0.82 (0.68,0.98)	0.83 (0.69,1.00)
0.6	0.58 (0.49,0.70)	0.60 (0.50,0.72)	0.61 (0.51,0.73)	0.63 (0.52,0.75)	0.64 (0.53,0.77)	0.66 (0.55,0.79)	0.67 (0.56,0.80)	0.69 (0.57,0.82)	0.70 (0.58,0.84)	0.71 (0.60,0.86)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.93)	0.79 (0.66,0.94)	0.80 (0.67,0.96)	0.82 (0.68,0.98)
0.65	0.57 (0.48,0.69)	0.59 (0.49,0.70)	0.60 (0.50,0.72)	0.61 (0.51,0.74)	0.63 (0.52,0.75)	0.64 (0.54,0.77)	0.66 (0.55,0.79)	0.67 (0.56,0.81)	0.69 (0.57,0.82)	0.70 (0.58,0.84)	0.72 (0.60,0.86)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.93)	0.79 (0.66,0.94)	0.80 (0.67,0.96)
0.7	0.56 (0.47,0.67)	0.58 (0.48,0.69)	0.59 (0.49,0.71)	0.60 (0.50,0.72)	0.62 (0.51,0.74)	0.63 (0.53,0.76)	0.65 (0.54,0.77)	0.66 (0.55,0.79)	0.67 (0.56,0.81)	0.69 (0.57,0.82)	0.70 (0.59,0.84)	0.72 (0.60,0.86)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.92)	0.79 (0.66,0.94)
0.75	0.55 (0.46,0.66)	0.57 (0.47,0.68)	0.58 (0.48,0.69)	0.59 (0.49,0.70)	0.61 (0.51,0.73)	0.62 (0.52,0.74)	0.63 (0.53,0.76)	0.65 (0.54,0.78)	0.66 (0.55,0.79)	0.68 (0.56,0.81)	0.69 (0.57,0.83)	0.70 (0.59,0.84)	0.72 (0.60,0.86)	0.73 (0.61,0.87)	0.74 (0.62,0.89)	0.76 (0.63,0.91)	0.77 (0.64,0.92)
0.8	0.54 (0.45,0.65)	0.55 (0.46,0.66)	0.57 (0.47,0.68)	0.58 (0.49,0.70)	0.60 (0.50,0.71)	0.61 (0.51,0.73)	0.62 (0.52,0.75)	0.64 (0.53,0.76)	0.65 (0.54,0.78)	0.66 (0.55,0.79)	0.68 (0.56,0.81)	0.69 (0.58,0.83)	0.70 (0.59,0.84)	0.72 (0.60,0.86)	0.73 (0.61,0.88)	0.74 (0.62,0.89)	0.76 (0.63,0.91)

Table 5: Statistical and Clinical Tipping Points for Overall Survival Analysis (when HR for overall survival of those with and without confounder set to 2)

Trt Effect HR (95% CI)

Assumed prevalence of unobserved confounder in SCA

Rows: Assumed prevalence of unobserved confounder in investigational arm

Inv. arm \ SCA	0.0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
0.0	0.76 (0.63,0.91)	0.80 (0.66,0.95)	0.83 (0.70,1.00)	0.87 (0.73,1.04)	0.91 (0.76,1.09)	0.95 (0.79,1.14)	0.99 (0.82,1.18)	1.02 (0.85,1.23)	1.06 (0.88,1.27)	1.10 (0.92,1.32)	1.14 (0.95,1.36)	1.17 (0.98,1.41)	1.21 (1.01,1.45)	1.25 (1.04,1.50)	1.29 (1.07,1.54)	1.33 (1.11,1.59)	1.36 (1.14,1.63)
0.05	0.72 (0.60,0.86)	0.76 (0.63,0.91)	0.79 (0.66,0.95)	0.83 (0.69,0.99)	0.87 (0.72,1.04)	0.90 (0.75,1.08)	0.94 (0.78,1.12)	0.97 (0.81,1.17)	1.01 (0.84,1.21)	1.05 (0.87,1.25)	1.08 (0.90,1.30)	1.12 (0.93,1.34)	1.16 (0.96,1.38)	1.19 (0.99,1.43)	1.23 (1.02,1.47)	1.26 (1.05,1.51)	1.30 (1.08,1.56)
0.1	0.69 (0.57,0.83)	0.72 (0.60,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.95)	0.83 (0.69,0.99)	0.86 (0.72,1.03)	0.90 (0.75,1.07)	0.93 (0.78,1.11)	0.96 (0.80,1.16)	1.00 (0.83,1.20)	1.03 (0.86,1.24)	1.07 (0.89,1.28)	1.10 (0.92,1.32)	1.14 (0.95,1.36)	1.17 (0.98,1.40)	1.21 (1.01,1.44)	1.24 (1.03,1.49)
0.15	0.66 (0.55,0.79)	0.69 (0.58,0.83)	0.73 (0.60,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.95)	0.82 (0.69,0.99)	0.86 (0.71,1.03)	0.89 (0.74,1.07)	0.92 (0.77,1.11)	0.96 (0.80,1.14)	0.99 (0.82,1.18)	1.02 (0.85,1.22)	1.05 (0.88,1.26)	1.09 (0.91,1.30)	1.12 (0.93,1.34)	1.15 (0.96,1.38)	1.19 (0.99,1.42)
0.2	0.63 (0.53,0.76)	0.66 (0.55,0.79)	0.69 (0.58,0.83)	0.73 (0.61,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.95)	0.82 (0.68,0.98)	0.85 (0.71,1.02)	0.88 (0.74,1.06)	0.92 (0.76,1.10)	0.95 (0.79,1.13)	0.98 (0.82,1.17)	1.01 (0.84,1.21)	1.04 (0.87,1.25)	1.07 (0.90,1.29)	1.11 (0.92,1.32)	1.14 (0.95,1.36)
0.25	0.61 (0.51,0.73)	0.64 (0.53,0.76)	0.67 (0.56,0.80)	0.70 (0.58,0.84)	0.73 (0.61,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.94)	0.82 (0.68,0.98)	0.85 (0.71,1.02)	0.88 (0.73,1.05)	0.91 (0.76,1.09)	0.94 (0.78,1.13)	0.97 (0.81,1.16)	1.00 (0.83,1.20)	1.03 (0.86,1.23)	1.06 (0.88,1.27)	1.09 (0.91,1.31)
0.3	0.58 (0.49,0.70)	0.61 (0.51,0.73)	0.64 (0.53,0.77)	0.67 (0.56,0.80)	0.70 (0.58,0.84)	0.73 (0.61,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.94)	0.82 (0.68,0.98)	0.85 (0.70,1.01)	0.87 (0.73,1.05)	0.90 (0.75,1.08)	0.93 (0.78,1.12)	0.96 (0.80,1.15)	0.99 (0.83,1.19)	1.02 (0.85,1.22)	1.05 (0.88,1.26)
0.35	0.56 (0.47,0.67)	0.59 (0.49,0.71)	0.62 (0.51,0.74)	0.65 (0.54,0.77)	0.67 (0.56,0.81)	0.70 (0.59,0.84)	0.73 (0.61,0.87)	0.76 (0.63,0.91)	0.79 (0.66,0.94)	0.81 (0.68,0.98)	0.84 (0.70,1.01)	0.87 (0.73,1.04)	0.90 (0.75,1.08)	0.93 (0.77,1.11)	0.95 (0.80,1.14)	0.98 (0.82,1.18)	1.01 (0.84,1.21)
0.4	0.54 (0.45,0.65)	0.57 (0.47,0.68)	0.60 (0.50,0.71)	0.62 (0.52,0.75)	0.65 (0.54,0.78)	0.68 (0.56,0.81)	0.70 (0.59,0.84)	0.73 (0.61,0.88)	0.76 (0.63,0.91)	0.79 (0.65,0.94)	0.81 (0.68,0.97)	0.84 (0.70,1.01)	0.87 (0.72,1.04)	0.89 (0.74,1.07)	0.92 (0.77,1.10)	0.95 (0.79,1.14)	0.97 (0.81,1.17)
0.45	0.52 (0.44,0.63)	0.55 (0.46,0.66)	0.58 (0.48,0.69)	0.60 (0.50,0.72)	0.63 (0.52,0.75)	0.65 (0.54,0.78)	0.68 (0.57,0.81)	0.71 (0.59,0.85)	0.73 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.94)	0.81 (0.68,0.97)	0.84 (0.70,1.00)	0.86 (0.72,1.03)	0.89 (0.74,1.06)	0.91 (0.76,1.10)	0.94 (0.78,1.13)
0.5	0.51 (0.42,0.61)	0.53 (0.44,0.64)	0.56 (0.46,0.67)	0.58 (0.48,0.70)	0.61 (0.51,0.73)	0.63 (0.53,0.76)	0.66 (0.55,0.79)	0.68 (0.57,0.82)	0.71 (0.59,0.85)	0.73 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.94)	0.81 (0.67,0.97)	0.83 (0.70,1.00)	0.86 (0.72,1.03)	0.88 (0.74,1.06)	0.91 (0.76,1.09)
0.55	0.49 (0.41,0.59)	0.51 (0.43,0.62)	0.54 (0.45,0.64)	0.56 (0.47,0.67)	0.59 (0.49,0.70)	0.61 (0.51,0.73)	0.64 (0.53,0.76)	0.66 (0.55,0.79)	0.68 (0.57,0.82)	0.71 (0.59,0.85)	0.73 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.94)	0.81 (0.67,0.97)	0.83 (0.69,1.00)	0.86 (0.71,1.03)	0.88 (0.73,1.05)
0.6	0.47 (0.39,0.57)	0.50 (0.41,0.60)	0.52 (0.43,0.62)	0.54 (0.45,0.65)	0.57 (0.47,0.68)	0.59 (0.49,0.71)	0.62 (0.51,0.74)	0.64 (0.53,0.77)	0.66 (0.55,0.79)	0.69 (0.57,0.82)	0.71 (0.59,0.85)	0.73 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.94)	0.81 (0.67,0.96)	0.83 (0.69,0.99)	0.85 (0.71,1.02)
0.65	0.46 (0.38,0.55)	0.48 (0.40,0.58)	0.51 (0.42,0.61)	0.53 (0.44,0.63)	0.55 (0.46,0.66)	0.57 (0.48,0.69)	0.60 (0.50,0.72)	0.62 (0.52,0.74)	0.64 (0.54,0.77)	0.67 (0.56,0.80)	0.69 (0.57,0.83)	0.71 (0.59,0.85)	0.74 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.94)	0.80 (0.67,0.96)	0.83 (0.69,0.99)
0.7	0.45 (0.37,0.53)	0.47 (0.39,0.56)	0.49 (0.41,0.59)	0.51 (0.43,0.61)	0.54 (0.45,0.64)	0.56 (0.46,0.67)	0.58 (0.48,0.69)	0.60 (0.50,0.72)	0.62 (0.52,0.75)	0.65 (0.54,0.77)	0.67 (0.56,0.80)	0.69 (0.58,0.83)	0.71 (0.59,0.85)	0.74 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.93)	0.80 (0.67,0.96)
0.75	0.43 (0.36,0.52)	0.45 (0.38,0.54)	0.48 (0.40,0.57)	0.50 (0.42,0.60)	0.52 (0.43,0.62)	0.54 (0.45,0.65)	0.56 (0.47,0.67)	0.58 (0.49,0.70)	0.60 (0.51,0.73)	0.63 (0.52,0.75)	0.65 (0.54,0.78)	0.67 (0.56,0.80)	0.69 (0.58,0.83)	0.71 (0.60,0.86)	0.74 (0.61,0.88)	0.76 (0.63,0.91)	0.78 (0.65,0.93)
0.8	0.42 (0.35,0.50)	0.44 (0.37,0.53)	0.46 (0.39,0.55)	0.48 (0.40,0.58)	0.51 (0.42,0.61)	0.53 (0.44,0.63)	0.55 (0.46,0.66)	0.57 (0.47,0.68)	0.59 (0.49,0.71)	0.61 (0.51,0.73)	0.63 (0.53,0.76)	0.65 (0.54,0.78)	0.67 (0.56,0.81)	0.69 (0.58,0.83)	0.72 (0.60,0.86)	0.74 (0.61,0.88)	0.76 (0.63,0.91)

CASE STUDY REFERENCES

Austin PC and Stuart EA. 2015a. Optimal full matching for survival outcomes: A method that merits more widespread use. Stat Med; 34: 3949–3967.

Austin PC and Stuart EA. 2015b. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in Medicine; 34(28):3661–3679.

Friends of Cancer Research whitepaper 2018: available online at https://www.focr.org/events/friends-cancer-research-annual-meeting-2018

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

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

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

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

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

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

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

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

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

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

Journal of Clinical Oncology®

Non-small cell lung cancer (NSCLC) case study examining whether results in a randomized control arm are replicated by a synthetic control arm (SCA).

An American Society of Clinical Oncology Journal

LUNG CANCER—NON-SMALL CELL METASTATIC

Ruthie Davi, Mark Chandler, Barbara Elashoff, Andrea Stern Ferris, Andrew Howland, David Lee, Antara Majumdar, Mark Stewart, Larry Strianese, Elizabeth Stuart, Xiang Yin, Antoine Yver

Medidata Solutions, New York, NY; Medidata Solutions, Inc., New York, NY; LUNGevity Foundation, Potomac, MD; Friends of Cancer Research, Washington, DC; Johns Hopkins University, Baltimore, MD; Daiichi Sankyo, Inc., Basking Ridge, NJ

9108

Background: The FDA's accelerated approval (AA) pathway provides conditional approval for an investigational product (IP) after positive effect on a surrogate endpoint has been provided, allowing patients earlier access to the therapy. Confirmation of a positive effect on the clinical endpoint after conditional approval is required and usually includes a randomized trial. However, such a trial is challenged by availability of the IP outside the trial. Recruitment becomes more difficult, and patients assigned to control are more likely to drop-out and use the non-assigned IP, which may bias the observed treatment effect. In AA settings we propose a SCA composed of patient level data from previous clinical trials to augment or replace the randomized control. Validity of this approach in one case study is assessed by examining if a SCA can replicate the outcomes of a target randomized control (TRC) from a recent NSCLC trial. **Methods:** The patients for the NSCLC SCA were required to have satisfied the key eligibility criteria of the target trial and were further selected using a propensity score-based approach to balance the baseline characteristics in the SCA and TRC. All patient selections were made without knowledge of patient outcomes. **Results:** The results show comparable balance in observed baseline characteristics of the SCA and TRC was achieved. Overall survival (OS) in TRC was replicated by SCA. The Kaplan Meier curves for OS in the SCA and TRC visually overlap. In addition, the log rank test (p = 0.65) and hazard ratio of 1.04 (95% CI: (0.88, 1.23)) were not statistically significant. **Conclusions:** If the SCA had been in place of the randomized control in this study, conclusions about the treatment effect would have been the same. While this may not hold when it is not possible to balance the groups on all confounders, this suggests that in some settings, SCA could augment or replace the randomized control in future trials easing recruitment, retention, and crossover challenges without compromising the understanding of the treatment effect. Future work should examine in what settings SCA is appropriate and consider the implications of potential unobserved confounders.

FRIENDS
of CANCER
RESEARCH

OPPORTUNITIES FOR COMBINATION DRUG DEVELOPMENT:
DATA SOURCES AND INNOVATIVE STRATEGIES TO ASSESS CONTRIBUTION OF COMPONENTS

Friends of Cancer Research (*Friends*) convened a multi-stakeholder meeting consisting of representatives from the U.S. Food and Drug Administration (FDA), the National Cancer Institute (NCI), the pharmaceutical industry, academia, and professional and patient advocacy organizations. This meeting served as a platform for characterizing key challenges and proposing forward-looking solutions in the development and regulation of combination therapies. This can include combinations of two or more investigational drugs, an investigational drug with a previously approved drug for a different indication, or two (or more) previously approved drugs for a different indication as a novel combination therapy. The roundtable discussion was segmented into two parts:

> **Part I: Innovative Methods to Facilitate Combination Drug Development**
>
> **Part II: Strategies for the Development of Unapproved PD-(L)1s Intended for Use in Combination Therapies**

The issue of combination therapy development is especially timely and important for patient access. As the number of combination therapies and codeveloped new investigational drugs increases, clinical trials are requiring increasingly complex study designs to accommodate more trial arms and the accrual of an extensive number of patients. Trial sponsors and regulators will need to balance the level of evidence needed for approval in the context of data that may already be available to ensure equipoise and expedite development. Innovative methods for assessing contribution of components in combination regimens are necessary to facilitate expedited approval. It must also be acknowledged that the goal of all stakeholders in the drug development process is to promote the rapid availability of safe and effective drug products, at the lowest possible cost, for the benefit of patients, while minimizing patients' exposure to

potentially ineffective and harmful agents.

This document is meant to facilitate ongoing discussions to further develop concepts extracted from the roundtable discussion as well as encourage additional input and proposals designed to facilitate the development of combination therapies.

PART I: INNOVATIVE METHODS TO FACILITATE COMBINATION DRUG DEVELOPMENT

The *Friends* multi-stakeholder roundtable began with two case-study presentations by representatives from Bristol-Myers Squibb (BMS) and Janssen. Below are key points from these presentations:

Case Study 1: Nivolumab-Ipilimumab Renal Cell Carcinoma Development Experience

Combination: Nivolumab (A) + Ipilimumab (B) v. Sunitinib (C)[1]

The nivolumab-ipilimumab combination was approved by the FDA for patients with intermediate or poor risk, previously untreated advanced renal cell carcinoma. The pivotal phase III trial was a randomized open-label study. It was randomized 1:1 and compared nivolumab plus ipilimumab with sunitinib. The primary endpoints were overall survival (OS) and objective response rate (ORR). Previous clinical trials investigating single agent efficacy and in combination had been conducted, which contributed to the safety and efficacy information on the contribution of each agent.[2,3,4] Bristol-Myers Squibb noted that the key challenges of combination development are understanding and demonstrating the additional benefit necessary for justifying added toxicity and demonstrating the contribution of each component of a combination.

Case Study 2: Daratumumab-Pomalidomide-Dexamethasone Multiple Myeloma Development Experience

Combination: Daratumumab (A) + Pomalidomide (B) + Dexamethasone (C)

The daratumumab-pomalidomide-dexamethasone (D-Pd) combination was evaluated in patients with relapsed/refractory multiple myeloma (MM) with ≥ 2 prior lines of therapy who were refractory to their last treatment. FDA approval was based on a non-randomized, multi-center, multi-cohort, phase 1b study. The treatment cohorts evaluated daratumumab in combination with multiple regimens. The primary endpoints included maximum tolerated dose (MTD) and ORR.

[1] Motzer R, Tannir N, McDermott D, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. N Engl J Med 2018;378:1277-1290.

[2] Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med 2015;373:23-34.

[3] Antonia SJ, López-Martin JA, Bendell J, et al. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. Lancet Oncol 2016;17:883-895.

[4] Hammers HJ, Plimack ER, Infante JR, et al. Safety and efficacy of nivolumab in combination with ipilimumab in metastatic renal cell carcinoma: the CheckMate 016 study. J Clin Oncol 2017;35:3851-3858.

Daratumumab had previously been approved as a monotherapy for the treatment of patients with heavily treated MM.[5] Pom-dex has also demonstrated progression free survival (PFS) benefit in patients with relapsed and refractory MM compared with pom alone.[6] External data supporting the findings from this single-arm combination study includes the results from two recently completed Phase 3 studies (POLLUX and CASTOR). The POLLUX phase 3 study, in which a combination of daratumumab with lenalidomide and dexamethasone and the CASTOR phase 3 study in which daratumamb plus bortezomib/dexamethasone (Vd) was evaluated against Vd alone induced a high ORR and significantly reduced the risk for disease progression and death in patients with relapsed or refractory MM compared with lenalidomide and dexamethasone. In an indirect comparison with historical data, D-Pd showed a clear benefit over individual components, existing therapy, and other historical datasets.

Table 1. Summary of Case Study Presentations.

Combination	Pivotal Trial Design	Use of External Data
Nivolumab (A) + Ipilimumab (B)	Nivolumab (A) + Ipilimumab (B) v. Sunitinib (C)	Previous clinical trials investigating single agent efficacy and in combination contributed to the safety and efficacy information on the contribution of each agent
Daratumumab (A) + Pomalidomide (B) + Dexamethasone (C)	Daratumumab (A) + Pomalidomide (B) + Dexamethasone (C)	Supported approval of a combination therapy based on a single-arm trial

These case studies were intended to frame the issue of combination therapy development and highlight real-world examples of the use of external data sources in the combination development and approval processes.

[5] US Food and Drug Administration. FDA approves Darzalex for patients with previously treated multiple myeloma [press release]. Available at: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm472875.htm. Accessed 18 January 2019.
[6] Richardson PG, Siegel DS, Vij R, et al. Pomalidomide alone or in combination with low-dose dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2 study. Blood. 2014;123(12):1826-1832.

INTRODUCTION

In 2013, the FDA released the "Codevelopment of Two or More New Investigational Drugs for Use in Combination" guidance for industry.[7] This guidance acknowledged that advances in the understanding of the pathophysiological processes underlying disease had, in many cases, necessitated the use of multiple targeted therapeutic agents to improve treatment response, reduce development of resistance, or minimize adverse events. The guidance was intended to be a high-level description of an approach for the development of two or more new investigational drugs[8] and describes criteria for determining when codevelopment is appropriate, recommends development strategies, and addresses certain regulatory concerns. This guidance also describes the necessity of demonstrating the contribution of each component to the effect of the novel combination therapy. While meeting this expectation is required, there may be more efficient processes and methods for data generation in scenarios involving the combination therapies discussed in this paper.

Although the FDA's 2013 guidance provides an overview of the development and regulatory processes for two or more new investigational drugs in combination, additional opportunities and learnings exist that may warrant exploration of whether a follow-on guidance specific to oncology is needed. Specifically, guidance may be warranted to complement the 2013 guidance, which was limited in scope to two or more novel drugs in combination. Opportunities to streamline combination therapy development programs, which may include the use of external data sources, are discussed. These discussions should be aimed at answering the following questions about demonstration of contribution of individual components to the effect of the combination in the context of a development program that investigates two or more agents that have not been previously approved for the indication:

1. Which trial design is most appropriate (e.g., factorial, adaptive, etc.)?
2. What is the biological rationale?
3. What are the endpoints and opportunities for earlier evidence aside from response?
4. What is the strength of external data needed to say that a therapeutic arm should not be included?

These questions served as the basis for identifying opportunities to use external data to inform development strategies and considerations for combination drug development.

---

[7] "Codevelopment of Two More New Investigational Drugs for Use in Combination" Guidance for Industry https://www.fda.gov/downloads/drugs/guidances/ucm236669.pdf

[8] Defined in the FDA's "Codevelopment of Two or More New Investigational Drugs for Use in Combination" guidance for industry as being a drug that has not been previously developed for any indication

[9] The concepts discussed in this whitepaper may be most applicable to combinations involving at least one previously approved agent

EARLY ISSUES TO ADDRESS IN A COMBINATION THERAPY DEVELOPMENT PROGRAM

The FDA guidance for the codevelopment of two or more new investigational drugs outlines four criteria that should be met by sponsors:

1. The combination is intended to treat a serious disease or condition
2. There is a strong biological rationale for use of the combination
3. It appears that the combination may provide a significant therapeutic advance over available therapy and is superior to the individual agents
4. There is a compelling reason why the new investigational drugs cannot be developed independently

It is important for sponsors pursuing the development of a combination to initiate conversations with the FDA early in their development programs to determine if they meet the above criteria and, if so, whether they are pursuing the most efficient path forward. These conversations are context-dependent and specific to the drug, indication, and need of the patient population at the time of development. For example, the risk-benefit of developing a novel combination with a relatively small improvement in objective response rate (ORR) in a disease setting where there is a high objective response rate to monotherapy would need to be discussed. While the FDA's 2013 guidance encourages early interaction between sponsors and the appropriate Center for Drug Evaluation and Research (CDER) review division, further work is needed to define the parameters of these early interactions between drug sponsors and the FDA and types of data that can inform strategies.

In addition to the need for context-dependent conversations between drug sponsors and the FDA, there is a need for the FDA to further clarify and provide guidance on strategies for demonstrating early activity for drugs being developed in combination and efficient design of development programs for combination therapy products.

Strategies for Demonstrating Early Activity for Combination Therapy Products

Guidance from the FDA on strategies for demonstrating early activity for combination therapy products is needed, especially as the prevalence of codeveloped immune therapies increases. This guidance should more clearly define how sponsors can demonstrate the biological rationale for use of the combination and that their combination has a significant therapeutic advance over existing therapeutic options.

The FDA's 2013 guidance establishes that "sponsors should develop evidence to support the biological rationale for the combination in an in vivo (preferable) or in vitro model relevant to the human disease or condition the product is intended to treat." Sponsor experiences have demonstrated, however, that findings in animal models are not easily translated to clinical predictions, indicating a need for better pre-clinical models. Additionally, while drug sponsors and the FDA have identified the indication of activity as being critical early in development, stakeholders have expressed uncertainty about what can be defined as "activity" or how to demonstrate this activity to the FDA.

Combination therapy products may have greater toxicities (including late on-set toxicities) than monotherapies, making it critical for sponsors to demonstrate a substantial improvement over other available therapies and individual agents on a clinically significant endpoint(s). The magnitude of benefit needed to justify increased toxicities, however, is not always apparent during the early evaluation of combination therapies and codeveloped new investigational drugs. Defining generally what is considered a clinically meaningful benefit and the level of added toxicity that is acceptable given this benefit will be context dependent.

Efficient Design of Development Programs for Combination Therapy Products

<u>Determining when a factorial design is necessary to demonstrate the contribution of each component</u>

In addition to making recommendations for how sponsors may demonstrate early activity of their combination products, the FDA should make recommendations around the efficient design of development programs for codeveloped combination therapies.

Traditionally, clinical trial designs of novel combinations intended for registration have demonstrated the effect of each of the individual components using a multi-arm Phase 3 trial that isolates the contribution of each drug to the overall treatment effect, including time-to event endpoints. To facilitate efficient drug development and better use of resources, FDA has recommended a common control arm when several drugs are being developed for the same population at the same time, and the ineffective investigational arms are discontinued earlier (discussed in more detail in the section of this manuscript titled "Sources of Data and Considerations"). Alternatively, an accelerated approval may be considered based on ORR and a regular approval based on OS results in the same trial.

In combination therapy development programs, the evaluation of the individual drugs as single agents often occurs in earlier phase trials. The utilization of these and other data sources creates opportunities to augment data collected in the pivotal study, reduce the number of patients randomized to a single-agent arm, or replace single-agent arms in phase III trials when appropriate. Situations when factorial designs are not necessary or not appropriate to demonstrate the contribution of each component should therefore be considered.

It has been proposed to develop criteria to assist with the decision-making process to determine when it may be permissible to pursue a more accelerated development strategy. The below suggested criteria could serve as the basis for these early conversations between sponsors and the FDA:

1. The combination shows activity in a population resistant to the individual agent(s)
2. The combination has biomarker driven/associated activity
3. The biological rationale for the combination differs from that of the single agent (this criterion would not be sufficient alone)
4. The combination is in a disease setting where there is no or very little single agent activity

These four suggested criteria are not intended to be binding as it is unlikely that a single combination therapy would meet all four criteria. Additionally, it will be important to consider criteria in the context of the specific disease setting in which the combination therapy is being developed. There may be more confidence in disease settings where the reported ORR has been low compared to where high ORRs have been documented. It may be easier to consider non-factorial trial designs when the single agents to be used in combination have demonstrated efficacy and safety in the randomized trials in that disease.

When a factorial design is not a viable option for trial design, alternative approaches must be pursued to demonstrate the contribution of individual components to the FDA and provide sufficient evidence to assess benefit-risk. These innovative approaches will ensure that an application for approval of a combination therapy will provide the required evidence of the contribution of the individual drugs to the effect of the combination. Alternative approaches will be discussed in further detail in the section of this paper titled "Sources of Data and Considerations."

Determining Which Endpoints Should be Selected

It is also important for the FDA to provide further guidance to sponsors on the appropriate selection of endpoints in clinical trials evaluating two or more drugs for use in combination in a new indication. Endpoint selection for clinical trials evaluating combination therapy products must depend on the research question and the intent of the study. Because combination therapy development may involve the use of external data sources that utilize different endpoints, endpoint selection can be challenging, and different endpoints may provide varying levels of information depending on the trial design. Potential primary and secondary endpoints traditionally used in oncology development are: ORR, PFS, OS, patient reported outcomes (PROs) or clinical outcome assessments, and complementary endpoints such as circulating tumor cells and biomarker-based endpoints (Table 1).

Table 2. Select Endpoints for Oncology Combination Development Trials

Endpoint	Comments
Overall Survival (OS)	Universally accepted "gold standard"; May require larger trial population and longer follow-up to show clinical benefit; Can be impacted by cross-over or subsequent therapies; incorporates impact of a drug's toxicities on survival
Progression-Free Survival (PFS)	Can require smaller patient populations; Definition may vary among trials and measurement may be subject to bias; Requires balanced timing of assessment among treatment arms
Objective Response Rate (ORR)	Can be assessed in single arm trials and requires smaller patient populations; Not a comprehensive measure of drug activity; In rare cancers or rare subpopulations of more common cancers, ORR/DoR may be appropriate
Biomarker-based endpoints	Can enable faster and more efficient clinical trials; Limited availability of validated biomarker-based endpoints
Clinical Outcome Assessments (COAs); Patient Reported Outcomes (PROs)	Useful for the assessment of toxicity and safety; assess whether clinical benefit impacts patient symptoms and quality of life; Can support direct or indirect evidence of treatment benefit; Need to minimize missing data points; Can be subject to bias and judgement

It is important for selected endpoints, such as ORR, to demonstrate the contribution of each component of the combination therapy. Additionally, clarity is needed to determine the strength of clinical evidence required to support the assessment of the contribution of each drug, including the number of patients and clinical trial data evaluating the drug(s) in other disease settings. The assessment of external data can be challenging when different endpoints may be utilized or criteria for defining response may differ among clinical trials.

Sources of Data and Considerations

A range of data sources exist that could help support combination therapy applications and regulatory decision-making. Again, it is imperative for sponsors to present data to the FDA demonstrating the contri-bution of individual components to the safety and efficacy of the combination therapy. These data sources include randomized pivotal clinical trials, randomized supportive trials, pivotal single-arm trials, supportive single-arm trials, patient registries, real-world data, and published clinical data. While randomized controlled trials remain the gold standard, opportunities may exist to use other data sources to augment clinical trial data to potentially reduce the necessary number of patients or control arms to support a more streamlined development path. As external data sources are considered, the advantages and limitations associated with the various sources and whether patient-level data is available will need to be evaluated when designing an

efficient combination development program.

It will also be important for drug sponsors and regulators to evaluate the population for which the combination therapy is being developed when making decisions about data sources to support the drug application. Sponsors should clarify with regulators if the combination is being developed for all relevant patients, a histology or site-specific indication, a particular disease stage, or a biomarker enriched population before committing to a development strategy. There will be different data quality measures such as data quantity, magnitude of effect, type of data, and directionality of data associated with each source that must be considered.

Appropriate statistical methods will also need to be utilized. Sponsors should present a pre-specified statistical analysis plan (SAP) that clearly lays out all hypotheses to be tested and the allocation of significance level for testing multiple hypotheses controlling the overall type I error rate. Principles for utilizing statistical methodologies for leveraging external data in a regulatory setting have also been described in a recent publication.[10] Methodology for augmenting clinical trial control arms is currently being explored by *Friends* and preliminary data is discussed in a recent whitepaper.[11] The necessary number of trial arms, the adequate number of patients, or decisions to remove an arm will be context dependent and will rely on the quality and type of data informing decisions.

Randomized pivotal clinical trials, randomized supportive trials, pivotal single-arm trials, supportive single-arm trials

A 2 x 2 factorial clinical trial design is an optimal design to isolate the treatment effect in a combination therapy (e.g. SOC vs. A vs. B vs. A+B). As mentioned previously, these trials can be inefficient and could produce duplicative data because the evaluation of individual drugs as single agents often occurs in earlier phase trials in combination therapy development programs or in pivotal trials that lead to a monotherapy approval in a different indication. Furthermore, based on the mechanisms of action, clinical activity of the monotherapy may not be anticipated, thus necessitating the initiation of combination therapy investigations earlier in development and making the factorial trial design unethical.

The benefits of randomized Phase 2 clinical trials were acknowledged by multiple stakeholders and were discussed as being optimal for demonstrating the contribution of individual components in a combination therapy. These trials allow sponsors to more readily produce data to identify when a drug or biologic is not going to be active. They also allow sponsors to more readily identify the more active arm to focus development efforts on.

Supportive single-arm trials may also be preferred by some sponsors due to the challenges associated with conducting randomized phase 2 clinical trials and translating their results into

---

[10] Lim, J., Walley, R., Yuan, J., et al. (2018). Minimizing Patient Burden Through the Use of Historical Subject-Level Data in Innovative Confirmatory Clinical Trials: Review of Methods and Opportunities. Therapeutic Innovation & Regulatory Science, 52(5), 546–559.
[11] Exploring whether a synthetic control arm can be derived from historical clinical trials: Case study in non-small cell lung cancer. https://www.focr.org/sites/default/files/pdf/SCA%20White%20Paper.pdf

predictions of Phase 3 benefit and risk. In the absence of randomized trials, however, a comprehensive evaluation of the contribution of each respective component in both preclinical and clinical data is needed. Additionally, in the absence of a randomized trial, time-to-event endpoints, such as OS, will likely not be informative.

As mentioned previously, the FDA has supported the use of a common control in clinical trials as seen in the I-SPY2 trial to minimize the time sponsors need to accrue patients and the number of patients assigned to a standard of care (SOC) control arm.[12] This clinical trial design was also previously discussed at a round-table co-hosted by *Friends* that focused on the optimal development of PD-1 inhibitors and included the proposal of a non-comparative collaborative trial to test multiple PD-1 inhibitors using a common control.[13] The utilization of these methods has been lacking; therefore, further work must be undertaken to describe optimal master protocol designs that collect high-quality data to increase sponsor uptake.

Patient Registries, Real-World Data, Patient-Level Data, and Published Clinical Data

Data collected outside of a traditional clinical trial is becoming more commonly explored for use in regulatory settings. In fact, the 21st Century Cures Act mandates FDA explore the potential use of real-world evidence (RWE) to help support regulatory decisions. FDA recently released their framework for implementing the RWE program.[14] Several challenges were noted with the use of these potential data sources:

- **Confidence in Data Source and Data Quality.** No uniform data standards or standardized definitions of real-world endpoints. Potential biases in data collection and variability in rigor of data collection and missingness.
- **Utilization of Different Endpoints.** Real-world endpoints are typically different than those utilized in clinical trials. There is a need for a better understanding of how real-world endpoints relate to traditional endpoints.
- **Recency of Data.** Age of data and relevance to current clinical practices are important.
- **Access to Patient Level Data.** Patient level data is helpful for propensity score to ensure comparability of patient populations, other statistical methods for making historical data more usable are needed.
- **Publication Bias.** Published data tends to reflect positive or supportive outcomes, which may not provide an accurate or complete picture.
- **Selection bias.** Patients captured by real-world data sources may come from socio-economically disadvantaged groups and there may be unobserved factors that could confound the results.

Patient registries, real-world data, and published clinical data present attractive opportunities for patients, regulatory decision-makers, and drug sponsors. Because these data sources are most often derived from broader populations, they are often more indicative of how a real-world patient population will respond to a given treatment.

The use of data collected outside of a traditional clinical trial is accompanied by multiple challenges. For regulatory decision-makers to be confident in these data it is important for sponsors to consider the age,

---

[12] https://www.ispytrials.org/i-spy-platform/i-spy2

[13] Friends of Cancer Research and Parker Institute for Cancer Immunotherapy Summit: Optimizing the Use of Immunotherapy. https://www.focr.org/events/friends-and-parker-institute-cancer-immunotherapy-summit-optimizing-use-immunotherapy

[14] Framework for FDA's Real-World Evidence Program. https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf

relevance, accuracy, intent, biases in collection, rigor of collection, and missingness of data. First, because rapid advancements are being made in science and medicine, older data may no longer be relevant. Second, caution must be taken to ensure patient populations are comparable between differing data sources. Although this method may not produce the same point estimates of component contribution of the combination therapy, it would be an indicator of whether the benefit exists. Additionally, time intervals between radiographic imaging, differences in dosing and scheduling, and endpoints used to assess treatment benefit may present challenges when aggregating and evaluating data. One possible approach to validating these endpoints would be to compare the SOC RWE results to the SOC results derived from clinical trials. Access to patient-level data from publications would also allow for more robust comparisons as opposed to relying on summary statistics.

Prior to implementation, sponsors should discuss the potential contribution(s) external data could play in regulatory decision-making taking into consideration the challenges cited in the preceding paragraph.

CONCLUSION

The codevelopment of two or more drugs for use in combination in a new indication presents challenges, but the growing availability of external data and development of innovative statistical methods create new opportunities. Improvements are critical to getting safer and more effective therapies to patients quickly and at a lower cost.

Several areas of opportunity were identified to help advance the concepts outlined in this discussion document:

- Define parameters and timing for conversations between FDA and sponsors evaluating two or more drugs for use in combination
- Outline types of data to demonstrate biologic rationale and early activity
- Establish general criteria for when factorial clinical trial designs are not needed and data that could inform this decision
- Provide guidance for the selection of endpoints and acceptable strength of clinical evidence needed to demonstrate contribution
- Organize and collect quality information on the use of external data sources to improve understanding and provide more sophisticated methodologies to more readily use these data sources (possible role for AI and machine learning)

PART II: STRATEGIES FOR THE DEVELOPMENT OF UNAPPROVED PD-(L)1s INTENDED FOR USE IN COMBINATION

INTRODUCTION

Sponsors seeking to develop novel drugs in combination with PD-(L)1s have approached the FDA citing problems accessing approved PD-(L)1s, which inhibit their development processes. These sponsors have noted challenges of partnering with drug sponsors of approved PD-(L)1 agents and thus have elected to develop their own novel PD-(L)1. While the extent of this issue is unknown, the challenges of maintaining equipoise and recruiting patients to an investigational PD-(L)1 arm is clear. One recent analysis of ongoing oncology trials identified 1,716 trials assessing PD-(L)1 immune checkpoint inhibitors in combination with other cancer therapies.[15] Based on accrual needs, more than 380,000 patients would be required for trials containing immunotherapy agents.

Potential trial design strategies for combinations containing an unapproved PD-(L)1

With noted challenges in mind, the following case study was proposed to inform potential trial design strategies:

> A PD-(L)1 checkpoint inhibitor is the approved SOC for the indication to be studied (Control = C). The study objective is to evaluate treatment effect of an unapproved PD-1 checkpoint inhibitor (A) in combination with another experimental drug (B). The treatment effect of A and B are unknown. The primary endpoint of the study is overall survival (OS). The intermediate endpoint is objective response rate (ORR).

Two potential proposals were discussed in the context of a randomized trial of an unapproved PD-(L)1 and an approved PD-(L)1 that incorporates an interim analysis to stop enrollment into one or more arms based on ORR (Figure 1).

Proposal 1: The interim analysis would be based on ORR and would evaluate A vs. A+B; B vs. A+B. Decision criteria would be utilized to stop enrollment into the arm containing monotherapy A, B, or both. Enrollment for either monotherapy arm would stop if shown to have significantly lower ORR than the combination. The final analysis would be based on OS; comparisons would be conducted in A+B vs. C (followed by B vs. C and A vs. C).

Proposal 2: The interim analysis would be based on ORR and would evaluate A vs. C, in which enrollment into arm C would stop if the ORR is similar within a pre-specified margin (no non-inferiority or biosimilar claim); interim analysis would also compare A+B vs B with decision criteria to stop enrollment into arm B, if shown to have significantly lower ORR than the combination. The control arm (C) would be dropped if shown to have equivalent ORR (based on

[15] Tang J. et al. The clinical trial landscape for PD1/PDL1 immune checkpoint inhibitors. Nature Reviews Drug Discovery. 17, pages 854–855 (2018)

a pre-specified margin of error) to monotherapy of same class (in this case, another PD(L)-1). Enrollment into the arm with the other monotherapy (B) would stop if shown to have significantly lower ORR than the combination. The final analysis would be based on OS; comparisons would be conducted in A+B vs A (followed by A vs. B, B vs. C and A vs. C).

Figure 1. Clinical trial design for approval consideration of a combination treatment (no monotherapy indication approval)
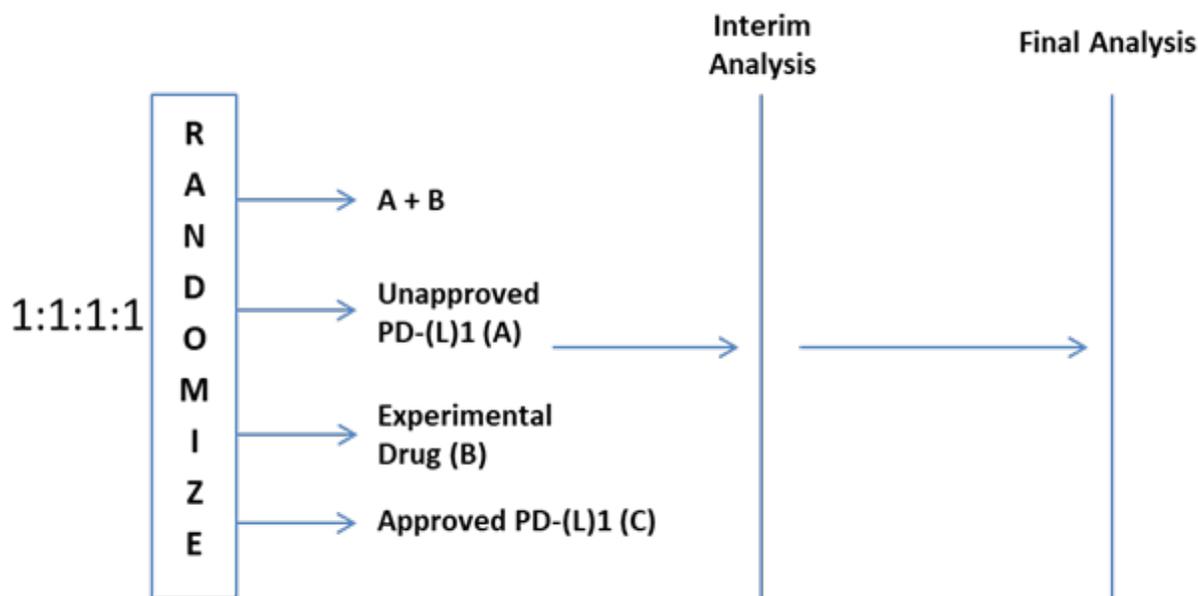
The proposed trial designs are not intended to make a superiority claim by comparison of the monotherapy arms. In addition, it would not support biosimilarity or exchangeability of the experimental PD-(L)1 with the approved PD-(L)1 nor would it support approval of an individual component of the combination.

In addition to clinical trial data, preclinical data to demonstrate the experimental PD-(L)1 is blocking the intended target is necessary. Safety was noted as not being a major concern among unapproved PD-(L)1s particularly given the similarities in the spectrum of toxicities across the approved PD-(L)1s. However, some challenges, in part due to the disease-area/tumor type, were noted in recent combination development programs due to safety concerns.[16] In addition, if the unapproved PD-(L)1 single agent arm is dropped in the clinical trial design proposed in proposal 1, it would limit the amount of long-term data available and would introduce some uncertainty as compared to approved PD-(L)1s. An additional concern was raised around relying on similarities between ORR at the interim analysis due to concern that it may not always translate to similarity in long-term outcomes.

[16] Keytruda: US FDA Reflects On Lessons Learned From Failed Myeloma Studies. https://pink.pharmaintelligence.informa.com/PS122072/Keytruda-US-FDA-Reflects-On-Lessons-Learned-From-Failed-Myeloma-Studies

CONCLUSION

The purpose of considering these proposed clinical trial strategies is to accelerate development of combination therapies that include an unapproved PD-(L)1 through regulatory flexibility, to accelerate the potential utilization of combination therapies across a more diverse range of tissue types, and to potentially alleviate noted challenges by some drug developers.

Additional considerations may also need to be explored to further facilitate the development of combination therapies containing immuno-oncology agents.

- Obtaining sufficient data on safety and efficacy will be important to consider both in the context of regulatory decision-making and in providing adequate data for patients and physicians who may be considering several therapeutic options.
- Improving the understanding of how preclinical analytical data or animal models can inform the toxicity profiles between an approved PD-(L)1 and an unapproved PD-(L)1 should be further defined.
- Creating incentives or policies to encourage greater collaboration between sponsors of approved PD-(L)1s and sponsors seeking to conduct combination studies with a PD-(L)1 backbone could be explored.

FUTURE CONSIDERATIONS

Areas that may require additional guidance:

- **Interactions Between FDA and Drug Sponsors.**
  - » Define parameters and timing for conversations between FDA and sponsors evaluating two or more drugs for use in combination.
  - » Define parameters for FDA input on adaptations or for the pre-specification of adaptations
- **Class Definition.**
  - » Define process for determining a drug class
  - » Demonstration of early activity
  - » Define how preclinical analytical data or animal models can inform the toxicity profiles between an approved PD-(L)1 and an unapproved PD-(L)1
  - » Suggest strategies for demonstrating early activity for drugs being developed in combination
  - » Suggest strategies for demonstrating the biological rationale for use of a combination
  - » Suggest strategies for demonstrating a combination has a significant therapeutic advance over existing therapeutic options
  - » Establish general criteria for when factorial clinical trial designs are not needed and data

that could inform this decision

» Provide clarity on the appropriate selection of endpoints in clinical trials evaluating two or more drugs for use in combination in a new indication

» Provide clarity around dosing strategies (i.e., could it possible to have a FIH dose in the monotherapy arm of a 2x2 factorial study?)

» Provide clarity on the strength of clinical evidence required to support the assessment of the contribution of each drug (i.e., number of patients)

- **External Data Sources in Regulatory Decision-Making.**
  » Indicate which data sources and methodologies are generally recommended or preferred by FDA
  » Provide clarity on how sponsors can incorporate RWE for the identification of the contribution of effect in a combination regimen and for augmenting clinical trial controls
  » Provide clarity on how efficacy data can be extrapolated from one disease setting to another and the value of single agent data in multiple disease settings in subsequent combination approvals in other disease settings
  » Provide clarity on the strength of clinical evidence generated in a clinical trial in different disease settings required to support the assessment of the contribution of each drug

Areas identified as needing further work by all stakeholders include:

- **Pre-competitive Collaborative Partnerships.**
  » Invest time and resources in the improvement of pre-clinical models
  » Discuss further how external data can be shared among sponsors (i.e., consortium opportunities). Patient level data will allow for the most robust comparisons
  » Formulate optimal master protocol designs that collect high-quality data to increase sponsor uptake
  » Consider incentives or policies to encourage greater collaboration between sponsors of approved PD-(L)1s and sponsors seeking to conduct combination studies with a PD-(L)1 backbone
- **Utility of External Data Sources.**
  » Invest resources in the evaluation of the utility of different external data sources
- **Impact Monitoring.**
  » Identify possible metrics for evaluating the impact of streamlined combination therapy development (i.e., opportunity cost, time, number of studies, number of patient participants)

FRIENDS of CANCER RESEARCH

PARKER INSTITUTE for CANCER IMMUNOTHERAPY

DESIGNING THE FUTURE OF CELL THERAPIES

INTRODUCTION

Advancements in cancer immunology and recent clinical experience with emerging cellular therapeutics such as tumor infiltrating lymphocytes (TILs), engineered T-cell receptor (TCR), and Chimeric Antigen Receptor (CAR) T-cell therapies are generating huge interest and activity both academically and industrially. Additional technologies, including cellular therapies based on natural killer (NK) and other immune cells as well as novel gene editing approaches have or will enter the clinic soon. These emerging therapeutics have the potential to rapidly change cancer treatment and may represent a new treatment paradigm.

To date, CAR T-cell therapies have only been approved by the U.S. Food and Drug Administration (FDA) for two types of cancers (certain types of leukemia and lymphoma); other T-cell based therapies have shown remarkable activity in a limited number of solid tumors but have not yet progressed to FDA approval.[1, 2, 3, 4, 5] There is great interest in exploring these new treatment modalities to encompass the treatment of solid tumors, which comprise 90% of all cancers and the majority of cancer deaths.[6] Currently, multiple challenges exist for the successful use of T-cell-based therapies in solid tumors, including issues related to antigen selectivity and expression, the immunosuppressive nature of the tumor microenvironment, tumor T-cell infiltration, and the phenomenon of T-cell exhaustion. Academia and industry are working on multiple ideas to address these barriers, and numerous T-cell-based product candidates are being developed, involving various cell sub-types, autologous and allogeneic approaches, various molecular manipulation strategies, and many different targets. However, due to the diversity of potential targets and the specificity of the human immune system, *in vivo* animal models are limited in their ability to predict product safety and efficacy for T-cell-based therapeutics.

[1] Stevanović S et al. Science 356, 200–205. April 2017
[2] Zacharakis N, et al. Nature Medicine 24, 724-730. June 2018
[3] Tran E et al. Science 344, 641-645. January 2014
[4] D'Angelo et al. Cancer Discovery 8:944. August 2018
[5] Brown et al. NEJM 375:2561-2569. December 2016
[6] SEER Cancer Stat Facts 2019. https://seer.cancer.gov/statfacts/html/all.html

ABOUT FRIENDS OF CANCER RESEARCH

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

ABOUT PARKER INSTITUTE FOR CANCER IMMUNOTHERAPY

The Parker Institute for Cancer Immunotherapy brings together the best scientists, clinicians and industry partners to build a smarter and more coordinated cancer immunotherapy research effort.

The Parker Institute is an unprecedented collaboration between the country's leading immunologists and cancer centers. The program started by providing institutional support to six academic centers, including Memorial Sloan Kettering Cancer Center, Stanford Medicine, the University of California, Los Angeles, the University of California, San Francisco, the University of Pennsylvania and The University of Texas MD Anderson Cancer Center. The institute also provides programmatic support for top immunotherapy investigators, including a group of researchers at Dana-Farber Cancer Institute, Robert Schreiber, PhD, of Washington University School of Medicine in St. Louis, Nina Bhardwaj, MD, PhD, of the Icahn School of Medicine at Mount Sinai, Philip Greenberg, MD, of the Fred Hutchinson Cancer Research Center, and Stephen Forman, MD, of City of Hope.

The Parker Institute network also includes more than 40 industry and nonprofit partners, more than 60 labs and more than 170 of the nation's top researchers focused on treating the deadliest cancers. The goal is to accelerate the development of breakthrough immune therapies capable of turning most cancers into curable diseases. The institute was created through a $250 million grant from The Parker Foundation.

To potentially help a much larger number of patients, in particular those patients with solid tumors and no remaining treatment options, it would be desirable to advance small, data-intensive clinical exploratory studies to differentiate which approaches warrant further focus. These studies would provide an opportunity to optimize the choice of candidates to advance into full product development by generating knowledge that cannot be gained using currently available nonclinical models. Small, early clinical studies also have the potential to facilitate better understanding of the biology of T-cell-based therapeutics and the product attributes driving efficacy and safety. However, clinical data can typically be obtained only after the compilation and submission of an investigational new drug application (IND) for each candidate to be evaluated. These IND procedural requirements can make it prohibitively slow and expensive to pursue this critical opportunity for more than a select few product candidates.

Furthermore, there can be varying interpretations of FDA guidance regarding phase appropriate current Good Manufacturing Practice (cGMP) requirements for manufacturing reagents, plasmids, vectors, and T-cell infusion products for use in the early investigational setting. In consequence, some institutions have imposed very strict cGMP requirements that are more applicable for later stage clinical development on all investigators, significantly increasing the cost and time to manufacture early investigational cell products. Likewise, while existing International Council for Harmonisation (ICH) guidance provide some direction, many of these documents were published at a time when cell therapy was in its infancy; while many of the concepts remain applicable, updated guidance specifically addressing the unique aspects of cellular therapies is needed. Due to the time required to manufacture most cellular therapies (encompassing plasmid and viral vector manufacturing and development of the cellular product manufacturing process and appropriate quality control testing), early clarity in their development is needed regarding the acceptability of a more phase appropriate cGMP approach to manufacturing for early clinical studies.

Ensuring that T-cell-based therapeutics are impactful for the greatest number of patients requires the adoption of new manufacturing technologies as more patients are treated and more clinical, translational, and product quality data is collected during a product lifecycle. This may require modifications to the manufacturing process throughout the different stages of a development program. As product and process knowledge increases, a regulatory strategy that enables adjustment of a process based on patient or patient-specific raw material information to maximize product quality for all patients will be necessary without conducting extensive costly and lengthy studies. This adds complexity to development as current regulatory requirements and processes may not readily allow for patient-level modifications, especially when the understanding of the linkage among product quality attributes, manufacturing processes, clinical efficacy, and safety continue to evolve late in development or after licensure. As product and process knowledge accumulate through the pivotal trial and post-market, an *adaptive* manufacturing process with the goal of generating a highly similar drug product from the patient-specific starting material should be enabled.

Part 1 of this paper outlines a number of regulatory opportunities to accelerate the development of these promising new therapeutics:

- **Opportunities to accelerate early discovery through IND flexibility**
  - o Expansion of the Exploratory IND paradigm to encompass early clinical studies of cell therapies
  - o Flexibility in the application of phase appropriate cGMPs to the manufacturing and testing of plasmids, viral vectors, ancillary materials and reagents, and T-cell-based infusion products for early exploratory clinical trials
  - o Opportunities for flexibility in cell processing and flexibility to permit the use of representative (e.g., high quality, pilot batch) viral vectors in cell product engineering runs
  - o Development of a "parent-child" IND framework to reduce the regulatory burden associated with entering the clinic to test multiple potential product candidates
- **Opportunities to accelerate the optimization of cell products during late stage development and post licensure**
  - o Establishment an adaptive manufacturing process for greatest patient benefit
  - o Develop additional guidance on classification of Chemistry, Manufacturing, and Controls (CMC) commercial process changes

Science- and risk-based approaches will be critical to mitigating and balancing any potential risk associated with either early clinical research or more flexible manufacturing paradigms versus the benefits of developing and optimizing these promising new therapeutics for patients with life-threatening cancers with limited or no therapeutic options. Many of the concepts outlined in this whitepaper may be broadly applicable to multiple types of immuno-oncology cell therapies. T-cell-based therapies, in particular CAR Ts, are used here to highlight specific examples.

Part 2 of the paper describes opportunities for research collaborations and data sharing to advance the cell and gene therapy field:

- **A scientific development consortium to share fundamental data and/or expedite investigational product development and testing processes**
  - o Establish a consortium to promote and facilitate prospective data collection
  - o Develop an exploratory adaptive platform study to evaluate the safety and efficacy of multiple clinical hypotheses and mechanistically defined cell and gene therapies
- **Establish agreed upon standard technologies to facilitate technology transfer between academic innovators and industry GMP producers**

The establishment of research collaborations and data sharing efforts can help facilitate harmonization of cell and gene therapy studies as well as allow for efficient implementation of manufacturing changes or modification of patient cohorts based on accruing clinical data.

PART 1: OPPORTUNITIES TO ACCELERATE EARLY DISCOVERY THROUGH IND FLEXIBILITY

1.1 Expansion of the Exploratory IND paradigm to encompass early clinical studies of cell therapies

FDA's 2006 Exploratory IND Guidance acknowledged the need "to reduce the time and resources expended on candidate products that are unlikely to succeed" and described "some early phase 1 exploratory approaches that are consistent with regulatory requirements while maintaining needed human subject protection, but that involve fewer resources than is customary, enabling sponsors to move ahead more efficiently with the development of promising candidates." This guidance also acknowledged that there is a great deal of flexibility in the amount of data that needs to be submitted with an IND application, depending on "the goals of the proposed investigation, the specific human testing proposed, and the expected risks." The stated purpose of exploratory INDs is to "assess feasibility for further development of a drug or biological product."[7]

Application of the exploratory IND concept to very early, small clinical studies for the purpose of candidate selection for T-cell-based therapeutics would facilitate the critical opportunities described above. However, certain modifications would be needed. The current Guidance explicitly states that an exploratory IND study is intended to involve "very limited human exposure" and to have "no therapeutic or diagnostic intent." Post-infusion expansion of cellular therapies, the durable nature of cellular products, and the ethical requirement to ensure clinical equipoise for patients with life-threatening cancers necessitate that they be dosed at therapeutic levels and with therapeutic intent. Nonetheless, a science-and risk-based approach to an expansion of the exploratory IND concept as it is applied to T-cell-based therapies, to facilitate the critical evaluation of the safety and activity of next generation T-cell-based therapeutics that could fundamentally improve their efficacy via small, data-intensive clinical studies, is possible and appropriate.

An expanded exploratory IND pathway would facilitate the efficient generation of clinical data on multiple T-cell-based product candidates or hypotheses in small (N generally less than 30 patients per cohort) studies, reducing the procedural regulatory burden for both the sponsor and the FDA reviewing division. To ensure patient protection, enrollment in exploratory cellular therapy INDs should be limited to patients with advanced cancers and limited or no treatment alternatives and the total numbers of patients to be treated under an exploratory IND should be limited to the number required to elucidate the hypotheses to be tested. The sponsor should thoroughly justify the number to be treated in the IND and/or protocol.

Exploratory phase protocols should be designed with a focus on patient safety and should incorporate opportunities to minimize risks. Evaluating the behavior of cellular products in humans is currently the most effective way to assess safety, since animal models have been unreliable and product quality attributes that predict safety have been difficult to identify.

---

[7] Guidance for Industry, Investigators, and Reviewers: Exploratory IND Studies. https://www.fda.gov/downloads/drugs/guidance-

Therefore, appropriate consideration should be given to protocol design features such as:

• Judicious dose escalation schemes, dose cohorts, and dose-limiting toxicity (DLT) windows
• Adequate dosing interval and safety assessments between patients enrolled at each dose level during the dose escalation phase
• Ongoing assessment of safety by a safety monitoring committee, including prior to dose escalations and expansion cohorts
• Consideration of incorporation of pre-specified safety, efficacy, and/or futility gates during the expansion phase, such as with a Simon 2-stage design, to ensure appropriate risk-benefit is maintained
• Pre-planned early reporting of safety results could be incorporated into the clinical plan, which could be agreed to during an INTERACT or pre-IND meeting or during IND review to avoid introducing unnecessary delay

Additional procedures to ensure patient protection could include explicit characterization of these INDs and the associated protocols as "exploratory," intended to support studies involving "very early clinical research," ensuring appropriate patient informed consent and IRB and FDA oversight with particular scrutiny applied to ensure that the appropriate patient population will be enrolled.

Such an approach is consistent with FDA's many expedited development programs that, while focused on later stages of development, explicitly acknowledge the need to balance the risks associated with early exposure to unproven investigational therapies against the potential bene-fits of early access to those therapies.

The following sections outline how phase appropriate cGMP compliance focused on product quality and patient safety, and a streamlined "parent-child" IND alternative to the current sin-gle IND per drug product process would further facilitate the conduct of these studies under an expanded exploratory IND paradigm. T-cell-based therapies are used as specific examples. We note that some of the proposals may be relevant for other types of gene-editing technologies or immune cell therapies.

1.2 Flexibility in the application of phase appropriate cGMPs to the manufacture and testing of plasmids, viral vectors, ancillary materials and reagents, and T-cell-based infusion products for early exploratory clinical trials

FDA's 2008 Guidance for Industry: cGMP for Phase 1 Investigational Drugs[8] provides a framework whereby more phase appropriate manufacturing can occur for early studies. The recognition that smaller scale manufacturing processes may be excluded from some of the controls required for later stages of development where larger numbers of patients are exposed to treatment or for

---

[8] Guidance for Industry CGMP for Phase 1 Investigational Drugs. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/%20Guidances/UCM070273.htm

commercialization is critical to innovative research and establishing a better understanding of the human biological impacts of new therapeutics in small investigational human studies. However, consistent understanding and interpretation of this guidance, especially as it would apply to exploratory cellular therapy INDs, is needed. We provide several key examples below where explicit alignment between FDA, academic and government institutions, and industry with flexible approaches would facilitate the early exploratory clinical studies described above.

Implicit in any approach for manufacturing Phase 1 appropriate materials is a focus on patient safety, and the concepts below are proposed with an emphasis on risk assessments and analytical testing to determine and manage potential impact to patient safety. As such, T-cell-based cellular products would undergo release testing following manufacture for standard safety attributes, such as sterility, absence of mycoplasma and endotoxin, viral integration elements (vector copy number), identity, purity, and potency.

The principle of a more flexible approach, if chosen, would be to ensure patient safety and to take steps to ensure that if the decision is made to pursue full product development, results obtained during the exploratory study would be similar to those for the subsequent investigational product used in the full development IND. However, reductions or deferral of testing relating to process consistency and long-term stability in these early "screening" studies would result in time and resource savings. Process optimization aspects of product development would be fully addressed during subsequent development for any candidate for which a decision has been made to move forward with full development. We note that sponsors may wish to assure that an adequate number of retain samples are obtained during the early product manufacturing to facilitate subsequent manufacturing comparability.

We note that if remarkable efficacy were seen for a product development candidate tested in an "exploratory IND, the requirement for a full IND with more standard manufacturing process development would still apply with the potential for associated delays. Sponsors may determine this risk is acceptable given the potential to save time and resources by eliminating product candidates that are destined to fail, resources that could be dedicated to intensive efforts to accelerate the development of the promising candidate. Finally, sponsors may decide to mitigate this risk by pursuing limited process development activities in parallel with clinical studies under an exploratory IND.

A risk-based approach to requirements for the production of raw materials and drug substance (DS) (e.g., viral vectors, including lentiviral vectors) for T-cell-based therapeutics could more rapidly lead research teams to better combinations of therapeutics, scFv alterations, novel manufacturing interventions, etc., which would lead to more robust products that don't fail in later stage development studies. Flexibility to permit the use of representative viral vectors in cell product engineering runs would result in significant time savings at little or no risk to patients. These opportunities could reduce the total time to manufacture investigational T-cell-based

therapeutic candidates for use in an early clinical study under an exploratory IND by approximately 50%, as depicted in Appendix 1: Section A and described in greater detail below.

1.2.1 Reduction in the infrastructure requirements for the manufacture of plasmids

Currently, production of plasmid DNA for downstream production of viral vectors and/or for gene editing tools is often outsourced to a limited number of companies, resulting in high costs and long manufacturing queues. Generally, sponsors and academic researchers have the technical capabilities to produce these plasmid DNA's, but interpretations of FDA guidance have led to institutional policies requiring cGMP grade plasmids for clinical studies. Due to the high infrastructure requirements (ISO-7 clean rooms, fully developed quality systems and cGMP trained personnel and associated resources) needed to produce cGMP grade plasmid DNA, many institutions have not invested in the development of the manufacturing and quality infrastructure to produce these raw materials internally. In the industry setting, the impression that cGMP grade plasmids may be required increases the cost and time associated with manufacturing investigational cellular products. Manufacture of cGMP grade plasmids for small, exploratory clinical trials of multiple early cellular product candidates would unnecessarily increase the cost and time to conduct these studies since it is expected that many of the candidates would not progress into full product development.

As an alternative to a requirement for cGMP grade plasmids, *high-quality* (HQ) *fit-for-purpose* plasmids may be acceptable. Plasmid DNA can be tested and sufficiently characterized to confirm its fit-for-purpose suitability for downstream use in early, exploratory clinical trials with little risk to patient safety.

For example, the regulatory burden associated with the manufacture of HQ DNA plasmids for exploratory clinical studies could be reduced by eliminating the need for an *E. coli* master cell bank (MCB). Note that a sponsor could also make a business decision to create the MCB and then freeze it, deferring the need for time consuming and expensive testing until a decision was made to go forward with full development with that product candidate. Manufacturing could occur with review of production protocols, analytical results, manufacturing batch records, and release tests could be performed by a second technical rather than quality assurance personnel. A certificate of testing (CoT) could be produced summarizing the test results and could include tests similar to those in Table 1 below. In essence, a CoT is similar to a certificate of analysis (CoA) but differs in a few key elements: 1) tests are mostly compendial and may not be fully qualified/validated; 2) tests may be peer reviewed by a technical expert (in lieu of a quality assurance resource); and 3) test results have a "Target Value" in lieu of "Acceptance Criterion." In addition, because the plasmid DNA materials are stable when frozen and anticipated to be used quickly in downstream manufacturing of viral vectors, at this stage the need to generate stability data could be weighed against the timing of use and available research data and in some cases, waived.

Table 1: Proposed "fit for purpose" testing of plasmid DNA for early phase clinical studies.		
Attribute	**Test Method**	**Target Value**
Appearance	Visual	Clear, colorless, no visible particulates
Concentration	Absorbance (A280)	Target +/- 10%
Purity	Absorbance (A260:280)	1.7 – 2.0
Safety	Endotoxin by LAL	< 25 EU/mg
Safety	Bioburden Testing	< 10 CFU/10mL
DNA Homogeneity	Gel Electrophoresis	>75% Supercoiled
Residual Host Protein	BCA Assay	Report result
Residual Host DNA	qPCR	Report result
Residual Host RNA	SYBR/Gel electrophoresis	Report result
Identity	Restriction digest and AGE	Conforms to reference
Identity	DNA sequencing	Confirm expected sequence at appropriate method sensitivity

The above proposals supported by appropriate documentation would facilitate the creation of greater manufacturing capacity by reducing the barriers to entry, permitting manufacturing of plasmid DNAs (for use in downstream manufacturing of viral vectors) at the academic or sponsor level, and further decompressing full-scale GMP manufacturing capacity for full product development manufacturing needs.

1.2.2 Use of phase appropriate vector testing strategies, including reductions in the replication competency testing requirements

In the context of early, exploratory clinical studies in patients with limited or no remaining treatment options and very poor long-term survival, the risk-benefit of earlier access to potentially beneficial T-cell-based therapeutic treatment is reasonable. Despite theoretical concerns, the risk of replication competency-related recombination events using 3rd generation viral vectors is extremely low as the elements required for virus replication are separated across 3 or 4 different plasmid DNAs and the 3' UTR portion of the transfer plasmids have been modified resulting in transcriptional inactivation of the LTR in the proviruses after integration. With respect to viral vectors currently used in cell therapy products, researchers have documented that, to date, no viral vector recombination events have been observed across hundreds of patient product tests.[9, 10]

9 Cornetta K et al. Molecular Therapy 26:1. January 2018
10 Cornetta K et al. Molecular Therapy: Methods & Clinical Development 10:371-378. September 2018

The current replication competency virus assay is based on testing vector supernatant or end of production cells on susceptible human cells over an 8-10-week period; this requirement adds significant expense and time to the overall product manufacture and release timelines. In order to address lengthy timelines required for viral vectors to be manufactured and released, elimination of the replication competency test for release of viral vector drug substance (DS) is proposed. In lieu of testing for the replication competency test in the viral vector DS material, it is proposed that a surrogate qualified/validated qualitative polymerase chain reaction (qPCR) test be done for the GAG and vesicular stomatitis virus G glycoprotein (VSV-G) or similar envelope gene sequences depending on the viral vector pseudotype, as has been recently suggested by Skrdlant et al.[11]

Vector and cellular drug product release decisions for such exploratory studies could be made on the basis of surrogate testing; if required, full, culture-based replication competency-based testing could be conducted in parallel in the background. The results of the full-culture testing would be available within the period of post-infusion patient follow up during which time patients would be followed for the development of treatment-related malignancy.

1.2.3 Use of risk-based approach for determining safety of reagents used in early clinical trials

Many reagents are employed in the production of viral vectors and therapeutic T-cells. Extensive manufacturing requirements for reagents (e.g., activation beads, selection reagents, cytokines, recombinant growth factors, etc.) create a time and cost burden in early development. Typically, these reagents are produced and stored frozen at higher concentrations to ensure greater stability. During manufacturing, a reagent would be thawed and diluted to the working concentration and then added to a much larger culture volume. Unless the reagent is used constantly throughout the entire manufacturing process, several rounds of washing, media changes, and formulation of the final cell product will significantly dilute the reagent. Similar to the manufacturing requirements for plasmid DNA, fit for purpose requirements (relying on science- and risk-based approaches to ensure patient safety and quality of the reagent) for HQ reagents used within the manufacturing process for early phase clinical studies would significantly reduce the cost and time burden associated with using innovative reagents. An emphasis on risk assessments to identify potential impact to patients (e.g., sterility/bioburden, products of animal origin, etc.) could provide guidance to academic researchers and industry partners. For non-pharmacopoeial reagents of non-biological origin, a review of a certificate of testing may provide assurance that a reagent is fit-for-purpose for use in the manufacturing of cellular products for small, early clinical studies. For reagents of biological origin (e.g., human serum), purchase from an accredited supplier, along with a certificate of analysis (source, sterility, endotoxin, infectious agents, mycoplasma) can confer suitability of use.

---

[11] Skrdlant LM et al. Molecular Therapy: Methods & Clinical Development Vol. 8 March 2018

Table 2 below provides an illustrative example of the approach to characterization of a novel recombinant cytokine, such as one that may be used as a media supplement in a representative T-cell-based therapeutic manufacturing process, which could form the basis of a "Certificate of Testing." These testing elements are based on the concepts provided in ICH Q6B and other regulatory guidance and represent an assessment of the reagent's identity, purity/impurity, potency and safety. Historical knowledge of production of the intended reagent should be utilized to set appropriate quantitative or qualitative science- and risk-based acceptance limits.

Table 2. Representative characterization testing of a recombinant protein cytokine reagent.

Attribute	Test Method	Target Value
Appearance	Visual	Clear, colorless, no visible particulates
Concentration	Appropriate methodology (protein-based BCA or other)	Target concentration +/- 20%
Purity	HPLC – SEC	$\geq$ 90% product peak
Safety	Endotoxin by LAL	$\leq$ 25 EU/mg
Safety	Sterility	USP <71>
Safety*	As needed	Report results
Residual Host Cell Protein	ELISA (if available)	Report result
Residual Host DNA	qPCR or PicoGreen Assay (for dsDNA)	Report result
Potency	Appropriate methodology (ELISA or activity assay)	Report result
Identity	Identity by MS	Confirmation of identity

*Additional test(s) may be required based on the source of the reagent (e.g., mammalian production

1.3 Opportunities for flexibility in cell processing

Given the resources required and complexity of manufacturing T-cell based therapeutic products, identifying similar flexibilities in the cell processing space would provide significant opportunities for innovation. While a robust discussion of the kinds of flexibility desired is out of scope for this document, a few examples and the anticipated impacts are offered below. Typically, a T-cell-based therapeutic is engineered using a relatively similar set of manufacturing unit operations: acquisition of patient starting material through apheresis/leukapheresis, isolation/purification of the T-cells through gradient, magnetic or alternative selection means, activation, and retroviral transduction to introduce the CAR or TCR, expansion of the engineered T-cells, and final harvest and cryopreservation. While there are variations on the above approach and a number of different pieces of equipment employed in various manufacturing processes, the general process lends itself to some potential flexibilities in the early development space.

1.3.1 Flexibility to permit the use of representative viral vectors in cell product engineering runs

For T-cell-based therapeutic products, current process development is often interpreted as requiring the use of GMP grade viral vector in the three engineering runs conducted to confirm the adequacy of the cellular product manufacturing process. Clarity that the use of "representative pilot" (i.e., development grade viral vectors manufactured in accordance with the final manufacturing process) would be acceptable, could result in significant time savings because the cellular product engineering runs could be run in parallel with the final GMP production runs for the viral vector. Additionally, because much of the development work for autologous cell therapies is done at scale, fewer engineering runs (e.g., 2) would be reasonable. As such, data from both development runs (e.g., in the process development lab) and engineering runs (e.g., in the GMP manufacturing facility) could be combined to demonstrate adequate control of the process.

1.3.2 Utilization of scale-models

Leveraging scale-down models is critical in examining variations in the manufacturing process and impact to T-cell phenotype and functionality. Currently, many of these experiments are often repeated numerous times at scale to demonstrate control of the process. This often requires significant investment in time, personnel, and reagent resources to accomplish. Given the significant patient-to-patient variability introduced by the various conditions of starting apheresis materials in many of these early clinical studies (e.g., age of patient, extent of prior lines of therapy, T-cell health and baseline population distribution, viability, etc.), it is difficult to precisely identify sources of variability. This exercise is challenging even in more mature areas, such as current approved CAR T therapies for hematological malignancies. Flexibility in the use

of scale-down models (as mentioned above in conjunction with a limited number of "at scale" development and engineering runs) would provide much needed ability to move promising pre-clinical programs into these early exploratory studies.

1.3.3 Phase appropriate release testing

For early phase exploratory trials, a focus on testing cell product components related to safety can provide flexibility. Safety would be assessed via testing for sterility, mycoplasma (via a rapid testing paradigm), endotoxin, etc., that are each important to demonstrating a lack of contamination of the cell product. Testing the cell product for elements of the viral engineering activity through assessment of integration of the vector into the T-cell genome can be done by determining the average vector copy number (VCN) via qPCR. Additionally, surrogate measures of viral replication competency can be done using qPCR with primers against various elements of the viral genome (discussed above as part of the relaxation of RCR/RCL testing above). Identity, purity, and potency are important release assays used to demonstrate that a particular manufacturing process was able to successfully yield the expected product. Identity is confirmed by flow cytometric staining for key cell surface markers, such as CD3, CD4, CD8, specific introduced CAR or TCR, etc., are typically used to provide assurance that the appropriate cell product was produced. This is of considerable importance if a manufacturing facility is involved in producing multiple products targeting different antigens. Many sponsors conduct additional characterization with numerous other cell surface markers to further understand their product, but these analyses should be focused on gathering additional data. Potency of CAR and TCR-based cell products is often demonstrated using either a cytokine release (e.g., IFN-gamma or TNF-alpha production) or cytolytic killing assay whereby the cell product is incubated with cells expressing the target and shown to bind and kill these "target" cells. Complexity and variability in both of these testing approaches in the early phase of development results in challenges in establishing numerical acceptance criteria. Additionally, limitations in the amounts of samples available, condition of the cell products (e.g., fresh testing vs. cryopreserved product testing) also contribute to variability and challenges with numerical acceptance criteria. Flexibility on the acceptance criteria would be advantageous and utilizing the report result verbiage for reporting could help move programs into the clinic faster.

1.4 Regulatory procedural flexibility – Development of a "parent-child" IND framework to reduce the regulatory burden associated with the clinical testing of multiple potential product candidates

Currently, outside of the area of non-engineered T-cells, sponsors must submit a new IND for each potential T-cell product candidate for which they wish to conduct clinical testing, and each IND requires significant time, resources, and expense both for the sponsor and the FDA reviewing division. In the setting of small, early data-intensive clinical studies intended to investigate the safety, feasibility, and mechanism of action of several closely-related T-cell-based can-

didates or related manufacturing process alterations (for example process alterations to maintain "stemness") a more efficient "parent-child" IND structure and process may be appropriate.

An exploratory "parent-child" IND is a feasible approach to reducing the regulatory procedural burden associated with evaluating multiple highly-related T-cell-based therapeutic constructs or manufacturing alterations in small clinical studies. The "parent" IND would contain common sections providing all of the common information relevant for the to-be-tested initial candidates or manufacturing alterations. For each candidate or manufacturing alteration, a "child" IND would also be submitted. This "child" IND would depend on heavy cross-referencing to the common sections in the "parent" IND while providing only the candidate or process specific information (e.g., CMC or nonclinical data) in separate sections (see Appendix 1: Section B). We note that cross-refencing to previously submitted information, with appropriate authorization, is an accepted practice.

At the time of initial IND submission, the "parent" and "child" IND could be assigned separate IND numbers, to facilitate safety reporting, etc., but reviewed in parallel within the standard 30-day IND review window. Each subsequent "child" IND would be subject to the normal 30-day review window. Consistency in approach to each "child" IND may be facilitated by assignment of the "parent" and all related "child" INDs to the same FDA review team.

The exploratory IND would include an explicit agreement by the sponsor that once the early testing of a particular construct or process is completed or discontinued, the associated exploratory "child" IND would be withdrawn. If the sponsor intends to proceed with full development of a candidate or manufacturing process, a new, traditional IND would be submitted for that candidate. Subsequent candidates or processes consistent with the common information in the original "parent" IND could subsequently be added as additional "children" to the original "parent", again relying heavily on cross-references. FDA would have an opportunity during the 30-day review to reject any proposed "child" as insufficiently related to the "parent" to justify acceptance. Ultimately, once the sponsor determines that no additional early candidates closely related to the original exploratory IND will be tested, the exploratory IND would be withdrawn.

The use of the parent-child IND approach would result in significant time and resource savings for sponsors and the FDA reviewing division and could facilitate the generation of critical knowledge regarding the safety, feasibility, and mechanism of action of many more T-cell-based therapeutic constructs and manufacturing alternations than is possible under the current regulatory paradigm and that cannot be generated in nonclinical studies. This reduced burden has the potential to be particularly significant for the most innovative academic and small biotech sponsors with limited resources.

Because the time and resource savings associated with the use of "parent-child" INDs would only be realized in situations where most of the information contained in the "parent" IND

would be relevant to all of the investigational candidates, the use of "parent-child" INDs would be limited to situations where the commonalities between the early cellular therapy candidates or manufacturing interventions are great enough to produce real gains in efficiency for both sponsor and the FDA reviewing division. For example, an exploratory IND might be limited to candidates directed at the same target. Whether a parent-child IND is appropriate for a particular set of candidates could be discussed in an INTERACT or pre-IND meeting or the justification could be provided in the IND itself (with an associated risk of delay if FDA disagrees).

1.5 Flexible Regulatory pathways to enable manufacturing and testing evolution during late stage development and post licensure

1.5.1 Regulatory opportunities to enable adaptive manufacturing processes for greatest patient benefit

In the case of T-cell-based therapeutics and other cell-based therapies, making these products impactful for the greatest number of patients may require adjusting manufacturing parameters for specific patient subsets. The first generation of engineered T-cell products treating patients with hematological malignancies (e.g., ALL, DLBCL, CLL, Multiple Myeloma) use the same manufacturing process for all patients. These products have made a meaningful impact over the standard of care in these diseases. The single manufacturing process framework was chosen for regulatory expediency and a lack of product knowledge to discriminate between patients. At the same time, patient-to-patient variability in the quality of T-cells from these patients leads to suboptimal drug product quality for a subset of patients, when a single manufacturing process is used for all patients. In order to increase the number of patients responding to these treatments, it may be necessary to adapt the manufacturing process for a subset of patients to increase the efficacy for the specific patient cohort, without impacting safety and efficacy for patients already responding using the original manufacturing process.

These new process parameter combinations for patient subsets are discovered during clinical development as more patients are treated and more clinical, translational, and product quality data are collected. As product and process knowledge increases, a regulatory strategy that adjusts a process based on patient or patient-specific raw material information to maximize product quality for all patients will be necessary without conducting extensive costly and lengthy clinical studies. This adds complexity to development as current regulatory requirements and processes may not readily allow for patient-level modifications, especially when the understanding of the linkage among product quality attributes, manufacturing processes, clinical efficacy, and safety continue to evolve late in development or after licensure.

Traditionally, manufacturing process lock is completed in advance of late-stage clinical trials to be able to repeatedly measure effect across many patients. Product and process knowledge is currently being generated to enable the development of an adaptive manufacturing process

with the goal of generating a highly similar drug product from the patient-specific starting material. The product and process knowledge to enable adaptive manufacturing in most cases will not emerge until a large number of patients are treated since the correlative analysis to discover the relationship is not available until enrollment of the pivotal trial. An example of this type of relationship includes the frequency of specific T-cell subtypes.[12] An example of emerging product knowledge and the rationale for an "adaptive" approach are discussed below.

Box 1. Example of Emerging Product Knowledge

A licensed product using a fixed manufacturing process leads to a durable response in 40% of patients. During clinical development, it is observed in a small subset (~10%) of non-responders that adapting the manufacturing process can convert these non-responders to responders. If a cohort can be identified with a control point and a separate set of process parameters that will meaningfully change product attributes to improve the biological activity of the product for a subset of patients, a regulatory mechanism permitting these adaptive changes would benefit patients in later trials and commercially. Running a prospective trial to support a supplemental approval for a very small subset of patients would not be viable. Existing guidance, such as ICH Q11 and ICH Q5, provide a framework for prospective process flexibility in the presence of strong product attribute understanding, including the application of Quality by Design (QbD) principles.

However, the challenge is that in the cellular therapy field, because of the small numbers and variability in patient-derived starting materials, product and process knowledge emerges only as clinical experience grows, which makes it difficult to plan into the prospective pivotal trial. In the case of cell therapies, an "adaptive" approach incorporating evolving product and process understanding is needed. Having to restart regulatory processes for each potential manufacturing adaptation is not feasible and has the unintended consequence of discouraging process improvements that could benefit patients.

[12] Fraietta JA, Lacey SF, Orlando EJ, et al. Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. Nature Medicine 24, 563-571 (2018).

1.5.2 Using Post Approval Lifecycle Management (PALM)-like plan for making manufacturing and testing changes

As we gain stronger product knowledge and process understanding and are able to correlate their impacts to clinical safety, efficacy, and durability results, the insights gained are likely to lead to improvements that can be made to the manufacturing process and/or quality control tests. For example, based on data gained during clinical development, a process adaptation (e.g., culture medium optimization, culture condition optimization) is identified, which modestly increases the efficacy or reduces adverse events (i.e., does not impact labeled dose). The magnitude of change in clinical profile may not be large enough to justify a full clinical development but is still beneficial to patients. For these changes, modifications could be managed via a pre-negotiated plan with health authorities (e.g., Post-Approval Lifecycle Management or Comparability Protocol). The filing requirements for the change may include a combination of an analytical comparability assessment, and/or a small clinical study, analogous to a bioequivalence study for a new process. A post-market commitment could be considered to demonstrate/confirm the efficacy of the new process.

1.5.3 Create CMC commercial process change reporting categories for cell-based therapies

FDA issued a draft guidance in December 2017 for CMC changes to an approved application intended to assist manufacturers of biological products in assessing the reporting category for CMC changes. This guidance provides a starting framework that can be further extended to T-cell-based therapies. As the cell-based therapeutic industry accumulates commercial manufacturing experience, sponsors can identify the most frequent manufacturing changes and propose recommended reporting categories based on risk assessment: Annual Reportable (AR), Changes Being Effected (CBE)-0, CBE-30, or Post Approval Supplement (PAS). Consistent with the fundamental guiding principle from the biologics guideline, the reporting category selected should be commensurate with the risk of an unintended outcome resulting from changes involving these elements. When assessing the impact of change on product quality, the historical product and process knowledge including experience gained during commercial manufacturing should be fully leveraged. Developing a best practice guide for cell therapy with specific examples of process and testing changes for the range of categorization would be a beneficial activity to be created by an industrial consortium.

However, it should be noted that the overall variability in cell-based therapy processes is predominantly influenced by the incoming patient-to-patient variability. Therefore, the traditional process performance qualification (PPQ) approach utilizing three healthy donor batches to qualify each change has limited applicability and instead a rigorous, continuous process verification (CPV) plays a larger role in demonstrating process control. Use of healthy donors to characterize process and analytical variability in theory is a good approach, but a significant number of healthy donors are potentially needed to quantify the variability contribution of the process and analytics. This consumes resources and manufacturing capacity that otherwise would be used to produce clinical or

commercial products. Hence, a concurrent qualification approach, where a change is introduced in manufacturing based on small scale data and is subject to verification through a CPV program during clinical/commercial use, is not only more efficient but would also allow the confirmation of change in the setting of real patients instead of healthy donors. In addition, standalone qualification of the specific process or manufacturing change without the need for end-to-end full PPQ may be sufficient in some cases (e.g., a change in a supplier of raw materials, reagents, and solvents that have a minimal potential to affect product quality) provided that the materials' specific use, physicochemical properties, impurity content, and acceptance criteria remain comparable could be validated offline and reported as an AR. Additionally, a change from a manual operation to an automated operation that does not change the process parameter set points could be addressed through automation qualification and reported as AR.

Lastly, in some cases demonstrating analytical comparability at the appropriate in-process intermediate level may be sufficient. For example, demonstration of comparability for the vector bulk material due to a process change in the vector manufacturing process should not require demonstration of final product comparability post-transduction. Analytical comparability of the bulk viral vector and, if needed, use of small-scale model to confirm transducibility of the cellular in-process product should be considered sufficient. The life cycle plan for process and method changes needs to be carefully sequenced so that potential impact of the changes is seen throughout the CPV program. Changes to process parameters outside of previously validated ranges should be assessed with respect to criticality to process performance and product quality.

Several other examples of post-approval changes are likely. The reporting categories and extent of requalification for these changes will be assessed keeping the above considerations in mind. A risk based approach to determine the extent and approach of qualification should be used which would determine if 1) qualification can be performed using small scale or whether full scale confirmation is needed; 2) qualification exercise can be limited to evaluating product attributes of the impacted intermediate or the final drug product; and 3) separate qualification is needed or if heightened CPV program for a period of time can be used. Given that many cell therapy companies are focused on early access to the promising therapies, several process improvements are deferred and become part of the post-approval life cycle plan. Examples of such deferred changes include: new primary packaging components for the final product to simplify ease of administration and enable more clinical sites; new activation reagents; introduction of a new media processing system to improve manufacturing robustness; a higher-grade of fetal bovine serum (FBS) to improve reliability; change of buffer manufacturer from in-house to an external manufacturer; automation of manual processing steps; automation of flow cytometry data analysis; increase in vector production scale to meet increasing demand; change to a rapid sterility method, rapid microbiologic testing, and change of vector manufacturing process to a suspension cell culture process; the addition of an identical manufacturing suite to double capacity for both vector and drug product; change in the antibiotic resistance in the vector cell bank/plasmid; improved potency method; and change to stability data for expiry extension.

1.5.4 Quality standards for ancillary materials used in the manufacturing of cell-cased therapy products intended to be developed as commercial products

Currently, sponsor companies are restrained by the limited numbers of GMP producers of these ancillary materials (e.g., recombinant proteins, growth factors, cytokines, and small molecules) because of the regulatory requirements associated with choosing novel reagents. For the fore-seeable future, the supply chain will be a critical path for product commercialization. The root cause for this supply chain is multi-factorial, but some modifications of applicable regulatory guidance could accelerate innovation.

In addition, stakeholders desire more uniform feedback from individual reviewers around qual-ity and testing standards for non-GMP ancillary materials. Stronger guidance on how to stratify quality and/or characterization requirements based on whether they are excipients, product contacting (primary) or secondary ancillary material (e.g., plasmids used in viral vector manufac-turing) or tertiary ancillary materials would be beneficial to the field. Moreover, greater health authority alignment with the principles published in USP <1043> or other guidance documents could result in greater consistency in CMC development across multiple phases.

1.5.5 Other regulatory opportunities to support cell-based therapies

The use of medical devices in the manufacturing of cell-based therapies: In the current gener-ation of engineered T-cell products, approved medical devices are used in the manufacturing of cell-based therapy products. These medical devices are sometimes used outside of their approved "intended use," and equipment validation is done by the biotechnology manufactur-ing sponsor. This usage outside of the approved "intended use" causes tension with the device manufacturer as they don't want to put their medical device license at risk due to a biotechnol-ogy application.

Regulatory guidance for new cell therapy digital platform: The digital platform is a unique and critical aspect of cell therapy manufacturing, and various components such as Chain of Identity (COI) must be described in the BLA. It will behoove the field to develop regulatory guidance akin to regulating the manufacturing facility where it would be inspectable at any time but operational changes under controlled procedures are allowed.

Additional unique cell therapy regulations – setting lot specific specifications: Adapting a mid- or late- stage trial to incorporate multiple products to patient subsets would improve the pace of development for patient-specific therapies. In the case of cell-based therapies, the ability to engineer change into the cell provides for innumerable therapeutic opportunities and the ability to overcome challenges. If a change to product attributes is identified as an important factor while in P2 or P3 development, that change could be made and reset to a "child" IND to quickly gain groundwork experience to advance to later development.

PART 2: DRIVING INNOVATION IN CELL AND GENE THERAPY FOR THE TREATMENT OF CANCER THROUGH RESEARCH COLLABORATIONS & DATA SHARING

2.1 A scientific development consortium comprised of academic, government, nonprofit, and industry could share fundamental data and/or expedite investigational product development and testing processes, in early stage development and characterization, to advance the cell and gene therapy field for cancer patients.

The lack of available patient and product data necessary for effective data mining to inform manufacturing and clinical trial design is a major impediment to the advancement of cell and gene therapy for treatment of cancer. Pooling of data is currently limited because data sets and product characteristics need to be standardized in order to enable cross-study comparisons and data analysis. The competitive nature of development and the need to protect commercial, confidential, and proprietary information further complicate entities' ability to pool data and hinder opportunities for prospective data harmonization efforts. To move the cell and gene therapy field forward in immune-oncology, efforts are needed to define taxonomy and standardize data collection and measurement processes for analysis while exploring the potential for data sharing through pre-competitive collaborative groups. The establishment of a multi-stakeholder group of experts to serve as an ad hoc consult group to consortia participants (academic, government, nonprofit, and commercial) would potentially facilitate the development, review, and implementation of standard processes within individual development programs (e.g., review interim manufacturing and clinical data and approve/advise on subsequent modifications) based upon existing datasets and findings. The consult group could refer to previous efforts as potential models, such as the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline on Genomic Sampling and Management of Genomic Data (ICH E18)[13], for the development of guidelines that facilitate harmonization of cell and gene therapy studies. Also, consortia participants could benefit from and be incentivized by having access to more real-time advice from technical experts, including FDA, in early stage development in exchange for implementing agreed upon processes for documentation and information sharing with other consortia members as appropriate. Additional topics that will need consideration include the merits of a single consortium vs multiple consortia linked by a common data structure that would enable cross-study analyses and what broad functions the consortium would be optimally positioned to perform on behalf of consortia members.

Collaborations that promote and facilitate prospective data collection using common data elements and controlled vocabularies to enable cross-study analyses are essential to significantly advance development of cell and gene therapies in oncology. Occurring well before commercialization, such collaborations would provide a proof-of-concept for generating standardized data to inform the early stages of investigational product development. The establishment of a common study platform would foster collaboration across multiple approaches with consistent design, standardized data collection, and analysis. For example, the Parker Institute for Cancer Immunotherapy (PICI) has pro-

---

[13] International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use. (2015). Guideline on Genomic Sampling and Management of Genomic Data (E18). Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E18_Step2.pdf

posed a pioneering exploratory adaptive platform study to evaluate the safety and efficacy of multiple clinical hypotheses and mechanistically-defined cell and gene therapies/combinations. The platform study would be designed to investigate one cancer indication **(Figure 1)** and/or one set of targets **(Figure 2)** with the collective input from study primary investigators, consortia members (academic centers and industry), PICI, and the FDA. It would consist of a core protocol where the shared study design is described, with several appendices (cohorts) elaborating on cohort-specific designs included, and would feature:

1) sharing of data analyses that could address common clinical, manufacturing, data, and regulatory issues; 2) prospectively planned modifications to one or more aspects of the investigational product based on accumulating data from participants in the trial; and 3) efficient implementation of changes based on clinical data after assessing the data by independent consortia and discussing with FDA.

Figure 1: Schema of a platform trial with a single target
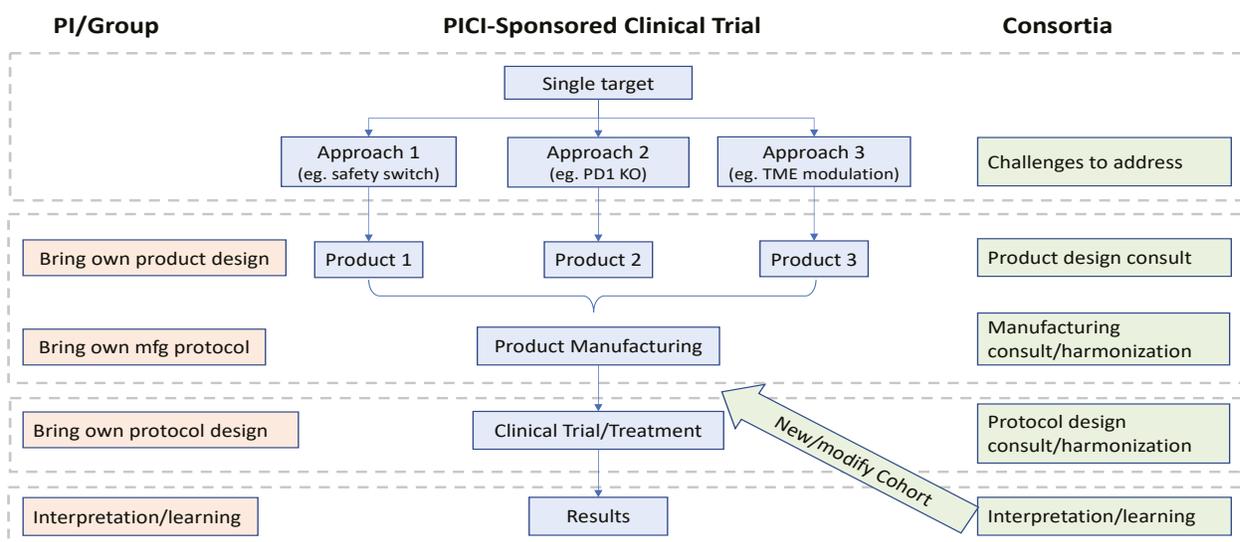
Schema of a platform study that is investigating one specific target but histology or modality agnostic. Each cohort will be independent, and products can come from different organizations. This design offers some standardization across different cohorts such as eligibility criteria, dose limiting criteria definition, and go/no go decisions. Emerging data will only be accessible to the organization that owns the product and to the sponsor of the study. With the permission of the sponsor, data that could inform future cell and gene therapy development will be shared with a group of experts who can make either general recommendations to inform the field, a communication to the FDA, or specific product recommendations. Recommendations could also be utilized for further optimization of the product and its development process.

Figure 2: Schema of a platform study with a single indication
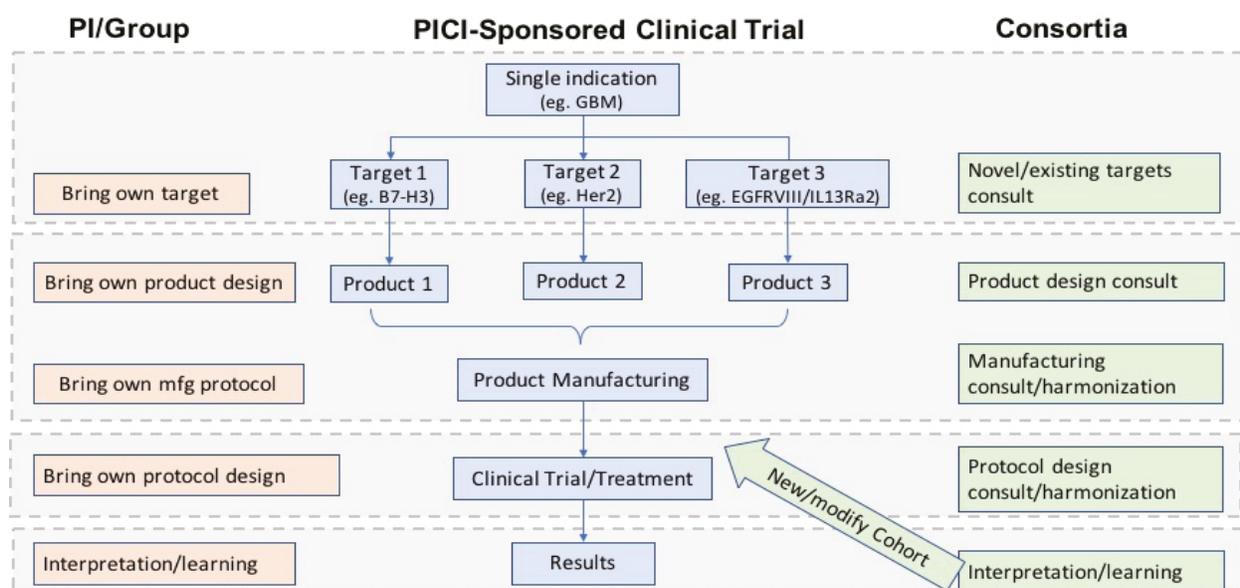
Schema of a platform study that is investigating one indication but allowing different products and modalities. The design and objectives are similar to the single target platform study described in Figure 1.

A key question in the field is: what are the features that characterize a safe, efficacious, and durable product? Therefore, as part of this platform, it will be important to establish harmonized strategies for collections and molecular profiling of the cells both before and after infusion. The variety of therapeutic approaches and indications that will be tested in a platform study provides tremendous opportunities to identify features of both the product and the manufacturing process, which lead to efficacious and safe therapies across a variety of contexts. These foundational learnings could be shared in a pre-competitive manner across consortium participants in order to accelerate the development of future therapies.

Another important initiative that is underway is a federally mandated and funded Regenerative Medicine Innovation Project (RMIP) established by the 21st Century Cures Act (Act).[14] The Act authorizes the appropriation of specific funds to NIH "for clinical research to further the field of regenerative medicine using adult stem cells, including autologous cells." Importantly, the Act requires that award recipients match, using non-federal contributions, in an amount at least equal to the federal award, which amplifies the federal investment and promotes collaboration across the public and private sectors. Moreover, the provision in the Act for the RMIP serves as a timely stimulus for NIH to work with NIST, FDA, DoD, and other partners in order to galvanize the field of cellular therapy in regenerative medicine (RM), foster major clinical advances, address key regulatory and technical issues in product development and clinical investigation, and ensure that RM clinical studies utilizing cell-based therapies are standardized, reproducible,

[14] 21st Century Cures Act, Pub. .L. No. 114-255, § 1001, 130 Stat. 1041 (2016).

and generalizable. To support the RMIP, NIH is establishing a Regenerative Medicine Innovation Catalyst (RMIC), as outlined in **Figure 3**, that will provide critical services to support RMIP clinical research, including development of common data elements for stem cell products and clinical outcomes, clinical data standards to facilitate preparation and commercialization of clinical grade stem cell products, and regulatory support. RMIC partners will be expected to make pre-competitive RM data and analyses publicly accessible to the broad biomedical community. Furthermore, the RMIC will perform prospective in-depth and

Figure 3: Framework for the NIH Regenerative Medicine Innovation Catalyst to facilitate clinical research and further the field of regenerative medicine
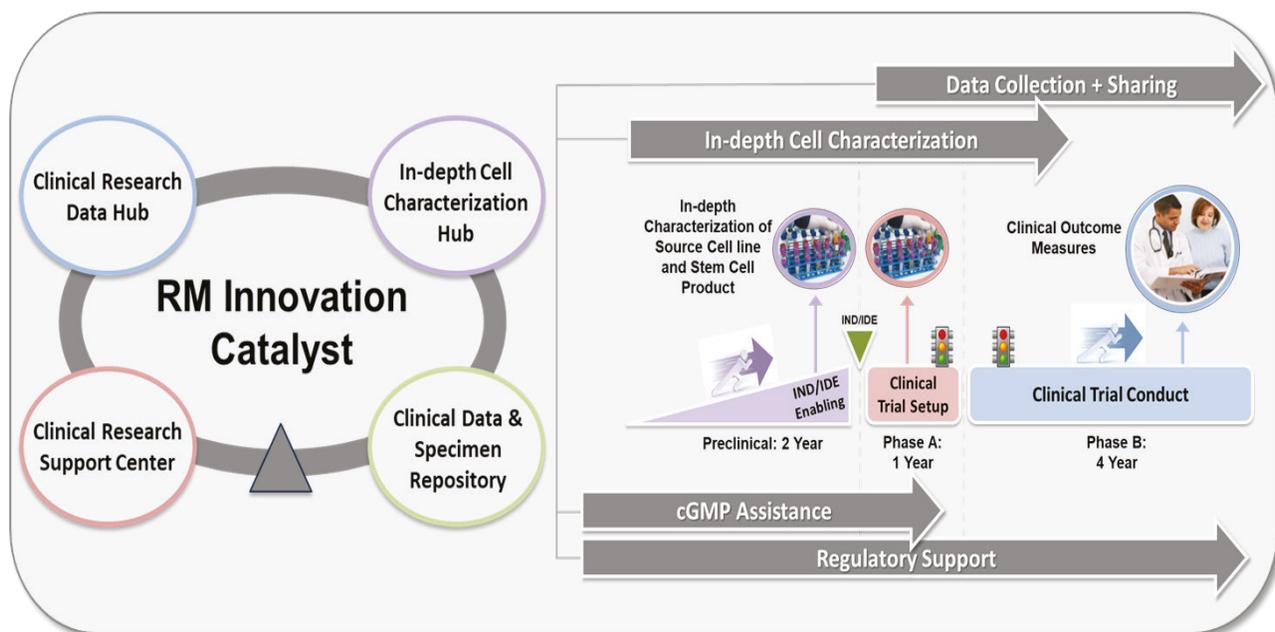

Figure 3 depicts the four major components of the Regenerative Medicine Innovation Catalyst (RMIC) and outlines the services and functions of the Catalyst throughout the RM pre-clinical development and clinical trial lifecycle. The RMIC consists of: (1) the Clinical Research Support Center, which will provide assistance in cGMP or phase-appropriate cell product manufacturing and regulatory support; (2) the In-depth Cell Characterization Hub which will coordinate the state-of-the art characterization of source stem cells as well as final clinical grade product and participate in development of common data elements describing cell products; (3) the Clinical Research and Data Standards Hub will develop, test, and implement common data elements for RM clinical research to enable cross-study analyses; and (4) the Clinical Data and Specimen Repository will provide both a controlled access database as well as a biorepository. The database will provide harmonized cell product data and clinical safety and efficacy data to facilitate correlation of cell characteristics with clinical outcomes. The biorepository will provide samples of source stem cell and cell products as well as a clinical biospecimens for subsequent analyses.

independent characterization of representative samples of source stem cells as well as final clinical-grade product and coordinate the storage and sharing of cell product characterization data linked to individual participant level outcomes data using cloud-based systems to help facilitate downstream correlation of key cell attributes to clinical safety and efficacy data. The RMIC is a pilot approach to providing critical support and data to the field of Regenerative Medicine, which, if successful, may be extended to all future NIH-sponsored RM clinical research. This new approach has the potential to address the major challenges for developing personalized cell-based therapies for cancer and many other diseases.

2.2 Establish agreed upon standard technologies (e.g., analytics for vectors, cell culture processes, potency assays for cells, simple manufacturing controls, and basic quality attributes) to facilitate technology transfer between academic innovators and industry GMP producers of these investigational therapies.

Difficulties with technology transfer from small academic institutional studies to larger, pharmaceutical company-sponsored trials are associated with an inability to expand trials beyond initial Phase 1 studies. Standard technologies are needed to understand the difficulty of the technology transfer process and guide design of smaller scale processes to enable replication and expansion to larger scale processes for further development by a commercial partner. The agreement upon a set of parameters for use by academic investigators that could enable rapid technology transfer would be mutually beneficial by adding value to the field for this

Box 2. Proposal to Facilitate Technology Transfer from Academic to Clinical Scale Industrial Process

STEP 1: Define and transfer the as-is process.

STEP 2: If starting with a lab scale academic process, the first step should be to mimic the scale production of the lab that developed the product and/or conducted the phase I study.

STEP 3: Develop the full-scale, clinical/commercial process – in a step-wise, operation by operation fashion if necessary.

SUCCESS FACTORS IN THE THREE STEP TECH TRANSFER MODEL
1. Establish Quality Attributes early in the tech transfer and use common analytical platforms to assess suitability across all stages.
2. Introduce and qualify GMP grade materials as early in the process as possible.
3. Careful consideration of plasmid and vector sourcing and manufacturing is needed at each stage. Final engineering runs should include clinical grade vector, if possible.
4. Conduct post-transfer proficiency testing to validate process and product controls.

therapeutic approach. Academics would have an asset with a more robust data package to help determine developability and risk/probability of success and companies would have an investigational product with a standardized data package and would be able to leverage a broader data set for evaluation of a specific program for developability. Further, it would enable leveraging of prior knowledge especially when using platform processes (e.g., same plasmid or vector with a different transgene). One way this could work would be for different industry producers of these therapies to agree upon non-proprietary common features that could subsequently be transitioned into their proprietary systems. These common features could then be provided to the academic innovators in the form of a toolkit or could even inform guidance around early stage clinical programs and a list of the Key Quality attributes that can/cannot be changed at a predetermined point during the Process Development Steps.

Several recommendations were identified to address key opportunities and help guide initial priorities for consortium-led efforts:

- Efforts should be undertaken to define taxonomy and standardize data collection and measurement processes for analysis.
- Pre-competitive collaborative groups should be formed to facilitate data sharing and include a multi-stakeholder group of experts to serve as an ad hoc consult group to consortia participants to facilitate the development, review, and implementation of standard processes within individual development programs.
- Non-profit clinical research organizations, as neutral and unbiased organizations, can play an integral role in harmonizing clinical trials and translational research. A platform study can offer commonality and opportunity for information sharing. This can lead to less redundancy and subjecting less patients to unnecessary risks.
- Collaborations that promote and facilitate prospective data collection using common data elements and controlled vocabularies should be formed to enable cross-study analyses.
- Deep molecular characterization of the cellular product will be key to identifying features of safe, efficacious, and durable therapies. Standardization of assays and collection strategies will provide opportunities to integrate data across a broad variety of indications and therapeutic strategies.
- Standard technologies should be developed to guide design of smaller scale processes to enable replication and expansion to larger scale processes for further development by a commercial partner.

CONCLUSIONS AND NEXT STEPS

This whitepaper outlines several opportunities and strategies to expedite T-cell based therapies into first in human studies, and to ensure that T-cell-based therapeutics are impactful for the greatest number of patients by creating a more "adaptive" manufacturing process that would allow the adoption of new manufacturing technologies as more patients are treated and more clinical, translational, and product quality data is collected during a product lifecycle. Moreover, efforts to encourage transparency, collaboration, and data sharing are needed so changes can be appropriately monitored and would allow the field to adapt to improvements efficiently. The proposals outlined in this whitepaper could be particularly useful in bringing cutting edge biological and genetic approaches forward to enhance the current generation of cell therapies in the highly complicated tumor microenvironment. This whitepaper is intended to provide high-level ideas to accelerate early cell therapies into clinical trials.

To fully consider and implement the proposals and strategies outlined in this whitepaper, key stakeholders will need to be called upon to continue the dialogue that has been initiated with this whitepaper. Formation of pre-competitive consortiums to standardize technologies, and the implementation of integrated platform studies would also help enable efficient development and collection of common data elements across trials. Additional areas, such as pre-clinical and clinical testing and the development of clinically relevant biomarkers to guide selection of the right patient population and detection of proof-of-concept in the clinical study, will require additional discussions and proposals to be considered.

APPENDIX 1: TABULAR SUMMARY OF EFFICIENCIES GAINED THROUGH EARLY STAGE MANUFACTURING AND IND FLEXIBILITY FOR T-CELL THERAPY EXPLORATORY CLINICAL TRIALS

a) Alternative manufacturing paradigm for early stage, exploratory trials

The potential time and cost savings for alternative approaches to use of R&D reagents, plasmid DNA, LVV manufacturing, and engineer run activities are outlined below.

CMC Activity	Typical Time[+] Investment	Areas of Proposed Flexibility	Potential Time[+] Savings	Potential Cost Savings
Use of R&D Reagents	3-6 months	Increasing options for use of R&D reagents and reducing cost and time to either enable or negotiate GMP manufacture of reagents	1-3 months	$ to $$$
Plasmid Manufacturing	4 months (+ 3 to 6 months in queue)	Reduced plasmid characterization & infrastructure requirements	5-7 months	$$
Viral Manufacturing	6 months (+ 9 to 12 months in queue)	Waive RCL testing in lieu of surrogate testing; reduced cGMP requirements for ancillary reagents	4 months	$
Cell Product Engineering Runs (3 runs)	3 months	Use representative pilot virus for parallel cell product engineering runs	2 months*	N/A

+ All time estimates are approximate
* There is some overlap in the time savings between the shortened LVV manufacturing timelines, and the engineering runs utilizing pilot materials. Overall, the ability to demonstrate process control using representative materials means that activities are not reliant upon manufacturing and release of LVV

Figure 1: Alternative Manufacturing Paradigm for Early Iterative Clinical Studies
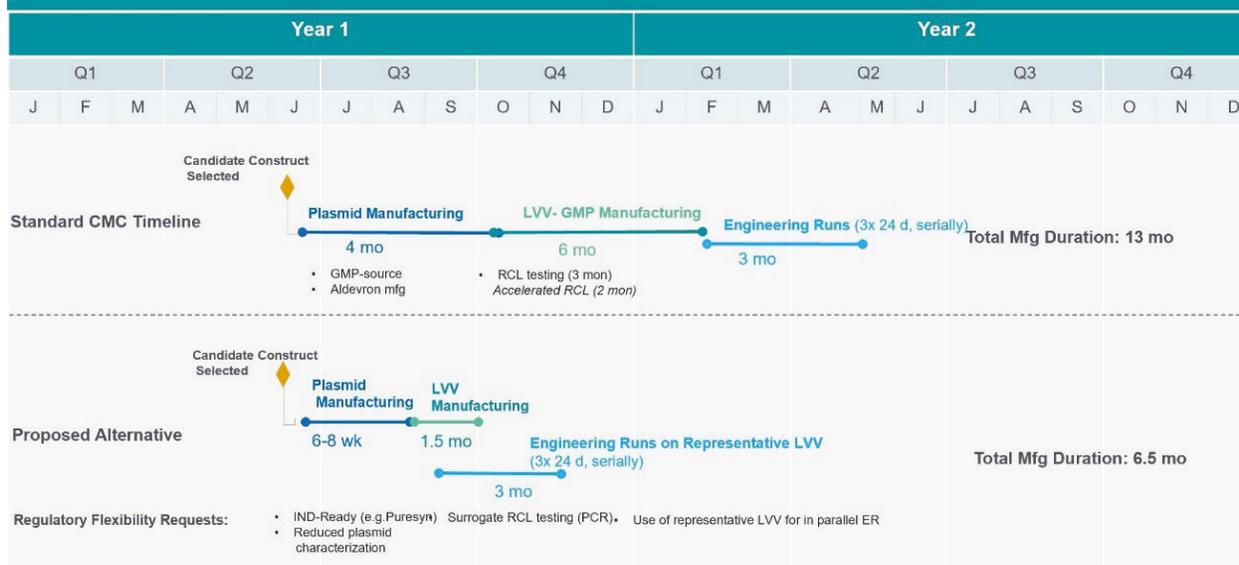
Expedited manufacturing of plasmid DNA and viral vectors coupled with cell product engineering run activities using representative viral vector could save time in getting into early phase clinical studies

b) "Parent-Child" IND paradigm for early stage, exploratory trials

Traditional development requires the submission of an IND for every product development candidate prior to the conduct of clinical trials. While the costs and time required to produce an IND vary significantly between sponsor types and experience, a reasonable estimate of the time and cost per IND is approximately 3-6 months of cross-functional document drafting and preparation and approximately $100,000 in medical writing and regulatory operational costs for the initial IND and approximately $25,000 per year in maintenance costs for the life of the IND. These time and cost estimates become prohibitive when a sponsor wishes to test several constructs or manufacturing process alterations.

A "parent-child" IND paradigm could result in significant savings in time and cost; the savings would increase with time and the number of constructs tested. An example table of contents of a "parent-child" IND is provided on the following page.

IND Module	Parent IND Contents	Required Section in Child IND	Child IND #2	Child IND #3
Module 1	1.1 Forms 1571 and 3674	Forms 1571 and 3674	Forms 1571 and 3674	Forms 1571 and 3674
	1.2 Cover letter			
	1.3 Transfer of obligations	–		
	1.4 References Letters of authorization Right of reference (to DMFs etc.)	Letter of cross reference to parent IND	Letter of cross reference to parent IND	Letter of cross reference to parent IND
	1.6 Meeting package			
	1.12 Environmental exclusion			
	1.14 Investigational brochure*	1.14 Investigational drug label for candidate #1	1.14 Investigational drug label for candidate #2	1.14 Investigational drug label for candidate #3
Module 2	2.2 CTD Introduction			
	2.3 Intro to Quality overall summary – all candidates			
	2.4 Nonclinical overview – all candidates			
	2.5 Clinical overview – target disease, population and common aspects of all candidates			
		2.6 Nonclinical summary – written and tabulated summary of nonclinical investigation of candidate #1	2.6 Nonclinical summary – written and tabulated summary of nonclinical investigation of candidate #2	2.6 Nonclinical summary – written and tabulated summary of nonclinical investigation of candidate #3
	2.7 Clinical pharmacology – all candidates	2.7 Aspects of clinical pharmacology specific to candidate #1	2.7 Aspects of clinical pharmacology specific to candidate #2	2.7 Aspects of clinical pharmacology specific to candidate #3
Module 3		3.2 Quality – specific to candidate / process #1	3.2 Quality – specific to candidate / process #2	3.2 Quality – specific to candidate / process #3
Module 4		4.2 Nonclinical study reports – specific to candidate #1	4.2 Nonclinical study reports – specific to candidate #2	4.2 Nonclinical study reports – specific to candidate #3
Module 5	5.0. Clinical study reports – Not applicable			