Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Real-world enrollment for a prospective clinico-genomic database using a pragmatic technology-enabled platform

Alexia Exarchos [a,*], Ariel B. Bourla [b], Maneet Kaur [b], Katja Schulze [a], Sophia Maund [a], Yi Cao [a], Yihua Zhao [b], Elizabeth H. Williams [c], Sarah C. Gaffey [c], Richard Zuniga [d], Shaily Lakhanpal [e], Vladan Antic [f], Michelle Doral [a], Johanna Sy [a], Neal J. Meropol [b], Anne C. Chiang [g]

[a] *Genentech, Inc., South San Francisco, CA, USA*
[b] *Flatiron Health, Inc., New York, NY, USA*
[c] *Foundation Medicine, Inc., Cambridge, MA, USA*
[d] *New York Cancer & Blood Specialists, New York, NY, USA*
[e] *Alabama Oncology, Birmingham, AL, USA*
[f] *F. Hoffmann-La Roche Ltd, Basel, Switzerland*
[g] *Yale School of Medicine, New Haven, CT, USA*

ARTICLE INFO

Keywords:
Prospective study
Circulating tumor DNA
Feasibility study
Electronic health records
Lung neoplasms

ABSTRACT

Background: Discovery and incorporation of predictive and prognostic biomarkers enhance outcomes for patients with cancer. Clinico-genomic datasets, which retrospectively link real-world clinical data to tumor sequencing data, are important resources for biomarker research, which has historically relied on robust research infrastructures exclusive to large academic centers. The objective was to evaluate the feasibility of a pragmatic, technology-enabled platform at community-based research sites for development of a prospective clinico-genomic database supported by centralized electronic health record (EHR)–based patient ascertainment and data processing.

Methods: Adults with stage IV or recurrent metastatic non-small cell lung cancer or extensive-stage small-cell lung cancer were enrolled at 23 US sites upon initiating a standard line of therapy. Enrollment rates were estimated from eligible populations at individual centers. Clinical data from routinely collected EHR documentation were centrally processed and normalized for quality control. Serial blood samples at pre-specified timepoints (baseline, during treatment and at disease progression/end of therapy) were used for circulating tumor DNA (ctDNA) genomic profiling.

Results: Between December 2019 and May 2021, 944 patients enrolled, representing ≈25 % of eligible patients. Eight-hundred seventeen of 944 (87 %), 406 of 606 (67 %) and 398 of 852 (47 %) participants provided qualifying samples for ctDNA testing at baseline, during treatment and at disease progression/end of therapy, respectively. Samples were provided at all three timepoints by 35 % of participants.

Conclusion: A community-based oncology patient cohort was rapidly enrolled, creating a real-world clinico-genomic dataset. This pragmatic study platform has potential research applications where prospective real-world data may contribute to evidence generation.

1. Introduction

The standard treatment approach for patients with non-small cell lung cancer (NSCLC) currently uses diagnostic testing for the presence of driver mutations and expression of PD-L1. This paradigm, enabled by development of treatments with high specificity for their cellular targets over the past two decades, has dramatically improved survival of patients with NSCLC and other malignancies [1–3]. These advances are enabled by biobanks with tissue and blood samples from patients paired with longitudinal clinical annotation. Such resources are critical for the discovery of new therapeutic targets and mechanisms of resistance, as well as the development of novel tools for risk stratification and monitoring of disease burden [4].

Prospective clinico-genomic datasets can supplement the clinical and biological data available from routine care with additional, intentionally

* Corresponding author. Genentech, Inc. 1 DNA Way South San Francisco, CA, 94080, USA.
E-mail address: exarchos.alexia@gene.com (A. Exarchos).

https://doi.org/10.1016/j.conctc.2025.101446
Received 17 July 2024; Received in revised form 15 October 2024; Accepted 1 February 2025
Available online 4 February 2025

Abbreviations:

AJCC	American Joint Committee on Cancer
CGDB	clinico-genomic database
CGP	comprehensive genomic profiling
ctDNA	circulating tumor DNA
ECOG	Eastern Cooperative Oncology Group
EHR	electronic health records
ES-SCLC	extensive-stage small cell lung cancer
FH	Flatiron Health
FHRD	Flatiron Health Research Database
FHRN	Flatiron Health Research Network
FMI	Foundation Medicine, Inc.
mNSCLC	metastatic non-small cell lung cancer
NSCLC	non-small cell lung cancer
NOS	not otherwise specified
PCG	Prospective Clinico-Genomic
SCLC	small cell lung cancer
SES	socioeconomic status
SOC	standard of care
TF	tumor fraction

collected information necessary for research advancements. Such prospective databases have linked clinical and genomic data, with data being derived from clinical trials, single-institution studies or national registries [4–8]. Complex clinical research infrastructures, historically centralized in large academic medical and research centers, have been required for patient identification, enrollment, data collection and follow-up. For instance, the American Association for Cancer Research's Project GENIE has developed a longitudinal, clinico-genomic repository of pan-cancer sequencing and clinical data aggregated across 18 participating large academic centers [4,6]. The Project GENIE protocol consists of targeted sequencing applied to solid tumor samples and narrow clinical data observations (e.g., cancer type, primary or metastatic tumor, age, sex, race) that rely on manual inputs [4]. Some studies have used data collected via Project GENIE to interrogate gene signatures across patients with specific cancers, but such federated approaches to data analysis may be limited by missingness and lags in recency of clinical data, incomplete treatment information and non-uniform timing of tumor sequencing [7].

Moreover, patient populations at participating academic centers often lack heterogeneity in age, race or ethnicity and socioeconomic status that would be characteristic of the general population. It is notable that while less than 10 % of US patients with cancer enroll in clinical trials, those who do generally are not representative of the broader population [9–11], and up to 85 % of patients with cancer receive care in community-based settings [4]. Thus, establishing an infrastructure to support access to research participation where most patients receive their care could accelerate evidence generation and improve the generalizability of research results.

Therefore, we sought to leverage the Flatiron Health Research Network (FHRN), sites related through electronic health record (EHR) integrations, to develop a platform for a community-based clinico-genomic study in which patient enrollment, sample collection and follow-up were seamlessly integrated into routine clinical workflows. We hypothesized that high-quality curation of information routinely collected during clinical care and entered into EHRs could be leveraged in the context of a prospective clinico-genomic study, since nearly all of the clinical information required was routinely collected during clinical care. Additionally, we posited that centralized data collection among a network of clinical practices would reduce the burden of participation for individual sites, along with limiting site activities to obtaining patient consent and collecting biospecimens (e.g., circulating tumor DNA [ctDNA]) and minimizing the need for activities that fall outside of routine care (i.e., "secondary" or "intentional" data collection for clinical research).

Flatiron Health and Foundation Medicine, Inc. (FMI) previously established the feasibility of generating a real-world clinico-genomic database by retrospectively linking longitudinal clinical data from the Flatiron Health real-world evidence database to the Foundation Medicine database of tumor sequencing results [8]. Building on this capability, the current study was undertaken to explore the feasibility of conducting a pragmatically designed, prospective, technology-enabled study in community oncology practices and to provide a proof of concept regarding the value of community-based biomarker research using the resulting clinico-genomic database. Specifically, this Prospective Clinico-Genomic (PCG) study (NCT04180176) was initiated to evaluate integration of a prospective data collection protocol into routine clinical practice to enable clinical and genomic insights into patients with metastatic NSCLC (mNSCLC) or extensive-stage SCLC (ES-SCLC) treated in primarily community settings.

2. Methods

2.1. Patients

This pragmatically designed, multicenter, low-interventional study enrolled adults (age ≥18 years) with a documented diagnosis of mNSCLC (Stage IV or recurrent metastatic disease) or ES-SCLC (or the equivalent stage per the American Joint Committee on Cancer, Version 8) [12] who were planning to initiate standard-of-care (SOC) systemic anti-cancer treatment, were able to comply with the study protocol in the investigator's judgment and provided written informed consent. The only exclusion criterion was participation in an interventional trial with at least one investigational medicinal product at the time of informed consent. The study was designed to enroll an initial "all-comer" cohort of approximately 950 individuals, with a planned re-enrollment phase of up to 50 individuals. Patients were enrolled from 23 participating sites from the FHRN (22 community-based sites and 1 academic medical center), which includes approximately 280 US cancer clinics (≈800 sites of care). Individuals were approached for participation at in-person visits only; screening and enrollment occurred during an SOC clinic visit and could occur on the same day. The PCG data model leverages many of the data fields incorporated in the nationwide deidentified Flatiron Health database [13,14] and Flatiron Health-Foundation Medicine clinico-genomic database (FH-FMI CGDB) [8].

Eligible patients were identified by site study teams with support from a clinical trial software platform that automates patient ascertainment through matching inclusion criteria to EHR data (OncoTrials™) [15].

This study was conducted in conformance with the International Council for Harmonization E6 guideline for Good Clinical Practice, the principles of the Declaration of Helsinki and the Code of Federal Regulations on the Protection of Human Subjects. All patients provided informed consent.

2.2. Primary objective

The primary objective was to evaluate the feasibility of a scalable, prospective research platform, as demonstrated by the following: (1) proportion of potentially eligible patients who were enrolled into the all-comer cohort—overall and stratified by site—and (2) proportion of enrolled patients who submitted sufficient blood samples viable for ctDNA testing at the following timepoints: enrollment, during treatment (specifically at the time of the first tumor assessment) and at the time of disease progression or the end of therapy. Instances of disease progression were determined based on documentation in the EHR by the treating clinician.

2.3. Proportion of enrolled patients

The study enrollment rate was defined as the number of enrolled patients divided by the estimated number of eligible patients. The number of eligible patients was estimated using the Flatiron Health Research Database (FHRD). A subset of patients in the FHRD were sampled for clinical abstraction of variables at each participating site [16], including variables used to determine study eligibility. This patient sample was then used to estimate the number of eligible patients using a hierarchical Bayesian approach and the observed probabilities of patients meeting the cohort criteria (see **Supplemental Methods** and Fig. S1).

2.4. Proportion of patients with sufficient blood samples

All patients were eligible to provide samples at the time of enrollment. Samples provided at the first tumor assessment qualified as during-treatment samples if there was no disease progression or if patients continued the treatment despite disease progression. If disease progression was observed at the first tumor assessment and treatment was discontinued within 14 days, samples provided at that time were qualified as progression/end-of-therapy samples (and those patients would not have during-treatment samples). If there was no progression at the first tumor assessment, patients would be eligible to provide the final progression/end-of-therapy sample at the point of observed disease progression or discontinuation of the therapy initiated at enrollment or for 30 days after. Patients were only eligible for all 3 timepoints if there was no disease progression or treatment discontinuation before the first tumor assessment.

Blood samples were solely collected during routine care visits, which could fall under the qualified timing for each sample type based on the following:

- Qualifying enrollment samples were collected within 28 days prior to or on the day of initiating the new line of therapy
- Qualifying during-treatment samples were collected within 14 days before or after the first tumor assessment performed during the line of therapy
- Qualifying progression/end-of-therapy samples were collected within 30 days after the progression or end-of-therapy event

If for any reason a patient was lost to follow-up—defined as two failed attempts by the site to contact the patient—the potential samples at any remaining timepoints were considered missing. These definitions were used to generate two measures of sample completeness: the number of non-missing blood samples at each of the three timepoints divided by the number of patients eligible for sample collection, and the number of enrolled patients with blood samples at all three timepoints divided by the number of patients eligible for three samples.

2.5. Clinical outcome assessments

Patients were assessed for real-world mortality, real-world progression and real-world response, which were obtained with technology-enabled abstraction by trained professionals [17,18]. Flatiron Health leveraged multiple data sources to generate a composite real-world mortality variable with high sensitivity and specificity. Real-world progression was defined as death or clinician-documented disease progression or growth or worsening of the tumor. Real-world response was defined as a clinician's assessment of change in disease burden during a line of therapy at timepoints associated with radiologic assessments. The clinical outcomes were reported elsewhere [19,20].

2.6. Socioeconomic status

Using the Yost Index methodology [21], the area-level socioeconomic status (SES) was measured per block groups from 2015 to 2019 based on each participant's most recent address. SES index quintile scores were based on the Census Bureau's American Community Survey using 2010 Census boundaries [22,23]. SES could not be measured for patients whose address was not geocoded or was otherwise unavailable.

2.7. Data and sample collection

Data elements included study status, enrollment and consent details, mortality, enhanced diagnosis details (e.g., diagnosis dates, histology, stage, additional malignancies, smoking status), enhanced pantumor details, progression data, response to lines of therapy, oral medications, study line of therapy confirmation and discontinuation, comorbidities and sample collection details.

Data and samples collected as part of the PCG study—in addition to treatments and clinical or radiological assessments of disease burden per SOC and local practice—are outlined in Fig. 1. Clinical data for the study were collected centrally via EHRs through Flatiron Health's data platform [13,24,25]. EHR data were entered by the provider in a combination of structured (e.g., laboratory results) and unstructured (e.g., physician notes) formats as part of routine clinical documentation. Structured data underwent a centralized mapping and normalization process to standard reference terminologies, when available, from the EHR to the study database. Unstructured data were processed separately from the EHR through technology-enabled chart abstraction by trained personnel (e.g., oncology nurses or tumor registrars) according to specific abstraction policies and procedures that are tested and optimized for reliability and reproducibility (e.g., data entry checks, duplicate abstraction of samples and reviews of data for possible discrepancies). The processed, amalgamated data were then transferred to a central study database that supported routine data management activities and generation of analytic data. Key components of this process included centralized quality control of study data and continuous access to all source data and documents throughout the study. As part of the data management and review plan, Flatiron Health conducted periodic site monitoring; quality checks of source data and abstracted data; and resolution of discrepancies with data processing, sites or vendors.

Blood samples intended for ctDNA profiling, performed by Foundation Medicine, were collected at each of the timepoints specified above (at enrollment, during therapy and at progression or end of therapy). Blood analysis consisted of comprehensive genomic profiling (CGP) using the FoundationOne®Liquid or FoundationOne®Liquid CDx (F1CDx®) test. Archival or fresh tumor specimens that were collected independent of the PCG study and optionally provided at enrollment underwent CGP via F1CDx®. Tissue and blood samples were also retained for potential use in protocol-related biomarker research and assay development, leveraging other profiling methodologies such as whole exome sequencing, whole genome sequencing, immunohistochemistry or gene expression analysis.

2.8. Assessment of trial operations

To evaluate the operative success of the PCG enrollment protocol, trial metrics were collected along with those from 12 interventional cancer trials sponsored by Genentech, Inc., conducted at sites also participating in the PCG study [26–37]. Trial metrics included consent rate, time to site activation, time to first patient in and screening failure rate. Of course, many fundamental differences exist between interventional trials and the current PCG study; therefore, these metrics were used for qualitative discussion rather than quantitative comparison.

2.9. Statistical analyses

The all-comer cohort was used for analyses. Continuous variables were summarized using means, standard deviations, medians and ranges. Categorical variables were summarized using counts and

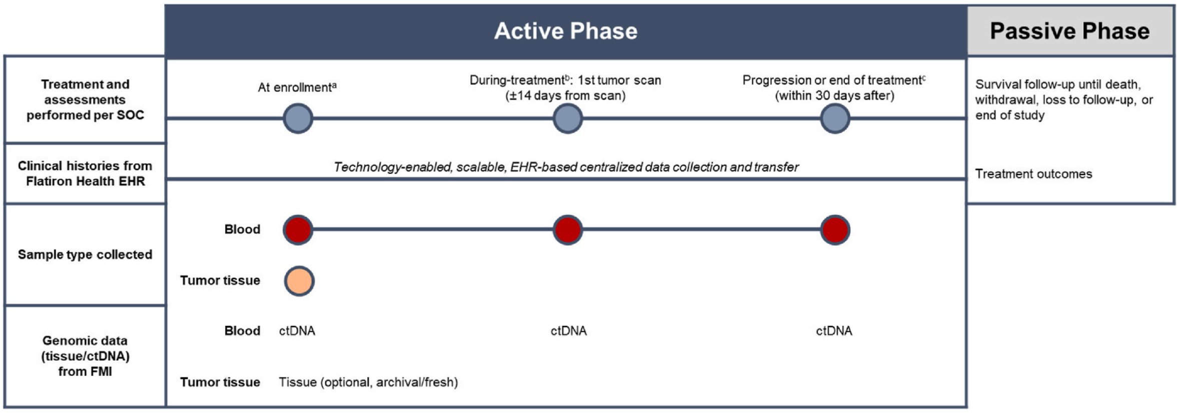
Active Phase
Passive Phase
Treatment and assessments performed per SOC
At enrollment^a
During-treatment^b: 1st tumor scan (±14 days from scan)
Progression or end of treatment^c (within 30 days after)
Survival follow-up until death, withdrawal, loss to follow-up, or end of study
Clinical histories from Flatiron Health EHR
Technology-enabled, scalable, EHR-based centralized data collection and transfer
Treatment outcomes
Sample type collected
Blood
Tumor tissue
Genomic data (tissue/ctDNA) from FMI
Blood
ctDNA
ctDNA
ctDNA
Tumor tissue
Tissue (optional, archival/fresh)

Fig. 1. PCG study overview.
ctDNA = circulating tumor DNA; EHR = electronic health record; FMI = Foundation Medicine, Inc.; SOC = standard of care.
[a]Enrollment sample could be obtained from Day −28 to Day 1 of the new line of therapy.
[b]Patients were only eligible to provide a during-treatment sample if there was no tumor progression at the 1st tumor scan, or if the line of therapy was continued for >14 days after the 1st scan despite progression.
[c]The visit where progression was determined or treatment was discontinued if it occurred >14 days after the 1st tumor scan.

proportions, as appropriate. Summaries were given overall and by each site. The generation of the analytic dataset and statistical analyses were conducted using R version 4.2 (R Foundation for Statistical Computing).

3. Results

3.1. Enrollment rates and proportions of eligible patients who were enrolled

Between December 13, 2019, and May 10, 2021, 944 patients were enrolled in the PCG study all-comer cohort across 23 study sites within the Flatiron network. Twenty patients were re-enrolled over the following 10 months. The overall enrollment rates relative to estimates of eligible patients for NSCLC, SCLC and the total population were 24.2 %, 19.3 % and 23.2 %, respectively (Table 1). Patient enrollment and attrition are illustrated in a flow diagram of the study (Fig. 2A), and the cumulative number of enrolled patients is shown in Fig. 2B. The enrollment rates were stable over the course of the study, including during the 2020 COVID-19 pandemic, and were comparable for NSCLC and SCLC (Fig. 2C). Across study sites, the estimated enrollment proportion ranged from <5 % to 50.9 % (Fig. 2D).

3.2. Patient characteristics

Patient demographics are shown in Table 2. The median age of all patients was 69 (interquartile range: 63–75) years, and 50.2 % were female. Of the study participants, 10.6 % were Black or African American, 1.0 % were Asian, 71.5 % were White, 9.2 % belonged to other races and the remaining 7.7 % had missing race data. All but seven patients were enrolled at community practices rather than academic medical centers. Of the overall population, 14.4 % were in the lowest quintile of area-level SES.

Table 1
Estimated enrollment rates.

Enrollment rate, % (95 % CI)	
NSCLC	24.2 (22.7–25.8)
SCLC	19.3 (16.2–22.7)
Total	23.2 (21.8–24.6)

NSCLC = non-small cell lung cancer; SCLC = small cell lung cancer.

3.3. Proportion of patients with sufficient blood samples

Completeness of sample collection for ctDNA profiling is shown in Fig. 2A and 3. Of the 944 patients enrolled in the primary study, 817 (86.5 %) had a qualifying enrollment sample (Fig. 2A and 3). Of 606 patients with during-treatment eligibility, 406 (67.0 %) had a qualifying sample, and of 852 patients with progression/end-of-therapy eligibility, 398 (46.7 %) had a qualifying sample. Ultimately, 565 patients were eligible for all three timepoints. Of those, 195 (34.5 %) had samples that qualified for each timepoint.

3.4. Assessment of trial operations

The current trial operations were measured against metrics derived from 12 interventional clinical studies to evaluate strengths of the PCG study platform (Table 3) [26–37]. The PCG study was initiated faster than interventional studies, with shorter median time to site activation (58 vs 130 days) and first patient in (8 vs 101 days). The PCG consent rate exceeded that for interventional trials (28 % vs 3 %), and the screening failure rate was lower (3 % vs 34 %).

4. Conclusion

This study demonstrated that it is feasible to execute a pragmatically designed, prospective, real-world, clinico-genomic study with data derived from a cohort predominantly enrolled from community oncology practices. The centralized clinical data collection facilitated patient participation by minimizing impact to routine care workflows and burden on site research staff. Approximately 25 % of eligible patients were accrued in the study, leading to 944 enrollees over 17 months.

Key strengths of this trial were the estimation of the eligible patient pool and evaluation of the overall accrual rate using EHR-derived FHRN data, as well as the availability of OncoTrials™ software at sites to assist with patient ascertainment. Patient data were centrally abstracted from routinely documented EHR clinical information, rather than requiring manual entry by sites into an electronic data capture system. Attainment of accrual goals in oncology clinical trials is an ongoing challenge, and 38 % of National Cancer Institute Cancer Therapy Evaluation Program–sponsored trials are discontinued based on failure to meet minimum accrual goals [38]. The pragmatic design and centralized operational activities of the PCG study platform could be applied for more efficient, lower-burden and lower-cost clinical trial recruitment in

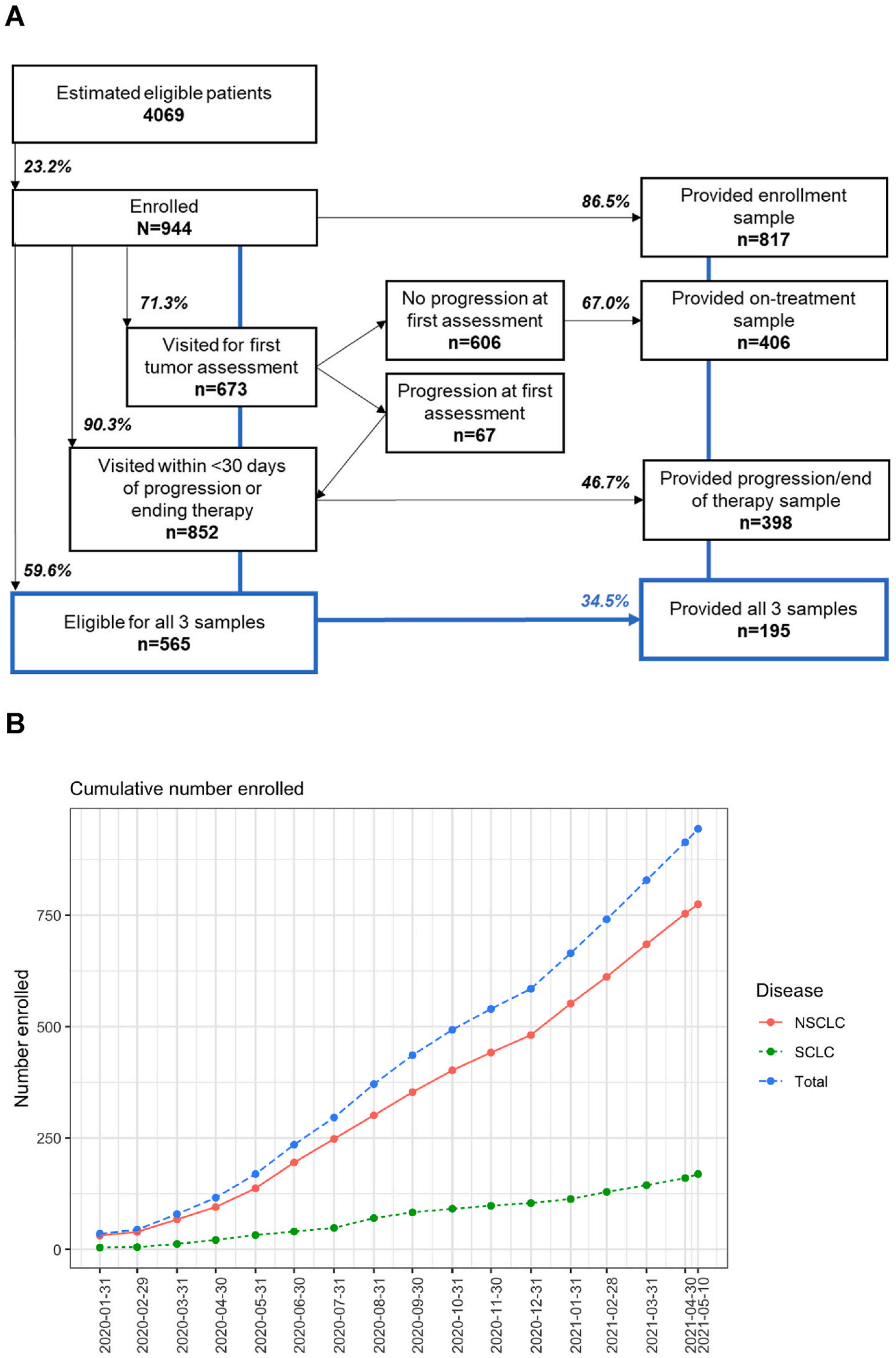

Fig. 2. (A) Patient enrollment and attrition, (B) cumulative number of patients enrolled over time, (C) continuous estimated enrollment rates, and (D) estimated total enrollment rates across study sites.[a]
NSCLC = non-small cell lung cancer; SCLC = small cell lung cancer.
[a]Sites are arranged in order of estimated enrollment rate.

a variety of clinical trial contexts.

Ongoing studies using Flatiron Health's real-world, prospective study platform include real-time, centralized, machine-assisted patient identification for studies, with notifications deployed to the study team based on routinely collected EHR data. We are assessing whether this intervention—aimed to further reduce site burden—will enhance the proportion of eligible patients who enroll in clinical trials. In addition, although this was not a treatment intervention study, ctDNA results were provided to treating clinicians, which may have increased the value of enrollment for the clinicians and their patients. The study period notably spanned the COVID-19 pandemic, and while many clinical trials resultantly suffered reductions in enrollment [39], rates of patient accrual in this study were consistently maintained. We posit that lower-burden, technology-enabled studies are less susceptible to site and global challenges that might limit a site's ability to direct resources from routine clinical care toward clinical trials.

Over 99 % of patients were recruited at community-based sites, indicating that a highly pragmatic study design with recruitment

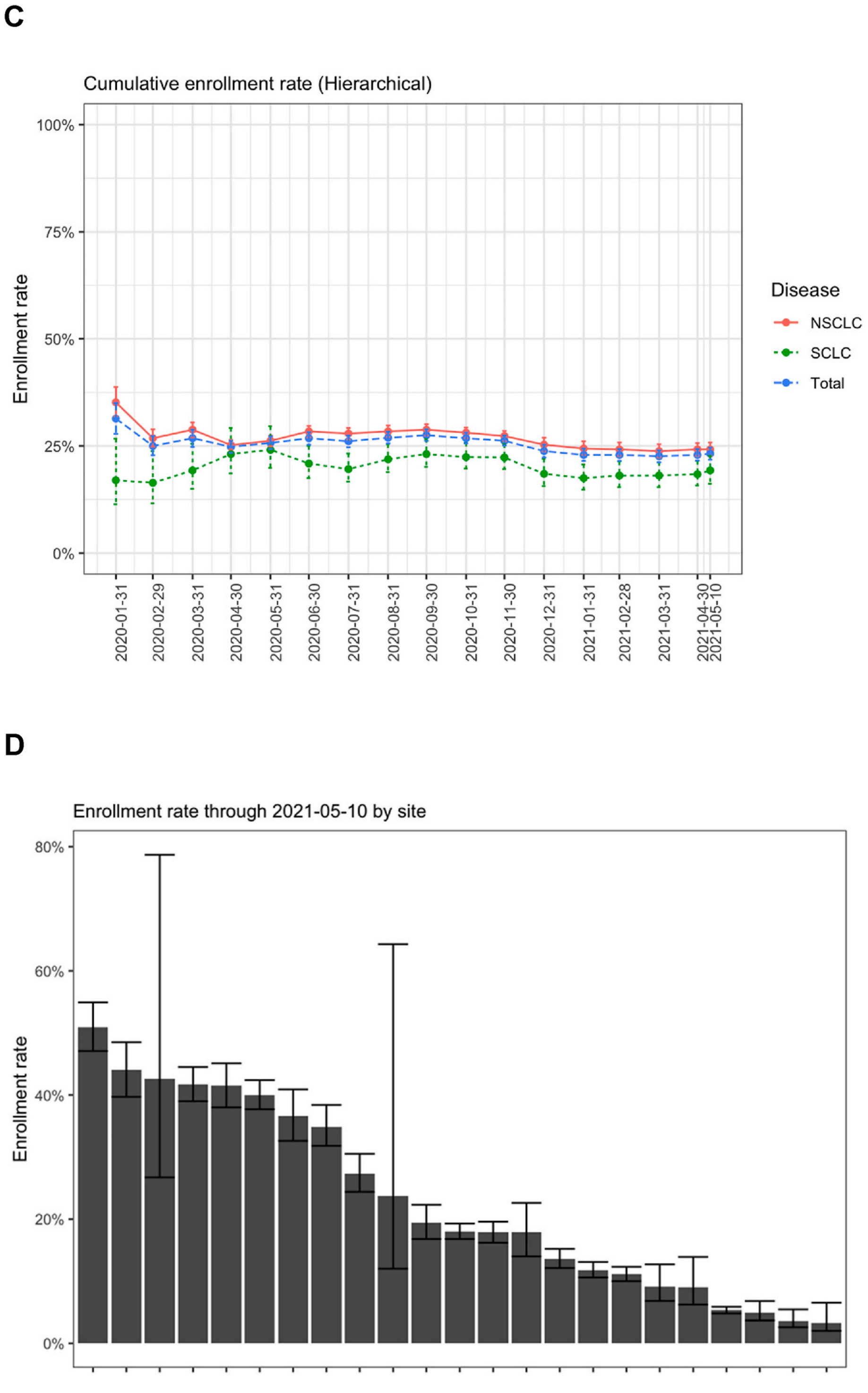

Fig. 2. (*continued*).

support and centralized data collection was feasible for collecting real-world clinical and biomarker data among oncology patients treated at these types of sites. The higher recruitment of patients with NSCLC vs SCLC in this study was expected as approximately 80 % of lung cancer diagnoses are NSCLC and 15 % are SCLC [40]. Lowering barriers to non-academic site enrollment is crucial for advancing cancer treatment, as 85 % of US patients with cancer receive community-based oncological care [4]. Modern genomic discoveries driving improvements in oncological care have largely been informed by consortia that exclude community sites including GENIE, which comprises 18 institutions and requires ≥500 unique genomic records for membership [4], and the National Cancer Institute's Cancer Genome Atlas, which is a collaboration between 20 institutions with 11,000 patients [41]. To build on the success of these platforms, there is a need for robust datasets composed of representative patient populations recruited from real-world, community-based clinics. Recent analyses continue to demonstrate that <5 % of patients with cancer in community practices take part in treatment studies [42,43]. An analysis of community oncology practices with low vs high research engagement found that low research engagement sites treated higher proportions of patients who were Latinx or Black [42]. These data highlight a need for enrollment strategies that are accessible and easily implementable at community sites.

Table 2
Demographics and clinical characteristics.

Characteristic	N = 944
Age, median (IQR), years	69 (63–75)
Sex, n (%)	
Female	474 (50.2)
Male	470 (49.8)
Race, n (%)	
Asian	9 (1.0)
Black or African American	100 (10.6)
White	675 (71.5)
Other	87 (9.2)
Missing	73 (7.7)
Practice type, n (%)	
Academic	7 (0.7)
Community	937 (99.3)
SES index 2015–2019, n (%)[a]	
1 (lowest)	136 (14.4)
2	175 (18.5)
3	191 (20.2)
4	206 (21.8)
5 (highest)	150 (15.9)
Smoking status, n (%)	
History of smoking	845 (89.5)
No history of smoking	87 (9.2)
Unknown	12 (1.3)
Disease, n (%)	
NSCLC	774 (82.0)
SCLC	170 (18.0)
AJCC stage at diagnosis, n (%)	
I	63 (6.7)
II	37 (3.9)
III	118 (12.5)
IV	717 (76.0)
Unknown/not documented	9 (1.0)
ECOG performance status, n (%)	
0	282 (29.9)
1	365 (38.7)
2	170 (18.0)
≥3	28 (3.0)
Unknown	99 (10.5)
NSCLC histology (n = 774), n (%)	
Non-squamous cell carcinoma	570 (73.6)
Squamous cell carcinoma	175 (22.6)
NSCLC histology NOS	23 (3.0)
Unknown	6 (0.8)
SCLC stage at diagnosis (n = 170), n (%)	
Extensive	139 (81.8)
Limited	23 (13.5)
Unknown	8 (4.7)

AJCC = American Joint Committee on Cancer; ECOG = Eastern Cooperative Oncology Group; ES-SCLC = extensive-stage small cell lung cancer; mNSCLC = metastatic non-small cell lung cancer; NSCLC = non-small cell lung cancer; NOS = not otherwise specified; SCLC = small cell lung cancer; SES = socioeconomic status.

[a] Using the Yost index methodology [21], the area-level SES was measured per block groups from 2015 to 2019 based on the participant's most recent address. SES index quintile scores were based on the Census Bureau's American Community Survey using 2010 Census boundaries [22,23].

Our community-based approach supported representative inclusion of individuals with low SES (roughly 20 % in each SES index quintile) and Black or African American individuals (11 % of the study population compared with 12 % of the US population [23]), which stands in contrast to cohorts historically recruited to clinical trials. The Surveillance, Epidemiology, and End Results Program (2016–2020) reported that non-Hispanic White and non-Hispanic Black individuals have the highest 5-year age-adjusted incidence rates for lung and bronchus cancer (56.4 and 53.8 per 100,000 population, respectively) [44]. Among 456,142 total cases of lung and bronchus cancer surveyed by Surveillance, Epidemiology, and End Results, 346,671 (76 %) of the individuals were non-Hispanic White and 49,130 (11 %) were non-Hispanic Black [44], which closely matches the demographic makeup of the PCG study. In Project GENIE, 72 % of patients indicated White as their primary race, which was similar to what was observed in PCG; 6 % indicated Black as their primary race (vs 11 % in PCG) and 5 % indicated Asian as their primary race (vs 1 % in PCG) [6]. In addition, of the >20,000 US-based clinical trials carried out between 2000 and 2020 that were assessed by Turner et al., only 43 % reported race and ethnicity data, and median combined enrollment of underrepresented groups fell well below census estimates [39]. The US Food and Drug Administration has draft guidance that recommends clinical trials have racial and ethnic representation that matches the US patient population and that diversity plans are developed for enrollment of historically underrepresented patients [45,46]. Using a community-based, prospective approach such as that adopted for the current PCG study may assist in meeting these goals.

It is notable that sample collection at sites was incomplete, with 35 % of participants providing samples at all eligible timepoints. In studies in which tissue submission is not an enrollment requirement, only 5 %–10 % sample completion is typically achieved (Genentech, Inc., internal data). Because there is no direct benefit to trial participants if they donate tissue samples, this comparison suggests that the PCG platform may have lowered barriers to sample completion. Also, participants were not required to make additional clinic visits just to provide samples; instead, sample collection was incorporated into their routine visits, which was a strength of the study design. However, if an individual experienced disease progression before their next visit after enrollment, they would miss the opportunity to provide an on-treatment sample. Notably, alerts were not implemented to remind site study teams to collect these specimens. In the future, automated interventions could be incorporated into EHR workflows to remind research staff of study activities and to support increased completeness of sample collection among enrolled patients.

The clinico-genomic database created in this study has enabled research evaluating the clinical utility of ctDNA profiling in assessing real-world response to first- or second-line antineoplastic treatment [19]. Results suggested that ctDNA tumor fraction (TF) was associated with longitudinal clinical outcomes, where individuals with decreased TF were responsive to therapy and those with increased TF had elevated likelihood of disease progression [19]. In another study, the clinical utility of liquid-based CGP was measured against that of tissue-based CGP in patients with advanced non-squamous NSCLC (N = 515) at the start of therapy, demonstrating that blood-based CGP can enhance detection of actionable driver alterations, especially when tissue CGP is unavailable [47]. Cultivating robust, real-world datasets that are suitable both to generate biological hypotheses and gain insight into clinical practices is critical for supporting the possibility of non-invasive tumor monitoring in patients with lung cancer.

The PCG study demonstrated the feasibility of rapidly enrolling a large and diverse real-world patient cohort to a non-treatment, low-interventional study in the community setting. Centralized data collection was designed to minimize the burden on local practice staff, and pragmatic elements, including broad inclusion criteria and minimal data collection outside of routine care documentation, bolstered the feasibility of this study. In addition to creation of cohorts for biomarker discovery and validation, a community-based, prospective, real-world study platform has potential applications to other research contexts, such as post-marketing studies, natural history studies and the establishment of external controls where randomization is not feasible.

CRediT authorship contribution statement

Alexia Exarchos: Writing – review & editing, Project administration. **Ariel B. Bourla:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maneet Kaur:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Katja Schulze:** Writing – review & editing, Conceptualization. **Sophia Maund:** Writing – review

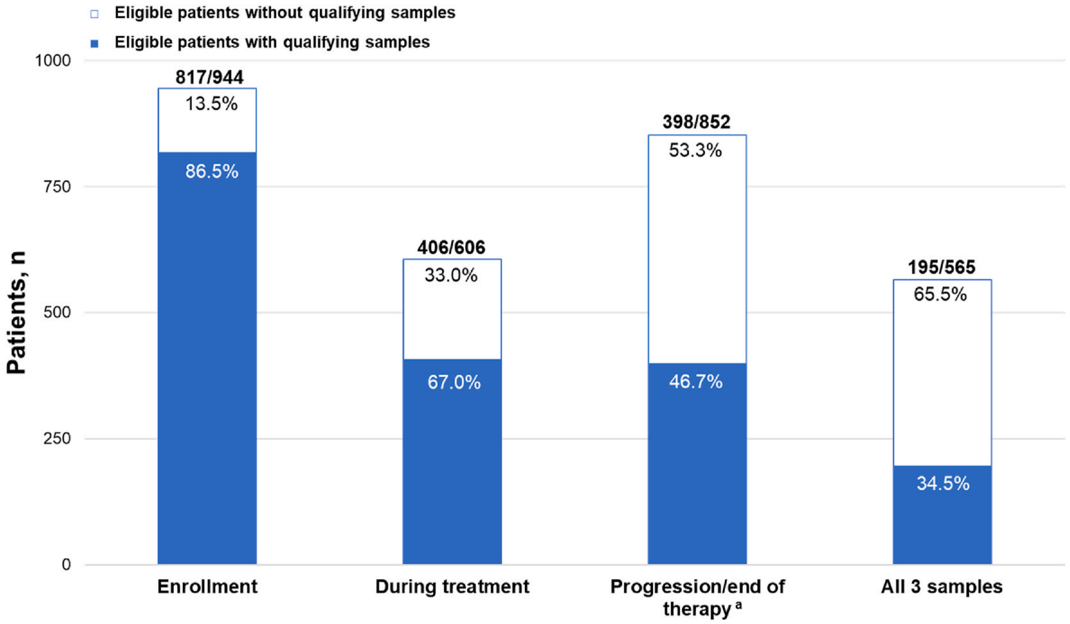

Fig. 3. Proportion of eligible patients at each timepoint with qualifying samples.
[a]Sixty-seven patients had disease progression and ended therapy at their first tumor assessment.

Table 3
Operations metrics.

Metric	PCG	Comparable clinical trials [26–37]
Consent rate, %	28	3
Time to site activation, median, days	58	130
Time to FPI, median, days	8	101
Screening failure rate, %	3	34

FPI = first patient in; PCG = Prospective Clinico-Genomic Study.

& editing, Methodology, Data curation, Conceptualization. **Yi Cao:** Writing – review & editing. **Yihua Zhao:** Writing – review & editing, Supervision. **Elizabeth H. Williams:** Writing – review & editing, Data curation. **Sarah C. Gaffey:** Writing – review & editing, Resources, Project administration. **Richard Zuniga:** Writing – review & editing, Resources. **Shaily Lakhanpal:** Writing – review & editing. **Vladan Antic:** Writing – review & editing, Validation, Conceptualization. **Michelle Doral:** Writing – review & editing, Conceptualization. **Johanna Sy:** Writing – review & editing. **Neal J. Meropol:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Anne C. Chiang:** Writing – review & editing, Resources.

Availability of data and material

The data that support the findings of this study have been originated by Flatiron Health, Inc., and Foundation Medicine, Inc. These de-identified data may be made available upon request and are subject to a license agreement with Flatiron Health and Foundation Medicine. Interested researchers should contact clinical.operations@foundationmedicine.com and dataaccess@flatiron.com to determine licensing terms.

Funding

This work was supported by Genentech, Inc. and F. Hoffmann-La Roche Ltd.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: **A Exarchos, K Schulze, S Maund, Y Cao, M Doral** and **J Sy** are employees of Genentech, Inc., and shareholders of F. Hoffmann-La Roche Ltd. **AB Bourla** was an employee of Flatiron Health, Inc., at the time of the study and is currently an employee and stockholder of Johnson & Johnson. **M Kaur** and **NJ Meropol** are employees of Flatiron Health, Inc., an independent member of F. Hoffmann-La Roche Ltd., and hold equity in F. Hoffmann-La Roche Ltd. **Y Zhao** was an employee of Flatiron Health, Inc., at the time of this study and is a stockholder of F. Hoffmann-La Roche Ltd. **EH Williams** and **SC Gaffey** are employees of Foundation Medicine, Inc., an independent member of F. Hoffmann-La Roche Ltd., and hold equity in F. Hoffmann-La Roche Ltd. **R Zuniga** and. **S Lakhanpal** have nothing to disclose. **V Antic** is an employee and shareholder of F. Hoffmann-La Roche Ltd. **AC Chiang** has received advisory board or consulting fees from AstraZeneca, Genentech, Inc., Janssen and Daiichi Sankyo, and has received research support from AbbVie, Amgen and Bristol Myers Squibb.

Acknowledgments

We thank the patients and their families. Medical writing assistance for this manuscript was provided by Emma Bouck, PhD, of Nucleus Global, and funded by F. Hoffmann-La Roche Ltd.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2025.101446.

References

[1] T. Mok, D.R. Camidge, S.M. Gadgeel, et al., Updated overall survival and final progression-free survival data for patients with treatment-naive advanced ALK-positive non-small-cell lung cancer in the ALEX study, Ann. Oncol. 31 (8) (2020) 1056–1064, https://doi.org/10.1016/j.annonc.2020.04.478.

[2] A. Drilon, C.H. Chiu, Y. Fan, et al., Long-term efficacy and safety of entrectinib in *ROS1* fusion-positive NSCLC, JTO Clin Res Rep 3 (6) (2022) 100332, https://doi.org/10.1016/j.jtocrr.2022.100332.

[3] D. Planchard, B. Besse, H.J.M. Groen, et al., Phase 2 study of dabrafenib plus trametinib in patients with BRAF V600E-mutant metastatic NSCLC: updated 5-year survival rates and genomic analysis, J. Thorac. Oncol. 17 (1) (2022) 103–115, https://doi.org/10.1016/j.jtho.2021.08.011.

[4] AACR Project GENIE Consortium, AACR Project GENIE: powering precision medicine through an international consortium, Cancer Discov. 7 (8) (2017) 818–831, https://doi.org/10.1158/2159-8290.CD-17-0151.

[5] A. Zehir, R. Benayed, R.H. Shah, et al., Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients, Nat. Med. 23 (6) (2017) 703–713, https://doi.org/10.1038/nm.4333.

[6] T.J. Pugh, J.L. Bell, J.P. Bruce, et al., AACR Project GENIE: 100,000 cases and beyond, Cancer Discov. 12 (9) (2022) 2044–2057, https://doi.org/10.1158/2159-8290.CD-21-1547.

[7] L.M. Smyth, Q. Zhou, B. Nguyen, et al., Characteristics and outcome of AKT1 E17K-mutant breast cancer defined through AACR Project GENIE, a clinicogenomic registry, Cancer Discov. 10 (4) (2020) 526–535, https://doi.org/10.1158/2159-8290.CD-19-1209.

[8] G. Singal, P.G. Miller, V. Agarwala, et al., Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database, JAMA 321 (14) (2019) 1391–1399, https://doi.org/10.1001/jama.2019.3241.

[9] V.H. Murthy, H.M. Krumholz, C.P. Gross, Participation in cancer clinical trials: race-, sex-, and age-based disparities, JAMA 291 (22) (2004) 2720–2726, https://doi.org/10.1001/jama.291.22.2720.

[10] W.B. Al-Refaie, S.M. Vickers, W. Zhong, H. Parsons, D. Rothenberger, E. B. Habermann, Cancer trials versus the real world in the United States, Ann. Surg. 254 (3) (2011) 438–442, https://doi.org/10.1097/SLA.0b013e31822a7047. ; discussion 442-443.

[11] W.B. Sateren, E.L. Trimble, J. Abrams, et al., How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials, J. Clin. Oncol. 20 (8) (2002) 2109–2117, https://doi.org/10.1200/JCO.2002.08.056.

[12] M.B. Amin, F.L. Greene, S.B. Edge, et al., The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging, Ca - Cancer J. Clin. 67 (2) (2017) 93–99, https://doi.org/10.3322/caac.21388.

[13] B. Birnbaum, N. Nussbaum, K. Seidl-Rathkopf, et al., Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research, arXiv (2020), https://doi.org/10.48550/arXiv.2001.09765.

[14] X. Ma, L. Long, S. Moon, B.J. Adamson, S.S. Baxi, Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR, medRxiv (2023), https://doi.org/10.1101/2020.03.16.20037143.

[15] OncoTrials®. Flatiron health. https://flatiron.com/clinical-research-solutions/oncotrials. (Accessed 9 January 2024).

[16] B. Adamson, M. Waskom, A. Blarre, et al., Approach to machine learning for extraction of real-world data variables from electronic health records, Front. Pharmacol. 14 (2023) 1180962, https://doi.org/10.3389/fphar.2023.1180962.

[17] Q. Zhang, A. Gossai, S. Monroe, N.C. Nussbaum, C.M. Parrinello, Validation analysis of a composite real-world mortality endpoint for patients with cancer in the United States, Health Serv. Res. 56 (6) (2021) 1281–1287, https://doi.org/10.1111/1475-6773.13669.

[18] S.D. Griffith, R.A. Miksad, G. Calkins, et al., Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small-cell lung cancer data set, JCO Clin Cancer Inform 3 (2019) 1–13, https://doi.org/10.1200/CCI.19.00013.

[19] A. Chiang, M. Kaur, Z. Assaf, et al., Circulating tumor DNA (ctDNA) changes in patients with lung cancer from a real-world prospective clinico-genomic (PCG) study. World Conference on Lung Cancer, 2022. August 6-9, 2022; Vienna, Austria.

[20] A. Vanderwalde, M. Lu, S. Maund, et al., ctDNA and real-world response (rwR) in patients with lung cancer from a prospective real-world clinico-genomic (PCG) study. Presented at: 2021 World Conference on Lung Cancer, Virtual, September .

[21] K. Yost, C. Perkins, R. Cohen, C. Morris, W. Wright, Socioeconomic status and breast cancer incidence in California for different race/ethnic groups, Cancer Causes Control 12 (8) (2001) 703–711, https://doi.org/10.1023/a:1011240019516.

[22] Geography Boundaries by Year. United States Census Bureau. Accessed November 1, 2023. https://www.census.gov/programs-surveys/acs/geography-acs/geography-boundaries-by-year.2010.html#list-tab-626530102.

[23] American community survey. United States Census Bureau. https://data.census.gov/. (Accessed 1 November 2023).

[24] X. Ma, L. Long, S. Moon, B.J. Adamson, S.S. Baxi, Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR, medRxiv (2020), https://doi.org/10.1101/2020.03.16.20037143. Preprint posted online May 30.

[25] P. Mpofu, S. Kent, P. Jónsson, et al., Evaluation of US oncology electronic health record real-world data to reduce uncertainty in health technology appraisals: a retrospective cohort study, BMJ Open 13 (10) (2023) e074559, https://doi.org/10.1136/bmjopen-2023-074559.

[26] ClinicalTrials.gov. A study of atezolizumab versus placebo in combination with paclitaxel, carboplatin, and bevacizumab in participants with newly-diagnosed stage III or stage IV ovarian, fallopian tube, or primary peritoneal cancer (IMagyn050). https://clinicaltrials.gov/study/NCT03038100. (Accessed 6 June 2024).

[27] ClinicalTrials.gov, A study of atezolizumab (MPDL3280A) compared with a platinum agent (cisplatin or carboplatin) + (pemetrexed or gemcitabine) in participants with stage IV non-squamous or squamous non-small cell lung cancer (NSCLC) [IMpower110], https://clinicaltrials.gov/study/NCT02409342. (Accessed 6 June 2024).

[28] ClinicalTrials.gov, A study of atezolizumab in combination with nab-paclitaxel compared with placebo with nab-paclitaxel for participants with previously untreated metastatic triple-negative breast cancer (IMpassion130). https://clinicaltrials.gov/study/NCT02425891. (Accessed 6 June 2024).

[29] ClinicalTrials.gov, A study of trastuzumab emtansine (kadcyla) plus pertuzumab (perjeta) following anthracyclines in comparison with trastuzumab (herceptin) plus pertuzumab and a taxane following anthracyclines as adjuvant therapy in participants with operable HER2-positive primary breast cancer. https://clinicaltrials.gov/study/NCT01966471. (Accessed 6 June 2024).

[30] ClinicalTrials.gov, A study of atezolizumab in participants with programmed death-ligand 1 (PD-L1) positive locally advanced or metastatic non-small cell lung cancer (NSCLC) [FIR], https://clinicaltrials.gov/study/NCT01846416. (Accessed 6 June 2024).

[31] ClinicalTrials.gov, A study of atezolizumab in participants with programmed death - ligand 1 (PD-L1) positive locally advanced or metastatic non-small cell lung cancer (BIRCH). https://clinicaltrials.gov/study/NCT02031458. (Accessed 6 June 2024).

[32] ClinicalTrials.gov, A study of MOXR0916 in combination with atezolizumab versus atezolizumab alone in participants with untreated locally advanced or metastatic urothelial carcinoma who are ineligible for cisplatin-based therapy. https://clinicaltrials.gov/study/NCT03029832. (Accessed 6 June 2024).

[33] ClinicalTrials.gov, Safety and pharmacokinetics (PK) of escalating doses of tiragolumab as a single agent and in combination with atezolizumab and/or other anti-cancer therapies in locally advanced or metastatic tumors. https://clinicaltrials.gov/study/NCT02794571. (Accessed 6 June 2024).

[34] ClinicalTrials.gov. A study of atezolizumab in combination with carboplatin plus (+) paclitaxel with or without bevacizumab compared with carboplatin+paclitaxel +bevacizumab in participants with stage IV non-squamous non-small cell lung cancer (NSCLC) (IMpower150). https://clinicaltrials.gov/study/NCT02366143. (Accessed 6 June 2024).

[35] ClinicalTrials.gov, A study of atezolizumab in combination with bevacizumab versus sunitinib in participants with untreated advanced renal cell carcinoma (RCC) (IMmotion151). https://clinicaltrials.gov/study/NCT02420821. (Accessed 6 June 2024).

[36] ClinicalTrials.gov. A study of atezolizumab in participants with locally advanced or metastatic urothelial bladder cancer (cohort 1). https://clinicaltrials.gov/study/NCT02951767. (Accessed 6 June 2024).

[37] ClinicalTrials.gov, Study to assess safety and efficacy of atezolizumab (MPDL3280A) compared to best supportive care following chemotherapy in patients with lung cancer [IMpower010], https://clinicaltrials.gov/study/NCT02486718. (Accessed 6 June 2024).

[38] S.K. Cheng, M.S. Dietrich, D.M. Dilts, A sense of urgency: evaluating the link between clinical trial development time and the accrual performance of cancer therapy evaluation program (NCI-CTEP) sponsored studies, Clin. Cancer Res. 16 (22) (2010) 5557–5563, https://doi.org/10.1158/1078-0432.CCR-10-0133.

[39] D.M. Waterhouse, R.D. Harvey, P. Hurley, et al., Early impact of COVID-19 on the conduct of oncology clinical trials and long-term opportunities for transformation: findings from an American Society of Clinical Oncology survey, JCO Oncol Pract 16 (7) (2020) 417–421, https://doi.org/10.1200/OP.20.00275.

[40] R.S. Herbst, J.V. Heymach, S.M. Lippman, Lung cancer, N. Engl. J. Med. 359 (13) (2008) 1367–1380, https://doi.org/10.1056/NEJMra0802714.

[41] The Cancer Genome Atlas Program (TCGA). Center for Cancer Genomics, National Cancer Institute. https://www.cancer.gov/ccg/research/genome-sequencing/tcga. (Accessed 26 September 2024).

[42] I. Altomare, X. Wang, M. Kaur, et al., Are community oncology practices with or without clinical research programs different? A comparison of patient and practice characteristics, JNCI Cancer Spectr. 8 (4) (2024) pkae060, https://doi.org/10.1093/jncics/pkae060.

[43] J.M. Unger, L.W. Shulman, M.A. Facktor, H. Nelson, M.E. Fleury, National estimates of the participation of patients with cancer in clinical research studies based on Commission on Cancer accreditation data, J. Clin. Oncol. 42 (18) (2024) 2139–2148, https://doi.org/10.1200/JCO.23.01030.

[44] SEER*Explorer, An interactive website for cancer statistics. Surveillance Research Program, National Cancer Institute, November 16, 2023. https://seer.cancer.gov/statistics-network/explorer/. (Accessed 19 December 2023).

[45] US Food and Drug Administration, Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials: guidance for industry. 2022. https://www.regulations.gov/document/FDA-2021-D-0789-0001. (Accessed 27 November 2023).

[46] P. Hyman, P.C. McNamara, Food and Drug omnibus reform act of 2022 Published January 21, 2023 https://www.thefdalawblog.com/wp-content/uploads/2023/01/HPM-FDORA-Summary-and-Analysis.pdf. (Accessed 6 June 2024).

[47] L.S. Schwartzberg, G. Li, K. Tolba, et al., Complementary roles for tissue- and blood-based comprehensive genomic profiling for detection of actionable driver alterations in advanced NSCLC, JTO Clin Res Rep 3 (9) (2022) 100386, https://doi.org/10.1016/j.jtocrr.2022.100386.