



OPEN ACCESS

EDITED BY

Lise Aagaard,
Independent Researcher, Copenhagen, Denmark

REVIEWED BY

Harlan Pittell,
Flatiron Health, United States
Arinjita Bhattacharyya,
University of Louisville, United States

*CORRESPONDENCE

Patrick Silva
✉ RickSilva@tamu.edu

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 31 March 2023

ACCEPTED 31 May 2023

PUBLISHED 06 July 2023

CITATION

Silva P, Janjan N, Ramos KS, Udeani G, Zhong L, Ory MG and Smith ML (2023) External control arms: COVID-19 reveals the merits of using real world evidence in real-time for clinical and public health investigations.
Front. Med. 10:1198088.
doi: 10.3389/fmed.2023.1198088

COPYRIGHT

© 2023 Silva, Janjan, Ramos, Udeani, Zhong, Ory and Smith. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

External control arms: COVID-19 reveals the merits of using real world evidence in real-time for clinical and public health investigations

Patrick Silva^{1*†}, Nora Janjan^{2†}, Kenneth S. Ramos¹, George Udeani³, Lixian Zhong⁴, Marcia G. Ory² and Matthew Lee Smith²

¹Institute of Bioscience and Technology and Department of Translational Medical Sciences, College Station, TX, United States, ²Center for Community Health and Aging, School of Public Health, Texas A&M University, College Station, TX, United States, ³Department of Clinical Pharmacy, School of Pharmacy, Texas A&M University, College Station, TX, United States, ⁴Department of Pharmaceutical Sciences, School of Pharmacy, Texas A&M University, College Station, TX, United States

Randomized controlled trials are considered the 'gold standard' to reduce bias by randomizing patients to an experimental intervention, versus placebo or standard of care cohort. There are inherent challenges to enrolling a standard of care or cohorts: costs, site engagement logistics, socioeconomic variability, patient willingness, ethics of placebo interventions, cannibalizing the treatment arm population, and extending study duration. The COVID-19 pandemic has magnified aspects of constraints in trial recruitment and logistics, spurring innovative approaches to reducing trial sizes, accelerating trial accrual while preserving statistical rigor. Using data from medical records and databases allows for construction of external control arms that reduce the costs of an external control arm (ECA) randomized to standard of care. Simultaneously examining covariates of the clinical outcomes in ECAs that are being measured in the interventional arm can be particularly useful in phase 2 trials to better understand social and genetic determinants of clinical outcomes that might inform pivotal trial design. The FDA and EMA have promulgated a number of publicly available guidance documents and qualification reports that inform the use of this regulatory science tool to streamline clinical development, of phase 4 surveillance, and policy aspects of clinical outcomes research. Availability and quality of real-world data (RWD) are a prevalent impediment to the use of ECAs given such data is not collected with the rigor and deliberateness that characterizes prospective interventional control arm data. Conversely, in the case of contemporary control arms, a clinical trial outcome can be compared to a contemporary standard of care in cases where the standard of care is evolving at a fast pace, such as the use of checkpoint inhibitors in cancer care. Innovative statistical methods are an essential aspect of an ECA strategy and regulatory paths for these innovative approaches have been navigated, qualified, and in some cases published.

KEYWORDS

external control arm, real world data, regulatory science, pharmacogenomics, oncology, pragmatic study

1. Introduction

Randomized controlled trials (RCTs) have been considered the ‘gold standard’ in clinical research, despite recent concerns about external generalizability and feasibility (1–3). Traditionally using 1:1 randomization to assign patients into the experimental (i.e., study intervention) and control (i.e., standard of care) study arms, RCTs often take years to complete. Further, RCTs are often impractical study designs for rare diseases. Generalizability can be limited by RCT enrollment criteria that often excludes common comorbidities. While multi-institutional trials are intended to overcome the inefficiencies, recruitment challenges, and costs of RCTs within an institution, these multi-institutional trials often do not expand the eligibility criteria of the Phase II efficacy trial to make the study more generalizable (4).

Clinical trial models have evolved to improve the sensitivity and efficiency of analysis. Reducing the overall number of study participants needed to evaluate trial outcomes statistically places fewer patients at risk for harm and reduces the number of patients randomized to a potentially less effective therapeutic control arm. The size of the intervention arm and the size of the control arm have different ethical and practical considerations but reducing each can have benefits to patients and study sponsors. In rare diseases, a control arm may be impractical, unethical, or otherwise impossible (5). Batten disease, a rare fatal inherited disorder, is a type of neuronal ceroid lipofuscinosis, in which the nervous system is unable to recycle certain degradation products. In clinical trial designs, the choice of a control arm is an especially critical aspect of trial design in which ethical and scientific issues are deeply entwined (6). A clinical trial for cerliponase alfa faced the challenge of very few patients available to enroll, and an ethical quandary about the best comparator to show efficacy with scientific rigor. The only practical control was to compare to disease progression in a historical standard of care cohort (7) using a curated global registry enabled this approach to clinical validation with cerliponase alfa (8).

The purposes of this paper are to: (1) highlight innovative statistical methods and document methodological challenges with clinical trials conducted during the COVID-19 pandemic in the US and internationally; and (2) offer recommendations for alternative approaches to capture real world data and utilize external control arms in clinical trials.

Abbreviations: ECA, external control arm; FDA, Food and Drug Administration; EMA, European Medicines Agency; RCT, randomized controlled trials; ICU, intensive care unit; OCEBM, Oxford Centre for Evidence-Based Medicine; VTE, venous thromboembolism; RWD, real-world data; CDRH, Center for Devices and Radiological Health; CBER, Center for Biologics Evaluation and Research; NGS, next-generation sequencing; SPSO, semi-parametric stochastic ordering; CRBM, Conditional Restricted Boltzmann Machine; CYP2C19, Cytochrome P450 2C19; CYP2D6, Cytochrome P450 2D6; rwTTD, real world time-to-treatment-discontinuation; rwOS, real world overall survival; UGT2B7, UDP glucuronosyltransferase family 2 member B7; UGT2B15, UDP glucuronosyltransferase family 2 member B15; PTGS1, prostaglandin endoperoxide synthase 1; CRP, C-reactive protein; PTGIR, Prostaglandin I2 receptor; HCRU, healthcare resource utilization.

2. Clinical trials during the COVID-19 pandemic

2.1. Changing best practices

Long-awaited reforms to generally accepted best practices for clinical trials were underscored during the COVID-19 pandemic. Using global technological advances, clinical outcomes were rapidly shared, often outside of formal clinical trials, due to the dire need to support patients and understand the pathophysiology of the disease. In contrast to the rapid identification and development of therapeutics and vaccine development for COVID-19, traditional development of therapeutics is highly inefficient before clinical validation and deployment is reached (9). This was particularly evident during the COVID-19 pandemic when the ethics of therapeutic randomization were questioned because best supportive care was ill-defined and rapidly changing (10). Many generally accepted clinical trial practices, were at odds with the immediate medical imperatives, and had to be reconsidered during the COVID-19 pandemic. Specifically for oncology, there was a call to rethink clinical trial dogma and revamp clinical trial design to increase efficiency, avoid highly restrictive eligibility criteria, and address the clinical needs of patients (11–13).

2.2. Limitations in traditional RCT approaches

Multiple attempts were made at conducting RCTs to validate therapeutics during the COVID-19 pandemic illustrating limitations in traditional RCT approaches. We provide several examples highlighting challenges in interpretation of findings:

(i) For example, patients ($n=596$) with moderate COVID-19 presentations (defined as having pulmonary infiltrates and room-air oxygen saturation of $>94\%$) were enrolled in a study from March 15 through April 18, 2020 at 105 hospitals in the United States, Europe, and Asia (14). Patients were randomized to receive remdesivir or standard care. Using a 1:1:1 ratio, patients either received a 10-day course of remdesivir ($n=197$), a 5-day course of remdesivir ($n=199$), or standard care ($n=200$). Four months later, the published results showed no statistically significant difference between a 10-day course of remdesivir and standard care. While a 5-day course of remdesivir resulted in a statistically significant improvement in status, the difference was of “uncertain clinical importance.” There were significant study limitations as hospital discharge rates varied greatly across regions, the open-label design potentially led to bias, some laboratory parameters were not collected, and viral loads were not assessed (15). This illustrates inherent challenge of achieving a perfect comparator group in a RCT. Put another way, an RCT is theoretically ideal but sometimes an imperfect and impractical assessment of a therapeutic in a real-world setting.

(ii) In another example, retrospective meta-analysis was performed using pooled data from 7 RCTs, conducted from February 26, 2020 to June 9, 2020, in 12 countries that evaluated the efficacy of three corticosteroid regimens in the treatment of COVID-19. The three corticosteroid regimens were systemic dexamethasone (3 trials, 1,282 patients, and 527 deaths), hydrocortisone (3 trials, 374 patients, and 94 deaths), and methylprednisolone (1 trial, 47 patients, and 26 deaths). The study included a total of 1,703 critically ill COVID-19

patients out of the 1,920 patients that were planned for statistical analysis. There was little inconsistency between individual study outcomes, resulting in few differences observed in the meta-analysis. All studies showed that systemic corticosteroids significantly lowered the 28-day all-cause mortality compared with usual care or placebo. The odds ratio (OR) for mortality with dexamethasone was 0.64 (95% CI, 0.50–0.82; $p < 0.001$), hydrocortisone was 0.69 (95% CI, 0.43–0.1.12; $p = 0.13$), and methylprednisolone was 0.91 (95% CI, 0.29–2.87; $p = 0.87$) (14).

(iii) In another case, an attempt was made to conduct a multicenter randomized double-blinded sequential trial in France that evaluated the effectiveness of hydrocortisone in COVID-19 intensive care unit (ICU) patients with respiratory failure. The primary outcome measure, treatment failure on day 21, was defined as death or persistent dependence on a ventilator or high-flow oxygen therapy. The study began March 7, 2020, and was stopped on June 29, 2020 when an interim analysis was performed. Although likely underpowered to identify a statistically and clinically important difference in the study endpoint, the trial was stopped early after 149 of the 290 patients were enrolled based on the strength of the interim results of this and other studies (16).

2.3. Studies evaluating quality of RCTs

We also note two studies explicitly evaluating the quality of randomized clinical trials. The first is a survey of the 516 COVID-19 randomized clinical trials, registered in [ClinicalTrials.gov](https://www.clinicaltrials.gov) and the World Health Organization International Clinical Trials Registry Platform between January 1 and April 9, 2020, conducted approximately a year later in October 2020. The survey evaluated whether the trials recruited 75% or more of their target sample size, stopped before reaching 75% recruitment, or continued to recruit patients but had not yet met the 75% accrual (either on schedule or delayed). Of the 516 randomized trials, only 53 (10.3%) of the 56 completed studies had been published (30%), had not started or were discontinued, and 24 (4.6%) were terminated early. The remaining studies were ongoing but only 126 (24.4%) were on schedule. Of the RCTs initiated in the first 100 days of the pandemic, 30% did not begin recruitment and only 10% had results reported by mid-October 2020. Of the 24 terminated and 46 discontinued trials, 14 RCT investigators communicated that the trials were discontinued due to decreasing COVID-19 cases (8 trials [57.1%]), emerging data regarding safety (6 trials [42.9%]), or futility (2 trials [14.3%]). Importantly, the statistically significant safety concerns, affirmed through peer-review, were derived from observational studies before a RCT could be completed (17). A high rate of multiplicity was noted during a June 8, 2020, review of the COVID-19 trials on [ClinicalTrials.gov](https://www.clinicaltrials.gov). This multiplicity was thought to enhance the likelihood of finding a positive result through chance alone resulting in the widespread administration of a potentially ineffective therapeutic. Fragmentation of efforts also lead to competition for study participants that compromised clinical trial accrual and the statistical power of all trials (18).

Second, a cross-sectional analysis was conducted of the characteristics and the strength of evidence of COVID-19 studies registered on [ClinicalTrials.gov](https://www.clinicaltrials.gov) on May 19, 2020. There were 640 observational and 664 RCTs. Over 75% ($n = 1,180$) of the RCTs and

observational COVID-19 studies were conducted at a single center, and only 29.1% of the COVID-19 studies (RCTs and observational) could potentially yield the OCEBM level 2 evidence, or the highest level of evidence (19). Among the RCTs, only 35.8% of studies had planned enrollment of more than 100 participants, and 17% involved at least 2 study centers, which is required for the highest level of evidence. Also, to fulfill the highest level of evidence criteria, only 29.1% of the RCTs were placebo-controlled, and only 11.3% of the placebo-controlled RCTs were blinded and conducted in at least two centers. Among the observational studies, 80.8% were conducted in a single center, and only 13.6% were prospective cohort studies that could yield the highest level of evidence (19).

2.4. Observational and therapeutic reports

By necessity, many of the COVID-19 studies were observational and single-arm therapeutic reports. These studies followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Guideline for Cohort Studies. Research consortia were also developed to evaluate the variabilities in therapeutic approaches, and to pool sufficient patient numbers for statistical analysis. An observational study of 2,483 consecutive admissions for confirmed COVID-19 was conducted from a 5-hospital health system. The outcomes of remdesivir (administered with or without a corticosteroid) were compared to matched-COVID-19 patients who did not receive remdesivir; matching was performed using time-invariant covariates and time-dependent covariates. From this study, remdesivir alone was associated with faster clinical improvement as evidenced by a 2-day shorter time to clinical improvement (adjusted hazard ratio 1.47; 95% CI 1.22–1.79) and a 7.7% (vs. 14.0%) 28-day mortality rate (adjusted hazard ratio 0.70; 95% CI 0.38–1.28) compared to matched controls (20). During the period of highest COVID-19-related mortality, from March 7, 2020, to June 17, 2020, a voluntary statewide collaborative initiative was established among 41% of the 92 noncritical access- nonfederal hospitals in Michigan to evaluate adherence to venous thromboembolism (VTE) anticoagulation regimens. Some of the hospitals included all COVID-19 patients, while high-volume hospitals used a pseudo-random sampling process to select cases given the limited availability of data abstraction resources. Pseudo-randomization, resulting from logistical constraints, involved evaluating all the potentially eligible patients; each day, one patient was selected for the study out of many eligible study patients based solely upon the time of hospital discharge (21).

3. Alternative approaches for external control arms

3.1. Real world data

In the best of circumstances, RCTs encounter difficulties in efficiently fulfilling study enrollment. During the restrictions and stressors of the COVID-19 pandemic, conducting and completing a RCT was made even more difficult. Providing information regarding the transmission and treatment of a novel virus, single-arm trials and the analysis of Real World Data (RWD) made significant scientific

contributions to understanding the epidemiology of severe disease, and the efficacy of an evolving therapeutic model during the COVID-19 pandemic.

When a RCT is not feasible for either rare diseases or ethical reasons, historical or external control arms, derived from RWD, have been used (22). Additionally, historical or external control arms often are more accurate comparators as RCTs are not perfectly representative of adherence and public health factors affecting how a therapy may be used in the real world. A combination of regulatory policy shifts, legislation, clinical data availability, and next-generation sequencing have catalyzed the feasibility of ECAs (23). Using the dynamic borrowing model, in which a historical or RWD control group closely matches a small concurrent control group in the study, there is minimal impact on the final study conclusions. This minimal impact is maintained when up to 80% of a trial’s control group consists of closely matched historical or RWD controls (22).

3.2. Regulatory guidance

In December 2016, Section 3022 of the 21st Century America Cures Act was enacted and directed the U.S. Food and Drug Administration (FDA) to accept statistical methods that include RWD and Real World Evidence (RWE) (23). The FDA’s Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER) issued guidance on August 31, 2017 regarding the submission of RWD and RWE to the FDA (24, 25). To improve the regulation of combination products, the Cures Act also directed the FDA to create inter-center institutes to coordinate initiatives in the development of drugs, biologics, and devices. The Oncology Center of Excellence was the first inter-center institute created.

After the FDA guidance was issued, the CDRH reported that the use of RWE in premarket and post-market regulatory decisions increased by 193% compared to the 2015 baseline (26). However, an

RCT was the statistical model used for approximately 60% of the studies investigating new therapeutic agents/devices (27). Given the inherent concerns for the safe administration of an agent/device, the enrolled clinical trial population may not be representative of the real world patient populations in which the agent is intended to be administered (Table 1). In general, the patient populations within RCTs are younger, more homogenous, and have fewer comorbidities than the actual patient populations (28, 29). Nearly 80% of clinical trials fail to meet the initial enrollment projection, and less than 10% of new therapeutic agents receive FDA commercial approval (30–33). RCTs may not always be representative of safety and efficacy within real world populations due to numerous and overlapping public health factors such as adherence, socioeconomic, genetics, polypharmacy and polychronic disease factors. Even after FDA approval based on an RCT, the marketing approval may be revoked or the FDA label changed (i.e., black box warnings) may occur because of unexpected adverse events within a real world population (34).

RWD and RWE have 4 roles as defined by regulatory guidance. In the first role, the FDA uses RWD and RWE to monitor post-market safety (adverse events) and efficacy that resulted in regulatory decisions. Second, to expedite approval of innovative therapeutic approaches, RWD and RWE support clinical trial designs, including approval of new indications for approved drugs. Third, RWD and RWE are used to: (a) develop therapeutic guidelines and decision support tools for clinical practice; and (b) to support healthcare insurance coverage and reimbursement decisions. Specific examples include: (a) generating and refining hypotheses to be tested in a prospective clinical study; (b) identify, demonstrate, or support the clinical validity of a biomarker; (c) support label expansion for approved drugs; and (d) for public health surveillance and policy efforts (17, 18, 35).

The sources of RWD are broad and incorporate both governmental and commercial data sources. Accepted sources of RWD include electronic health records (EHRs), claims and billings databases, product and disease registries, patient-generated data, and health

TABLE 1 Practical pros and cons of external control groups.

	RCT control groups		External control groups		
	Single institution	Multi-institution	Historical	Traditional	Contemporaneous
Allow for changes in standards of care	N	N	P	Y	Y
Decreased patient number required for clinical trial	N	N	Y	Y	Y
Ability to use digital twins	P	P	Y	Y	Y
Decreased time required for clinical trial	N	N	Y	Y	Y
Control arms receive current standard of care	Y	Y	P	Y	Y
Ethical concerns regarding inferior care in control arm	P	P	N	N	N
Receive treatment in real world healthcare setting	P	P	Y	Y	Y
Interim analysis possible	Y	Y	N	N	Y
Acceptance by FDA	Y	Y	Y	Y	Y
Increased time in getting novel treatments to clinic	Y	Y	N	N	N

Y represents-yes; N represents-no; P represents-perhaps.

status data gathered from other sources, including mobile devices. The analysis of RWD provides information about, or enables comparisons of, the clinical outcomes associated with a therapeutic intervention and characterizes standards-of-care. These therapeutic outcomes generate RWE from a wide range of study designs that include data mining, observational, and interventional studies.

To improve access to innovative medical devices, the FDA formally approved the use of RWD and RWE to reduce premarket data collection, and to monitor safety and efficacy in the post-market setting (36, 37). As in all clinical investigations, the FDA requires that all data be reliable, accurate, and verifiable with robust quality control. RWD is also required to be large enough to evaluate the medical product in the specific regulatory context. Three determining factors for the use of RWD are whether: (1) sufficient patient numbers and detail is available for analysis; (2) confounding factors can be addressed; and (3) the database is generalizable to the involved population. For example, RWD may need to account for the off-label use of an agent.

In clinical trials, reliable and relevant RWD can be applied when traditional prospective data collection models are impractical, and ethical issues arise regarding treatment assignment (38–40). The FDA guidance emphasizes the role of external controls derived from RWD in the regulatory evaluation of medical devices, and for rare or life-threatening disease (41–44). The guidance also states that randomization does not guarantee the absence of bias; rather, it mitigates the known sources of bias. Multiple international agencies and academic groups have also advocated for and accepted non-randomized evidence within clinical trials (45, 46). The use of RWE has expanded along with next-generation sequencing (NGS)-based testing and other forms of high-content bioanalysis (47, 48). The population paradox is defined as understanding the clinical relevance of molecular differences within a population. The population paradox enables the use of clinicogenomics databases for therapeutic development, clinical trials design, and therapeutic decision making for individualized medical care (49–52). Evidence gaps, however, remain for the use of NGS with RWE in health insurance coverage decision-making with the exception of pharmacogenomics testing, suspected pediatric genetic disorders, and oncology (48).

To mitigate potential bias, the use of RWD also requires a well-defined study design, clinical endpoints, bias mitigation strategy, and statistical analysis that is comparable to traditional RCTs (53, 54). Especially when multiple data sources are used, a more diverse patient population may be possible. However, the data source must be reliable, and the data should be anonymized to protect the privacy of the patient privacy. If patient-specific data is needed, such as a biomarker result, the data can be tokenized through an intermediary to link the result to the anonymous data point. When RWE is submitted to the FDA, both the RWD and RWE must conform to recognized data standards for file formats and data structures using standardized variables and definitions (55).

3.3. External control arm

Medical paradigms are changing. The intended use and labeling of therapeutics are increasingly agnostic of medical subspecialty, target tissue or organ, or pathologic characterization (56). Inter-connecting biological pathways now define risks for the development of disease or a disease process, and the corresponding therapeutic strategy. Relevant

outcomes are expanding beyond survival (57, 58). Instead, outcomes of interest now reflect the importance of controlling the symptoms and trajectory of the disease (58), quality of life indicators, and/or downstream healthcare utilization (59).

Greater emphasis is also placed on outcomes that occur after completion of the clinical trial and in real world populations (60). Real world outcomes can be used to better estimate the cost-of-illness from both the patient and societal perspective (61). Innovative statistical approaches are increasingly accepted within and outside the construct of RCTs, resulting in an evolution in the design and efficiency of conducting clinical trials (62). External control arms (ECAs), derived from RWD, are increasingly used to reduce the number of patients and time required for study completion (5). ECAs and historical control arms have been used for decades to evaluate new therapeutics in rare diseases (63). Most importantly, from an ethical perspective, ECAs also place fewer patients within a clinical trial at risk for adverse events and potentially limited therapeutic benefit (64).

3.4. Innovative statistical methods

Statistical methods have been developed to minimize the number of patients randomized to a control arm, while retaining sufficient power for primary endpoints and safety evaluations. Inefficiencies in conducting clinical trials often slow the availability of breakthrough therapies for patients. Chapple and Thall (65) introduced a novel and more efficient semi-parametric stochastic ordering (SPSO) model for Phase I-II trials. This new statistical model uses a flexible monotone increasing the toxicity model and a semi-parametric stochastic ordering model for efficacy probabilities. A novel sensitivity analysis of the range of correlations between toxicity and efficacy was then performed. As a previously unreported finding, they found that all prospective clinical study designs performed worse when there was a negative correlation between toxicity and efficacy.

Digital twins are commonly used in industry to build predictive models for clinical trial performance and the real-world implementation of a new therapy. The approach involves developing a statistical model using external data from a patient population that can predict the trajectory of the disease for a standard of care or control arm. Reducing the number of patients accessioned to the study by 30–50%, that standard of care arm is then used as a synthetic control that is compared to a single-armed interventional clinical trial. Comprehensive forecasting of Alzheimer's Disease progression was accomplished using the unsupervised Conditional Restricted Boltzmann Machine (CRBM) learning model that incorporated 44 clinical variables in 1909 patients with mild manifestations of disease. The generated synthetic patient data, including the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) score, accurately reflected the means, standard deviations, and correlations of each variable over time. The accuracy of the Alzheimer's Disease trajectory from the actual data could not be distinguished from the synthetic control by a logistic regression (66). Biological markers are used with greater frequency to personalize therapeutics and reduce risk from futile care. In oncology an ECA, derived from a de-identified clinico-genomic database, closely replicated the control arm from the randomized IMblaze370 study of metastatic colorectal cancer (67, 68). Such databases can not only match patients within the control arm but, potentially more important, may also more closely match patients randomized to the therapeutic arm of an RCT.

Confounders, such as comorbidities, may limit the generalizability of RCT results to the population identified within the FDA approved product label. Once FDA approved, unless specifically excluded in the FDA product label, a confounding factor may result in significant unanticipated adverse events that could outweigh the derived therapeutic benefit. The use of ECAs could support subsequent additional prospective single-arm Phase II-III studies that have a more expansive eligibility criteria in a limited patient cohort (69, 70). Substantially reducing the number of patients at risk, these prospective studies could provide critical toxicity and efficacy information more rapidly than a RWE retrospective review of large populations. This more rapid evaluation of relevant populations might find therapeutic benefit to a patient subpopulation that was not included in the foundational RCT, and/or prevent adverse events by identifying previously unknown risk factors. Using pharmacogenomic registry data, the genetic and pharmacologic co-variants of efficacy, safety and tolerability could also be determined (71). Known variants of CYP2C19 and CYP2D6 (72, 73), emergent α 36 variants of ER (74), and drug–drug interactions with serotonin reuptake inhibitors (75) all have clinically relevant impacts on outcomes of breast cancer patients treated with tamoxifen. Covariates, that influence clinical outcomes, can be identified in clinical trials that use ECAs and a relatively small number of patients. Identifying covariates within product registration and Phase IV clinical trial designs can improve overall response and reduce toxicity rates.

External control arms and single-arm trials are increasingly accepted by regulators for drug approval, and Health Technology Assessment (HTA) bodies (63). ECAs have also been used to provide clinical context. A 13-fold increase in the use of ECAs was observed among the 433 single-arm trial submissions between 2011 and 2019 (76). Between 2015 and 2019, the use of ECAs increased 22%, and approximately half (52%; 226/433) of the submissions included ECAs (76). The ECAs included historical controls from prior clinical trials (24%; 104), and from RWD (20%; 87), while 40% (175) of single-arm trials did not incorporate an ECA (76). The overall acceptance rate for single-arm trial submissions was 48%, increasing to 59% with RWD ECAs. The acceptance rate of single-arm trials increased from 41% from 2015 to 2017 to 61% in 2018–2019 (76). Between 2015 and 2019, the acceptance rate for single-arm trials with historical controls as ECAs decreased by 10%, and with no ECAs decreased by 329 (16%) (54).

A proof-of concept study showed that the control arms in RCTs could be replicated by ECAs from curated electronic health record data. Applying study eligibility criteria, 9 advanced non-small-cell lung cancer trials were evaluated. Key aspects of the trials ranged from biomarker availability, study start dates, and overall survival as an endpoint. A comparison of the log hazard ratios among all RCTs and ECAs resulted in a 0.86 Pearson correlation coefficient (77). As an example, advanced non-small-cell lung cancer trials have used ECAs derived from de-identified contemporaneous electronic health records to determine overall survival within single-arm trials.

3.5. Contemporaneous external control arm

External control arms are useful as a comparator to validate the control arm of a RCT by revealing potential biases inherent in the study environment or enrolled population. Within an RCT, the goal is for the control group to mirror the experimental arm cohort as much as

possible. With restricted study eligibility criteria, the RCT control and experimental arms may be matched to each other in terms of demographics and comorbidities, but these cohorts may not always be representative of the real-world patient population. Variations in outcomes may occur based on access to care, insurance payer, practice patterns, and health-system resources. Indeed, zip code is a major determinant of clinical trial variations in recruitment (78) and health outcomes (79). RCTs conducted within a single institution in a single location, versus a multi-Institutional study, will generally not account for population diversity. Diversity in outcomes, especially related to healthcare resource utilization and cost, are often attributable to the provider, payer and/or population factors. For example, many factors differ between academic medical centers in large urban areas vs. community oncology clinics, including health system resources and the social determinants of health.

As a specific illustration, outcomes were compared among de-identified patients with advanced non-small-cell lung cancer treated with programmed death 1/programmed death-ligand 1 inhibitors. The comparison involved RWD from an insurance claims database and EHRs obtained from 6 healthcare organizations (70). The datasets ranged from 269 to 6,924 patients. Correlations between real world time-to-treatment-discontinuation (rwTTD), time-to-next-treatment, and overall survival (rwOS) ranged between 0.6 and 0.9, with rwTTD being the most consistent endpoint. Real world endpoints were also consistent between institutions. At 1-year, rwOS ranged between 40 and 57%; these results were within the range of the median OS values from published clinical trials (69, 70). Two key conclusions were determined from this study: (1) real world endpoints are valid and should be used to support regulatory and payer decision-making; and (2) observed differences likely reflected true differences between real world and clinical trial populations and practices. Being conscious of these differences can inform risk mitigation for downstream clinical development and post-market surveillance.

As pharmacogenomics practice expands, the relative frequency of pharmacogene variants (drug-gene interactions), polypharmacy permutations (drug–drug interactions), and correlations related to drug-class adverse reactions might inform rare adverse effects. These toxicities, that are often difficult to detect within an RCT, might not be identified prior to FDA approval. Several drugs have been withdrawn from the market despite compelling RCT results resulting in FDA approval. In some cases, patients suffered great harm due to rare pharmacogenomics or drug–drug interactions that might not be previously known (80). For example, the cardiovascular toxicities of rofecoxib are suspected as being attributable to one or a combination of polymorphisms in a number of genes: 372 UGT2B7; (81) UGT2B15; (81) PTGS1; (12) CRP; (12) and PTGIR (82).

Combinations of rare but clinically consequential genotypes could easily be omitted from representation in the treatment (and control) arm of an RCT, but an ECA with genotype data might lend comparative insight before adverse events arise in the general population post FDA approval. The Observational Health Data Sciences and Informatics dataset contains 250 million cases. Within the dataset, 10% of diabetes, 24% of hypertension, and 11% of depression patients had unique treatment pathways (81). This then yields a daunting number of permutations in drug combinations that might materially impact safety or efficacy outcomes in RCTs. Compared to traditional RCT control arms, ECAs can validate the degree of similarity to the general population and identify factors for prospective pharmacovigilance later in drug

development or during post market monitoring. RCTs are often considered to be an impractical means to validate clinical decision-making given an increasing number of pharmacogenomics variants (83), and that RWD is an important source of evidence in support of advancing the field of pharmacogenomics (50). Treating many rare diseases, oncology practice has been significantly influenced by RWD from registries and clinicogenomics databases for clinical decision support (84).

While variability exists within RCTs, variability also exists with ECAs. There are 3 types of ECAs which are accepted by regulatory agencies and based on time of cohort acquisition. Burcu and colleagues (64) characterized the 3 types of ECAs as historical, hybrid, and contemporaneous. Using their definition, historical ECAs almost exclusively use retrospective data, while the ECAs designated as hybrid and contemporaneous collect data both retrospectively and prospectively. A hybrid ECA, collecting both retrospective and prospective data, is often used to augment a control arm in an RCT. Hybrid ECAs are established either at the end of the single-arm trial using retrospective data from the time frame that the clinical trial was conducted, and the start of a single- arm trial with real-time follow-up. Contemporaneous ECAs generally provide a matched cohort from RWD for the control arm during the conduct of the clinical trial (85).

The control arms of RCTs seek to match the experimental arm patient demographics and comorbidities, and both arms are enrolled in the study at the same time. Most RCTs, however, do not match socioeconomic factors, such as the type of healthcare insurance payer, the type of health system, or the region of the country where care is delivered. Socioeconomic factors influence health behaviors, healthcare practice patterns, availability of medications and procedures, healthcare resource utilization (HCRU), and costs. RCTs conducted in a single-institution are poised to provide the least patient variability within the study, including HCRU, costs, and other socioeconomic factors, with all patients likely receiving treatment from the same facilities and health system procedures. In contrast, multi-institutional RCTs do not usually account for socioeconomic, health system, and regional variables within the randomization schema. Cooperative group trials rarely identify or statistically account for outlier institutions. Socioeconomic, health system and regional variables become more apparent when more diverse populations are included within a clinical trial or with RWD. Social determinants of

health were highly evident during the COVID pandemic in which members of racial and ethnic minority groups, had higher rates of COVID-19 positivity, disease severity, and outcomes (86).

Defined by the 3 types of ECAs, historical, traditional, and contemporaneous, outcomes are dependent on expanded criteria that encompass the socioeconomic variables of a diverse RWE population (Table 2) (54). All three types of ECAs match the individual patient demographics and comorbidities for both the experimental and control arms within a RCT. However, regional variability can reflect access to care, rural vs. urban, academic medical center vs. community hospital or clinic. Most important to outcomes, however, are social determinants of health, including the type of health insurance, race, ethnicity and socioeconomic status.

Patients enrolled on an RCT generally have more access to care and are treated within a healthcare institution or system. Historical ECAs are either used for context as the standard of care for a selected time-period, or are retrospectively matched to the enrolled patients upon completion of a RCT. Especially when derived from datasets that are socioeconomically diverse and that include underserved populations, historical ECAs may not be an accurate cohort for the treatment arm of a clinical trial. In a large study, such as a multi-institutional RCT, the variances among institutions may not be significant enough to impact outcomes or compromise statistical significance. Most multi-institutional studies are conducted within academic medical centers or private institutes that foster medical research. In most cases, there is little difference among the variables associated with socioeconomic determinants of health in multi-institutional RCTs.

Much like the hybrid ECA, the demographics of patients within a traditional ECAs are matched during an interim analysis or at the end of the RCT. The traditional ECA is actively included within the study methods. Depending on the RCT population, the dataset from which the traditional ECA is derived can be specifically selected to include socioeconomic factors in the dataset. The contemporaneous ECA is more granular than the traditional ECA, and is the only control group that can match all the confounding variables of demographics, comorbidities, enrolled at the same time as the experimental arm patient, type of healthcare insurance, type of healthcare system delivery and access to care, and regional parameters. Using a contemporaneous ECA, theoretically, could yield the most exact comparison to the experimental arm of a RCT.

TABLE 2 Comparison of control groups in randomized controlled trials (RCT) by single institution, multi-institution, traditional external control groups, and contemporaneous external control groups.

	RCT Control groups		External control groups		
	Single institution	Multi-institution	Historical	Traditional	Contemporaneous
Individual patient demographics	Y	Y	Y	Y	Y
Individual patient comorbidities	Y	Y	Y	Y	Y
Enrolled at the same time as the experimental arm	Y	Y	N	N	Y
Type of health insurance/Payer	P	N	N	P	Y
Health systems matched	Y	N	N	P	Y
Region matched ¹	Y	N	N	P	Y

Y represents yes; N represents no; P represents perhaps. ¹Regions within the United States: NorthEast/New England [ME, NH, VT, MA, CT, RI], NorthEast/Middle Atlantic [NY, PA, NJ], South/South Atlantic [WV, MD, DE, DC, VA, NC, SC, GA, FL], South/East South Central [KY, TN, AL, MS], South/West South Central [AR, LA, TX, OK], Midwest/East North Central [MI, WI, IL, IN, OH], Midwest/West North Central [MN, IA, MO, ND, SD, NE, KS], West/Mountain [MT, ID, WY, NV, UT, CO, AZ, NM], West/Pacific [WA, OR, CA, AK, HI].

4. Discussion of limitations

The randomized, controlled prospective trial is the gold standard in rigorous science for several reasons. Prospective randomization of patients to a standard of care control and highly structured experimental protocol mitigates (but does not always eliminate) bias and enables for control of certain variables (but not all variables) that might confound a treatment effect (87). One study of RCT trials underlying 143 cancer drug approvals, 17% had suboptimal control arms (87). It can be argued that the use of genomic biomarkers can bias a trial by design in that patient selection strategies are geared toward a more specific hypothesis, so alignment of the hypothesis and study design with methodology is important to mitigate the inherent biases these tools can introduce (88).

Depending on how heterogenous a population is and whether a powered trial can be fully accessioned, a prospective, randomized controlled trial may not be feasible, cost-effective, or practical. For example, the use of pharmacogenomics to guide clopidogrel use is complex. Several biotransformation mechanisms are involved in the pharmacology of clopidogrel: CYP2C19 (1st step), CYP3A4 (2nd step), and CYP2C9 (89), implicating several hundred currently known variants of these three pharmacogenes. Clopidogrel is used for a number of different cardiovascular indications. Commercial genotyping of these pharmacogenes is currently limited to a small number of the most prevalent variants in commercially available tests. Thus, the RCTs that have been conducted for pharmacogenomic-guided clopidogrel use are very narrow use cases involving a single indication and a limited number of variants. The costs of validating each drug-pharmacogene-disease triad permutation of pharmacogenomic-guided use of clopidogrel in an RCT across so many domains would be prohibitive. The use of real-world clinical outcome data from an ECA (e.g., an ungenotyped approach) can serve as a useful, albeit imperfect, comparator to assess utility of pharmacogenomic-guided use of drugs like clopidogrel across many domains, however, limited outcome data can itself result in biases (90). This is a more pragmatic approach to pilot and assess RCT feasibility or as an alternative form of evidence, by enabling the use of existing data instead of the costs of enrolling a participating control cohort. Pooling data is an approach that has been used to confront sample size challenges in single health systems but data governance barriers across organizational barriers make this approach laborious (49, 51, 52).

There are limitations to the utility of ECAs, that vary depending on disease, clinical specialty, and importantly, the outcomes of interest. Sometimes, the clinical outcomes measures may not be in the EMR or measured with research-robust methodologies, or data does not meet GXP (e.g., GCP, GLP, GMP) and CDISC standards for use in regulatory filings, in cases where FDA review of the clinical study is required. Even in instances where the ECA data comes from a clinical trial such as the CRBM model that created digital twins to augment the control group (66), data limitations can limit ECA cohort size. This challenge is especially acute in rare diseases and involving registry data. When using data from an EMR this limitation can arise if data access to multiple health systems is impractical to obtain. ECAs can be a useful clinical study strategy for comparative effectiveness, genomic medicine, and pilot studies where regulatory hurdles are less burdensome. Use of an ECA requires some capacity to account for and mitigate bias, which inherently requires a data set with detailed clinical annotation, in

addition to the capacity to do subanalysis for age, sex, stage of disease and other major determinants of outcomes. Comparisons of the ECA to the treatment arm may not be feasible for all clinical important clinical outcomes as a result of informatics silos or medical practices (5). The entire case-level natural history of disease would be an idealized data source (49), but in reality all data sets have significant blindspots across time and geography (91). The ECA need not be an alternative to a prospective internal control of standard of care (i.e., limited to single arm trials), should a strong validation of external validity be justified, which could make sense for phase 2 trials used to assess the merit of a phase 3 or label expansion program. The FDA provides guidance on the use of external control arms in drug and biologic product development (92).

5. Conclusion

Research paradigms are changing, and the pace of change was accelerated by the realities of conducting clinical trials during the COVID-19 pandemic (93). While control arms within RCTs seek to match the demographic and comorbidities of the experimental arm cohort, randomization procedures rarely account for socioeconomic factors or broader differences in the clinical setting. Important factors include the type of healthcare insurance, healthcare delivery system, access to care, and regional variations. There is a growing body of literature showing that ECAs are a valuable approach to increasing efficiency, reducing patient risk, and accounting for social determinants of health. Various ECA approaches are necessary to help capture RWD in ways that maximize the knowledge gained when a study is conducted. Multiple contemporaneous control arms, that include the disparate social determinants of health, could theoretically be incorporated within a clinical trial to anticipate the real-world results of the new therapeutic in Phase IV analysis. More than any other type of control arm, contemporaneous ECAs allow a granular match with the experimental arm cohort and are likely to become a preferred approach in clinical research as data quality and annotation improve real-world practice. The production of high quality RWD in medical specialties such as pharmacogenomics and oncology have great promise to achieve validation and accelerate implementation of new innovations that can reduce the costs and duration of clinical trials and improve care, while illuminating causes of health disparities.

Author contributions

NJ and MS conceived of the work. NJ and PS made significant contributions to the literature review, research, drafting and revising of the manuscript. KR, GU, LZ, MS, and MO provided critical input on revisions, content, and organization. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Kessler R, Glasgow RE. A proposal to speed translation of healthcare research into practice. *Am J Prev Med.* (2011) 40:637–44. doi: 10.1016/j.amepre.2011.02.023
- Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard — lessons from the history of RCTs. *N Engl J Med.* (2016) 374:2175–81. doi: 10.1056/NEJMms1604593
- Moffat KR, Cannon P, Shi W, Sullivan F. Factors associated with recruitment to randomized controlled trials in general practice: protocol for a systematic review. *Trials.* (2019) 20. doi: 10.1186/s13063-019-3354-z
- Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in Cancer research: barriers, evidence, and strategies. *Am Soc Clin Oncol Educ Book.* (2016) 35:185–98. doi: 10.1200/EDBK_156686
- Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol.* (2020) 12:457–67. doi: 10.2147/CLEP.S242097
- de Melo-Martín I, Sondhi D, Crystal RG. When ethics constrains clinical research: trial design of control arms in “greater than minimal risk” pediatric trials. *Hum Gene Ther.* (2011) 22:1121–7. doi: 10.1089/hum.2010.230
- Schulz A, Ajayi T, Specchio N, de Los RE, Gissen P, Ballon D, et al. Study of intraventricular cerliponase alfa for CLN2 disease. *N Engl J Med.* (2018) 378:1898–907. doi: 10.1056/NEJMoa1712649
- Johnson TB, Cain JT, White KA, Ramirez-Montealegre D, Pearce DA, Weimer JM. Therapeutic landscape for batten disease: current treatments and future prospects. *Nat Rev Neurol.* (2019) 15:161–78. doi: 10.1038/s41582-019-0138-8
- Saini KS, de Las HB, Plummer R, Moreno V, Romano M, de Castro J, et al. Reimagining global oncology clinical trials for the postpandemic era: a call to arms. *JCO Global Oncol.* (2020) 6:1357–62. doi: 10.1200/GO.20.00346
- London AJ, Omotade OO, Mello MM, Keusch GT. Ethics of randomized trials in a public health emergency. *PLoS Negl Trop Dis.* (2018) 12. doi: 10.1371/journal.pntd.0006313
- Araujo DV, Watson GA, Siu LL. The day after COVID-19—time to rethink oncology clinical research. *JAMA Oncol.* (2021) 7:23–4. doi: 10.1001/jamaoncol.2020.4240
- Germaine CGS, Bogaty P, Boyer L, Hanley J, Engert JC, Brophy JM. Genetic polymorphisms and the cardiovascular risk of non-steroidal anti-inflammatory drugs. *Am J Cardiol.* (2010) 105:1740–5. doi: 10.1016/j.amjcard.2010.01.352
- Nabhan C, Choueiri TK, Mato AR. Rethinking clinical trials reform during the COVID-19 pandemic. *JAMA Oncol.* (2020) 6:1327–9. doi: 10.1001/jamaoncol.2020.3142
- The WHOREAFC-TWG. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. *JAMA.* (2020) 324:1330–41. doi: 10.1001/jama.2020.17023
- Spinner CD, Gottlieb RL, Criner GJ, Arribas López JR, Cattelan AM, Soriano Viladomiu A, et al. Effect of Remdesivir vs standard care on clinical status at 11 days in patients with moderate COVID-19: a randomized clinical trial. *JAMA.* (2020) 324:1048–57. doi: 10.1001/jama.2020.16349
- Dequin P-F, Heming N, Meziani F, Plantefève G, Voiriot G, Badié J, et al. Effect of hydrocortisone on 21-day mortality or respiratory support among critically ill patients with COVID-19: a randomized clinical trial. *JAMA.* (2020) 324:1298–1306. doi: 10.1001/jama.2020.16761
- Janiaud P, Axfors C, Ioannidis JPA, Hemkens LG. Recruitment and results reporting of COVID-19 randomized clinical trials registered in the first 100 days of the pandemic. *JAMA Netw Open.* (2021) 4:1193. doi: 10.1001/jamanetworkopen.2021.0330
- Kouzy R, Abi Jaoude J, Garcia CJ, El Alam MB, Taniguchi CM, Ludmir EB. Characteristics of the multiplicity of randomized clinical trials for coronavirus disease 2019 launched during the pandemic. *JAMA Netw Open.* (2020) 3. doi: 10.1001/jamanetworkopen.2020.15100
- Pundi K, Perino AC, Harrington RA, Krumholz HM, Turakhia MP. Characteristics and strength of evidence of COVID-19 studies registered on ClinicalTrials.gov. *JAMA Intern Med.* (2020) 180:1398–1400. doi: 10.1001/jamainternmed.2020.2904
- Garibaldi BT, Wang K, Robinson ML, Zeger SL, Bandeen-Roche K, Wang M-C, et al. Comparison of time to clinical improvement with vs without Remdesivir treatment in hospitalized patients with COVID-19. *JAMA Netw Open.* online, rapid electronic publication (2021) 4:e213071. doi: 10.1001/jamanetworkopen.2021.3071
- Vaughn VM, Yost M, Abshire C, Flanders SA, Paje D, Grant P, et al. Trends in venous thromboembolism anticoagulation in patients hospitalized with COVID-19. *JAMA Netw Open.* (2021) 4. doi: 10.1001/jamanetworkopen.2021.11788
- Dron L, Golchi S, Hsu G, Thorlund K. Minimizing control group allocation in randomized trials using dynamic borrowing of external control data - an application to second line therapy for non-small cell lung cancer. *Contemp Clin Trials Commun.* (2019) 16:100446. doi: 10.1016/j.conctc.2019.100446
- Carrigan G, Bradbury BD, Brookhart MA, Capra WB, Chia V, Rothman KJ, et al. External comparator groups derived from real-world data used in support of regulatory decision making: use cases and challenges. *Curr Epidemiol Reports.* (2022) 9:326–37. doi: 10.1007/s40471-022-00305-9
- FDA. Real-world evidence: Real-world data (RWD) and real-world evidence (RWE) are playing an increasing role in health care decisions. Available at: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (2020)
- FDA. Use of real-world evidence to support regulatory decision-making for medical devices. Guidance for industry and Food and Drug Administration staff. Available at: <https://www.fda.gov/media/99447/download> (2017)
- CDRH. Center for Devices and Radiological Health (CDRH) 2016–2017 Strategic/Priorities/Accomplishments. Available at: <https://www.fda.gov/media/95317/download> (Accessed January 2, 2023) (2016).
- DiMasi JA, Grabowski HG, Hansen RW. The cost of drug development. *N Engl J Med.* (2015) 372:1972. doi: 10.1056/NEJMc1504317
- Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health.* (2017) 20:858–65. doi: 10.1016/j.jval.2017.03.008
- The clinical trial is open. The elderly need not apply. Available at: <https://www.nytimes.com/2018/04/13/health/elderly-clinical-trials.html> (2018).
- Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics.* (2019) 20:273–86. doi: 10.1093/biostatistics/kxx069
- Huss R. *The high Price of failed clinical trials: Time to rethink the model.* Clinical Leader Newsletter (2019) Available at: <https://www.clinicalleader.com/doc/the-high-price-of-failed-clinical-trials-time-to-rethink-the-model-0001>
- Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov.* (2019) 18:495–6. doi: 10.1038/d41573-019-00074-z
- Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp Clin Trials Commun.* (2018) 11:156–64. doi: 10.1016/j.conctc.2018.08.001
- Frank C, Himmelstein DU, Woolhandler S, Bor DH, Wolfe SM, Heymann O, et al. Era of faster FDA drug approval has also seen increased black-box warnings and market withdrawals. *Health Aff.* (2014) 33:1453–9. doi: 10.1377/hlthaff.2014.0122
- Nabhan C, Klink A, Prasad V. Real-world evidence—what does it really mean? *JAMA Oncol.* (2019) 5:781–3. doi: 10.1001/jamaoncol.2019.0450
- FDA. U.S. Food and Drug Administration. Balancing premarket and Postmarket data collection for devices subject to premarket approval. Available at: <http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm393994.pdf> (Accessed January 2, 2023) (2015).
- FDA. U.S. Food and Drug Administration. The least burdensome provisions of the FDA modernization act of 1997: Concept and principles. Available at: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm085999.pdf> (Accessed January 2, 2023) (2019).
- Gagne JJ, Thompson L, O’Keefe K, Kesselheim AS. Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ.* (2014) 349:g6802. doi: 10.1136/bmj.g6802
- Gaddipati H, Liu K, Pariser A, Pazdur R. Rare Cancer trial design: Lessons from FDA approvals. *Clin Cancer Res.* (2012) 18:5172–8. doi: 10.1158/1078-0432.CCR-12-1135
- Nardini C. The ethics of clinical trials. *Eancermedscience.* (2014) 8:387. doi: 10.3332/ecancer.2014.387
- Berger T, Rozovski U, Moshe Y, Yaari S, Frisch A, Hellmann I, et al. Midostaurin in combination with intensive chemotherapy is safe and associated with improved remission rates and higher transplantation rates in first remission—a multi-center historical control study. *Ann Hematol.* (2019) 98:2711–7. doi: 10.1007/s00277-019-03795-8
- Shadman M, Maloney DG, Storer B, Sandmaier BM, Chauncey TR, Smedegaard Andersen N, et al. Rituximab-based allogeneic transplant for chronic lymphocytic

- leukemia with comparison to historical experience. *Bone Marrow Transplant.* (2020) 55:172–81. doi: 10.1038/s41409-019-0660-8
43. Shah AN, Flaum L, Helenowski I, Santa-Maria CA, Jain S, Rademaker A, et al. Phase II study of pembrolizumab and capecitabine for triple negative and hormone receptor-positive, HER2-negative endocrine-refractory metastatic breast cancer. *J Immunother Cancer.* (2020) 8:e000173. doi: 10.1136/jitc-2019-000173
44. William WN Jr, Tsao AS, Feng L, Ginsberg LE, Lee JJ, Kies MS, et al. Single arm, phase II study of cisplatin, docetaxel, and Erlotinib in patients with recurrent and/or metastatic head and neck squamous cell carcinomas. *Oncologist.* (2018) 23:526–e49. doi: 10.1634/theoncologist.2017-0661
45. Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The role of real-world evidence in FDA-approved new drug and biologics license applications. *Clin Pharmacol Ther.* (2022) 111:135–44. doi: 10.1002/cpt.2474
46. Lewis JRR, Kerridge I, Lipworth W. Use of real-world data for the research, development, and evaluation of oncology precision medicines. *JCO Precis Oncol.* (2017) 1:1–11. doi: 10.1200/PO.17.00157
47. Gor AS, Patel KB, Rabara V, Clinton N. The role of real-world evidence (RWE) registries using NGS testing to address cancer health disparities (CHD) in small community oncology practices. *J Clin Oncol.* (2022) 40:e18519. doi: 10.1200/JCO.2022.40.16_suppl.e18519
48. Deverka PA, Douglas MP, Phillips KA. Use of real-world evidence in US payer coverage decision-making for next-generation sequencing-based tests: challenges, opportunities, and potential solutions. *Value Health.* (2020) 23:540–50. doi: 10.1016/j.jval.2020.02.001
49. Silva P, Dahlke DV, Smith ML, Charles W, Gomez J, Ory MG, et al. An idealized Clinicogenomic registry to engage underrepresented populations using innovative technology. *J Pers Med.* (2022) 12:713. doi: 10.3390/jpm12050713
50. Silva P, Jacobs D, Kriak J, Abu-Baker A, Udeani G, Neal G, et al. Implementation of pharmacogenomics and artificial intelligence tools for chronic disease Management in Primary Care Setting. *J Pers Med.* (2021) 11:443. doi: 10.3390/jpm11060443
51. Silva PJ, Ramos KS. *Precision medicine at the academic-industry interface. Precision medicine for investigators, practitioners and providers.* Elsevier; (2020). 545–560
52. Silva PJ, Schaibley VM, Ramos KS. Academic medical centers as innovation ecosystems to address population—omics challenges in precision medicine. *J Transl Med.* (2018) 16:28. doi: 10.1186/s12967-018-1401-2
53. Edelman D, Habermehl C, Schlenk RF, Benner A. Adjusting Simon's optimal two-stage design for heterogeneous populations based on stratification or using historical controls. *Biom J.* (2019) 62:311–29. doi: 10.1002/bimj.201800390
54. Collignon O, Schritz A, Senn SJ, Spezia R. Clustered allocation as a way of understanding historical controls: components of variation and regulatory considerations. *Stat Methods Med Res.* (2020) 29:1960–71. doi: 10.1177/0962280219880213
55. Feinberg BA, Gajra A, Zettler ME, Phillips TD, Phillips EG, Kish JK. Use of real-world evidence to support FDA approval of oncology drugs. *Value Health.* (2020) 23:1358–65. doi: 10.1016/j.jval.2020.06.006
56. Silva PJ, Ramos KS. Chapter 53 – Precision medicine at the academic-industry interface In: J Faintuch and S Faintuch, editors. *Precision medicine for investigators, practitioners and providers.* Cambridge MA: Academic Press (2020). 545–60.
57. Chen EY, Haslam A, Prasad V. FDA acceptance of surrogate end points for Cancer drug approval: 1992–2019. *JAMA Intern Med.* (2020) 180:912–4. doi: 10.1001/jamainternmed.2020.1097
58. Zhao F. Surrogate end points and their validation in oncology clinical trials. *J Clin Oncol.* (2016) 34:1436–7. doi: 10.1200/JCO.2016.66.4581
59. Rupp T, Zuckerman D. Quality of life, overall survival, and costs of Cancer drugs approved based on surrogate endpoints. *JAMA Intern Med.* (2017) 177:276–7. doi: 10.1001/jamainternmed.2016.7761
60. Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Intern Med.* (2015) 175:1992–4. doi: 10.1001/jamainternmed.2015.5868
61. Haier J, Schaefer J. Economic perspective of Cancer care and its consequences for vulnerable groups. *Cancers (Basel).* (2022) 14:3158. doi: 10.3390/cancers14133158
62. Liu F, Demosthenes P. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol.* (2022) 22:287. doi: 10.1186/s12874-022-01768-6
63. Jahanshahi M, Gregg K, Davis G, Ndu A, Miller V, Vockley J, et al. The use of external controls in FDA regulatory decision making. *Ther Innov Regul Sci.* (2021) 55:1019–35. doi: 10.1007/s43441-021-00302-y
64. Burcu M, Dreyer NA, Franklin JM, Blum MD, Critchlow CW, Perfetto EM, et al. Real-world evidence to support regulatory decision-making for medicines: considerations for external control arms. *Pharmacoepidemiol Drug Saf.* (2020) 29:1228–35. doi: 10.1002/pds.4975
65. Chapple AG, Thall PF. Comparison of phase I-II designs with parametric or semi-parametric models using two different risk-benefit trade-off criteria. *Contemp Clin Trials.* (2020) 97:106099. doi: 10.1016/j.cct.2020.106099
66. Fisher CK, Smith AM, Walsh JR, Simon AJ, Edgar C, Jack CR, et al. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci Rep.* (2019) 9:13622. doi: 10.1038/s41598-019-49656-2
67. (CHMP) CfMPfHU. *Qualification opinion for prognostic covariate adjustment (PROCOVA™)* European Medicines Agency (2022).
68. Schröder C, Lawrance M, Li C, Lenain C, Mhatre SK, Fakih M, et al. Building external control arms from patient-level electronic health record data to replicate the randomized IMblaze370 control arm in metastatic colorectal Cancer. *JCO Clin Cancer Inform.* (2021) 5:450–8. doi: 10.1200/CCI.20.00149
69. Zauderer MG. Practical application of real-world evidence in developing cancer therapies. *JCO Clin Cancer Inform.* (2019) 3:1–2. doi: 10.1200/CCI.19.00071
70. Stewart M, Norden AD, Dreyer N, Henk HJ, Abernethy AP, Chrischilles E, et al. An exploratory analysis of real-world end points for assessing outcomes among immunotherapy-treated patients with advanced non-small-cell lung Cancer. *JCO Clin Cancer Inform.* (2019) 3:1–15. doi: 10.1200/CCI.18.00155
71. Aponte J, Silva PJ, Ramos KS. Multianalyte tests in clinical care In: S Ps, editor. *Comprehensive precision medicine.* London, UK: Elsevier (2023)
72. Sim S, Lövrot J, Lindh JD, Bergh J, Xie H. Effect of CYP2C19 and CYP2D6 genotype on tamoxifen treatment outcome indicates endogenous and exogenous interplay. *Pharmacogenomics.* (2018) 19:1027–37. doi: 10.2217/pgs-2018-0089
73. He W, Grassmann F, Eriksson M, Eliasson E, Margolin S, Thorén L, et al. CYP2D6 genotype predicts tamoxifen discontinuation and prognosis in patients with breast Cancer. *J Clin Oncol.* (2020) 38:548–57. doi: 10.1200/JCO.19.01535
74. Shi L, Dong B, Li Z, Lu Y, Ouyang T, Li J, et al. Expression of ER- α 36, a novel variant of estrogen receptor α , and resistance to tamoxifen treatment in breast cancer. *J Clin Oncol.* (2009) 27:3423–9. doi: 10.1200/JCO.2008.17.2254
75. Kelly CM, Juurlink DN, Gomes T, Duong-Hua M, Pritchard KI, Austin PC, et al. Selective serotonin reuptake inhibitors and breast cancer mortality in women receiving tamoxifen: a population based cohort study. *BMJ.* (2010) 340:c693. doi: 10.1136/bmj.c693
76. Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, et al. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health.* (2021) 24:1118–25. doi: 10.1016/j.jval.2021.01.015
77. Carrigan G, Whipple S, Capra WB, Taylor MD, Brown JS, Lu M, et al. Using electronic health records to derive control arms for early phase single-arm lung Cancer trials: proof-of- concept in randomized controlled trials. *Clinical Pharmacology & Therapeutics.* (2019) 107:369–77. doi: 10.1002/cpt.1586
78. Feyman Y, Provenzano F, David FS. Disparities in clinical trial access across US urban areas. *JAMA Netw Open.* (2020) 3. doi: 10.1001/jamanetworkopen.2020.0172
79. Graham GN. Why your ZIP code matters more than your genetic code: promoting healthy outcomes from mother to child. *Breastfeed Med.* (2016) 11:396–397. doi: 10.1089/bfm.2016.0113
80. Zhang W, Roederer MW, Chen W-Q, Fan L, Zhou H-H. Pharmacogenetics of drugs withdrawn from the market. *Pharmacogenomics.* (2012) 13:223–31. doi: 10.2217/pgs.11.137
81. Zhang JY, Zhan J, Cook CS, Ings RM, Breau AP. Involvement of human UGT2B7 and 2B15 in Rofecoxib metabolism. *Drug Metab Dispos.* (2003) 31:652–8. doi: 10.1124/dmd.31.5.652
82. Arehart E, Stitham J, Asselbergs FW, Douville K, MacKenzie T, Fetalvero KM, et al. Acceleration of cardiovascular disease by a dysfunctional prostacyclin receptor mutation: potential implications for cyclooxygenase-2 inhibition. *Circ Res.* (2008) 102:986–93. doi: 10.1161/CIRCRESAHA.107.165936
83. Huddart R, Sangkuhl K, Whirl-Carrillo M, Klein TE. Are randomized controlled trials necessary to establish the value of implementing pharmacogenomics in the clinic? *Clin Pharmacol Ther.* (2019) 106:284–6. doi: 10.1002/cpt.1420
84. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol.* (2019) 16:312–25. doi: 10.1038/s41571-019-0167-7
85. Sun Z, Niman SM, Pagani O, Partridge AH, Azim HA Jr, Peccatori FA, et al. Estimation of historical control rate for a single arm de-escalation study - application to the POSITIVE trial. *Breast.* (2020) 53:1–7. doi: 10.1016/j.breast.2020.05.012
86. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic-review and Meta-analysis. *JAMA Netw Open.* (2021) 4:e2134147. doi: 10.1001/jamanetworkopen.2021.34147
87. Hilal T, Sonbol MB, Prasad V. Analysis of control arm quality in randomized clinical trials leading to anticancer drug approval by the US Food and Drug Administration. *JAMA Oncol.* (2019) 5:887–92. doi: 10.1001/jamaoncol.2019.0167
88. Committee for Medicinal Products for Human Use (CHMP). *Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection.* EMA/446337/2011. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP) (2011).
89. Sangkuhl K, Klein TE, Altman RB. Clopidogrel pathway. *Pharmacogenet Genomics.* (2010) 20:463–5. doi: 10.1097/FPC.0b013e3283385420
90. van der Wouden CH, Böhringer S, Cecchin E, Cheung K-C, Dávila-Fajardo CL, Deneer VHM, et al. On behalf of the ubiquitous pharmacogenomics

consortium. Generating evidence for precision medicine: considerations made by the ubiquitous pharmacogenomics consortium when designing and operationalizing the PREPARE study. *Pharmacogenet Genomics*. (2020) 30:131–44. doi: 10.1097/FPC.0000000000000405

91. Mishra-Kalyani PS, Amiri Kordestani L, Rivera DR, Singh H, Ibrahim A, DeClaro RA, et al. External control arms in oncology: current use and future directions. *Ann Oncol*. (2022) 33:376–83. doi: 10.1016/j.annonc.2021.12.015

92. *Considerations for the design and conduct of externally controlled trials for drug and biological products*. Food and Drug Administration. Center for Drug Evaluation and Research; Center for Biologics Evaluation and Research; oncology Center of Excellence. (2023); Docket- FDA-2022-D-2983.

93. Chevret S, Timsit J-F, Biard L. Challenges of using external data in clinical trials—an illustration in patients with COVID-19. *BMC Med Res Methodol*. (2022) 22:321. doi: 10.1186/s12874-022-01769-5