Pharmaceutical Statistics

WILEY

REVIEW

Do You Want to Stay Single? Considerations on Single-Arm Trials in Drug Development and the Postregulatory Space

Yulia Dyachkova[1] | Cornelia Dunger-Baldauf[2] | Nathalie Barbier[2] | Jenny Devenport[3] | Stefan Franzén[4] | Gbenga Kazeem[5] | Thomas Künzel[3] | Pierre Mancini[6] | Giacomo Mordenti[7] | Knut Richert[8] | Antonia Ridolfi[6] | Daniel Saure[9]

[1]Merck Healthcare KGaA, Darmstadt, Germany | [2]Novartis AG, Basel, Switzerland | [3]F Hoffman-La Roche, Basel, Switzerland | [4]AstraZeneca UK Ltd, Cambridge, UK | [5]GlaxoSmithKline, Stevenage, UK | [6]Sanofi, Paris, France | [7]Daiichi Sanko Europe GmbH, Munich, Germany | [8]Bayer AG, Berlin, Germany | [9]Boehringer Ingelheim Europe GmbH, Ingelheim, Germany

Correspondence: Yulia Dyachkova (yulia.dyachkova@merckgroup.com)

Received: 6 December 2023 | **Revised:** 3 April 2024 | **Accepted:** 23 May 2024

Keywords: causal inference | principled decision-making | single-arm trial

ABSTRACT

Single-arm trials (SATs), while not preferred, remain in use throughout the drug development cycle. They may be accepted by regulators in particular contexts (e.g., in oncology or rare diseases) when the potential effects of new treatments are very large and placebo treatment is unethical. However, in the postregulatory space, SATs are common, and perhaps even more poorly suited to address the questions of interest. In this manuscript, we review regulatory and HTA positions on SATs; challenges posed by SATs to address research questions beyond regulators, evolving statistical methods to provide context for SATs, case studies where SATs could and could not address questions of interest, and communication strategies to influence decision making and optimize study design to address evidence needs.

1 | Introduction

Single-arm trials (SATs)—those in which every included participant receives the same treatment (or more generally intervention)—have been present in medical literature since the beginning of clinical trials. However, SATs often fail to provide conclusive answers to clinical questions. For a relevant historical example, in 1946, a large trial was initiated by the US Veterans Affairs Administration to evaluate the efficacy and safety of streptomycin as a treatment for tuberculosis [1]. Ultimately, this SAT was inconclusive. This contrasted with its contemporary randomized clinical trial (RCT) sponsored by the UK Medical Research Council, generally considered to be the first published instance of a blinded RCT, which demonstrated the efficacy and safety of streptomycin [1]. Nowadays, RCTs are the mainstay of the pharmaceutical industry. Demanded by health authorities and other decision-makers, the double-blind RCT is the gold standard of designs to enable causal conclusions to be drawn about the safety and efficacy of new treatments.

The last decade has seen an evolution in regulations to allow for other types of trial designs, including SATs, to support drug development. The 21st Century Cures Act (Cures Act), signed into law on December 13, 2016, in the USA, is designed to accelerate drug development and bring innovations and new advances faster and more efficiently to the patients who need them [2]. To meet these expectations, the FDA has created a framework for evaluating the potential use of real-world evidence (RWE) to help support some of the approvals of a new indication or to help support or satisfy drug postapproval study requirements [2]. Observational studies, often designed as prospective noninterventional SATs, fall into the category of RWE. SATs can be also designed as interventional based on the study goal and the phase of drug development. While health authorities may accept SATs in certain circumstances, HTA positions, while evolving, seem to view SATs as more problematic [3, 4].

The limitations of SATs as stand-alone evidence are well-described in the statistical literature, by regulatory bodies, and

by other stakeholders. Nonetheless, SATs remain in use across the stages of drug development and especially in the postmarketing space. While many experts are involved in the design and conduct of clinical trials, the trial statistician is a key partner and well-equipped to guide the team and influence choices on the best approach to address scientific questions. In this paper, we will discuss the use of SATs throughout the drug development lifecycle and review statistical and stakeholder perspectives on the deficiencies of SATs for various purposes (Section 2); critically examine why SATs are still considered, particularly in the postregulatory space, via the presentation of some case studies where SATs have been used or not used effectively to address questions of interest (Section 3); provide advice to statisticians who need to influence principled decision making during evidence generation planning (Section 4) or analyze data and communicate results from SATs (Section 5); and conclude with a discussion (Section 6).

2 | Limitations and Guidance on SATs From Various Perspectives

Phase III pivotal trials are usually randomized, controlled, double-blind studies in which both the investigator and patient are unaware of which drug or treatment is being administered [2]. This design has very desirable properties from a statistical perspective. Randomization prevents bias in the allocation of the intervention by creating treatment groups balanced for risk factors for the targeted outcome and has been considered a critical element in establishing a causal relationship between the treatment and health care outcomes [2]. Further, the inclusion of a concurrent control group ensures that many systematic effects of time, including regression to the mean, are balanced and canceled out in the comparison. Blinding removes subjectivity in assessing and reporting of both efficacy and safety evaluations and consequently minimizes bias in the outcome. In SATs, there is generally insufficient information within the study to rule out other factors that might have caused the observed outcome. Additionally, everyone knows what treatment is being given. This knowledge can impact patient selection to participate, the study assessments, dropouts, conduct of the trial, and of course, the outcomes. In other words, without randomization, blinding, and a control group, it is difficult to causally attribute the observed results to the treatment.

If considering an SAT rather than RCT to address study objectives, one needs to be aware of multiple pitfalls that can jeopardize or even invalidate study results. Multifunctional discussions should take place to rule out potential issues. Based on the study objectives, potential deficiencies of an SAT versus an RCT might be even more numerous than the following listed from different perspectives (Table 1).

While RCTs are the gold standard, SATs are used in a variety of contexts throughout the drug development lifecycle [16], even in the regulatory settings (Figure 1). For example, in situations where the disease is rare, or giving alternative treatment is unethical, and the new medication fulfills an unmet medical need, exceptions are made with accelerated approval by FDA [6, 17], conditional approval by EMA [18, 19], and other agencies. The common context of the SAT use includes:

- Early development—signal seeking (change in the right direction), early safety evaluations (e.g., finding maximum tolerated dose), biomarkers identification.
- Natural history—course of a disease over time with respect to demographic, genetic, environmental, and other factors. Identifying candidate/potential biomarkers or subgroups for further investigation. Identify current treatment patterns, standard of care (SoC), and unmet clinical needs. Helping to make pipeline decisions [20].
- Late phase—primarily oncology [21, 22] (e.g., when spontaneous response without intervention is not expected) and rare disease (e.g., in contexts where randomization might be deemed unethical due to lack of SoC) [9, 23].
- Postapproval including medical affairs—including treatment pattern studies: treatment sequencing, dosing, adherence to guidelines and labels. Resource use and costs. Identifying additional risks and benefits (e.g., novel endpoints), performance in RW and special populations underrepresented in RCTs [24]. Many SATs are small and done soon after launch in countries not included or underrepresented in RCTs to provide input into local negotiations of market access, reimbursement, and price.

EMA [7] and FDA [8] issued draft guidance documents in 2023 related to external controls and SATs. These documents highlight well-known limitations and biases of SAT and discuss ways of potentially minimizing them either via external control arm or prespecified statistical methodology. Many other agencies contributed ideas to address specific needs in market access and HTA areas. Table 2 summarizes key guidance documents related to SAT.

3 | Case Studies of Successful and Unsuccessful Use of SAT

Example 1. Regulatory approvals based on an SAT.

There are multiple studies in oncology leading to a successful drug registration based on SAT [32, 33]. For example, in a biomarker-driven subset of patients with RET fusion-positive non–small cell lung cancer (NSCLC) which represents about 1% of the NSCLC patients, that is relatively rare, a new drug, selpercatinib [11], was approved by FDA and EMA in 2020 based on a single-arm phase 1/2 study. Approval of this medication was granted on the basis of an SAT because of the very high observed response rates (85% in patients on first-line treatment [34]), the poor prognosis in this population of patients with advanced or metastatic lung cancer, and the substantially lower response rates without the target therapy (19% in patients in first line on standard chemotherapy [35]). The phase 3 study positive results were published in October 2023 [36], demonstrating that patients have gained access to the innovative drug 3 years earlier due to accelerated/conditional approval provisions.

Example 2. SAT used to extend the label.

Eltrombopag was initially approved in Europe in March 2010 under the trade name Revolade for the treatment of adult

TABLE 1 | Perspectives on single-arm trials.

Statistical perspective	• Cannot support a causal interpretation (there are other plausible explanations that cannot be easily ruled out) ○ Selection bias (outcomes are due to the included patient characteristics rather than the treatment) ○ History (events concurrent with treatment such as changes in treatment patterns, access policies, treatment guidelines) ○ Time-bias (due to the definitions of index dates, treatment initiation, or follow-up periods) [5] ○ Maturation—naturally occurring changes over time (e.g., aging) ○ Regression to the mean—occurs with selection bias (e.g., inclusion criteria require a certain level of disease activity), where change is more likely to occur in one direction independent of treatment ○ A Hawthorne effect: participants might behave differently if they know they are observed • Potentially biased results in the absence of blinding due to subjectivity in evaluation and knowledge of the therapy • Treatment effect (effect size and variability of the observed outcomes) is difficult to interpret without a reference (a proper control group). • External comparators are always different in some way and so may not be able to salvage desired causal conclusions due to residual confounding and lack of consistency. • Transparency and scrutiny of research (e.g., prespecification): inappropriately interpreted analyses increasingly publicly challenged [6]
Regulatory perspective	• Can support approvals only in rare circumstances [7–9], such as high unmet need (e.g., in oncology, hematology, orphan diseases), or predictable control group outcomes (e.g., for monotherapy epilepsy treatments [10] or late line of oncology treatment [11, 12]) • Often requires subsequent definite RCT [9]
Payer perspective	• Comparative effectiveness information is needed to estimate incremental cost-effectiveness ratio (ICER) • Health technology assessment authorities usually do not consider single-arm trials for nonorphan diseases • Many countries' authorities are reluctant to consider or give very low value (Germany and France [4]) to SATs. This prohibits access to medication.
Clinical scientist/academic perspective	• Want a conclusive answer: The results of SAT can be put in context by nonanalytically comparing (eyeballing) to other SAT or appropriate RCT but this has several limitations. • Ascertainment bias due to different data collection standards if using external reference to the treatment effect • Patients in an (external) control group might be systematically different from patients in SAT, especially if historical control is used. This might be due to change in clinical practice, change in assessment methods, unavailability of a biomarker of interest in the control group, other factors that impact the prognosis • Last choice for Cochrane reviews and top journals [13] • Difficult to interpret safety results in the absence of a control arm (see example in Section 3) • Reputational risk [14] • Allegations of inappropriate reasons [14, 15] for conducting trial, especially if no real scientific objective or if it cannot be credibly achieved by trial

patients with primary immune thrombocytopenia (ITP) lasting 6 months or longer from diagnosis and who are refractory to other treatments (e.g., corticosteroids and immunoglobulins).

After the approval, growing RWE on the successful use of thrombopoietin receptor agonists (TPO-RAs) in patients with newly diagnosed and persistent ITP as well as updated ITP treatment guidelines prompted the sponsor to propose a label extension to allow ITP patients who are refractory to other treatments to be treated with eltrombopag irrespective of time from initial diagnosis. At that time, the single-arm phase II study TAPER (CETB115J2411, 2018-000452-18, EudraCT Number, [37]) was ongoing.

This single-arm study evaluated the safety and efficacy of eltrombopag and its ability to induce sustained response after treatment discontinuation. The study enrolled 105 adult patients with ITP who were refractory or relapsed after first-line corticosteroid treatment. There were no restrictions with regards to disease stage (i.e., restrictions to newly diagnosed, persistent or chronic ITP), and approximately half of the enrolled population was patients with newly diagnosed ITP.

To support the proposed label extension in adult ITP, an ad hoc analysis of the ongoing TAPER study was done using a subset of patients ($n = 85$) with at least 6 months of data. This ad hoc

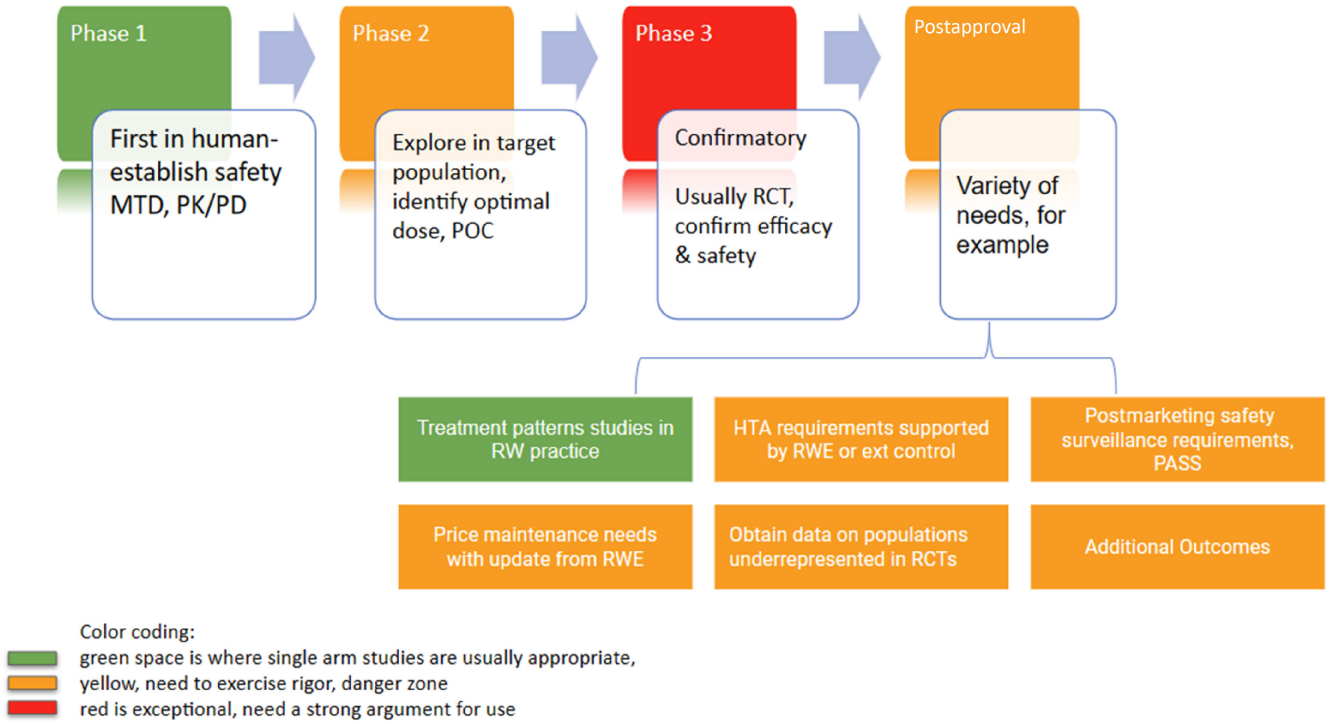

FIGURE 1 | SAT use in drug development cycle.

analysis, together with subgroup analysis of data from the completed ITP registration studies, a literature review and updated popPK/PD analysis indicated effective and safe use of eltrombopag independent of disease stage, and were submitted to EMA. The agency issued a positive opinion in September 2022 (with EC Decision in October 2022) on the proposed label extension and agreed to remove the restriction "lasting 6 months or longer from diagnosis" from the adult ITP indication. The newly approved EU indication for adult ITP patients is as follows: "Revolade is indicated for the treatment of adult patients with primary immune thrombocytopenia (ITP) who are refractory to other treatments (e.g. corticosteroids, immunoglobulins)."

In May 2023, CHMP further granted a positive Opinion of the EU procedure for the addition of TAPER primary analysis data to support efficacy claims related to tapering and sustained response off-treatment. The results are reflected in the Revolade EU Summary of Product Characteristics (SmPC), section 5.1.

Example 3. Delayed approval because of SAT results.

Foradil, a long-acting β_2 agonist formoterol furoate, dry powder inhaler currently used as a bronchodilator in the management of asthma and chronic obstructive pulmonary disease, was submitted for an NDA on June 24, 1997, for Asthma Indication. During the submission process, an SAT's results were published in 1998 [38]. Although the results of this SAT provided good insight in terms of inhaler ease of use by patients, good efficacy rated by most physicians, and improvement in the use of rescue medication following the treatment, 8 deaths among 1380 patients were reported. Among other issues seen in the submission dossier, the FDA became concerned on May 24, 2000 [39], with the postmarketing data available and some fatal exacerbation episodes. Some of these fatal exacerbation episodes were coming from the SAT published in 1998. A new RCT study was initiated in 2000 [40], the EFORA study, with more than 6000 patients and 2 (formoterol) to 1 (control) randomization. The study reported two deaths in each of the treatment arms. The FDA approval was granted on February 16, 2001, with all the open questions solved. The SAT safety results [39] contributed to the 4-year approval delay.

Example 4. Impact of an observational postmarketing SAT.

Not all studies are focused on estimating relative benefit of a new treatment. Understanding disease management and patient care in real-world settings, including drug utilizations (treatment patterns, adherence), acute and long-term safety, and/or effectiveness, is often of importance. Further, a postmarketing study offers the opportunity to include a broader patient population and to estimate their outcomes.

LUMINOUS, a prospective, 5-year, multicenter, observational, noninterventional study to evaluate the long-term safety, effectiveness, treatment patterns, and health-related quality of life of ranibizumab for the treatment of eye diseases in routine clinical practice belongs to this category. The SAT design was chosen as the objective was noncomparative and the vision of patients without treatment is known to decline rapidly. Worldwide more than 30,000 patients were recruited in the indications for which ranibizumab was approved. We review the results for the largest group, 6421 ranibizumab-naïve patients with neovascular age-related macular degeneration [41]. Beyond the requirement of at least one visit per year, visits were scheduled to the discretion of the treating physicians. Patients were treated by injections into the eye according to local ranibizumab labels. The primary effectiveness variable was gain in visual acuity (letters) from the study start. Safety was assessed based on ocular and nonocular

TABLE 2 | Guidance documents related to SAT.

Issued by	Name	Highlights
ICH E10, May 2001 [25]	ICH guidance for industry E10 Choice of Control Group and Related Issues in Clinical Trials (May 2001).	The choice of control group for a trial affects the inferences, ethical acceptability, degree of bias, type and pace of recruitment, relevant endpoints, public and scientific credibility, and acceptability by regulatory authorities. "If the course of a disease were uniform in a given patient population, or predictable from patient characteristics…results of treatment could simply be compared with the known outcome without treatment."
EMA, April 2023 draft [7]	Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorization	Discusses potential issues with SATs, particularly challenges with the estimand components. Particular importance is placed on the prespecification of estimands, data collection, study conduct, data handling, analysis methods, etc. Table 1 provides a number of sources of potential biases along with explanations and potential bias reduction strategies. Having a good external control is a key for mitigating some of the issues.
FDA, February 2023, draft [8]	Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products (guidance for industry)	Suitability of externally controlled trial design warrants case-by-case assessment informed by heterogeneity of disease, preliminary evidence of investigational product, outcome of interest, and goal of the trial (superiority or noninferiority). Particular challenges presented by external control arms created from clinical trials and RWD are discussed. Considerations for assessing comparability of data: time periods, geographic region, diagnostic criteria, prognosis, treatments, other treatment-related factors, time (designation of index date, follow-up periods), intercurrent events, outcome, missing data are provided
Centre for Drug Evaluation (CDE) in China, published in 2023 [26]	Guidelines for Design and Protocol Framework of Real-World Studies of Drugs	Provides step-by-step instructions on the protocol requirements for a real-world study. Each item is supplemented with an explanation of why it is important and risks that need to be mitigated within the context of an RW study. It includes SAT, since 'external control is usually established based on real-world data.'
EUnetHTA 21 (2022) [27]	Methodological guideline D4.3.2 DIRECT AND INDIRECT COMPARISONS	Gives detailed recommendations about how evidence synthesis should be conducted to assess the relative efficacy or effectiveness of a new intervention compared to one or more existing interventions. RCTs are referred to as the gold standard for informing estimates of the new treatment's effectiveness vs comparators. Direct comparisons based on RCTs are clearly preferred. SATs belong to nonrandomized evidence for which very stringent requirements must be fulfilled, including individual patient data of all data sources, prespecified adjustment for potential confounders, and a large treatment effect estimate to rule out a conclusion of effectiveness due to bias.

(Continues)

TABLE 2 | (Continued)

Issued by	Name	Highlights
NICE (Dec 2016) [28]	DSU TECHNICAL SUPPORT DOCUMENT 18: METHODS FOR POPULATION-ADJUSTED INDIRECT COMPARISONS IN SUBMISSIONS TO NICE	Examines methods for population-adjusted indirect comparisons, in which individual patient data (IPD) in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome by the means of the Matching-Adjusted Indirect Comparison (MAIC) and the Simulated Treatment Comparison (STC).
NOMA (updated October 2021) [29]	Guidelines for the submission of documentation for single technology assessment (STA) of pharmaceuticals	Section 6.1.5 has a provision for use of MAIC and STC in case of the need to use data from SAT. Ideally, IPDs are available for at least 1 study to enable adjusted comparisons.
CADTH (May 2023) [30]	Guidance for reporting real-world evidence	Fit for purpose design, transparent reporting with rationale for the choice of design. Readers are referred to Hernan 2020 "Causal Inference. What if" [31]. Studies with causal research should consider modern causal inference framework, such as target trial emulation, to guide their study designs and methods

treatment-emergent adverse events. The mean visual acuity gain was 3.1 letters compared with 11.3 and 7.2 letters in the registration trials.

Mean visual acuity gains by baseline visual acuity ranged from 12.6 (lowest baseline) to −3 (highest baseline), with increasing mean numbers of injections (4.3–5.5). On the other hand, groups with increasing numbers of injections had slightly increasing visual acuity gains. Loading doses, that is, three injections within the first 90 days of treatment, were associated with higher mean vision gain.

The majority of LUMINOUS patients received less than six injections over the course of a year, lower than anticipated with the common treatment regimens used in clinical practice. In the registration trials, patients had been treated monthly. Even under the more flexible treat and extend regimen patients would receive at least seven injections during the first year, suggesting potential underdosing in LUMINOUS. Based on the descriptive results, the authors recommended starting treatment with three consecutive monthly ranibizumab injections (loading doses) and generally avoiding underdosing.

LUMINOUS achieved the objective of providing insight into the safety, effectiveness, and real-world treatment patterns of ranibizumab. As visits were less frequent in LUMINOUS than in RCTs, comparisons of incidences of safety events are difficult to interpret; however, the safety event profiles were similar and no new safety signals were found. In the absence of a comparator the gain in visual acuity can be put into the perspective of, for example, the registration RCT MARINA, where the placebo (sham) group suffered an average letter loss of 10 letters, in line with natural history findings. This suggests a confirmation of the effectiveness of ranibizumab.

Findings involving baseline visual acuity or the number of injections are difficult to interpret as the comparisons are nonrandomized and the number of injections is itself a random variable depending on responses at previous visits. There may be other factors contributing to vision gain. The magnitude of effect observed in LUMINOUS in the context of the registration RCTs is likewise difficult to interpret due to the lack of comparability of the patient population and the treatment conditions. Although the recommendation to increase the number of injections in clinical practice is plausible and appropriate, it is not, strictly speaking, supported by a causal conclusion.

4 | Advice for Statisticians to Influence the Study Design Decision

The study design and planning process is usually a multifunctional collaboration, which hopefully begins with the specification of objectives and continues with the honing of clear, testable, research questions. Sometimes teams will approach statisticians directly with the idea of running an SAT. Table 3 characterizes frequent discussion points for SATs in the postapproval space. Potential responses are provided with the intent to influence stakeholders to define their research questions prior to targeting a particular study design.

Once the research questions are refined, appropriate designs must be evaluated for the context and their ability to answer the questions. Sometimes, after careful consideration, an SAT design is selected. We now consider some reasons for utilizing an SAT to address questions of interest and highlight examples in Table 4. As mentioned in regulatory guidelines, SATs may be an acceptable option in the case of diseases that have high and predictable mortality or progressive morbidity, life-threatening and severely debilitating with an unmet medical need [42]; where placebo treatment is unethical [43] or there is no standard active comparator; and when the uncertainty or variability of control outcomes is small relative to the treatment effect. The use of SAT may be more easily accepted in orphan diseases where there are very few patients in the world available for study, though the preceding criteria still apply.

TABLE 3 | Sample dialogs about SATs in the postregulatory setting.

Request for SAT	Potential response and considerations
We just want to be able to talk about the treatment or drug in a more pragmatic setting.	Yes, can we consider the opportunity we have with this study to answer questions that your customers have? What would they like to know?
We can always say we are confirming what we saw in Ph III	Maybe Ph III is not confirmed in the new study? Study conditions will likely be different, and without a comparator, it will be difficult to explain differences in results. What does "confirming" mean in this context?
Do we need to generate data for an active comparator? There is the risk we do not come out favorably.	This is a valid concern. Can we explore different scenarios and weigh risks vs possible advantages?
This is exploratory. Let's just see what we get.	There must be some knowledge gap or other reason that motivates you to plan the study. Could you explain what would you like to know better about the novel drug?
But we have always done it this way and it was hugely successful	Very interesting. How do you define success? What made the approach a success? Can we go from there?

TABLE 4 | General SAT usage summary.

Goal/objective of the study	Recommended	Pros	Cons
Characterize burden and epidemiology of disease; characterize the patients with unmet medical needs (globally or locally)	Yes	Noncomparative, cheaper, and faster data collection	
Drug early development, dose-finding, signal detection	Yes	Allows early signal detection, preliminary safety signal, maximum tolerated dose	Does not give an estimate of treatment effect or relative safety profile
Evaluating efficacy and safety in a pivotal trial	No	Potentially faster, useful when standard of care does not exist and it is unethical to use placebo	Cannot evaluate relative efficacy or safety. Often requires further trials.
How are newly diagnosed patients being treated in clinical practice? (or generically, treatment patterns)	Yes	Reflect RW practice (e.g., treatment modalities and concomitant medications)	Have to avoid naive comparisons of safety and effectiveness
PASS, safety surveillance	No	It might be that all eligible patients with the indication receive new treatment, no control available.	Cannot directly or reliably compare safety with other cohorts or RTC

Another context in which SATs may be appropriate is when the question or interest is noncomparative. For example, in proof of concept studies, it may be adequate to run an SAT to see if the changes over time occur in the expected direction. Similarly, in early-phase expansion studies in oncology, SATs may be used to quickly study multiple doses or mechanisms. The speed and simplicity of an SAT in these contexts may trump greater certainty. Finally, observational data to evaluate the patient journey, the current burdens of treatment, and clinical practice patterns may be adequately detailed and far more convenient and cost-effective than a formal trial.

However, there are contexts in which the conduct of SATs is potentially debatable. For example, RCTs are often large and quite costly. Therefore, teams may pitch SATs to make projects more economically viable. Some local health authorities might require additional data from an SAT to evaluate efficacy in their own country when the drug has already been approved in larger countries by the usual RCT pivotal trials. While certainly convenient, the interpretation of results is still ambiguous and subject to the statistical and clinical deficiencies noted above. Some postmarketing trials are conducted to provide physicians a chance to gain experience with the new treatment. It is not entirely clear

why a clinical trial is the appropriate mechanism for this, or whether a training program might be more appropriate.

When considering an SAT as a potential design option for the objective, study teams should evaluate alternative designs. For example, a retrospective study using existing registries or RWE databases might be usually both faster and cheaper to achieve the objectives despite data limitations [44]. As an alternative to RCT, pragmatic randomized trials [45] and randomized registry trials [46, 47] have been broadly used in real-world patient populations (i.e., with less restrictive inclusion/exclusion criteria), allowing for more flexible visit schedules, and fewer study assessments–particularly in the postregulatory setting. As such, they may be easier and cheaper to conduct and recruit with potentially higher external validity than typical RCTs without fully sacrificing scientific rigor.

In the situation where the patient population is limited and having a control arm is expected to hinder enrollment or prolong the recruitment time, randomized trials with unbalanced allocation might provide a better solution than an SAT. Loss in power from unequal allocation is small: at the same time, a contemporary, randomly allocated control group within a trial is more valid, and can be supplemented with dynamic borrowing from an external data source to increase power.

5 | Suggestions for the Analysis and Interpretation of SAT

Once an SAT has been executed and analyzed, it must be interpreted with appropriate caution. One of the biggest criticisms of the SAT is overinterpretation of the results, especially in terms of implicit or explicit comparisons and causal inference. Statisticians can help their teams to avoid misleading comparisons and facilitate discussion and appropriate interpretation of the results. Table 5 gives a few examples of naive and improved interpretations of results from SATs.

Because there is no internal control in SATs, external information is usually needed to contextualize the results. In the simplest case, we can identify other relevant studies to which the results can be related [48]. Beyond that, we can attempt more involved analytic comparisons by either trying to match the study to already existing ones using simple standardization, completing matched indirect comparisons (MAIC), or attempting to create matched external control populations. The use of external controls is common in oncology [7, 11, 12, 17, 21, 33, 49], hematology [33, 50, 51], and rare diseases [45]. EMA [7] and FDA [17] provide good summaries of the SAT and give recommendations on selecting control data or designing studies with external control.

FDA provides detailed guidance of constructing contemporary external control arm for pivotal studies [17]. This guidance document as well as the EMA guidance [7] are applicable to many other types of SAT, especially studies with the goal to demonstrate effectiveness and safety. Indeed, in exceptional cases where SAT is the only option, an external control arm has to be used. The use of external information in the analysis or interpretation of an SAT is a crucial design element and should be prespecified in the study protocol [17]. Alternatively, external information may be used to establish a threshold for efficacy that can be demonstrated to fulfill the conditions that support isolating a treatment effect [17, 26].

If one has to construct an external control arm, plan for quantitative bias analysis and report it with the estimates of relative effects. The statistical analysis plan should be finalized before study initiation to avoid modifications and amendments due to unblinded nature of the data in SAT [52]. Potential intercurrent events should be discussed and an estimand framework [53] followed in the study planning stage. Particular attention should be paid to missing data and attrition bias since these are challenging to evaluate in the absence of a contemporary control arm.

Multiple techniques and software solutions are being developed to match, weight or otherwise balance patients between SAT and the external comparator [3, 12, 31]. We will briefly mention the following strategy:

- Unadjusted nonanalytic contextualization relating the results to other studies
- Standardization of outcomes, for example, standardized mortality rates
- Adjusted contextualization using, for example, MAIC [54, 55]
- Contextualization using an external control arm [56] or synthetic arm [57]
- Threshold crossing [58]

Unadjusted nonanalytic contextualization happens when the results from an SAT are compared side-by-side with results from relevant RCTs and other SATs. If the population and design are sufficiently similar, comparison can be meaningful to some extent, but one needs to be aware of the inherent difference between an RCT and SAT, as well as the possibility of differences between SATs that are not obvious from reading the publications or even the protocols. Nonetheless, the unadjusted contextualization is an improvement from interpretation without any additional context.

Standardization and MAIC utilize individual person data (IPD) from the present study and weights that reflect the distribution of variables seen in an external study based on summary statistics of the external study. Standardization is simple to perform and is applied to subgroups of patients while MAIC relies on a logistic regression being applied to the study data to derive individual analysis weights. Both techniques estimate the outcome one would see if the study had the same distribution of patient characteristics as the reference. While standardization is limited to include a small number of such characteristics, MAIC can handle a larger number of them.

External and synthetic control arms have seen increases in popularity and use external data sources (e.g., real data from another RCT, RWD, or synthetically generated data; [59, 60]). The controls can be selected to match the study patients using standard matching techniques such as propensity scores [61]. Controls based on synthetic data have the advantage of maintaining patient privacy and consequently do not necessarily

TABLE 5 | SAT interpretation examples.

Available outcome measure	Naive interpretation	Improved alternative interpretation
Change from baseline	"The non-zero change from baseline demonstrates the real-world effectiveness of drug X."	"The change from baseline on drug X was comparable to that seen in the similar population of the Phase 3 trials (or other similar SAT). This may suggest drug X is effective in this real-world setting. A key limitation is that our study lacks a control group making it difficult to attribute the changes from baseline to drug X. However, we believe..." OR "The change from baseline on drug X was substantially larger than that seen in any placebo group of any clinical trials in similar populations. This suggests drug X may be effective in this setting. A key limitation is that our study lacks a control group so that we rely on informal comparisons to placebo groups in previous trials. However, we believe..."
Evaluation relative to an unstated threshold	"Overall disease activity was very low in patients treated with drug X over the course of Y years, and no new safety signals were observed. This further corroborates the positive risk/benefit ratio..."	"Overall disease activity was low (citation of validated disease activity level (DAL) threshold) in patients treated with drug X over the course of Y years, and no new safety signals were observed. Limitations of our study include that there were no DAL thresholds required for study entry and our design lacks a concurrent control group making it difficult to attribute the results to drug X. However, we believe that based on the low dropout rate, high compliance, and..."
Comparison to some predefined threshold Y is performed	"The response rate was greater than threshold Y, thus, drug X is effective."	"The 95% confidence interval for the response rate on drug X lies above Y. Thus, we conclude that drug X is effective in this setting. The threshold of Y is based on the fact that this is more than twice the response rate seen in previous placebo controlled trials. A limitation of our findings is that these previous trials are more than 5 years old, but we believe..."
Comparison of change from baseline to external controls	"The greater change from baseline than in external controls confirms real-world effectiveness of drug X."	"The greater change from baseline than in the external controls in a propensity score matched (*or other adjustment method that can minimise confounding*) analysis shows the real-world effectiveness of drug X. A limitation of our study is that some key covariates such as Y and Z were not available, which may lead to a bias in a comparison to external controls. However, the observed differences are larger than what could be expected as a result of such confounding."

experience the same challenges for data sharing as real data, making it far easier to co-analyze with the study data. However, external controls based on synthetic data may raise additional questions from health authorities, though positions could evolve as global privacy regulations increase. Both external and synthetic control can be set up to reflect the patients in the SAT

and, therefore, estimate the effect for that population (which is different from MAIC).

Finally, there is a threshold crossing. This is not a statistical method per se, but rather a conceptual framework in which a threshold for the observed outcome is defined such that exceeding it is interpreted as evidence of a causal effect of the treatment. The argument being that crossing the threshold in the absence of a causal effect is considered impossible or sufficiently unlikely [18, 26, 58]. Thresholds of this nature are more credible (even outside of regulatory settings) when they are prespecified in the protocol and analysis plan.

6 | Discussion

SATs have utility in certain contexts: to look for signals in early research and to collect high-quality data about what is happening in clinical contexts, as contrasted with secondary NIS where observations can be characterized (a combination of clinical practice and the quality of the data source), for special regulatory applications in rare disease with objective endpoints, to supplement or contextualize information as part of larger development programs, and so forth. In this manuscript, we have reviewed the contexts in which SATs are used from early development, late phase pivotal trials in molecular-driven oncology treatments and rare disease, market access, and medical affairs. While the disadvantages of SATs are evident throughout drug development, SATs are often put forward as a "low risk" design in the postregulatory space to reduce costs via smaller sample size, to avoid generating data that detracts from pivotal trial success, to facilitate the study of a broader patient population, or to avoid generating data that can be used by or for competitors. However, in the post-regulatory space, some of the most pertinent stakeholder questions to ensure the appropriate adoption of new treatments into clinical practice are comparative in nature. Thus, SATs might be a low-risk/low-reward compromise.

This manuscript also briefly reviews approaches to reduce bias in SATs such as the use of external controls and special analysis methods, which are continually evolving. However, we note that such tools may not be applicable in all contexts, come with some strong assumptions (no hidden confounding and consistency of outcomes [definition and assessment]) and may not fully remove bias inherent in SATs to address certain questions. Therefore, we also provided recommendations of how statisticians can and should try to influence decision-makers to select trial designs best suited to address their questions when the context permits it. The allure of a quick win with a simple design and "good enough data" are formidable foes in science. Statisticians should strive to influence the clarity and precision of the research question to translate it into a scientifically rational design that is appropriate for the context. Ultimately, this may yield the most value and impact.

Author Contributions

The Launch & Lifecycle eSIG agreed upon the idea for the article. Y.D. led the literature review and writing with substantial support from C.D., N.B., and J.D. All authors were involved in all contributions, authoring, and approval of this article.

Acknowledgements

Martin Blogg contributed to early discussions around single-arm trials in medical affairs. Table 5 was inspired by work presented by unpublished work of Bjoern Bornkamp. We also acknowledge the insightful feedback of an anonymous reviewer that improved the final version.

Conflicts of Interest

The authors declare no conflicts of interest. All authors worked together as an informal collaborative cross-company project (unfunded).

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. M. L. Meldrum, "A Brief History of the Randomized Controlled Trial. From Oranges and Lemons to the Gold Standard," *Hematology/Oncology Clinics of North America* 14, no. 4 (2000): 745–760, vii, https://doi.org/10.1016/s0889-8588(05)70309-9.

2. fda.gov, "Framework for FDA's Real-World Evidence Program," 2018, https://www.fda.gov/media/120060/download.

3. D. Patel, F. Grimson, E. Mihaylova, et al., "Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials," *Value in Health* 24, no. 8 (2021): 1118–1125, https://doi.org/10.1016/j.jval.2021.01.015.

4. K. M. Facey, P. Rannanheimo, L. Batchelor, M. Borchardt, and J. de Cock, "Real-World Evidence to Support Payer/HTA Decisions About Highly Innovative Technologies in the EU–Actions for Stakeholders," *International Journal of Technology Assessment in Health Care* 36, no. 4 (2020): 459–468, https://doi.org/10.1017/S026646232000063X.

5. S. Suissa, "Single-Arm Trials With Historical Controls: Study Designs to Avoid Time-Related Biases," *Epidemiology* 32, no. 1 (2021): 94–100, https://doi.org/10.1097/EDE.0000000000001267.

6. T. B. Ribeiro, L. E. Colunga-Lozano, A. P. V. Araujo, C. L. Bennett, I. Hozo, and B. Djulbegovic, "Single-Arm Clinical Trials That Supported FDA Accelerated Approvals Have Modest Effect Sizes and Were at High Risk of Bias," *Journal of Clinical Epidemiology* 148 (2022): 193–195, https://doi.org/10.1016/j.jclinepi.2022.01.018.

7. EMA/CHMP, "Reflection Paper on Establishing Efficacy Based on Single5 Arm Trials Submitted as Pivotal Evidence in a Marketing Authorisation: Considerations on Evidence From Single-Arm Trials (Draft)," April 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing_en.pdf.

8. fda.gov, "Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products," 2023, https://www.regulations.gov/document/FDA-2022-D-2983-0002.

9. S. Goring, A. Taylor, K. Müller, et al., "Characteristics of Non-Randomised Studies Using Comparisons With External Controls Submitted for Regulatory Approval in the USA and Europe: A Systematic Review," *BMJ Open* 9, no. 2 (2019): e024895, https://doi.org/10.1136/bmjopen-2018-024895.

10. J. A. French, S. Wang, B. Warnock, and N. Temkin, "Historical Control Monotherapy Design in the Treatment of Epilepsy," *Epilepsia* 51, no. 10 (2010): 1936–1943, https://doi.org/10.1111/j.1528-1167.2010.02650.x.

11. C. Rolfo, L. M. Hess, M. H. Jen, et al., "External Control Cohorts for the Single-Arm LIBRETTO-001 Trial of Selpercatinib in RET+ Non-Small-Cell Lung Cancer," *ESMO Open* 7, no. 4 (2022): 100551, https://doi.org/10.1016/j.esmoop.2022.100551.

12. S. Popat, S. V. Liu, N. Scheuer, et al., "Addressing Challenges With Real-World Synthetic Control Arms to Demonstrate the Comparative Effectiveness of Pralsetinib in Non-Small Cell Lung Cancer," *Nature Communications* 13, no. 1 (2022): 3500, https://doi.org/10.1038/s41467-022-30908-1.

13. S. Goodacre, "Uncontrolled Before-After Studies: Discouraged by Cochrane and the EMJ," *Emergency Medicine Journal* 32 (2015): 507–508.

14. H. C. Sox and D. Rennie, "Seeding Trials: Just Say "No" [Published Correction Appears in Ann Intern Med. 2008 Sep 16;149(6):439]," *Annals of Internal Medicine* 149, no. 4 (2008): 279–280, https://doi.org/10.7326/0003-4819-149-4-200808190-00012.

15. K. P. Hill, J. S. Ross, D. S. Egilman, and H. M. Krumholz, "The ADVANTAGE Seeding Trial: A Review of Internal Documents," *Annals of Internal Medicine* 149, no. 4 (2008): 251–258, https://doi.org/10.7326/0003-4819-149-4-200808190-00006.

16. A. Li and R. C. Bergan, "Clinical Trial Design: Past, Present, and Future in the Context of Big Data and Precision Medicine," *Cancer* 126 (2020): 4838–4846, https://doi.org/10.1002/cncr.33205.

17. fda.gov, "Clinical Trial Considerations to Support Accelerated Approval of Oncology Therapeutics (Draft)," 2023, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-considerations-support-accelerated-approval-oncology-therapeutics.

18. EMA, "PRIME: Priority Medicines," accessed November 12, 2023, https://www.ema.europa.eu/en/human-regulatory/research-development/prime-priority-medicines.

19. EMA, "Conditional Marketing Authorisation," accessed November 12, 2023, https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/conditional-marketing-authorisation.

20. M. Hashmi, J. Rassen, and S. Schneeweiss, "Single-Arm Oncology Trials and the Nature of External Controls Arms," *Journal of Comparative Effectiveness Research* 10, no. 12 (2021): 1052–1066, https://doi.org/10.2217/cer-2021-0003.

21. B. Mintzes and A. Vitry, "Flawed Evidence Underpins Approval of New Cancer Drugs," *BMJ* 366 (2019): l5399, https://doi.org/10.1136/bmj.l5399.

22. S. Englert, F. Mercier, E. A. Pilling, et al., "Defining Estimands for Efficacy Assessment in Single Arm Phase 1b or Phase 2 Clinical Trials in Oncology Early Development," *Pharmaceutical Statistics* 22, no. 5 (2023): 921–937, https://doi.org/10.1002/pst.2319.

23. M. Jahanshahi, K. Gregg, G. Davis, et al., "The Use of External Controls in FDA Regulatory Decision Making," *Therapeutic Innovation & Regulatory Science* 55, no. 5 (2021): 1019–1035, https://doi.org/10.1007/s43441-021-00302-y.

24. fda.gov, "Postmarketing Approaches to Obtain Data on Populations Underrepresented in Clinical Trials for Drugs and Biological Products," 2023, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarketing-approaches-obtain-data-populations-underrepresented-clinical-trials-drugs-and?utm_medium=email&utm_source=govdelivery.

25. EMA/CHMP, "ICH Topic E 10: Choice of Control Group in Clinical Trials," 2001, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf.

26. Center for Drug Evaluation (CDE), China, "Guidelines for Design and Protocol Framework of Real-World Studies of Drugs (for Trial Implementation)," 2023.

27. EUnetHA, "D4.3.2 DIRECT AND INDIRECT COMPARISONS (version1)," 2022, https://www.eunethta.eu/wp-content/uploads/2022/08/EUnetHTA-21-Deliverable-D4.3.2-Methodological-Guideline-on-Direct-and-indirect-comparisons-V1.0.pdf.

28. D. M. Phillippo, A. E. Ades, S. Dias, et al., "NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE," 2016, https://www.sheffield.ac.uk/nice-dsu/tsds/population-adjusted.

29. NOMA, "Guidelines for the Submission of Documentation for Single Technology Assessment (STA) of Pharmaceuticals," accessed November 12, 2023, https://www.legemiddelverket.no/globalassets/documents/english/public-funding-and-pricing/documentation-for-sta/guidelines-20.05.2020.pdf.

30. CADTH, "Guidance for Reporting Real-World Evidence," 2023, https://www.cadth.ca/guidance-reporting-real-world-evidence.

31. M. A. Hernán and J. M. Robins, *Causal Inference: What If* (Boca Raton: Chapman & Hall/CRC, 2020).

32. T. B. Ribeiro, C. L. Bennett, L. E. Colunga-Lozano, A. P. V. Araujo, I. Hozo, and B. Djulbegovic, "Increasing FDA-Accelerated Approval of Single-Arm Trials in Oncology (1992 to 2020)," *Journal of Clinical Epidemiology* 159 (2023): 151–158, https://doi.org/10.1016/j.jclinepi.2023.04.001.

33. A. J. Hatswell, G. Baio, J. A. Berlin, A. Irs, and N. Freemantle, "Regulatory Approval of Pharmaceuticals Without a Randomised Controlled Study: Analysis of EMA and FDA Approvals 1999-2014," *BMJ Open* 6, no. 6 (2016): e011666, https://doi.org/10.1136/bmjopen-2016-011666.

34. A. Drilon, G. R. Oxnard, D. S. W. Tan, et al., "Efficacy of Selpercatinib in RET Fusion-Positive Non-Small-Cell Lung Cancer," *The New England Journal of Medicine* 383, no. 9 (2020): 813–824, https://doi.org/10.1056/NEJMoa2005653.

35. L. Gandhi, D. Rodríguez-Abreu, S. Gadgeel, et al., "Pembrolizumab Plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer," *The New England Journal of Medicine* 378, no. 22 (2018): 2078–2092, https://doi.org/10.1056/NEJMoa1801005.

36. C. Zhou, B. Solomon, H. H. Loong, et al., "First-Line Selpercatinib or Chemotherapy and Pembrolizumab in RET Fusion-Positive NSCLC," *The New England Journal of Medicine* 389, no. 20 (2023): 1839–1850, https://doi.org/10.1056/NEJMoa2309457.

37. N. Cooper, W. Ghanima, N. Vianelli, et al., "Sustained Response off Treatment in Eltrombopag-Treated Patients With ITP Who Are Refractory or Relapsed After First-Line Steroids: Primary Analysis of the Phase II TAPER Trial," *HemaSphere* 6 (June 2022): 193–194, https://doi.org/10.1097/01.HS9.0000844060.38812.e6.

38. A. M. Clauzel, M. Molimard, V. Le Gros, E. Lepère, N. Febvre, and F. B. Michel, "Use of Formoterol Dry Powder Administered for Three Months via a Single-Dose Inhaler in 1,380 Asthmatic Patients," *Journal of Investigational Allergology & Clinical Immunology* 8, no. 5 (1998): 265–270.

39. fda.gov, "Foradil (Formoterol Fumarate Inhalation) Aerolizer," Approval Date: February 16, 2001, accessed November 12, 2023, https://www.accessdata.fda.gov/drugsatfda_docs/nda/2001/20831_Foradil.cfm.

40. C. Brambilla, V. Le Gros, I. Bourdeix, and Efficacy of Foradil in Asthma (EFORA) French Study Group, "Formoterol 12 Microg BID Administered via Single-Dose Dry Powder Inhaler in Adults With Asthma Suboptimally Controlled With Salmeterol or On-Demand Salbutamol: A Multicenter, Randomized, Open-Label, Parallel-Group Study," *Clinical Therapeutics* 25, no. 7 (2003): 2022–2036, https://doi.org/10.1016/s0149-2918(03)80202-3.

41. F. G. Holz, M. S. Figueroa, F. Bandello, et al., "Ranibizumab Treatment in Treatment-Naive Neovascular Age-Related Macular Degeneration: Results From LUMINOUS, a Global Real-World Study," *Retina* 40, no. 9 (2020): 1673–1685, https://doi.org/10.1097/IAE.0000000000002670.

42. fda.gov, "FDA Guidance for Industry: Rare Diseases—Common Issues in Drug Development," February 2019 (Revision 1), https://www.fda.gov/media/120091/download.

43. D. Sissoko, C. Laouenan, E. Folkesson, et al., "Experimental Treatment With Favipiravir for Ebola Virus Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea [Published Correction Appears in PLoS Med. 2016 Apr;13(4):e1002009. Dortenzio, Eric [Corrected to D'Ortenzio, Eric]] [Published Correction Appears in PLoS Med. 2016 Jun;13(6):e1002066]," *PLoS Medicine* 13, no. 3 (2016): e1001967, https://doi.org/10.1371/journal.pmed.1001967.

44. EMA, "Guideline on Registry-Based Studies," October 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf-0.

45. M. S. Barnish and S. Turner, "The Value of Pragmatic and Observational Studies in Health Care and Public Health," *Pragmatic and Observational Research* 8 (2017): 49–55, https://doi.org/10.2147/POR.S137701.

46. M. S. Lauer and R. B. D'Agostino, Sr., "The Randomized Registry Trial—The Next Disruptive Technology in Clinical Research?" *The New England Journal of Medicine* 369, no. 17 (2013): 1579–1581, https://doi.org/10.1056/NEJMp1310102.

47. B. Karanatsios, K. H. Prang, E. Verbunt, J. M. Yeung, M. Kelaher, and P. Gibbs, "Defining Key Design Elements of Registry-Based Randomised Controlled Trials: A Scoping Review," *Trials* 21, no. 1 (2020): 552, https://doi.org/10.1186/s13063-020-04459-z.

48. J. Lim, R. Walley, J. Yuan, et al., "Minimizing Patient Burden Through the Use of Historical Subject-Level Data in Innovative Confirmatory Clinical Trials: Review of Methods and Opportunities," *Therapeutic Innovation & Regulatory Science* 52, no. 5 (2018): 546–559, https://doi.org/10.1177/2168479018778282.

49. P. S. Mishra-Kalyani, L. Amiri Kordestani, D. R. Rivera, et al., "External Control Arms in Oncology: Current Use and Future Directions," *Annals of Oncology* 33, no. 4 (2022): 376–383, https://doi.org/10.1016/j.annonc.2021.12.015.

50. A. Rambaldi, J. M. Ribera, H. M. Kantarjian, et al., "Blinatumomab Compared With Standard of Care for the Treatment of Adult Patients With Relapsed/Refractory Philadelphia Chromosome-Positive B-Precursor Acute Lymphoblastic Leukemia," *Cancer* 126, no. 2 (2020): 304–310, https://doi.org/10.1002/cncr.32558.

51. M. C. Ozelo, J. Mahlangu, K. J. Pasi, et al., "Valoctocogene Roxaparvovec Gene Therapy for Hemophilia A," *The New England Journal of Medicine* 386, no. 11 (2022): 1013–1025, https://doi.org/10.1056/NEJMoa2113708.

52. S. Kent, M. Salcher-Konrad, S. Boccia, et al., "The Use of Nonrandomized Evidence to Estimate Treatment Effects in Health Technology Assessment," *Journal of Comparative Effectiveness Research* 10, no. 14 (2021): 1035–1043, https://doi.org/10.2217/cer-2021-0108.

53. EMA/CHMP, "ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials.".

54. D. M. Phillippo, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams, and N. J. Welton, "Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal," *Medical Decision Making* 38, no. 2 (2018): 200–211, https://doi.org/10.1177/0272989X17725740.

55. J. E. Signorovitch, V. Sikirica, M. H. Erder, et al., "Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research," *Value in Health* 15, no. 6 (2012): 940–947, https://doi.org/10.1016/j.jval.2012.05.004.

56. M. Burcu, N. A. Dreyer, J. M. Franklin, et al., "Real-World Evidence to Support Regulatory Decision-Making for Medicines: Considerations for External Control Arms," *Pharmacoepidemiology and Drug Safety* 29, no. 10 (2020): 1228–1235, https://doi.org/10.1002/pds.4975.

57. K. Thorlund, L. Dron, J. J. H. Park, and E. J. Mills, "Synthetic and External Controls in Clinical Trials—A Primer for Researchers," *Clinical Epidemiology* 12 (2020): 457–467, https://doi.org/10.2147/CLEP.S242097.

58. H. G. Eichler, B. Bloechl-Daum, P. Bauer, et al., ""Threshold-Crossing": A Useful Way to Establish the Counterfactual in Clinical Trials?" *Clinical Pharmacology and Therapeutics* 100, no. 6 (2016): 699–712, https://doi.org/10.1002/cpt.515.

59. A. Gonzales, G. Guruswamy, and S. R. Smith, "Synthetic Data in Health Care: A Narrative Review. PLOS Digit," *Health* 2, no. 1 (2023): e0000082, https://doi.org/10.1371/journal.pdig.0000082.

60. H. U. Burger, C. Gerlinger, C. Harbron, et al., "The Use of External Controls: To What Extent Can It Currently Be Recommended?" *Pharmaceutical Statistics* 20, no. 6 (2021): 1002–1016, https://doi.org/10.1002/pst.2120.

61. N. Loiseau, P. Trichelair, M. He, et al., "External Control Arm Analysis: An Evaluation of Propensity Score Approaches, G-Computation, and Doubly Debiased Machine Learning," *BMC Medical Research Methodology* 22, no. 1 (2022): 335, https://doi.org/10.1186/s12874-022-01799-z.