Visualized Cancer Medicine 2024, **5**, 5
© The Authors, published by EDP Sciences, 2024
https://doi.org/10.1051/vcm/2024006

Available online at:
https://vcm.edpsciences.org

PERSPECTIVE ARTICLE OPEN ACCESS

Quasi-experimental design for external control arm studies alongside single arm trials for regulatory purposes

John Bian* and Chao-Nan Qian

Guangzhou Concord Cancer Center, 9 Ciji Road, Guangzhou, Guangdong 510555, PR China

Received 4 February 2024, Accepted 2 May 2024, Published online 10 June 2024

Abstract – To accelerate new drug, biologic, and medical device development and to improve efficiency of delivery of the latest breakthroughs of innovative, life-saving treatments to millions of patients, single-arm trial (SAT) applications of rare diseases or conditions supplemented by their external control arm (ECA) evidence for regulatory approvals have been surging since 2016. However, there have been increasing concerns over potential biases threatening the internal validity of these applications from regulatory authorities, payers, and research community. There are two main sources of potential biases. The first source is heterogeneity between two arms at the level of patients, and the second one at the level of systems (e.g., two entirely different sets of hospitals from which patients in a SAT and patients in an ECA are drawn separately). The currently commonly used study design is a post-intervention measurement only design that though mitigating the first source of bias, is utterly unable to control for the second one. This perspective article will propose a quasi-experimental design as an alternative that may mitigate the second source of bias, aiming to improve the internal validity of SAT and ECA studies. We will start summarizing the two main sources of biases that may impede the causal inference of these studies. Two approved therapies supported by SAT and ECA studies will be used as an example to illustrate these biases in detail. We will then introduce the intuition of the quasi-experimental design, underlying assumptions and data requirements, and empirical strategies for estimating interventional effects. We will conclude this article by discussing caveats of applying this alternative design for SAT and ECA studies.

Key words: Food and drug administration, Single-arm trial, External control arm, Standard-of-care, Quasi-experimental design.

Background

Single-arm trial (SAT) applications supplemented by evidence from their accompanied external control arm (ECA) studies for regulatory approvals have been surging since the passage of the 21st Century Cures Act of 2016, part of which was to help accelerate new drug, biologic, and medical device development and improve efficiency of delivery of the breakthroughs of innovative, life-saving treatments to millions of patients [1, 2]. Specifically, a SAT alongside an ECA (SAT/ECA) study has been deemed necessarily for investigating efficacy and safety of an intervention for treating rare diseases or conditions of which adequacy sample size is unattainable for a randomized controlled trial (RCT), or from which patients are inflicted tremendous suffering because of lack of effective treatments in the current practice.

However, applications of SAT/ECA studies for regulatory purposes have been controversial. While RCTs remain as the gold standard for examining a cause-effect relationship for regulatory approvals, comparative evidence from SAT/ECA studies is subject to potential biases that may impede causal inference [1, 3]. The study design currently, commonly adopted in almost all SAT/ECA studies is a head-to-head comparison using a post-intervention measurement only design. (A SAT and an ECA, in the absence of a RCT, are non-equivalent or imbalanced.) As a result, this design may suffer potential threats to its internal validity (see a summary in Table 1). To preserve the scientific rigors of SAT/ECA studies, regulatory authorities, including U.S. Food and Drug Administration (FDA), European Medicines Agency (EMA), and Chinese National Medical Products Administration (NMPA) have clarified their stands and issued guidance on use of real-world data for SAT/ECA studies [4–6].

Using two approved therapies partially supported by two SAT/ECA studies as an example, this perspective article will firstly identify the two main sources of potential biases in these two studies, both of which employed a post-intervention measurement only design. We will then propose a quasi-experimental design as an alternative that will mitigate one of the two sources of biases. In the end, we will discuss caveats of using this alternative design for SAT/ECA studies.

*Corresponding author: Johnbian514@gmail.com

Table 1. Two main sources of biases threatening internal validity of SAT/ECA studies.

Patient-level	Inconsistent measurement of endpoints (e.g., uneven follow-up time of long-term progression events and overall survival)/prognostic risk factors between SATs and ECAs.
	Non-random missingness of data.
	Differences in observed baseline characteristics that make it challenging to define SoC for ECAs as a comparator.
	Differences in unobserved baseline characteristics.
	Use of historical (vs. concurrent) ECAs* in which practice may deviate from that in SATs.
System-level	Underlining heterogeneity among the systems**
	−Population-level disease burdens.
	−Practice style, experience, reimbursement policy, clinical guideline.

* Historical ECAs refer to patients in SATs and ECAs are drawn from difference time periods.
** For example, one system may consist of a set of hospitals from which SAT patents are drawn whereas the other consists of an entirely different set of hospitals from which ECA patients are drawn.
SAT: single arm trial; ECA: external control arm; SoC: standard-of-care.

Comparison of two study designs

Post-intervention measurement only design

This design incorporates a non-equivalent standard-of-care (SoC) ECA into an interventional SAT. A straightforward, yet naïve, estimate of an effect (Δ) of this intervention with or without adjustment for patient-level covariates may be potentially biased, shown in scenario A (Figure 1A). The biases may come from heterogeneity between the two arms at the patient and/or system level (Table 1). Most statistical efforts for mitigating biases in this design focus on adjustment for patient-level heterogeneity. However, this design does not allow for any meaningful adjustment for system-level heterogeneity.

We discuss here three scenarios (Figures 1B–1D) in which no bias or biases of the naïve effect estimated in Figure 1A may occur. Assume a cancer ECA/SAT study in which (1) the patient-level heterogeneity is well adjusted, (2) the endpoints of interest prior to the intervention such as complete response (CR) in the SoC are hypothetically known to both arms, and (3) the system-level differences between the two arms would have remained time-invariant (or unchanged over the entire pre/post-intervention period) if the intervention had been absent. (Note that the unit of analysis is a patient, meaning that there is no repeated measurement at the patient level.) The degree of bias may highly depend on potential differences in pre-intervention measurement of CR in the SoC between the two arms. In Figure 1B, there is likely no bias because the prior CR rates between the SAT and ECA were the same. In Figures 1C and 1D, however, the naïve effect (Δ) in Figure 1A would be subject to biases at the system level largely because the prior CR rates presumably differed.

Axicabtagene ciloleucel and tisagenlecleucel, an autologous chimeric antigen receptor (CAR) T-cell therapy, were approved for relapsed or refractory diffuse large B-cell lymphoma after two or more prior lines of therapy. The approvals for the CAR-T products were primarily based on the evidence from the SAT/ECA studies [7–10]. Two corresponding ECA studies were used as supplemental comparative effectiveness evidence for their regulatory approvals, both of which applied a post-intervention measurement only design [9, 10]. Table 2 briefly summaries the two SAT/ECA studies, the strategies for mitigating biases at the patient-level, and potential residual confounding biases. Using the tisagenlecleucel as an example, the observed endpoints of the ECA patients drawn from the LYSARC could have been drastically different from those of the SAT patients drawn from non-LYSARC in the absence in a RCT. As purposely illustrated in Figures 1C and 1D, the reported efficacy could potentially be under/over-estimated without taking account of time-invariant differences in effectiveness between the two arms prior to the intervention.

Quasi-experimental design

This alternative design is to eliminate any biases as a result of time-invariant system-level differences that cannot be controlled in the post-intervention measurement only design. Though it may be novel to SAT/ECA studies, this design has been widely used in healthcare evaluation studies to adjust for time-invariant differences between interventions and controls [11–13].

Compared to a post-intervention measurement only design, this design (or pre/post-intervention measurement with a non-equivalent control group) incorporates pre-intervention measurement to SAT/ECA studies, similarly showed in Figures 1B–1D [14]. The intuition of this design is that the pre-intervention differences in outcomes of SoC between the two arms (or systems) are appropriate estimates of what the post differences in the outcomes of intervention vs. SoC would have been if the intervention had not occurred. (A visualized justification for a quasi-experimental design in SAT/ECA studies is introduced in Video 1.) Table 3 summarizes assumptions and data requirements for a quasi-experimental design. In addition, we generally assume high quality of data (e.g., consistent measurement and non-missingness) used for the SAT/ECT studies as a *priori*.

Empirical estimation strategies include unadjusted or adjusted approaches, using a difference-in-differences (DD) specification [14]. An unadjusted estimate of the intervention effects (Δ) is illustrated in Table 4. Multivariable regression models may be used for additional adjustment. A modeling specification may be considered as follows: regress an outcome

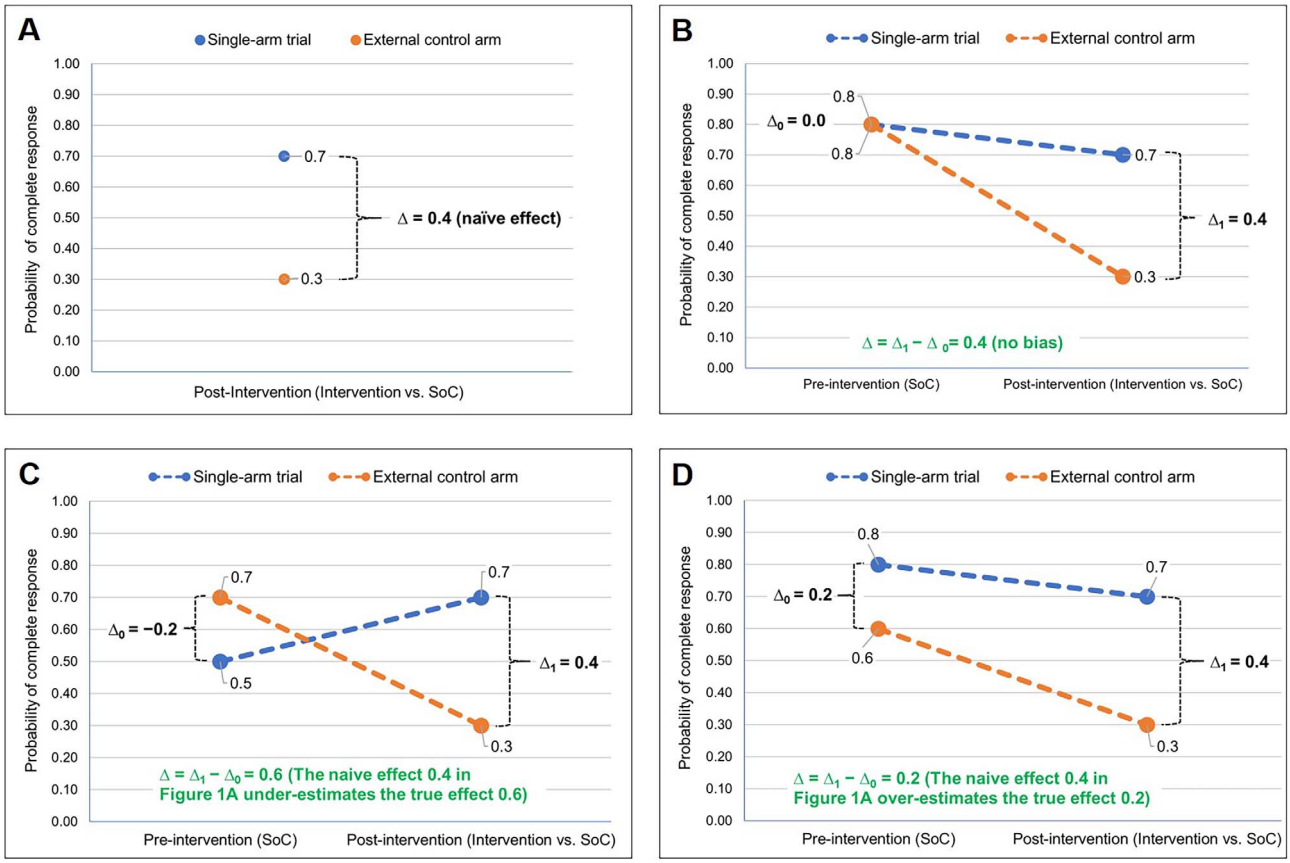


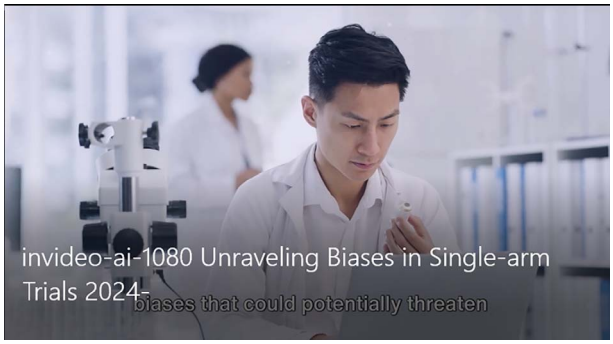
Figure 1. Comparisons of two different study designs under various scenarios. A, post-intervention mean complete response comparison (naïve scenario A). B, mean complete response with hypothetical pre-intervention measurement (scenario B). C, mean complete response with hypothetical pre-intervention measurement (scenario C). D, mean complete response with hypothetical pre-intervention measurement (scenario D).

Table 2. Two approved CAR-T therapies using SAT/ECA studies for relapsed or refractory diffuse large B-cell lymphoma with ≥2 prior lines of therapy.

	Axicabtagene ciloleucel	Tisagenlecleucel
SAT	ZUMA-1 study (n = 101) [6]	JULIET study (n = 115) [7]
	Almost 2 dozen of medical centers in U.S. and Israel.	Over a dozen of medical centers from U.S. and other countries.
ECA	SCHOLAR-1 study (n = 636) [8]	CORAL study (n = 297) [9]
	Observational cohorts from two large medical centers and participants of two previous phase-III RCTs.	Observational cohorts from the Lymphoma Academic Research Organization (LYSARC) in France.
Key endpoints	Complete response, overall response, progression events, and overall survival.	
Bias mitigation	Multivariable regression including propensity score analysis.	
	Sensitivity analysis of definition of SoC.	
Residual confounding biases		
Patient-level	Unable to control time trends due to use of a historical ECA.	
	Concern over unobserved confounders.	
	Non-random missingness in ECAs.	
System-level	Unable to adjust because of use of the post-intervention measurement only design.	

CAR-T: chimeric antigen receptor T-cell; SAT: single arm trial; ECA: external control arm; SoC: standard-of-care.

(e.g., binary CR) on (1) an interaction term of pre/post time period dummy and SAT/ECA dummy (of key interest that captures the intervention effect), (2) patient-level characteristics, (3) system-level time-varying characteristics (e.g., changes in practice experience) if available, (4) SAT/ECA dummy (i.e., controlling for all time-invariant system-level heterogeneity), and (5) pre/post time period dummy (i.e., controlling for secular trends affecting SAT and ECA patients similarly).



Video 1. The necessity of quasi-experimental design in clinical trials. This video was generated by using a commercially available artificial intelligent platform Invideo AI. https://vcm.edpsciences.org/10.1051/vcm/2024006#V1

In the tisagenlecleucel SAT/ECA study shown in Table 2, the investigators would have adopted a quasi-experimental design in practice. They may use four groups of patients instead of two by collecting two additional sets of endpoints along with corresponding covariates, one for each of the two arms, during the pre-intervention period. The resulting estimated effects in the DD specification could alleviate the regulatory concern over any potential system-level, time-invariant biases arising from the post-intervention measurement only design.

Discussion

A quasi-experimental design that has been commonly applied to evaluate health policies and large-scale population-level interventions may also potentially strengthen the internal validity of SAT/ECA studies, compared to a post-intervention measurement only design. This design is especially useful in eliminating time-invariant heterogeneity at the system level, concerned by regulatory authorities, payers, and research community.

However, readers need to be aware of some limitations in applications of this design. Firstly, this design is still subject to biases, compared to a well-designed and executed RCT. For example, unobserved patient-level as well as time-varying system-level confounders may lead to biased estimates, and secular trends (e.g., cross-system changes in clinical guideline or reimbursement policies) that affect a SAT and an ECA differently may also introduce additional biases. Second, significantly more resources may be required to apply a quasi-experimental design to conduct SAT/ECA studies largely because of the requirement for increased sample size and of more complicated statistical analysis (e.g., estimates of the interaction terms). While the increased costs of a quasi-experimental design could be a constraint to many SAT/ECA studies, it may be a good practice to at least perform some unadjusted comparisons of key effectiveness outcomes and prognostic risk factors between the two arms prior to interventions. This type of ad hoc comparisons may provide trial sponsors and decision makers with additional insights of likelihood and magnitude of potential biases from a post-intervention measurement only design.

As SAT/ECA studies have become integral part of new drug, biologic, and device applications for regulatory approvals, the quasi-experimental design, an alternative to a post-intervention measurement only design, should be considered to strengthen the scientific rigors of future SAT/ECA studies.

Table 3. Assumptions and data requirements for quasi-experimental design.

Assumptions
−Underlying differences among systems from which patients of SAT/ECA studies are drawn remain unchanged over time.
−Any secular changes (e.g., clinical guidelines) over time affect the SATs and ECAs similarly.
Data requirements
−The systems from which SAT patients are drawn are different from those from which ECA patients are drawn; patients drawn within-system prior to and post intervention were different patients.
−Similar (if not identical) eligibility criteria (e.g., prognostic risk factors) are applicable to all patients.
−Except the SAT group, all information on the other 3 groups (i.e., the ECA with post-intervention measurement, ECA with pre-intervention measurement, and SAT with pre-intervention measurement) is extracted retrospectively via chart reviews.
−Concurrent timing (i.e., the pre-intervention period is similar between a SAT and an ECA; so is the post-intervention period).

Table 4. Illustration of quasi-experimental design with DD specification for estimating intervention effect (Δ) in a cancer study.

	Pre-intervention period	Post-intervention period	DD
SAT at the system level	CR_{10}	CR_{11}	
ECA at the system level	CR_{00}	CR_{01}	
1st difference	$\Delta_0 = CR_{10} - CR_{00}$	$\Delta_1 = CR_{11} - CR_{01}$	$\Delta = \Delta_1 - \Delta_0$

SAT: single arm trial; ECA: external control arm; CR: (average) complete response (0/1); DD: difference-in-differences.

Acknowledgements

The authors would like to thank Professor Ming Matthew Wang at Grand Rapids Community College, Michigan, USA, for his assistance in generating the video of this article by using an artificial intelligent platform.

Funding

This research received no external funding.

Conflict of interest

Authors declare that they have no competing interests.

Data availability statement

This article has no associated data generated and/or analyzed.

Author contribution statement

Study conception and design: John Bian.

Data collection, analysis and interpretation of results: John Bian and Chao-Nan Qian.

Manuscript preparation: John Bian and Chao-Nan Qian.

Ethics approval

Ethical approval was not required.

References

1. Jaksa A, Louder A, Maksymiuk C, et al. A comparison of 7 oncology external control arm case studies: Critiques from regulatory and health technology assessment agencies. Value Health. 2022;25(12):1967–1976.
2. Wang XM, Dormont F, Lorenzato C, et al. Current perspectives for external control arms in oncology clinical trials: Analysis of EMA approvals 2016–2021. J Cancer Policy. 2023;35:100403.
3. Lambert J, Lengline E, Porcher R, et al. Enriching single-arm clinical trials with external controls: Possibilities and pitfalls. Blood Adv. 2023; 7(19):5680–5690.
4. FDA. Considerations for the design and conduct of externally controlled trials for drug and biological products guidance for industry. Available at: https://www.fda.gov/media/164960/download.
5. EMA. Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing 6 authorisation: 7 Considerations on evidence from single-arm trials. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing-authorisation_en.pdf.
6. Li P, Su Wang S, Chen YW, Use of real-world evidence for drug regulatory decisions in China: current status and future directions. Ther Innov Regul Sci. 2023;57(6):1167–1179.
7. Neelapu SS, Locke FL, Bartlett N, et al. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. N Engl J Med. 2017;377(26):2531–2544.
8. Schuster SJ, Bishop MR, Tam CS, et al. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma. N Engl J Med. 2019;380(1):45–56.
9. Neelapu SS, Locke FL, Bartlett N, et al. Comparison of 2-year outcomes with CAR T cells (ZUMA-1) vs salvage chemotherapy in refractory large B-cell lymphoma. Blood Adv. 2021;5(20):4149–4155.
10. Maziarz RT, Zhang J, Yang H, et al. Indirect comparison of tisagenlecleucel and historical treatments for relapsed/refractory diffuse large B-cell lymphoma. Blood Adv. 2022;6(8):2536–2547.
11. Dimick JB, Ryan AM, Methods for evaluating changes in health care policy: The difference-in-differences approach. JAMA. 2014;312(22): 2401–2402.
12. Bian J, Cristaldi KK, Summer AP, et al. Associations of a school-based, asthma-focused telehealth program with emergency department visits among children enrolled in South Carolina Medicaid. JAMA Pediatr. 2019;173(11):1041–1048.
13. Bian J, Chen B, Hershmen D, et al. Effects of FDA boxed warning of erythropoietin-stimulating agents on utilization and adverse outcome. J Clin Oncol. 2017;35(17):1945–1951.
14. Miller CJ, Smith SN, Pugatch M, Experimental and quasi-experimental designs in implementation research. Psychiatry Res. 2020;283:112452 .

Cite this article as: Bian J & Qian C-N. Quasi-experimental design for external control arm studies alongside single arm trials for regulatory purposes. Visualized Cancer Medicine. 2024; **5**, 5.