Literature Review of Statistical Methods Comparisons Through Simulations When Using External Control Arm for Regulatory or HTA Submissions

Binglin Yue,[1] Kaifeng Lu,[1] Sheng Xu,[2] Shanmei Liao[2]

[1]BeiGene USA, Inc, San Mateo, CA, USA; [2]BeiGene (Shanghai) Co, Ltd, Shanghai, China

INTRODUCTION

- The use of external control arms (ECAs) to augment or use in place of a concurrent internal control in randomized trials can be particularly useful in studies for which patient recruitment may be difficult (eg, rare diseases) or an ethical treatment comparator is not available[1,2]
- In February 2023, the US Food and Drug Safety Administration released draft guidance on externally controlled trials for drug and biological products[3]
  - The draft guidance highlights the importance of developing an appropriate statistical plan in advance, accounting for bias, and assessing comparability between study arms
- However, there was no consensus regarding the selection of the appropriate statistical methods when constructing an ECA

OBJECTIVE

- To perform a targeted literature search of statistical methods used in the analysis of ECA-supported trials and identify the most robust approaches

METHODS

- For identification of statistical methods used in ECA-supported trials, a targeted literature search was conducted on Google and PubMed using the search string "external control statistical method propensity score Bayesian"
  - Articles published in English between 2011 and July 2022 were considered; congress abstracts were excluded
  - A total of 122 articles were retrieved and manually curated for relevance

RESULTS

Targeted literature search

- Of the 122 articles identified in our literature search, 15 statistical articles suitable for summary were selected (**Table 1**)

Table 1. Summary of Targeted Literature Search

Citation	Method	Outcomes	Example/case	Conclusion
Baron et al. *J Biopharm Stat*. 2022[5]	Novel method that integrates PS and Bayesian divide-and-conquer techniques (3 priors used: double hierarchical prior, robust mixture prior, or power prior) to combine stratum-specific parameters and estimate the parameter of interest	Versatile	Simulation	▪ When there was an imbalance in the covariates of the external and current trial data, only the proposed method using a hierarchical prior approach resulted in a smaller bias compared with the nonstratified versions ▪ When covariates of the external and current trial data were balanced, the nonstratified approaches performed better
Lin and Lin. *J Biopharm Stat*. 2022[6]	Review of PS-based methods under the Bayesian framework, including recommendations for reporting in clinical studies	Versatile	Cardiovascular and oncology examples	▪ Different scenarios incorporating external data with or without RCTs are discussed: using RWD as prior information, as augmented controls, as sensitivity analysis, and as a standalone source of evidence ▪ Incorporating PS for evidence synthesis under a Bayesian framework is a promising approach to leverage RWD for clinical studies ▪ Challenges and limitations of this approach include data quality, selection bias, confounding bias, transportability bias, and computational complexity
Lin et al. *Pharm Stat*. 2022[7]	2 Methods to ensure exchangeability based on either PS matching or Bayesian approach with discounting by coefficient of overlap	Not outcome driven	Simulations	▪ The proposed methods can produce more realistic and informative priors that reflect the variability and uncertainty in the historical data, while maintaining exchangeability with the current data
Sawamoto et al. *Pharm Stat*. 2022[8]	Bayesian adaptive randomization design that incorporates PS-matched historical controls to adjust for covariate imbalance	Time-to-event outcomes (eg, PFS)	Simulation	▪ The proposed design can achieve comparable or better power and type I error rate than conventional designs while reducing sample size and bias ▪ Practical guidance on how to implement the proposed design in real trials was provided
Wang et al. *J Biopharm Stat*. 2022[4]	PS adjustment methods (eg, stratification, weighting, matching) integrated with Bayesian priors (eg, power prior, commensurate prior) for augmented controls	Binary	Simulation and oncology example	▪ PS adjustment integrated with Bayesian commensurate priors can be a useful tool for augmenting control designs when historical control data are available and relevant
Wang et al. *arXiv* (preprint). 2022[9]	Method that integrates PS adjustment and Bayesian dynamic borrowing using power prior	Continuous or binary	Simulation and oncology example	▪ The proposed method is a useful tool for augmenting small or imbalanced RCTs with historical control data using Bayesian methods while accounting for covariate imbalance and heterogeneity
Liu et al. *Stat Med*. 2021[10]	Novel method that integrates stratification on the PS and Bayesian meta-analytic-predictive prior	Versatile	Simulation and oncology example	▪ The proposed method better accounted for heterogeneity between external and current trial data compared with the PS-power prior method
Roychoudhury et al. *Stat Med*. 2020[11]	Bayesian meta-analytic approach to leverage historical control data based on a robust hierarchical model for piecewise exponential time-to-event data, which allows for flexible modeling of the hazard function over time and accounts for heterogeneity and uncertainty across historical sources	Time-to-event endpoint (eg, survival)	Oncology examples	▪ The proposed method is a novel and flexible framework that can handle various sources of historical data with different levels of relevance and compatibility ▪ Practical guidance for implementing the method and evaluating its performance was provided
Lin et al. *Pharm Stat*. 2019[12]	2 Matching schemes based on PS estimated through generalized boosted methods to incorporate external data into Bayesian analysis of clinical trials with disproportionate allocation	Binary	Antibacterial drug development example	▪ The proposed method can be useful for designing and analyzing trials with augmented control groups, especially when internal data are limited or imbalanced
Lim et al. *Ther Innov Regul Sci*. 2018[13]	Review of frequentist and Bayesian approaches as used in drug studies that included historical control data	N/A	N/A	▪ High-level summary of statistical approaches targeted toward an industry and regulatory audience
Lin et al. *Pharm Stat*. 2018[14]	2 PS-matching methods (pair matching and nearest-neighbor matching adjusted by caliper) to augment current control data from an RCT with those from a historic control	Binary	Antibacterial drug development example	▪ Both PS-matching schemes to augment control data can reduce bias in estimating treatment response and increase efficiency compared with random sampling of historic data
Zhao et al. *Health Serv Outcomes Res Methodol*. 2016[15]	PS-matching and Bayesian commensurate prior integrated method	Binary	Simulation and antiretroviral example	▪ This method has utility for adaptive borrowing of data from observational studies to augment data from an RCT
Viele et al. *Pharm Stat*. 2014[16]	Comparison and review of methods for borrowing historical control data	Binary	Simulation	▪ The borrowing behavior of different methods is compared (by power, type I error, and MSE) based on changes in user-controlled parameters
Hobbs et al. *Bayesian Anal*. 2012[17]	Extension of commensurate power priors[18] for general linear and general linear mixed models for Gaussian and non-Gaussian responses	Versatile	Simulation and oncology examples	▪ These methods can improve the bias-variance tradeoff when using data from multiple historic sources
Hobbs et al. *Biometrics*. 2011[19]	Multiple classes of hierarchical Bayesian models that incorporate a measure of how commensurate historical Gaussian data are with current Gaussian data (eg, commensurate power priors)	Continuous	Simulation and oncology examples	▪ Adaptive borrowing in Gaussian settings based on commensurability between current and historical data can obtain more power than no borrowing while controlling for type I error

MSE, mean squared error; N/A, not applicable; PFS, progression-free survival; PS, propensity score; RCT, randomized controlled trial; RWD, real-world data.

- Overall, articles discussed or proposed ECA construction methods that focused on propensity score (PS), Bayesian methods with different priors, or integrated approaches
  - However, most articles did not examine which approach was the most robust
- We identified 1 comprehensive study by Wang et al[4] that compared the properties of PS ± Bayesian methods in different combinations; the findings of this study are summarized below

Case study of Wang et al.[4]

- PS ± Bayesian integrated approaches in addition to 2 naive methods (no borrowing; full borrowing) were evaluated in varying combinations (**Table 2**)

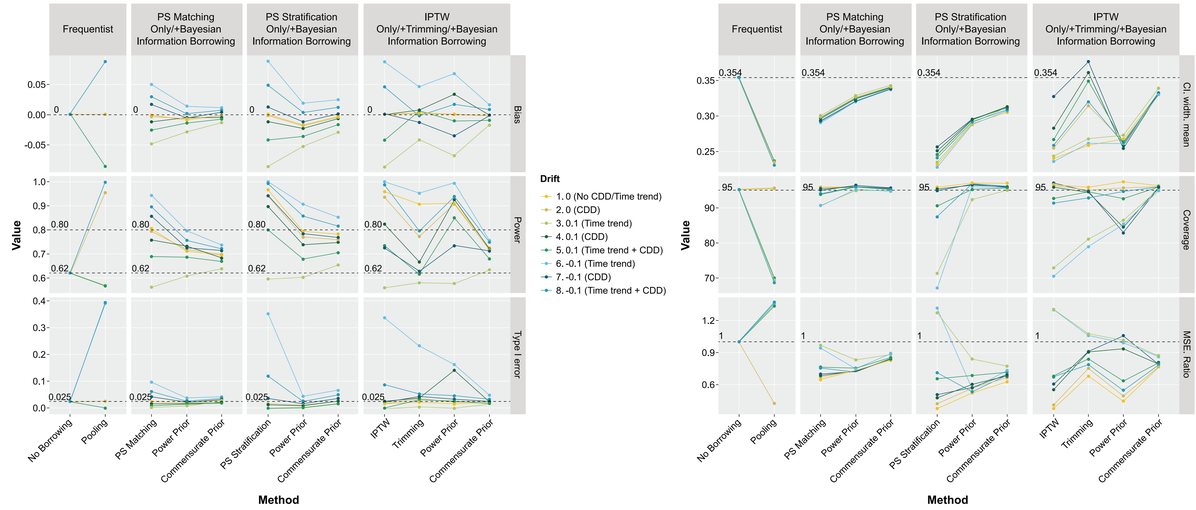
Table 2. Summary of Methods Evaluated[4]

	PS methods				Bayesian information borrowing	
	PSM	PSS	IPTW	+ Trimming	PP	CP
Naive methods						
No borrowing						
Full borrowing (pooling)						
PS methods or two-stage integrated methods						
PS matching methods	■				■	■
PS stratification methods	■	■			■	■
IPTW methods	■	■	■	■	■	■

Each row represents a different set of methods evaluated.
CP, commensurate prior; IPTW, inverse probability of treatment weighting; PP, power prior; PS, propensity score; PSM, propensity score matching; PSS, propensity score stratification.

- The performance of these different approaches was compared using simulated data of a phase 2 randomized (2:1) controlled trial of treatment group E (n=80) vs control group CD (n=40) and a historical control group CH (n=300; all from 1 study)
- The impact on treatment effect estimation was assessed across multiple simulation scenarios
  - Different treatment effect sizes
  - Drift in outcome from time trend, covariate distribution difference (CDD), or both
  - Unmeasured vs measured confounding
- Common criteria to compare the different methods, including bias, type I error, and power, were assessed for each scenario (**Figure 1**)
- For the nonfrequentist approaches in simulations when there was no drift in treatment outcome (**Figure 1**, scenarios 1 and 2)
  - Estimates were almost unbiased when PS modeling alone or combined with commensurate prior was used
  - Simulated type I errors were close to 0.025, mean squared error ratios were all <1, and coverage was ≥95
  - Mean CI widths were lower and greater power was obtained for the PS-only methods vs the 2-stage approaches
  - The most power was achieved using inverse probability of treatment weighting (IPTW) only or PS stratification only
- For the nonfrequentist approaches in simulations when there was drift in treatment outcome (**Figure 1**, scenarios 3-8)
  - Biases tended to be greater when PS-only vs 2-stage methods were used; however, the biases approached 0 in scenarios 4 and 7 when drift was attributed to CDD alone
    - Drift from time trend, whether alone or in combination with CDD (scenarios 3, 5, 6, and 8), resulted in larger biases
  - Type I error was generally lower with the 2-stage vs PS-only methods
    - Type I error was especially pronounced when drift from time trend alone was negative (scenario 6) and PS-stratification only or IPTW-only methods were used
  - Mean CI widths were lower for the PS-matching and PS-stratification only methods vs the 2-stage approaches
    - Greater variation in mean CI width was observed across scenarios for the IPTW-only and IPTW + trimming approaches
  - Coverage was lowest when there was drift from time trend alone (scenarios 3 and 6) and PS-stratification only or IPTW-only methods were used

Figure 1. Performance of Statistical Approaches Across Multiple Simulation Scenarios[4]



Wang X, Suttner L, Jemielita T, Li X. Propensity score-integrated Bayesian prior approaches for augmented control designs: a simulation study. *J Biopharm Stat*. 2022;32(1):170-190. Reprinted by permission of the publisher Informa UK Limited trading as Taylor & Francis Ltd, http://www.tandfonline.com
CDD, covariate distribution difference; IPTW, inverse probability of treatment weighting; MSE, mean squared error; PS, propensity score.

CONCLUSIONS

- There is a lack of consensus regarding which statistical method is best for constructing ECAs, particularly when outcomes differ between the current and historical trial(s)
- This literature review provided valuable information on different statistical methods used to construct ECAs, such as PS, Bayesian with different priors, or integrated approaches
- The analysis of Wang et al[4] demonstrated that PS-Bayesian integrated methods tended to result in lower bias and type I error than PS-only methods when outcome distributions between the current control and historical control were similar
- As highlighted by Wang et al,[4] these approaches, whether PS only or 2 stage, are not recommended for scenarios in which treatment efficacy substantially differs between the current study and the ECA
  - Higher response rates in the historical vs current control may lead to inaccurate conclusions that the treatment does not have an effect
- The findings of Wang et al[4] suggest that trial-specific simulations can help find a "sweet spot" that balances all criteria to help devise a statistical approach for ECA-supported trials

REFERENCES

1. Jahanshahi M, et al. *Ther Innov Regul Sci*. 2021;55(5):1019-1035.
2. Hall KT, et al. *Clin Pharmacol Ther*. 2021;109(2):343-351.
3. US Food and Drug Administration. Accessed March 27, 2023. https://www.fda.gov/media/164960/download.
4. Wang X, et al. *J Biopharm Stat*. 2022;32(1):170-190.
5. Baron E, et al. *J Biopharm Stat*. 2022;32(1):75-89.
6. Lin J, Lin J. *J Biopharm Stat*. 2022;32(1):53-74.
7. Lin J, et al. *Pharm Stat*. 2022;21(2):327-344.
8. Sawamoto R, et al. *Pharm Stat*. 2022;21(5):1074-1089.
9. Wang J, et al. *arXiv*. Preprint posted online October 4, 2022.
10. Liu M, et al. *Stat Med*. 2021;40(22):4794-4808.
11. Roychoudhury S, Neuenschwander B. *Stat Med*. 2020;39(7):984-995.
12. Lin J, et al. *Pharm Stat*. 2019;18(2):223-238.
13. Lim J, et al. *Ther Innov Regul Sci*. 2018;52(5):546-559.
14. Lin J, et al. *Pharm Stat*. 2018;17(5):629-647.
15. Zhao H, et al. *Health Serv Outcomes Res Methodol*. 2016;16(3):154-171.
16. Viele K, et al. *Pharm Stat*. 2014;13(1):41-54.
17. Hobbs BP, et al. *Bayesian Anal*. 2012;7(3):639-667.
18. Hobbs BP, et al. *Biometrics*. 2011;67(3):1047-1056.

DISCLOSURES

BY, KL, SX, and **SL** are all employees of BeiGene and own stock in BeiGene

ACKNOWLEDGMENTS

Medical writing support was provided by Jenna M. Gaska, PhD (ArticulateScience, LLC), and was supported by BeiGene Co, Ltd