Evaluating Bayesian Borrowing Methods for Treatment Effect Extrapolation: A Simulation-Based Study

Authors: Marie Génin, MSc[1], Tristan Fauvel, PhD[1], **Antoine Movschin, MSc[1]**, Billy Amzal, MBA, MPH, MSc, PhD[1]
Affiliation: [1]Quinten Health, Paris, France

MSR91
ISPOR
Europe
2025

INTRODUCTION

- In some contexts, such as paediatrics or rare disease, direct evidence generation may be challenging. Indirect evidence from different settings is sometimes available and can help inform the research question.
- Bayesian borrowing methods are increasingly used to extrapolate treatment effect from a source context to a target context[1].
- However, there is limited understanding of how these characteristics are influenced by key factors such as the drift between source and target treatment effects[2].

OBJECTIVES

- **Evaluate and compare the performance of Bayesian methods for treatment effect extrapolation under realistic trial scenarios through a comprehensive simulation study.**
- **Provide guidance for method selection in small population or extrapolation settings (e.g., pediatrics, rare diseases).**

METHODS

- A **large-scale simulation study** was performed under **multiple realistic scenarios** inspired from 6 **real use cases**[4] (see table 1 for details on the use cases and table 2 for varying parameters in the simulation study).
- **Various state-of-the-art methods**[4] were implemented to extrapolate the treatment effect (see table 3).
- The quality of the extrapolation was assessed via a **comprehensive set of metrics**[4]:
  - **Frequentists operating characteristics**: Type I error (TIE), Power, Mean Squared Error (MSE), Precision (half-width of the 95% credible interval - CI), Coverage of the 95% CI.
  - **Prior effective sample size** (ESS).

Table 1: Use cases inspiration for scenarios

Use case	Therapeutic area	Outcome type
Botox	Lower limb spasticity	continuous
Dapagliflozin	Type II diabetes	continuous
Aprepitant	Postoperative nausea	binary
Belimumab	Systemic lupus erythematosus	binary
Teriflunomide	Multiple sclerosis	time-to-event
Mepolizumab	Severe asthma	recurrent events

Table 2: Varying parameters in the simulation study

Parameter	Variations
Type of outcome	Continuous, Binary, Time-to-event, Recurrent events (see use cases)
Target sample size	$N_T = N_S, N_S/2, N_S/4, N_S/6$
Treatment effect magnitude	Consistent, partially consistent, null
Drift in treatment effect	Ranging from no to large drift (up to treatment effect in the source study)
Ratio of variance between studies	$\sigma_T/\sigma_S = 1$, $\sigma_T/\sigma_S = 2$

Table 3: Implemented extrapolation methods in the simulation study

Extrapolation methods	Description
Test-then-pool (TtP) (equivalence/difference test)	Performs a preliminary test for study consistency and pools data only if no significant heterogeneity is detected
Conditional Power Prior (CPP)	Borrows information from the source study through a power prior with a fixed weight, calibrated to preserve nominal type I error
P-value based Power Prior (p-PP)	Modulates the amount of borrowing according to the similarity between studies, using a function of the p-value for consistency
Empirical Bayes Power Prior (EBPP)	Estimates the power parameter directly from the data via maximum likelihood, allowing data-driven borrowing adaptivity
Normalized Power Prior (NPP)	Scales the power prior by a normalization constant to ensure a proper posterior distribution while controlling borrowing strength
Commensurate Power Prior (Com.PP)	Links source and target parameters through a hierarchical prior that adaptively downweights borrowing when between-study drift is large
Robust Mixture Prior (RMP)	Combines an informative prior (from the source) with a vague prior in a mixture model, enabling automatic discounting in case of prior-data conflict

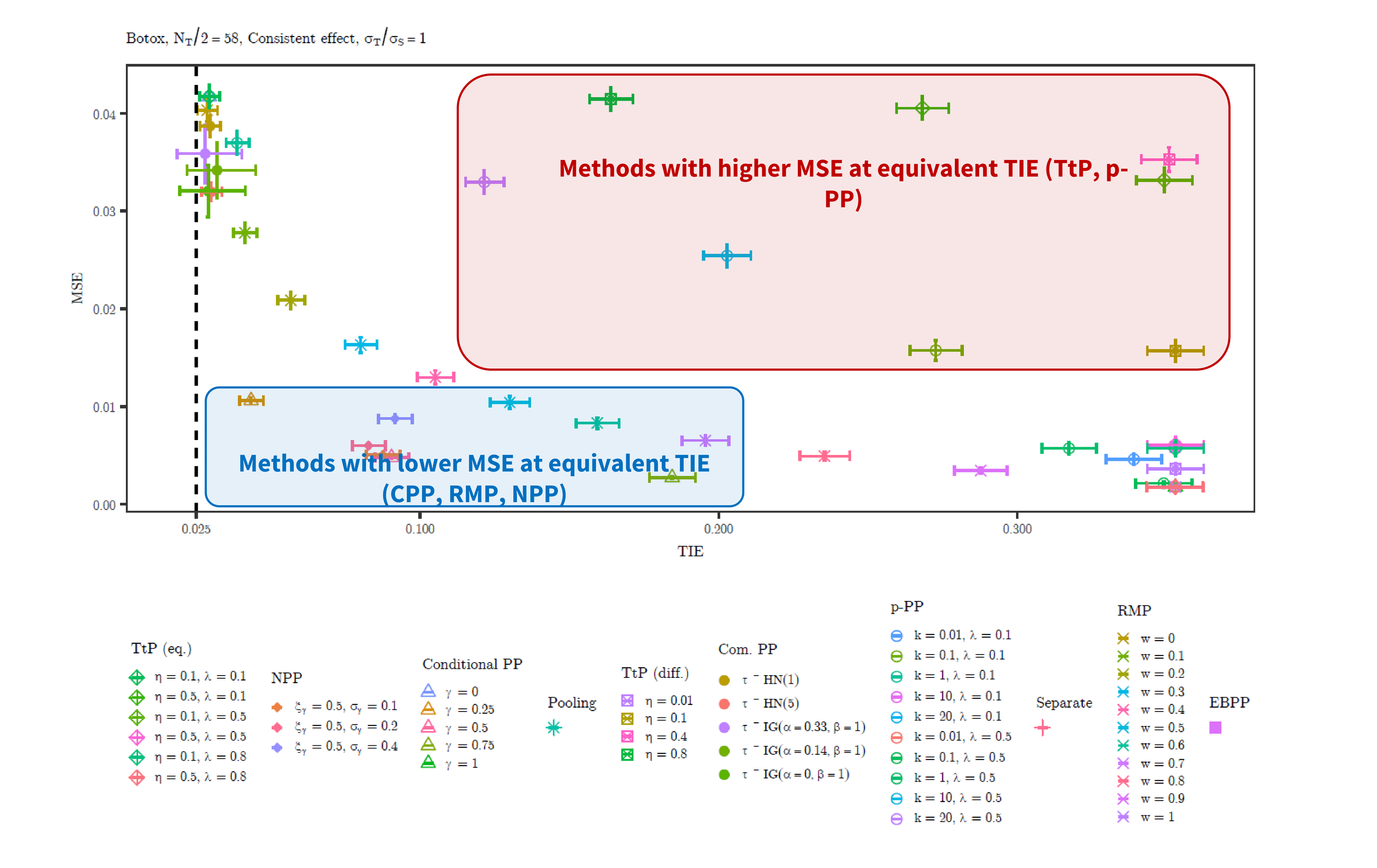
RESULTS

- We performed a systematic comparison of the main Bayesian borrowing methods in a unified framework .
- All Bayesian borrowing methods induce type I error inflation, and do not allow for type I error control.
- Overall, CPP and RMP seem to perform better, while p-PP and TtP yield lower performance.
- Because of the uncertainty associated with the performance of these methods, it is recommended to run extensive simulation studies tailored to the problem and data at hand, to understand their sensitivity to data drift and prior specification.

Bias and precision

- Static or adaptive borrowing can reduce MSE, especially if the drift is small.
- With borrowing, MSE is increased in case of drift and reduced in the absence of drift.
- Inconsistent treatment effects can induce bias; dynamic borrowing approaches can partially mitigate this risk by reducing the ESS (i.e. the amount of borrowed information).
- However, the adaptiveness of these methods was limited: they rarely drop borrowing entirely event with large drift.
- Overall, CPP and RMP achieved better MSE at equivalent type I error, while TtP and p-PP seemed to incur higher MSE.

Figure 3: MSE vs TIE for all methods for the Botox use case



For most methods, the MSE decreases as the TIE increases. The best performing methods are RMP, NPP and CPP. TtP and p-PP achieve suboptimal performance, with higher MSE at equivalent TIE. Overall, TIE increases with higher levels of borrowing.

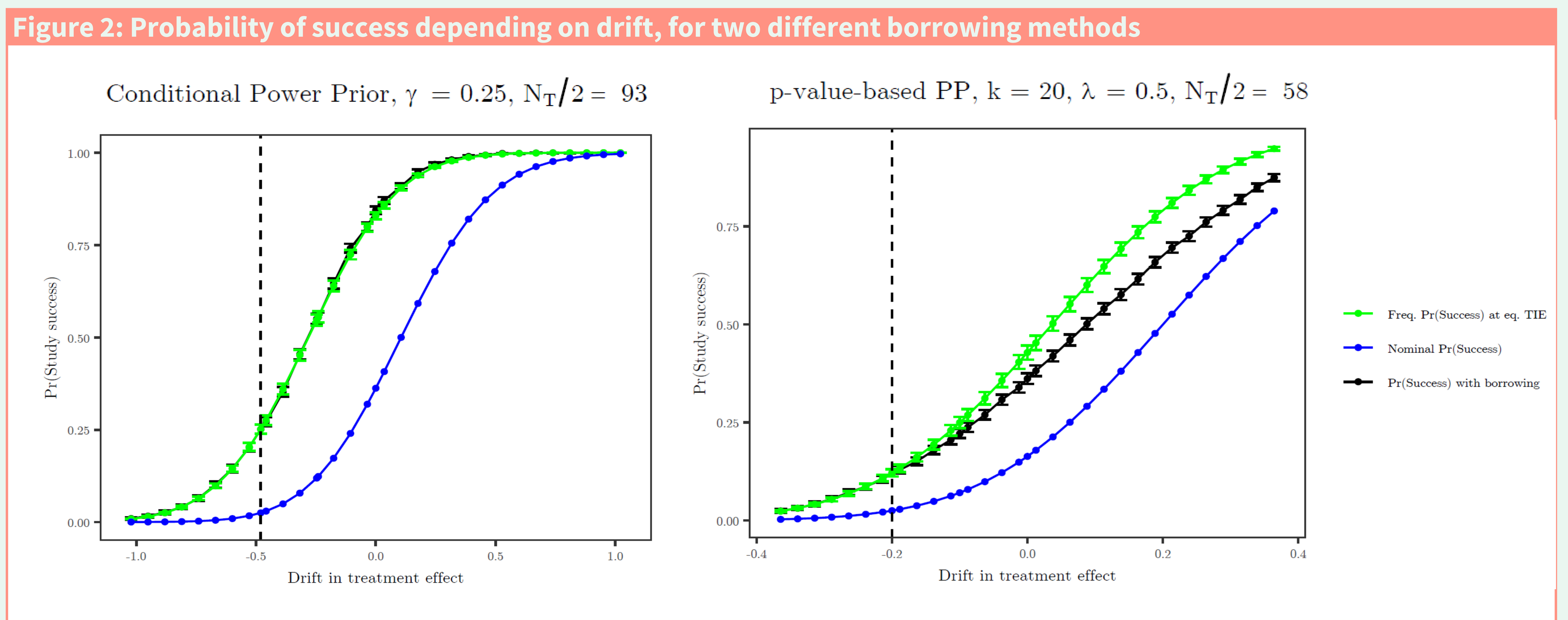
Power and type I error

- Borrowing may increase power but at the expense of type 1 error inflation, which confirms the findings or prior work[3]. When adjusting for type I error, no true power gain can be expected.

Coverage

- Coverage decreased with larger drift.
- Overall, CPP achieved better coverage at equivalent type I error, while TtP and p-PP performed worse.

Figure 2: Probability of success depending on drift, for two different borrowing methods



The probability of success of a borrowing method is at most equivalent (CPP - left figure) to that of a t-test at the equivalent type I error rate; in some cases, it is even lower (p-PP - right figure).
Vertical dotted line: drift value corresponding to a null treatment effect in the target population. Blue line: probability of success of a t-test at a nominal TIE of 0.25 as a function of the drift. Black line: probability of success of a Bayesian borrowing approach; the resulting TIE is inflated (the curve crosses the dotted line well above 0.25). Green line: probability of success of a t-test at the equivalent inflated TIE.

CONCLUSION

- **Borrowing treatment effects via Bayesian methods can improve power at the cost of type I error inflation.**
- **Across scenarios, Conditional Power Prior and Robust Mixture Prior emerged as the most reliable in balancing bias, precision, and error control.**
- **Simulation-based calibration remains essential before using any borrowing approach in confirmatory clinical settings.**

REFERENCES:
1. Viele, K. et al. "Use of historical control data for assessing treatment effects in clinical trials." Pharmaceutical Statistics 13.1 (2014), 41–54.
2. Best, N., Price, R. G., Pouliquen, I. J., and Keene, O. N. "Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing". Pharmaceutical Statistics 20.3 (2021), 551–562.
3. Kopp-Schneider, A., Calderazzo, S., and Wiesenfarth, M. "Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control". Biometrical Journal. Biometrische Zeitschrift 62.2 (2020), 361–374.
4. Fauvel, T., Tanniou, J., Godbillot, P., Génin, M., & Amzal, B. "Comparison of Bayesian methods for extrapolation of treatment effects : A large scale simulation study" (Version 2). arXiv (2025).

ABBREVIATIONS: TtP: Test-then-Pool; CPP: Conditional Power Prior; p-PP: p-value based Power Prior; EBPP: Empirical Bayes Power Prior; NPP: Normalized Power Prior; Com.PP: Commensurate Power Prior; RMP: Robust Mixture Prior; TIE: Type I Error; MSE: Mean Squared Error; CI: Credible Interval; ESS: Effective Sample Size

ACKNOWLEDGEMENT: This project was funded through the reopening of competition no. 02 under framework contract following procurement procedure EMA/2020/46/TDA (Lot 3). This poster expresses the opinion of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of Quinten Health or the European Medicines Agency or one of its committees or working parties.
CONFLICT OF INTEREST: NA

CONTACT INFO: a.movschin@quinten-health.com