Discussing "Data fusion under weak identifiability in heterogeneous treatment effect modelling"

Stavros Nikolakopoulos

Department of Psychology, University of Ioannina

Dutch Medicines Evaluation Board (CBG-MEB)

O'Bayes, Athens
11 June 2025

University of Ioannina

CBG MEB | Medicines Evaluation Board

Disclaimer

All views expressed are my own and do not represent the Dutch Medicines Evaluation Board or the European Medicines Agency

The presentation is based solely on publicly available information

Acknowledgment

This project is carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI).
Project Number: 015467

Combining Observational and RCT Data

- Regulators like (and should like) RCTs
- Regulators (really) like strong control of FWER
- Regulators do not like combining
- Regulators do not like Bayesian
- However, the tide is turning
- Abundance of RWD - can not be ignored, strategic goal of the EMA (Arlett et al, 2021)
- Cases where borrowing is actually required

Borrowing of External data

Cases when it can be discussed, conditions that need to be fulfilled

- Infeasibility of RCT must be adequately justified
- Rare diseases
- Unethical/impossible randomization to Placebo/Control
- Borrowing of Control data (Registries, historical RCTs, natural history studies etc)

Cases when it is actively sought as a relevant solution

- *Extrapolation* of treatment effects in populations where RCTs are unethical/infeasible
- Pediatrics
- Borrowing of treatment effect data (*Borrowing from RCTs in adults*)

Borrowing data in RCTs - Methods

- When it comes to (dynamic) borrowing, most developed methods are about borrowing from past RCTs in concurrent RCTs
- Control group or contrast-based, in the form of a prior
- Dynamic: adjust amount of borrowing based on *conflict* between historical data (D_0) and concurrent data (D_1) so that D_1 is not dominated
- Most methods are based on **power priors** and **robust mixture priors**

Borrowing data in RCTs - Methods

Power Prior

$$p(\theta \mid D_0, \eta) \propto L(\theta; D_0)^{\eta} \cdot p_0(\theta)$$

where η is either fixed or based on some discrepancy measure between D_0 and D_1

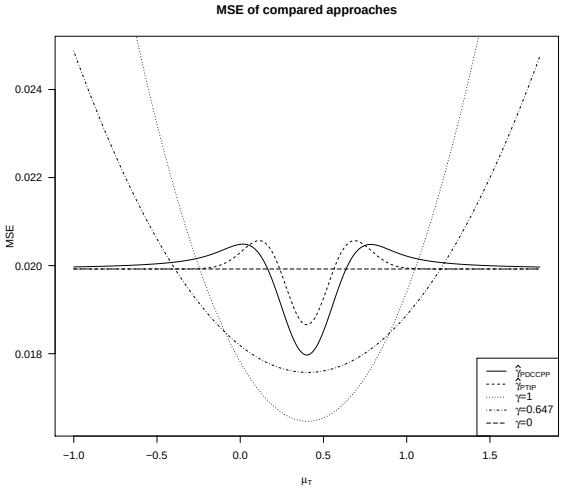
Robust Mixture Priors

$$p(\theta \mid D_1) = p(\theta \mid D_1, \text{predictive prior from } D_0) \cdot w + p(\theta \mid D_1, \text{vague}) \cdot (1 - w)$$

where $w = p(\text{predictive prior from } D_0 \mid D_1)$ and $1 - w = p(\text{vague} \mid D_1)$

A large scale simulation study commissioned by the EMA compared most methods w.r.t. frequentist operating characteristics (Fauvel et al, 2025)

Bias - Variance trade off



1

[1] Figure from Nikolakopoulos et al, 2018

Power Prior

$$p(\theta \mid \text{obs}, \eta) \propto L(\theta; \text{obs})^{\eta} \cdot p_0(\theta)$$

Allows learning about model parameters conditional on the model
Most work on simple ATE, focus on effects on T1E

Power Likelihood

Joint Likelihood

$$L_{\text{exp,obs}}(\theta) = L_{\text{exp}}(\theta) \cdot L_{\text{obs}}(\theta)^{\eta}$$

Much more flexible, allows for learning about the model
Can learn about CATE, not just ATE
Can handle different sets of covariates

Assessing similarity of RCT and RWD

Target Trial Emulation (Hernán & Robins, 2016)

- Use (Big) Observational data to emulate a target RCT
- Define inclusion criteria, treatments, outcomes, think very carefully of follow up times, intercurrent events
- Estimate ATE (?) using PS
- ATO (AT in the overlap population, where PS is close to 0.5 **Overlap Weighting**) is shown to improve covariate balance and estimation precision (Li et al, 2018)
- Could perhaps improve properties of $\hat{\pi}(\mathbf{x_i})$?
- Current research employs TTE to assess similarity between RCT and RWD

Comments and Questions

- Importance of unbiased *choice of data*
  - Could the model include both aggregate (ie published) data?
- Reasoning behind 50-50 proposal? How can we reflect our belief that RCT is the true model and adjust borrowing based on that?
- Assessment of FWER essential for confirmatory inference - translate to subgroup effects and multiple testing for CATE - BCF might be a challenge in that respect?

Upcoming EMA workshop

References

- Arlett, P., Kjaer, J., Broich, K., & Cooke, E. (2021). Real-world evidence in EU medicines regulation: enabling use and establishing value. *Clinical pharmacology and therapeutics*, 111(1), 21.
- Fauvel, T., Tanniou, J., Godbillot, P., Génin, M., & Amzal, B. (2025). Comparison of Bayesian methods for extrapolation of treatment effects: a large scale simulation study. *arXiv preprint* arXiv:2504.01949.
- Nikolakopoulos, S., van der Tweel, I., & Roes, K. C. (2018). Dynamic borrowing through empirical power priors that control type I error. *Biometrics*, 74(3), 874-880.
- Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8), 758-764.
- Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521), 390-400.

Thank you!