

Applying Bayesian Model Averaging to modelling longitudinal treatment effects within Alzheimer’s Disease clinical trials

Ben Jones¹, Ashwini Venkatsubramaniam², Doug Thompson², Dave Lunn², Aris Perperoglou², Dave Inman¹

¹Department of Biostatistics, GSK; ²Quantitative Sciences Innovation, GSK



Bayesian Model Averaging delivers modest power gains with controlled Type I error whilst also lowering the bias associated with misspecified models

Digital poster



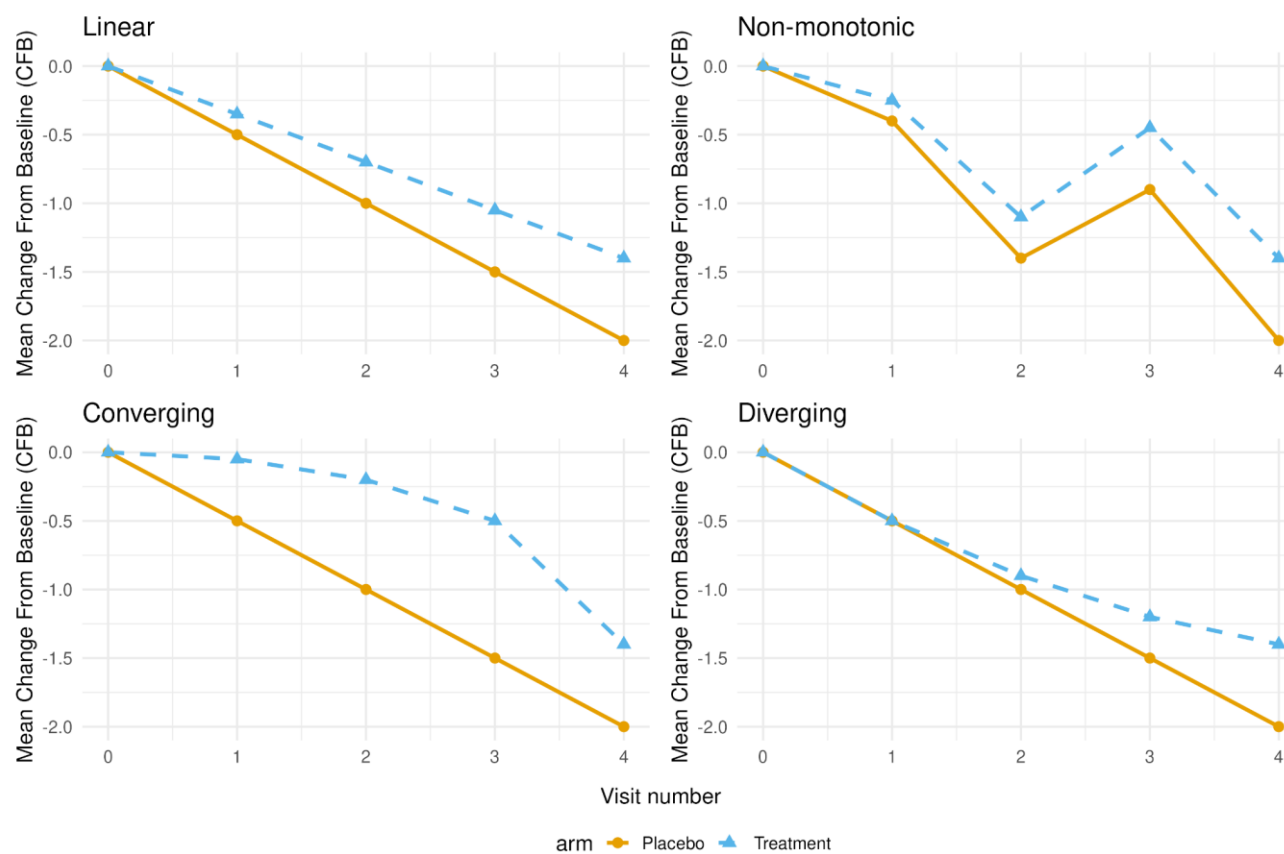
Scan Me

Background

- Pre-specifying an appropriate model for analysis of longitudinal trial data is challenging, particularly in Alzheimer’s Disease where delayed treatment effects can induce deviation from linearity within trial follow-up.
- MMRM is typically preferred, particularly in a regulatory setting, due to the lack of assumptions relied upon, despite the tendency for alternative continuous-time models (e.g. linear) to be more efficient.
- We conducted a proof-of-concept simulation study to explore BMA as a solution to such modelling uncertainties, implemented using pseudo-BMA^[1], noting similarities with the principles of MCP-MOD^[2], commonly used in dose selection, which has broad regulatory acceptance^[3,4].

Methods

Figure 1: Data generating mechanisms



ADEMP ^[5]	Description
Aims	• To evaluate whether reported efficiency gains from single continuous fits are retained within a BMA framework and if unbiased estimation can be achieved whilst ensuring Type I error control.
Data Generating Mechanisms (DGMs)	• N = 400 (200/arm), SD = 2, AR(0.5) correlation structure • Between-arm difference at final timepoint: 0.6 • Trajectories: Linear, Non-monotonic (NM), Converging, and Diverging (Figure 1) + null scenarios • Missing data: complete, low (+2.5% at each timepoint); high (+7.5% at each timepoint)
Estimands and targets	• Between-arm difference in CFB at the final visit • Associated null hypothesis
Methods	• MMRM • BMA, implemented using “pseudo-BMA” with Bayesian bootstrapping, which derives weights for each candidate model using relative predictive accuracy^[1]. Model set: Linear, Quadratic, Spline regression (k=1,2,3 knots)
Performance Measures	• Relative bias, coverage of the 95% CrI, power and type I Error

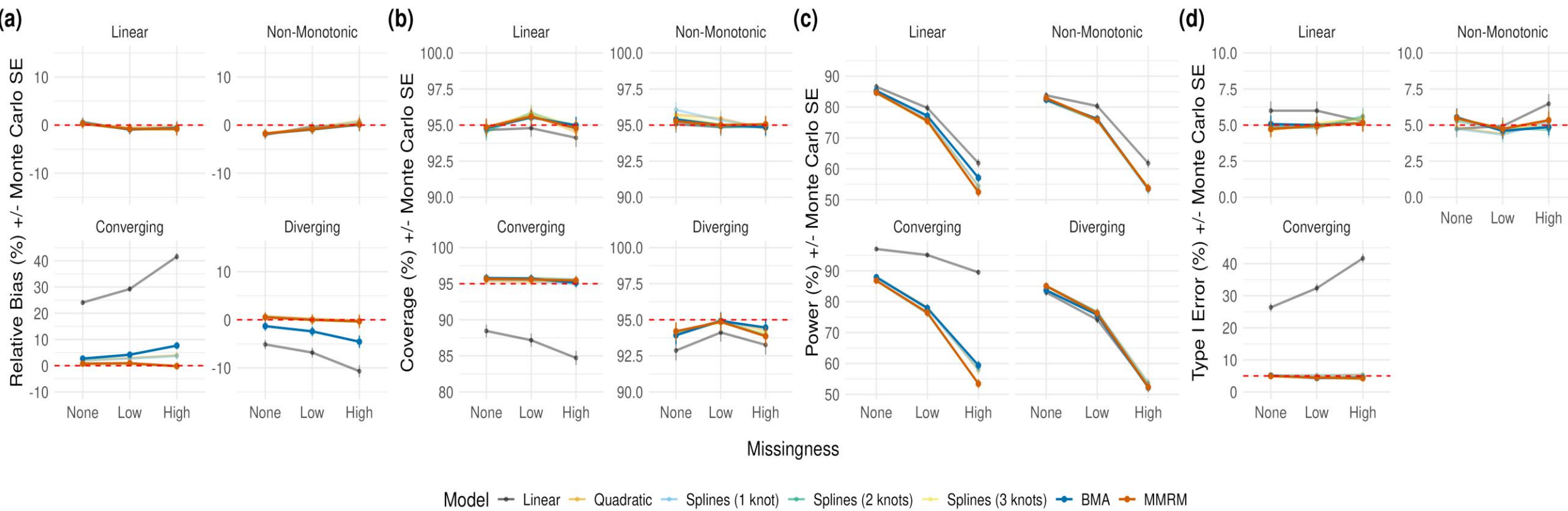
Results

- The pseudo-BMA approach assigns weights to the models best aligned with the DGMs (Table 1).
- Both BMA and MMRM provide unbiased treatment effect estimation (Figure 2a) and nominal coverage (Figure 2b) under the linear and NM DGMs. However, BMA introduces bias under both the converging and diverging DGMs, although coverage is broadly retained at nominal levels.
- BMA achieves modest power gains relative to MMRM under linear and converging DGMs, with no differences under the NM or diverging DGMs (Figure 2c). Type I error is well controlled for both BMA and MMRM under all DGMs (Figure 2d).

Table 1: Weights derived from pseudo-BMA

Scenario	Data Generating Mechanism				
	Linear*	NM*	Converging	Converging (Null)	Diverging
Linear	53.8%	0.0%	12.2%	0.5%	49.1%
Quadratic	15.6%	0.0%	27.7%	15.6%	17.6%
Splines (k=1)	15.4%	0.0%	27.8%	15.4%	17.4%
Splines (k=2)	7.7%	50.1%	16.3%	34.3%	8.0%
Splines (k=3)	7.5%	49.9%	16.0%	34.2%	8.0%

Figure 2: Performance measure results for (a) relative bias; (b) coverage; (c) power; and (d) type I error.



Conclusions

Implementation of BMA can, in scenarios where stronger assumptions are valid (e.g. linear), result in modest power gains whilst providing unbiased estimation and maintaining Type I Error Control.

However, the procedure can allocate non-trivial weights to models with strong but mis-specified assumptions, reducing but not eliminating bias and Type I error relative to single model fits.

Further work is required to validate the utility of the methodology in the regulatory context, through exploration of assigning conservative priors to weights in favour of models with a stronger underlying assumptions. Future work will also assess implementation in real-life case studies and explore utility of BMA in the context of instream decision making.

Abbreviations

AR: Autoregressive; BMA: Bayesian Model Averaging; CrI: Credible Interval; DGM: Data Generating Mechanism; MCP-MOD: Multiple Comparison Procedure – Modelling; MMRM: Mixed Model for Repeated Measures; NM: Non-monotonic; SD: Standard Deviation.

References

- [1] Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)," Bayesian Analysis, Bayesian Anal. 13(3), 917-1007, (September 2018)
- [2] Pinheiro, J., Bornkamp, B., Glimm, E. and Bretz, F. (2014). Model-based dose finding under model uncertainty using general parametric models. Statist. Med., 33: 1646-1661. <https://doi.org/10.1002/sim.6052>
- [3] EMA, Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty (2014)
- [4] FDA, Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty (2015)
- [5] Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019;38:2074–2102. <https://doi.org/10.1002/sim.8086>

Disclosures

This study was funded by GSK. B Jones, AS, DT, DL, AP and DI are employees of GSK and hold financial equities in GSK.