MSR130

Comparison of Software for Bayesian Network Meta-analysis: a Case Study of Binomial Outcomes in Ulcerative Colitis

Ainsworth C[1], Youn J-H[2], Petersohn S[2], Gittfried A[2], Jevdjevic M[2], Piena M[2]

[1]OPEN Health Evidence & Access, Manchester, UK; [2]OPEN Health Evidence & Access, Rotterdam, Netherlands

OPEN HEALTH

INTRODUCTION

- Numerous software are available for performing Bayesian network meta-analysis (NMA), however WinBUGS remains the preferred option, with code provided in the National Institute of Health and Care Excellence Technical Support Documents widely adopted throughout the industry.[1]
- Given that WinBUGS and its successor, OpenBUGS, are no longer being developed,[2,3] alternative software have become available. Stan, a programming language released in 2012 and in active development today, is one of these.[4]
- In the NMA setting, where timelines are often pressured and the quantity of analyses sizeable, understanding the benefits alternative software may offer with regards to performance, running time and user experience is of high importance and interest.

OBJECTIVES

- Compare results using three different software (WinBUGS, OpenBUGS and Stan) to perform multiple NMAs of binomial outcomes assessing relative effects of treatments in patients with ulcerative colitis.

METHODS

Software

- Previously published NMAs[5] conducted using OpenBUGS were replicated using the R packages R2WinBUGS[6] and multinma[7] to run WinBUGS and Stan, respectively, and their results were compared.
- BUGS software (WinBUGS and OpenBUGS) utilise Gibbs sampling to sample from probability distributions,[2,3] whereas Stan uses Hamiltonian Monte Carlo (HMC) sampling.[4] Therefore it would be expected that, if differences were to exist between the software, WinBUGS and OpenBUGS would be more similar to one another than to Stan.

Evidence

- Where data allowed, analyses comprised the following outcomes at induction and maintenance: response, remission, mucosal healing, and discontinuations due to adverse events, and were stratified by prior exposure to anti-tumour necrosis factor (anti-TNF) therapy.[5]
- All networks included only placebo-controlled RCTs, thereby only including direct evidence and resulting in a star-shaped network.

Model Specification

- NMAs were performed using both fixed effect (FE) and random effects (RE) models for treatment effects.
- In line with the previous publication,[5] models assumed binomial distributions and used a logistic link function with a burn-in of 20,000 iterations, total of 60,000 iterations, and uninformative priors.

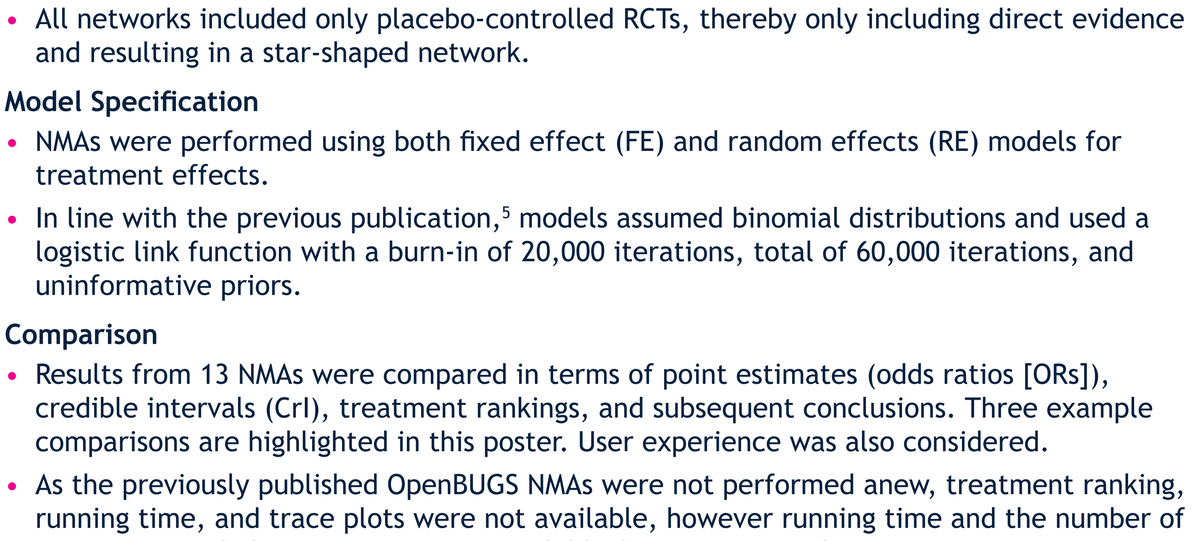
Comparison

- Results from 13 NMAs were compared in terms of point estimates (odds ratios [ORs]), credible intervals (CrI), treatment rankings, and subsequent conclusions. Three example comparisons are highlighted in this poster. User experience was also considered.
- As the previously published OpenBUGS NMAs were not performed anew, treatment ranking, running time, and trace plots were not available, however running time and the number of iterations needed to converge were available for WinBUGS and Stan.

RESULTS

- NMAs of clinical response, clinical remission, and mucosal healing at induction in the population of anti-TNF naïve patients are presented in this poster, all with the same network diagram (Figure 1). Seven studies reported data for clinical response and remission, six studies reported data for mucosal healing.
- Outputs were similar across software for all NMAs. Figure 2 presents forest plots of mean ORs and 95% CrIs versus placebo by software, for each of the three endpoints of interest. Further, conclusions aligned across software, with no changes to treatment rankings based on probability of best estimates (Figure 3).

Figure 1. Network Diagram for Clinical Response and Remission at Induction



Abbreviations: ADA, adalimumab; GLM, golimumab; IFX, infliximab; PBO, placebo; VDZ, vedolizumab.

- In the pairwise comparison of adalimumab versus infliximab, median ORs were consistent across the different software in the FE NMA of response in anti-TNF naïve patients at induction: 2.19 [95% CrI: 1.35-3.55] using OpenBUGS vs. 2.52 [1.50-3.94] using WinBUGS vs. 2.44 [1.51-3.95] using Stan.
- All networks were too small to give reliable estimates for RE models using OpenBUGS[5] and WinBUGS, but comparable results to FE analyses were observed using Stan (adalimumab versus infliximab, median OR [95% CrI]: 2.45 [1.34-4.59]).
- Running times were similar between software for all networks, likely due to the size of the networks. Stan required less running time and iterations to converge. Further, the shinyStan functionality was noted as more user-friendly in diagnosing NMA models.

Figure 2. Median ORs versus Placebo from FE NMA of Clinical Response, Clinical Remission, and Mucosal Healing at Induction by Software

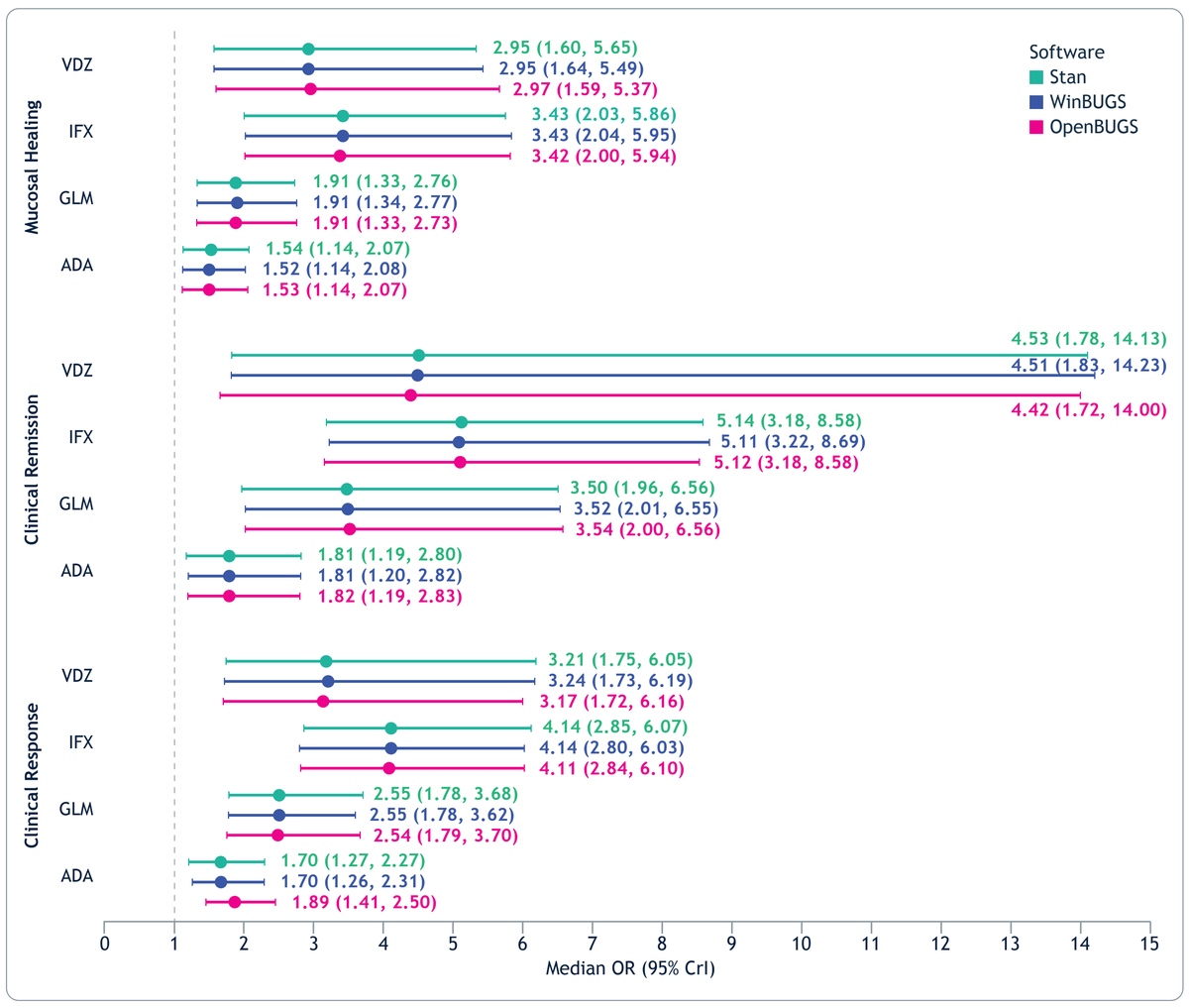
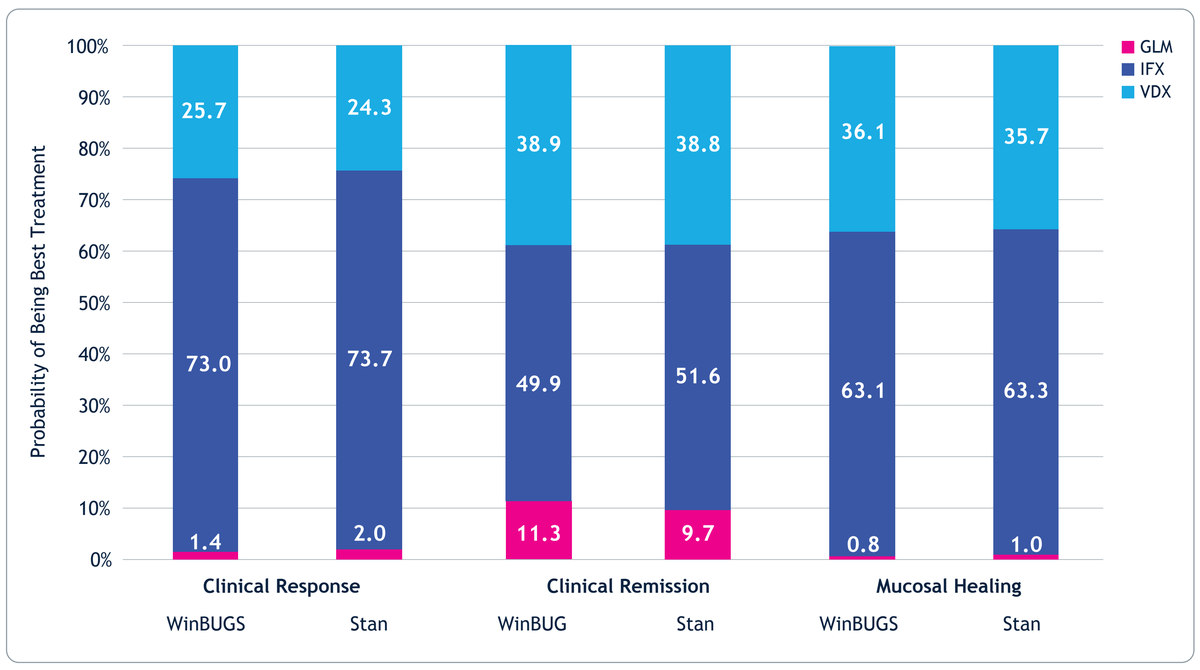


Figure 3. Probability of Best Treatment from FE NMA of Clinical Response, Clinical Remission, and Mucosal Healing at Induction by Software



Note: ADA and PBO had 0% probability of being best treatment across all outcomes and software.

CONCLUSIONS

- Comparing software resulted in consistent outcomes and interpretations. While these results may not be generalizable to all network shapes and sizes, this study found that in small networks with no indirect evidence, all three software return comparable estimates for median ORs and CrIs, in addition to treatment rankings.
- Differences were observed in software performance under RE models, with BUGS software unable to produce reliable results, whilst Stan produced results comparable to FE analyses. Analysts should consider software choice, alongside model selection, in the presence of between-study heterogeneity.
- This demonstrates that Stan provides a valid alternative and may be advantageous both in terms of performance and user experience.

REFERENCES

1. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011
2. Lunn, D, Spiegelhalter, D, Thomas, A, et al. The BUGS project: Evolution, critique and future directions. *Stat Med.* 2009;28(25):3049-67.
3. Goudie RJB, Turner RM, De Angelis D, et al. MultiBUGS: A Parallel Implementation of the BUGS Modelling Framework for Faster Bayesian Inference. *J Stat Softw.* 2020 Oct 7;95:7.
4. Stan Modeling Language Users Guide and Reference Manual [program], 2022
5. Vickers AD, Ainsworth C, Mody R, et al. Systematic Review with Network Meta-Analysis: Comparative Efficacy of Biologics in the Treatment of Moderately to Severely Active Ulcerative Colitis. *PLoS One*. 2016;11(10):e0165435.
6. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A Package for Running WinBUGS from *R. J Stat Softw.* 2005;12(3):1–16.
7. Phillippo DM. multinma: Bayesian Network Meta-Analysis of Individual and Aggregate Data. 2022. https://doi.org/10.5281/zenodo.3904454

ACKNOWLEDGEMENTS

We want to thank Mark Lambton and Niels Dunnewind for their software support, Claire Simons for her review of this poster, and Sonja Kroep for her contributions conceptualizing this project.