A comparative study of alternative software to conduct hazard ratio-based network meta-analysis

Perera C[1], Hirst A[1], Heron L[1]

[1]Adelphi Values PROVE™, Bollington, Cheshire SK10 5JB, United Kingdom

Adelphi
ADELPHI VALUES

PROVE
Expertise in Access and Value Evidence Outcomes

MSR97

Introduction

- The increasing demand for comparative evidence in health technology assessments (HTAs) underscores the need for efficient and reliable analytical methods, especially when direct head –to-head clinical trials is absent.[1]
- Network meta-analysis (NMA) serves as a crucial tool for indirect comparisons across multiple interventions.
- Bayesian approaches to NMA are particularly advantageous due to their flexibility in incorporating different types of data and accounting for uncertainty.
- The choice of software for conducting Bayesian NMA can significantly impact the efficiency, user experience, and ultimately the feasibility of timely analyses.
- Traditional software like WinBUGS has been extensively utilized, but newer platforms such as Stan and Just Another Gibbs Sampler (JAGS) offer potential improvements in computation and usability.[2-4]

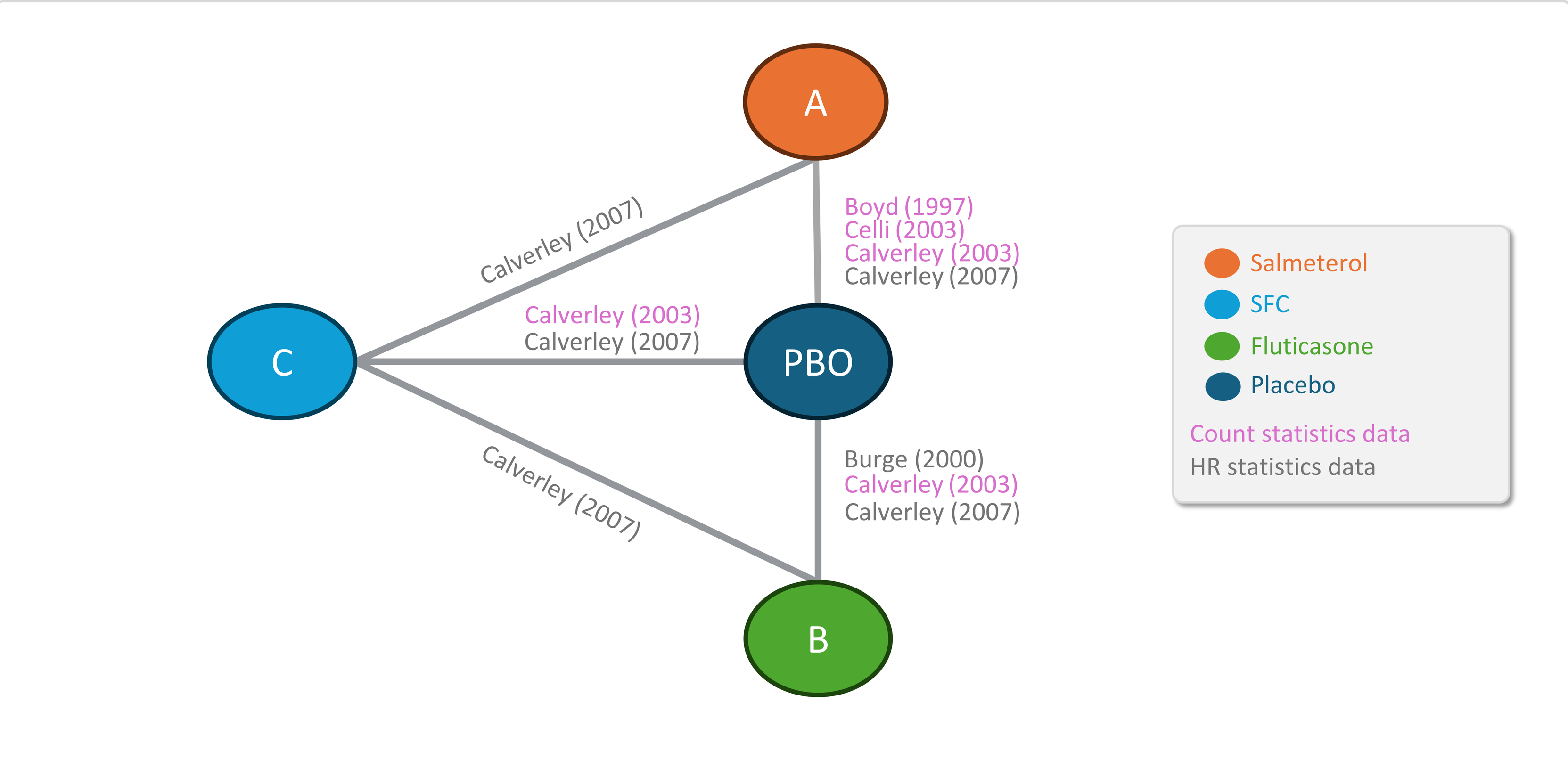
Study aim and objectives

- This study aims to compare three software packages—WinBUGS, Stan, and JAGS—in performing hazard ratio-based NMAs by replicating a published analysis.
- The objective is to evaluate differences in results, computational efficiency, and user experience to inform optimal software selection for future NMAs in HTA.

Methods

- A published Bayesian NMA, by Woods et al., focusing on survival endpoints was replicated using WinBUGS, Stan, and JAGS.[5]
- The original analysis combined both count data and hazard ratio statistics on the hazard ratio scale, encompassing a network of evidence from three randomized controlled trials comparing three treatment regimens.
- Each software was employed to program the Bayesian NMA model, estimating median hazard ratios and corresponding 95% credible intervals (CrIs).
- To assess consistency across software, we compared the hazard ratio estimates and CrIs obtained from each platform.
- Identical analyses were run on the three software with the parameters reported in Table 1.

Figure 1. Network of evidence.



Abbreviations: SFC, salmeterol and fluticasone propionate

Table 1. General NMA parameters.

Parameters	Value
Number of iterations	200,000
Burn-in	40,000
Thinning parameter	20
Number of HR observations	5
Number of binary observations	8
Number of treatment	4
Number of studies	5

Computational efficiency

- To assess the computational efficiency quantitatively, the sampling run-time required for each analysis will be recorded using the microbenchmark package:
- Each analysis will be run ten times, and the mean run time will be recorded.

User experience

- A qualitative assessment of efficiency focused on the software ease-of-use, including aspects of programming, model implementation, and debugging processes.

Results

Figure 2. (A) Violin plot showing differences in sampling time across software across 10 models runs. (B) Forest plot for fixed effect model.

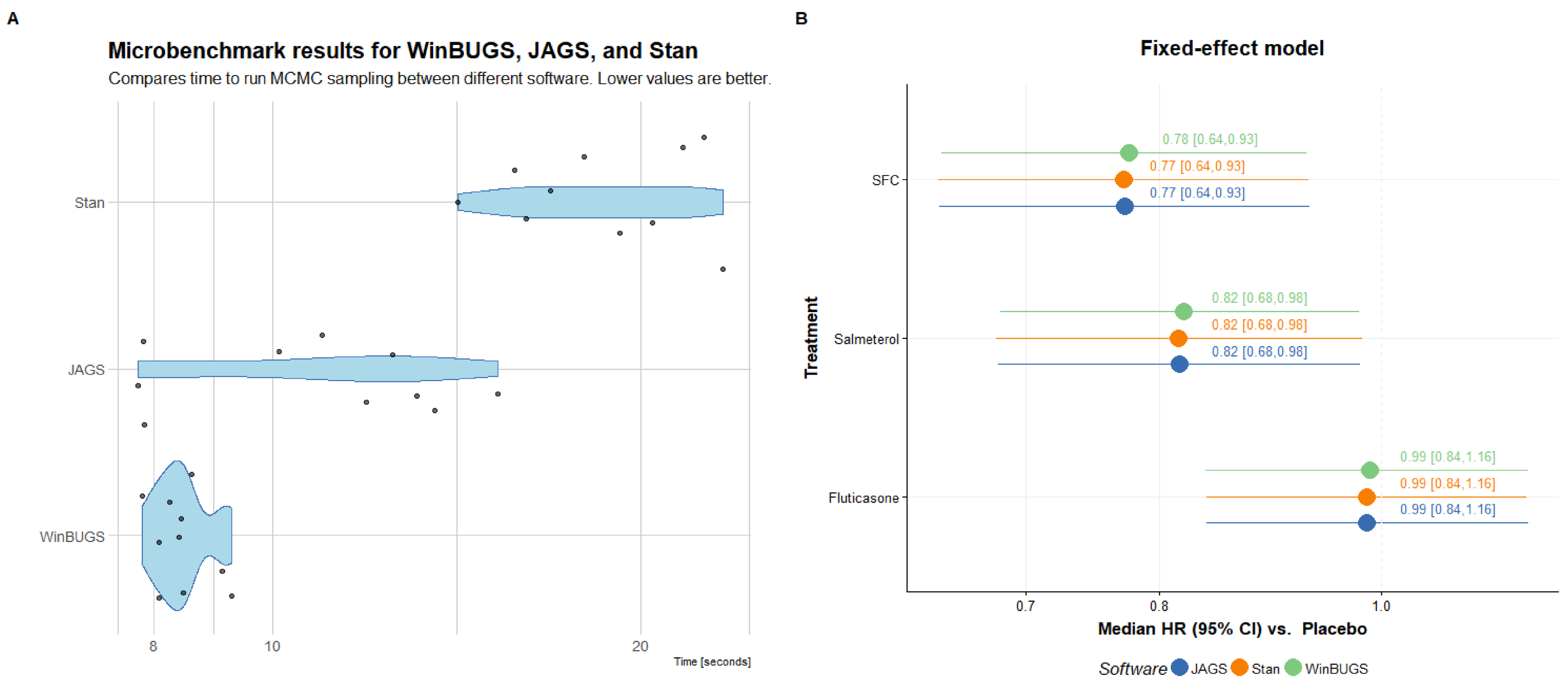


Table 2. Hazard ratio results from network meta-analysis across different software.

Treatment	WinBUGS HR (95% CrI)	Stan HR (95% CrI)	JAGS HR (95% CrI)
SFC	0.776 (0.643 - 0.928)	0.772 (0.643 - 0.928)	0.772 (0.642 - 0.929)
Salmetrol	0.820 (0.682 - 0.978)	0.817 (0.680 - 0.978)	0.815 (0.679 - 0.979)
Fluticasone	0.989 (0.838 - 1.159)	0.987 (0.840 - 1.158)	0.985 (0.838 - 1.159)

Abbreviations: CrI, credible interval; HR, hazard ratio; SFC, salmeterol and fluticasone propionate

- The median hazard ratio estimates across WinBUGS, Stan, and JAGS exhibited minimal differences, ranging between 0.004 and 0.006.
- This consistency indicates that all three software packages produced comparable findings in terms of effect estimates.
- The 95% CrIs were also closely aligned across platforms, showing negligible variations that did not affect the overall interpretation of results.
- Stan and JAGS yielded median hazard ratio estimates that were generally lower than those obtained from WinBUGS.
- This slight discrepancy may be attributed to the different Markov Chain Monte Carlo (MCMC) sampling algorithms employed by each software, which can influence convergence and estimation.
- In terms of computational efficiency, WinBUGS and JAGS demonstrated superior performance over Stan.
- The mean sampling time over 10 model runs was 8.44 seconds for WinBUGS, 11.44 for JAGS, and 18.58 for Stan.
- Stan's longer computation time is counterbalanced by its enhanced user experience; it offers a more intuitive programming environment and efficient debugging capabilities, largely due to its seamless integration within the RStudio integrated development environment (IDE).

Conclusions

- This comparative analysis of WinBUGS, Stan, and JAGS revealed that all three software packages produce consistent and reliable results for hazard ratio-based NMAs.
- While WinBUGS and JAGS offer faster computation times, Stan provides a superior user experience in terms of programming ease and debugging efficiency, facilitated by its integration with RStudio.
- The findings from this study were in accordance with studies investigating alternative software used to conduct NMA.[6]
- These findings suggest that the choice of software for Bayesian NMA can be tailored to the specific needs of the analysis.
- For time-sensitive projects where computational speed is paramount, WinBUGS or JAGS may be preferred in the presence of a simple network of evidence.
- Conversely, for analyses that benefit from an enhanced programming interface and ease of model manipulation, Stan emerges as a favourable option.
- Complex analyses that may require additional computational power may see that Stan provides a valid alternative given its No-U-Turn Sampling algorithm.
- Efficient execution of NMAs will be critical to support the growing need for rapid comparative evidence in the European Union Joint Clinical Assessment.
- Future research assessing the computation efficiency of other software with more complex NMA methods is required to understand where to optimize performance.
- The ability to produce timely and reproducible results enhances the validation and communication of findings, ultimately contributing to more informed decision-making in healthcare policy and practice.

References

1. Kisser A, Knieriemen J, Fasan A, et al. Towards compatibility of EUnetHTA JCA methodology and German HTA: a systematic comparison and recommendations from an industry perspective. The European Journal of Health Economics. 2021:1-16.; **2.** Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing. 2000;10:325-337.; **3.** Carpenter B, Gelman A, Hoffman MD, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 01/11 2017;76(1):1 - 32. doi:10.18637/jss.v076.i01.; **4.** Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Vienna, Austria; 2003:1-10.; **5.** Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. BMC.; **6.** Medical Research Methodology. 2010/06/10 2010;10(1):54. doi:10.1186/1471-2288-10-54.; **7.** Jevdjevic M, Youn JH, Petersohn S, Gittfried A, Ainsworth C, Piena M. MSR110 A Comparison of Stan Versus WinBUGS Software for Conducting Bayesian.; **8.** Hazard Ratio-Based Network Meta-Analysis. Value in Health. 2022;25(12):S371. doi:10.1016/j.jval.2022.09.1841