A COMPARATIVE REVIEW OF BAYESIAN NETWORK META-ANALYSIS: A USAGE CASE STUDY

Author: Krypton Ma - 2024 Data & Analytics Statistics Summer Intern. Contact Info: kma1@unc.edu

Collaborator: Xiaofei Wang, Michael Sonksen

CONTENTS

BACKGROUND

From Meta Analysis (MA) to Bayesian Network Meta Analysis (BNMA)

Meta Analysis (MA) is a type of evidence-synthesis study by aggregating results from multiple research results to generate a comprehensive assessment of treatment outcomes in pharmaceutical research. It combines facts quantitatively throughout studies addressing comparable research questions, thereby enhancing statistical energy and precision compared to individual studies. It commonly requires systematic literature reviews, statistical pooling, and using software to evaluate heterogeneity and potential bias. This method is pivotal in evaluating the efficacy and protection of medication across diverse patient populations and settings, presenting valuable insights for regulatory selection-making, scientific exercise guidelines, and healthcare coverage.

Network Meta Analysis (NMA), also known as s multiple-treatments meta-analysis or mixed-treatment comparison[12], extends meta-analysis by allowing comparison for indirect evidence and therefore enhance the breadth and depth of proof synthesis in pharmaceutical studies. It leverages advanced statistical strategies to integrate evidence between any study. This tool widens the flexibility of meta-analysis and offers a nuanced knowledge of comparative remedy effects to support scientific decision making.

At Lilly, NMA is applied by many cross functional teams such as Real World Evidence and Payer, Reimbursement, Market Access. The ranking of treatments from NMA is then used for building health economics and outcome model, which is a pivotal piece of evidence for healthcare technology and assessment (HTA) submission. Such findings are important to inform decisions regarding market entry, reimbursement, pricing, and clinical guidelines.

Among NMA analysis, Bayesian Network Meta-Analysis (BNMA) is of a more prevalent and flexible method over Frequentist's approach. BNMA allows for the incorporation of prior to generate posterior distribution of relative effects interested by Markov chain Monte Carlo (MCMC) approximation.

BATMAN NMA is the destination to carry out BNMA, and it is one of the multiple functions from "BAyesian Tool for the Meta-Analysis of Networks plus" ("BATMAN+", version 2.2, 2024-6-14). BATMAN+ is a tool developed and maintained by SIC@Lilly, and it is the in-house solution and a comprehensive tool for developing all evidence-based synthesis models currently in use. It has been validated in a joint venture with MDIT to make sure the tool aligns with guidelines from NICE (National Institute for Health and Care Excellence). BATMAN+ can perform:

- Indirect Comparisons Meta Analysis (ICMA)

- Matching adjusted Indirect Comparisons (MAIC)

- Arm-Based Network Meta Analysis (ABNMA)

- Contrast-Based Network Meta Analysis (CBNMA)

- Model-Based Network Meta Analysis (MBNMA) for longitudinal dosage/ time

- Multi-level Network Meta Analysis (MultiNMA) for a mixture of patient-level (IPD) and aggregated (AgD) data

- Multivariate Network Meta Analysis (MVNMA)

The focus of this script is Bayesian Network Meta Analysis (BNMA), which falls mainly under Contrast-Based Network Meta Analysis (CBNMA), but also includes ABNMA and MultiNMA. Next chapter I will introduce BNMA candidates and explain package selection criteria.

Overview of R packages for NMA

CRAN provides a holistic summary of packages and tools for meta-analyses.

There are two packages with general NMA analyses: *NMA* and *meta*. Some niche packages are great for specific usage cases and more information can be found upon the previous reference. For the purposes of this study, they will not be included. However, researchers are highly recommended to look for possible substitute functions to enhance the functionality of BATMAN. For example:

- *nmaILNA* provides NMA with meta-regression capabilities with integrated nested Laplace approximation (INLA) instead of traditional MCMC sampling method.
- *rankinma* provides treatment ranking in NMA.
- *rnmamod* can carry out NMA analysis and address (aggregate) missing participant outcome data.
- *multinma* takes from IPD/AgD level data, or a mixture of both. This package was used in BATMAN pacman analysis.
- *netmeta* focuses on frequentists' approach and has been used in BATMAN feasibility section.
- *metapack* allows easy access to regression-modeling of the variances (of the treatment effects) and response covariance matrices. However, it lacks comprehensive graphical option output and is designed to serve as an efficient sampler algorithm.
- *NMA* can check transitivity (similarity between study settings to allow comparison).

For this manual, we will narrow down to five R CRAN packages that carry out BNMA as their core functions: *multinma*, *pcnetmeta*, *bnma*, *gemtc*, and *BUGSnet*. Although *multinma* and *pcnetmeta* are used in BATMAN+ for ABNMA (arm-based) and MultiNMA ('pacman'), we still want to include these two packages to compare on Bayesian functionality.

Note: Package *multinma* is archived from CRAN, but one can still import the lastest version on Rstudio IDE.

COMPARATIVE REVIEW

A NMA analysis begins with feeding individual patient data (IPD) or aggregated data (AgD), exploration from the network plot to validate feasibility, and model specification followed by assessment of convergence and model fit.

Here, we present the functionalities of different important features: Table 1 to 5 summarize features of BNMA available in the latest versions of *multinma* (version 0.7.1), *pcnetmeta* (version 2.8), *bnma* (version 1.6.0), *gemtc* (version 1.0-2), *BUGSnet* (version 1.1.0) packages. Red highlighted text represents the uniqueness of functionality for a particular package.

Table 1 presents high-level summaries for each package specialty and limitation. All five packages were updated in recent years. We find that BATMAN+ is a greatly comprehensive tool. In all the following aspects of an NMA analysis, BATMAN NMA provides fine and customized model tuning parameters, extra features to fit special study requirements, and yields a complete list of outputs both numerically and graphically.

Table 1: general information

	multinma	**pcnetmeta**	**bnma**	**gemtc**	**BUGSnet**
publish year (update year)	2020 (2024)	2014 (2022)	2020 (2024)	2012 (2023)	2019 (2021)
highlight	- flexible model specification - baseline model	- arm-based - tidy output - one step to fit - pairwise measures	- auto convergence check (error if fail) - baseline model - reasonable and dispersed initial values if unspecified - comprehensive output - exhaustive inconsistency modelling	- comprehensive output	- comprehensive output for feasibility - easy network setup - concise coding statements
limitation	limited flexibility in MCMC setup	- limited graphical output - B/W TFL	- cannot fit covariate	highly inflexible data read in	current version does not support categorical covariates with > 3 levels

When looking at the posterior parameter estimates, analysts should keep in mind that BNMA analysis does not automatically correct the data, so formatting data is crucial. Table 2 presents the functionalities of the input data and feasibility output. In addition, *pcnetmeta* only supports arm-based NMA (which corresponds to BATMAN ARM); other packages can use either contrast-level or arm-level data from the built-in functions to convert arm-level data into contrast-level data. All packages supported data from a multi-arm study. However, note that BATMAN NMA analysis only takes in either IPD/AgD data, but not both at the same time (the PACMAN function can apply).

Table 2: data input and network plotting functionality

<table>
<tr><th>Tasks</th><th>Features</th><th>BATMAN</th><th>multinma</th><th>pcnetmeta</th><th>bnma</th><th>gemtc</th><th>BUGSnet</th></tr>
<tr><td><b>Forms of input data</b></td><td>Arm-level</td><td colspan="6" rowspan="3">Y</td></tr>
<tr><td></td><td>Contrast-level</td></tr>
<tr><td><b>Class of outcome</b></td><td>Binomial</td></tr>
<tr><td></td><td>Multinomial</td><td>Y</td><td>Y</td><td>N</td><td>Y</td><td>N</td><td>N</td></tr>
<tr><td></td><td>Count</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr>
<tr><td></td><td>Continuous</td><td>Y</td><td>Y</td><td>N</td><td>Y</td><td>Y</td><td>Y</td></tr>
<tr><td></td><td>Time-to-event</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>Y</td></tr>
<tr><td></td><td>Survival</td><td>Y</td><td>Y</td><td>Y</td><td>N</td><td>Y</td><td>Y</td></tr>
<tr><td><b>Link function</b></td><td>Identity (normal)</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr>
<tr><td></td><td>Cloglog (poisson)</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>Y</td></tr>
<tr><td></td><td>Logit (binomial/ multinomial)</td><td colspan="6">Y</td></tr>
<tr><td></td><td>Probit (binomial/ multinomial)</td><td>Y</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>N</td></tr>
<tr><td></td><td>Reciprocal</td><td>N</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>N</td></tr>
<tr><td></td><td>Log</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr>
<tr><td><b>Descriptive measures</b></td><td>Number of studies</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr>
<tr><td></td><td>Sample size</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr>
</table>

	Multi-arm studies	N	N	N	N	Y	Y
	Covariate	N	N	N	N	N	Y
	treatments	Y	N	N	N	Y	Y
Feasibility panels	Network plot	Y	Y	N	Y	Y	Y

BATMAN accommodates a wide range of outcome responses, even though binary outcome and survival data are mostly seen in the NMA literature. Another highlight of BATMAN is that it yields a variety of plot panels (adopted from package *netmeta*, frequentist NMA approach) for feasibility checks (Figure 1):

- network plot (static/ interactive)

- funnel plot

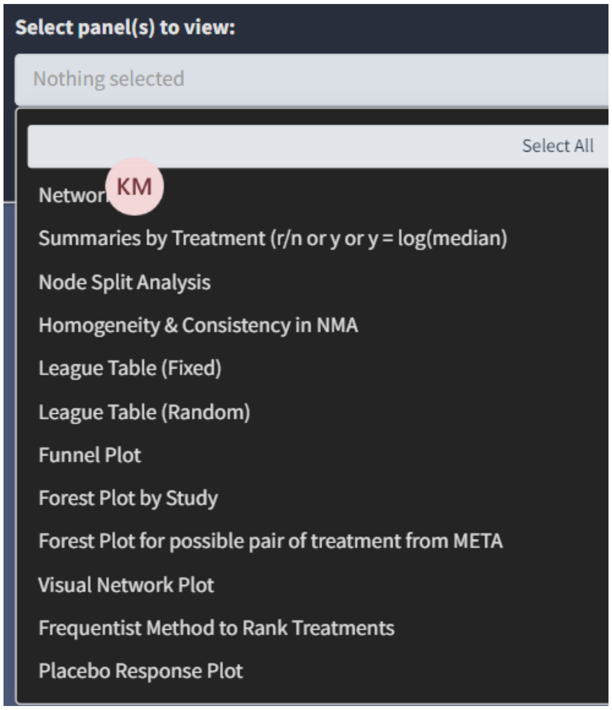
- node splitting analysis (forest plot for each pair if analyzed from direct or indirect evidence)

- forest plot (by study/ pairwise)

- placebo response plot

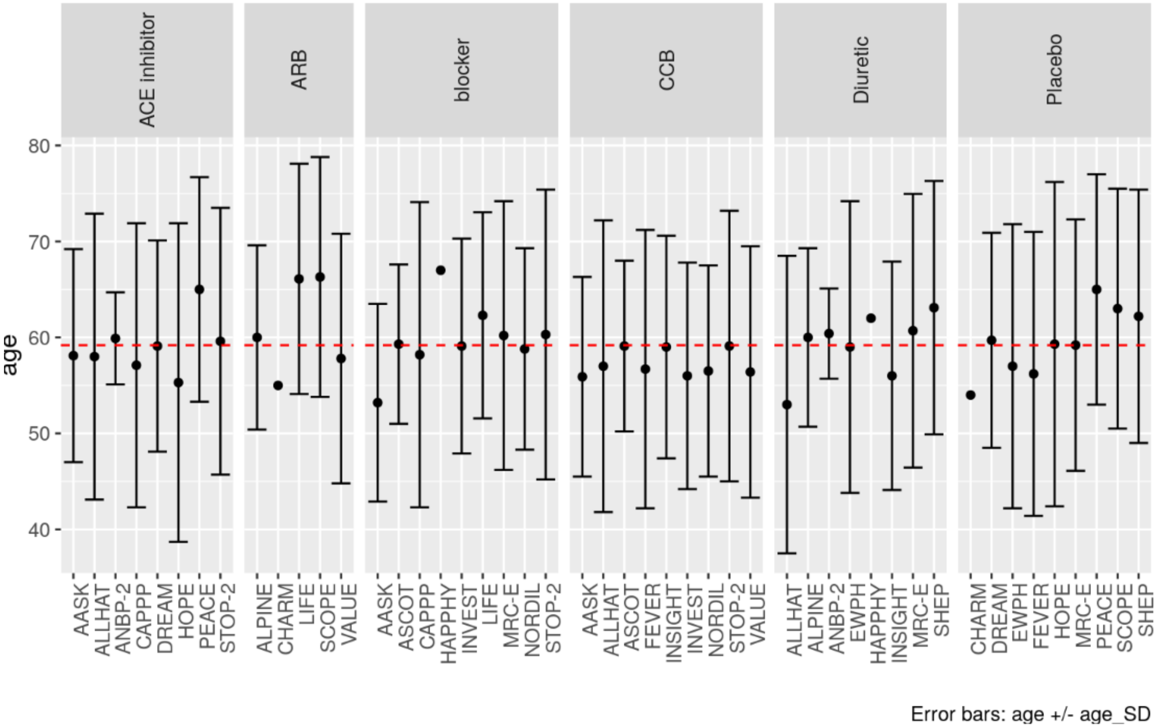
- other numerical summaries such as league table

Figure 1: Comprehensive feasibility panel in BATMAN



BUGSnet has many functions to summarize study and patient information and provides summary statistics on the response to interventions (both singular and pairwise). For example, Figure 2[1], borrowed from the *BUGSnet* vignette, is a vertical forest plot showing the mean and range of age among patients for each study, grouped by different interventions. While a network plot is an essential diagram, descriptive statistics from *BUGSnet* can be visually helpful to check the validity of patient data and help researchers decide on the inclusion or exclusion of certain studies for subgroup analysis.

Figure 2: Forest plot with patient information on BUGSnet



For arms with mixed treatment (for example, supportive care + *Etanercept* 25 mg), *multinma* was able to color the mixed treatment in the network plot, which allows the identification of any arm involved with any specific treatment (function '*plot.nma_data()*' with argument *'show_trt_class=TRUE*').

Table 3 presents the options for fixed and random model effect modeling, inconsistency modelling, and covariate inclusion. Fixed models are ideal for small sample size (patient size). If there is heterogeneity among multiple trials, random effect modelling is recommended to account for between-trial variation. Package *pcnetmeta* is arm based and it offers three different types of models *("hom_eqcor", "het_eqcor", or "het_cor"*): the former two has the option to decide the common correlation coefficient between treatments. However, the package is arm based and this functionality is not comparable to others.

Different heterogeneity parameters basically allow for finer variance-covariance control of the treatment effect. The same function and customization are allowed in the package *metapack*, which is mentioned earlier in the introduction, but not included in this review. Some packages have limited support for fitting covariates. For example, *BUGSnet* does not support categorical covariates with 3 or more levels.

Table 3: model specification

Tasks	Features	BATMAN	multinma	pcnetmeta	bnma	gemtc	BUGSnet
Model	Fixed Model	Y					
	Random Model						
	Baseline Model	Y	Y	N	Y	N	N
Covariates inclusion	Meta-regression	Y	Y	N	Y	Y	Y

While users carry out random effect relative treatment model, prior distribution options for in-between trial variance (heterogeneity parameter from random models) differed across packages. Packages *bnma* and *pcnetmeta* provide users with the option to specify the Wishart distribution for a multinomial response, and *BUGSnet* provides a Beta distribution. This is not currently supported by the BATMAN NMA, and neither multi-categorical response data are often applied. In addition, multinma has the most complete list of prior distribution selection. In *multinma*, there are three types of priors that allow flexibility: intercept, treatment effect, and heterogeneity (in-between trial standard deviation) parameter. The following distributions are all available in this package as an option for choosing heterogeneity prior: normal, half normal, log normal, Cauchy, half Cauchy, Student t, half Student t, log Student t, Exponential, Flat (non-informative prior).

A side note is if the prior is hard to determine, BATMAN can provide defaults for different likelihoods, but the user should conduct a short sensitivity analysis for these priors before running MCMC sampling.

After sampling, package *bnma* allows for finer control over deriving posterior samples by setting a convergence limit before running MCMC. It automatically checks for convergence using Gelman-Rubin diagnostics before sampling the full iteration amount and provides ERROR if the sequence fails to converge.

Table 4: MCMC and convergence assessment

Tasks	Features	BATMAN	multinma	pcnetmeta	bnma	gemtc	BUGSnet
Baseline and relative effect parameters prior	Normal distribution with heuristic initial values	Y	-	Y	Y	-	Y
	Restricted to change variance only	-	Y	-	-	Y	-
Heterogeneity prior	Uniform	Y	N	Y	Y	Y	Y
	Inverse Gamma	Y	N	Y	Y	Y	Y
	Half normal	Y	Y	N	Y	N	N
	Log normal	Y	N	N	N	Y	N
	Wishart	N	N	Y	Y	N	N
	Beta	N	N	N	N	N	Y
	Normal	N	Y	N	N	N	N
	Cauchy	N	Y	N	N	N	N
	Student t	N	Y	N	N	N	N
MCMC sampler	JAGS	Y	N	Y	Y	N	Y
	Stan	N	Y	N	N	N	N
	OpenBUGS	N	N	N	N	Y	N
	WinBUGS	N	N	N	N	Y	N
Control over posterior samples		See Table 6					
	multiple chain?	Y					
Assess convergence	trace plot	Y	N	N	Y	Y	Y
	Gelman-Rubin	Y	N	N	Y	Y	Y
	Automatic?	N	N	N	Y	N	N

Some packages, such as *multinma* and *bnma*, have comprehensive testing on inconsistency assumption checking, including several sensitivity models: node splitting, unrelated mean effects (UME), and inconsistent models. BATMAN offers an inconsistency checking option before modelling, and provides results when consistency is violated. Quantitatively, one can also check the goodness of fit using the DIC value. *BUGSnet* provides other measurements such as pD (effective number of parameters, the difference between the posterior mean deviance and the deviance at the posterior mean parameters) and Dres (deviance residual, the difference between the observed deviance and the expected deviance under the posterior predictive distribution). "pD" offers a way to penalize model complexity and "Dres" assesses the model fit with lower values indicating better fit.

While assessing the model's goodness of fit, the package *BUGSnet* does not support the calculation of leverage for zero cells.

As for effect measures, the following abbreviations were used: risk ratio (RR), odds ratio (OR), risk difference (RD), mean difference (MD), absolute risk (AR), event rate (ER), absolute standardized difference (ASD), hazard ratio (HR), and number needed to treat (NNT).

While BATMAN provides the essential SUCRA plots for each treatment, outputs from packages such as *multinma* and *rankinma* have more customized methods to show rankings. For example, *multinma* yields one comprehensive SUCRA and *rankinma* can visualize the probabilities of treatments among possible ranks in terms of line and stacked bar charts.

Table 5: Assumption testing and output

Tasks	Features	BATMAN	multinma	pcnetmeta	bnma	gemtc	BUGSnet
GOF	DIC	Y					

	pD						
	Dres						
Assumptions-homogeneity	Q-statistic	Y	N	N	N	N	Y
	Global I^2	N	N	N	N	Y	Y
Assumptions-consistency	Residual deviance	Y	Y	N	Y	N	N
	Inconsistency model	Y	Y	N	Y	N	Y
	Unrelated Mean Effect model	Y	Y	N	Y	Y	Y
	Unrelated Study Effect model	N	N	N	N	Y	N
	Node split model	deprecated	Y	N	Y	Y	N
Effect size measures	RR	Y	N	Y	Y	Y	Y
	OR	Y					
	RD	Y	N	Y	Y	N	Y
	HR	Y	Y	N	N	N	Y
	MD	Y	N	N	N	Y	Y
	AR	Y	Y	Y	N	N	N
	NNT	Y	N	N	Y	N	N
	ER	Y	N	Y	Y	N	N
	ASD	Y	N	N	N	N	N
Ranking	Probabilities	Y	Y	Y (1st only)	Y	Y	Y
	Rankograms	Y	Y	N	Y	Y	Y
	SUCRA	Y	Y	N	Y	N	Y
Others	Himalayan	Y	N	N	N	N	N
	Ridge	Y	N	N	N	N	N
	Forest	Y					

CASE STUDY SET-UP

For the case study, we applied the Baker 2009 dataset[11]. This data set comes from a systematic review of randomized controlled trials on pharmacologic treatments for chronic obstructive pulmonary disease (COPD) (Baker et al., 2009).

The primary outcome, the occurrence of one or more episodes of COPD exacerbation, was binary (yes/no). Five single drug treatments (*Fluticasone*, *Budesonide*, *Salmeterol*, *Formoterol*, and *Tiotropium*) and two mixed treatment arms (*Fluticasone + Salmeterol, Budesonide + Formoterol)* were compared to placebo. There were 39 studies with 29 2-arm studies and 10 multi-arm studies. Data from 28235 patients were collected.

To account for the limitation of functionalities across packages and to maximize the exchangeability between results, a random effect model without adjustment on baseline risk or inclusion of covariate is fitted. Since the response is binary, we apply the "logit" link function.

For the prior choice in random effect model, we standardized all are fitted with a treatment effect prior drawn from a normal distribution with mean 0 and standard deviation of 100, and the prior for in-between trial variation (or heterogeneity parameter) was chosen to be either uniform distribution between 0 and 10 (BATMAN, *bnma*, *pcnetmeta*, *gemtc*), or half normal distribution with a scale factor of 5 (*multinma*), depending on the distribution availability for in-between trial variation that each package supports.

The following standardized MCMC setup was applied:

Table 6: MCMC simulation setup

Package/ tool	adaptation	burn-in	iteration	maximum run if not converge	chain	thin factor
BATMAN	4,000	4,000	20,000	NA	4	1
multinma	NA	8,000	20,000	NA	4	1
pcnetmeta	5,000	3,000	20,000	NA	4	1
bnma	NA	NA	28,000	28,001	4	1
gemtc	8,000	NA	20,000	NA	4	1
BUGSnet	5,000	3,000	20,000	NA	4	Auto

RESULTS

During the case study, *pcnetmeta* results failed to converge. 200,000 iteration trial was tested out again but still showed lagging speed that led me difficulties to draw meaningful conclusions from it. The root cause is that *pcnetmeta* uses an arm-based algorithm so multivariate normal matrix invert operation is immensely complicated and computationally burdensome.

For this section, ranking and speed results from *pcnetmeta* are not comparable to the output yielded from other candidates or BATMAN, and they should be interpreted with caution because of unconverging samples.

Ranking

The following results apply treatment abbreviation in terms of arm labels. The label for each is composed of the first and last letter.

Arm	**Arm label**
Fluticasone	Fe
Budesonide	Be
Salmeterol	Sl
Formoterol	Fl
Tiotropium	Tm
Fluticasone + Salmeterol	Fe+Sl
Budesonide + Formoterol	Be+Sl
Placebo	PCB

Arm ranking heatmap

Table 7 is the arm rankings from different packages based on posterior probability ranking data, with the MCMC set up in Table 6. The rank is based on SUCRA plot when applicable or based on odds ratio point estimates otherwise.

Table 7: Treatment ranking (* represents a relatively wide 95% CrI)

Package	**Ranking (1st to 8th best to worst)**							
BATMAN	Be+Fl	Tm	Fe+Fl	Be	Sl	Fe	Fl	PCB
multinma	Be+Fl	Tm	Fe+Fl	Be	Sl	Fe	Fl	PCB
pcnetmeta	Fl*	Fe*	Tm*	Be*	Fe+Fl*	Sl*	Be+Fl*	PCB
bnma	Be+Fl	Tm	Fe+Fl	Be*	Sl	Fe	Fl	PCB
gemtc	Be+Fl	Tm	Fe+Fl	Sl	Be	Fe	Fl	PCB
BUGSnet	Be+Fl	Tm	Fe+Fl	Be	Sl	Fe	Fl	PCB

	BATMAN	**multinma**	**pcnetmeta**	**bnma**	**gemtc**	**BUGSnet**
Be+Fl	1	1	7	1	1	1
Tm	2	2	3	2	2	2
Fe+Fl	3	3	5	3	3	3
Be	4	4	4	4	5	4
Sl	5	5	6	5	4	5
Fe	6	6	2	6	6	6
Fl	7	7	1	7	7	7
PCB	8	8	8	8	8	8

From the heatmap (Figure 3), BATMAN has its ranking aligned perfectly with the ranks from *multinma, bnma, BUGSnet.* Package *gemtc* yields highly similar ranking results with packages above except for the opposite ranking for *Budesonide* and *Salmeterol* (noted as “Be” and “Sl”) in Table 7, and this can be explained by the small point estimate different between two drugs across all packages as well: the odds ratio point estimates of *Budesonide* and *Salmeterol* do not differ until the thousandth, and Budesonide shows a larger standard deviation, causing it ranking behind Salmeterol in *gemtc* simulation.

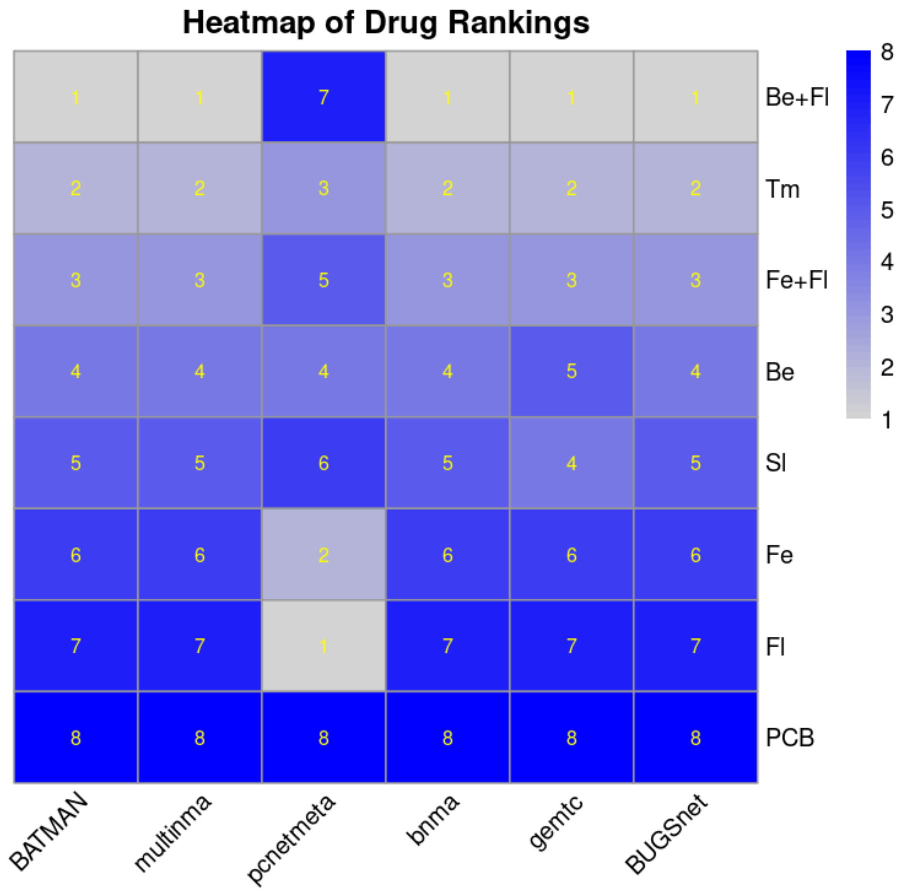


Figure 3: Treatment ranking heatmap

Relative effect forest plot

The forest plot (Figure 4) visualizes the relative effect sizes of seven treatment arms compared to the placebo. The x-axis depicts the odds ratio estimates from posterior distributions, with 95% credible intervals, and the y-axis lists the package with spacing applied between different drug classes for clarity. The plot uses color-coding for treatment arms. We also included a vertical dashed line at a null ratio of 1. Package *pcnetmeta* yields estimates with significantly different width for intervals, but otherwise results are highly consistent across packages.

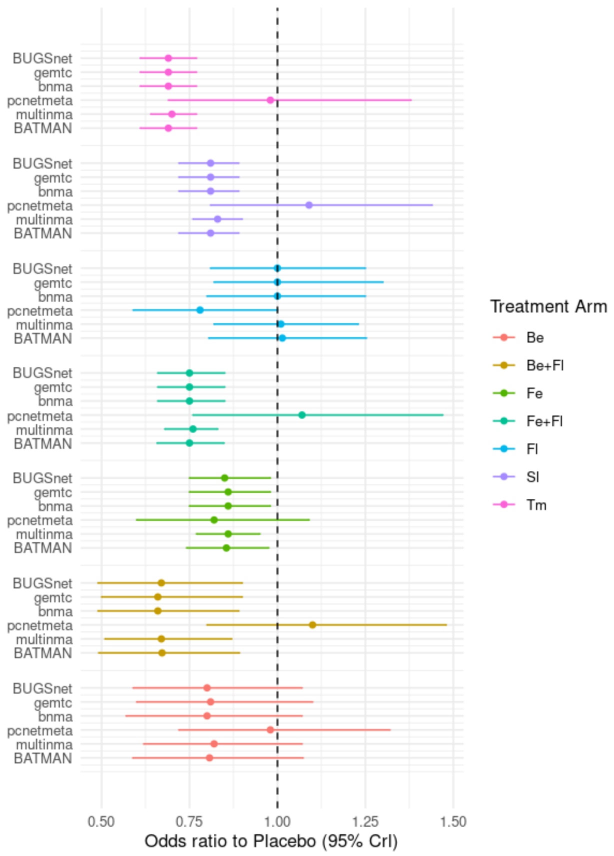


Figure 4: Relative effect forest plot by treatment

Computation speed

To assess computational efficiency of different packages, we incorporated the following in-line code in R to time MCMC sampling process:

```
start_time = sys.time()
[MCMC sampling code]
end_time = sys.time()
end_time – start_time
```

Time difference is printed as above and standardized across 10 runs. R package *pcnetmeta* is excluded from comparison because of its significant difference in computational efficiency, failure to complete adaptation and convergence.

Table 8: MCMC average run time

Average run time in minutes (standard deviation)
[*pcnetmeta* excluded]

BATMAN	multinma	bnma	gemtc	BUGSnet
0.6028 (0.0012)	0.9172 (0.021)	1.3413 (0.148)	0.861 (0.053)	0.6767 (0.008)

Table 8 and Figure 5 show the average run time, measured in minutes, from 5 package tools. Outliers are shown in red. BATMAN outperformed all and showed the greatest computation efficiency, with minimal difference from the second fastest package *BUGSnet*. However, while interpreting the data, there are several limitation and consideration to take in:

1. Number of studies: Computational time is influenced by the number of studies being analyzed. In this usage case study, there are 39 studies and 94 different arms in total. This variability can affect the reported times and comparisons across different programs.
2. Modeling differences: For BATMAN, the computational time is reflected in the relative effect model from the independent modeling that gets tested. In contrast, for other packages, the computational time is calculated holistically, including preprocessing steps, since the sampling function cannot be separated. This means that packages with less computational time demonstrate much superior computational efficiency compared to BATMAN, as preprocessing steps are accounted for in the total run time.
3. Variability from high-performance computing (HPC) environment: When running computations in a HPC environment, computational time can be affected by various factors such as cluster core availability, network load, and other concurrent processes. These factors introduce variability in timing measurements, which may not fully capture the efficiency of the programs under real-world conditions. It is advised to interpret run time by caution.

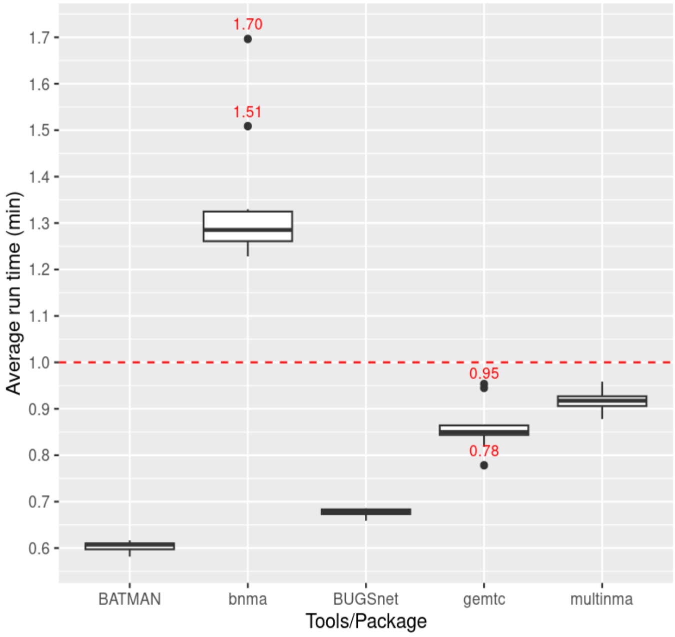


Figure 5: Average computation time for 28,000 samples (boxplot)

DISCUSSION AND SUMMARY

Advantages

BATMAN+ has strong advantages over other R packages when it comes to functionality. The biggest advantage of BATMAN+ is TFL synthesis: users can generate highly customized plots, reports, and PowerPoint slide in a few clicks. As for plots, users can select from Himalayan, ridge, forest, and network plot and select the treatment effects, measures of effects, and baseline effects. "Custom report" panel performs cross modeling comparison. Model variety stems from model specification, where users are prompted to select random or fixed for treatment effect and baseline effect and decide to fit baseline model and treatment effect model together (simultaneous) or separately (independent). Sometimes, researchers may find it time consuming to pull up data and relevant graphs from different model output folders, and this feature gives further advanced plots showing the measurement effects between treatments automatically. Besides automated process of TFL system, most output from BATMAN+ is in *'_output_'* folder, one can select the model and go to HTA subfolder to find more such as stacked bar chart on treatment ranks and rankograms.

In terms of data entry, before feasibility analysis and after data is loaded, users can freely edit cell entries and manipulate dataset.

To sum up, BATMAN+ offers great flexibility to adjust for different study specifications: versatile classes of outcome, data types (support both IPD and AgD data formats), model variety, and sensitivity analysis on different factors.

Table 9: Highlights for BATMAN

Features	Advantages
High-level	automated TFL system
	customized output (batch forest plots)
	efficiency: parameter file for future use
	cross model comparison reports
Data input	comprehensive class of outcome
	binomial to nominal conversion
	individual patient data (IPD) & aggregated (AgD) friendly
	flexible data manipulation
	simultaneous/ independent modelling option
	auto check validity on different models
Model Specification	random/ fixed baseline model
	baseline risk adjustment
Sensitivity	exclusion of studies
	exclusion of outlier

Specific Improvements

Table 10 listed improvements suggested from the review, a supplement to the insights drawn from the review above, which highlighted specific packages about their unique usage and functionality. For some, the package recommended to check out is in brackets if applied.

Recent main BATMAN updates (version 2.2) have included the feature to allow users to filter output files for meta-data in panel "custom reporting", session sharing (Beta version), and inclusion of different-level covariates within a single study.

Our next steps towards output reporting can be in two ways: 1) set filters to allow users choose BATMAN output, and 2) automate the process to generate Tables, Figures, and Listings (TFLs) based on current templates.

Feasibility panels can be extended from features in *BUGSnet*, which are explained earlier. To make it better for users, we should summarize different SUCRAs to one where coloring denotes the difference between treatment arms and provide a way to automatically check for convergence and give errors if failing. Such automation is achieved in *bnma*.

Figure 5 below is extracted from package *netmeta* (function *'netheat()'*) with a way to quantify "the change in inconsistency between direct and indirect evidence in a network estimate in the row after relaxing the consistency assumption for the effect of one design in the column"[12]. It includes an option to show the square-root of the between-study variance τ^2 for a random effects model between different studies as well as pairs of studies.

Table 10: Improvement suggestions for BATMAN

Features	What to do next?
Feasibility	add study information on network plot (now showing study count only) (*gemtc*)
	more feasibility panels (*BUGSnet*)
Model Specification	add Wishart distribution option for heterogeneity priors
Model Output	summarize SUCRA plots in one panel
	set output file format filters (.png, .txt, .pdf) to custom results
	create standard Tables, Figures, and Listings based on templates for Health Technology Assessments (HTA) submissions
Diagnostics	automated validation for convergence (*bnma*)
	provided reasons of failed run
	quantified inconsistency heatmap (*netmeta*, Figure 6)
User Interface/ User Experience	See S.3

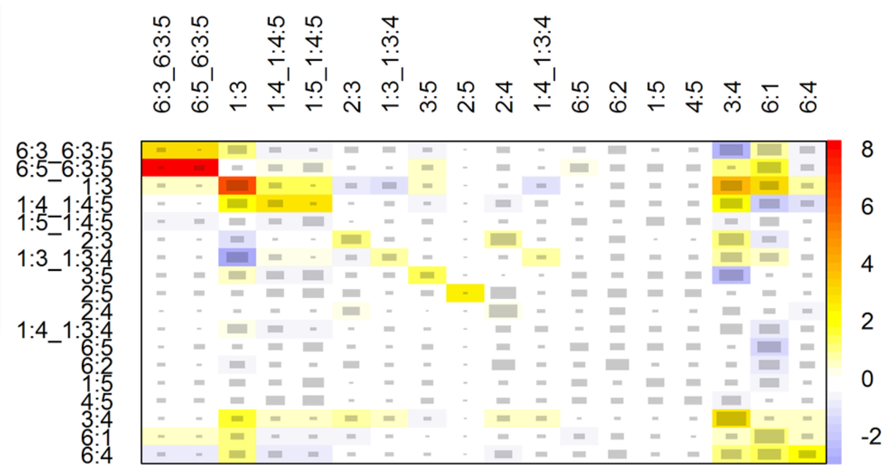


Figure 6: Inconsistency detecting heatmap (from *netmeta*)

Future Steps

Current pain points with the computational speed for IPD data and survival data exist. Previous usage of BATMAN has suggested improvement of the burden of computational time, and running time is manageable for reduced iterations. Moreover, testing on the susceptibility and reproducibility of results (ranking and relative effects measures) for different initial values, and seeding options is a next step to further validate result's consistency for BATMAN.

SUPPLEMENTARY

S.1: Simulation code on binary, Baker[11] data

Please refer to the deliverable folder.

S.2: NMA packages applied in BATMAN - more than NMA

Section	Feature	Package used
MA	fitting simple Bayesian meta-analysis models (continuous/ binary only. Author: Michael Sonksen)	*ma*
Arm-based analysis	Arm-Based Network Meta-Analysis for Binary Outcomes	*pcnetmeta*
	Arm-Based Network Meta-Analysis for Continuous Outcomes	
	Arm-Based Network Meta-Analysis for Binary Outcomes with Follow-up Time Reported	
	Arm-Based NMA for Count Datasets with Exposure Time in Person-Years Reported	
	treatment-specific absolute effect sizes plot	
	treatment-specific contrast effect sizes plot	
	network plot	
	treatment rank probability plot	
Feasibility	conduct network meta-analysis with contrast-based data	*netmeta*
	network plot	
	'comparison-adjusted' funnel plot	
	forest plot for network meta-analysis (network estimate)	
	forest plot for direct, indirect, and network estimate	
	create league table (square matrix showing all pairwise comparisons) with NMA results	
	probabilities of each treatment being at each possible rank and the SUCRAs	
	Frequentist method to rank treatments in network	
	design-based decomposition of Cochran's Q in NMA (assess the homogeneity/consistency)	
pacman	Bayesian NMA and NMR (covariate adjust) with Individual patient data, aggregate data, or both	*multinma*
	Set up a network containing individual patient data (IPD)	
	Set up a network containing arm-based aggregate data (AgD)	
	Set up a network containing contrast-based aggregate data (AgD)	
	Set up a network containing aggregate survival data (AgD)	
	Combine multiple data sources (*set_ipd, set_agd_arm, set_agd_contrast*) into one network	
	Add numerical integration points to aggregate data	
	fits network meta-analysis and (multilevel) network meta-regression models in Stan	
	Produce plots comparing the prior and posterior distributions of model parameters	
	Matrix of plots for a *stan_nma* object	
	Calculate the DIC for a model fitted using the *nma*() function	
	Obtain predictions of absolute effects from NMA models fitted with *nma*()	
	Produce posterior treatment rankings and rank probabilities from a fitted NMA model	
	produce population-adjusted relative effects	

S.3 User Interface, User Experience suggestions for BATMAN

1. Connect the big steps from session manager to custom results using arrows to indicate workflow.

2. In data cleaning step, the usage of "transform Binary-logit Input file to Ordered Categorical" is a niche use case. Therefore, it should better be put under a switch bar that asks user if they want to convert binary outcome to multinomial order. If the toggle is switched on, expand further options.

3. Under feasibility tab, the panel of “summaries by Treatment (r/n or y or y = log(median)” does not have its brackets closed properly. I would suggest renaming to “treatment descriptive statistics” and add a note about what this means under the panel tab.

REFERENCES

1. BUGSnet: Bayesian inference using Gibbs sampling to conduct network meta-analysis. Retrieved July 9, 2024, from https://bugsnetsoftware.github.io/vignettes
2. Tallarita, M., De Iorio, M., & Baio, G. (2019). A comparative review of network meta-analysis models in longitudinal randomized controlled trials. Statistics in Medicine, 38(16), 3053–3072. https://doi.org/10.1002/sim.8169
3. Liu, Y., Béliveau, A., Wei, Y., Chen, M. Y., Record-Lemon, R., Kuo, P. L., … Chen, G. (2022). A gentle introduction to Bayesian network meta-analysis using an automated R package. Multivariate Behavioral Research, 58(4), 706–722. https://doi.org/10.1080/00273171.2022.2115965
4. Dias, S., Welton, N. J., Sutton, A. J., et al. (2014). NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. London: National Institute for Health and Care Excellence (NICE). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK310366/
5. Neupane, B., Richer, D., Bonner, A. J., Kibret, T., & Beyene, J. (2014). Network meta-analysis using R: A review of currently available automated packages. PLOS ONE, 9(12), e115065. https://doi.org/10.1371/journal.pone.0115065
6. Phillippo, D. M. (2024). multinma: Bayesian network meta-analysis of individual and aggregate data (R package version 0.6.0). https://doi.org/10.5281/zenodo.3904454. Retrieved from https://dmphillippo.github.io/multinma/
7. Seo, M., & Schmid, C. (2024). bnma: Bayesian network meta-analysis using 'JAGS' (R package version 1.6.0). Retrieved from https://CRAN.R-project.org/package=bnma
8. Béliveau, A., Boyne, D. J., Slater, J., Brenner, D., & Arora, P. (2019). BUGSnet: An R package to facilitate the conduct and reporting of Bayesian network meta-analyses. BMC Medical Research Methodology, 19, 196. https://doi.org/10.1186/s12874-019-0798-0
9. Van Valkenhoef, G., & Kuiper, J. (2023). gemtc: Network meta-analysis using Bayesian methods (R package version 1.0-2). Retrieved from https://github.com/gertvv/gemtc
10. Lin, L., Zhang, J., Hodges, J. S., & Chu, H. (2017). Performing arm-based network meta-analysis in R with the pcnetmeta package. Journal of Statistical Software, 80(5), 1–25. https://doi.org/10.18637/jss.v080.i05
11. Baker, W. L., Baker, E. L., & Coleman, C. I. (2009). Pharmacologic treatments for chronic obstructive pulmonary disease: A mixed-treatment comparison meta-analysis. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 29, 891–905. https://doi.org/10.1592/phco.29.8.891
12. Balduzzi, S., Rücker, G., Nikolakopoulou, A., Papakonstantinou, T., Salanti, G., Efthimiou, O., & Schwarzer, G. (2023). netmeta: An R package for network meta-analysis using frequentist methods. Journal of Statistical Software, 106(2), 1–40. https://doi.org/10.18637/jss.v106.i02