*Research Synthesis Methods* (2026), **17**: 1018–1024
doi:10.1017/rsm.2026.10089

Research Synthesis Methods

RESEARCH-IN-BRIEF

# Personalized treatment hierarchies in Bayesian network meta-analysis*

Augustine Wigle and Erica E. M. Moodie

Epidemiology, Biostatistics and Occupational Health, McGill University, Canada

**Corresponding author**: Augustine Wigle; Email: augustine.wigle@mcgill.ca

**Received:** 17 November 2025; **Revised:** 19 February 2026; **Accepted:** 25 February 2026; published online 6 May 2026

**Abstract**

Network meta-analysis (NMA) is an increasingly popular evidence synthesis tool that can provide a ranking of competing treatments, also known as a treatment hierarchy. Treatment–covariate interactions (TCIs) can be included in NMA models to allow relative treatment effects to vary with covariate values. We show that in an NMA model that includes TCIs, treatment hierarchies should be created with a particular covariate profile in mind. We outline the typical approach for creating a treatment hierarchy in standard Bayesian NMA and show how a treatment hierarchy for a particular covariate profile can be created from an NMA model that estimates TCIs. We demonstrate our methods using a real network of studies for the treatment of major depressive disorder.

**Highlights**

**What is already known?**

- Network meta-analyses (NMAs) that include treatment–covariate interactions, such as individual participant data (IPD) NMAs, provide relative treatment effects conditional on covariate values.
- Treatment hierarchies created from ranking metrics are a popular way to summarize the results of a network meta-analysis.

**What is new?**

- We describe how existing ranking metrics can be easily adapted to provide personalized treatment hierarchies from an IPD NMA.

**Potential impact for RSM readers**

- Practitioners should consider creating tools that can compute ranking metrics in IPD NMAs to support personalized clinical decision-making.

## 1. Introduction

Network meta-analysis (NMA) is a statistical methodology that facilitates the estimation of relative effects for multiple treatments from a set of clinical trials. A convenient method for summarizing the results of a standard NMA is to construct a treatment hierarchy, that is, a list of the treatments ranked in order of preference. Treatment hierarchies are constructed using ranking metrics that consider the

This article was awarded Open Materials badge for transparent practices. See the Data availability statement for details.
*This article has been updated since original publication. A notice detailing the change has been published.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Society for Research Synthesis Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-ShareAlike licence (https://creativecommons.org/licenses/by-sa/4.0), which permits re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited.

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

magnitude and uncertainty of relative effect estimates, such as the Surface Under the Cumulative RAnking curve (SUCRA) in Bayesian NMA.[1,2] A treatment hierarchy is a desired output of many NMAs.[3]

When effect modifiers are present, incorporating treatment–covariate interactions (TCIs) into the NMA model can reduce heterogeneity and provide more precise estimates.[4–6] TCIs allow the expected relative treatment effects, and therefore treatment hierarchies, to differ between individuals with different covariate values, presenting an opportunity to compute personalized hierarchies. Authors have called for the reporting of treatment hierarchies at different covariate levels when TCIs are included in an NMA,[4–7] but this is rarely done in practice. In this article, we show how Bayesian NMAs with TCIs can produce personalized treatment hierarchies, and demonstrate the methods using a network of studies for major depressive disorder (MDD).

## 2. Treatment hierarchies in standard NMA models

In a standard NMA with no TCIs, the parameters representing treatment performance are given by $\psi_{20}^{(1)}, \psi_{30}^{(1)}, \ldots, \psi_{G0}^{(1)}$, where $\psi_{g0}^{(1)}$ is the expected relative effect of treatment $g$ compared to treatment 1, and the index 0 indicates that the parameter represents the main effect of the treatment and not an interaction, which will become important in the following section. Additionally, $G$ is the number of treatments in the network, and $\psi_{10}^{(1)} = 0$. We assume that the expected relative effect of treatment $g$ versus $h$, $\psi_{g0}^{(h)}$, is given by $\psi_{g0}^{(1)} - \psi_{h0}^{(1)}$ for any $g, h$ in $1, \ldots, G$. This is referred to as the consistency assumption in NMA literature.[8] In the remainder of the article, we suppress the superscript (1) and emphasize that all parameters are relative to treatment 1.

SUCRA values are calculated from the relative effects to obtain a treatment hierarchy that is relevant to the population represented by the NMA. The SUCRA value for treatment $g$ is calculated in a standard Bayesian NMA as

$$\begin{aligned} \text{SUCRA}(g) &= \frac{\sum_{s=1}^{G-1} \sum_{r=1}^{s} p_{gr}}{G-1} \\ &= \frac{G - \text{E}\left[rank(g)\right]}{G-1}, \end{aligned} \tag{2.1}$$

where $p_{gr}$ is the probability that treatment $g$ takes rank $r$, that is, it has a more favorable response than exactly $r$ treatments, and $\text{E}\left[rank(g)\right]$ is the average rank of treatment $g$.[2,9]

SUCRA values can be computed in a standard Bayesian NMA using samples from the posterior distribution. Assume that we have a sample of size $M$ from the posterior distribution of $\psi_{10} = 0, \psi_{20}, \ldots, \psi_{G0}$, and let $\psi_{g0(m)}$ represent the value of $\psi_{g0}$ in the $m^{\text{th}}$ posterior sample. Note that $\psi_{10} = 0$ by definition for every posterior sample. Then we can calculate SUCRA values as follows:

1. For each sample $m = 1, \ldots, M$ and each treatment $g = 1, \ldots, G$, determine the treatment rank according to

$$rank(g)_{(m)} = G - \sum_{h \neq g} I\left\{\psi_{g0(m)} > \psi_{h0(m)}\right\},$$

   where $I\{\cdot\} = 1$ when $\cdot$ is true and 0 otherwise.
2. Calculate the average rank for each treatment,

$$\text{E}\left[rank(g)\right] = \frac{1}{M} \sum_{m=1}^{M} rank(g)_{(m)}.$$

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

3. Calculate SUCRA$(g)$ using

$$\text{SUCRA}(g) = \frac{G - \text{E}\left[rank(g)\right]}{G-1}.$$

The SUCRA of treatment $g$ in a standard NMA is interpreted as the proportion of competing treatments that $g$ is expected to beat,[2] or the inverse-scaled average rank of treatment $g$,[9] *when it is employed in the population represented by the NMA*. The resulting treatment hierarchy is not personalized to any particular patient and depends on the covariate distribution in the population.[10]

## 3. Treatment hierarchies in NMA models with TCIs

TCIs can be estimated using aggregate data NMA models with meta-regression and Individual Participant Data (IPD) NMA models, including recently proposed Individualized Treatment Rule (ITR) NMA models.[11,12] In Supplementary Section S1, we describe the suitability of each model type for creating personalized treatment hierarchies. Briefly, we do not recommend using aggregate data NMA meta-regression to create personalized treatment hierarchies due to its risk of ecological bias and confounding.[11,13–15] IPD NMA can be used to produce personalized treatment hierarchies when care is taken in specifying the model.[6,7,16]

In an NMA model that includes TCIs, the interpretation of the parameter $\psi_{g0}$ does not correspond to the expected relative effect on average in the population. Instead, $\psi_{g0}$ corresponds to the expected relative effect of treatment $g$ compared to treatment 1 in the reference group. To illustrate, suppose there are $Q$ effect-modifying covariates included as TCIs in an NMA model. The expected relative effect of treatment $g$ versus treatment 1 for an individual with covariate values $X_1 = x_1, X_2 = x_2, \ldots, X_Q = x_Q$ is given by

$$\psi_{g0} + \psi_{g1}x_1 + \psi_{g2}x_2 + \cdots + \psi_{gQ}x_Q, \tag{3.1}$$

where $\psi_{g1}, \ldots, \psi_{gQ}$ are the coefficients representing the interaction between each covariate and treatment $g$ compared to treatment 1, that is, $\psi_{gq}, q > 0$ is the change in the relative effect of treatment $g$ compared to treatment 1 for a one unit increase in covariate $q$, and $\psi_{g0}$ is the expected relative effect when all $x_1, \ldots, x_Q$ equal 0.

Now, consider creating a treatment hierarchy from a model that includes TCIs. If we were to use only $\psi_{g0}$, $g = 1, \ldots, G$ to create the hierarchy for such a model, the hierarchy would apply to patients whose covariate values are all equal to zero. The hierarchy will be meaningless for patients with a different covariate profile. Clearly, for a treatment hierarchy to be meaningful in an NMA with TCIs, it should be intentionally created for a particular set of covariate values.

Let $\text{E}\left[rank(g) \mid x_1, \ldots, x_Q\right]$ be the average rank of treatment $g$ given covariates $x_1, \ldots, x_Q$. The average ranks are determined using the expected relative effects defined in equation (3.1). They can be calculated using samples from the posterior of $\psi_{gq}$, $g = 1, \ldots, G$ and $q = 0, \ldots, Q$

1. For each sample $m = 1, \ldots, M$ and each treatment $g = 1, \ldots, G$, determine the expected relative treatment effect given $x_1, \ldots, x_Q$ via

$$RE_{1g(m)}(x_1, \ldots, x_Q) = \psi_{g0(m)} + \psi_{g1(m)}x_1 + \psi_{g2(m)}x_2 + \cdots + \psi_{gQ(m)}x_Q \quad \text{for } g = 2, ..., G,$$

and $RE_{11(m)}(x_1, \ldots, x_Q) = 0$, where $RE_{1g(m)}(x_1, \ldots, x_Q)$ represents the expected relative effect of treatment $g$ versus 1 given covariates $x_1, \ldots, x_Q$ in sample $m$.

2. For each sample $m = 1, \ldots, M$ and each treatment $g = 1, \ldots, G$, determine the treatment rank according to

$$rank(g)_{(m)} = G - \sum_{h \neq g} I\left\{RE_{1g(m)}(x_1, \ldots, x_Q) > RE_{1h(m)}(x_1, \ldots, x_Q)\right\}.$$

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

3. Calculate the average rank for each treatment given the covariates,

$$\mathrm{E}\left[rank(g) \mid x_1, \ldots, x_Q\right] = \frac{1}{M}\sum_{m=1}^{M} rank(g)_{(m)}.$$

4. Calculate $\mathrm{SUCRA}(g)$ using

$$\mathrm{SUCRA}(g) = \frac{G - \mathrm{E}\left[rank(g) \mid x_1, \ldots, x_Q\right]}{G-1}.$$

The treatment hierarchy produced in this way is personalized to a patient with the covariate profile defined by $x_1, \ldots, x_Q$.

## 4. Example: Pharmacological treatments for MDD

We illustrate the use and impact of personalized treatment hierarchies using data from three studies investigating treatment for MDD. A systematic review process was not followed to select the studies; rather, the studies were selected based on the availability of IPD. Details on the study designs are given in Supplementary Section S2. The data provide information on six pharmacological treatments: Bupropion, Citalopram + Bupropion, Citalopram + Buspirone, Escitalopram, Sertraline, and Venlafaxine. The response is defined as the negative of the 17-question Hamilton Rating Scale for Depression (HRSD-17), so that larger values are preferred. The following covariates, considered to be potential effect modifiers for depression, were included as TCIs in the analysis: age, sex, marital status, years of education, employment status, household size, age of depression onset, number of depressive episodes, chronicity of current MDD episode, and HRSD-17 before receiving any treatment in the study.[17,18] We fit a two-stage common effect Bayesian ITR NMA model with IPD to estimate expected relative treatment effects in the MDD network,[12] described in detail in Supplementary Section S3.

We computed personalized treatment hierarchies using SUCRA for two hypothetical patients, Patient A and Patient B, each with unique covariate profiles: Patient A is employed while Patient B is unemployed, Patient A has experienced three or fewer depressive episodes while Patient B has experienced more than three, Patient A is male and Patient B is not, and Patient A lives in a household with two more members than Patient B. The specific covariate values for each patient are described in Supplementary Section S3.3.

### *4.1. Results*

A plot showing point estimates and 95% credible intervals for treatment effects and TCIs is shown in Supplementary Section S3.4. All credible intervals contain zero. This is not too surprising, since only three studies were included in the analysis. In Figure 1, we show the posterior distribution of the expected relative effects as defined in equation (3.1) of each treatment compared to Sertraline for Patients A and B. It is clear that there are differences in the posterior densities between the patients. The SUCRA values for Patients A and B reflect the differences in the posteriors, and are compared in Table 1. For example, Citalopram + Bupropion is ranked first for Patient A, while it is ranked only fourth out of the six treatments for Patient B. Table 1 also compares the SUCRA and treatment rankings that would result from using a naive method that only considered the main treatment effects, $\psi_{g0}$. The naive method gives the same hierarchy for both patients, and this differs from both personalized hierarchies, particularly for Patient B, who deviates more from the reference profile.

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

***Table 1.*** *SUCRA values and ranking of each treatment using a naive nonpersonalized approach, and using a personalized approach for Patient A and Patient B.*

| Treatment | Naive | | Patient A | | Patient B | |
|---|---|---|---|---|---|---|
| | SUCRA | Rank | SUCRA | Rank | SUCRA | Rank |
| Bupropion | 0.30 | 5 | 0.22 | 6 | 0.59 | 3 |
| Citalopram+Bupropion | 0.81 | 1 | 0.96 | 1 | 0.50 | 4 |
| Citalopram+Buspirone | 0.79 | 2 | 0.66 | 2 | 0.63 | 2 |
| Escitalopram | 0.20 | 6 | 0.33 | 4 | 0.10 | 6 |
| Venlafaxine | 0.53 | 3 | 0.52 | 3 | 0.87 | 1 |
| Sertraline | 0.38 | 4 | 0.30 | 5 | 0.31 | 5 |

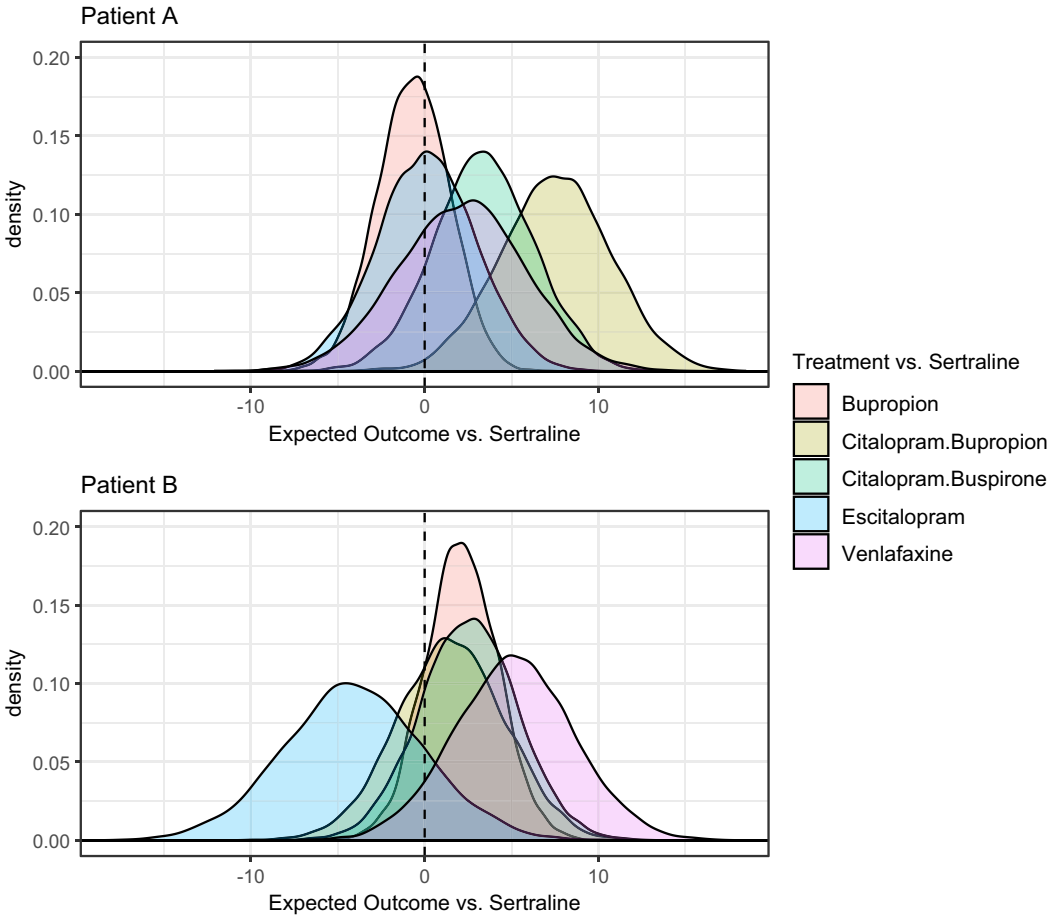



***Figure 1.*** *Posterior distributions of the relative effects of each treatment for Patients A and B.*

## 5. Discussion

Personalized treatment hierarchies can be computed from IPD NMA models that include TCIs. We demonstrated the methods using an ITR NMA of treatments for MDD, where the TCIs were selected based on previous studies on MDD effect modifiers. The posterior distributions of the expected relative

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

effects changed for different patient profiles and resulted in different treatment hierarchies, suggesting possible value in personalizing recommendations.

We caution against relying on the statistical significance of TCI parameters to decide when it would be beneficial to create personalized treatment hierarchies. First, TCIs that exist in practice but have not yet been estimated precisely enough to reach statistical significance are still important to include in the model, and a personalized treatment hierarchy could be useful in this context. Additionally, although the $\psi_{gq}$ parameters may not be statistically significantly different from zero, these parameters represent interactions compared to the reference treatment, and it is possible that other interaction terms may be statistically significant for other treatment comparisons, for example, the credible interval for the interaction of age with Bupropion versus Sertraline may contain zero, but the credible interval for the interaction of age with Bupropion versus Venlafaxine may be well away from zero. Therefore, an interaction that does not seem important at first glance may indeed be important for other treatment comparisons.

On the other hand, including TCIs that do not exist in practice will increase the variance of the posterior distribution of conditional relative effects without improving the accuracy of treatment recommendations. The use of the SUCRA ranking metric to create the hierarchy will mitigate the additional variance to some extent, as SUCRA summarizes the entire posterior distribution of relative effects,[19] but personalized treatment hierarchies may still be misleading if the included TCIs do not exist. For this reason, we believe personalized treatment hierarchies should be created when there is additional reason to believe that the included TCIs do exist in practice, besides only the NMA results at hand.

We focused on creating treatment hierarchies from the expected relative effects. In random-effects NMA models, treatment effects and interactions are allowed to vary across studies, and a personalized treatment hierarchy that better reflects uncertainty could be obtained by considering this additional predictive variability. A Bayesian approach to creating predictive treatment hierarchies in standard NMAs has been proposed[20] and could be adapted to personalized treatment hierarchies.

We echo previous calls to report treatment hierarchies for different covariate profiles when possible and when there is evidence that TCIs exist.[4–7] If many covariates are effect modifiers, it may not be feasible to report hierarchies for all covariate patterns of interest. Some recent IPD NMAs have included online tools that allow individuals to compute expected relative effects for custom covariate values to aid in decision-making.[21,22] With little additional effort, personalized treatment hierarchies could be included in the output of such tools. We encourage researchers conducting IPD NMAs to consider creating tools that can provide personalized treatment hierarchies to support clinical decision-making.

**Acknowledgments.** Data and/or research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Dataset identifier(s): #2199 (EMBARC), #2153 (REVAMP), #2148 (STAR*D). This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA.

**Author contributions.** Conceptualization: A.W., E.E.M.M.; Data curation: A.W.; Formal analysis: A.W.; Methodology: A.W.; Software: A.W.; Supervision: E.E.M.M.; Visualization: A.W.; Writing—original draft: A.W.; Writing—review and editing: A.W., E.E.M.M.

**Competing interest statement.** The authors declare that no competing interests exist.

**Data availability statement.** Code and data to reproduce the results in this article are available at https://doi.org/10.5281/zenodo.19007228.

**Funding statement.** A.W. is supported by the Natural Sciences and Engineering Research Council of Canada (Grant No. PDF-598932-2025). E.E.M.M. is a Canadian Institutes of Health Research Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine.

**Ethical standards.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

**Supplementary material.** To view supplementary material for this article, please visit http://doi.org/10.1017/rsm.2026.10089.

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press

## References


[1] Salanti G, Nikolakopoulou A, Efthimiou O, Mavridis D, Egger M, White IR. Introducing the treatment hierarchy question in network meta-analysis. *Am J Epidemiol*. 2022;191: 930–938.

[2] Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol*. 2011;64: 163–171.

[3] Papakonstantinou T, Salanti G, Mavridis D, Rücker G, Schwarzer G, Nikolakopoulou A. Answering complex hierarchy questions in network meta-analysis. *BMC Med Res Methodol*. 2022;22: 47.

[4] Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: individual patient data may be beneficial if only for a subset of trials. *Stat Med*. 2013;32: 914–930.

[5] Freeman SC, Fisher D, Tierney JF, Carpenter JR. A framework for identifying treatment-covariate interactions in individual participant data network meta-analysis. *Res Synth Methods*. 2018;9: 393–407.

[6] Riley RD, Dias S, Donegan S, et al. Using individual participant data to improve network meta-analysis projects. *BMJ Evid Med*. 2023;28: 197–203.

[7] Riley RD, Debray TP, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Stat Med*. 2020;39: 2115–2137.

[8] Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: continuing controversies and recent developments. *Res Synth Methods*. 2024;15: 702–727.

[9] Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15: 58.

[10] Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc*. 2020;183: 1189–1210.

[11] Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*. 2008;27: 1870–1893.

[12] Shen J, Moodie EEM, Golchi S. Two-stage Bayesian network meta-analysis of individualized treatment rules for multiple treatments with siloed data. *Stat Methods Med Res*. 2026;35: 3–20.

[13] Donegan S, Williamson P, D'Alessandro U, Smith CT. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. *Stat Med*. 2012;31: 3840–3857.

[14] Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med*. 2002;21: 371–387.

[15] Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol*. 2004;57: 683–697.

[16] Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out {W}ithin-trial and {a}cross-trial information. *Stat Med*. 2017;36: 772–789.

[17] Kessler RC, van LHM, Wardenaar KJ, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci*. 2017;26: 22–36.

[18] Perlman K, Benrimoh D, Israel S, et al. A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *J Affect Disord*. 2019;243: 503–515.

[19] Chiocchia V, Nikolakopoulou A, Papakonstantinou T, Egger M, Salanti G. Agreement between ranking metrics in network meta-analysis: an empirical study. *BMJ Open*. 2020;10: e037744.

[20] Rosenberger KJ, Duan R, Chen Y, Lin L. Predictive P-score for treatment ranking in Bayesian network meta-analysis. *BMC Med Res Methodol*. 2021;21: 213.

[21] Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive Behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*. 2021;78: 361–371.

[22] Furukawa TA, Suganuma A, Ostinelli EG, et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. *Lancet Psychiatry*. 2021;8: 500–511.


**Cite this article:** Wigle, A, Moodie EEM. Personalized treatment hierarchies in Bayesian network meta-analysis. *Research Synthesis Methods*. 2026;17: 1018–1024. https://doi.org/10.1017/rsm.2026.10089

https://doi.org/10.1017/rsm.2026.10089 Published online by Cambridge University Press