

基于贝叶斯模型平均的元分析：原理与实现

任子伟¹, 刘铮^{2*}, 胡传鹏^{1*1}

(¹ 南京师范大学心理学院, 南京, 210097)

(² 香港中文大学(深圳)人文社科学院, 深圳, 518172)

摘要 元分析作为一种综合已有研究成果的重要统计方法, 在量化研究中广泛应用。然而, 研究者在分析过程中常面临模型选择的困境: 在处理研究间异质性时, 需在固定效应与随机效应模型之间做出选择; 在应对潜在发表偏倚时, 又需要从多种校正模型中选定一种。目前, 这些模型选择仍缺乏统一标准。贝叶斯统计框架下的贝叶斯模型平均 (Bayesian Model Averaging, BMA) 为解决这些问题提供了新的思路: 它将不同统计模型纳入同一模型空间, 量化各模型获得数据支持的程度, 并据此对效应量进行加权, 从而规避单一模型选择带来的不确定性。基于 BMA 的元分析能同时检验三个关键假设(效应是否存在、异质性是否存在以及发表偏倚是否存在), 并以模型平均的方式实现对效应量的稳健估计。该方法可通过开源软件 JASP 或 R 语言实现, 为研究者实施元分析提供了新选择。

关键词 贝叶斯模型平均 元分析 模型不确定性 发表偏倚 稳健贝叶斯元分析

1 引言

元分析(meta-analysis, 又译为荟萃分析)通过对多个独立研究中的效应量进行统计整合, 实现对某一主题研究进展的阶段性、量化总结(任志洪, 赖丽足, 2023; Glass, 1976)。该方法将相同主题但来自不同研究的结果转化为统一的效应量指标(如标准化均值差、相关系数等), 从而评估效应的整体方向、大小及其异质性来源。自 Glass 于 1976 年首次提出以来, 元分析作为累积证据的关键方法, 已广泛应用于心理学、教育学、医学等科学领域。

元分析本质上是分层模型(hierarchical model), 纳入元分析的不同研究被视为同一个模型中的不同层级。理想情况下, 模型的最底层是单个被试的观测数据。然而, 由于原始文献通常仅报告研究层面的汇总数据(如效应量及其标准误), 而非被试的个体数据。元分析模型将这些含有抽样误差的估计值作为输入, 并利用标准误差量化因抽样随机性导致的研究内变异。图 1 展示了最典型的两层元分析的结构。在实际应用中, 当同一研究报告多个效应量或存在其他依赖结构时, 模型可进一步扩展为多层元分析(multilevel meta-analysis; Van den Noortgate et al., 2013, 2015), 本文暂不展开讨论。这种“数据嵌套于研究”的结构, 在经典的频率主义框架和贝叶斯框架下均可以实现。

¹ *通讯作者: 胡传鹏, hcp4715@hotmail.com; 刘铮, zhengliu2@link.cuhk.edu.cn

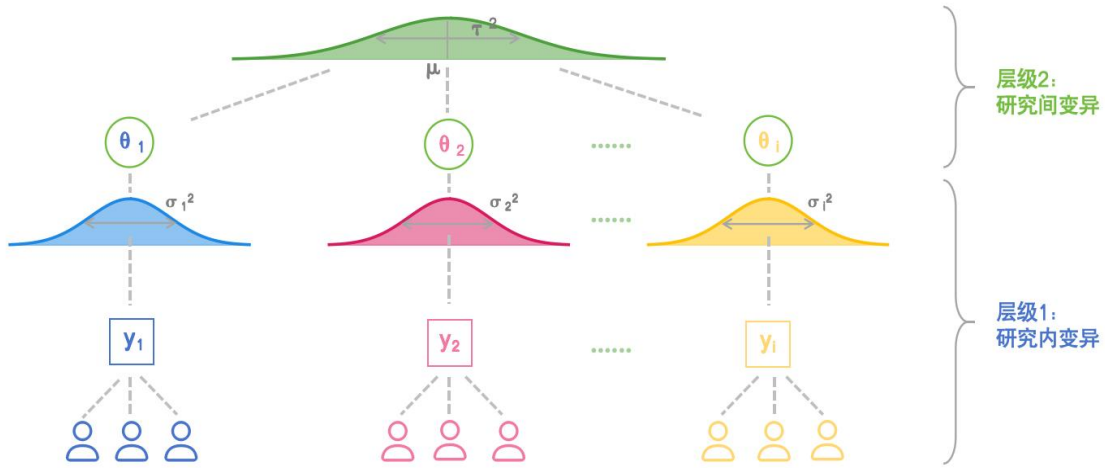


图 1 元分析分层模型结构示意图

注：该示例模型包含两个变异层级：(1) 研究间变异(层级 2)：顶部绿色曲线表示效应量的总体分布 $N(\mu, \tau^2)$ ，其中 μ 为总体平均效应， τ^2 为研究间异质性。每个研究的真实效应量 θ_i (圆形) 被视为从该总体分布中随机抽取的样本。(2) 研究内变异(层级 1)：底部代表各研究中的个体被试，个体数据汇聚成观测到的汇总效应量 y_i (方块)。中间的彩色曲线表示各研究的抽样分布：给定真实效应 θ_i 时，观测值 y_i 的抽样分布 $y_i \sim N(\theta_i, \sigma_i^2)$ 。这反映了由于样本量有限而产生的随机误差，其宽度由方差 σ_i^2 决定。假设不存在发表偏倚，各独立研究的观测效应量 y_i 以总体平均效应 μ 为中心呈正态分布波动，变异性由抽样误差 σ_i^2 和研究间异质性 τ^2 共同构成，其概率密度函数表示为：

$$f(y_i | \mu, \tau^2; \sigma_i^2) = \frac{\phi\left(\frac{y_i - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right)}{\sqrt{\sigma_i^2 + \tau^2}}$$

元分析既可以在传统频率主义框架之下进行，也可以在贝叶斯框架下实现。在频率主义视角下，总体效应量被视为固定但未知的真实参数(Lehrer et al., 2023)，元分析的目标是通过综合已有的研究来估计该总体效应。其推断逻辑建立在“重复抽样”的假说之上，将纳入元分析的每项研究视为对该参数的一次带有噪声的估计。这种框架在处理元分析的分层结构时存在局限：它将研究间异质性 τ 估算为一个固定值，容易忽略 τ 本身的不确定性。相比之下，贝叶斯框架将总体效应量 μ 和异质性 τ 均视为随机变量，通过分布进行表示。这些控制个体研究分布的参数被称为总体参数，因其所处的分层逻辑层级在个体参数之上，亦被称为超参数(hyper-parameters)，可以将各层不确定性统一在一个概率

模型中([Grant & Di Tanna, 2025](#))。这种全概率框架能够将异质性估算的不确定性完整地保留在效应量估计中([Harrer et al., 2021](#); [Kruschke & Liddell, 2018](#))。

研究者在实施元分析时通常会面临两个需要进行主观选择的难题：纳入研究的异质性与潜在的发表偏倚。前者指不同研究间效应量的真实差异([Sedgwick, 2015](#))，后者指统计显著或符合预期方向的研究结果更容易被发表，从而导致效应量高估([Ferguson & Brannick, 2012](#))。在处理异质性时，研究者需在固定效应模型(假设 $\tau^2 = 0$)与随机效应模型(假设 $\tau^2 > 0$)之间选择；在处理发表偏倚时，又需在 PET-PEESE(precision-effect test and precision-effect estimate with standard error; [Stanley & Doucouliagos, 2014](#))、发表筛选模型(selection models)([Iyengar & Greenhouse, 1988](#); [McShane et al., 2016](#))、剪补法([Duval & Tweedie, 2000](#))等多种校正方法中做出选择([Carter et al., 2019](#); [Hedges, 1992](#))。然而，这些模型的选择缺乏广泛认同的标准，不同方法可能得出不一致的结论([Almalik, 2025](#); [Carter et al., 2019](#); [Ioannidis, 2008](#))。

贝叶斯模型平均(Bayesian Model Averaging, BMA)为解决上述问题提供了新思路：将所有合理的模型纳入同一模型空间，计算各个模型的后验概率，并以此为权重对不同模型所估计出的总体效应值进行加权，从而将模型选择的不确定性直接整合到统计推断中([Berkhout et al., 2024](#); [Fragoso et al., 2018](#); [Hinne et al., 2020](#))。针对异质性问题，BMA 将固定效应与随机效应模型纳入同一框架进行加权平均；针对发表偏倚问题，BMA 进一步将无偏倚模型、PET-PEESE 模型以及发表筛选模型等多种校正方法同时纳入。这一思想发展出的元分析方法被称为稳健贝叶斯元分析(Robust Bayesian Meta-Analysis, RoBMA)，可通过 JASP 和 R 软件包实现([Bartoš et al., 2022, 2023](#); [Maier et al., 2023](#))。

本文首先简述贝叶斯元分析的基础，进而重点介绍 BMA 的方法与原理。在此基础上，以 Liu 等(2024)的公开数据为例，演示在 JASP 中进行 RoBMA 的完整操作流程，包括数据准备、模型设定、结果解读与可视化呈现，并在补充材料中展示 R 语言的实现。最后，探讨该方法的局限性并列出使用 RoBMA 时建议报告的内容。需要指出的是，元分析的严谨性受到学术界的高度重视，相关学术组织及期刊已发展出规范性报告指南，如医学领域的 PRISMA 报告规范([Moher et al., 2010](#))和心理学领域的 PRIOR-MA 报告清单([刘宇 等, 2021](#); [Appelbaum et al., 2018](#))。本文聚焦于统计方法层面的改进，所介绍的贝叶斯模型平均方法可与上述报告指南结合使用。

2 贝叶斯元分析基础

贝叶斯推断与频率推断的核心差异在于对概率的定义：频率主义视概率为长期重复抽样中事件发生的客观频率(e.g., [Hennig, 2024](#))；贝叶斯主义则将概率理解为随机变量，其分布随着观测数据的输入而发生变化。例如，在获得任何信息之前，对某一参数的信念本身就是一个分布(即先验分布)；在获得观测数据之后，依据贝叶斯定理将先验分布与观测数据相结合，即可得到后验分布(e.g., [van de Schoot et al., 2021](#))。该更新过程的数学表述为：

$$p(\theta | \text{data}) = \frac{p(\text{data} | \theta)p(\theta)}{p(\text{data})}$$

其中, $p(\theta)$ 是为先验分布, 表示在观测数据之前对参数 θ 的概率分布; $p(\text{data}|\theta)$ 为似然函数, 即给定参数 θ 下数据集的概率或概率密度; $p(\text{data})$ 为归一化常数; $p(\theta|\text{data})$ 为后验分布, 即结合观测数据后对参数 θ 概率分布的更新。贝叶斯推断基于后验分布展开, 使得研究结论不再局限于接受或拒绝零假设, 而是能够对参数估计和模型比较提供直接的概率性解释。

当处理心理学研究中常见的嵌套数据(如试次嵌套于被试、被试嵌套于实验条件)时, 传统的独立参数估计往往忽略数据的群聚特征。层级模型通过引入分层架构, 有效整合了这类嵌套关联, 解决了这一问题: 个体水平的参数不再被视为互不相关的独立取值, 而是被假定共同服从于一个更高层级的群体分布。在数学上, 这一拓展将原有的单一先验分布进一步分解(Si et al., 2025)。假设 θ 代表各独立的个体参数, Φ 代表控制其分布的总体参数(即超参数, 如总体均值 μ 和异质性 τ), 联合后验分布(省略归一化常数)可表示为:

$$p(\theta, \Phi | \text{data}) \propto p(\text{data} | \theta) p(\theta | \Phi) p(\Phi)$$

其中, $p(\theta, \Phi | \text{data})$ 代表联合后验分布; $p(\text{data}|\theta)$ 为似然函数; $p(\theta|\Phi)$ 称为层级先验, 将个体参数 θ 约束在由总体参数 Φ 决定的群体分布之中; $p(\Phi)$ 则是总体参数的先验分布称为超先验, 代表对群体参数 Φ 的初始信念。

层级模型通过群体层面的分布来汇总被试间的信息, 从而有效估计随机的被试效应。个体层面的参数估计受到群体分布的约束: 个体的观测数据越极端、参数估计越不确定, 其最终的后验估计值就越向群体均值靠拢(Lee, 2011; Veenman et al., 2024)。这种局部池化能够有效缓解个人层面极值的影响, 通常产生更准确的个体层面估计。然而, 由于贝叶斯层级模型结构复杂, 联合后验分布通常缺乏闭式解析解, 研究者需依赖马尔可夫链蒙特卡洛(Markov Chain Monte Carlo, MCMC)方法进行近似求解。MCMC 在参数空间中迭代抽样, 经过充分迭代与预热后达到收敛, 最终以样本集合描绘后验分布的形态。近年来, 已有成熟的工具包支持贝叶斯层级模型的实现(胡传鹏, 王继贤, 2025; 潘晚珂 等, 2023)。

贝叶斯框架的优势不仅体现在单个模型的层级化构建上, 更在于它能将模型本身视为具有概率分布的随机变量, 从而纳入统计推断。例如, 传统元分析要求研究者在固定效应与随机效应模型之间做出非此即彼的确定性决策; 而贝叶斯框架则允许将这两个模型视为概率空间中的离散随机变量, 通过计算似然评估观测数据对不同模型的支持程度。具体而言, 该方法量化各模型的后验模型概率, 将其作为权重对各模型下的参数估计值进行加权整合。这一过程将模型选择的不确定性直接转化为权重分配, 并纳入最终的效应量估计中, 从而实现基于模型平均的元分析。这种方法避免了单一模型选择

可能带来的偏差，提升了元分析结果的稳健性(Gronau et al., 2021; Hinne et al., 2020; Maier et al., 2023)。下文将介绍模型平均的思想如何在元分析中应用。

3 元分析中的模型选择困境与模型平均的解决方案

元分析实践中面临两个主要的模型选择困境：第一，是否假定研究间存在异质性(固定效应 vs. 随机效应)；第二，是否假定存在发表偏倚以及采用何种方法进行校正。传统的做法是将这两个问题分开处理：先根据异质性检验选定元分析模型，再主观择取某种偏倚校正方法对其参数估计进行校正。贝叶斯模型平均(Bayesian Model Averaging, BMA)不强行选择某个模型，而是将所有合理的模型纳入模型空间，根据数据对各模型的支持程度(模型后验概率)进行加权平均。

BMA 在元分析中的应用，体现为模型空间根据其处理不确定性维度增加而不断拓展。基础 BMA 通过整合“效应是否存在”和“异质性是否存在”两个核心维度，构建四个候选模型组成的模型空间，旨在解决研究间的异质性问题。RoBMA 则进一步把发表偏倚校正的不确定性纳入模型空间，将多种发表偏倚校正的模型与 BMA 的四种基础模型相结合，构建出更广阔的模型空间。不仅实现了对“是否存在发表偏倚”的概率化推断，还依据各模型的后验概率加权，得到更稳健的综合效应量估计(见图 2)。下文将循序渐进，先探讨 BMA 对异质性问题的解决(3.1 节)，再阐述 RoBMA 如何将这一思想延伸至发表偏倚校正领域(3.2 节)。

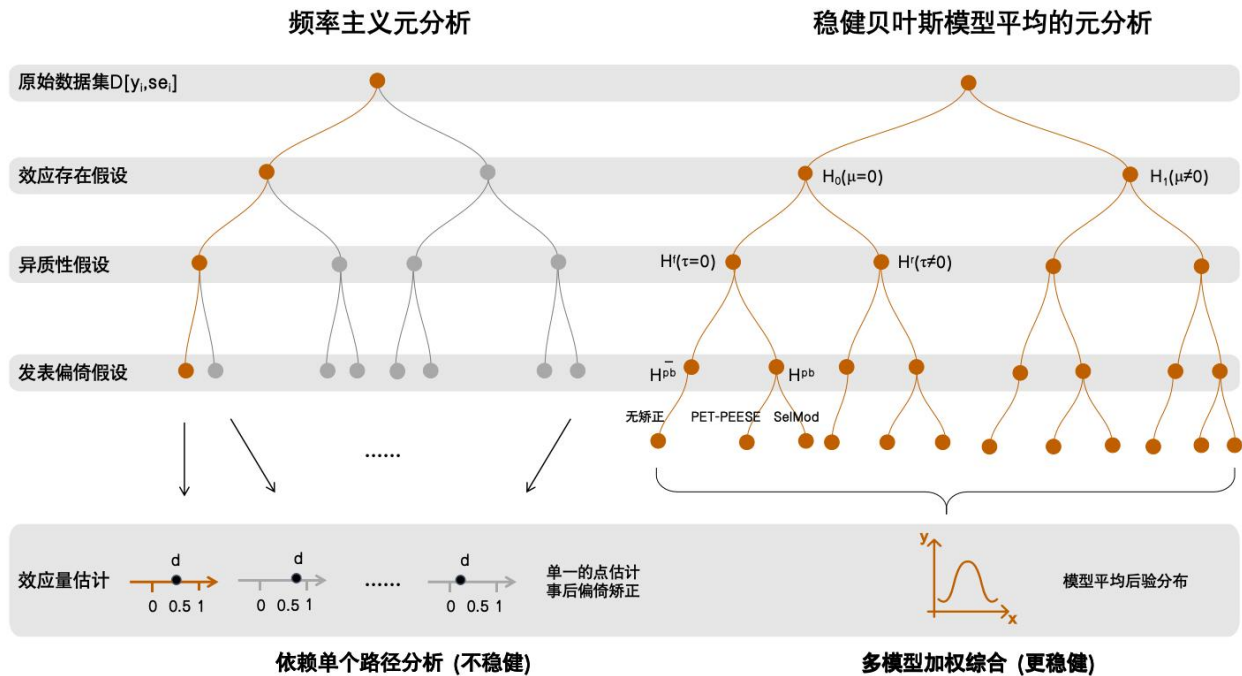


图 2 RoBMA 与传统频率主义元分析在分析路径不确定性上的对比

注：在元分析中，研究者从原始数据出发，在效应假设、异质性及偏倚校正等数据分析节点上均存在多种选择，形成多条潜在的平行分析路径。传统频率主义元分析(左侧)仅在多条路径中人为选定一条(如橙色的路径)，输出该路径下的点估计值，忽略其他可能同样具有合理性的分析路径；而 RoBMA 方法(右侧)则将所有可能的多个路径均纳入统一的模型空间，并通过 BMA 实现对全部路径信息的加权整合，输出更稳健的效应量后验估计，避免由于“人为选定单一路径”而导致不稳健的结论。

3.1 文献异质性带来的模型选择问题及 BMA 解决方案

由于纳入元分析中的研究可能存在异质性，传统方法需要研究者在固定效应与随机效应模型之间进行选择。这种选择具有一定的主观性，且忽略了模型本身的不确定性。与传统方法强行选定单一模型不同，BMA 将模型层面的不确定性转化为概率权重，从而在统一推断过程中实现对固定效应与随机效应估计的加权整合(Himne et al., 2020)。BMA 的推断流程可概括如下：

首先，定义模型空间。依据元分析中总体效应(层级模型超参中的 μ)和研究间异质性(层级模型超参中的 τ)是否存在不同假设进行交叉组合，模型空间可以划分为以下 4 个模型(Gronau et al., 2021)：

模型 M1(H_0^f)：固定效应零模型($\mu = 0, \tau = 0$)；

模型 M2(H_1^f)：固定效应备择模型($\mu \sim g(\cdot), \tau = 0$)；

模型 M3(H_0^r)：随机效应零模型($\mu = 0, \tau \sim h(\cdot)$)；

模型 M4(H_1^r)：随机效应备择模型($\mu \sim g(\cdot), \tau \sim h(\cdot)$)。

这 4 个模型同属元分析层级模型，差异表现在对效应存在性与异质性的参数先验设定(各模型的数学形式见附录 1)。

其次，设定模型与参数的先验分布。在贝叶斯框架下，各模型成立的可能性即为先验模型概率。由于各模型代表了互斥的理论假设，其概率分布为离散概率分布。如果研究者在分析前对上述四个模型缺乏特定偏好，通常遵循无偏好先验原则，赋予四个模型相等的先验概率(各 0.25)。反之，若明确理论依据，也可调整先验权重(如设定 M4 的概率为 0.4，其余三者均为 0.2)。由于四个模型均为元分析层级模型，因此各模型内的未知参数也需预设先验分布，才能进行贝叶斯推断。在零假设模型(M1 和 M3)中，关键参数(总体效应 μ 和研究间异质性 τ)采用位于零点的尖峰分布 Spike(0)作为先验，即将其在实数域内所有概率质量分配给 0；而在备择假设模型(M2 和 M4)中，参数被赋予连续分布先验： μ 默认采用标准正态分布 Normal(0, 1)， τ 则默认使用基于实证数据推导的逆伽马分布 Inverse-Gamma(1, 0.15)(Gronau, Van Erp, et al., 2017)。除上述默认先验外，BMA 也支持研究者基于领域专家知识设定有信息的先验分布。模型与参数的先验设定共同构成研究者的初识认知，并将通过贝叶斯更新过程与观测数据融合，推导出对应的后验分布。

接着，各模型与参数的先验依据贝叶斯定理更新为后验。在模型层面，每个模型后验概率 $p(H_k | Data)$ 的计算遵循贝叶斯公式：

$$p(H_k | Data) = \frac{p(Data | H_k) \cdot p(H_k)}{\sum_{j=1}^4 p(Data | H_j) \cdot p(H_j)}$$

其中， k 指向当前待评估的模型， j 是求和过程中的临时变量，用于累加所有模型的联合概率。 $p(Data | H_k)$ 是当前待评估模型的似然，衡量该模型预测数据的准确性； $p(H_k)$ 是先验模型概率，通常默认设为等权重的 0.25。分母为归一化常数，确保所有后验概率之和为 1。该环节输出的后验概率值量化了数据对各模型的相对支持程度，并作为核心权重用于后续的模型平均推断。在参数层面，每个备择模型(M2 和 M4)的先验分布通过观测数据的似然函数更新，生成参数的后验分布 $P(\mu | data, \mathcal{H}_k)$ 。这一过程实质上是贝叶斯定理在连续参数空间上的应用，实践中通常采用 MCMC 算法进行模拟抽样以获得近似解。

随后，利用包含式贝叶斯因子(Inclusion BF)评估效应及异质性的存在性。贝叶斯推断通过贝叶斯因子(Bayes factor, BF)量化观测数据对不同统计假设的相对证据强度(胡传鹏 等, 2018; 吴凡 等, 2018)，本质上是两个模型似然之比，反映模型对观测数据预测能力的差异(王允宏 等, 2023)。在 BMA 框架下，由于涉及多个模型，需采用包含式贝叶斯因子整合整个模型空间的信息。

当检验“总体效应是否为零”时，包含式贝叶斯因子 BF_{10} 定义为备择假设模型集合 (H_1) 与零假设模型集合 (H_0) 的后验几率与先验几率之比：

$$BF_{10} = \frac{P(H_1^f | data) + P(H_1^r | data)}{P(H_0^f | data) + P(H_0^r | data)} \bigg/ \frac{P(H_1^f) + P(H_1^r)}{P(H_0^f) + P(H_0^r)}$$

该指标的解读与传统用于假设检验的贝叶斯因子相似，其数值大小直接反映证据的连续强度(胡传鹏 等, 2018; 王允宏 等, 2023; Schad et al., 2023)。 $BF_{10} > 1$ 表示数据相对更支持备择假设 H_1 ，而 $BF_{10} < 1$ 则倾向于支持零假设 H_0 。具体而言， $1 < BF_{10} < 3$ 视为对 H_1 的弱证据， $3 < BF_{10} < 10$ 视为中等证据， $BF_{10} > 10$ 视为强证据(其倒数对应于对 H_0 的同等支持强度)。 $BF_{10} \approx 1$ 表示数据提供的信息强度微弱，不具有诊断性。

当检验“是否存在异质性”时，包含贝叶斯因子 BF_{rf} 评估所有假定存在异性(随机效应模型)与所有假定不存在异质性(固定效应模型)的相对支持程度：

$$BF_{rf} = \frac{P(H_0^r | data) + P(H_1^r | data)}{P(H_0^f | data) + P(H_1^f | data)} \bigg/ \frac{P(H_0^r) + P(H_1^r)}{P(H_0^f) + P(H_1^f)}$$

最后，基于模型后验概率对总体效应量进行加权估计。研究者可根据推断目标选择不同的估计策略：若研究逻辑遵循“先检验后估计”，即先基于包含式贝叶斯因子确认观测数据支持效应存在，

则通常报告条件模型平均的结果——仅在备择假设模型集合(H_1)进行加权, 其效应量 μ 的后验分布表示为:

$$P(\mu | \text{data}, \mu \neq 0) = \sum_{k \in \{\mathcal{H}_1^f, \mathcal{H}_1^r\}} P(\mu | \text{data}, \mathcal{H}_k) \cdot w_k$$

其中, $P(\mu | \text{data}, \mathcal{H}_k)$ 是在特定模型下 μ 的后验分布, 权重 $w_k = \frac{P(\mathcal{H}_k | \text{data})}{P(\mathcal{H}_1^f | \text{data}) + P(\mathcal{H}_1^r | \text{data})}$ 为经过归一化的模型后验概率。由此获得的综合效应量分布可提取中位数及 95%最高后验密度区间(HDI), 作为在“效应存在”假设下的稳健点估计与区间估计(van de Schoot et al., 2021)。然而, 若需全面反映模型空间所包含的所有不确定性, 则采用全模型平均——不再预设效应是否存在, 而是加权混合所有备择模型与零模型的后验分布。全模型平均更能真实反映参数推断中的模型不确定性, 避免在证据不足时过度估计效应量(Bartoš, Maier, Stanley, et al., 2025)。

与其他贝叶斯分析相似(胡传鹏 等, 2018), 敏感性分析是评估结果稳健性的必要环节。研究者需要系统改变关键参数的先验分布并重复分析, 以检查主要结论对先验选择的依赖程度。此外, 贝叶斯框架支持序贯更新——当新观测数据出现后, 可将当前的后验分布作为新的先验进行迭代推断, 动态展示证据积累过程。由于其推断逻辑不依赖于重复抽样的频率假设, 因而能避免传统频率主义框架下多重检验带来的 α 错误。

BMA 的输出使研究者能够以概率化、多层次的方式呈现结果: 既可通过贝叶斯因子提供关于效应存在与否、异质性有无的证据强度, 又可依据模型平均的后验分布给出兼顾模型不确定性的总体效应估计值, 帮助研究者完成更为稳健的推断。

3.2 校正发表偏倚的多个潜在模型与 RoBMA 解决方案

发表偏倚校正模型的选择是元分析中另一个难题。目前常用的校正方法主要包括两类: 其一为 PET-PEESE 回归校正, 主要用于处理小样本效应; 其二为发表筛选模型, 主要用于校正因统计显著性导致研究被选择性发表的筛选机制。

PET-PEESE 利用元回归校正小样本偏倚, 其核心预设是: 如果不存在发表偏倚, 效应量与标准误(由样本量决定)应相互独立; 反之, 二者会呈现相关性。该方法遵循两步式条件决策规则: 第一步是 PET 模型(精度效应检验), 构建效应量对标准误的线性回归。若回归系数 β 在 $\alpha = 0.10$ 水平下显著, 则判断存在小样本偏倚。第二步进行决策: 若 PET 检验不显著, 直接以 PET 模型的截距 μ 作为校正后的效应量估计; 若显著, 则转用 PEESE 模型(基于标准误的精度效应估计)。PEESE 模型通过建立效应量对方差的回归, 使用截距 μ 作为排除偏倚影响后的真实效应(Stanley & Doucouliagos, 2014)。

发表筛选模型则直接对论文评审与发表过程中的筛选机制进行建模，其模型包括理想状态下的常规元分析统计模型和受发表过程影响的筛选过程模型(Citkowicz & Vevea, 2017)。常规元分析模型假设在没有选择偏倚的理想情况下，观测值 y_i 分布由总体平均效应 μ 、研究间异质性 τ^2 和抽样误差 σ_i^2 共同定义的概率密度函数 $f(y_i | \mu, \tau^2; \sigma_i^2)$ 所描述。筛选过程模型则是在常规元分析统计模型中引入一个权重函数 $w(p_i)$ ，来表征不同显著性水平的研究被观测到的相对概率。模型通过 $w(p_i)$ 修正标准元分析的概率密度函数，构建加权概率密度函数：

$$f_{weighted}(y_i) = \frac{w(p_i) \cdot f(y_i | \mu, \tau^2; \sigma_i^2)}{\int_{-\infty}^{\infty} w(p_i) \cdot f(y_i | \mu, \tau^2; \sigma_i^2) dy_i}$$

其中分母项为归一化常数，确保加权后的全概率总和为 1。当 $w(p_i) = 1$ (即所有研究无论 p 值大小，被发表的概率均等)，该加权模型退化为标准元分析模型。估计阶段采用最大似然法，基于观测数据估计上述加权概率密度函数的未知参数，使模型与观测数据达到最优拟合。

筛选模型的核心在于构建一个能够表征统计显著性(或效应量大小)与研究发表概率之间关系的权重函数。不同的筛选模型在权重的定义方式上存在差异，其中较为常用的是基于 p 值的阶梯权重函数模型(Step-weight-function selection models)(Hedges, 1992)。该方法将权重函数 $w(p_i)$ 定义为分段常数函数，根据预先设定的显著性阈值(如 0.001, 0.01, 0.05 等)将纳入元分析的单个研究的 p 值划分为 k 个区间，并设定各区间的相对发表权重 w_k 。通常将显著性最高的区间权重固定为 $w_i = 1$ 作为参照，其余区间的权重作为待估参数，通过上述加权的元分析模型进行估计。该模型将同时对真实效应量、异质性及各区间权重参数进行联合迭代估计(Vevea & Hedges, 1995)。值得注意的是，该模型中未知参数较多，如果纳入的文献数量较少，可能导致某个区间(如 $p > 0.05$)缺乏观测数据，此时传统最大似然估计将因为信息不足而无法收敛，研究者可能需要采用合并相邻区间(例如合并 $0.05 < p < 0.10$ 与 $0.10 < p < 0.50$)或改用基于先验分布的贝叶斯方法才能获得各未知参数的估计。

PET-PEESE 回归校正与发表筛选模型方法存在共同的局限性：不同方法得出的结论可能不一致，目前缺乏广泛认可的标准来选择校正方法；每种方法均依赖于特定模型预设，模型预设不成立时，校正可能失效甚至引入新的偏差(Carter et al., 2019; Simonsohn et al., 2014; Sterne et al., 2011)。这种模型选择的困境在 BMA 框架下不复存在，研究者只需要在元分析的模型空间中引入第三个维度——偏倚校正。此时，贝叶斯元分析的模型空间涵盖三个维度：总体效应量是否为零、纳入文献是否存在异质性、以及是否进行发表偏倚校正。纳入发表偏倚考量的 BMA 即构成 RoBMA，其实施步骤如下：

首先，构建模型空间。RoBMA 系统组合三个核心维度的预设：

总体效应维度：存在 $\mu = 0$ vs 不存在 $\mu \neq 0$ ；

异质性维度: 存在 $\tau^2 > 0$ vs 不存在 $\tau^2 = 0$;

发表偏倚校正维度: 无校正 \bar{p}^b vs 校正 p^b 。

在“发表偏倚校正”维度上, RoBMA 纳入 9 种偏倚矫正方案: 包括“无校正”条件、PET-PEESE 矫正条件(2 种: PET 和 PEESE)、基于 p 值权重的发表筛选模型(6 种: SelMod1-6, 见表 1)。这些方案将分别与 BMA 的 4 个模型进行组合, 最终形成由 36 个独立模型构成的模型空间(全部模型列表见附录 2):

模型 M1-M4: 对应 BMA 的 4 个基础模型, 为无校正的基准模型;

模型 M5-M12: 在 M1-4 上分别进行 PET 或 PEESE 校正, 共 8 个模型;

模型 M13-M36: 在 M1-4 上分别进行 6 种发表筛选模型 (SelMod1-6) 校正, 共 24 个模型。

表 1 RoBMA 中发表筛选模型的设定

模型	检验方向	截断点	p 值区间划分	对应权重参数	假设与含义
SelMod1	双侧	0.05	$[0, 0.05] > [0.05, 1]$	$\omega_1 > \omega_2$	无论效应方向, 显著结果的发表概率高于非显著结果。
SelMod2	双侧	0.05, 0.10	$[0, 0.05] > [0.05, 0.10] > [0.10, 1]$	$\omega_1 > \omega_2 > \omega_3$	无论效应方向, 发表概率随显著性水平递减: 显著 > 边缘显著 > 非显著。
SelMod3	单侧	0.05	$[0, 0.05] > (0.05, 1]$	$\omega_1 > \omega_2$	仅预期方向上的显著结果发表概率更高; 非显著与反向结果的发表概率同等降低。
SelMod4	单侧	0.025, 0.05	$[0, 0.025] > (0.025, 0.05] > (0.05, 1]$	$\omega_1 > \omega_2 > \omega_3$	仅预期方向上, 发表概率按显著性层级递减: 极显著 > 一般显著 > 非显著。
SelMod5	单侧	0.05, 0.50	$[0, 0.05] > (0.05, 0.50] > (0.50, 1]$	$\omega_1 > \omega_2 > \omega_3$	仅预期方向上, 显著结果 > 方向正确但不显著的结果 > 反向结果。
SelMod6	单侧	0.025, 0.05, 0.50	$[0, 0.025] > (0.025, 0.05] > (0.05, 0.50] > (0.50, 1]$	$\omega_1 > \omega_2 > \omega_3 > \omega_4$	仅预期方向上, 发表概率按极显著 > 一般显著 > 方向正确但不显著 > 反向结果逐级递减。

注: 1.检验方向: “双边”指发表概率仅取决于双侧显著性水平, 与效应方向无关; “单边”指发表概率仅对预期方向(如正向)的结果敏感, 反向结果与非显著结果共享最低发表概率权重。2.截断点对应: 单侧 0.025 对应双侧 $p < 0.05$ (常规显著); 单侧 0.05 对应双侧 $0.05 \leq p < 0.10$ (边缘显著); 单边 0.50 对应效应方向与假设一致但不显著的情形。双侧模型直接以双侧 p 值划分区间。3.权重参数(ω): 代表各区间相对最显著区间的发表概率, 最显著区间通常固定为 $\omega_i=1$ 。模型假

设 p 值越小越容易发表($\omega_1 > \omega_2 > \dots > \omega_k$)。4. 偏倚程度: ω 随区间衰减得越快, 说明发表偏倚越严重; 若各区间 $\omega \approx 1$, 则表明不存在明显的发表偏倚。

其次, 设置模型与参数的先验分布。在模型设定阶段, 需为各模型分配先验模型概率。RoBMA 默认采用无信息先验策略(Maier et al., 2023): 在三个维度中相互对立的模型组赋予相等的先验概率(各为 0.5), 体现了模型选择前的“公平竞争”原则。具体而言: 总体效应是否存在的两类模型先验概率均为 0.5; 异质性是否存在的两类模型先验概率均为 0.5; 是否进行发表偏倚校正的两组模型先验概率也均为 0.5。

对于“进行发表偏倚校正”这一组模型(总概率 0.5), 其概率进一步在两类校正方法之间平均分配: 一类是基于 p 值的发表筛选模型(含 6 种权重函数), 另一类是 PET-PEESE 回归模型(含 PET 和 PEESE 两种)。每类校正方法获得 0.25 的先验概率, 这 0.25 的概率在其各自类别内的具体模型间进行了均匀分配。经过组合后, M1-M4 的先验概率为 $0.5 * 0.5 * (0.5) = 0.125$, M5-M12 的先验概率为 $0.5 * 0.5 * (0.25/2) \approx 0.031$, M13-M36 的先验概率为 $0.5 * 0.5 * (0.25/6) \approx 0.010$ 。若研究者基于理论或前期证据对特定假设有较强信心, 也可调整相应模型的先验权重。

在参数层面, 各待估参数的先验设定如下(Bartoš et al., 2022): PET-PEESE 回归系数 β 均采用截断于正数范围的半柯西分布——PET 系数设定为 $\text{Cauchy}(0, 1)$, PEESE 系数设定为 $\text{Cauchy}(0, 5)$ 。该先验约束 β 为正值, 以体现“小样本效应”的正向相关性(即标准误越大, 观测到的效应量倾向于越高), 同时借助柯西分布的重尾特性保留弱信息先验的特征。对于发表筛选模型, 其核心参数——不同 p 值区间的相对发表权重 ω ——采用单位累积狄利克雷先验(Cumulative Dirichlet prior)。该先验对不同 p 值区间的发表概率施加了单调性约束(如统计显著研究的发表概率不低于不显著研究), 其参数维度会随 p 值截断点的数量而调整: 当模型设定 1、2、3 个截断点(对应划分 2、3、4 个 p 值区间)时, 权重先验分别设定为 $\text{CumDirichlet}(1, 1)$ 、 $\text{CumDirichlet}(1, 1, 1)$ 、 $\text{CumDirichlet}(1, 1, 1, 1)$ 。参数 μ 与 τ 的先验设置与前述 BMA 的默认设定一致。

接着, 执行模型更新与参数推断。在设定先验后, RoBMA 对模型空间内的每个候选模型分别执行 MCMC 采样, 同步实现两个维度的推断: 在参数层面, 采样直接获取各模型下参数的后验分布; 在模型层面, 采样得到的后验样本被用于估计各模型的边际似然。由于高维参数空间的积分求解较为困难, RoBMA 基于 MCMC 后验样本采用桥采样(bridge sampling)等方法对边际似然进行数值近似(Bartoš et al., 2026; Gronau, Sarafoglou, et al., 2017; Gronau et al., 2020)。各模型的边际似然结合先验概率, 依据贝叶斯定理, 计算出后验模型概率作为后续模型平均的核心权重。

随后, 计算包含式贝叶斯因子。基于各模型的模型后验概率与参数的后验分布, RoBMA 可计算三类包含式贝叶斯因子, 分别用于检验以下三个假设: (1)总体效应是否存在(BF_{10}); (2)研究间异质

性是否存在(BF_{rj}); (3)发表偏倚是否存在(BF_{pb})。三类包含式贝叶斯因子的计算逻辑与 BMA 相同, 其核心区别在于进行比较的模型集合划分不同, 具体的模型集合构成及完整的公式推导详见附录 3。

以量化发表偏倚证据强度的包含式贝叶斯因子(BF_{pb}) 为例, 其计算公式为:

$$BF_{pb} = \frac{\sum_{m \in \mathcal{M}^{bias}} p(\mathcal{H}_m | data)}{\sum_{n \in \mathcal{M}^{no-bias}} p(\mathcal{H}_n | data)} \bigg/ \frac{\sum_{m \in \mathcal{M}^{bias}} p(\mathcal{H}_m)}{\sum_{n \in \mathcal{M}^{no-bias}} p(\mathcal{H}_n)}$$

\mathcal{M}^{bias} 代表所有包含发表偏倚校正的模型(即在基础模型上扩展了发表偏倚参数的模型),

$\mathcal{M}^{no-bias}$ 代表未包含校正的基础模型。根据本节的模型构建, 在 36 个模型中, $\mathcal{M}^{no-bias}$ 包含

M1-M4, 共 4 个模型(即 BMA 中 4 个无发表偏倚矫正的基础模型); \mathcal{M}^{bias} 包含 M5-M36, 共 32 个模型(包括经 PET-PEESE 方法矫正的 8 个模型, 以及经发表筛选模型矫正的 24 个模型, 均基于 M1-M4 构建)。在默认先验设定中, “存在发表偏倚”与“不存在发表偏倚”被赋予相等的先验概率(各 0.5), 即

$\sum_{m \in \mathcal{M}^{bias}} p(\mathcal{H}_m) = \sum_{n \in \mathcal{M}^{no-bias}} p(\mathcal{H}_n) = 0.5$, 即先验概率比为 1。此时 BF_{pb} 的值直接等于后验概率比, 反映了数据实际支持“存在发表偏倚”的程度。

最后, 进行效应量的加权估计。得到各模型的模型后验概率及参数后验分布后, RoBMA 默认采用无条件估计逻辑, 即通过贝叶斯模型平均对所有 36 个模型中总体效应量 μ 的后验分布进行整合, 以获得稳健的估计。具体而言, 模型平均后的综合后验分布定义为各模型参数后验分布按其模型后验概率的加权平均:

$$p(\mu | data) = \sum_{k=1}^{36} p(\mu | \mathcal{H}_k, data) \times p(\mathcal{H}_k | data)$$

该综合分布即为元分析综合效应量的稳健估计, 可以基于该分布计算效应量 μ 的点估计(如中位数或后验均值)及其 95%最高密度可信区间(HDI)。

与 BMA 类似, RoBMA 中同样推荐研究者进行敏感性分析, 对关键的模型参数使用不同的参数先验分布重复上述分析流程, 并报告关键结果(贝叶斯因子与效应量估计)是否保持一致。

综上所述, RoBMA 通过贝叶斯模型平均框架, 将总体效应是否存在、文献间异质性、文献发表偏倚三个维度的不确定性均纳入统计模型, 为元分析提供了一种更为稳健、透明的推断工具。下文将以一项真实的元分析公开数据为例, 展示如何在 JASP 软件中使用 RoBMA 进行元分析, 并与传统频率主义元分析方法的结果进行对比。

4 JASP 进行贝叶斯模型平均元分析

为展示 RoBMA 在 JASP 中的实际操作，本节将基于一项已发表元分析(Liu et al., 2024)公开数据进行分析。该元分析旨在检验认知心理学中的“自我参照效应”(孙淑婷 等, 2023; Liu et al., 2025; Sui & Humphreys, 2015)能否推广至教学领域，即当学习信息与自我建立联系后，学习效果是否更优。研究共纳入 13 篇原始文献中的 20 个独立研究($N = 1082$)，采用传统频率学派元分析方法中的随机效应模型进行效应量综合，结合 Egger 检验(Egger et al., 1997; Vevea & Woods, 2005)和等高线增强漏斗图(Peters et al., 2008)对发表偏倚进行评估，并通过剪补法(Duval & Tweedie, 2000)进行校正。未经校正的元分析结果显示，自我参照条件下学习效果更优(Hedges' $g = 0.40$, 95% CI [0.18, 0.62])。异质性检验(I^2)表明纳入研究存在高度异质性。剪补法校正后，效应量下降至 $g = 0.23$, 95% CI [-0.02, 0.48]，提示原始结果可能存在发表偏倚。

通过 JASP 进行基于模型平均的贝叶斯元分析操作步骤如下(对应的 R 代码参见附录 4):

首先，下载该元分析的数据(文件名: Liu_2024_meta_data.csv, 下载地址: osf.io/g23kh/), 在 JASP(版本 0.95.4)中打开该 csv 文件。随后，点击 JASP 右上角蓝色加号，选择“Meta-Analysis”，将该模块添加到最上方功能区(图 3)。添加后点击该按钮，在“Meta-Analysis”的“Bayesian”子菜单中选择“Robust Bayesian Meta-Analysis”，即可打开下图灰色区域中的图形交互界面。

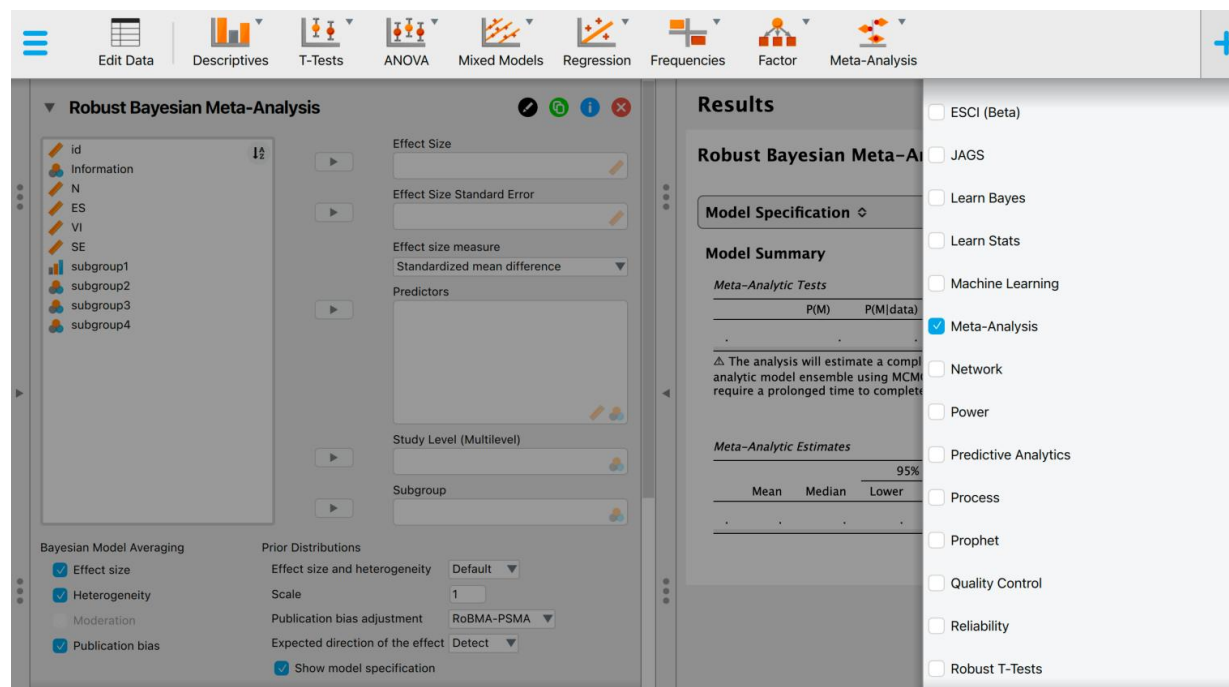


图 3 JASP 添加“元分析”模块并进入 RoBMA 分析界面

在界面左侧的变量选择框中，将效应量数据(ES)拖入 "Effect Size" 框中，将对应的标准误数据 (SE)拖入 "Effect Size Standard Error" 框中。在“Prior Distributions”选项下的“Effect Size and heterogeneity”下拉菜单可选择“Default”使用默认分布或选择“Custom”自定义先验分布。如图 4 左侧面板所示，RoBMA 允许用户根据理论预期灵活设定先验分布(图中参数为默认分布参数，未作修改)。勾选“Prior Distributions”下的 "Model Specification", 界面右侧面板将展示 RoBMA 如何通过组合不同的先验来构建备选模型空间。

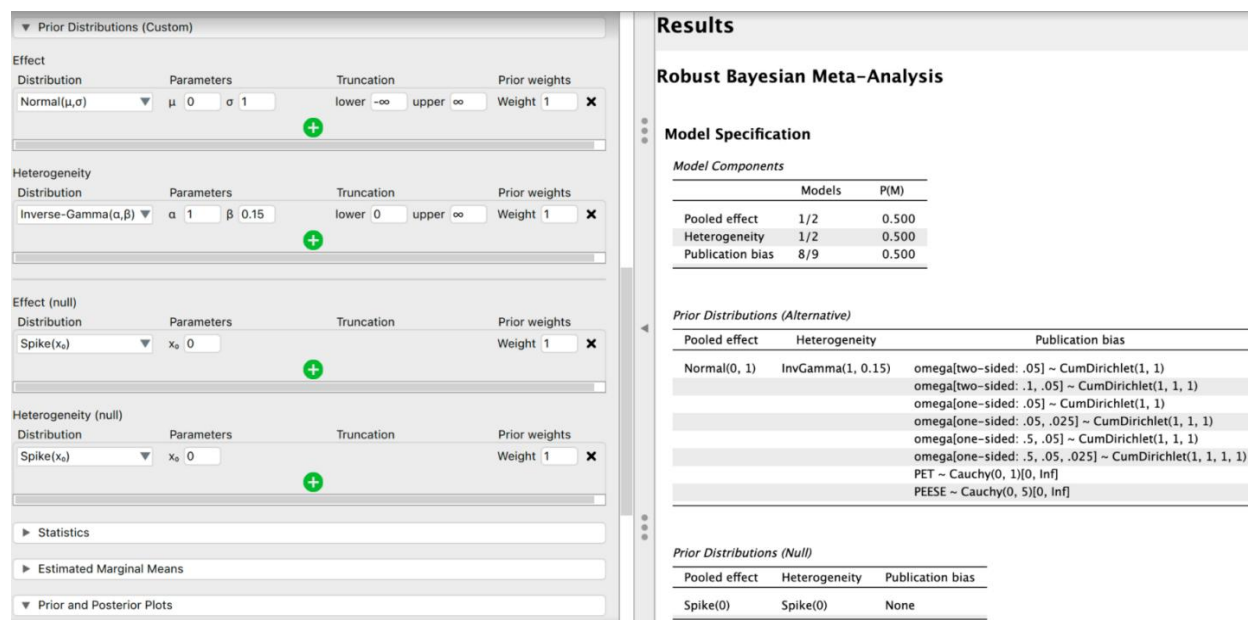


图 4 RoBMA 模块的先验设定界面

完成上述设置后，JASP (0.95.4) 将自动运行 MCMC 采样。在本示例的数据规模及硬件环境(8核 CPU、8GB 内存)下，该过程运行时间约为 2 秒，峰值内存占用约为 430 MB。如图 5 中左侧面板所示，在“Advanced”选项中，用户可以自行设置 Adaptation(适配期)、Burnin(也称 Warmup，即在正式分析中不考虑的 MCMC 样本)及 Samples(采样数)等参数，建议设置足够的采样数(如图中所示的 10000 次以上)以确保模型收敛。运行期间，界面右侧的“Diagnostics”将实时输出诊断汇总表(包含 R-hat、ESS 等收敛指标)以及各项参数的轨迹图(Trace Plot)，用于监测采样的平稳性。若需要进一步的可视化，用户可以继续点击带有 "Plot" 的选项卡，生成先验/后验分布图、森林图等。

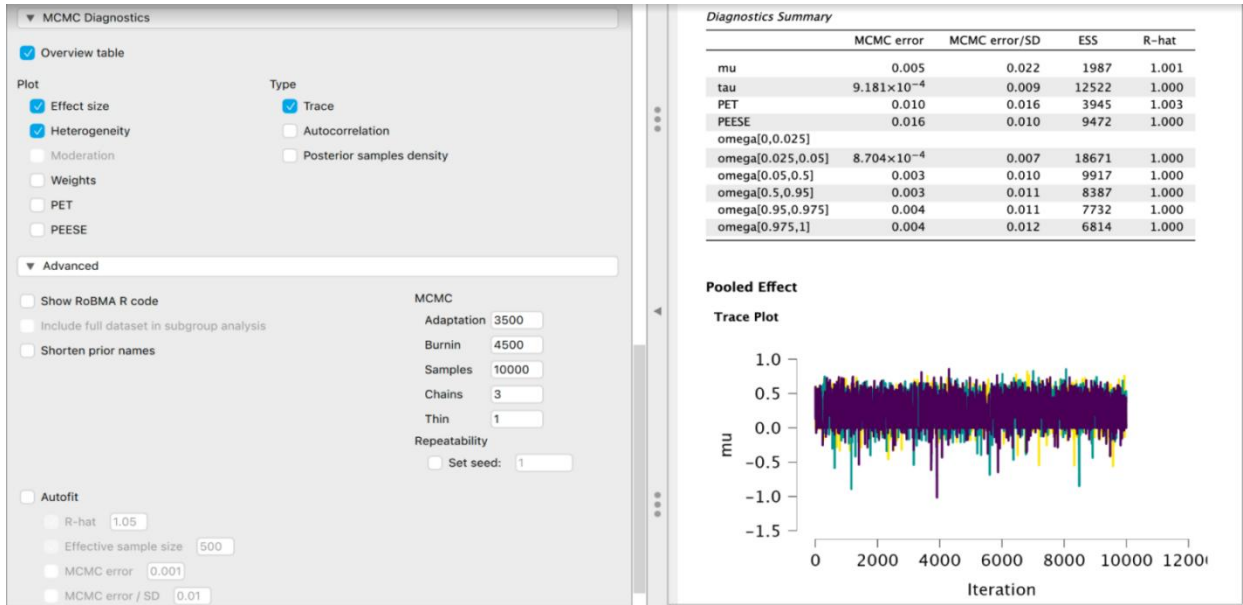


图5 RoBMA 模块的MCMC 诊断与高级参数设置界面

本研究通过 R-hat、有效样本量(ESS)及轨迹图对本例的 RoBMA 模型收敛性进行了检验。图5的结果显示，所有参数的 R-hat 均小于 1.01(低于 1.05 的保守阈值)，且有效样本量均远大于 1000；同时，效应量的轨迹图显示多条采样链混合良好，呈现平稳的随机波动。以上诊断结果一致表明MCMC 算法已实现理想收敛，后验估算结果稳健可靠。

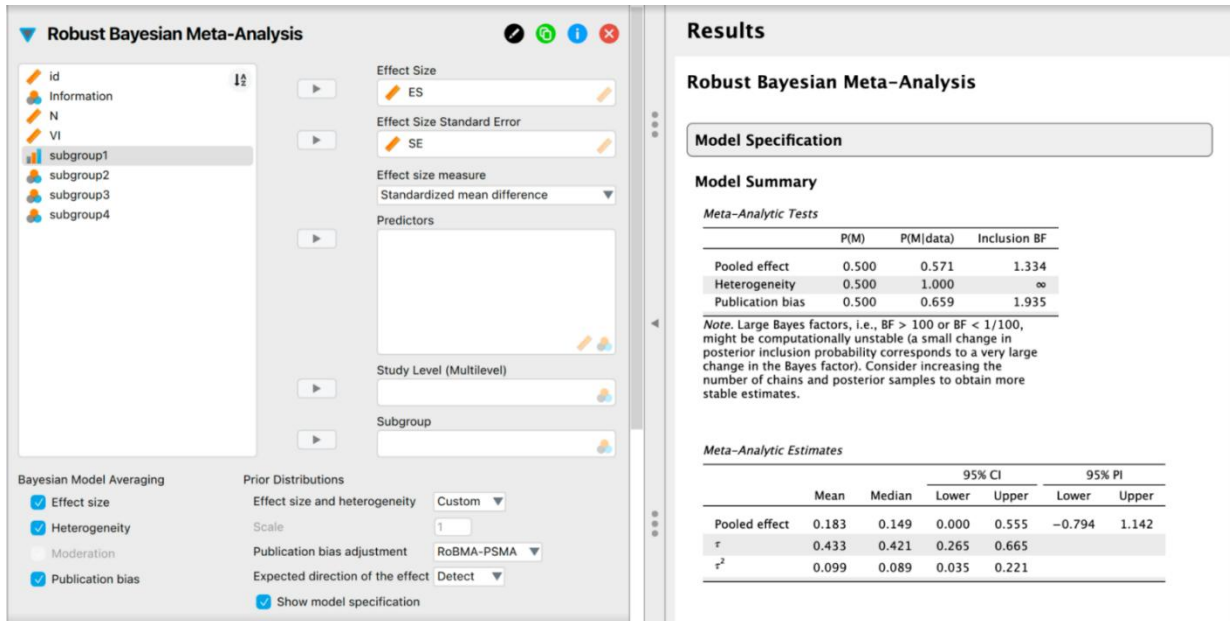


图6 RoBMA 模块的输出结果界面

RoBMA 的结果如图 6 所示，在证据强度的推断上，“Model Summary”中的“元分析检验(Meta-Analytical Tests)”模块显示综合效应量(Pooled effect)的纳入贝叶斯因子 BF_{10} 为 1.334，对应的后验概率 $p(M|data)$ 为 0.571，表明数据仅能提供弱证据支持“存在效应”的模型。与之对应，在“元分析参数估计(Meta-Analytic Estimate)”表中，可以看到考虑了所有潜在模型(包括校正发表偏倚的模型)后的综合效应量(pooled effect)为 中位数 = 0.149, 95% 可信区间(Credible Intervals, CrI)为[0, 0.555]。值得注意的是，“元分析检验(Meta-Analytical Tests)”表中的异质性检验(Heterogeneity)提供了极强的证据支持($BF_{H^2} = \infty$)，“元分析参数估计(Meta-Analytic Estimate)”表中异质性估计的中位数为 $\tau = 0.421$ (95% CrI [0.265, 0.665])，表明研究间的效应量存在极高的离散性。这意味着自我参照编码的干预效果高度依赖于具体的学科领域、教学情境及学生群体，单一的综合效应量难以全面概括其在多样化教育场景中的实际表现。此外，元分析检验表中发表偏倚(Publication bias)的证据强度则相对较弱($BF_{pbpb} = 1.935$)，表明虽然存在“显著结果更易被发表”的潜在倾向，但该偏倚对结果的干扰尚未达到较高水平。

上述结果与原研究基本一致，但更趋于保守。RoBMA 估算的效应量(0.149)低于原研究校正后的结果(0.23)，且其可信区间([0.00, 0.555])相较于原研究的置信区间([-0.02, 0.48])更为宽阔。这表明在纳入了模型选择、异质性及发表偏倚的多重不确定性后，RoBMA 给出更谨慎的效应估计。在异质性方面，RoBMA 的极强证据支持与原研究中 $I^2 = 80.32\%$ 的高异质性结论高度契合。此外，RoBMA 对发表偏倚的弱证据支持($BF_{pbpb} = 1.935$)也与原研究中 Egger 检验不显著($p = 0.375$)的发现互为印证。

以上示例表明，基于贝叶斯模型平均的元分析方法所得到的结果与传统元分析中较为谨慎的做法(异质性分析与敏感性分析)基本一致，且在统计推断上提供出更为保守的参数估计。同时，基于贝叶斯的模型平均方法将多个研究假设整合在统一的框架之下，在一次分析中即可提供多维度的结果，为研究者提供了更为全面且简洁的统计证据。

5 结论与展望

作为心理学及相关领域的重要方法，高质量的元分析能够为研究者及从业者提供参考。然而，传统元分析中的模型选择常让研究者面临抉择困难，且目前缺乏统一的标准(Borenstein et al., 2010; Rice et al., 2018)。基于贝叶斯模型平均的元分析方法将“选择”变为“整合”：不再预设单一模型，而是根据观测数据的表现，为模型空间内所有合理的假设分配权重。这种“加权平均”的处理方式能够自动纳入模型选择带来的不确定性，从而产出更为稳健、客观的推断结论。

相比于 $p < 0.05$ 这种“显著或不显著”的二元论推断法，基于贝叶斯模型平均的元分析以贝叶斯因子度量证据强度，能清晰回应元分析关注的核心问题——效应是否存在。更重要的是，该框架可以同时回答三个关键问题：效应是否存在、纳入的研究是否存在异质性、以及研究领域是否存在发表偏倚，从而简化了方法上的流程。

基于贝叶斯模型平均的元分析在处理发表偏倚方面的优势尤为突出，充分体现了其统一模型的优越性。不同于传统做法中通过 Egger 检验(Egger et al., 1997; Vevea & Woods, 2005)或等高线增强漏斗图(Peters et al., 2008)等先对发表偏倚进行评估，再以剪补法等方法进行“补丁式”事后校正，基于贝叶斯模型平均的方法将发表偏倚的校正逻辑直接纳入模型构建之中，自然地将发表偏倚的影响整合进效应量的估计过程。此外，由于贝叶斯框架的固有特性，它天然支持“动态元分析”(Oravecz et al., 2017)：当新研究出现时，其证据可直接用于更新现有的后验分布；同时，它允许将领域内的历史数据或专家经验转化为先验知识，从而提升统计推断的信息效率。

但基于贝叶斯模型平均的元分析在实际应用中存在其特有的挑战与局限。首先，该方法的实现依赖于高维积分与 MCMC 等复杂算法，其计算成本高于传统方法。在本文的案例中，在个人计算机(8 核 CPU, 8GB 内存)运行 JASP 中传统元分析的“Model Summary”部分，耗时约为 1 秒；而同等硬件环境下运行 JASP 中贝叶斯模型平均元分析的“Model Summary”部分，耗时约为 14 秒。如果纳入研究的数据量更大，可以预期时间上的差异将更加明显。其次，RoBMA 输出的结果比传统的 p 值或置信区间要丰富得多，这意味着使用者需要更多的知识储备以正确解读其结果。从多个模型的后验概率分布到包含式贝叶斯因子，再到校正后的各维度参数估计，这些信息对研究者和读者都提出了更高要求。在传播科学结论时，需要清晰、准确地解读这些非二元论的证据。

此外，作为贝叶斯推断框架，基于贝叶斯模型平均的元分析的结论——尤其是 BF 的精确数值——往往对先验分布的设定较为敏感。虽然可以通过设置不同的先验量级进行敏感性分析，但如何论证先验的合理性并透明地报告这些选择，仍是研究者需要审慎考量的难点。如果先验设定不当，很可能导致先验信息“主导”了数据信息，从而掩盖了真实的科学发现。在“小样本”或“极端异质性”的情境下，数据提供的信息量非常有限，后验推断会表现出极强的不稳定性，结论往往高度依赖于先验设定。在这种情况下，基于贝叶斯模型平均的元分析也很难给出高度确定性的结论。为方便研究者解读与报告基于贝叶斯模型平均的元分析结果，我们结合贝叶斯分析报告指南(Kruschke, 2021; van Doorn et al., 2021)，建议报告如下关键信息(见表 2)。

表 2 RoBMA 结果报告与解读规范指南

维度	关键指标	解读与报告建议
分析计划	研究性质	明确标注是基于预注册计划的验证性分析，还是基于现有数据的探索性重分析。
	分析目标	阐明研究旨在通过 BF 验证效应的存在性(检验)，并通过后验分布确定校正后的效应量大小(估计)。
	效应方向设定	根据研究目标设定单侧或双侧检验，并提供理论正当性说明。

模型设置	模型假设及数量	应说明模型空间涵盖了固定/随机效应、有/无效应、以及有/无发表偏倚的各种组合，并报告模型总数。
	模型先验概率 $p(H)$	通常对每种假设为 0.5，表示在观测数据前，研究者持中立、公正的立场。
	参数先验分布	报告所选先验分布及具体参数，如 $\mu \sim \text{Normal}(0, 1)$ 或 $\tau \sim \text{Inv-Gamma}(1, 0.15)$ 。若使用非默认先验，必须提供理论或历史证据的合理解释。
	软件环境与设置	注明 JASP 或 R 软件版本，以及 R 包版本号。报告 MCMC 采样设置(如链数、正式采样数、预热期)以确保可复现性。
	发表偏倚矫正	说明采用了何种校正方案(如 RoBMA 包中的 PSMA 架构，包含发表筛选模型与 PET-PEESE 模型)。
模型表现	收敛性(\hat{R} 与轨迹图)	\hat{R} 应小于 1.05(建议 < 1.01)；同时确认轨迹图呈现健康的“多毛毛虫”状，无趋势性偏移。表明 MCMC 采样链已达到平稳分布，结果可靠。
	误差百分比 (Error %)	建议报告数值误差。低于 20%的误差百分比被认为是数值估计达到足够稳定性的标志。
	有效样本量 (ESS)	说明 ESS 是否充足，确保后验分布的估计精度。
假设检验	效应存在(BF_{10})	核心指标。衡量数据支持某种假设(如效应存在、异质性存在、偏倚存在)相对于反向假设的强度。
	异质性存在(BF_{H^2})	1-3 倍：证据极弱； 3-10 倍：证据中等； >10 倍：证据很强。
	发表偏倚存在(BF_{pb})	
参数估计	综合效应量(μ)	报告 μ 的中位数，其对偏态分布比均值更稳健。代表排除发表偏倚等后的真实平均效果。
	效应量可信区间(95% CrI)	报告 μ 的 95% CrI，表示真实效应量有 95% 的概率落在该范围内。若包含 0，意味着数据尚无法排除零效应的可能性。
	异质性(τ)	报告 τ 的中位数。描述研究间效应的波动情况。 τ 值越大说明效果越不稳定，越需要关注调节变量的影响。
	异质性可信区间(95% CrI)	报告 τ 的 95% CrI，代表真实异质性 τ 有 95% 的概率落在该区间内。较大的区间宽度反映了对研究间变异程度估计的不确定性。
稳健性与透明度	先验敏感性分析	讨论不同先验宽度对 BF 值的显著影响，以评估结论是否稳健。
	开放科学实践	鼓励公开托管 .jasp 和 .R 文件及原始数据，附带 README 告知软件和依赖包版本，以便进行复现和替代分析。

基于贝叶斯模型平均的元分析为元分析中的模型不确定性提供了系统的解决方案，但要充分发挥其潜力，需要方法与操作规范的进一步优化。在方法学层面，可以与开放科学实践结合，通过预注册和代码共享提升结果的可重复性。同时，应根据实证基础选择先验，减少主观设定对推断的过度影响。对研究者而言，关键在于合理使用与严谨报告。建议在模型不确定性突出、需量化证据强度或校正发表偏倚的场景中优先选用该方法。应用时可以开展敏感性分析，检验结论对不同先验的稳健性，并遵循透明报告准则，完整说明模型设定、先验依据、核心结果(如贝叶斯因子、后验估计)及分析代码。

目前，基于贝叶斯模型平均的元分析仍在扩展。例如，稳健贝叶斯多层元分析能够处理研究内效应量的依赖性(Bartoš et al., 2026)，而基于贝叶斯模型平均的元回归则为评估不同类型的调节变量提供了更为稳健的工具(Bartoš, Maier, Stanley, et al., 2025)。基于贝叶斯的模型平均框架对新统计模型具有开放性与包容性，未来研究可以将更多经典的偏倚诊断逻辑(如 Egger 检验的贝叶斯变体或非线性偏倚模型)进行统计模型化，并将其作为候选路径吸纳至贝叶斯模型平均的模型空间中，从而实现复杂偏倚机制的自动加权与识别。总而言之，基于贝叶斯模型平均的元分析为元分析提供一种更细致处理不确定性的工具，而非替代传统方法，其合理使用有助于推动更稳健、更可信的科学证据。

参考文献

- 胡传鹏, 孔祥祯, Wagenmakers E.-J., Ly A., 彭凯平. (2018). 贝叶斯因子及其在 JASP 中的实现. *心理科学进展*, 26(6), 951.
- 胡传鹏, 王继贤. (2025). 贝叶斯统计及其在心理学中的应用. Zenodo. doi: <https://doi.org/10.5281/zenodo.17623003>
- 刘宇, 陈树铨, 樊富珉, 邸新, 范会勇, 封春亮, ... 胡传鹏. (2021). 开放式荟萃分析的规范化报告. *中国科学: 生命科学*, 51(6), 764 – 778.
- 潘晚珂, 温秀娟, 金海洋. (2023). 贝叶斯混合效应模型: 基于 brms 的应用教程. *心理技术与应用*, 11(10), 577 – 598.
- 任志洪, 赖丽足. (2023). *心理学元分析方法: 理论、案例与操作*. 中国人民大学出版社.
- 孙淑婷, 王楠, 温佳慧, 胡传鹏. (2023). 自我参照的神经成像认知本体论数据集. *中国科学数据 (中英文网络版)*, 8(3).
- 王允宏, Bergh D. V. D., Aust F., Ly A., Wagenmakers E.-J., 胡传鹏. (2023). 贝叶斯方差分析在 JASP 中的实现. *心理技术与应用*, 11(9), 528 – 541.
- 吴凡, 顾全, 施壮华, 高在峰, 沈模卫. (2018). 跳出传统假设检验方法的陷阱——贝叶斯因子在心理学研究领域的应用. *应用心理学*, 24(3), 195 – 202.
- Almalik, O. (2025). Adjusting for Publication Bias in Meta-Analysis with Continuous Outcomes: A Comparative Study. *Mathematics*, 13(21), 3487.
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3 – 25.
- Bartoš, F., Maier, M., Quintana, D. S., & Wagenmakers, E.-J. (2022). Adjusting for publication bias in JASP and R: Selection models, PET-PEESE, and robust bayesian meta-analysis. *Advances in Methods and Practices in Psychological Science*, 5(3), 25152459221109259.
- Bartoš, F., Maier, M., Stanley, T. D., & Wagenmakers, E.-J. (2025). Robust Bayesian meta-regression: Model-averaged moderation analysis in the presence of publication bias. *Psychological Methods*. Advance online publication.
- Bartoš, F., Maier, M., & Wagenmakers, E.-J. (2026). Robust Bayesian multilevel meta-analysis: Adjusting for publication bias in the presence of dependent effect sizes. *Behavior Research Methods*, 58(6), 165.
- Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, 14(1), 99 – 116.
- Bartoš, F., Maier, M., Wagenmakers, E.-J., Goosen, J., Denwood, M., & Plummer, M. (2025). RoBMA: Robust Bayesian Meta-Analyses (Version 3.6.1) [R package].
- Berkhout, S. W., Haaf, J. M., Gronau, Q. F., Heck, D. W., & Wagenmakers, E.-J. (2024). A tutorial on bayesian model-averaged meta-analysis in JASP. *Behavior Research Methods*, 56(3), 1260 – 1282.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97 – 111.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115 – 144.
- Citkowitz, M., & Vevea, J. L. (2017). A parsimonious weight function for modeling publication bias. *Psychological Methods*, 22(1), 28 – 41.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455 – 463.

- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629 – 634.
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120 – 128.
- Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1), 1 – 28.
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3 – 8.
- Grant, R., & Di Tanna, G. L. (2025). *Bayesian meta-analysis: A practical introduction* (1st ed.). Chapman and Hall/CRC.
- Gronau, Q. F., Heck, D. W., Berkhout, S. W., Haaf, J. M., & Wagenmakers, E.-J. (2021). A primer on bayesian model-averaged meta-analysis. *Advances in Methods and Practices in Psychological Science*, 4(3), 25152459211031256.
- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ... Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81, 80 – 97.
- Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R Package for Estimating Normalizing Constants. *Journal of Statistical Software*, 92, 1 – 29.
- Gronau, Q. F., Van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., & Wagenmakers, E.-J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: the case of felt power. *Comprehensive Results in Social Psychology*, 2(1), 123 – 138.
- Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman and Hall/CRC.
- Hedges, L. V. (1992). Modeling Publication Selection Effects in Meta-Analysis. *Statistical Science*, 7(2), 246 – 255.
- Hennig, C. (2024). Probability Models in Statistical Data Analysis: Uses, Interpretations, Frequentism-as-Model. In *Handbook of the History and Philosophy of Mathematical Practice* (pp. 1411 – 1458). Cham: Springer International Publishing.
- Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science*, 3(2), 200 – 215.
- Ioannidis, J. P. A. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice*, 14(5), 951 – 957.
- Iyengar, S., & Greenhouse, J. B. (1988). Selection Models and the File Drawer Problem. *Statistical Science*, 3(1), 109 – 117.
- Kruschke, J. K. (2021). Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, 5(10), 1282 – 1291.
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178 – 206.
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55(1), 1 – 7.
- Lehrer, E. J., Wang, M., Sun, Y., & Zaorsky, N. G. (2023). An Introduction to Meta-Analysis. *International Journal of Radiation Oncology, Biology, Physics*, 115(3), 564 – 571.
- Liu, Z., Hu, M., Zheng, Y., Sui, J., & Chuan-Peng, H. (2025). A multiverse assessment of the reliability of the self-matching task as a measurement of the self-prioritization effect. *Behavior Research Methods*, 57(1), 37.
- Liu, Z., Wen, J., Liu, Y., & Hu, C.-P. (2024). The effectiveness of self: A meta-analysis of using self-referential encoding techniques in education. *British Journal of Educational Psychology*, 94(1), 112 – 137.

- Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, 28(1), 107 – 122.
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. *Perspectives on Psychological Science*, 11(5), 730 – 749.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, 8(5), 336 – 341.
- Oravecz, Z., Huentelman, M., & Vandekerckhove, J. (2017). Sequential Bayesian updating for big data. In *Big data in cognitive science* (pp. 13 – 33). Routledge/Taylor & Francis Group.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991 – 996.
- Rice, K., Higgins, J. P. T., & Lumley, T. (2018). A Re-Evaluation of Fixed Effect(s) Meta-Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(1), 205 – 227.
- Schad, D. J., Nicenboim, B., Bürkner, P.-C., Michael Betancourt, & Vasishth, S. (2023). Workflow techniques for the robust use of bayes factors. *Psychological Methods*, 28(6), 1404 – 1426.
- Sedgwick, P. (2015). Meta-analyses: what is heterogeneity? *BMJ*, 350, h1435.
- Si, S., Gu, J., & Tian, M. (2025). Leveraging single-case results to Bayesian hierarchical modelling. *Computational Statistics*, 40(2), 795 – 819.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science*, 9(6), 666 – 681.
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60 – 78.
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., ... Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343.
- Sui, J., & Humphreys, G. W. (2015). The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends in Cognitive Sciences*, 19(12), 719 – 728.
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., ... Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2), 576 – 594.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: a multilevel approach. *Behavior Research Methods*, 47(4), 1274 – 1294.
- van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., ... Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813 – 826.
- Veenman, M., Stefan, A. M., & Haaf, J. M. (2024). Bayesian hierarchical modeling: an introduction and reassessment. *Behavior Research Methods*, 56(5), 4600 – 4631.
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419 – 435.
- Vevea, J. L., & Woods, C. M. (2005). Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions. *Psychological Methods*, 10(4), 428 – 443.

Meta-Analysis Based on Bayesian Model Averaging: Principles and Implementation

Ren Ziwei¹, Liu Zheng², Hu Chuanpeng¹

⁽¹⁾School of Psychology, Nanjing Normal University, Nanjing, 210097)

⁽²⁾School of Humanities and Social Sciences, The Chinese University of Hong Kong(Shenzhen), Shenzhen, 518172)

Abstract Meta-analysis serves as an essential statistical methodology to synthesize independent empirical findings. It is deeply embedded across various quantitative disciplines. However, researchers routinely confront significant dilemmas regarding model selection throughout the analytical workflow. When managing across-study heterogeneity, researchers face a strict choice between fixed-effects and random-effects frameworks. Similarly, to mitigate the threat of potential publication bias, researchers choose from a diverse array of disparate correction models. At present, the field lacks a unified, standardized criterion for selecting the single optimal model configuration. Consequently, traditional approaches rely on a single chosen model and neglect model uncertainty. This omission causes overconfident statistical inferences, underestimated standard errors, and potentially biased estimates of the true population effect size. In turn, these estimation errors reduce the replicability of empirical findings.

To resolve these deep-seated methodological vulnerabilities, Bayesian Model Averaging (BMA) offers an innovative and mathematically rigorous paradigm that operates within the broader Bayesian statistical framework. Instead of an arbitrary, ex-ante selection of a single statistical model, BMA directly accommodates model uncertainty. It embeds all theoretically plausible candidate models into a single, cohesive model space. Specifically, BMA treats the model itself as a random variable within a probability space. This formulation effectively manages uncertainty in both the effect size and heterogeneity. By establishing prior probabilities for each model configuration, BMA leverages empirical data to calculate posterior model probabilities. This process quantifies the precise degree of empirical data support that each individual model receives. Ultimately, the final effect size is a weighted average across all models, where posterior probabilities directly determine individual model weights. This approach avoids the inherent biases of single-model selection. It yields a continuous measure of cumulative evidence rather than a binary accept-reject decision.

Researchers operationalize this rigorous approach in meta-analysis as robust Bayesian meta-analysis (RoBMA). RoBMA inherits the core model-averaging principles of BMA. It systematically incorporates multiple publication bias correction models, such as selection models and PET-PEESE (precision-effect test and precision-effect estimate with standard error), into the Bayesian inference framework. It avoids a forced selection of a single optimal model. Instead, the framework weights competing hypotheses within the model space based directly on empirical data. This mechanism provides rigorous quantitative evidence, specifically inclusion Bayes factors (BF), to evaluate 3 critical hypotheses concurrently to govern the validity of literature synthesis. These hypotheses test whether a true population effect exists, whether study-level heterogeneity is present, and whether publication bias contaminates the literature. RoBMA constructs a multi-dimensional model space to map every possible combination of these 3 dimensions (e.g., presence vs. absence of an effect, presence vs. absence of heterogeneity, and presence vs. absence of publication bias). As a consequence, the framework outputs model-averaged posterior estimates that fully incorporate model uncertainty, delivering an exceptionally robust and reliable evaluation of the primary effect size.

Beyond its conceptual and theoretical advantages, the practical execution of BMA-based meta-analysis has become highly viable due to contemporary computational breakthroughs and software integration. This paper comprehensively outlines the concrete, accessible pathways for applied researchers to deploy these advanced statistical techniques in real-world research scenarios. Specifically, the open-source statistical software JASP seamlessly executes the entire analytical pipeline, which provides an intuitive and user-friendly graphical user interface (GUI) for point-and-click execution. For researchers who require scriptability and reproducibility, specialized open-source packages make the methodology fully available into the R programming language. These programming tools facilitate automated reporting and extensive sensitivity analyses. They effectively lower the technical barriers to advanced Bayesian inference for non-statisticians. This paper effectively bridges the historical divide between intricate Bayesian theory and practical application. It delivers a definitive roadmap for researchers to maximize transparency and credibility in meta-analytic conclusions. Therefore, this methodology offers broad applicability for enhancing the robustness of evidence synthesis across psychological science and various other disciplines.

Key words bayesian model averaging, meta-analysis, model uncertainty, publication bias, robust bayesian meta-analysis

附录

附录 1: 基于概率密度函数的四种 BMA 模型公式

BMA 四个模型的概率密度函数(Probability density function, PDF)数学表达式如下:

一、模型 M1: 固定效应零模型(H_0^f)

假设不存在总体效应($\mu = 0$), 且不存在研究间异质性($\tau = 0$)。PDF 公式为:

$$f(y_i|0, 0; \sigma_i^2) = \frac{\phi\left(\frac{y_i}{\sigma_i}\right)}{\sigma_i}$$

二、模型 M2: 固定效应备择模型(H_1^f)

假设存在总体效应($\mu \neq 0$), 但不存在研究间异质性($\tau = 0$)。PDF 公式为:

$$f(y_i|\mu, 0; \sigma_i^2) = \frac{\phi\left(\frac{y_i - \mu}{\sigma_i}\right)}{\sigma_i}$$

三、模型 M3: 随机效应零模型(H_0^r)

假设不存在总体效应($\mu = 0$), 但存在研究间异质性($\tau \neq 0$)。PDF 公式为:

$$f(y_i|0, \tau^2; \sigma_i^2) = \frac{\phi\left(\frac{y_i}{\sqrt{\sigma_i^2 + \tau^2}}\right)}{\sqrt{\sigma_i^2 + \tau^2}}$$

四、模型 M4: 随机效应备择模型(H_1^r)

假设存在总体效应($\mu \neq 0$), 且存在研究间异质性($\tau \neq 0$)。对应完整的元分析层级模型, 其公式为:

$$f(y_i|\mu, \tau^2; \sigma_i^2) = \frac{\phi\left(\frac{y_i - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right)}{\sqrt{\sigma_i^2 + \tau^2}}$$

注: 在上述公式中, y_i 为第 i 个研究的观测效应量, σ_i^2 为该研究对应的已知抽样误差方差, $\phi(\cdot)$ 代表标准正态分布的概率密度函数。

附录 2: RoBMA 模型列表及其对应假设

一、4个无发表偏倚模型，对应BMA下的4个基础模型并增加了“无需发表偏倚校正”的假设：

M1 ($H_0^{fp\bar{b}}$): 固定效应零模型 ($\mu = 0, \tau = 0$)

M2 ($H_1^{fp\bar{b}}$): 固定效应备择模型 ($\mu \sim g(\cdot), \tau = 0$)

M3 ($H_0^{rp\bar{b}}$): 随机效应零模型 ($\mu = 0, \tau \sim h(\cdot)$)

M4 ($H_1^{rp\bar{b}}$): 随机效应备择模型 ($\mu \sim g(\cdot), \tau \sim h(\cdot)$)

二、8个PET-PEESE校正模型。在4个基础模型上分别应用PET或PEESE回归校正小样本效应。

A. 基于固定效应零模型：

M5 (H_0^{fpb}): M1 + PET 校正

M6 (H_0^{fpb}): M1 + PEESE 校正

B. 基于随机效应零模型

M7 (H_0^{rpb}): M3 + PET 校正

M8 (H_0^{rpb}): M3 + PEESE 校正

C. 基于固定效应备择模型

M9 (H_1^{fpb}): M2 + PET 校正

M10 (H_1^{fpb}): M2 + PEESE 校正

D. 基于随机效应备择模型

M11 (H_1^{rpb}): M4 + PET 校正

M12 (H_1^{rpb}): M4 + PEESE 校正

三、24个发表筛选模型校正。在4个基础模型上分别应用6种不同的基于 p 值的权重函数(SelMod1–SelMod6)进行校正。

A. 基于固定效应零模型

M13 (H_0^{fpb}): M1 + SelMod1 (双侧 .05)

- M14 (H_0^{fpb}): M1 + SelMod2 (双侧 .10 & .05)
- M15 (H_0^{fpb}): M1 + SelMod3 (单侧 .05)
- M16 (H_0^{fpb}): M1 + SelMod4 (单侧 .025 & .05)
- M17 (H_0^{fpb}): M1 + SelMod5 (单侧 .50 & .05)
- M18 (H_0^{fpb}): M1 + SelMod6 (单侧 .50, .05, .025)

B. 基于随机效应零模型

- M19 (H_0^{rpb}): M3 + SelMod1 (双侧 .05)
- M20 (H_0^{rpb}): M3 + SelMod2 (双侧 .10 & .05)
- M21 (H_0^{rpb}): M3 + SelMod3 (单侧 .05)
- M22 (H_0^{rpb}): M3 + SelMod4 (单侧 .025 & .05)
- M23 (H_0^{rpb}): M3 + SelMod5 (单侧 .50 & .05)
- M24 (H_0^{rpb}): M3 + SelMod6 (单侧 .50, .05, .025)

C. 基于固定效应备择模型

- M25 (H_1^{fpb}): M2 + SelMod1 (双侧 .05)
- M26 (H_1^{fpb}): M2 + SelMod2 (双侧 .10 & .05)
- M27 (H_1^{fpb}): M2 + SelMod3 (单侧 .05)
- M28 (H_1^{fpb}): M2 + SelMod4 (单侧 .025 & .05)
- M29 (H_1^{fpb}): M2 + SelMod5 (单侧 .50 & .05)
- M30 (H_1^{fpb}): M2 + SelMod6 (单侧 .50, .05, .025)

D. 基于随机效应备择模型

- M31 (H_1^{rpb}): M4 + SelMod1 (双侧 .05)
- M32 (H_1^{rpb}): M4 + SelMod2 (双侧 .10 & .05)
- M33 (H_1^{rpb}): M4 + SelMod3 (单侧 .05)

M34 (H_1^{rpb}): M4 + SelMod4 (单侧 .025 & .05)

M35 (H_1^{rpb}): M4 + SelMod5 (单侧 .50 & .05)

M36 (H_1^{rpb}): M4 + SelMod6 (单侧 .50, .05, .025)

附录 3: RoBMA 假设所对应的模型集合及三种包含贝叶斯因子的计算公式

RoBMA 共包含 36 个具体模型。为了量化不同假设的证据强度, RoBMA 将这 36 个模型按照效应(μ)、异质性(τ)以及发表偏倚的有无进行交叉分类(附表 1), 并通过计算三类“包含式贝叶斯因子”来比较这些模型集合的后验概率。

附表 1 RoBMA 36 个预设模型的假设组合与分类

假设集合	效应假设	异质性假设	发表偏倚假设	包含的具体模型编号	模型数量
\mathcal{H}_0^{fpb}	$\mu = 0$	$\tau = 0$	无	M1	1
\mathcal{H}_1^{fpb}	$\mu \neq 0$	$\tau = 0$	无	M2	1
\mathcal{H}_0^{rpb}	$\mu = 0$	$\tau \neq 0$	无	M3	1
\mathcal{H}_1^{rpb}	$\mu \neq 0$	$\tau \neq 0$	无	M4	1
\mathcal{H}_0^{fpb}	$\mu = 0$	$\tau = 0$	有	PET/PEESE: M5-6 发表筛选模型: M13-18	8
\mathcal{H}_1^{fpb}	$\mu \neq 0$	$\tau = 0$	有	PET/PEESE: M9-10 发表筛选模型: M25-30	8
\mathcal{H}_0^{rpb}	$\mu = 0$	$\tau \neq 0$	有	PET/PEESE: M7-8 发表筛选模型: M19-24	8
\mathcal{H}_1^{rpb}	$\mu \neq 0$	$\tau \neq 0$	有	PET/PEESE: M11-12 发表筛选模型: M31-36	8

基于上述模型分类, RoBMA 通过计算后验概率与先验概率的比值(即包含式贝叶斯因子)来量化三大核心假设的证据强度。计算公式基于上述表格中的模型集合展开:

1. 总体效应包含式贝叶斯因子 (BF_{10})

BF_{10} 通过比较所有假设效应存在的模型集合 (\mathcal{M}_{effect}) 与所有假设效应不存在的模型集合 ($\mathcal{M}_{no-effect}$) 得出。其计算公式为:

$$BF_{10} = \frac{\sum_{m \in \mathcal{M}_{effect}} p(\mathcal{H}_m | data)}{\sum_{n \in \mathcal{M}_{no-effect}} p(\mathcal{H}_n | data)} \bigg/ \frac{\sum_{m \in \mathcal{M}_{effect}} p(\mathcal{H}_m)}{\sum_{n \in \mathcal{M}_{no-effect}} p(\mathcal{H}_n)}$$

根据上表， BF_{10} 的计算公式也可展开写为：

$$BF_{10} = \frac{P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_1^{r_{pb}}|data) + P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_1^{r_{pb}}|data)}{P(\mathcal{H}_0^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data) + P(\mathcal{H}_0^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data)} \bigg/ \frac{P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_1^{r_{pb}}) + P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_1^{r_{pb}})}{P(\mathcal{H}_0^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}}) + P(\mathcal{H}_0^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}})}$$

效应存在 vs. 不存在 后验概率比 效应存在 vs. 不存在 先验概率比

2. 异质性包含式贝叶斯因子 (BF_{rf})

BF_{rf} 通过比较所有假设异质性存在的随机效应模型(\mathcal{M}_{random})与所有假设异质性不存在的固定效应模型(\mathcal{M}_{fixed})的后验概率之和得出。其计算公式为：

$$BF_{rf} = \frac{\sum_{m \in \mathcal{M}_{random}} p(\mathcal{H}_m|data)}{\sum_{n \in \mathcal{M}_{fixed}} p(\mathcal{H}_n|data)} \bigg/ \frac{\sum_{m \in \mathcal{M}_{random}} p(\mathcal{H}_m)}{\sum_{n \in \mathcal{M}_{fixed}} p(\mathcal{H}_n)}$$

根据上表， BF_{rf} 的计算公式也可展开写为：

$$BF_{rf} = \frac{P(\mathcal{H}_1^{r_{pb}}|data) + P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data) + P(\mathcal{H}_0^{f_{pb}}|data)}{P(\mathcal{H}_1^{r_{pb}}|data) + P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data) + P(\mathcal{H}_0^{f_{pb}}|data)} \bigg/ \frac{P(\mathcal{H}_1^{r_{pb}}) + P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}}) + P(\mathcal{H}_0^{f_{pb}})}{P(\mathcal{H}_1^{r_{pb}}) + P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}}) + P(\mathcal{H}_0^{f_{pb}})}$$

随机效应 vs. 固定效应 后验概率比 随机效应 vs. 固定效应 先验概率比

3. 发表偏倚包含式贝叶斯因子 (BF_{pb})

BF_{pb} 通过比较所有包含发表偏倚校正的模型(\mathcal{M}_{bias})与所有不包含校正的基础模型($\mathcal{M}_{no-bias}$)的相对证据强度来计算。其计算公式为：

$$BF_{pb} = \frac{\sum_{m \in \mathcal{M}_{bias}} p(\mathcal{H}_m|data)}{\sum_{n \in \mathcal{M}_{no-bias}} p(\mathcal{H}_n|data)} \bigg/ \frac{\sum_{m \in \mathcal{M}_{bias}} p(\mathcal{H}_m)}{\sum_{n \in \mathcal{M}_{no-bias}} p(\mathcal{H}_n)}$$

根据上表， BF_{pb} 的计算公式也可展开写为：

$$BF_{pb} = \frac{P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_1^{r_{pb}}|data) + P(\mathcal{H}_0^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data)}{P(\mathcal{H}_1^{f_{pb}}|data) + P(\mathcal{H}_1^{r_{pb}}|data) + P(\mathcal{H}_0^{f_{pb}}|data) + P(\mathcal{H}_0^{r_{pb}}|data)} \bigg/ \frac{P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_1^{r_{pb}}) + P(\mathcal{H}_0^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}})}{P(\mathcal{H}_1^{f_{pb}}) + P(\mathcal{H}_1^{r_{pb}}) + P(\mathcal{H}_0^{f_{pb}}) + P(\mathcal{H}_0^{r_{pb}})}$$

发表偏倚存在 vs. 不存在 后验概率比 发表偏倚存在 vs. 不存在 先验概率比

附录 4: R 软件(RoBMA 包)进行贝叶斯模型平均元分析的基础代码示例与说明

在 R 环境下可以使用 RoBMA3.6.1 进行分析(Bartoš, Maier, Wagenmakers, et al., 2025)。模型设定为标准的 PSMA 架构(包含发表筛选模型与 PET-PEESE)，所有先验分布均采用软件包的默认设置。

```

library(RoBMA)
df <- read.csv("Liu_2024_meta_data.csv")
fit <- RoBMA(
  d          = df$ES,
  se         = df$SE,
  model_type = "PSMA",
  algorithm  = "ss",          # 使用 Spike-and-Slab 算法
# 采样精度设置(这里与 JASP 保持一致)
  Chain      = 3,
  adapt      = 5000,         # 适应期
  burnin     = 5000,         # 预热期
  sample     = 10000,        # 正式样本量
  parallel   = TRUE,         # 开启并行计算提高速度
  seed       = 1234,         # 设置随机种子以确保结果一致
  autofit    = FALSE        # 关闭自动拟合
)

```

```

# 查看模型汇总, 包括 BF 值和校正后的效应量
summary(fit)

```

```

# Robust Bayesian meta-analysis

```

```

# Components summary:

```

#	Prior prob.	Post. prob.	Inclusion	BF
# Effect	0.500	0.550		1.225
# Heterogeneity	0.500	1.000		Inf
# Bias	0.500	0.672		2.046

```

# Model-averaged estimates:

```

#	Mean	Median	0.025	0.975
# mu	0.177	0.098	0.000	0.551
# tau	0.432	0.421	0.266	0.661
# omega[0,0.025]	1.000	1.000	1.000	1.000
# omega[0.025,0.05]	0.960	1.000	0.552	1.000
# omega[0.05,0.5]	0.853	1.000	0.202	1.000

```
# omega[0.5,0.95] 0.811 1.000 0.118 1.000
# omega[0.95,0.975] 0.817 1.000 0.118 1.000
# omega[0.975,1] 0.826 1.000 0.118 1.000
# PET 0.321 0.000 0.000 1.946
# PEESE 0.490 0.000 0.000 5.714
# The estimates are summarized on the Cohen's d scale (priors were specified on the Cohen's d scale).
# (Estimated publication weights omega correspond to one-sided p-values.)
```