

· 循证研究与临床转化 ·

# 网络Meta分析研究进展系列（二）：网络Meta分析统计模型及模型拟合软件选择

张天嵩<sup>1</sup>, 孙凤<sup>2</sup>, 董圣杰<sup>3</sup>, 杨智荣<sup>4</sup>, 武珊珊<sup>5</sup>, 田金徽<sup>6</sup>

【摘要】本文系统梳理了目前网络Meta分析（NMA）的常用统计模型、建模策略、分析策略和统计软件，并给出合理选择NMA模型和软件的建议，以期提高NMA制定者和使用者规范实施和解读NMA的能力。

【关键词】网络Meta分析；统计模型；贝叶斯策略；频率学策略；统计软件

【中图分类号】R4

【文献标志码】A

开放科学（源服务）标识码（OSID）



**Advances in methodology of network meta-analysis (2) : the selection of statistical models and softwares for network meta-analysis** Zhang Tiansong\*, Sun Feng, Dong Shengjie, Yang Zhirong, Wu Shanshan, Tian Jinhui. \*Department of TCM, Jing'an District Central Hospital, Fudan University, Shanghai 200040, China.

Corresponding authors: Zhang Tiansong, E-mail: zhangtiansong@fudan.edu.cn; Sun Feng, E-mail: sunfeng@bjmu.edu.cn

【Abstract】This paper reviews the popular statistical models, modeling strategies, analysis approaches, and statistical softwares for network meta-analysis (NMA), and provides some recommendations for how to select and use the models, so as to improve the abilities of NMA makers and users to perform and interpret an NMA correctly.

【Key words】Network meta-analysis; Statistical models; Bayesian approaches; Frequentist approaches; Statistical softwares

Meta分析可以定量、科学地整合研究结果，已在许多科学领域取得显著成果<sup>[1]</sup>，在医学领域中常用于比较不同干预措施有益还是有害。在临床中，针对同一种疾病常存在多种干预措施，医师决策时需根据所有可获得的证据对不同干预措施进行全面比较，但在实际情况下，干预措施的疗效间往往缺乏可靠证据或尚无直接比较证据，该问题可能需要通过两种途径解决：一是开展同时评估多种干预措施利弊的随机对照临床试验，但存在诸多困难，可行性较差；二是通过网络Meta分析（network meta-analysis, NMA）在一个证据体中同时评估多个干预措施，通过合成直接和间接比较的证据，可获得证据体中任一对比较的干预措施之间更为精确的相对效应估计，并根据测量结局（如有效性和安全性）的优劣进行排序<sup>[2]</sup>，由于NMA的上述优势，越来越多方法学团队聚焦于直接和间接证据整合理论和方法，自

Lumley<sup>[3]</sup>提出NMA的概念及Meta回归分析策略以来，基于不同的学术观点，已有多个研究团队提出各自的NMA模型和分析策略，如贝叶斯分层模型、频率学方法、图论法等<sup>[4]</sup>，并可由诸多强力软件实现<sup>[5]</sup>，如WinBUGS、R、Stata、SAS、Stan等。本文通过文献梳理NMA相关理论进展，对NMA统计模型、建模策略、分析策略和统计软件进行总结，并给出选择建议和注意事项。

## 1 建模数据

**1.1 观测数据和缺失数据** 大多数NMA统计模型（如经典Lu & Ades模型<sup>[6]</sup>），均以观测到的数据（observed data）建模；而Zhang等<sup>[7]</sup>认为，一般情况下，纳入NMA中的每个研究只是对所有干预措施中感兴趣的子集（如2个或3个干预措施）进行比较，因此表示成为典型“研究×干预措施”矩阵的数据结构非常稀疏，并以一项NMA研究<sup>[8]</sup>为例说明这一问题。该研究共纳入117个随机对照研究来比较12种新型抗抑郁药物的疗效，其中114个是两臂研究，3个是三臂研究，很显然是具有显著缺失的不完全块结构数据。如果忽略了非完全随机产生的缺失数据，仅依赖观测到或未观测到的数据信息，则估计会产生偏倚<sup>[9]</sup>。Zhang等建议可以从不完全数据中获得附加信息，把研究中未观测到的臂作为缺失数据，然后借用其他研究中观测到的相关信息，基于马尔可夫链蒙特卡洛（Markov chain Monte Carlo, MCMC）算法填补，从而建立了一种新的缺失数据框架建模方法<sup>[7,9]</sup>。

**1.2 数据类型与效应尺度选择** 与经典Meta分析

基金项目：2019年上海市静安区医学科研课题(2019ZY01)；国家自然科学基金(71673003)；北京大学教学新思路(2019YX007, 71009Y0026)

作者单位：<sup>1</sup> 200040 上海,复旦大学附属静安区中心医院中医科；<sup>2</sup> 100191 北京,北京大学循证医学中心 北京大学公共卫生学院流行病学与卫生统计学系；<sup>3</sup> 264000 烟台,烟台市烟台山医院骨关节科；<sup>4</sup> CB1 8RN英国剑桥大学临床医学院初级医疗中心；<sup>5</sup> 100050 北京,首都医科大学附属北京友谊医院国家消化系统疾病临床医学研究中心；<sup>6</sup> 730000 兰州,兰州大学循证医学中心 甘肃省循证医学与临床转化重点实验室

共同通讯作者：张天嵩, E-mail: zhangtiansong@fudan.edu.cn;

孙凤, E-mail: sunfeng@bjmu.edu.cn

doi: 10.3969/j.issn.1674-4055.2020.07.01

相同<sup>[10]</sup>，根据测量结局的不同，NMA的数据类型一般也分为二分类数据（dichotomous data）、连续型数据（continuous data）、有序数据（ordinal data）、计数数据（count data）和时间事件数据（time-to-event data）等五大类。二分类数据是NMA中最常用的数据类型，一项调查研究表明<sup>[11]</sup>，共纳入186个NMA研究，其中测量结局数据类型报告为二分类数据者111个（60%）、连续型数据者53个（28%）、时间事件数据者17个（9%）、有序数据者5个（3%）；二分类数据效应量选择比值比（odds ratio, OR）者66个（35%）、危险比（risk ratio, RR）者44个（23%）、危险差（risk difference, RD）者0个（0%）、同时报告OR、RR、RD者1个（1%），连续型数据效应量选择均数差（mean difference, MD）者43个（23%）、标化均数差（standardized mean difference, SMD）者9个（5%）、均数比（ratio of means, RoM）者1个（1%），时间事件数据效应量选择风险比（hazard ratio, HR）者17个（9%），计数数据效应量选择率比（rate ratio, RR）者5个（3%）。

目前尚无明确的选择效应量的理论原则可供参考。一般认为，根据不同测量结局数据类型，从可解释性和数学特性等角度考虑，选择合适的效应尺度<sup>[11]</sup>。以二分类数据为例，因OR具有良好的数据特性，Coory等<sup>[13]</sup>推荐在间接比较中选择使用OR，一些学者则推荐使用RR<sup>[14]</sup>，Caldwell<sup>[11]</sup>则认为OR有时会因呈现大的效应量出现错误解释，建议采用HR（如果事件时间研究的结果是二分类数据，也可由四格表计算出HR）。

建议在进行NMA时，可以基于各种因素合理选择效应量：①测量结局数据类型；②干预效应的可交换性及可加性，在进行Meta分析时，效应尺度假设观测到的干预效应是线性可加并假设在研究间是可交换的<sup>[15]</sup>；③科学依据，如有时获益危险比（RRb）和损害危险比（RRh）可能提供不同量级和方向的结果<sup>[11]</sup>，再如在合并不同基线风险的研究时，选择RR、OR和RD等不同效应量可能会影响干预排序等<sup>[16]</sup>；④直接比较的研究间异质性，如RD比RR、OR更容易出现异质性等；⑤必要时可选择不同效应量，作为敏感性分析。

**1.3 数据格式** NMA中待分析的数据，按格式一般分为标准化数据（normalized data）和非标准化数据（non-normalized data）两种，适用于不同建模模型和拟合模型软件要求。前者又称为长数据或窄数据，是每个研究中每个臂占一行，主要变量有研究名称、每个臂发生事件人数和总人数；后者称为短数据或宽数据，是每个研究占一行，主

要变量有研究名称、所有干预措施的事件发生数和总人数。

因为NMA部分相关术语没有完全统一，为了便于后续讨论，在此我们介绍和区别几个术语：①干预（treatment）和臂（arm）：前者是指待评估的药物或医疗设备，而后者是指关于单个研究中被随机分配到的药物或医疗设备的患者资料（组）。②参照干预（reference treatment）和基线干预（baseline treatment）：前者是指标准对照干预如安慰剂，可用于与其他阳性干预进行比较；后者是指在每个研究中被指定为对照臂的干预。两者在大多数情况下相同，但也可能不同。③设计（design）：是仅指某试验中相比较干预措施组合，如Hasselblad<sup>[17]</sup>研究分析过的“AHCPR Smoking Cessation Guideline Panel”戒烟数据中共含有24个研究，有“未干预（No intervention）”、“自助（Self-help）”、“个体辅导（Individual counselling）”、“小组辅导（Group counselling）”等四种干预措施，假设分别以A、B、C、D表示，则Gritz 1992研究中含有干预措施A、B，则称为AB设计；Cottraux 1983研究中含有干预措施A、C、D，则称为ACD设计，依次类推，该概念在设计×干预交互模型（design-by-treatment interaction model）中非常重要。

## 2 统计模型

**2.1 基本原理** 假设NMA中共含有S个研究，T个干预措施，N个直接配对比较，干预措施k相对于干预措施c（中 $k, c=1, 2, \dots, T$ ）的相对效应为 $\mu_{ck}$ ，实际上并不需要计算所有 $\mu_{ck}$ ，只需要估计部分基本参数即可，基本参数 $\mu_t$ （ $t=1, 2, \dots, T-1$ ）表示（-1）个相互独立的合并干预效应。简单方法，通过选择T个干预措施中某一个为参照（如A），则每个 $\mu_t$ 代表干预措施t相对于A（ $t=1, 2, \dots, T; t \neq A$ ），所以 $\mu_t \equiv \mu_{At}$ ，NMA需要估计所有 $\mu_t$ S，因此，所有其他比较（功能参数）的合并效应则可以通过一致性等式 $\mu_{ck} = \mu_{Ak} - \mu_{Ac} = \mu_k - \mu_c$ 获得。

**2.2 主要的统计模型** 简单的网络结构为由A、B、C三个干预措施构成的星形结构，假设有“A vs. C”和“B vs. C”直接比较的证据，则可以通过校正间接比较（adjusted indirect comparison, AIC）方法按NMA一致性等式获得“A vs. B”的证据，即是最初用NMA统计分析的Bucher法。后来逐渐发展起来并在当前最为流行的NMA数据统计分析模型如层次模型、多元Meta分析模型、Meta回归模型、两步策略线性模型等都是基于NMA基本原理中一致性等式，最小化需要待估计的参数数量，虽然模型各异，实质上是等效的<sup>[18]</sup>，基本上是广义线性模型的拓展，没有优劣

之分。在实际中，应根据软件实现模型的能力及研究者专业知识和技能来选择的相应的模型，基于此，本文主要介绍两种最为流行的结构模型（structural models）：层次模型和多元Meta分析模型。

**2.2.1 层次模型** 层次模型（hierarchical models）是NMA最常用的数据分析模型<sup>[11,18]</sup>，由Lu与Ades<sup>[6]</sup>、Salanti<sup>[19]</sup>、Dias<sup>[20]</sup>等提出并不断完善，建模灵活、易于扩展，是目前众多贝叶斯NMA的基础，核心为广义线性模型（generalized linear model, GLM），可以通过不同的连接函数拟合服从二项分布、正态分布、泊松分布等指数分布的数据。英国国家健康与临床研究所（National institute for health and clinical excellence, NICE）决策支持机构（Decision Support Unit, DSU）制作的技术支持文档（Technical Support Documents, TSDs）中提供了层次模型针对不同数据类型的贝叶斯NMA代码，有兴趣的读者可以从DSU官网（nicedsu.org.uk）上下载学习。

设有  $N$  ( $i=1,2,\dots,N$ ) 个研究，每个研究有  $k$  ( $k=1,2,\dots,K$ ) 个干预措施（臂），第  $i$  个研究中的干预措施组合为  $D_i$ （即是前文所谓的设计），第  $i$  个研究第  $k$  个臂中感兴趣的参数为  $\eta_{ik}$ ，该参数可由不同度量模型（measurement model）用来描述数据，对于二类数据，以  $p_{ik}$  表示第  $i$  个研究第  $k$  个干预措施组事件发生概率，总人数  $n_{ik}$ ，发生的事件人数为  $r_{ik}$ ， $r_{ik}$  服从二项分布： $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$ ，而概率  $p_{ik} = \frac{\exp(\eta_{ik})}{1 + \exp(\eta_{ik})}$ 。

层次模型有两种建模策略，第一种为Ades、Salanti、Dias等提出的基于对比（contrast-based）模型<sup>[21-23]</sup>，指定每个研究中的臂  $b_i$  作为基线参照，则有  $n_{ik} = \mu_{ib_i} + \delta_{ib_k}$  ( $k=D_i$ )。式中  $\mu_{ib_i}$  称为研究截距，是固定效应，针对二分类数据表示第  $i$  个研究中臂  $b_i$  的发生事件（戒烟数据中戒烟成功）比数的自然对数。 $\delta_{ib_k}$  表示第  $i$  个研究中臂  $k$  ( $k=D_i$ ) 相对于  $b_i$  比较的特定比较效应，相应为比数比的自然对数（lnOR），当  $k=b_i$  时， $\delta_{ib_k}=0$ ；若  $k \neq b_i$ ，则随机效应模型为  $\delta_{ib_k} \sim N(\mu_{b_i k}, \sigma^2_{b_i k}) = N(\mu_{1k} - \mu_{1b_i}, \sigma^2_{b_i k})$ ，模型中  $\mu_{1k}$  关键参数，表示如果以干预措施1为参照干预，对于  $k > 1$ （表示在1之后）的臂  $k$  与1比较的平均干预效应，显然  $\mu_{11}=0$ 。若为固定效应模型，则有  $\sigma^2_{b_i k}=0$ ， $\delta_{ib_k} = \mu_{b_i k} = \mu_k = \mu_{b_i}$ 。

研究间异质性（方差） $\sigma^2_{kc}$  有几种不同的假定，如研究间同方差（ $\sigma^2$ ）或异方差结构（ $\sigma^2_{kc}$ ），要注意的是，当纳入的研究为多臂研究时，各个臂之间效应存在相关性，必须指定效应值的方差-协方差矩阵。设第  $i$  个研究的随机效

应的估计向量为  $\delta_i$ ，假设其服从多元正态分布，简单处理，假设研究间同方差为  $\sigma^2$ ，则有<sup>[2,23,24]</sup>：

$$\delta_i = \begin{pmatrix} \delta_{i12} \\ \vdots \\ \delta_{i1a_i} \end{pmatrix} \sim N_{a_i-1} \left( \begin{pmatrix} \mu_{i12} \\ \vdots \\ \mu_{i1a_i} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \dots & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & \dots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \dots & \sigma^2 \end{pmatrix} \right)$$

式中  $a_i$  ( $a_i=2,3,\dots$ ) 表示第  $i$  研究的臂的数， $\mu_{i1l} = \mu_{i1k} - \mu_{i1l}$ ；由  $\delta_i$  的方差-协方差矩阵可知，对于任意两个臂之间的协方差为  $\sigma^2/2$ 。根据多元正态分布的条件分布公式可得每一个  $\delta_{i,tk}$  条件分布为：

$$\delta_{i,tk} \mid \begin{pmatrix} \delta_{i12} \\ \vdots \\ \delta_{i1(k-1)} \end{pmatrix} \sim N \left( (\mu_{i1k} - \mu_{i1t}) + \frac{1}{k} \sum_{j=1}^{k-1} [\delta_{i1j} - (\mu_{i1j} - \mu_{i1t})], \frac{k}{2(k-1)} \sigma^2 \right)$$

第二种为Hong<sup>[9]</sup>、Hawkins等<sup>[23]</sup>提出的基于臂（arm-based）建模策略。仍假设有  $N$  ( $i=1,2,\dots,N$ ) 个研究，每个研究有  $k$  ( $k=1,2,\dots,K$ ) 个干预措施（臂），第  $i$  个研究中臂的总数为  $K$ ， $s[k]$  表示研究中臂  $k$ ， $t[k]$  表示该臂施加给患者的干预措施， $d_t$  表示干预措施  $t$  相对于参照干预 ( $t=1$ ) 的干预效应，显然有  $d_1=0$ ，臂  $k$  事件发生数与未发生数的比值对数为  $\eta_k$ ， $\mu_{s[k]}$  表示参照干预的研究特定测量结果，根据异质性假设不同可分为固定效应模型和随机效应模型<sup>[23]</sup>：一是，假定不同随机效应方差发生个体干预措施水平则为具有干预措施特定方差随机效应模型（Random treatment effects with treatment-specific variance）， $\eta_k = \mu_{s[k]} + \delta_{s[k],t[k]}$ ，式中  $\delta_{s[k],t[k]} \sim N(d_{t[k]}, \frac{\sigma^2_{t[k]}}{2})$ ；二是，假定所有比较间随机效应方差相同，则为共方差随机干预效应模型（random treatment effects with common variance）， $\eta_k = \mu_{s[k]} + \delta_{s[k],t[k]}$ ，式中  $\delta_{s[k],t[k]} \sim N(d_{t[k]}, \frac{\sigma^2}{2})$ ；二是，如果研究间无异质性，则为固定效应模型  $\eta_k = \mu_{s[k]} + d_{t[k]}$ 。

**2.2.2 多元Meta分析模型** 多元Meta分析模型（multivariate meta-analysis model）由White等<sup>[25]</sup>在Salanti等工作基础上提出，假设整个网络中含有  $T$  个干预措施，研究  $i$  均为两臂研究，其观察效应为  $y_i$ ，设  $T-1$  个基本参数  $\mu_i$  表示  $iA$  间比较的效应，为每个研究中的系列子集，随机项为  $\delta_i$ ，随机误差为  $\epsilon_i$ ；如果某些研究中不含参照干预措施A，可采用数据填补技术给予臂A一个极小的数据信息，NMA中所有比较均可通过基本对比来表达，每个研究中报告一个或多个测量结局  $y_{iA} = y_{it}$ ，则一致性模型为  $(y_{i1}, \dots, y_{i(T-1)}) = (\mu_{i1}, \dots, \mu_{i(T-1)}) + (\delta_{i1}, \dots, \delta_{i(T-1)}) + (\epsilon_{i1}, \dots, \epsilon_{i(T-1)})$ 。该模型适合任意测量结局，如MD、SMD、lnRR、lnOR等。可以看出，对于比较组不含臂A的研究，通过数据填补变为至少含有三臂的研究，因此在每个研究中不同

干预措施间的 $(\delta_{1...} \delta_{i(T-1)})$ 和 $(\epsilon_{1...} \epsilon_{i(T-1)})$ 满足方差-协方差的多元正态分布<sup>[26]</sup>。

设计×干预交互模型基于本模型发展而来<sup>[25,27]</sup>，衍生出固定效应和随机效应不一致性模型等。仍假设某NMA网络含有T个干预措施A, B, C等等，令A作为共同参照，首先，假设纳入Meta分析的研究中均含有A（对于不含臂A的研究，可通过数据填补）；令 $d=1, 2, \dots, D$ 表示设计类型，假设在第d个设计（ $d=1$ 至D）第i个试验（ $i=1$ 至nd）中，干预措施J（ $J=B, C, \dots$ ）与A比较的结果为 $y_{di}^{AJ}$ ， $y_{di}^{AJ}$ 可表示任一测量结局，如MD, SMD, logRR, logOR等等，则不一致性模型为<sup>[28]</sup>： $y_{di}^{AJ} = \delta_{di}^{AJ} + \beta_{di}^{AJ} + \omega_{di}^{AJ} + \epsilon_{di}^{AJ}$ ， $J=B, C, \dots$ 。

模型中， $\delta^{AJ}$ 表示J与A间对比的总效应，可视作为固定参数，为感兴趣的参数。 $\beta_{di}^{AJ}$ 表示异质性，是指同一设计真实治疗效应的变异，一般视其为随机效应： $\beta_{di}^{AJ} : N(0, \Sigma)$ ，该模型对 $\Sigma$ 无约束，作为非结构化模型，在模型中，J-A对比研究间方差为 $\text{var}(\beta_{di}^{AJ}) = \Sigma^{JJ}$ ，K-J对比（ $J, K \neq A$ ）研究间方差为 $\text{var}(\beta_{di}^{AK} - \beta_{di}^{AJ}) = \Sigma^{JJ} - 2\Sigma^{JK} + \Sigma^{KK}$ ，不同对比研究间方差确实各异。对于 $\Sigma$ 还存在两种可能的结构化异质性模型，在全结构化异质性模型中，假设所有干预对比有相同的研究间方差 $\tau^2$ ，考虑J-A对比研究间方差，对于所有J允许 $\Sigma^{JJ} = \tau^2$ ，考虑K-J对比研究间方差，同样允许 $\Sigma^{JK} = \tau^2/2$ （ $J \neq K$ ），因此有 $\Sigma = \tau^2 P(0.5)$ ，此处 $P(\rho)$ 为矩阵，其对角元素等于1，远离对角元素等于 $\rho$ 。 $\omega_{di}^{AJ}$ 项表示设计间不一致性，一般作为固定效应处理<sup>[25]</sup>，也可以作为随机效应模型处理<sup>[28]</sup>，假定 $\omega_{di} : \sim Nc(0, \Sigma_{\omega})$ ，式中 $\omega_{di} = (\omega_{di}^{AB}, \omega_{di}^{AC}, \dots, \omega_{di}^{AK})^T$ ， $\Sigma_{\omega}$ 为矩形矩阵，其对角线数值为 $\tau_{\omega}^2$ ，对角线外其他数据为 $\tau_{\omega}^2/2$ ，在一致性模型中，设 $\omega_{di}^{AJ} = 0$ 。 $\epsilon_{di}^{AJ}$ 表示研究内误差，假设 $\epsilon_{di} : N(0, S_{di})$ ，而假定 $S_{di}$ 已知。

NMA中同质性假设和一致性假设是两个非常重要的假设。基于同质性假设衍生出固定效应模型（fixed-effect model）和随机效应模型（random-effects model），基于一致性假设衍生出一致性模型（consistency model）和不一致性模型（inconsistency model），这些不同的模型体现在结构模型中，在实际应用中，可以结合数据特点、模型特点及拟合数据的不同模型拟合优度来选择。

**2.3 建模策略** 虽然NMA统计模型有多种，但建模策略即是NMA的度量模型（measurement model）分为两种：基于臂（arm-based, AB）和基于对比（contrast-based, CB），为了避免引起术语混淆，本文采用AB似然和CB似然，两者的主要区别有：①两者假设不同，Shuster<sup>[29]</sup>介绍了Meta分

析中研究随机（studies at random, SR）和效应随机（effects at random, ER）两种随机效应方式，前者假设纳入Meta分析中的每个研究是独立的，后者假定每个研究的效应是随机的。CB似然需要ER类假设，而AB需要SR类假设<sup>[7]</sup>。②两者估计方法不同，CB似然采用近似正态来产生每一相对干预合并效应量及其方差；而AB似然则可直接对每个臂的绝对效应如ln(odds)建模<sup>[30]</sup>。③AB似然处理随机干预效应为臂水平治疗反应的变异<sup>[23]</sup>。

虽然这两种建模策略一直存在争议<sup>[22]</sup>，且从理论上讲AB似然可能破坏了随机化，但似乎并非实质问题，因为一些研究表明，AB和CB建模策略均适用于NMA数据建模分析<sup>[22]</sup>，两者获得的结果基本上相等<sup>[23]</sup>。

**3 分析策略**

贝叶斯统计学与经典统计学是统计学两个主要学派，两者对概率定义有所区别，而NMA也主要有贝叶斯和频率学两大分析策略。

**3.1 贝叶斯分析框架** 贝叶斯统计综合了未知参数的总体信息、样本信息及先验信息，根据贝叶斯定理，获得未知参数的后验分布，进而对未知参数进行统计推断，因其灵活性强，广泛应用于科学研究中。贝叶斯NMA全面考虑了参数的不确定性，可以用直接概率来描述这些参数（如，一种干预措施优于另一种干预措施的概率）。贝叶斯NMA中主要采用MCMC和集成嵌套拉普拉斯近似（integrated nested Laplace approximations, INLA）两种算法，在使用MCMC算法时要注意几个主要问题<sup>[2]</sup>：

一是对模型中的参数合理指定先验分布。先验分布可以允许利用先验信息确定，但在先验信息或知识缺乏的情况下，常用的处理方法<sup>[31,32]</sup>：对绝对或相对效应测量（位置参数）选择指定无信息或模糊正态先验分布；根据测量结局类型，对研究间异质性标准差 $\tau$ （尺度参数）指定为弱信息通常是均匀先验分布。在NMA文献中<sup>[32]</sup>，无论是固定效应模型还是随机效应模型，对所有研究中每个干预效应（如 $\mu_{ib}$ ）和基本参数（如假定以A为参照， $d_{AB}, d_{AC}, \dots, d_{AK}$ ）通常指定的先验分布为正态分布 $\text{Normal}(0, 10^{3-5})$ ；对于随机效应模型中，研究间异质性标准差 $\tau$ 的先验分布指定为均匀分布 $\text{uniform}(0, 2 \sim 10)$ ，如果配对比较的研究稀疏，则可以指定半正态分布 $\text{HN}(0.5)$ 或 $\text{HN}(1)$ <sup>[33]</sup>。二是若在后验分布中使用马尔科夫蒙特卡罗链模拟，则判断检验马尔科夫链收敛十分重要；需要指出的是，与MCMC算法相比，INLA算法费时短且不需要收敛性诊断。三是通常使用偏差信息准则（Deviance information criterion, DIC）对模型进行

比较并评估其拟合度，公式为： $DIC = p_D + \bar{D}$ ，式中， $p_D$ 又称为参数的有效数量， $\bar{D}$ 用于测量模型复杂程度，用于测量模型的拟合优度，由偏差 $D(\theta)$ 度量。DIC值越小说明模型越好，一般认为不同模型的DIC的差值在3个或以上单位即为重要差异<sup>[2,20]</sup>。四是贝叶斯模型因其复杂性和指定先验分布潜在的主观性，可能会影响到研究的结果。

**3.2 频率学分析框架** 频率学分析框架广泛应用于经典Meta分析中，但近年来涌现出的Meta回归及多元Meta分析等NMA模型和方法等，均可由频率学分析框架实现，且具有计算速度快、不需要为参数指定先验、可避免蒙特卡洛误差分析、可由Stata、R、SAS等通用软件轻松实现等优点，也得到广泛应用<sup>[2,26]</sup>。在NMA中，频率学分析方法的估计和推断常是基于最大似然（maximum likelihood）等算法<sup>[21]</sup>。频率学分析框架是观察数据在假设参数值的样本分布下的概率，得到的参数估计比贝叶斯方法更间接<sup>[34]</sup>。

因为NMA许多技术及相关统计软件，最初建立和开发时均基于贝叶斯方法，当前大多数NMA文献都是采用贝叶斯分析框架；同时，相对于频率学方法，贝叶斯方法在处理复杂或稀疏数据时更有价值<sup>[34]</sup>，因此，在NMA的数据分析时，建议首选贝叶斯方法，以频率学方法作为敏感性分析，但对于非统计学专业研究者而言可能会存在不少困难。

**4 统计软件**

可实NMA的软件或程序大体可以分为三类：一是，贝叶斯统计分析软件，如WinBUGS、OpenBUGS、JAGS、Stan等，可以实现贝叶斯NMA；二是，通用分析软件，如Stata、R、SAS等，既可以实现贝叶斯NMA，也可以实现频率学NMA；三是NMA专用分析包或程序，如ADDIS、GeMTC、ITC、某些Excel插件如NetmetaXL和

MetaXL等。决定进行NMA后，需要根据数据类型、效应量、证据类型等情况确定合适的统计模型和分析框架，然后再选择合适的计算软件，具体如表1所示<sup>[2,26]</sup>。

根据建模数据、软件功能特点及易用性等方面考虑，建议WinBUGS、Stata、R等三个主要NMA软件两两配合使用或三者联用：

(1) WinBUGS是NMA中最为常用的软件包<sup>[34,35]</sup>，但对于初学者而言，学习和理解其模型代码是一个挑战，且该软件在数年操作、数据注释和绘图制图等方面的功能有所欠缺<sup>[35]</sup>。建议使用JAGS，它是一个采用MCMC算法拟合贝叶斯层次模型的程序，与BUGS语言同源，但与OpenBUGS和WinBUGS有些小差异，并可由其他软件如R调用。

(2) Stata软件的mvmeta命令和network组命令均可拟合多元Meta分析模型<sup>[2,26]</sup>，其中mvmeta功能强大，但在数据预处理、模型参数设置、结果图示化等方面，需要一定的数学及统计学技能，初学者不易掌握；network组命令提供了数据准备、数据分析、绘制相关图形、干预措施效果排序等功能，特别是最新版本可通过调用mvmeta命令或WinBUGS软件等实现频率学及贝叶斯NMA，分别拟合，操作方便，建议使用network组命令。这两个命令均可在联网情况下，在Stata命令输入窗口键入命令：net from <http://www.homepages.ucl.ac.uk/~rmjwiww/stata/>，按提示完成安装，具体使用方法，可以阅读相关文献<sup>[2,26,36]</sup>。

(3) R软件除了有通过调用WinBUGS（如R2WinBUGS）、OpenBUGS软件（R2OpenBUGS、BRugs、rbugs）、JAGS软件（如R2jags、rjags、runjags）、Stan软件（如rstan）进行NMA的扩展包外，还有gemtc、pcnetmeta、bnma、netmeta等几个NMA专用包。

表1 网络Meta分析模型及常用实现软件

模型或方法	软件	扩展包或命令
Bucher法（逐步策略）	R软件	meta包、metafor包、rmeta包
	Stata软件	indirect命令
	WinBUGS、OpenBUGS	/
Meta回归模型	R软件	nlme包、metafor包
	Stata软件	metareg命令
	WinBUGS、OpenBUGS	/
多元Meta分析模型	R软件	mvmeta包、metafor包、mmaINLA包
	Stata软件	mvmeta命令、network组命令
	WinBUGS、OpenBUGS	/
基于图形原理（图论）法 层次模型	R软件	netmeta包
	WinBUGS、OpenBUGS	/
	ADDIS/GeMTC	/
	R软件	gemtc包、pcnetmeta包、bnma包
两步法线性模型	SAS软件	genmod程序、glimmix程序、MCMC程序
	WinBUGS、OpenBUGS	/
	Stata软件	mvmeta命令
	SAS软件	mixed程序

gemtc包基于层次模型、采用对比似然建模策略，但因检查出的错误未及时修改而于2020年5月28日从R综合档案网络（comprehensive R Archive Network, CRAN）镜像仓库中移除。pcnetmeta包是基于臂建模策略<sup>[36]</sup>，通过调用JAGS软件，可以对二分类数据、连续型数据、计数数据等不同数据类型进行NMA，如针对二分类数据，可以计算绝对风险（absolute risk, AR）、RD、lnOR、lnRR、OR、RR等效量，指定Probit连接函数建模<sup>[37,38]</sup>。bnma包用于拟合Dias<sup>[20]</sup>提出的层次模型，基于对比策略建模，通过调用JAGS软件进行NMA，但不需要用户编写JAGS代码和设定初始值等，使用方便。

## 5 结语

本文针对典型数据的NMA主流统计模型进行了梳理和介绍，并给出相关的选择和使用建议，不正之处，敬请斧正。NMA作为一种新技术和方法发展迅速，在实施时会面临着更多的挑战，一些重要问题尚未完美解决，如，NMA满足同质性、相似性（可传递性）、一致性等假设的证明问题；效应调节因子影响干预效应大小、干预措施疗效排序等问题。总之，NMA可以对多个干预措施的有效性和安全性等进行比较，是一种具有广阔前景的技术和方法，但研究者在应用时也必须注意其不足之处。

## 参考文献

- Gurevitch J, Koricheva J, Nakagawa S, *et al.* Meta-analysis and the science of research synthesis[J]. *nature*, 2018, 555(7691):175–82.
- 张天嵩. 网络Meta分析. 见王吉耀主编. 循证医学与临床实践[M]. 4版. 北京: 科学出版社, 2019, 133–82.
- Lumley T. Network meta-analysis for indirect treatment comparisons[J]. *Stat Med*, 2002, 21(16):2313–24.
- Lifeng Lin. Quantifying and presenting overall evidence in network meta-analysis[J]. *Stat Med*, 2018, 37(28):4114–25.
- Xu C, Niu Y, Wu J, *et al.* Software and package applicating for network metaanalysis: A usage-based comparative study[J]. *J Evid Based Med*, 2018, 11(3):176–83.
- Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons[J]. *Statistics in Medicine*, 2004, 23:3105–24.
- Zhang J, Chu H, Hong H, *et al.* Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness[J]. *Stat Methods Med Res*, 2017, 26(5):2227–43.
- Cipriani A, Furukawa TA, Salanti G, *et al.* Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis[J]. *Lancet*, 2009, 373(9665):746–58.
- Hong H, Chu H, Zhang J, *et al.* A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons[J]. *Res Synth Methods*, 2016, 7(1):6–22.
- Higgins J, Thomas J (Senior Editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*. Cochrane, 2019. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).
- Nikolakopoulou A, Chaimani A, Veroniki AA, *et al.* Characteristics of Networks of Interventions: A Description of a Database of 186 Published Networks[J]. *PLoS One*, 2014, 9(1):e86754.
- Caldwell DM, Welton NJ, Dias S, *et al.* Selecting the Best Scale for Measuring Treatment Effect in a Network Meta-Analysis: A Case Study in Childhood Nocturnal Enuresis[J]. *Res Synth Methods*, 2012, 3(2):126–41.
- Coory M, Jordan S. Frequency of effect modification affecting indirect comparisons[J]. *PharmacoEconomics*, 2010, 28(9):723–32.
- Altman DG, Deeks JJ, Sackett DL. 1998. Odds ratios should be avoided when events are common (letter)[J]. *British Medical Journal*, 317(7168):1318.
- van Valkenhoef G, Ades AE. Evidence synthesis assumes additivity on the scale of measurement: response to "Rank reversal in indirect comparisons" by Norton *et al.*[J]. *Value Health*, 2013, 16(2):449–51.
- Norton EC, Miller MM, Wang JJ, *et al.* Rank reversal in indirect comparisons[J]. *Value Health*, 2012, 15(8):1137–40.
- Hasselblad V. Meta-analysis of multitreatment studies[J]. *Med Decis Making*, 1998, 18(1):37–43.
- Efthimiou O, Debray TP, van Valkenhoef G, *et al.* GetReal in network meta-analysis: a review of the methodology[J]. *Res Synth Methods*, 2016, 7(3):236–63.
- Salanti G, Higgins JP, Ades AE, *et al.* Evaluation of networks of randomized trials[J]. *Statistical Methods in Medical Research*, 2008, 17:279–301.
- Dias S, Sutton AJ, Ades AE, *et al.* Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials[J]. *Med Decis Making*, 2013, 33(5):607–17.
- Hoaglin DC, Hawkins N, Jansen JP, *et al.* Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2[J]. *Value Health*, 2011, 14(4):429–37.
- White IR, Turner RM, Karahalios A, *et al.* A Comparison of Arm-Based and Contrast-Based Models for Network Meta-Analysis[J]. *Stat Med*, 2019, 38(27):5197–213.
- Hawkins N, Scott DA, Woods B. 'Arm-based' parameterization for network meta-analysis[J]. *Res Synth Methods*, 2016, 7(3):306–13.
- 董圣杰. 贝叶斯Meta分析. 见詹思延主编. 系统综述与Meta分析[M]. 北京: 人民卫生出版社, 2019, 223–56.
- White IR, Barrett JK, Jackson D, *et al.* Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression[J]. *Research Synthesis Methods*, 2012, 3(2):111–25.
- 张天嵩, 田金徽, 孙凤, 等. 网状Meta分析. 见詹思延主编. 系统综述与Meta分析[M]. 北京: 人民卫生出版社, 2019, 79–107.
- Higgins JPT, Jackson D, Barrett JK, *et al.* Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies[J]. *Research Synthesis Methods*, 2012, 3(2):98–110.
- Jackson D, Barrett JK, Rice S, *et al.* A design-by-treatment interaction model for network meta-analysis with random inconsistency effects[J]. *Stat Med*, 2014, 33(21):3639–54.
- Shuster JJ, Guo JD, Skyler JS. Meta-analysis of safety for low event-rate binomial trials[J]. *Res Synth Methods*, 2012, 3(1):30–50.
- Gü nhan BK, Friede T, Held L. A design-by-treatment interaction model for network meta-analysis and meta-regression with integrated nested Laplace approximations[J]. *Res Synth Methods*, 2018, 9(2):179–94.
- Lambert PC, Sutton AJ, Burton PR, *et al.* How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS[J]. *Stat Med*, 2005, 24(15):2401–28.
- van Valkenhoef G, Lu G, de Brock B, *et al.* Automating network meta-analysis[J]. *Res Synth Methods*, 2012, 3(4):285–99.