

· 循证研究与临床转化 ·

网络Meta分析研究进展系列（七）：多水平网状Meta回归

刘凤琪¹, 石舒原¹, 杨智荣², 武珊珊³, 张天嵩⁴, 田金徽⁵, 董圣杰⁶, 孙凤¹

【摘要】本文简要介绍了一种对个体水平数据和集合水平数据的混合证据体进行分析的新方法：多水平的网状Meta回归，阐述了该方法的产生背景、显著优势、模型公式、实施步骤、案例解读，并对未来的发展和应用前景进行了展望。

【关键词】网络Meta分析；多水平的网状Meta回归；个体水平数据；效应修饰因子

【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 (7): multi-level network meta-regression Liu

Fengqi*, Shi Shuyuan, Yang Zhirong, Wu Shanshan, Zhang Tiansong, Tian Jinhui, Dong Shengjie, Sun Feng.

*Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China.

Corresponding authors: Sun Feng, E-mail: sunfeng@bjmu.edu.cn

[Abstract] This paper briefly introduces a new method for analyzing the mixed evidence body of individual-level data and aggregate-level data: multi-level network meta-regression. It explains the background, significant advantages, model formulas, implementation steps, and case analysis of this method, and prospects for future development and application.

[Key words] Network Meta-analysis; Multi-level network Meta-regression; Individual patient data (IPD); Effect modifier

1 前言

临床干预措施在人群中的效果或安全性常会受到诸多因素的影响，有些因素可能与干预措施的效应产生交互作用，即某一干预措施与结局指标的关联强度可能随该因素的水平不同而改变，此现象称为效应修饰作用，该因素被称为效应修饰因子。

随机对照试验（RCT）是获得干预措施效果无偏估计的金标准。通过随机化，可认为研究人群的所有基线指标的分布在各比较组中实现了平衡。但研究同一临床问题的多个RCT可能因为各自纳入排除标准的不同，而导致所纳入的人群之间效应修饰因子的分布存在差异。标准的间接比较或网状Meta分析（NMA）^[1,2]是建立在假设RCT人群之间所有效应修饰因子都分布均匀的条件下去进行的。因此，当不同RCT间的分布存在差异时，传统间接比较和NMA就自动忽略了这些差异的存在，而可能产生有偏估计。

经典的Meta分析及NMA均来自于聚合水平的数据（AgD），虽然在证据分级系统被看作是最高等级的证据，但其局限性也日益突显^[3]，除上述这一问题外，效应的不一致性、研究间异质性、不确定性（可信区间较宽）及发表偏倚等等，均可能产生有偏的或者难以解释的结果。近年来，有学者提出收集个体水平数据（IPD）资料，基于目标人群执行校准的Meta分析方法^[4]。根据IPD效应修饰因子的联合分布，对AgD的分布进行调整，从而实现二者之间的分布均衡，再将效应值进行合并以产生无偏估计。目前此类方法越来越普遍，各国决策机构也越来越支持采纳其相关证据^[5]。

2 理论进展

早期的基于人群校准方法是由间接比较衍生而来^[4]。常见校准方法包括：匹配调整间接比较（MAIC）^[6]；模拟治疗比较（STC）^[7]等。以上方法仅适用于包含2个研究的间接比较，要求1个研究获得IPD水平资料，另1个研究获得AgD水平资料，2种干预措施通过某一共有桥梁进行间接比较^[8]，其结果仅适用于其中提供AgD资料的试验人群。优势之处在于，当1个RCT拥有IPD数据，另1个RCT只有AgD，在这种情况下能够最大程度的利用已有数据，通过校准初步实现了效应修饰因子在2个RCT间的分布均衡，与标准间接比较相比降低了偏倚风险。但上述方法应用范围较窄，不适用于复杂证据体，且默认将提供

基金项目：国家自然科学基金(72074011);北京大学医学部教改课题(2019YB39);北京大学2020年度“教育大数据研究项目”(2020YBC02)

作者单位：¹ 100191 北京,北京大学公共卫生学院流行病与卫生统计学系;² CB1 8RN 剑桥,英国剑桥大学临床医学院初级医疗中心;³ 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;⁴ 200040 上海,复旦大学附属静安区中心医院中医科;⁵ 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;⁶ 264000 烟台,烟台市烟台山医院骨关节科

通讯作者：孙凤, E-mail: sunfeng@bjmu.edu.cn

doi: 10.3969/j.issn.1674-4055.2020.12.01

AgD资料的人群视作决策目标人群，其真实世界外推性有限，无法在指定的目标人群中进行效果预测^[9]。

随着2002年NMA的提出^[2]，极大拓展了间接比较的框架。由于绝大多数疾病或症状的干预措施都远不止两种，决策者也提出同时对多个干预措施进行综合评估的需求，并开始探讨有闭环的，既有直接比较，又有间接比较的复杂证据体的疗效估计。在复杂证据体中，如果获得的都是AgD资料，则经典的NMA^[2]已经可以很好地处理；如果存在重要的效应修饰因子，后续发展的网状Meta回归^[10]也能进行调整；如果获得了证据体中全部的IPD资料，则可以用理想的IPD的NMA^[11]估计疗效。上述方法均可给出证据体中任意两个干预措施的两两比较的结果并进行排序，近十年来很受青睐。但上述方法在解决效应修饰因子分布这一问题上，网状Meta回归可能会因缺乏个体信息而产生聚合偏倚，而获得证据体中全部IPD进行IPD的NMA往往十分困难。

因此，Phillippo等^[12-14]2019年开发了一种多水平的网状Meta回归（ML-NMR）来执行基于IPD校准的NMA。ML-NMR源自于生态学参考文献^[15]中介绍的一种方法，该方法原本用于合并个体水平调查资料和集合水平的队列数据。2012年Jansen等首先将此模型应用于NMA中IPD和AgD的合并^[16]，当时仅局限于二分类结局和二分类协变量。ML-NMR将上述方法进行了总结归纳和拓展。

已得到证明，即便只能获得证据体中部分IPD资料或RCT基线特征每个水平联合分布的结果，ML-NMR也可以根据人群的效应修饰因子分布进行校准，将间接比较或NMA校准到某一目标人群中^[17]。该方法的稳健性已在Phillippo博士^[14]论文中进行了充分模拟和实例论证，并已在两次国际大会上进行了交流^[12,13]，与随机效应NMA相比，ML-NMR模型具有更好的拟合性，即降低了不确定性，也避免了聚合偏倚，模型的可解释性也更强。

3 ML-NMR模型公式及实施步骤

ML-NMR模型主要包括两个子模型，一个子模型用于IPD资料，另一个子模型用于AgD资料。假设研究者需要比较K种干预措施，共纳入了J个RCT。在获得IPD资料的试验中，用 y_{ijk} 和 x_{ijk} 分别代表在j试验中接受了k干预措施的个体i的结局指标和协变量水平。在AgD资料的试验中，“.”表示为集合水平指标，用 $y_{.jk}$ 代表在j试验中接受k干预措施者的汇总结局水平， $f_{jk}(x)$ 代表在j试验中接受k干预措施者的协变量分布情况。

两个子模型的具体公式如下：

IPD资料模型：

$$y_{ijk} \sim \pi_{\text{Ind}}(\theta_{ijk})$$

$$g(\theta_{ijk}) = \eta_{jk}(x_{ijk}) = \mu_j + x_{ijk}^T (\beta_1 + \beta_{2,k}) + \gamma_k$$

AgD资料模型：

$$y_{.jk} \sim \pi_{\text{Agg}}(\theta_{.jk})$$

$$\theta_{.jk} = \int g^{-1} \{ \eta_{jk}(x) \} f_{jk}(x) dx$$

$\pi_{\text{Ind}}(\theta_{ijk})$ 和 $\pi_{\text{Agg}}(\theta_{.jk})$ 分别代表IPD和AgD数据的似然分布。 θ_{ijk} 和 $\eta_{jk}(x_{ijk})$ 分别代表在j试验中接受了k干预措施的具有协变量向量 x_{ijk} 的个体i在协变量条件下的平均结局和线性预测因子。 $\theta_{.jk}$ 是在j试验中接受了k干预措施的边际平均结局。 $g(\cdot)$ 是一个适当的连接函数， μ_j 代表j试验的基线水平，是该研究的特定截距， $\beta_{2,k}$ 代表效应修饰因子针对干预措施k的系数，即效应修饰因子与干预措施k的交互作用， β_1 代表其他协变量的回归系数。 γ_k 代表与第1种干预措施相比，k干预措施的效果，即 $\gamma_1=0$ 。 $f_{jk}(x)$ 代表在j试验中接受了k干预措施组的协变量分布情况。

ML-NMR的执行过程可大致分为5步：①由IPD水平的分布获得AgD水平分布情况；②构建AgD水平回归模型；③共有效应修饰因子效应恒定假设；④评估残留的异质性和不一致性；⑤在特定目标人群中产生效应估计。

另外需注意的是：ML-NMR中的共有效应修饰因子假设，即假设效应修饰因子与干预措施效果的交互作用在某一类干预措施中是完全一致的，或者说是可交换的，在实际分析时，通常根据多个干预措施的作用原理或机制，将其分为若干组，在每组中对效应修饰因子的交互作用进行估计。这一假设与标准NMA或其他校准方法的假设相比，其限制性更小，假设本身也更加合理。

ML-NMR可使用Stan软件开发的代码进行实现，Stan是一个用于进行哈密顿蒙特卡罗模拟的贝叶斯统计软件包^[18]，其基于Stan之前为进行NMA分析而开发的框架进行了拓展，包括了基于人群特征调整和纳入IPD数据的模型建立和分析。2019年Phillippo等开发了操作性强的R软件包（<https://github.com/dmphilippo/multinma>），并在2020年发表的相关案例中提供了数据和代码^[17]。除此之外，WinBUGS和JAGS也能实现ML-NMR的运行。

4 案例解读

Phillippo等^[17]以中重度斑块状牛皮癣的6种干预措施为例，对ML-NMR方法的执行过程及结果质量进行了最新的验证。

研究纳入了4个持续时间12周以上的Ⅲ期临床试验（UNCOVER-1, UNCOVER-2, UNCOVER-

3, FIXTURE), 4个RCT中包含的干预措施包括: 依那西普、依克珠单抗(2周一次)、依克珠单抗(4周一次)、苏金单抗(150 mg)、苏金单抗(300 mg)和安慰剂。其中UNCOVER系列的3个RCT均获得了IPD资料, FIXTURE试验为AgD资料(PASI)量表评分的提高比例是否达到75%、90%或100%(分别表示为PASI 75、PASI 90和PASI 100)。其中5个临床相关指标被认为可能具有效应修饰作用, 分别为: ①牛皮癣占身体表面积百分比; ②牛皮癣患病持续时间; ③先前是否接受过牛皮癣系统治疗; ④是否患有银屑病关节炎; ⑤体重。上述5个潜在效应修饰因子的信息既包含在UNCOVER系列的IPD资料中, 也在FIXTURE试验的AgD资料中。该研究纳入了4个RCT构成了包含6种干预措施的网状证据体, 其证据结构如图1所示。

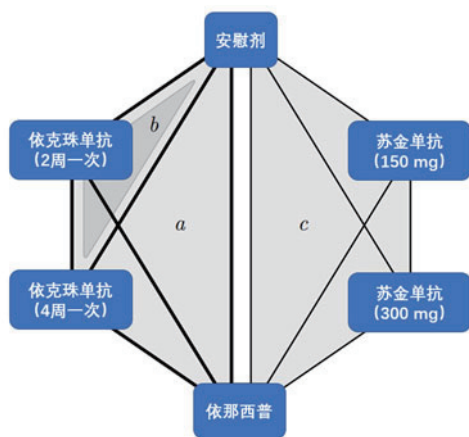


图1 牛皮癣案例的网状证据结构图

该研究采用准蒙特卡洛抽样法(QMC), 从5维Sobol序列^[19]中共提取了10 000个点, 将AgD的RCT的效应修饰因子联合分布映射成个体水平数据, 即作为ML-NMR模型的个体水平数据。研究者还根据UNCOVER系列3个试验中的效应修饰因子分布, 对5个因素是否存在相关关系进行了分析, 例如患病持续时间越长者越可能曾接受过牛皮癣系统治疗。

然后, 研究者对PASI 75这一结局构建ML-NMR模型, 应用Stan软件, 采用贝叶斯法, 为每个参数指定无信息先验分布 $N(0, 100^2)$ 。PASI 75是一个基于PASI量表评分产生的二分类变量, 通过一个连接函数, 将模型参数和干预措施效果均表示为PASI量表评分的标准化均值差(SMDs)。鉴于依克珠单抗和苏金单抗均作为白介素17A受体抑制剂, 具有共同的作用模式, 故假设在二者的所有剂量下, 效应修饰因子均具有相同的交互作用。在以上共有效应修饰因子的假设前提下, 经过4个并行链分别运行2000次迭代来对模型进

行构建, 并使用潜在尺度缩减因子评估其收敛性。上述构建的模型可认为是一个以效应修饰因子水平为自变量, 以结局效应值为因变量的回归模型, 根据所构建的ML-NMR模型, 在四个RCT人群中分别计算各干预措施两两比较的结果。在提供IPD数据的UNCOVER系列3个RCT中, 根据每个个体的效应修饰因子水平, 即可获得基于每个个体的效应值, 求均值可得到结局效应值。在提供AgD资料的FIXTURE试验中, 根据效应修饰因子的联合分布情况, 即可获得结局效应值。根据Stan软件开发的代码可自动完成上述一系列计算, 其具体的计算原理及数理公式在Phillippo等发表的方法学文献^[17]中有详细解读。

最终, 研究者获得了6种干预措施在PASI 75这一结局上, 分别在UNCOVER-1、UNCOVER-2、UNCOVER-3和FIXTURE这4个人群中的所有两两比较的结果。结果显示, 干预措施两两比较的结果在这4个人群之间存在微小的差异, 例如: 与安慰剂相比, 依那西普在FIXTURE人群中似乎比在UNCOVER人群中更加有效。

在该研究的4个人群中, 对于PASI 75这一结局, 5种阳性干预措施与安慰剂相比都是有效的, 其中依克珠单抗(2周一次)在4个人群中均显示达到PASI 75的患者占比最高, 约为85.9%~90.3%, 该组与其他干预措施进行比较的SMD的95%可信区间也均不包含0, 故可认为是在PASI 75结局中, 依克珠单抗(2周一次)为表现最佳的干预措施。

研究者还对ML-NMR的结果质量进行了评价。以FIXTURE人群中依克珠单抗(2周一次)和苏金单抗(300 mg)的疗效比较为例, 研究者分别比较了ML-NMR计算出的、MAIC法计算出的、标准间接比较和NMA得出的SMD。ML-NMR显示二者SMD为0.34(0.10, 0.58), MAIC法得出的二者SMD为0.28(0.00, 0.56), 标准的间接比较结果显示SMD为0.37(0.12, 0.63), 随机效应NMA得出的SMD为0.45(-0.02, 0.92)。其中ML-NMR与MAIC这两种基于人群校准的方法所得出的结果差异不大, 与MAIC法相比, ML-NMR纳入了研究的所有可用信息, 降低了不确定性。与标准NMA相比, ML-NMR比随机效应NMA具有更好的拟合度, 并且由于解释了因效应修饰因子导致的试验内变异而提高了估计的精度, 残差值更小。

在该案例中, 研究者采用ML-NMR对中重度斑块状牛皮癣的多种干预措施效果进行比较, 结果表明, 在考虑共有效应修饰因子的假设成立, 且没有忽略重要的效应修饰因子的情况下, ML-NMR确实表现出较好的性能。与固定效应模型或

随机效应模型NMA相比, ML-NMR对RCT中效应修饰因子的作用进行了分析解释, 构建的模型拟合度更佳。与早期校准方法相比, ML-NMR应用范围更加广泛, 可以处理复杂证据体, 并且可以获得在证据体中每个RCT人群中的比较结果。

这一案例详细解释了ML-NMR的执行过程和所得结果, 但该研究只对PASI 75这1个二分类结局进行了分析, 作者也提到ML-NMR在连续型变量、多分类变量、生存资料等数据类型中的应用仍需探索。

5 总结与展望

ML-NMR是一种分析IPD和AgD混合证据体的新方法, 其优势主要是能够调整效应修饰因子的分布。总的来说, ML-NMR具有以下优点: ①干预措施的比较仍基于随机分组, 避免了指示征混杂; ②不需要全部试验的IPD资料; ③可依据人群特征校准RCT结果, 在任一已知分布的目标人群中进行效果预测; ④可产生药企研发、临床医生、临床指南制定者、卫生技术评估及决策制定机构等共同关注的结果。

RCT结果外推到真实世界的目标人群时, 就可能出现人群代表性不足导致的药物效力-效果差距^[20], 影响RCT结果在实际决策目标人群中的实现^[21]。根据多个RCT构成的复杂证据体进行ML-NMR后, 不仅可获得基于每个RCT的协变量分布所得出的比较结果, 还可以在其他任一协变量分布已知的人群中进行效果估计^[17], 从而提高RCT结果在决策目标人群中的适用性。这一方法为药物和治疗措施的评价提供了更多的可行性, 有助于推动RCT与真实世界证据的桥梁和融合, 使RCT结果最大可能地为“真实世界”决策而服务, 最大程度提高决策目标人群的健康收益。

综上, ML-NMR仍需进一步发展和验证。该方法对目标人群效果的估计是基于目标人群的协变量分布均已知的的前提下, 尚未考虑到人群协变量分布估计可能存在的问题和不确定性^[22]。另外, ML-NMR在其他类型的数据中的应用仍需进一步的案例分析, 例如多分类数据、生存分析数据等, 其统计分析过程和结果的稳定性也需大量模拟研究进行比较和验证。最后, ML-NMR能否将观察性研究或单臂研究等也整合到证据体中进行合并, 也是未来需要解决的重要问题。

参考文献

[1] Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials[J]. *Journal of clinical epidemiology*, 1997, 50(6):683-91.
 [2] Lumley T. Network meta-analysis for indirect treatment comparisons[J]. *Statistics in medicine*, 2002, 21(16):2313-24.

[3] Egger M, Moons KG, Fletcher C, et al. GetReal: from efficacy in clinical trials to relative effectiveness in the real world[J]. *Res Synth Methods*, 2016, 7(3):278-81.
 [4] Phillippo DM, Dias S, ElSada A, et al. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals[J]. *Int J Technol Assess Health Care*, 2019, 35(3):221-8.
 [5] Wang SV, Schneeweiss S, Gagne JJ, et al. Using Real-World Data to Extrapolate Evidence From Randomized Controlled Trials[J]. *Clinical pharmacology and therapeutics*, 2019, 105(5):1156-63.
 [6] Jiang Y, Ni W. Performance of unanchored matching-adjusted indirect comparison (MAIC) for the evidence synthesis of single-arm trials with time-to-event outcomes[J]. *BMC Med Res Methodol*, 2020, 20(1):241.
 [7] Caro JJ, Ishak KJ. No head-to-head trial? simulate the missing arms[J]. *Pharmacoeconomics*, 2010, 28(10):957-67.
 [8] Phillippo DM, Dias S, Ades AE, et al. Equivalence of entropy balancing and the method of moments for matching-adjusted indirect comparison[J]. *Res Synth Methods*, 2020, 11(4):568-72.
 [9] Maksymowych WP, Strand V, Nash P, et al. Comparative effectiveness of secukinumab and adalimumab in ankylosing spondylitis as assessed by matching-adjusted indirect comparison[J]. *European journal of rheumatology*, 2018, 5(4):216-23.
 [10] 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.
 [11] Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology[J]. *Res Synth Methods*, 2015, 6(4):293-309.
 [12] Phillippo DM, Dias S, Ades AE, et al. Synthesis of individual and aggregate level data using multilevel network meta-regression: extension to general likelihoods[J]. In: *Abstract from Royal Statistics Society International Conference*, 2019, Belfast, United Kingdom, 2019.
 [13] Phillippo D, Dias S, Ades T, et al. Synthesis of individual and aggregate level data using multilevel network meta-regression: extension to general likelihoods, <http://www.iscb2019.info/>[J]. In: *40th Annual Conference of the International Society for Clinical Biostatistics-KU Leuven*, Leuven, Belgium, 2019.
 [14] Phillippo DM. Calibration of treatment effects in network meta-analysis using individual patient data[D]. University of Bristol, 2019.
 [15] Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data[J]. *Stat Med*, 2006, 25(12):2136-59.
 [16] Jansen JP. Network meta-analysis of individual and aggregate level data[J]. *Res Synth Methods*, 2012, 3(2):177-90.
 [17] Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons[J]. *Journal of the Royal Statistical Society Series A*, 2020, 183(3):1189-210.
 [18] Carpenter B, Gelman A, Hoffman MD, et al. Stan: A Probabilistic Programming Language[J]. *Journal of Statistical Software*, 2017, 76(1):1-29.
 [19] Sobol IM. On the distribution of points in a cube and the approximate evaluation of integrals[J]. *USSR computational mathematics and mathematical physics*, 1967, 7(4):86-112.
 [20] Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response[J]. *Nature reviews Drug discovery*, 2011, 10(7):495-506.
 [21] Bothwell LE, Podolsky SH. The Emergence of the Randomized, Controlled Trial. *The New England journal of medicine*, 2016, 375(6):501-4.
 [22] Phillippo DM, Dias S, Ades AE, et al. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study[J]. *Stat Med*, 2020.

本文编辑: 孙竹