

· 循证研究与临床转化 ·

网络Meta分析研究进展系列（十七）：网络Meta分析中缺失结局数据的处理策略

张天嵩^{1,2}, 董圣杰³, 杨智荣⁴, 武珊珊⁵, 田金徽⁶, 孙凤⁷

【摘要】本文主要介绍了在网络Meta分析中如何处理缺失结局数据的策略，以期提高NMA制定者和使用者规范实施和解读缺失数据NMA的能力。

【关键词】网络Meta分析；缺失结局数据；统计模型；模式混合模型

【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 (17): Strategies for addressing missing outcome data in network meta-analysis Zhang Tiansong*, Dong Shengjie, Yang Zhirong, Wu Shanshan, Tian Jinhui, Sun Feng. *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 introduces the strategies for addressing missing outcome data in network meta-analysis, so as to improve the abilities of NMA makers and users to perform and interpret an NMA of missing outcome data correctly.

[Key words] Network meta-analysis; Missing outcome data; Statistical models; Pattern-mixture model

缺失数据（missing data）在科学研究中非常普遍，是指因各种原因应该得到而未得到的数据^[1]，单个研究中如果出现缺失数据结果，可能导致研究精度降低和效应量估计偏倚^[2]。缺失数据在系统评价和Meta分析中也很常见，它有多种含义，一般是指缺失结局数据（missing outcome data, MOD）、缺失感兴趣的干预组、缺失研究（missing studies）、缺失统计量（missing statistics）、缺失整个研究结果、缺失研究水平的基本特征等^[2,3]。MOD在实施网络Meta分析（Network meta-analysis, NMA）普遍存在，研究者们尝试数据处理（data-manipulation）和建模（modelling）等策略对MOD进行处理^[4]，本文梳理和介绍几种主要的分析方法和策略，以期为研究人员了解相关知识和进行NMA提供参考。

1 缺失结局数据

MOD这一名词广泛应用临床研究文献中，是指某些研究对象的结局数据不存在的情境^[5]。不

管研究者采用何种严格的方法学，纳入Meta分析的随机对照试验在收集每个研究对象相关结局数据时，几乎不可避免地会出现失败的情况^[6]而造成数据缺失，但系统评价员有时候很难判断MOD的原因是原始研究对结局未进行测量还是未进行报告^[3]。

近年来，研究者开始关注系统评价中的数据缺失问题，按照主要结局数据类型，既往文献关注MOD较多的是二分类结局数据缺失和连续型结局数据缺失（missing continuous outcome data, MCOD）。如一项针对387个NMA的调查研究表明^[7]，共273个NMA或明确或含蓄地指明纳入分析的研究存在数据缺失，仅141（41%）个NMA明确指出拟采用的分析计划，但实际上只有95个（35%）NMA真正实施；以二分类数据为主要结局的NMA共127个，在以连续型数据为主要结局的92个NMA中有5个NMA针对纳入的每个研究的每个臂MOD情况进行报告。

2 结局数据缺失机制

数据缺失导致的偏倚风险依赖于数据缺失机制，思考为什么数据会出现缺失非常重要^[3]。数据缺失机制一般是用来描述数据缺失依赖于研究对象的特征及结局的倾向如何，一般可分为^[5]：

2.1 完全随机缺失 完全随机缺失（missing completely at random, MCAR）是指对于所有研究对象而言缺失结局的概率是相同的，不依赖于研究对象的特征，如因时间安排困难造成的研究对象失去预约机会。MCAR假设的意思是指提供数

基金项目：国家自然科学基金面上项目(72074011);2019年上海市静安区医学科研课题(2019ZY01)

作者单位：¹ 200040 上海,复旦大学附属静安区中心医院中医科;² 200072 上海,上海市静安区中医医院;³ 264000 烟台,烟台市烟台山医院骨关节科;⁴ CB1 8RN 剑桥,英国剑桥大学临床医学院初级医疗中心;⁵ 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;⁶ 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;⁷ 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

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.1674-4055.2021.10.01

据的某组研究对象是研究总体的随机抽样，但在实际中常常难以成立。

2.2 随机缺失 随机缺失 (missing at random, MAR) 是指数据缺失的倾向性与研究对象有关，但缺失结局的概率与结局本身无关。如，假定初级中学随机分配接受不同的干预措施，以症状严重程度量表来评价能否降低学校相关性焦虑，越是年龄小的学生有可能提供的信息越少，因为他们可能对量表条目理解困难。在Meta分析中，因获得的是聚合数据、且没有足够的脱落人群信息，MAR假设常难以验证。

2.3 非随机缺失 非随机缺失 (not missing at random, NMAR) 是指缺失与缺失的数据相关，即是数据缺失的概率与未观察到的结果数据有关，如因干预措施真正的疗效较差而导致患者出现脱落。

3 结局数据缺失处理策略

在Meta分析中常应用数据处理和建模策略来处理缺失数据^[4,8]。

3.1 数据处理策略 主要包括排除MOD或基于退化概率分布的填补等两种方法，因其简单、直观且不需要复杂的统计软件而在经典的Meta分析和NMA中广泛使用^[4,7,9]。另一方面，在应用时要注意其因简单易行而带来的挑战：①若MOD数量比较多、缺失机制不可忽略，则排除MOD有可能会产生偏倚的结果^[6,10]；②数据填补因采用退化概率分布常缺乏合理性。

3.1.1 排除策略 排除 (exclusion) 策略是指忽略缺失数据，仅分析可以利用、结局数据完整的研究数据，称为完整病例Meta分析 (Complete cases meta-analysis)，是许多Meta分析采用的最普通参考策略^[5]。在应用时要注意满足该策略的MCAR假设，如果数据是NMAR，则数据缺失率越高则分析结果越不可靠^[11]。

3.1.2 填补策略 数据填补 (data imputation) 策略是指采用适当的估计方法，以适合的数据填补代替缺失数据。适用于数据填补策略的情况^[1]：如果数据缺失属于非MCAR；缺失率较小 (15%以下)；含有缺失值的变量对于研究问题有非常重要的意义；注意事项：一般要遵循保守原则；要根据缺失数据的类型、机制，选择不同的方法必要时采用多种方法进行填补，并对不同的数据填补方法产生的结论进行敏感性分析。

试验中针对缺失数据常用填补数据的方法来处理，最常用两种策略^[5]：一是以提供数据的研究对象测量结局的平均值来替代缺失值，是简单填补法；二是把最后一次观测值作替代缺失值，称为末次访视结转 (last observation carried

forward, LOCF)。还可采用多重热层填补法、趋势得分法、多重回归填补法、数据扩增法等多重填补法，通过随机生成值来替代缺失值得到多个原始数据集的衍生数据集进行分析，需要涉及贝叶斯理论、MCMC方法等理论和算法^[1]。

3.2 建模策略 建模策略是以观察到的结局、缺失结局和结局调节因素等联合似然为中心进行处理数据。与数据处理策略相比，建模策略具有概念和统计上的优势，它可以定量化处理缺失结局和观察到的结局之间关联，而不是在分析前对数据进行调整；且可合并关联的不确定性^[6,10,12]。因此，每个研究可以对MOD进行效应量及其标准误的校正，并将这些校正传递到Meta分析对效应量的合并估计中。常用建模策略有模式混合模型 (pattern-mixture model) 和选择模型 (selection model)。

4 模式混合模型

4.1 模型简介^[13,14] 在系统评价中，模式混合模型是处理MOD最常用的方法^[13]。以二分类数据为例，简单介绍一步法模式混合模型：假定纳入NMA中有N个研究含有T个干预措施 (也称为臂)，在第i个研究中第k(k=1,2,K,a_i)个臂中，事件发生人数为r_{ik}，缺失人数m_{ik}，总人数为n_{ik}，令观察到的事件发生人数与缺失人数是来自相关二项式分布的随机抽样^[12]，则有：r_{ik}~Bin(p_{ik}^o, n_{ik}-m_{ik})和m_{ik}~Bin(q_{ik}, n_{ik})，式中，p_{ik}^o为完成观察者事件发生的条件概率，q_{ik}为MOD的概率。

模型混合模型可以用描述研究中研究观察完成者和缺失个体的分布^[10,12]：在第i个研究中第k个臂事件发生概率p_{ik}可基于事件是否观察到或缺失来进行条件建模，则有：p_{ik}=p_{ik}^o·(1-q_{ik})+p_{ik}^m·q_{ik}，式中，p_{ik}^m第i个研究中第k个臂事件缺失个体事件发生的条件概率。则在模型中一个非常重要的参数——缺失信息OR (informative missingness odds ratio, IMOR) 参数可定义为： $\delta_k = \frac{p_{ik}^o / (1 - p_{ik}^o)}{p_{ik}^m / (1 - p_{ik}^m)}$ ，式中ln(δ_{ik})=φ_{ik}~N(ω_{ik}, σ_{ik}²)，因数据缺失机制不明，假定为MAR是合理的^[14]，则令φ_{ik}~N(0,1)。

通过logit函数连接可得随机效应NMA模型：logit(p_{ik})=μ_i+θ_{ik}，θ_{ik}~N(μ_{t_{ik}}, τ²)。式中，μ_i=logit(p_{i1})表示第i个研究基线臂的事件发生比数的对数 (ln(odds))，θ_{ik}是指第i个研究干预臂k(k≠1)相对于基线臂1事件发生的比值的对数ln(OR)；t_{ik}表示第i个研究中第k个臂中干预措施，因此有t_{ik}∈{A,B,K}；如果在多臂研究中 (研究中a_i≥3) 则因与同一个参照干预措施相比的其他干预措施的ln(OR)间存在相关性，需要指定服从多元正态分布；τ²在实践中，常假设为一个常数，相当于在

多臂研究中两个 $\theta_{ik}(k \neq 1)$ 之间的相关系数为0.5。

假定在纳入NMA中的干预措施 $T=\{A, B, C, K\}$ 中, 令 A 为参照干预措施, 有第 j 和 l 个干预措施, 且 $j \neq k \in T \setminus \{A\}$, 它们之间的相对效应可根据一致性原理^[15], 通过基本参数计算所得: $\mu_l = \mu_j - \mu_A$ 。

4.2 适用数据类型 该模型可用于MOD为二分类数据^[14]和连续型数据的NMA^[16]。针对MOD为二分类数据模型(一步策略)如上文所述。Spineli^[16]等认为, 相比于填补策略, 从概念和统计学上而言复杂的策略处理MCOD更为合理, 分析和比较Mavridis等^[17]提出的模式混合模型两步策略, 指出了不足之处, 一是, 把研究内干预效应及相应标准误固定到缺失信息参数的均数和方差的分布中; 二是没有允许观察到的数据贡献到缺失信息参数估计中。在其二分类MOD模式混合模型基础上, 建立了MCOD的分析模型, 可获得均数差、标准化均数差、均数比等效应量, 其建模思路与二分类MOD相同, 本文不再赘述, 感兴趣的读者可以阅读相关文献^[16]。

4.3 模型拟合策略 该模型可通过一步(one-stage)或两步(two-stage)策略来实现。一步策略如上所述, 两步法思路是: 以二分类数据为例, 第一步采用模式混合模型校正研究内 $\ln(OR)$; 第二步采用Dias^[18]等提出基于对比策略建模。简而言之, 一步策略就是从每个研究每个臂中提取的信息作为输入数据, 并把模式混合模型融入到NMA的层次模型中; 二步策略是把通过模式混合模型获得的研究内的效应估计值作为输入数据来实施NMA^[14]。

一步策略和两步策略均可在贝叶斯框架下实现, WinBUGS、R等软件包可以选用。文献^[14, 16]中提供了分析数据示例及R软件实现代码, 可以下载学习。

4.4 选择模型 针对二分类MOD, 也可采用通过对观察到的数据建模的选择模型来处理^[10, 13, 19]。假定纳入NMA中有 N 个研究含有 T 个干预措施(臂), 在第 i 个研究中第 $k(k=1, 2, K, a_i)$ 个臂中, 事件发生人数为 r_{ik} , 缺失人数 m_{ik} , 总人数为 n_{ik} , 令其服从正项分布: $(r_{ik}, n_{ik} - r_{ik} - m_{ik}, m_{ik})^T \sim M(p_{1,ik}, p_{2,ik}, p_{3,ik}, n_{ik})$, 式中, $p_{1,ik} = (1 - c_{1,ik})g p_{ik}$, $p_{2,ik} = (1 - c_{0,ik})g(1 - p_{ik})$, $p_{3,ik} = c_{1,ik}g p_{ik} + c_{0,ik}g(1 - p_{ik})$ 。 $p_{1,ik}$ 表示观察到的潜在事件的边缘概率, $p_{2,ik}$ 表示未观察到的潜在事件的边缘概率, 表示MOD的实际概率(如 $p_{3,ik} = q_{ik}$)。该模型对缺失数据是否经历事件进行条件建模, $c_{1,ik}$ 和 $c_{0,ik}$ 分别表示第 i 个研究第 k 臂中MOD基于有无潜在事件研究对象的条件概率, 该模型贝叶斯框架下实现^[19], 因在模型中只有参数 q_{ik} 是从数据估计所得, 其他参数需要指定先验分布。

5 结语

系统评价员应当重视NMA时数据缺失情况, 但目前尚无处理MOD的“最佳”策略^[13]。如果在对MOD进行NMA时, 应注意的事项: ①应尽可能和原始研究者联系, 索要所缺失的数据; ②虽然排除和填补策略简单易行, 但注意其明确的假设; ③建议模式混合模型一步策略处理MOD; ④建议采用基于不同假设、不同方法分析数据, 进行敏感性分析。

参考文献

- [1] 张天嵩,董圣杰,周支瑞. 高级Meta分析方法-基于Stata实现[M]. 上海:复旦大学出版社,2015,247-60.
- [2] Mavridis D,White IR. Dealing with missing outcome data in meta-analysis[J]. Res Synth Methods,2020,11(1):2-13.
- [3] Higgins JPT,Thomas J,Chandler J,et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021.Available from www.training.cochrane.org/handbook.
- [4] Spineli LM,Kalyvas C. Comparison of exclusion, imputation and modelling of missing binary outcome data in frequentist network meta-analysis[J]. BMC Med Res Methodol,2020,20(1):48.
- [5] Mavridis D,Chaimani A,Efthimiou O,et al. Addressing missing outcome data in meta-analysis[J]. Evid Based Mental Health, 2014,17(3):85-9.
- [6] Higgins JPT,White IR,Wood AW. Imputation methods for missing outcome data in meta-analysis of clinical trials[J]. Clinical Trials,2008,5(3):225-39.
- [7] Spineli LM,Yepes-Nunez JJ,Sch ü nemann HJ. A systematic survey shows that reporting and handling of missing outcome data in networks of interventions is poor[J]. BMC Med Res Methodol,2018,18:115.
- [8] Akl EA,Kahale LA,Agoritsas T,et al. Handling trial participants with missing outcome data when conducting a meta-analysis:a systematic survey of proposed approaches[J]. Syst Rev,2015,4:98.
- [9] Gamble C,Hollis S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis[J]. J Clin Epidemiol,2005,58:579-88.
- [10] White IR,Higgins JP,Wood AM. Allowing for uncertainty due to missing data in meta-analysis--part 1:two-stage methods[J]. Stat Med,2008,27:711-27.
- [11] Wood AM,White IR,Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals[J]. Clin Trials Lond Engl,2004,1:368-76.
- [12] Turner NL,Dias S,Ades AE,Welton NJ. A Bayesian framework to account for uncertainty due to missing binary outcome data in pairwise meta-analysis[J]. Stat Med,2015,34:2062-80.
- [13] Spineli LM. An empirical comparison of Bayesian modelling strategies for missing binary outcome data in network meta-analysis[J]. BMC Med Res Methodol,2019,24,19(1):86.
- [14] Spineli LM,Papadimitropoulou K,Kalyvas C. Pattern-mixture model in network meta-analysis of binary missing outcome data:one-stage or two-stage approach?[J]. BMC Med Res Methodol,2021,21(1):12.
- [15] Lu G,Ades AE. Assessing evidence inconsistency in mixed treatment Comparisons[J]. J Am Stat Assoc,2006,101(474):447-59.
- [16] Spineli LM,Kalyvas C,Papadimitropoulou K. Continuous(ly) missing outcome data in network meta-analysis:A one-stage pattern-mixture model approach[J]. Stat Methods Med Res,2021,30(4):958-75.

(下转1161页)