

· 循证研究与临床转化 ·

网络Meta分析研究进展系列（九）：预测模型的网络Meta分析

武珊珊¹, 杨智荣², 董圣杰³, 张天嵩⁴, 田金徽⁵, 孙凤⁶

【摘要】本文简要介绍了一种针对预测模型进行网络Meta分析的新方法即两阶段Meta回归方法, 阐述了该方法的产生背景、具体理论基础和案例解读, 并对该方法的应用注意事项及未来的发展应用前景进行了讨论。

【关键词】网络Meta分析; 预测模型; 预后评分; 外部验证

【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 (9): Network Meta-Analysis of prediction models

Wu Shanshan*, Yang Zhirong, Dong Shengjie, Zhang Tiansong, Tian Jinhui, Sun Feng. *Department of Clinical Epidemiology and EBM, National Clinical Research Center for Digestive Diseases, Beijing Friendship Hospital, Capital Medical University, Beijing, 100050, China.

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

[Abstract] This paper briefly introduces a new method for network meta-analysis of prediction models. It explains the background, basic theory, implementation and case analysis of this method, discusses the points for attention and prospects for future development and application.

[Key words] Network meta-analysis; Prediction model; Prognostic score; External validation

临床预测模型或预后评分近年来应用越来越广泛, 它们通过对个体某些特征指标如实验室检验或症状、体征等的利用来估计现在或未来发生某特定事件的风险, 为医疗决策提供一定的科学依据^[1,2]。目前, 在很多疾病领域针对同一个研究结局都已发表了为数不少的预测模型或预后评分, 但大多数预测模型长期处于“多数被建立, 少数被验证, 极少被应用”的情况^[2,3]。以心血管疾病发生的预测模型为例, 截止2013年6月已有363个预测模型被建立, 其中仅36%的预测模型进行了外部验证; 而在进行了外部验证的预测模型中, 仅65%和58%的研究分别报告了模型的区分度和校准度, 大多数模型都未被临床广泛应用^[4]。一方面是因为大多数预测模型缺乏足够的外部验证, 且外部验证研究大多存在研究设计的缺陷或报告不规范, 导致模型的外推性不高; 另一方面是因为针对同一个研究结局的模型太多且缺乏对不同模型之间预测准确度的比较研究, 无法识别各个模型的预测准确度孰优孰劣, 从而导致临床

的选择困难^[5]。

为解决这一问题, Haile等学者^[6]对经典的网络Meta分析(network meta-analysis, NMA)进行扩展, 提出了针对预测模型的NMA, 即(Multiple score comparison network meta-analysis, MSC), 该方法根据每一个原始队列研究中验证比较的模型分为不同的组别, 各组间选择同一个模型作为共同参照, 采用Lu等^[7]提出的两阶段Meta回归方法, 先计算每个组内直接比较Meta分析的结果, 再将第一阶段计算出的各组直接比较结果采用Meta回归的方法进行合并, 从而得到针对同一结局的所有模型之间预测准确度的比较结果。本文将介绍MSC方法的具体理论基础, 并通过案例对该方法的结果进行详细解读。

1 理论基础

1.1 预测模型研究的效应指标简介 预测模型准确度包含区分度和校准度两个方面。一个好的预测模型, 不仅要有很好的区分度, 同时还应具备良好的校准度。所谓区分度, 指模型能够把未来发病风险高低不同的人群正确的区分开来, 通过设置一定风险界值, 高于界值则判断为发病, 低于界值则判断为不发病, 从而正确区分个体是否会发生某结局事件。评价预测模型区分能力的指标, 最常用的就是大家非常熟悉的ROC曲线下面积(Area under Receiver Operating Characteristics curve, AUC), 也叫C统计量(C-statistics)。AUC越大, 说明预测模型的判别区分能力越好。一般AUC<0.60认为区分度较差, 0.60~0.75认为

基金项目: 国家自然科学基金(71673003, 72074011); 北京大学教育基金项目(2019YB39, 2020ZD001, 2020YBC02)

作者单位: ¹ 100050 北京, 首都医科大学附属北京友谊医院临床流行病学与循证医学研究室, 国家消化系统疾病临床医学研究中心; ² CB1 8RN 英国剑桥大学临床医学院初级医疗中心; ³ 264000 烟台, 烟台市烟台山医院骨关节科; 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 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

通讯作者: 孙凤, E-mail: sunfeng@bjmu.edu.cn

doi: 10.3969/j.issn.1674-4055.2021.02.02

模型有一定区分能力，>0.75认为区分能力较好。校准度则是定量评价一个疾病风险模型预测未来某个个体发生结局事件概率精确性的重要指标，反映了模型预测风险与实际发生风险的一致程度。校准度好，提示预测模型的精确性高；校准度差，则模型有可能高估或低估疾病的发生风险。通常用拟合优度检验（Hosmer-Lemeshow good of fit test）、Brier评分、校准图来进行判断^[8]。此外，还可从其他方面如基于确定的cut-off值来评价各模型的灵敏度、特异度、阴性预测值、阳性预测值、净重新分类指数（net reclassification improvement, NRI）和综合判别改善指数（integrated discrimination improvement, IDI）等^[9]。因此，在预测模型NMA中，效应指标通常为区分度指标如AUC，也可为校准度指标或其他指标如NRI等指标。本研究以不同预测模型间的AUC差值即ΔAUC作为效应指标，进行后续基础理论的介绍和案例解读。本文以下所提及的预测模型准确度默认为AUC。

1.2 预测模型NMA的证据结构 在经典干预性研究NMA中，我们通常按照纳入的每个随机对照试验（randomized controlled trial, RCT）的干预措施来进行分组，进而绘制网状证据图，以明确整个NMA中不同干预措施之间是否有直接比较，从而判断某两种干预措施之间的疗效比较是间接或混合比较的结果。而预测模型的外部验证研究多为基于某个队列人群来验证一个或多个已有预测模型的准确度，并没有传统意义上的干预措施。因此，在预测模型NMA中，我们将每个原始队列研究（本文中均指预测模型的外部验证研究）验证比较的预测模型暂定为NMA中的“干预措施”，将预测模型准确度（如AUC）作为合并的效应指标，只纳入验证比较的预测模型数目≥2个的队列研究（即类似于传统NMA纳入的两臂研究或多臂研究），最终所有队列研究中验证比较的预测模型的并集即为该NMA中所有的“干预措施”，进而根据各队列研究中验证比较的模型组合分为不同的组别，每个组别内的预测模型之间即认为存在直接比较，并可在此基础上进一步绘制网状证据图。

1.3 两阶段Meta回归模型 Haile等学者^[6]对经典的NMA进行了扩展，在Lu（2011）等^[7]的基础上提出了采用两阶段Meta回归模型对预测模型NMA进行统计分析。该方法根据每一个原始队列研究中验证比较的模型分为不同的组别，各组之间选择同一个模型作为共同参照，采用“两阶段”策略进行分析。具体理论基础如下：

设预测模型NMA证据网络中共纳入N

（i=1,2,……N）个队列研究，M（j=1,2, 3,……M）种预测模型（M通常≥3），N个队列研究根据验证模型组合的不同可分为G（g=1, 2,……G）个组别。令M_g表示第G_g个组别中预测模型（即“干预措施”）的数目，N_g表示第G_g个组别中纳入的队列研究的个数，则以只包含X、Y和Z三种预测模型共计44个队列研究的NMA为例，可整理为如下表1，即该NMA中N=44，M=3，G=3，其中N₁=20，N₂=16，N₃=8，M₁=M₂=2，M₃=3。

表1 X、Y和Z三种预测模型共计44个队列研究的NMA示例

组别 (G _g)	队列研究数目 (N _g)	预测模型数目 (M _g)	预测模型
G1	20	2	X, Y
G2	16	2	X, Z
G3	8	3	X, Y, Z

① I 阶段

选择预测模型X作为共同参照，则可分别计算出第1组和第2组每个队列i中预测模型Y和Z与共同参照预测模型X的ΔAUC及其方差，分别记为Δ_{ixY}、Var(Δ_{ixY})和Δ_{ixZ}、Var(Δ_{ixY})。而在第3组中，由于涉及X、Y和Z三种预测模型，类似于经典NMA中多臂研究之间存在一定的相关性，我们还需估计与共同参照预测模型X之间ΔAUC的协方差，即：

$$\Delta_{i3} = \begin{pmatrix} \Delta_{ixY} \\ \Delta_{ixZ} \end{pmatrix}$$

$$Var(\Delta_{i3}) = \begin{pmatrix} (\Delta_{ixY})^2 & Cov(\Delta_{ixY}, \Delta_{ixZ}) \\ Cov(\Delta_{ixY}, \Delta_{ixZ}) & (\Delta_{ixZ})^2 \end{pmatrix}$$

在以上研究集合数据的层面上，我们可基于传统Meta分析的倒方差法对每个组别内的N_g个队列研究的结果进行加权合并，从而获得每个组别内各预测模型与共同参照模型X之间的预测效果差值ΔAUC及其方差，以第3组为例，其加权合并公式为：

$$\Delta_3 = (\Delta_{XY}, \Delta_{XZ})^T = \frac{\sum_i W_{i3} \Delta_{i3}}{\sum_i W_{i3}}, W_{i3} = var(\Delta_{i3})^{-1}$$

② II 阶段

基于I阶段计算出的各组直接比较的效应值和方差，采用Meta回归的方法即加权最小二乘法进行合并，从而得到针对同一结局的各预测模型与共同参照模型预测准确度的混合比较结果。具体公式如下，其中y为I阶段计算出的各组直接比较的效应值，var(y)为I阶段计算出的各组直接比较效应值的方差，X为根据不同的预测模型比较所设置的哑变量，不同X哑变量对应的系数d_b，即为我们所要估计的各预测模型与共同参照模型预测准确度的混合比较结果。

$$y = Xd_b + \varepsilon, \varepsilon \sim N(0, W^{-1}), W^{-1} = var(y)$$

1.4 异质性、相似性和一致性的评估 异质性、相似性和一致性是NMA中非常关键的三个假设,只有数据满足这三个假设的情况下,我们才能进行下一步的NMA分析,NMA分析的结论才是可靠的。在预测模型NMA中,三个关键假设的含义和检验方法与经典NMA类似。①异质性:在I阶段直接比较Meta分析的计算过程中,各组别中各个队列研究之间的异质性因子 τ^2 的计算公式与传统Meta分析类似,可根据计算结果采用各组别特异的 τ^2_{group} 来进行后续分析,也可如经典NMA分析一般采用结构化方差模型即采用统一的异质性因子 τ^2_{pooled} 进行后续分析^[7]。Haile SR等通过慢性阻塞性肺疾病(COPD)患者死亡预测模型的NMA实例研究^[6]显示,在可获取多个队列研究的个体资料数据的情况下,使用统一的异质性因子 τ^2_{pooled} 进行后续分析更容易得到无偏的结果。②相似性:Haile SR等建议从临床角度选取一些可能会对预测模型准确度产生影响的关键因素,建议采用Meta回归的方法评估这些因素对模型预测准确度的影响,并对其中可产生影响的因素进行了单因素方差分析,若各因素在各组间分布均衡($P>0.05$),即可认为符合相似性假设;反之,若单因素方差分析显示 P 值 <0.05 ,则可能该NMA不符合相似性假设。③一致性:可采用Q统计量(即直接比较和间接比较的残差平方和)对整个NMA的进行全局一致性检验,Q统计量服从自由度为 $N-K+1$ 的卡方分布,其中 N 为各组别中比较的数目之和, K 为NMA中预测模型的总数,若 $P>0.05$,则可认为全局满足一致性假设^[7];可采用节点拆分法(Node-splitting method)^[10]进行局部的一致性检验,以评估各预测模型之间的直接比较与间接比较的结果是否有显著性差异。

2 案例解读

以COPD患者的死亡预测模型研究为例,基于COPD国际队列联盟(Cohorts Collaborative International Assessment consortium, 3CIA)的个体资料数据^[11,12],进行NMA分析,以获得各预测模型之间的预测准确度比较结果。该NMA共纳入24个队列研究,包含15 762例COPD患者,共计随访42 203人·年,其中1871例COPD患者发生了死亡。各队列在研究地区、样本量、结局事件数及研究对象的重要临床特征上存在一定的差异:如COPD患者的平均年龄范围跨度较大,在58~72岁;平均一秒用力呼气容积(Forced expiratory volume in one second, FEV1)的范围在30%~70%之间;改良呼吸困难评估(Modified Medical Research Council)量表mMRC评分平均值的范围在1.0~2.8之间;6分钟步行试验的平均距

离在218~487米。各队列的随访时间不一,本次NMA分析以最短随访时间即3年的死亡作为研究结局。24个队列研究涉及GOLD、GOLD(2011)、ADO、BODE、BODEupd、eBODE、BODEx、DOSE、SAFE和BAED共计10个预测模型^[13-18]。根据验证比较的预测模型组合不同可将24个队列研究分为6组,各组分别包含4、5、3、1、7和4个研究,各组比较的预测模型、总样本量、死亡事件数等具体情况详见表2。

应用R软件ggplot2软件包绘制网状证据图(图1),每个组别中验证比较的各预测模型之间用直线连接,不同组别之间的线条用不同的颜色区分,6个组别共计有6条颜色各不相同的连线。例如:粉色线条代表组别1,该线条连接了GOLD和ADO,即表示Copenhagen、HUNT、Japan和SEPOC四个队列研究验证比较了GOLD和ADO两个预测模型的准确度;紫色线条代表组别6,该线条连接了GOLD、GOLD(2011)、ADO、BODE、BODEupd、eBODE、BODEx、DOSE和BAED九个预测模型,即表示La Princesa Madrid、Requena II、Tenerife和Terrassa II四个队列研究均验证比较了前述9个预测模型的准确度。该网状证据图中不同线条的粗细与各组别总计的死亡事件数成正比,即采用死亡事件数加权,死亡事件数越多,则线条越粗。当然,该作图较为灵活,线条的粗细也可采用各组别总计的样本量进行加权。由图1可见,该NMA涉及10个预测模型之间的共计45个比较的结果,其中包含41个直接比较,另外SAFE与eBODE、BODEx、DOSE和BAED四个模型之间尚无直接比较结果,只有间接比较结果。

关于NMA三个关键假设的评估:①异质性:表2展示了各组别中各个队列研究之间的异质性因子 τ 的数值,由表2可见各组的异质性因子大小不一,在本研究的NMA分析中,我们采用了结构化方差模型,即各组之间采用统一的异质性因子 $\tau^2_{\text{pooled}}=0.00015$,提示该NMA的异质性较小,可认为满足同质性假设;②相似性:研究者从临床角度选取了7个可能会对预测模型准确度产生影响的关键因素,包括平均年龄及方差、样本量、死亡比例、肺部阻塞的程度FEV1%的预测值及其方差和运动后肺功能,采用Meta回归的方法评估了这些因素对模型预测准确度的影响,并对其中可产生影响的5个因素(肺部阻塞的程度FEV1%的预测值及其方差、平均年龄、死亡比例和运动后肺功能)进行了单因素方差分析,结果显示以上5个关键因素在各组间分布均衡,即可认为符合相似性假设;③一致性:采用与传

统NMA类似的一致性检验方法，全局检验结果显示Q统计量=22.1, $P=0.14$ ；节点拆分法（Node-splitting method）结果显示各预测模型之间的直接比较与间接比较结果无显著性差异，故可认为该NMA符合一致性假设。

在满足异质性、相似性和一致性三个关键假设的基础上，选取GOLD预测模型作为共同参照，利用24个队列研究中每个预测模型与共同参照GOLD模型之间的 ΔAUC 及其方差和协方差矩阵作为原始数据，其中根据3CIA联合队列的个体资料数据进行协方差的计算，采用随机效应模型的两阶段Meta回归进行统计分析，具体代码可参考文献附件^[12]，运行结果见表3。

由表3可见，前6行为I阶段每个组别直接比较的Meta分析结果，第7行为II阶段混合比较的结果。GOLD模型作为共同参照，其预测COPD患者3年死亡率的准确度即AUC的范围为0.481至0.731，中位数为0.614（四分位间距：0.587, 0.641）。由混合比较的结果可见，所有模型中ADO模型对COPD患者3年死亡率的预测准确度

最高，其准确度显著高于GOLD模型，与GOLD模型的差值 ΔAUC 为+0.083（95%CI: 0.069, 0.097）；其次为BODEupd、eBODE和BODE模型，与GOLD模型预测准确度的差值 ΔAUC 分别为+0.072（95%CI: 0.051, 0.093）、+0.069（95%CI: 0.044, 0.093）和+0.064（95%CI: 0.045, 0.082）；SAFE、BODE_x、DOSE和GOLD(2011)四个模型的表现也显著优于GOLD模型， ΔAUC 的范围为+0.014至+0.052；BEAD模型与GOLD模型的预测准确度之间无显著性差异， ΔAUC 为+0.016（95%CI: -0.007, 0.038）。在满足一致性假设的情况下，根据表3中II阶段混合比较结果中的9个基本参数，我们则可以推算出NMA中的所有功能参数，即9个模型中任意两个预测模型之间的预测准确度差值 ΔAUC 及其95%CI，因篇幅有限不再展示相应结果，有兴趣的读者可自行去推导计算。

3 讨论

临床预测模型或预后评分近年来在临床实践中的应用越来越广泛，目前在很多疾病领域针对

表2 COPD死亡预测模型NMA研究分组情况

组别	预测模型	队列研究	样本量	死亡数	Q统计量	自由度df	异质性因子 τ
1	GOLD-ADO	Copenhagen, HUNT, Japan, SEPOC	4323	378	2.8	3	0
2	GOLD-ADO-BODE-BODEupd	Barmelweid, Basque, Galdakao, Pamplona, Zaragoza I	1208	215	15.5	12	0.014
3	GOLD-GOLD(2011)-ADO-BODE-BODEupd	Mar de Plata Argentina, PACECOPD, Son Espases Mallorca	556	61	10.9	8	0.025
4	GOLD-GOLD(2011)-ADO-BODE-BODEupd-SAFE	COPDgene	4484	337	7.46E-29	0	NA
5	GOLD-GOLD(2011)-ADO-BODE _x -DOSE-BAED	Genkols, ICE COLD ERIC, Initiatives BPCO, Sevilla, Terrassa I, Terrassa III, Zaragoza II	4346	722	48.1	30	0.011
6	GOLD-GOLD(2011)-ADO-BODE-BODEupd-eBODE-BODE _x -DOSE-BAED	La Princesa Madrid, Requena II, Tenerife, Terrassa II	845	125	34.5	24	0.014

备注：GOLD: Global initiative for chronic Obstructive Lung Disease; BODE: Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODEupd: BODE updated; ADO: Age, Dyspnoea, airflow Obstruction; e-BODE: severe acute exacerbation of COPD plus BODE; BODE_x: Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE: Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE: Saint George’s Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D: Body-mass index, Acute Exacerbations, Dyspnoea; NA: not available.

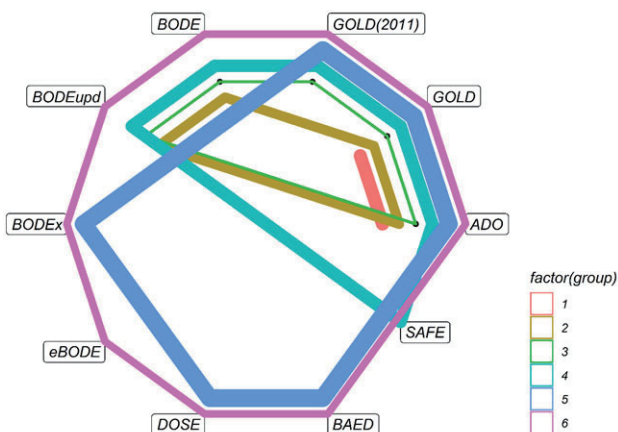


图1 COPD死亡预测模型NMA证据结构图

同一个研究结局都已发表了为数不少的预测模型或预后评分，但大多数模型缺乏外部验证研究，且缺乏各预测模型之间准确度的相互比较结果，难以确定最佳预测模型，从而导致临床实践中的选择困难，出现大多数预测模型长期处于“多数被建立，少数被验证，极少被应用”的情况。Haile SR等^[6]学者提出的基于两阶段Meta回归的预测模型NMA解决了这一问题，该方法将队列研究中验证的多个预测模型作为“干预措施”，借用经典NMA的间接比较理论，获得针对同一结局的各预测模型准确度（如AUC）两两比较的结果，从而确定最佳预测模型，以指导临床实践，辅助临床决策，具有重要的临床意义。

表3 两阶段Meta回归的MSC结果（以GOLD作为共同参照，效应指标为ΔAUC）

阶段	组别	ADO	BODEupd	eBODE	BODE	SAFE	BODEx	DOSE	BEAD	GOLD(2011)
I	1	0.097 (0.07, 0.123)								
I	2	0.098 (0.057, 0.139)	0.124 (0.078, 0.17)		0.098 (0.059, 0.137)					
I	3	0.044 (-0.03, 0.117)	0.023 (-0.054, 0.099)		0.019 (-0.054, 0.091)					-0.011 (-0.053, 0.03)
I	4	0.042 (0.01, 0.074)	0.043 (0.008, 0.078)		0.049 (0.017, 0.081)	0.037 (0.005, 0.069)				-0.008 (-0.038, 0.022)
I	5	0.099 (0.076, 0.123)					0.056 (0.035, 0.076)	0.036 (0.015, 0.057)	0.032 (0.005, 0.058)	0.028 (0.008, 0.047)
I	6	0.076 (0.027, 0.126)	0.043 (-0.006, 0.092)	0.048 (0.004, 0.093)	0.043 (0.001, 0.085)		0.030 (-0.015, 0.074)	0.021 (-0.023, 0.065)	-0.017 (-0.079, 0.045)	0.008 (-0.031, 0.047)
II		0.083 (0.069, 0.097)	0.072 (0.051, 0.093)	0.069 (0.044, 0.093)	0.064 (0.045, 0.082)	0.052 (0.022, 0.082)	0.045 (0.029, 0.061)	0.027 (0.010, 0.045)	0.016 (-0.007, 0.038)	0.014 (0.001, 0.028)

备注：GOLD: Global initiative for chronic Obstructive Lung Disease; BODE: Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODEupd: BODE updated; ADO: Age, Dyspnoea, airflow Obstruction; e-BODE: severe acute exacerbation of COPD plus BODE; BODEx: Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE: Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE: Saint George’s Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D: Body-mass index, Acute Exacerbations, Dyspnoea; MSC: multiple score comparison, 多评分比较; AUC: area under the receiver operating characteristics curve ROC曲线下面积; CI: confidence interval, 置信区间

预测模型NMA是在经典NMA基础上的扩展，但又与经典NMA有着不同之处，主要表现在以下三个方面。首先，两者的效应指标不同，经典NMA的效应指标多为logOR、logRR或均数差（mean difference, MD）等，来表示不同干预措施之间的疗效差别；而预测模型NMA的效应指标一般为ΔAUC，即表示不同预测模型之间准确度的差值。其次，经典NMA纳入的RCT研究中试验组和对照组的研究对象通常是完全独立的；而预测模型NMA纳入的原始研究则通常是基于同一个队列人群同时验证两个或多个预测模型，换言之，不同预测模型准确度的结果所基于的研究对象是完全重复的。第三，考虑到可行性及统计把握度和样本量的问题，经典NMA纳入的RCT研究大多数为2臂研究，3臂及以上的研究较少；而预测模型NMA纳入的队列研究通常会基于该队列同时验证多个不同的预测模型，验证预测模型个数≥4的队列研究很常见，即在预测模型NMA中“多臂研究”更为常见。

鉴于预测模型的外部验证研究中研究对象的完全重复利用及“多臂研究”的情况，在进行预测模型NMA时需要特别关注各预测模型准确度结果之间的相关性，即前述方差-协方差矩阵的计算是关键。由于各个队列人群之间人口学及临床特征往往存在一定的差异，因而解决这一问题的最佳办法是利用多个大型队列研究整合在一起的个体资料数据进行相关系数的计算，如同上述COPD患者死亡预测模型的NMA研究一样，利用3CIA整合队列^[11]的个体资料数据进行了各预测模型ΔAUC之间相关系数的计算，从而获得各“多

臂研究”的方差-协方差矩阵，进一步通过两阶段Meta回归模型进行混合比较的计算。当然，若无法获取到整合队列的个体资料数据，则可采用经典NMA中类似的假设，如假定相关系数R=0.4或0.5来进行方差-协方差矩阵的计算。此时，建议采用不同的相关系数分别进行敏感性分析，通过判断不同相关系数假设下的结果是否一致来衡量结果的稳定性。

此外，本研究中只用ΔAUC作为效应指标进行了NMA分析。事实上，我们还可用其他指标如NRI或校准度指标等作为效应指标进行分析，以从不同方面对各预测模型的优劣进行评价，帮助我们筛选出最佳预测模型，从而进一步指导临床实践^[6]。

4 结语

本文针对预测模型NMA实现的基础理论进行了介绍，并通过案例进行了详细解读。该方法以所有预测模型的外部验证研究为基础，选择同一个预测模型作为共同参照，通过大规模个体资料数据来计算各预测模型结果之间的相关性，进而采用两阶段Meta回归模型，可得到针对同一结局的任意两个模型之间预测准确度的比较结果，对最终选择最佳预测模型应用到临床实践以更好地指导临床决策有较大意义。然而，该方法目前还不太成熟，尚处在发展阶段，Haile SR等开发的mscpredmodel R package目前只能针对个体资料数据的预测模型MSC进行分析，且关于发表偏倚、一致性假设等各方面的方法学还在摸索中。笔者于Web of Science数据库中检索，迄今为止该论文^[11]被引用4次，均为COPD死亡预测模型的

相关研究，目前未见到其他疾病领域预测模型 NMA 的应用实例，一方面可能与个体资料的队列数据不易获取无法计算各预测模型结果之间的相关性有关，另一方面可能与该方法需要事先进行较为复杂的数据预处理，尤其是方差-协方差的计算有关，未来可加快各疾病领域的数据共享和队列整合，并开发相应的程序包来更高效快捷地进行数据预处理，以促进该方法的推广应用，从而更好地辅助临床实践。

参考文献

[1] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation[J]. *Eur Heart J*, 2014, 35(29):1925-31.

[2] Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration[J]. *Ann Intern Med*, 2015, 162(1):W1-W73.

[3] Grady D, Berkowitz SA. Why is a good clinical prediction rule so hard to find [J]? *Arch Intern Med*, 2011, 171(19):1701-2.

[4] Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review[J]. *BMJ*, 2016, 353:i2416.

[5] Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting[J]. *BMC Med Res Methodol*, 2014, 14:40.

[6] Haile SR, Guerra B, Soriano JB, et al. Multiple Score Comparison: a network meta-analysis approach to comparison and external validation of prognostic scores[J]. *BMC Med Res Methodol*, 2017, 17(1):172.

[7] Lu G, Welton NJ, Higgins J, et al. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach[J]. *Res Synth Methods*, 2011, 2:43-60.

[8] Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of

Clinical Prediction Models: Users' Guides to the Medical Literature [J]. *JAMA*, 2017, 318(14):1377-84.

[9] 于莉莉, 武颂文, 夏结来. 利用净重新分类指数与整体鉴别指数评价一种新危险因素的补充预测能力[J]. *中国卫生统计*, 2017, 34(5):761-3.

[10] Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis[J]. *Stat Med*, 2010, 29:932-44.

[11] Soriano JB, Lamprecht B, Ramirez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data[J]. *Lancet Respir Med*, 2015, 3:443-50.

[12] Guerra B, Haile SR, Lamprecht B, et al. Large-scale external validation and comparison of prognostic models: an application to chronic obstructive pulmonary disease[J]. *BMC Med*, 2018, 16(1):33.

[13] Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease[J]. *N Engl J Med*, 2004, 350:1005-12.

[14] Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE index[J]. *Am J Respir Crit Care Med*, 2009, 180:1189-95.

[15] Azarisman MS, Fauzi MA, Faizal MP, et al. The SAFE (SGRQ score, airflow limitation and exercise tolerance) index: a new composite score for the stratification of severity in chronic obstructive pulmonary disease[J]. *Postgrad Med J*, 2007, 83:492-7.

[16] Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: the B-AE-D indices[J]. *Eur Respir J*, 2016, 47:1635-44.

[17] Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index[J]. *Lancet*, 2009, 374(9691):704-11.

[18] Soler-Cataluña JJ, Martínez-García MA, Sánchez LS, et al. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients[J]. *Respir Med*, 2009, 103:692-9.

本文编辑：孙竹

欢迎投稿， 欢迎订阅！

中国科技核心期刊， 科技论文统计源期刊

E-mail: ebcvm_cj@126.com; 联系电话: 010-84008479

网址: www.ebcvm.cn/www.ebcvmcj.com/www.ebcvmcj.org