·论著·方法学研究·

预测模型系统评价的方法与实践：基于慢性乙型肝炎患者肝癌发生风险预测模型的案例分析

刘昊睿[1]，周业胜[1]，李彧格[1]，于　爽[1]，孙　凤[2]，武珊珊[1]

1. 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（北京 100050）
2. 北京大学公共卫生学院流行病与卫生统计学系（北京 100191）

【摘要】本文旨在介绍预测模型系统综述与 Meta 分析的研究方法，对其六个关键步骤（研究问题的提出、检索策略的制定、文献筛选与偏倚风险评价、数据提取和准备、证据合并及 Meta 分析、结果解释及报告）、常用的工具（PROBAST 偏倚风险评估工具以及 TRIPOD–SRMA 报告规范）进行详细阐述。并以慢性乙型肝炎相关肝细胞癌风险预测模型为例，对 REACH–B 模型的 3 年预测效能进行 Meta 分析，以为临床预测模型的证据合成提供系统性解决方案和技术路径参考。

【关键词】预测模型；系统综述；Meta 分析；偏倚风险评估；肝细胞癌；慢性乙型肝炎

【中图分类号】R 318.14　　【文献标识码】A

Systematic review and Meta-analysis of prediction models: a case study on the risk prediction model for hepatocellular carcinoma in patients with chronic hepatitis B

LIU Haorui[1], ZHOU Yesheng[1], LI Yuge[1], YU Shuang[1], SUN Feng[2], WU Shanshan[1]

1. Department of Clinical Epidemiology and Evidence-Based Medicine Research, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

2. Department of Epidemiology and Health Statistics, School of Public Health, Peking University, Beijing 100191, China

Corresponding author: WU Shanshan, Email: shanshanwu@ccmu.edu.cn

【Abstract】This article aims to introduce the research methodology for systematic reviews and Meta-analyses of prediction models, detailing its six key steps (research question formulation, search strategy development, literature screening and risk of bias assessment, data extraction and preparation, evidence synthesis and Meta-analysis, and result interpretation and reporting), as well as commonly used tools (PROBAST risk of bias assessment tool and TRIPOD-SRMA reporting guidelines). This article uses the chronic hepatitis B-related hepatocellular carcinoma risk prediction model as an example to conduct a Meta-analysis of the 3-year predictive performance of the REACH-B model. This provides a systematic solution and technical approach for evidence synthesis in clinical prediction models.

【Keywords】Prediction models; Systematic review; Meta-analysis; Risk of bias assessment; Hepatocellular carcinoma; Chronic hepatitis B

DOI: 10.12173/j.issn.1004-5511.202503192

基金项目：国家自然科学基金青年科学基金项目（82204119）

通信作者：武珊珊，博士，研究员，教授，博士研究生导师，Email: shanshanwu@ccmu.edu.cn

临床预测模型又称临床预测规则，是指利用医学征兆、症状或其他临床发现预测特定疾病或结局发生的概率，包括诊断模型和预后模型[1]。按照研究目的，预测模型类研究可分为模型开发研究、模型验证研究和开发验证同时进行的研究，其中模型验证又可分为内、外部验证，取决于与模型开发使用的数据集是否相同。内部验证为使用相同的数据集验证模型，但常因过度拟合而高估模型性能，其核心目的是评估模型过拟合风险和优化参数选择，通过重抽样技术检验模型在训练数据中的稳定性；外部验证为使用模型开发以外的数据集验证模型，其核心目标是检验模型泛化能力和明确临床适用边界。由于开发与验证数据集存在异质性（如人群特征、测量方法差异），外部验证结果通常较内部验证保守，但更能反映真实场景下的实际效用。在临床领域，对于同一疾病和同一结局，不同预测模型往往针对不同的人群和情境建立。针对慢性阻塞性肺疾病（COPD）多种结局，已有超过 408 个预测模型被开发[2]。例如，ADO指数（age，dyspnea，and airflow obstruction）基于年龄、呼吸困难程度和气流受限程度预测 3 年死亡率[3]，而 B-AE-D 指数（BMI，acute exacerbations，dyspnea）则结合体重指数、既往急性加重史与呼吸困难评分，用于预测 COPD 患者 2 年死亡风险[4]。

尽管模型数量众多，但受变量选择、建模方法、验证方式和风险结局定义不一致，导致不同模型间适用性和稳定性存在显著异质性[2]。此外，多数模型缺乏外部验证，约 75% 的模型未进行内部验证，严重制约其在真实临床中的应用推广。这种“模型爆炸”与“证据碎片化”的现象导致临床医生面临选择困境[2]。为此，需根据研究目的识别、评估、合成证据，并形成证据概要表，进行针对预测模型的系统综述和 Meta 分析，总结预测模型表现证据，并进行证据分级，为循证医学和临床指南制订提供支持[5-6]。本文重点介绍针对同一个预测模型多个外部验证研究的系统综述与Meta分析的制作步骤，包括研究问题的提出、检索策略的制定、文献筛选与偏倚风险评价、数据提取和准备、证据合并及 Meta 分析、结果解释及报告 6 个步骤，并以慢性乙型肝炎相关肝细胞癌（hepatocellular carcinoma，HCC）风险预测模型为例，进行 Meta 分析，旨在为临床预测模型的证据合成提供系统性解决方案和技术路径参考。

1 研究问题的提出

提出合适的问题是系统综述和 Meta 分析的基础，也是最重要的环节之一。选题一般来自临床和科研实践，是对目前仍有争议的问题进行研究。在预测模型的系统综述与 Meta 分析研究中，研究问题的提出应依据 PICOTS 原则：①研究的目标人群（population），指在具备哪些特征的研究对象中评价预测模型的效果；②研究的指示模型（index model），指研究者特别关注的某个或某几个预测模型；③研究的比较模型（comparator model），类似于传统研究中的对照组，该指标仅在评价多个预测模型表现时才使用，若只针对同一个预测模型的外部验证研究，对于该指标可不做要求；在预测模型的系统综述与 Meta 分析研究中，指示模型与比较模型通常区分并不明确，可以互为指示与比较模型；④研究的结局指标（outcome），指指示模型与比较模型预测的结局事件；⑤时间（timing），即预测的结局事件发生的时间或未来模型使用的时间点，该指标在诊断和预后模型中应用不同，诊断模型关注“什么时候可以使用待评价预测模型”，预后模型关注“在哪个期间的临床结局才被预测”；⑥研究所处的场景（setting），通常指预测模型开发或未来应用的场景或人群，如社区一般人群的预测或临床专病人群的预测等。

2 检索策略的制定

系统综述与 Meta 分析的可靠结论来自于对已有研究做出的综合评价与分析，因此需尽可能完整检索与提出问题相关的文献，在本研究中（仅针对预测模型外部验证研究的系统综述与 Meta 分析）主要指检索针对特定研究人群、场景和结局事件的某个模型的全部模型验证研究，模型开发研究的结果通常不纳入。首先，根据 PICOTS 制定检索策略，确定检索词和需要检索的数据库。常用检索词为疾病名称、结局事件名称、指示模型与比较模型名称以及预测模型相关研究术语，应根据具体情况调整检索策略。如相同疾病不同命名、同一结局事件不同命名以及不同的简写方式等，以保证能检索所有文献。检索范围包括各种电子数据库、期刊、会议论文及其他未发表资

料库等，对于一些原始研究的参考文献也需检索，甚至已经发表的同类型系统综述与 Meta 分析，其引用的文献也可以参考。附表 1 为 Geersing 等 [7] 建议的灵敏度和特异度均较高的预测模型相关研究的检索策略（以 PubMed 检索为例），仅供参考。研究者可根据自身需求调整和优化检索策略，提高预测模型研究的查全率和查准率。

3 文献筛选与偏倚风险评价

检索获得的题录一般借助于文献管理软件（如 NoteExpress 或 EndNote 等）进行管理，并根据事先制定的文献纳入和排除标准进行筛选。纳入排除标准均依据 PICOTS 原则制定，也可在此基础上进一步限制，如针对特定种族或地区、年龄、有无合并症的研究人群进行限制。与传统系统综述和 Meta 分析一样，在制定好纳入和排除标准后，一般先阅读题目和摘要筛选文献，再阅读全文复筛。筛选过程须两人进行，遇到不一致时，可以商量或寻求第三者解决。

由于原始研究的潜在偏倚影响研究结果，因此进行偏倚风险评估是非常重要的。在预测模型的系统综述和 Meta 分析中，对原始预测模型进行偏倚风险评价的工具为 PROBAST（Prediction Model Risk of Bias Assessment Tool），官网为 https://www.probast.org/。PROBAST 工具通过系统评估预测模型研究的四个关键要素（研究对象、预测变量、结局和统计分析）共 20 个问题，评价诊断和预后类预测模型开发 / 验证研究中的潜在偏倚风险：首先评估各要素中相关问题（如人群代表性、预测因子测量时间、缺失数据处理方法等）的偏倚风险情况，若任一关键要素中存在高风险，则整体判定为高风险；适用性评价则旨在独立评估研究设计与系统综述目标的一致性（如人群特征匹配度）；最终结果用于敏感性分析和证据分级 [8]。当然，PROBAST 不只用于预测模型的系统综述和 Meta 分析，也可作为预测模型原始研究的通用评价工具。

4 数据提取和准备

预测模型的系统综述和 Meta 分析中，以预测模型研究系统评价的关键评估和数据提取清单（Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies，CHARMS）作为参考 [9]，原始研究提取信息主要包括：①研究基本信息，如作者、发表年份等；②研究人群基本特征，如种族、年龄、疾病诊断标准、有无接受治疗、某种合并症的比例等；③研究中验证预测模型的信息，包括纳入的预测因子（分类或连续变量等）、模型公式（系数、截距等）、预测结局事件、预测结局发生的时间、样本量及结局事件发生数等；④模型诊断或预测效能相关指标，包括区分度（discrimination）、校准度（calibration）等；其中区分度指标通常指受试者操作特征（ROC）曲线及其曲线下面积（AUC）或一致性指数（C-index），校准度指标通常指 total O:E ratio，即结局事件实际发生数（observed events）和模型预测结局事件发生数（expected events）的比值 [9]；需要注意的是，仅有上述区分度和校准度相关指标的点估计值是无法进行后续 Meta 分析的，还需提取上述两方面指标的区间估计值，即 95% 置信区间（confidence interval，CI）。

以上 4 类需提取的信息中，模型诊断或预测效能相关指标的点估计值和区间估计值是后续分析的核心。在原始研究中，预测模型的区分度和校准度指标一般以表格形式列出，可直接获取，但也有部分研究需通过简单计算获得。有时，符合纳入标准的文献中缺乏具体信息，尤其是校准度指标缺乏较多（即大多数研究中未报告模型预测结局事件发生数，因而无法计算 total O:E ratio），因此需与原文作者联系获取，在多次联系无果的情况下，该文章在进行校准度的 Meta 分析时可以剔除。此外，需要注意的是，预测模型验证类研究常以校准曲线图的形式展示校准度，该图通常将所有研究对象分为 5~10 组呈现，在这种情况下可通过联系原文作者获取每个组别的模型预测结局事件发生数，进而获得总的预测结局事件数。此外，在针对某些预测因子有多种定义和测量方法的情况下，研究者也可根据是否进行亚组分析，提取预测因子的定义与测量方法等指标。

5 证据合并与Meta分析

与传统系统综述和 Meta 分析类似，如针对某个已开发的模型有足够（通常 ≥ 5 个）的模型验证研究结果 [10]，预测模型的系统综述和

Meta 分析也可采用定量的方法对纳入的多个模型验证研究的区分度和校准度数据进行合并，其分析本质与传统 Meta 分析类似，即求解区分度或校准度的加权平均值。由于预测模型的研究通常为观察性研究，研究间异质性较大，故一般采用随机效应模型进行统计合并。通常使用 Hartung-Knapp-Sidik-Jonkman（HKSJ）方法进行随机效应 Meta 分析，将模型区分度和校准度估计值分别汇总为加权平均值，其中权重由研究的标准误差和样本数量决定[10]。在计算平均性能的 95%CI 时，建议采用最大似然估计和 HKSJ 方法的随机效应模型。

预测模型研究间的异质性主要来自三个方面：①研究对象的差异，如研究对象的来源、人口学特征、疾病严重程度、病程、合并症、治疗方案等方面存在一定差异；②研究设计的差异，如在预测因素选择及形式、结局事件的诊断标准、随访时间及测量方法等研究设计的多个要素方面存在不同；③统计分析方法或选择性报告等方面的差异，如回归分析方法、偏倚风险、研究规模等方面存在一定差异。因此，在进行异质性评估时，通常采用 HKSJ 方法估计合并后区分度和校准度的预测区间。预测区间反映模型在新人群中的预期表现，不仅考虑了样本数据中的变异性，还考虑了未来观测值本身的不确定性，因此通常宽于 CI。通过计算预测区间，可估计模型在新研究中的潜在性能表现。若预测区间显著宽于 CI，则表明原始研究间存在显著的异质性，应采用亚组分析或 Meta 回归的方法进一步解释、探索异质性的来源及大小。在个体参与者数据可获取时，也可通过个体受试者数据的 Meta 分析检验亚组因素的效应修饰作用，其精确性与可信度更高，还可避免生态偏倚导致的错误。

此外，与其他系统综述和 Meta 分析类似，对于偏倚风险较低或较高的研究，研究者应行敏感性分析，如可仅纳入偏倚风险较低的研究再次进行 Meta 分析，并与之前 Meta 分析的结果进行比较，避免因个别低质量的研究影响 Meta 分析结果的稳定性和可靠性。

6 结果解释及报告

预测模型的系统综述和 Meta 分析在结果解释时应考虑：①是否提供了关于预测模型 PICOTS 要素和性能的所有必要信息；②预测模型的汇总校准度和区分度如何；③这些模型中每个模型的汇总证据在特定人群和特定结果方面的确定性如何。研究者可使用 GRADE 工具评估预测模型系统评价证据的质量，但用于预测模型的 GRADE 专用工具正在开发中，可借鉴和修改采用预后研究和预测因子研究的 GRADE 工具（如将关联强度测量更改为模型的性能指标，预后因素的探索和验证更改为模型的开发和验证）[11]。

在预测模型报告方面，2015 年已有学者制定了《个体预后或诊断多因素预测模型报告规范》（Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis，TRIPOD），该报告规范从标题、摘要、前言、方法、结果、讨论以及其它七个方面，提出了 22 个条目，并一一进行了充分的说明和举例，为预测模型的开发、验证和更新提供了报告规范，在进行预测模型的系统综述和 Meta 分析制作时也可进行参考，从而提高研究质量[12]。为了提高预测模型系统综述和 Meta 分析研究质量，2019 年 Cochrane 预后研究方法学组着手编制报告指南，经多轮德尔菲调查和会议讨论，并借鉴 PRISMA 2020 和其他报告指南，形成了预测模型系统综述和 Meta 分析的报告清单，并命名为 TRIPOD-SRMA（TRIPOD for Systematic Reviews and Meta-analysis）[13]。该报告清单分为标题、摘要、前言、方法、结果、讨论和其他信息 7 部分，共包含 26 个条目、34 个次级条目，适用于所有类型的预测模型研究的系统综述（包括使用或未使用 Meta 分析），此处“预测模型”特指根据若干预测因素预测单一的结局或者个体罹患风险的多变量模型。这类综述通常用于：①确定特定临床领域内所有预测模型；②确定特定总体的所有预测模型；③确定特定结局的所有预测模型；④针对特定预测模型，概括其预测性能，或比较多个竞争模型的预测性能。

7 案例介绍

7.1 案例背景

我国慢性乙型肝炎患病率高，并发症重。在我国肝硬化和肝癌患者中，由乙肝引起的比例分别为 60% 和 80%[14]。最新全国体检数据显示我国农村 21~49 岁男性人群中乙型肝炎表面抗原阳性率为 6%[15]。据此推算，全国约有 2 500 万人为慢

性乙肝患者；按照每年 1.6%~4.0% 的发病率计算，每年约有 40~100 万患者可发展为代偿性肝硬化。抗病毒治疗可降低乙肝相关并发症的发生率和病死率，但即使经过有效的抗病毒治疗仍有部分患者会出现疾病进展，包括门脉高压相关并发症和肝癌并导致死亡。因此，实现临床终点事件的精准预测并加强干预是降低病死率的关键措施。

目前国内外已有较多预测慢性乙肝患者临床终点事件的风险预测模型，主要以 HCC 为预测结局，但被临床广泛应用的模型较为少见。很多预测模型在开发过程中存在一定的方法学缺陷，导致模型的精确度和外推性不高，从而出现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。因此，急需系统检索慢性乙肝患者 HCC 发生风险的预测模型研究，并对现有证据进行合并，评估并比较各 HCC 预测模型的效果，为临床慢性乙肝患者 HCC 的精准预测提供高级别的循证医学证据 [16]。

7.2 案例解读与整体评价

系统检索 PubMed、Embase 和 The Cochrane Library 数据库中公开发表的慢性乙肝患者 HCC 发生风险的预测模型研究，包括模型开发及模型外部验证研究。经一系列筛选，共纳入 30 篇研究 [17-46]，其中 1 篇 [17] 为仅进行模型开发的研究，11 篇 [18-28] 为进行模型开发 + 外部验证的研究，18 篇 [29-46] 为仅进行模型外部验证的研究。共 12 篇 [17-28] 文献涉及模型开发，包含 14 个模型，其预测模型构建的基本特征详见表 1 和附表 2。

14 个预测模型的建立均采用队列研究，其中 13（93%）个模型基于亚洲人群构建，仅 PAGE-B 模型 [23] 是基于多个欧洲国家的高加索人种构建。各模型构建人群样本量范围为 192~23 851，中位数为 1 035，结局事件发生中位数为 56，其中 CAMD 模型 [26] 构建人群样本量超过 2 万，结局事件发生数目最多（596 例 HCC）。

关于各模型慢性乙肝患者临床终点事件的风险预测因素，大致可分为 3 类：①传统流行病学危险因素，包括年龄、性别、HCC 家族史、饮酒、糖尿病合并症、肝硬化等信息；②临床检测指标，包括丙氨酸氨基转移酶、白蛋白、血小板、总胆红素、甲胎蛋白、肝硬度值、乙型肝炎病毒 e 抗原、乙型肝炎病毒 DNA 等指标；③遗传易感性检测指标，包括核心启动子突变等指标，见附表 2。

在模型构建方法学层面，大部分模型（13/14，93%）采用了 COX 比例风险回归来构建预测模型，少数（1/14，7%）采用了 Logistic 回归方法，未见其他统计学方法的使用。所有模型均采用 AUC 或 C-index 指标进行模型区分度的评价，但近半数（6/14，43%）模型未进行校准度的评价和报告，限制了后续校准度指标的证据合并与 Meta 分析。此外，13 个基于 COX 比例风险回归构建的预测模型中，50% 的模型（如 CU-HCC[19]、GAG-HCC[17]、LSM-HCC[21] 等）未报各预测因素的系数及基础无病生存率（disease free probability），1 个基于 Logistic 回归构建的预测模型也未报告截距项和（或）各预测因素的系数，使这些模型的临床应用受到一定的限制，也无法基于其他外部队列对这些模型的校准度进行外部验证。由此可见规范预测模型的产生过程及报告方法，不仅有助于提升预测模型本身的质量，也能为后续预测模型的 Meta 分析及广泛验证提供可能。

在模型外部验证方面，14 个模型中有 11 个（79%）至少经过 1 次外部验证。值得注意的是，各模型构建 / 验证人群的关键特征（是否接受抗病毒治疗与肝硬化状态）差异性较大。GAG-HCC[17]、NGM-HCC[18] 及 REACH-B[20] 模型是基于未进行抗病毒治疗的慢性乙肝患者，mREACH-B[22]、PAGE-B[23]、mPAGE-B[25]、CAMD[26]、AASL-HCC[27]、REAL-B[28] 等模型是基于抗病毒治疗的慢性乙肝患者，CU-HCC[19]、LSM-HCC[21] 及 RWS-HCC[24] 模型则同时包含了上述两类患者，抗病毒治疗患者的比例范围为 15%~36%。在肝硬化方面，REACH-B[20] 模型是唯一针对非肝硬化患者构建的预测模型，其余模型的构建人群则同时包含了肝硬化和非肝硬化的慢性乙肝患者，肝硬化患者的比例范围为 15%~47%。由此可见，各模型开发时目标人群的关键特征差异较大，因而在后期不同临床场景下不同目标人群中应用的效果也不尽相同，各模型的外推性有待进一步验证，也是后续 Meta 分析中异质性产生的重要来源。因此，在后续基于预测模型的 Meta 分析中，应根据这些关键特征（是否接受抗病毒治疗、肝硬化状态）分别进行相应的亚组分析，以探讨异质性的来源，并进一步明确各模型在不同特征人群中的预测效能。

表1 慢性乙型肝炎患者肝细胞癌预测模型构建及验证情况

Table 1. Construction and validation of prediction models for HCC in patients with chronic hepatitis B

模型名称	发表年份	国家/地区	样本量（开发/验证）	终点指标	结局事件数（开发/验证）	预测HCC发生时间（年）*	肝硬化比例	抗病毒治疗比例	建模方法	模型开发	
										区分度	校准度
GAG-HCC[17]	2009	中国香港	820/16 325	HCC	40/899	5、10	15%	0%	COX回归	AUC	无
NGM1-HCC[18]	2010	中国台湾	2 435/3 323	HCC	103/131	5、10	–	0%	COX回归	AUC	相关系数
NGM2-HCC[18]	2010	中国台湾	2 435/3 323	HCC	103/131	5、10	–	0%	COX回归	AUC	相关系数
CU-HCC[19]	2010	中国香港	1 005/18 236	HCC	105/1 097	10	38%	15%	COX回归	AUC	无
REACH-B[20]	2011	中国台湾	3 584/20 565	HCC	131/1 307	3、5、10	0%	0%	COX回归	AUC	校准图
LSM-HCC[21]	2014	中国香港	1 035/3 158	HCC	38/170	5	32%	36%	COX回归	AUC	无
mREACH-BI[22]	2014	韩国	192/–	HCC	15/–	3	47%	100%	COX回归	AUC	无
mREACH-BII[22]	2014	韩国	192/–	HCC	15/–	3	47%	100%	COX回归	AUC	无
PAGE-B[23]	2016	希腊\意大利\西班牙\荷兰\土耳其	1 325/62 055	HCC	51/2 793	5	20%	100%	COX回归	Harrell C-index	校准图
RWS-HCC[24]	2016	新加坡	538/3 726	HCC	42/242	10	15%	18%	Logistic回归	AUC	无
mPAGE-B[25]	2018	韩国	2 001/40 367	HCC	132/2 177	5	19%	100%	COX回归	AUC	校准图
CAMD[26]	2018	中国台湾	23 851/19 321	HCC	596/478	3	27%	100%	COX回归	AUC	校准图
AASL-HCC[27]	2019	韩国	944/298	HCC	56/24	3、5	39%	100%	COX回归	AUC	校准图/HL检验
REAL-B[28]	2020	美国\中国\新西兰\南韩\日本	5 365/2 683	HCC	378/202	3、5、10	20%	100%	COX回归	AUC	校准图/HL检验

注：HL检验.Hosmer-Lemeshow检验；AUC.曲线下面积（area under the curve）；HCC.肝细胞癌（hepatocellular carcinoma）；*指模型预测HCC发生的时间点，部分研究预测了多个时间点HCC的发生（如GAG-HCC研究分别预测了5年和10年HCC的发生）。

在经外部验证的预测模型中，REACH-B[20]、CU-HCC[19]、PAGE-B[23] 和 mPAGE-B[25] 四个模型被广泛验证，外部验证样本量分别为 20 565 例（1 307 例 HCC）、18 236 例（1 097 例 HCC）、62 055 例（2 793 例 HCC）和 4 0367 例（2 177 例 HCC）。偏倚风险评价结果显示，大多数研究在研究对象（数据来源、纳入排除标准等）、预测变量（定义、测量方法及时间点等）、结局事件（诊断标准等）三方面的偏倚风险较低，40% 的研究在统计分析方面存在较大的偏倚风险，主要表现为结局事件数较少、多因素统计分析方法报告不规范、缺失数据处理欠佳、模型性能指标展示不足（尤其是校准度指标）等问题。

根据各模型外部验证的原始研究，预测 HCC 这一结局事件的时间设定为 3 年、5 年、7 年和 10 年，在不同时间点分别对各模型外部验证的区分度和校准度指标进行 Meta 分析（包括仅进行模型外部验证的研究，共 29 篇文献）。Meta 分析结果显示，区分度方面，大多数模型对慢性乙肝患者 HCC 发生风险的预测效果较好，尤其是以抗病毒治疗人群开发的 HCC 预测模型（如 REAL-B[28]、CAMD[26] 模型），其预测准确度更高；校准度方面，大多数模型的校准度表现欠佳，如 REACH-B[20] 模型通常低估 HCC 的发生风险，而 mPAGE-B[25] 模型则显示会高估 HCC 的发生风险，CAMD[26] 模型的校准度最佳，但因报告校准度指标的外部验证研究较少（已通过邮件索要了部分研究的校准度数据），Meta 分析合并的校准度指标 CI 较宽，故仅作为参考，该指标的解读需谨慎。

本研究对 HCC 预测模型 4 个时间点（3 年、5 年、7 年和 10 年）的区分度和校准度指标分别进行了 Meta 分析，工作量较大，但每次统计分析的过程是类似的，故本文仅选取了 REACH-B 这一模型对 3 年 HCC 发生风险的预测效果（包括区分度和校准度）进行展示，详细软件操作过程见附件。

8 结语

本文分别从研究问题提出、检索策略制定、文献筛选与偏倚风险评价、数据提取和准备、证据合并及 Meta 分析、结果解释及报告共计 6 个方面对预测模型系统综述与 Meta 分析的制作进行了介绍，并以慢性乙型肝炎患者 HCC 风险预测模型为例，详细介绍了从文献检索、数据提取到 Meta 分析的方法学步骤，为相关研究提供了实践参考。值得注意的是，本研究预测模型的系统综述和 Meta 分析的内容并非将多个不同模型在原始研究中的数据进行 Meta 分析，而是针对同一个模型的所有外部验证研究结果进行汇总。

附件见《医学新知》官网附录（https://yxxz.whuznhmedj.com/futureApi/storage/appendix/202503192.pdf）

伦理声明：不适用
作者贡献：文献查阅：周业胜、李彧格、于爽、刘昊睿；文章撰写：刘昊睿、武珊珊；文章修改：刘昊睿；文章审阅：武珊珊、孙凤；基金支持：武珊珊
数据获取：不适用
利益冲突声明：无
致谢：不适用

参考文献

1 van Smeden M, Reitsma JB, Riley RD, et al. Clinical prediction models: diagnosis versus prognosis[J]. J Clin Epidemiol, 2021, 132: 142-145. DOI: 10.1016/j.jclinepi.2021.01.009.
2 Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obst ructive pulmonary disease: systematic review and critical appraisal[J]. BMJ, 2018, 367: l5358. DOI: 10.1136/bmj.l5358.
3 Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstru ctive pulmonary disease: the updated BODE index and the ADO index[J]. Lancet, 2009, 374(9691): 704-711. DOI: 10.1016/S0140-6736(09)61301-5.
4 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(6):1635-1644. DOI: 10.1183/13993003.01485-2015.
5 Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and Meta-analysis of prediction model performance[J]. BMJ, 2017, 356: i6460. DOI: 10.1136/bmj.i6460.
6 Damen JA, Moons KG, Smeden M, et al. How to conduct a systematic review and Meta-analysis of prognostic model studies[J]. Clin Microbiol Infect, 2023, 29(4): 434-440. DOI: 10.1016/j.cmi.2022.07.019.
7 Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews[J]. PLoS One, 2012, 7(2): e32844. DOI: 10.1371/journal.pone.0032844.

8 Wolff RF, Moons KG, Riley RD, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies[J]. Ann Intern Med, 2019, 170(1): 51–58. DOI: 10.7326/M18–1377.

9 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist[J]. PLoS Med, 2014, 11(10): e1001744. DOI: 10.1371/journal.pmed.1001744.

10 Debray TP, Damen JA, Riley RD, et al. A framework for Meta–analysis of prediction model studies with binary and time–to–event outcomes[J]. Stat Methods Med Res, 2019, 28(9): 2768–2786. DOI: 10.1177/0962280218785504.

11 Foroutan F, Guyatt G, Zuk V, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks[J]. J Clin Epidemiol, 2020, 121: 62–70. DOI: 10.1016/j.jclinepi.2019.12.023.

12 Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement[J]. BMJ, 2015, 350: g7594. DOI: 10.1136/bmj.g7594.

13 Snell K, Levis B, Damen J, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and Meta–analyses (TRIPOD–SRMA)[J]. BMJ, 2023, 381: e73538. DOI: 10.1136/bmj–2022–073538.

14 Wang F, Fan J, Zhang Z, et al. The global burden of liver disease: the major impact of China[J]. Hepatology, 2014, 60(6): 2099–2108. DOI: 10.1002/hep.27406.

15 Liu J, Zhang S, Wang Q, et al. Seroepidemiology of hepatitis B virus infection in 2 million men aged 21–49 years in rural China: a population–based, cross–sectional study[J]. Lancet Infect Dis, 2016, 16(1): 80–86. DOI: 10.1016/s1473–3099(15)00218–2.

16 Wu S, Zeng N, Sun F, et al. Hepatocellular carcinoma prediction models in chronic hepatitis B: a systematic review of 14 models and external validation[J]. Clin Gastroenterol Hepatol, 2021, 19(12): 2499–2513. DOI: 10.1016/j.cgh.2021.02.040.

17 Yuen MF, Tanaka Y, Fong DY, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B[J]. Journal of hepatology, 2009, 50(1): 80–88. DOI: 10.1016/j.jhep.2008.07.023.

18 Yang HI, Sherman M, Su J, et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection[J]. J Clin Oncol, 2010, 28(14): 2437–2444. DOI: 10.1200/jco.2009.27.4456.

19 Wong VW, Chan SL, Mo F, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers[J]. J Clin Oncol, 2010, 28(10): 1660–1665. DOI: 10.1200/jco.2009.26.2675.

20 Yang HI, Yuen MF, Chan HL, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH–B): development and validation of a predictive score[J]. Lancet Oncol, 2011, 12(6): 568–574. DOI: 10.1016/s1470–2045(11)70077–8.

21 Wong GL, Chan HL, Wong CK, et al. Liver stiffness–based optimization of hepatocellular carcinoma risk score in patients with chronic hepatitis B[J]. J Hepatol, 2014, 60(2): 339–345. DOI: 10.1016/j.jhep.2013.09.029.

22 Lee HW, Yoo EJ, Kim BK, et al. Prediction of development of liver–related events by transient elastography in hepatitis B patients with complete virological response on antiviral therapy[J]. Am J Gastroenterol, 2014, 109(8): 1241–1249. DOI: 10.1038/ajg.2014.157.

23 Papatheodoridis G, Dalekos G, Sypsa V, et al. PAGE–B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5–year antiviral therapy[J]. J Hepatol, 2016, 64(4): 800–806. DOI: 10.1016/j.jhep.2015.11.035.

24 Poh Z, Shen L, Yang HI, et al. Real–world risk score for hepatocellular carcinoma (RWS–HCC): a clinically practical risk predictor for HCC in chronic hepatitis B[J]. Gut, 2016, 65(5): 887–888. DOI: 10.1136/gutjnl–2015–310818.

25 Kim JH, Kim YD, Lee M, et al. Modified PAGE–B score predicts the risk of hepatocellular carcinoma in Asians with chronic hepatitis B on antiviral therapy[J]. J Hepatol, 2018, 69(5): 1066–1073. DOI: 10.1016/j.jhep.2018.07.018.

26 Hsu YC, Yip TC, Ho HJ, et al. Development of a scoring system to predict hepatocellular carcinoma in Asians on antivirals for chronic hepatitis B[J]. J Hepatol, 2018, 69(2): 278–285. DOI: 10.1016/j.jhep.2018.02.032.

27 Yu JH, Suh YJ, Jin YJ, et al. Prediction model for hepatocellular carcinoma risk in treatment–naive chronic hepatitis B patients receiving entecavir/tenofovir[J]. Eur J Gastroenterol Hepatol, 2019, 31(7): 865–872. DOI: 10.1097/meg.0000000000001357.

28 Yang HI, Yeh ML, Wong GL, et al. Real–world effectiveness from the Asia Pacific rim liver consortium for HBV risk score for the prediction of hepatocellular carcinoma in chronic hepatitis B patients treated with oral antiviral therapy[J]. J Infect Dis, 2020, 221(3): 389–399. DOI: 10.1093/infdis/jiz477.

29 Yang HI, Tseng TC, Liu J, et al. Incorporating serum level of hepatitis B surface antigen or omitting level of hepatitis B virus DNA does not affect calculation of risk for hepatocellular carcinoma in patients without cirrhosis[J]. Clin Gastroenterol Hepatol, 2016, 14(3): 461–468. e462. DOI: 10.1016/j.cgh.2015.10.033.

30 陈薇 , 陈靖 , 郑琦 , 等 . 乙型肝炎病毒相关性肝细胞癌预测模型的验证 [J]. 中华肝脏病杂志 , 2015, 23(7): 507–511. [Chen W, Chen J, Zheng Q, et al. Validation study of prediction models of hepatitis B virus–related hepatocellular carcinoma[J]. Chinese Journal of Hepatology, 2015, 23(7): 507–511.] DOI: 10.3760/cma.j.issn.1007–3418.2015.07.008.

31 Kim WR, Loomba R, Berg T, et al. Impact of long–term tenofovir disoproxil fumarate on incidence of hepatocellular carcinoma in patients with chronic hepatitis B[J]. Cancer, 2015, 121(20): 3631–3638. DOI: 10.1002/cncr.29537.

32 Seo YS, Jang BK, Um SH, et al. Validation of risk prediction models for the development of HBV–related HCC: a retrospective multi–center 10–year follow–up cohort study[J]. Oncotarget, 2017,

8(68): 113213–113224. DOI: 10.18632/oncotarget.22375.

33 Jeon MY, Lee HW, Kim SU, et al. Feasibility of dynamic risk prediction for hepatocellular carcinoma development in patients with chronic hepatitis B[J]. Liver Int, 2018, 38(4): 676–686. DOI: 10.1111/liv.13583.

34 Kim MN, Hwang SG, Rim KS, et al. Validation of PAGE–B model in Asian chronic hepatitis B patients receiving entecavir or tenofovir[J]. Liver Int, 2017, 37(12): 1788–1795. DOI: 10.1111/liv.13450.

35 Ahn J, Lim JK, Lee HM, et al. Lower observed hepatocellular carcinoma incidence in chronic hepatitis B patients treated with entecavir: results of the ENUMERATE study[J]. Am J Gastroenterol, 2016, 111(9): 1297–1304. DOI: 10.1038/ajg.2016.257.

36 Wong GL, Chan HL, Chan HY, et al. Accuracy of risk scores for patients with chronic hepatitis B receiving entecavir treatment[J]. Gastroenterology, 144(5): 933–944. DOI: 10.1053/j.gastro.2013.02.002.

37 Coffin CS, Rezaeeaval M, Pang JX, et al. The incidence of hepatocellular carcinoma is reduced in patients with chronic hepatitis B on long–term nucleos(t)ide analogue therapy[J]. Aliment Pharmacol Ther, 2014, 40(11–12): 1262–1269. DOI: 10.1111/apt.12990.

38 Arends P, Sonneveld MJ, Zoutendijk R, et al. Entecavir treatment does not eliminate the risk of hepatocellular carc inoma in chronic hepatitis B: limited role for risk scores in Caucasians[J]. Gut, 64(8): 1289–1295. DOI: 10.1136/gutjnl–2014–307023.

39 Kim GA, Lee HC, Kim MJ, et al. Incidence of hepatocellular carcinoma after HBsAg seroclearance in chronic hepatitis B patients: a need for surveillance[J]. J Hepatol, 2015, 62(5): 1092–1099. DOI: 10.1016/j.jhep.2014.11.031.

40 Papatheodoridis GV, Dalekos GN, Yurdaydin C, et al. Incidence and predictors of hepatocellular carcinoma in Caucasian chronic hepatitis B patients receiving entecavir or tenofovir[J]. J Hepatol, 2015, 62(2): 363–370. DOI: 10.1016/j.jhep.2014.08.045.

41 Abu–Amara M, Cerocchi O, Malhi G, et al. The applicability of hepatocellular carcinoma risk prediction scores in a North American patient population with chronic hepatitis B infecti on[J]. Gut, 2016, 65(8): 1347–1358. DOI: 10.1136/gutjnl–2014–309099.

42 Tawada A, Chiba T, Saito T, et al. Utility of prediction scores for hepatocellular carcinoma in patients with chronic hepatitis B treated with nucleos(t)ide analogues[J]. Oncology, 2016, 90(4): 199–208. DOI: 10.1159/000444392.

43 Brouwer WP, van der Meer AJP, Boonstra A, et al. Prediction of long–term clinical outcome in a diverse chronic hepatitis B population: role of the PAGE–B score[J]. J Viral Hepat, 2017, 24(11): 1023–1031. DOI: 10.1111/jvh.12727.

44 Daheim M, Lang S, Goeser T, et al. Real–world risk score for hepatocellular carcinoma risk prediction in CHBV: a validation outside of Asia[J]. Gut, 2017, 66(7): 1346–1347. DOI: 10.1136/gutjnl–2016–312993.

45 Riveiro–Barciela M, Tabernero D, Calleja JL, et al. Effectiveness and safety of entecavir or tenofovir in a Spanish cohort of chronic hepatitis B patients: validation of the Page–B score to predict hepatocellular carcinoma[J]. Dig Dis Sci, 2017, 62(3): 784–793. DOI: 10.1007/s10620–017–4448–7.

46 Yip TC, Wong GL, Wong VW, et al. Reassessing the accuracy of PAGE–B–related scores to predict hepatocel lular carcinoma development in patients with chronic hepatitis B[J]. J Hepatol, 2020, 72(5): 847–854. DOI: 10.1016/j.jhep.2019.12.005.

收稿日期：2025 年 03 月 25 日 修回日期：2025 年 05 月 09 日
本文编辑：李绪辉 曹 越

引用本文：刘昊睿, 周业胜, 李彧格, 等. 预测模型系统评价的方法与实践：基于慢性乙型肝炎患者肝癌发生风险预测模型的案例分析[J]. 医学新知, 2025, 35(8): 861–869. DOI: 10.12173/j.issn.1004–5511.202503192.
Liu HR, Zhou YS, Li YG, et al. Systematic review and Meta–analysis of prediction models: a case study on the risk prediction model for hepatocellular carcinoma in patients with chronic hepatitis B[J]. Yixue Xinzhi Zazhi, 2025, 35(8): 861–869. DOI: 10.12173/j.issn.1004–5511.202503192.