

慢性肾脏病发病风险预测模型研究进展

曾芷青¹ 杨淞淳² 余灿清^{1,3} 张路霞^{4,5} 吕筠^{1,3} 李立明^{1,3}

¹北京大学公共卫生学院流行病学与卫生统计学系,北京 100191;²中南大学湘雅医院皮肤科,长沙 410008;³北京大学公众健康与重大疫情防控战略研究中心,北京 100191;⁴北京大学健康医疗大数据国家研究院,北京 100191;⁵北京大学第一医院肾内科/北京大学肾脏病研究所,北京 100034

通信作者:吕筠,Email:lvjun@bjmu.edu.cn

【摘要】 慢性肾脏病(CKD)是全球重要的公共卫生问题,严重危害人群健康。利用预测模型对人群未来一段时间的CKD发病风险进行分层,针对高危人群采取干预措施是实现CKD一级预防的重要途径。世界范围内已经开发出了二十多个CKD发病风险预测模型,我国学者也开发出了4个适用于中国人群的预测模型,但目前的临床指南中尚未推荐使用任何专门的CKD风险预测模型。现有模型在结局定义、预测因子、缺失数据处理和建模方法选择方面仍有局限。在未来,新兴生物标志物和多基因风险评估的应用以及机器学习方法的发展将为继续改进模型提供更多可能。

【关键词】 慢性肾脏病; 风险预测模型; 一级预防

基金项目:国家自然科学基金(82192904,82192901,82192900);中央高校基本科研业务费医-X种子基金(BMU2022MX025)

Progress in research of risk prediction model for chronic kidney disease

Zeng Zhiqing¹, Yang Songchun², Yu Canqing^{1,3}, Zhang Luxia^{4,5}, Lyu Jun^{1,3}, Li Liming^{1,3}

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ²Department of Dermatology, Xiangya Hospital, Central South University, Changsha 410008, China; ³Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing 100191, China; ⁴National Institute of Health Data Science of Peking University, Beijing 100191, China; ⁵Department of Nephrology, Peking University First Hospital/Institute of Nephrology, Peking University, Beijing 100034, China

Corresponding author: Lyu Jun, Email: lvjun@bjmu.edu.cn

【Abstract】 Chronic kidney disease (CKD) is an important global public health problem that greatly threatens population health. Application of risk prediction model is a crucial way for the primary prevention of CKD, which can stratify the risk for developing CKD and identify high-risk individuals for more intensive interventions. By now, more than twenty risk prediction models for CKD have been developed worldwide. There are also four domestic risk prediction models developed for Chinese population. However, none of these models have been recommended in clinical guidelines yet. The existing risk prediction models have some limitations in terms of outcome definition, predictors, strategies for handling missing data, and model derivation. In the future, the applications of emerging biomarkers and polygenic risk scores as well as advances in machine learning methods will provide more possibilities for the further improvement of the model.

【Key words】 Chronic kidney disease; Risk prediction model; Primary prevention

Fund programs: National Natural Science Foundation of China (82192904, 82192901, 82192900); Peking University Medicine Seed Fund for Interdisciplinary Research Supported by "the Fundamental Research Funds for the Central Universities" (BMU2022MX025)

DOI:10.3760/cma.j.cn112338-20220908-00771

收稿日期 2022-09-08 本文编辑 万玉立

引用格式:曾芷青,杨淞淳,余灿清,等.慢性肾脏病发病风险预测模型研究进展[J].中华流行病学杂志,2023,44(3):498-503. DOI: 10.3760/cma.j.cn112338-20220908-00771.

Zeng ZQ, Yang SC, Yu CQ, et al. Progress in research of risk prediction model for chronic kidney disease[J]. Chin J Epidemiol, 2023, 44(3):498-503. DOI: 10.3760/cma.j.cn112338-20220908-00771.



在过去的二十年间,慢性肾脏病(chronic kidney disease, CKD)作为全球重要的公共卫生问题受到广泛关注^[1-2]。CKD具有患病率高、知晓率低、预后差和医疗费用高的特点,患者发生终末期肾病、心血管病和过早死亡的风险显著增加^[3-6],给个人和社会带来了沉重的疾病和经济负担。2017年,全球CKD患者总数超过6.97亿,CKD所造成的死亡达120万例^[7]。在1990-2019年全年年龄段人群中CKD所引起的伤残调整寿命年(disability adjusted life years, DALYs)从第29位上升到第18位,是上升速度较快的疾病之一^[8]。利用预测模型对人群CKD发病风险进行分层,针对高危人群采取干预措施是CKD一级预防的重要途径^[4,9]。本文在介绍国内外主要CKD风险预测模型的基础上,重点概述现有模型在建模时的局限性以及今后继续优化模型的方向。

一、国外主要CKD发病风险预测模型

CKD发病风险预测模型研究最早开始于欧美国家^[10],随后泰国^[11]、日本^[12]和韩国^[13]等亚洲国家也陆续开展。目前国外已有20多个基于一般人群建立的CKD发病风险预测模型,这些模型通常具有以下特征:①研究人群年龄范围多在20~75岁之间,人群基线特征存在较大差异;②均基于前瞻性队列研究构建,但样本量和结局事件数差异较大;③大多数研究采用单次肾小球滤过率(glomerular filtration rate, GFR)低于 $60 \text{ ml} \cdot \text{min}^{-1} \cdot (1.73 \text{ m}^2)^{-1}$ 来定义CKD结局事件,部分研究同时将阳性蛋白尿作为诊断标准。模型中常见的预测因子包括年龄、性别、BMI、SBP、吸烟状况、糖尿病和高血压^[5,14],部分研究额外纳入了临床检验指标和多种疾病史作为预测因子。预测模型区分度多在0.7~0.9之间,仅有部分模型经过外验证。本文重点介绍其中具有代表性的3项研究。

2010年,Hippisley-Cox和Coupland^[15]在接近160万名35~74岁英国人群中分性别建立并验证了用于预测5年CKD发病风险的QKidney模型。男性模型包括年龄、种族、BMI、SBP、吸烟状况、糖尿病、高血压等预测因子;女性模型则在男性模型的基础上额外纳入系统性红斑狼疮和肾结石病史,两个模型内验证及外验证Harrell's C指数均高于0.87。QKidney模型区分度和校准度均较高,并于2012-2016年在英国多个大型数据库中得到了广泛外验证^[16-18]。

2012年,Alssema等^[19]在荷兰6 019名28~85岁研究对象中建立了一个用于预测7年慢性心血管代谢疾病风险的模型(Alssema模型)。研究结局定义为随访期间出现以下任意一种诊断:心血管病、心血管病所致死亡或猝死、2型糖尿病以及CKD。Alssema模型共纳入7个预测因子:年龄、BMI、腰围、吸烟状况、高血压用药史、心血管病家族史、糖尿病家族史;在男性中受试者工作特征曲线下面积(area under the curve, AUC)为0.82,女性中AUC为0.81。这是一项同时对多种慢性心血管代谢疾病进行预测的研究;此模型不需要实验室检查的结果,可以简化基层实践中心血管

代谢疾病的预防工作^[19],已被纳入荷兰全科医生指南^[20]。随后,此模型分别于澳大利亚和伊朗人群中进行了外验证,结果显示Alssema模型具有较高的区分度,但校准度不佳^[20-21]。

2019年,Nelson等^[22]在来自CKD预后联盟(Chronic Kidney Disease Prognosis Consortium, CKD-PC)的31个不同国家队列、超过440万的无糖尿病研究对象中建立了一个5年CKD发病风险预测模型(CKD-PC模型)。模型共纳入9个预测因子:年龄、性别、种族、BMI、吸烟状况、GFR(估算)、高血压、心血管病、尿白蛋白/肌酐比值。该模型于多国、大样本量的人群中建模并进行外验证,相较于在单一人群中建立的CKD风险预测模型,具有较好的普适性^[22],其模型性能也优于既往模型^[23-24]。

二、国内CKD发病风险预测模型

目前在中国人群中建立的CKD发病风险预测模型共有4个。除此之外,CKD-PC模型建模时纳入的31个不同国家队列中包含2个中国人群队列。

2010年,Chien等^[23]在5 168名接受健康体检的研究对象中建立了一个简单的CKD风险预测模型(Chien模型)。Chien模型仅纳入5个预测因子(年龄、BMI、DBP、2型糖尿病、心血管病),其信息易获取;模型在建模队列中的Harrell's C指数为0.768,但在外验证队列中出现明显下降(0.667)。该研究中位随访时间仅2.2年,在预测更长时间范围的CKD发病时可能受到限制。2013年,孙凤等^[25]在18 275名参加中国台湾地区美兆健康体检人群中建立了CKD的5年发病风险预测模型(美兆模型)。除年龄、文化程度、糖尿病和高血压外,美兆模型还纳入了LDL-C、TG和血清肌酐,在外验证队列中模型AUC为0.768,区分度较高。上述两个预测模型均是在自愿参加健康体检的人群中建立的,长期参加健康体检者与很少或不参加者可能存在社会人口学等特征方面的差异,限制了模型的外推能力。

2020年,Wen等^[26]利用中国邯郸眼病研究的3 266名研究对象数据构建了一个适用于农村人群的预测模型(邯郸模型)。邯郸模型仅纳入性别、文化程度、腰围、SBP和糖尿病作为预测因子,模型AUC为0.717。由于该模型纳入的预测因子简单、不依赖于实验室检测,可以更容易地整合到农村初级卫生保健系统中快速地识别高CKD发病风险人群,具有较高的实用性。

2021年,Yu等^[27]在上海市“高峰计划”人群队列与生物标本库中分性别构建了3年CKD发病风险预测模型(上海高峰模型),研究人群年龄在20~74岁间,共纳入4 117名男性和5 932名女性研究对象。这项研究将女性生殖因素(包括初潮年龄、经期长度、怀孕次数和绝经状态)纳入预测模型的研究,其结果显示初潮年龄与女性CKD发病风险升高有关。上海高峰模型区分度很高,在女性中的Harrell's C指数达到0.951,男性中则为0.934,但其纳入的预测因子多为临床检验指标,一定程度上限制了其应用场所和可推广性。

构建CKD-PC模型时纳入的两个中国人群队列分别是

“北京队列”和“台湾美兆健检队列”^[22]。“北京队列”的研究对象是在北京市古城社区卫生服务中心服务范围内的、自愿参与研究的 40 岁以上城市居民^[28]。“台湾美兆健检队列”的研究对象是 1996–2004 年在美兆机构接受健康检查的 ≥ 20 岁成年人^[29]。CKD-PC 模型在“北京队列”中的 Harrell's C 指数(0.789)低于 31 个队列的综合水平(0.845),而在“台湾美兆健检队列”中的 Harrell's C 指数(0.917)高于综合水平^[22]。CKD-PC 模型在更广泛的中国人群中的预测表现有待更多研究确认。

三、现有模型的局限性

恰当的建模方法是模型良好预测能力的基础。然而,2013 年一项系统综述指出,2011 年 9 月之前发表的 CKD 预测模型研究通常在缺失数据处理、建模方法选择以及模型性能评估等方面存在不足,并且普遍存在报告信息不全面的情况,影响了预测模型的可靠性^[14]。结合此系统综述所提出的问题及对 2011 年 9 月之后发表的研究进行梳理,对现有 CKD 发病风险预测模型存在的局限性总结为以下几方面。

1. 结局定义:根据“改善全球肾脏病预后组织”(Kidney Disease:Improving Global Outcomes,KDIGO)工作组 2012 年制定的 CKD 临床实践指南,CKD 的定义为肾脏结构或功能异常超过 3 个月^[30],其重要诊断标准是 $GFR < 60 \text{ ml} \cdot \text{min}^{-1} \cdot (1.73 \text{ m}^2)^{-1}$ 。GFR 测定的“金标准”是计算外源性物质(如菊粉、放射性同位素等)的清除率^[31],但其操作繁琐、费时费力,不适合在流行病学研究中应用,目前的研究均是采用公式来估算 GFR。

2013 年之前发表的研究都是基于肾脏疾病饮食改良(modification of diet in renal disease,MDRD)公式来估算 GFR。但已有证据显示,MDRD 公式在实际 GFR 高于诊断阈值的情况下会低估 GFR,导致对 CKD 的“过度诊断”;并且 MDRD 公式在不同种族人群中的准确性存在差异^[5,32]。2014 年更新的英国国家卫生与临床优化研究所指南推荐使用 CKD 流行病学研究公式(chronic kidney disease epidemiology collaboration,CKD-EPI)^[33]。此公式已在成年人中得到了广泛验证,证实其能减少估算偏差,具有比 MDRD 公式更好的精度^[33]。研究人员仍致力于开发出更加准确的计算公式,目前已有针对老年人群的柏林倡议研究公式^[34]、适用于一般人群的 Lund-Malmö 公式^[35]和全年龄范围公式^[36]。

$GFR < 60 \text{ ml} \cdot \text{min}^{-1} \cdot (1.73 \text{ m}^2)^{-1}$ 持续 3 个月以上才能诊断 CKD。然而,目前除一项研究采用随访期间至少两次连续 GFR 估算结果来定义结局外^[13],其余研究都是采用单次 GFR 估算结果,不能排除某些急性可逆性肾损伤的患者,可能高估 CKD 的发病率^[37]。如果大型人群队列研究通过匹配现有的电子病历系统,根据出院诊断中的国际疾病分类编码来确定 CKD,可能存在灵敏度低而特异度高的特点,即漏掉一些处于疾病早期的患者,从而低估 CKD 的发病率^[38]。对 CKD 结局的定义如何兼顾真实性和获取数据的可行性是一个难点。

2. 预测因子:现有的 CKD 风险预测模型所纳入的预测因子数从 4 个到 17 个不等,模型中的预测因子并不是越多越好,纳入的预测因子越多,在建模时越容易出现过拟合的情况。同时,预测因子信息越是方便获取或便于测量,模型的可推广性越好。

为了方便预测模型的实际应用,目前的研究倾向于将连续变量处理为分类变量,然而如何设定截断值以及分类的数量往往没有统一的标准,这样的处理方式会造成信息的丢失和检验效能的降低^[14]。2015 年发表的《个体预后与诊断预测模型研究报告规范》(transparent reporting of a multivariable prediction model for individual prognosis or diagnosis,TRIPOD)声明指出,预测模型研究需明确报告每个预测因子的处理方法,包括将连续变量转换为分类变量时的截断值界定方法、原始数据的转换方法(如对数转换等),以及预测因子与结局之间是线性还是非线性关系^[39]。若模型为非线性回归模型,应报告采用的方法,如多项式回归或限制性立方样条等;若为线性回归模型,则建议报告预测因子与结局变量呈线性关系的研究假设^[39]。另外,部分研究没有报告连续型预测因子(如年龄)的范围,例如 CKD-PC 模型^[22]、Chien 模型^[23]和邯郸模型^[26]都只报告了研究对象的平均年龄。超出模型适用年龄范围的应用可能会导致有偏的风险估计^[14]。为明确预测模型适用的目标人群、准确判断研究背景,TRIPOD 声明同时建议研究者提供所有预测因子和结局指标等重要变量的分布情况信息^[39-40]。

3. 缺失数据:对于缺失值的处理方法是 CKD 发病风险预测模型研究中经常遗漏报告的部分。上述 7 项研究中,仅有 2 项研究使用插补法进行缺失数据的填补^[15,19],另外 1 项研究明确报告了有缺失值的个体数并将其排除^[22],而剩余 4 项研究未报告缺失值的处理情况^[23,25-27]。研究者倾向于排除有缺失数据的个体,从而进行所谓的“完整案例分析”,一方面,这种方法往往会导致样本量的减少,降低统计效能;另一方面,纳入分析的研究对象不一定能代表原始研究人群,造成了选择偏倚^[5,14]。TRIPOD 声明指出,预测模型研究应详细描述研究数据的缺失情况以及处理方法,包括具体插补的变量及插补次数;更推荐采用多重插补法进行缺失数据处理,因为其能保留缺失数据的随机性,相较单一插补法具有更高的准确率^[39]。

4. 建模方法:logistic 回归模型和 Cox 比例风险回归模型是用于构建预测模型的两种常见方法:前者方法简单、对数据资料要求较为宽松,通常用于横断面研究数据(例如构建疾病诊断模型)和短期风险(例如 30 d 住院死亡率)的预测;而后者可以有效利用随访时间信息并对截尾数据进行分析,适用于较长时间范围内风险(例如 5 年、10 年发病风险)的预测^[39]。在慢性病发病风险预测领域,采用 logistic 回归模型来代替 Cox 比例风险回归模型的前提是研究者需要证明队列随访中不存在截尾数据^[39]。然而,目前大部分 CKD 发病风险预测模型仍是基于 logistic 回归模型构建的,

并且未报告采用该方法的原因,在未来的研究中应更注重建模方法的选择以及研究报告的规范。

5. 模型应用:2019年10月,KDIGO召开了以“CKD早期发现与干预”为主题的共识会议,提出CKD-PC模型可用于评估人群CKD发病风险并确定个体化再筛查频率,建议高风险个体每1~2年进行一次筛查,而较低风险个体可适当延长间隔^[41]。除此之外,目前的临床指南中尚未推荐使用任何专门的CKD风险预测模型。缺乏客观、无偏的外验证是影响模型实际应用的原因之一^[14]。部分模型没有在独立的外部人群中进行验证,或在进行外验证时模型的性能大幅度下降^[25,27];另一部分研究是将同一个数据集随机分成建模数据集和验证数据集来进行内验证^[10,15],除非数据集样本量非常庞大,这种验证方法往往会得到有偏的结果^[14]。如果确实难以找到独立的外验证数据集,推荐使用自抽样法(bootstrap)来进行内验证^[42]。自抽样法可以利用全部的研究数据,能改善参数估计、降低模型过拟合的可能性^[14]。

此外,除CKD-PC模型之外^[22],上述CKD风险预测模型大多是在局部地区人群中进行建模,这些模型能否直接应用于不同国家和种族的人群,包括现有中国人群的模型能否应用于中国其他地区,都有待进一步的验证,未来的研究应纳入具有不同遗传背景和环境特征的研究对象^[5]。

四、未来模型优化方向

1. 探索可继续改进模型的预测因子:

(1)生物标志物:近年来由于转录组学、蛋白质组学和代谢组学的发展,研究者发现了越来越多与肾脏疾病有关的生物标志物,包括不对称二甲基精氨酸、对称二甲基精氨酸、尿调节素、肾损伤分子1、中性粒细胞明胶酶相关脂质运载蛋白等^[43]。有可能作为CKD发病风险预测因子的生物标志物应具有以下特点:生物学变异性小,不受年龄、营养状况或共病状况的影响,方便测量等^[44-45]。

新的生物标志物能否改善现有预测模型的性能,可以通过计算净重新分类指数(net reclassification index, NRI)或整体鉴别指数(integrated discrimination improvement)等指标来评价^[24,46]。例如,Fox等^[44]探究了C-反应蛋白、醛固酮、肾素、B型尿钠肽、纤溶酶原激活物抑制剂-1、纤维蛋白原和同型半胱氨酸等生物标志物能否改善CKD发病风险的预测。结果显示,在原有CKD发病风险预测模型中额外纳入具有统计学意义的生物标志物,Harrell's C指数提升了0.012, NRI为6.9%,说明增加这些生物标志物可以一定程度提升模型的性能^[44]。未来随着更多相关生物标志物的发现,将为改善CKD风险预测模型性能、准确预测人群患病风险提供更广阔的前景。

(2)多基因风险评分:近年来使用多基因风险评分(polygenic risk score, PRS)来预测CKD发病风险也是研究的热点方向^[47]。然而,向传统风险预测模型中加入PRS似乎并不能显著提升模型的性能。早在2012年,O'Seaghdha等^[48]利用16个与3期CKD相关的单核苷酸多态性

(single-nucleotide polymorphisms, SNPs)位点构建PRS,结果发现额外纳入PRS作为预测因子的模型与原模型相比,Harrell's C指数的差异无统计学意义。其原因可能是所使用的SNPs位点数目太少,仅能解释1.43%的遗传变异^[48];且选取的SNPs来源于白人人群,可能并不适用于其他人群^[49]。2021年Yu等^[47]使用LDpred法在英国生物银行(UK Biobank, UKB)队列中构建了包含超过150万个SNPs位点的PRS。仅将PRS作为预测因子来预测CKD的发病风险时,模型AUC为0.58;使用传统预测因子(包括性别、年龄、文化程度、BMI、吸烟状况、糖尿病、高血压、冠心病史)的模型AUC为0.685;在此基础上额外纳入PRS的模型AUC略有增加(0.690),但其差异仍无统计学意义^[47]。PRS在CKD风险预测方面的应用价值尚难以定论。现有全基因组关联研究在欧美白人以外的其他种族人群中仍十分有限,是限制这一方向发展的重要瓶颈。

2. 采用新的模型构建算法:机器学习方法在处理高维度、大样本数据时具有独特优势^[50],可以克服传统建模方法在处理预测因子与结局之间非线性关系方面的局限^[51],有可能挖掘庞大的电子健康档案(electronic health records, EHR)数据库中的有用信息,正越来越多地被应用于疾病风险预测领域^[52]。Krishnamurthy等^[53]利用中国台湾地区医保索赔数据库构建了用于预测6个月或12个月CKD发病风险的机器学习模型。其研究结果显示,卷积神经网络模型预测能力最强,AUC分别为0.957和0.954,双向长短期记忆网络模型次之;年龄、糖尿病、痛风、慢性肾小球肾炎、高血压以及磺胺类药物和血管紧张素类药物用药史是最强的预测因子^[53]。Luo等^[54]采用轻量梯度提升机算法在UKB数据库中探究了37种生活方式因素(包括膳食模式、体力活动、睡眠、吸烟和饮酒等)对CKD发病风险的影响,并构建了一个生活方式评分模型。面包摄入量、体力活动水平、饮茶量和饮水量是与CKD关联最密切的因素;纳入多种生活方式因素的模型Harrell's C指数为0.710^[54]。

尽管机器学习模型已取得不错的区分度,但其在模型解读及外推性方面仍存在局限。设法弄清机器学习模型中的“黑箱”^[52]、开展对外验证研究是今后该领域发展的重要方向^[55]。

五、小结

预测个体未来发生疾病的风险是现代健康管理的关键一步。构建和应用CKD风险预测模型,识别高危人群,开展分级管理,可促进CKD的预防关口前移。现有CKD风险预测模型的构建环节仍存在局限性,今后的研究应提高对结局定义、预测因子选择和缺失数据处理的关注度;除此之外,寻找合适的建模及验证队列人群、提高模型的可推广性仍是今后研究的重点和难点所在。未来,随着健康医疗大数据平台以及多组学研究的发展,可以尝试优化建模算法,探索可继续改进模型的预测因子,以期进一步提高模型的预测能力。

利益冲突 所有作者声明无利益冲突

参 考 文 献

- [1] Taal MW. Chronic kidney disease 10 years on: what have we learned? [J]. *Curr Opin Nephrol Hypertens*, 2012, 21(6):607-611. DOI:10.1097/MNH.0b013e328358a30e.
- [2] Yang C, Gao BX, Zhao XJ, et al. Executive summary for China kidney disease network (CK-NET) 2016 annual data report[J]. *Kidney Int*, 2020, 98(6):1419-1423. DOI: 10.1016/j.kint.2020.09.003.
- [3] di Angelantonio E, Chowdhury R, Sarwar N, et al. Chronic kidney disease and risk of major cardiovascular disease and non-vascular mortality: prospective population based cohort study[J]. *BMJ*, 2010, 341:c4986. DOI:10.1136/bmj.c4986.
- [4] Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review[J]. *JAMA*, 2019, 322(13):1294-1304. DOI:10.1001/jama.2019.14745.
- [5] Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review[J]. *PLoS Med*, 2012, 9(11): e1001344. DOI: 10.1371/journal.pmed.1001344.
- [6] Weiner DE, Tighiouart H, Amin MG, et al. Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies[J]. *J Am Soc Nephrol*, 2004, 15(5):1307-1315. DOI: 10.1097/01.asn.0000123691.46138.e2.
- [7] GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. *Lancet*, 2020, 395(10225): 709-733. DOI:10.1016/S0140-6736(20)30045-3.
- [8] GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019[J]. *Lancet*, 2020, 396(10258): 1204-1222. DOI: 10.1016/S0140-6736(20)30925-9.
- [9] Taal MW, Brenner BM. Predicting initiation and progression of chronic kidney disease: developing renal risk scores[J]. *Kidney Int*, 2006, 70(10):1694-1705. DOI: 10.1038/sj.ki.5001794.
- [10] Kshirsagar AV, Bang H, Bombardieri AS, et al. A simple algorithm to predict incident kidney disease[J]. *Arch Intern Med*, 2008, 168(22): 2466-2473. DOI: 10.1001/archinte.168.22.2466.
- [11] Saranburut K, Vathesatogkit P, Thongmung N, et al. Risk scores to predict decreased glomerular filtration rate at 10 years in an Asian general population[J]. *BMC Nephrol*, 2017, 18(1):240. DOI:10.1186/s12882-017-0653-z.
- [12] Umesawa M, Sairenchi T, Haruyama Y, et al. Validity of a risk prediction equation for CKD after 10 years of follow-up in a Japanese population: the Ibaraki prefectural health study[J]. *Am J Kidney Dis*, 2018, 71(6): 842-850. DOI:10.1053/j.ajkd.2017.09.013.
- [13] Lee C, Yun HR, Joo YS, et al. Framingham risk score and risk of incident chronic kidney disease: a community-based prospective cohort study[J]. *Kidney Res Clin Pract*, 2019, 38(1):49-59. DOI:10.23876/j.krcp.18.0118.
- [14] Collins GS, Omar O, Shanyinde M, et al. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods[J]. *J Clin Epidemiol*, 2013, 66(3): 268-277. DOI:10.1016/j.jclinepi.2012.06.020.
- [15] Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney[®] scores[J]. *BMC Fam Pract*, 2010, 11: 49. DOI: 10.1186/1471-2296-11-49.
- [16] Collins G, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney[®] scores using a primary care database[J]. *Br J Gen Pract*, 2012, 62(597): e243-250. DOI:10.3399/bjgp12X636065.
- [17] Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study[J]. *BMJ Open*, 2014, 4(8):e005809. DOI: 10.1136/bmjopen-2014-005809.
- [18] Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK[J]. *BMC Med*, 2016, 14:104. DOI: 10.1186/s12916-016-0650-2.
- [19] Alssema M, Newson RS, Bakker SJ, et al. One risk assessment tool for cardiovascular disease, type 2 diabetes, and chronic kidney disease[J]. *Diabetes Care*, 2012, 35(4):741-748. DOI:10.2337/dc11-1417.
- [20] Rauh SP, Rutters F, van der Heijden AAWA, et al. External validation of a tool predicting 7-year risk of developing cardiovascular disease, type 2 diabetes or chronic kidney disease[J]. *J Gen Intern Med*, 2018, 33(2):182-188. DOI: 10.1007/s11606-017-4231-7.
- [21] Asgari S, Moosaie F, Khalili D, et al. External validation of the European risk assessment tool for chronic cardio-metabolic disorders in a Middle Eastern population[J]. *J Transl Med*, 2020, 18(1): 267. DOI: 10.1186/s12967-020-02434-5.
- [22] Nelson RG, Grams ME, Ballew SH, et al. Development of risk prediction equations for incident chronic kidney disease[J]. *JAMA*, 2019, 322(21): 2104-2114. DOI: 10.1001/jama.2019.17379.
- [23] Chien KL, Lin HJ, Lee BC, et al. A prediction model for the risk of incident chronic kidney disease[J]. *Am J Med*, 2010, 123(9):836-846. DOI: 10.1016/j.amjmed.2010.05.010.
- [24] O'Seaghdha CM, Lyass A, Massaro JM, et al. A risk score for chronic kidney disease in the general population[J]. *Am J Med*, 2012, 125(3):270-277. DOI:10.1016/j.amjmed.2011.09.009.
- [25] 孙凤, 郁凯, 陶庆梅, 等. 中国台湾省 35~74 岁体检人群慢性肾病 5 年发病风险预测模型[J]. *中国预防医学杂志*, 2013, 14(1): 47-52. DOI: 10.16506/j.1009-6639.2013.01.006.
- [25] Sun F, Yu K, Tao QM, et al. Estimated risks on developing chronic kidney disease within 5 years for adults aged 35-74 based on Taiwan MJ health-checkup Database[J]. *Chin Prev Med*, 2013, 14(1): 47-52. DOI: 10.16506/j.1009-6639.2013.01.006.
- [26] Wen JP, Hao J, Zhang Y, et al. Risk scores for predicting incident chronic kidney disease among rural Chinese people: a village-based cohort study[J]. *BMC Nephrol*, 2020, 21(1):120. DOI:10.1186/s12882-020-01787-9.
- [27] Yu YT, Zhao Q, Jiang YG, et al. Prediction models and nomograms of 3-year risk of chronic kidney disease in China: a study from the Shanghai Suburban Adult Cohort and Biobank (2016-2020) [J]. *Ann Transl Med*, 2021, 9(22):1690. DOI:10.21037/atm-21-5647.

- [28] Zhang LX, Zuo L, Xu GB, et al. Community-based screening for chronic kidney disease among populations older than 40 years in Beijing[J]. *Nephrol Dial Transplant*, 2007, 22(4):1093-1099. DOI:10.1093/ndt/gfl763.
- [29] Wen CP, Cheng TYD, Tsai MK, et al. All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan[J]. *Lancet*, 2008, 371(9631): 2173-2182. DOI: 10.1016/S0140-6736(08)60952-6.
- [30] Levin A, Stevens PE, Bilous RW, et al. Kidney disease: improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease[J]. *Kidney Int Suppl*, 2013, 3(1):1-150. DOI:10.1038/kisup.2012.73.
- [31] 邱玲, 国秀芝. 肾小球滤过率评估方程的研究现状和面临的挑战[J]. *中华检验医学杂志*, 2014, 37(6):406-409. DOI: 10.3760/cma.j.issn.1009-9158.2014.06.003.
- Qiu L, Guo XZ. The status of research on estimated glomerular filtration rate and its challenges[J]. *Chin J Lab Med*, 2014, 37(6):406-409. DOI:10.3760/cma.j.issn.1009-9158.2014.06.003.
- [32] Taal MW. Chronic kidney disease: towards a risk-based approach[J]. *Clin Med (Lond)*, 2016, 16 Suppl 6:s117-120. DOI:10.7861/clinmedicine.16-6-s117.
- [33] Carville S, Wonderling D, Stevens P. Early identification and management of chronic kidney disease in adults: summary of updated NICE guidance[J]. *BMJ*, 2014, 349: g4507. DOI:10.1136/bmj.g4507.
- [34] Schaeffner ES, Ebert N, Delanaye P, et al. Two novel equations to estimate kidney function in persons aged 70 years or older[J]. *Ann Intern Med*, 2012, 157(7):471-481. DOI:10.7326/0003-4819-157-7-201210020-00003.
- [35] Björk J, Grubb A, Larsson A, et al. Accuracy of GFR estimating equations combining standardized cystatin C and creatinine assays: a cross-sectional study in Sweden [J]. *Clin Chem Lab Med*, 2015, 53(3): 403-414. DOI: 10.1515/cclm-2014-0578.
- [36] Pottel H, Hoste L, Dubourg L, et al. An estimated glomerular filtration rate equation for the full age spectrum[J]. *Nephrol Dial Transplant*, 2016, 31(5): 798-806. DOI:10.1093/ndt/gfv454.
- [37] Glasscock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls[J]. *Nat Rev Nephrol*, 2017, 13(2): 104-114. DOI: 10.1038/nrneph.2016.163.
- [38] Huang YM, Xu DM, Long JY, et al. Spectrum of chronic kidney disease in China: a national study based on hospitalized patients from 2010 to 2015[J]. *Nephrology (Carlton)*, 2019, 24(7):725-736. DOI:10.1111/nep.13489.
- [39] Moons KGM, 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):W1-73. DOI:10.7326/M14-0698.
- [40] 曹煜隆, 单娇, 龚志忠, 等. 个体预后与诊断预测模型研究报告规范——TRIPOD 声明解读[J]. *中国循证医学杂志*, 2020, 20(4):492-496. DOI:10.7507/1672-2531.201912032.
- Cao YL, Shan J, Gong ZZ, et al. Interpretation of the TRIPOD statement: a reporting guideline for multivariable prediction model for individual prognosis or diagnosis[J]. *Chin J Evid-Based Med*, 2020, 20(4): 492-496. DOI: 10.7507/1672-2531.201912032.
- [41] Shlipak MG, Tummala SL, Boulware LE, et al. The case for early identification and intervention of chronic kidney disease: conclusions from a Kidney Disease: improving global outcomes (KDIGO) controversies conference[J]. *Kidney Int*, 2021, 99(1):34-47. DOI:10.1016/j.kint.2020.10.012.
- [42] Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples[J]. *J Clin Epidemiol*, 2003, 56(5): 441-447. DOI: 10.1016/s0895-4356(03)00047-7.
- [43] Rysz J, Gluba-Brzózka A, Franczyk B, et al. Novel biomarkers in the diagnosis of chronic kidney disease and the prediction of its outcome[J]. *Int J Mol Sci*, 2017, 18(8): 1702. DOI:10.3390/ijms18081702.
- [44] Fox CS, Gona P, Larson MG, et al. A multi-marker approach to predict incident CKD and microalbuminuria[J]. *J Am Soc Nephrol*, 2010, 21(12): 2143-2149. DOI: 10.1681/ASN.2010010085.
- [45] Provenzano M, Serra R, Garofalo C, et al. OMICS in chronic kidney disease: focus on prognosis and prediction[J]. *Int J Mol Sci*, 2021, 23(1):336. DOI:10.3390/ijms23010336.
- [46] Grams ME, Coresh J. Assessing risk in chronic kidney disease: a methodological review[J]. *Nat Rev Nephrol*, 2013, 9(1):18-25. DOI:10.1038/nrneph.2012.248.
- [47] Yu Z, Jin J, Tin A, et al. Polygenic risk scores for kidney function and their associations with circulating proteome, and incident kidney diseases[J]. *J Am Soc Nephrol*, 2021, 32(12):3161-3173. DOI:10.1681/ASN.2020111599.
- [48] O'Seaghdha CM, Yang Q, Wu HS, et al. Performance of a genetic risk score for CKD stage 3 in the general population[J]. *Am J Kidney Dis*, 2012, 59(1): 19-24. DOI: 10.1053/j.ajkd.2011.08.030.
- [49] Köttgen A, Pattaro C, Böger CA, et al. New loci associated with kidney function and chronic kidney disease[J]. *Nat Genet*, 2010, 42(5):376-384. DOI:10.1038/ng.568.
- [50] 吴行伟, 刘馨宇, 龙恩武, 等. 机器学习在临床药物治疗中的研究进展[J]. *中国全科医学*, 2022, 25(2):254-258. DOI: 10.12114/j.issn.1007-9572.2021.01.309.
- Wu XW, Liu XY, Long EW, et al. Research progress of machine learning in clinical drug therapy[J]. *Chin Gen Pract*, 2022, 25(2): 254-258. DOI: 10.12114/j.issn.1007-9572.2021.01.309.
- [51] D'Ascenzo F, de Filippo O, Gallone G, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets[J]. *Lancet*, 2021, 397(10270): 199-207. DOI:10.1016/S0140-6736(20)32519-8.
- [52] Collins GS, Moons KGM. Reporting of artificial intelligence prediction models[J]. *Lancet*, 2019, 393(10181): 1577-1579. DOI:10.1016/S0140-6736(19)30037-6.
- [53] Krishnamurthy S, Kapelesh K S, Dovgan E, et al. Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan[J]. *Healthcare (Basel)*, 2021, 9(5):546. DOI:10.3390/healthcare9050546.
- [54] Luo WJ, Gong LL, Chen XJ, et al. Lifestyle and chronic kidney disease: a machine learning modeling study[J]. *Front Nutr*, 2022, 9: 918576. DOI: 10.3389/fnut.2022.918576.
- [55] Gulamali FF, Sawant AS, Nadkarni GN. Machine learning for risk stratification in kidney disease[J]. *Curr Opin Nephrol Hypertens*, 2022, 31(6): 548-552. DOI: 10.1097/MNH.0000000000000832.