

DOI:10.13288/j.11-2166/r.2022.16.002

基于体病相关探讨临床预测模型的构建

吴涛涛¹, 王琦², 李岩¹, 李文乐¹, 高惠贤¹, 崔殿坤¹, 李英帅^{2✉}

1. 北京中医药大学中医学院, 北京市朝阳区北三环东路 11 号, 100029; 2. 北京中医药大学国家中医体质与治未病研究院

[摘要] 疾病的早期诊断、早期治疗、早期预防对提高人群生存质量、减轻医疗负担至关重要, 而建立相关疾病的预测模型对此有重要实践意义。通过对中医体质学体病相关理论和临床预测模型原理的探讨、体病相关流行病学文献的检索和对体病相关预测模型的案例分析, 认为以中医体质为预测变量构建相关疾病的临床预测模型是可行的, 并通过总结体病预测模型的相关研究现状, 分析了在构建体质相关疾病预测模型时应注意确立研究问题、选择研究对象、筛选预测变量、控制量表测量偏倚、体质变量赋值等方面的问题。

[关键词] 临床预测模型; 中医体质学; 体病相关

随着现代生活方式和饮食习惯等诸多方面的改变, 人的体质发生了很大的变化, 随之而来的慢性非传染性疾病、代谢相关疾病的发病率也在日益攀升, 已成为影响人民健康的公共卫生问题^[1-2]。在卫生保健中, 预测模型是为了预测一个个体是否存在一个特定的临床结局, 如是否患某种疾病(诊断模型), 或将来是否会发生某种临床结局(预后模型)^[3]。临床预测模型广泛应用于医疗决策、医疗质量管理及医疗资源配置。中医体质学认为, 体质是疾病发生的土壤, 在疾病发生发展过程中发挥了重要作用^[4]。针对疾病高危体质人群, 建立相关疾病的数学诊断与预后模型, 可实现疾病的早预防、早诊断、早治疗, 对降低相关疾病发病率、改善生活质量具有重要实践意义。

1 基于体病相关构建临床预测模型的可行性

1.1 体质是预测疾病发生发展的重要因素

中医体质学认为, 体质受先天遗传和后天获得影响, 是人类在生长、发育过程中形成的与自然、社会环境相适应的人体个性特征, 表现为形态结构、生理功能和心理状态方面综合的、相对稳定的固有特质^[4]。体质的差异性可导致个体对致病因素有着不同的易感性、易罹性及发病倾向性。如痰湿

质易发代谢综合征、肥胖等^[5]。此外, 体质因素影响病机, 如平和质者, 感受寒邪则为寒病, 感受湿邪则为湿病; 阴虚质者受邪后多从热化, 阳虚质者受邪后多从寒化^[4]。因此, 体质状态是预测疾病发生发展的重要因素。

1.2 中医体质可作为临床预测模型的预测变量

预测变量是指能够预测现有临床结局或临床事件将来发生与否的变量, 一般指疾病的危险因素。预测变量的类别范围可以从人群流行病特征、病史和体格检查结果, 延伸到影像学检查结果, 电生理检查, 血、尿或机体组织检测和疾病分期/特征, 再到组学和其他新兴生物测量结果。预测变量应当具有低偏倚风险和良好的适用性^[6]。

中医体质不同于传统的预测变量, 其通过中医体质量表(CCMQ)测量得出。中医体质量表^[7]按照中医体质理论和中医体质类型设计, 包括了形体特征、心理特征、常见表现、对外界环境的适应能力和发病倾向五个方面, 条目代表性好、独立性强、敏感性高, 能够体现人群的生理、心理特征及发病倾向。体质评分作为预测变量, 具备以下三个优点: 1) 与结局密切相关, 预测性能好。体质是疾病发生的土壤, 体质状态是预测疾病发生发展的重要因素, 如痰湿体质易患冠心病、糖尿病^[8-9]等, 这些疾病是体质评分作为预测变量的应用重点。有研究^[10]显示, 在西医常规危险因素的基础上, 加入中医体质, 能使模型的预测效能显著增

基金项目: 中国工程院战略研究与咨询项目(2021-XBZD-1); 北京中医药大学一流学科建设項目

✉ 通讯作者: liyingshuai2013@163.com

加。2) 量表效度较好, 测量稳定。朱燕波等^[11]以中医体质量表与健康状况调查问卷 SF-36 的相关系数评价中医体质量表的效标效度, 结果表明中医体质量表效标效度和区分效度良好。对其中 130 例进行重复测量的结果显示, 中医体质量表的重测信度好, 测量结果稳定。3) 测量方便, 成本不高, 接受度高。体质量表共包含 60 个条目, 可通过纸质版或手机、计算机等 APP、网页等多种形式作答, 量表的回答时间短, 回答率高^[11]。综上, 中医体质作为预测变量具有低偏倚风险和良好的适用性。

1.3 体病相关的流行病学研究可为模型构建提供现有数据

模型构建的数据来源分为现有数据和研究数据, 自 2009 年《中医体质分类与判定》^[12] 标准发布以来, 研究者们采用这一标准实施了大量的流行病学研究, 目前已经积累了丰富的体病相关数据。检索中国知网 (CNKI)、万方数据库 (WF)、维普数据库 (VIP)、PubMed 和 Embase 数据库收录的建库至 2019 年 2 月 25 日发表的体病相关研究, 共得到 1441 篇文献, 包括 1300 篇横断面研究、81 篇病例对照研究和 60 篇队列研究^[13]。其中横断面研究数据可用于构建诊断模型, 队列研究数据可用于构建预后模型。

2 体病预测模型相关研究现状

2.1 体病预测模型研究的基本特点

在 CNKI 以“中医体质”“预测模型”为关键词检索得到 4 项将体质作为预测变量构建预测模型的研究^[10,14-16], 其中 3 项为诊断模型研究, 1 项为预后模型研究。涉及的疾病均为与体质密切相关的疾病, 如代谢综合征、糖尿病、缺血性中风等。研究对象均来自医院或诊所的就诊患者和体检人群, 其中诊断模型的数据来源于病例对照研究及健康体检数据库, 预后模型的数据来源于注册登记式研究。4 项研究体质变量的赋值方法也有不同: 2 项研究^[15-16]纳入的研究对象均为痰湿体质人群, 所以使用了痰湿体质得分来作为预测变量; 1 项研究^[14]将九种体质编码为 0~8 九个数字; 1 项研究^[10]中并未报告体质变量赋值方法。危险因素的筛选和模型的拟合均使用了 Logistic 回归分析方法, 此外, 1 项研究^[16]还使用了 Bayes 判别分析法和支持向量机法拟合预测模型。4 项研究均使用内部验证的方法评估模型, 其中 1 项研究^[16]使用 80% 数据拟合

模型, 20% 数据验证模型; 1 项研究^[14]按年份划分训练集和验证集; 另外 2 项研究^[10,15]没有划分训练集和验证集, 使用全部数据进行模型的拟合和验证。4 项研究均使用区分度和校准度指标来评价模型, 其中 2 项研究^[14,16]还使用了灵敏度、特异度等指标, 结果显示各指标均表现良好。

2.2 体病预测模型研究的优点与不足

4 项研究均纳入中医体质这一无创、可快速获得的因素作为预测变量, 融合了中医特色, 且模型的区分度与校准度等指标都表现优秀, 说明预测性能良好。4 项研究使用到了 Logistic 回归分析方法、Bayes 判别分析法和支持向量机法构建模型。Logistic 逐步回归模型、Bayes 判别分析模型使用数学公式组合展示, 解释性强、易于理解; 支持向量机模型使用机器语言展示, 解释性和理解性较差, 但判断准确率更高。与既往只使用一种方法开发模型相比, 应用范围和前景广阔, 更具有使用价值。

目前的研究也存在一定的局限性, 表现在: 1) 研究对象均为就诊患者, 纳入病例不能代表全部病例, 难以避免选择偏倚, 会导致应用场景有限。2) 部分信息需要研究对象通过回忆既往暴露史获取, 难以避免回忆偏倚, 影响数据的可靠性。3) 2 项研究^[14,16]划分了训练集和验证集, 但并未使用如随机拆分验证、交叉验证以及 Bootstrap 重抽样等常用的内部验证方法, 划分依据可能缺乏科学性; 2 项研究^[10,15]未划分训练集和验证集, 使用全部数据拟合和验证模型, 虽然可以最大程度地利用样本, 但可能导致模型稳定性变差。当应用场景发生变动, 模型的预测能力可能下降。4) 研究成果的报告未完全遵循《个体预后与诊断的多变量预测模型透明报告》(TRIPOD)^[17], 完整性有所欠缺。如 3 项研究^[10,14-15]未报告缺失值的处理方法。虽说处理缺失值最好的方法是防止出现缺失值, 但缺失值是任何研究都无法回避的问题。当缺失的样本例数大时, 直接剔除不仅可能引入选择偏倚, 而且导致信息丢失, 样本量减少, 把握度下降。

3 基于体病相关论构建临床预测模型应注意的问题

3.1 确立研究问题

临床预测模型回答的是疾病的诊断或预后问题, 确立研究问题是构建临床预测模型的第一步。研究问题一定要和临床实践相联系, 一般来说, 只有那些给社会和个人带来严重负担, 且可通过相关

手段有效预防的疾病,建立的预测模型才能在临床实践和公共卫生领域中得以广泛应用^[18];而体质是多种慢病如高血压、糖尿病等疾病发生的“土壤”,通过调理偏颇体质可以实现未病先防、欲病早治、既病防变^[19-20]。因此,在建立体病相关的预测模型时,应选择具有重大临床价值且当体质作为预测变量时预测性能好的疾病,即某种体质的好发疾病。如阴虚质、气虚质、痰湿质是罹患糖尿病的体质危险因素,因此可以阴虚质、气虚质、痰湿质这三种体质为预测变量构建糖尿病的诊断模型^[14];痰湿质和血瘀质是缺血性中风病复发的体质危险因素,因此可以痰湿质和血瘀质为预测变量来构建缺血性中风病复发的预后模型^[10]。

3.2 选择研究对象

研究对象的选取应该根据研究问题来确定。诊断模型的研究目的是基于研究对象的临床症状和特征估计当前发生某临床结局的概率,多见于横断面研究;预后模型的研究目的是估计在目前健康状况下(健康或患病),未来一段时间内发生某临床结局(发病、复发、死亡,伤残及并发症)的概率,多见于队列研究^[21]。在构建体病相关的预测模型时,除了一般预测模型的纳入、排除标准外,应对受试者的体质类型加以限定。当某疾病的高危体质只有一种,研究目的是预测该特定体质类型发生某临床结局的风险时,研究对象只纳入该体质类型;当某疾病的高危体质有多种,或者研究目的是评估每一种体质发生某临床结局的风险时,则可纳入所有体质类型人群。

3.3 筛选预测变量

预测模型的变量筛选,应结合文献报道、统计学方法以及专家经验。此外,在确定预测模型的预测变量时,一些实际的因素如指标测量的难易度、测量成本、应用的难易度等也应考虑在内。合适的筛选策略可以避免模型过度拟合以及预测变量的筛选偏倚。在构建体病相关的预测模型前,研究者应该系统检索相关文献,收集整理已报道的高危体质类型作为备选预测变量;然后通过统计方法,并结合医学认识和专家经验,从备选变量中选出最终纳入模型的预测变量。9种体质类型可作为9个预测变量,但是一般来说,疾病的高危体质类型才可纳入预测模型,如代谢综合征诊断模型仅将9种体质中的痰湿质作为预测变量^[15-16],糖尿病诊断模型将9种体质中的阴虚质、气虚质、痰湿质作为预测

变量^[14]。

3.4 控制量表测量偏倚

常见的预测变量多由仪器或设备测量得出,测量精确,偏倚较小。而中医体质量表为自填量表,由于受试者可能存在文化程度低、思考问题的角度偏差及依存性差等问题,会出现自我评价的结果与实际情况有所出入,导致测量偏倚。因此,在填写量表时,可以由专业人员辅助,对一些难以理解的条目进行解释,但不能主观引导受试者的选择。此外,应当进一步修订完善中医体质量表,使得条目尽量口语化、简单易懂^[22]。

3.5 体质变量赋值

预测变量一般包括连续变量和分类变量。分类变量一般用数字0~9来赋值。连续变量的赋值比较复杂,如果连续变量和结果之间的关系是线性的,则可以在回归方程中包含连续变量。如果没有,可以将其转换为二分变量或有序分类变量,然后将它们放入回归方程中,但应尽可能将原始的连续变量包含在回归模型中^[23]。体质类型为分类变量,体质评分为连续变量。在构建体病相关的预测模型时,若体质评分与发生结局的风险呈线性关系,应使用中医体质量表测得的该体质评分来赋值;否则,体质变量应当按照分类变量来赋值,如9种体质类型可赋值为数字0~8。

4 总结与展望

中医体质是预测疾病发生发展的重要因素,目前已有许多关于疾病预测模型的研究成果,但把中医体质作为预测变量的疾病预测模型很少。一方面体病相关的流行病学文献虽然很多,但高质量研究较少^[13];另一方面,由于体质量表本身的局限性,导致将体质作为预测变量仍存在一定的不足。目前临床上广泛使用的中医体质量表为60条目完整版,作答比较费时;同时体质量表属于自评量表,评分带有主观成分,可能导致测量结果存在偏差,有一定的偏倚风险^[22,24]。但需要指出的是,这种主观性和不确定性主要来源于调查过程中研究对象填写量表的认知和态度,而非量表本身^[25],未来希望能使用可靠的简化量表^[26-28]或找到能代表体质特征的标志性指标。同时应开展高质量、大样本的临床研究来为体质相关疾病预测模型的构建提供数据支撑;并对已建立的模型采用不同时间、不同地域的数据集来进行外部验证,进一步优化、完善预测模

型。以此来进一步充实体质三级预防体系,更好地实现疾病的“早筛查、早预警、早预防”。

参考文献

- [1]胡建平,饶克勤,钱军程,等.中国慢性非传染性疾病经济负担研究[J].中国慢性病预防与控制,2007,16(3):189-193.
- [2]中华医学会糖尿病学分会代谢综合征研究协作组.中华医学会糖尿病学分会关于代谢综合征的建议[J].中华糖尿病杂志,2004,12(3):5-10.
- [3]RANSTAM J, COOK JA, COLLINS GS. Clinical prediction models[J]. Br J Surg, 2016, 103 (13):1886.
- [4]王琦.中医体质学[M].北京:人民卫生出版社,2008:44.
- [5]李竹青,秦静波,孟翔鹤,等.辨痰湿体质论治体病相关疾病的临床思路[J].天津中医药,2020,37(10):1142-1146.
- [6]WOLFF RF, MOONS KGM, 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.
- [7]王琦,朱燕波,薛禾生,等.中医体质量表的初步编制[J].中国临床康复,2006,10(3):12-14.
- [8]王前飞,王前奔,王琦.肥胖人痰湿型体质变异的分布及其与冠心病的关系[J].云南中医杂志,1992,13(2):4-6.
- [9]刘艳骄,王琦.肥胖人痰湿体质与糖尿病相关性研究[J].山东中医学院学报,1993,17(2):34-39.
- [10]王桂倩,谢雁鸣,易丹辉,等.基于体质的缺血性中风病复发风险评估模型建立研究[J].中华中医药杂志,2019,34(10):4860-4866.
- [11]朱燕波,王琦,折笠秀树.中医体质量表的信度和效度评价[J].中国行为医学科学,2007,16(7):651-654.
- [12]中华中医药学会.中医体质分类与判定[J].世界中西医结合杂志,2009,4(4):303-304.
- [13]罗辉.中医体质学体病相关临床研究的系统评价和方法学研究[D].北京:北京中医药大学,2019.
- [14]张颖,章亦莹,杨瑞文,等.基于中医体质辨识的糖尿病风险评估模型建立与验证[J].中国全科医学,2020,23(10):1261-1266.
- [15]沈昆.成年女性痰湿体质人群易发代谢综合征的影响因素及预警模型构建研究[D].北京:北京中医药大学,2019.
- [16]梁雪.痰湿体质人群易发代谢综合征的危险因素筛选和诊断模型构建研究[D].北京:北京中医药大学,2017.
- [17]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):W1-73. doi: 10.7326/M14-0698.
- [18]张蕊,郑黎强,潘国伟.疾病发病风险预测模型的应用与建立[J].中国卫生统计,2015,32(4):724-726.
- [19]王琦.痰湿体质系列研究及在代谢性慢病防控中的应用[J].天津中医药,2020,37(1):4-8.
- [20]方旖旎,王琦,张国辉,等.中医体质学在“治未病”中的应用研究[J].中医杂志,2020,61(7):581-585.
- [21]谷鸿秋,周支瑞,章仲恒,等.临床预测模型:基本概念、应用场景及研究思路[J].中国循证心血管医学杂志,2018,10(12):1454-1456,1462.
- [22]李英帅,杨寅,李玲孺,等.中医体质量表应用中的疑难问题解读[J].中医杂志,2015,56(10):844-846.
- [23]ZHOU ZR, WANG WW, LI Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R[J]. Ann Transl Med, 2019, 7(23):796.
- [24]辛海,吴剑坤,郭昕,等.中医九型体质量表临床应用中的问题探讨[J].中华中医药杂志,2014,29(6):1841-1843.
- [25]罗辉,王琦.中医体质与疾病相关性临床研究的方法学挑战和设计实施建议[J].中医杂志,2020,61(1):20-26.
- [26]朱燕波,虞晓含,王琦,等.简短版中医体质量表的初步设置与考评[J].中国全科医学,2017,20(7):879-885.
- [27]朱燕波,王琦,虞晓含,等.中医体质量表-41条目简短版的结构效度和反应度评价[J].中国全科医学,2017,20(26):3282-3286.
- [28]朱燕波,王琦,史会梅,等.中医体质量表30条目简短版的制定与评价[J].中医杂志,2018,59(18):1554-1559.

(下转第1531页)