

论著

文章编号:1000-5404(2013)11-1144-04

应用分类树模型构建糖尿病肾病蛋白尿进展的风险预测模型

赵文波,李明,唐 骅,司美君,刘 迅,娄探奇 (510630 广州,中山大学附属第三医院肾内科)

[摘要] **目的** 应用分类树模型构建糖尿病肾病由微量蛋白尿到大量蛋白尿进展风险的预测模型,并评价其应用价值。**方法** 选取本院临床资料完善的糖尿病肾病住院患者 256 例,根据 GFR 分期及尿白蛋白定量诊断分为糖尿病微量蛋白尿组(183 例)和糖尿病大量蛋白尿组(73 例)。记录患者的相关临床资料,采用 Exhaustive CHAID 分类树算法建立糖尿病肾病进展风险的预测模型,采用错分概率 Risk 值、索引图评价模型的应用价值。**结果** 所建立的分类树模型包括 3 层,共 11 个结点,共筛选出 4 个解释变量:胱抑素水平、高血压病程、腰臀比、白蛋白水平;其中最为重要的预测因素为血清胱抑素水平和腰臀比水平。模型错分概率 Risk 值为 0.141,模型拟合的效果较好。**结论** 分类树模型不仅能有效地拟合糖尿病肾病由微量蛋白尿向大量蛋白尿进展的风险预测,还可以有效地筛选变量间的交互作用效应。

[关键词] 糖尿病肾脏疾病;分类树;危险因素;预测模型

[中图分类号] R181.23; R587.2; R696.2

[文献标志码] A

Construction of a risk prediction model for proteinuria progression in diabetic nephropathy by classification tree

Zhao Wenbo, Li Ming, Tang Hua, Si Meijun, Liu Xun, Lou Tanqi (Department of Nephrology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong Province, 510630, China)

[Abstract] **Objective** To establish a risk prediction model for progression of microalbuminuria to proteinuria in diabetic nephropathy, and to evaluate its value for type 2 diabetes patients with early renal damage. **Methods** Totally 256 patients with confirmed type 2 diabetes who were inpatients or outpatients in our hospital from January to June 2012 were enrolled in this study. According to estimated glomerular filtration rate (eGFR) installments and urine albumin quantitative analysis, they were divided into proteinuria group ($n=73$) and early diabetic kidney damage group ($n=183$). The clinical data of the patients were recorded to analyze the main factors for the progression of diabetic nephropathy using the Exhaustive CHAID classification tree algorithm. The value of the established model was evaluated by the Risk statistics and index map. **Results** The model had 3 stratum and 11 nodes. There were 4 explanatory variables screened out in the model, that is, CysC level, history of hypertension, waist-hip ratio and albumin level. The most important risk factors were CysC level and waist-hip ratio. The risk value of misclassification probability of the model was 0.141, and the classification tree model fitted the actuality very well. **Conclusion** Classification tree model can not only properly predict the progression of microalbuminuria to proteinuria in diabetic nephropathy, but also reveal the complex interaction effects among the factors.

[Key words] diabetic nephropathy; classification tree; risk factors; prediction model

Supported by the National Key Technology R&D Program(2011BAI10B05) and the Project of Sci & Tech Research of Guangdong Province (2009B030801196). Corresponding author: Lou Tanqi, Tel: 86-20-85252865, E-mail: Lou.tq@163.com

2007 年美国国家肾脏基金会(NKF)制定的“KDOQI 糖尿病和慢性肾脏病临床实践指南和推

荐”^[1]提出糖尿病肾脏疾病(diabetic kidney disease, DKD)取代以往糖尿病肾病(diabetic nephropathy, DN),并提出应用尿蛋白结合 MDRD 公式评估的肾小球滤过率(estimated glomerular filtration rate, eGFR)对糖尿病患者进行评估。目前 DKD 患者较少临床行肾穿病理检查。既往多采用 Logistic 回归法对 DKD 危险因素进行分析,虽然对单个危险因素自身影响作用分

[基金项目] 国家科技支撑计划(2011BAI10B05);广东省科技计划项目(2009B030801196)

[通信作者] 娄探奇,电话:(020)85252865, E-mail: Lou.tq@163.com

[优先出版] <http://www.cnki.net/kcms/detail/51.1095.R.20130228.1715.012.html> (2013-02-28)

析充分,但不能充分分析危险因素间存在的交互作用,容易忽视危险因素间的层次关系。分类树模型能较好地分析危险因素存在的交互作用,并可分析各个危险因素影响的具体人群特征,能够提示具有何种特征的DKD患者群更容易出现蛋白尿增多、病情进展,从而可对该特征人群进行早期干预。分类树模型在疾病危险因素研究中已有报道^[2-3],刘建平^[4]用分类树构建了脑卒中的发病风险预测模型,但少见应用于DKD进展风险预测的报道。本研究分析2012年1-6月在本院内分泌科及肾内科住院或门诊随诊的2型糖尿病患者的一般临床资料和实验室检查资料,应用尿微量白蛋白测定结合eGFR对糖尿病患者进行评估,排除可能的非糖尿病肾脏疾病(non-diabetic renal disease, NDRD),使用SPSS 13.0统计软件中分类树模型,探讨糖尿病微量蛋白尿进展为大量白蛋白尿的相关危险因素,建立预测模型,并评价其应用价值。

1 资料与方法

1.1 研究对象

收集2012年1-6月在本院内分泌科及肾内科住院或门诊随诊的2型糖尿病患者,临床资料详细完整者256例,其中男性135例,女性121例,年龄20~91岁。分为微量白蛋白尿组183例,男性135例,女性121例,年龄(61.11±13.04)岁;大量白蛋白尿组73例,男性135例,女性121例,年龄(61.52±10.40)岁。2组年龄、性别差异无统计学意义($P>0.05$)。MDRD公式计算eGFR,微量白蛋白尿组为(119.60±49.31) mL/(min·1.73 m²)。大量白蛋白尿组为(56.89±39.52) mL/(min·1.73 m²)。诊断及入选标准:①按WHO 1999年糖尿病的诊断标准确诊^[5];②DKD诊断标准:NKF-KDOQI指南^[1]提出的按照eGFR和尿蛋白定量进行。糖尿病微量白蛋白尿诊断标准(DKD早期肾损害):eGFR>30 mL/(min·1.73 m²)并且伴有微量白蛋白尿为20~200 μg/min或微量白蛋白肌酐比值(mALB/Cr)为30~299 mg/g;大量白蛋白尿诊断标准:尿蛋白排泄率(UREA)>200 μg/min或mALB/Cr>300 mg/g,不论eGFR水平。排除标准:①1型糖尿病;②糖尿病急性并发症:糖尿病酮症、高渗性昏迷等;③原发性肾病或其他继发性肾脏疾病;④入院时并发影响尿检或肾功能的常见疾病;⑤NKF-KDOQI指南^[1]提出的按照eGFR和尿蛋白定量,无白蛋白尿,eGFR<60 mL/(min·1.73 m²);或微量白蛋白尿,eGFR<30 mL/(min·1.73 m²)。

1.2 观测指标

1.2.1 一般临床资料 性别、年龄、糖尿病病程、糖尿病家族史、饮酒史、吸烟史、收缩压、舒张压、身高、体质量、BMI、腰臀比、高血压病史。

1.2.2 并发疾病 包括糖尿病及其他并发症(视网膜病变、周围神经病变、周围血管病变、糖尿病足)以及脂肪肝、慢性胆囊炎、胆囊结石、冠心病、脑梗死病史。

1.2.3 实验室检查指标 取入院后次日清晨肌酐(Scr)、尿素氮(BUN)、尿酸(UA)、总胆固醇(TC)、甘油三酯(TG)、高密度脂蛋白(HDL)、低密度脂蛋白(LDL)、脂蛋白a(Lpa)、胱

抑素(CysC)、血红蛋白(HGB)、糖化血红蛋白(HbA1c)、空腹血糖(FPG)、餐后2h血糖(2hPG)以及尿蛋白排泄率(24h尿)或晨尿微量白蛋白/肌酐。

1.3 研究方法

采用横断面研究,以内分泌科、肾内科住院或门诊随诊的2型糖尿病患者住院病历为原始资料,对临床资料、并发疾病情况进行数据录入,分析糖尿病肾损害进展的相关危险因素,并建立风险模型。

1.4 统计学分析

采用SPSS 13.0统计软件进行Exhaustive CHAID分类树模型分析,模型中因素数值的拆分与合并的显著性检验标准均为0.05,最大生长深度设定为3层,为充分发现影响糖尿病肾损害蛋白尿进展的潜在危险因素及危险因素间的关系,设定父结点最小样本含量为20,子结点为10。

2 结果

2.1 模型的建立

糖尿病早期肾脏损害发病风险模型所建立分类树模型包括3层,共11个结点,其中终末结点7个(图1),共筛选出4个解释变量:血清CysC、高血压病程、腰臀比、血ALB水平(P 值分别是0.000、0.002、0.002和0.005)。

2.2 模型的分析

从图1可见,第1层是按照研究对象血清CysC进行拆分,所以,本组中发病风险影响最大的危险因素为血清CysC水平,其中,CysC>1.53 mg/L组的糖尿病肾损害进展为大量白蛋白尿发生概率明显高于CysC≤1.13 mg/L组及CysC为>1.13~1.53 mg/L组,是进展的高危人群。而在具有不同CysC水平的人群中筛选出糖尿病蛋白尿进展的危险因素也不相同,由CysC≤1.13 mg/L组为终末结点、>1.13~1.53 mg/L组、>1.53 mg/L组分别是高血压病程和腰臀比。在CysC>1.53 mg/L人群中,如果同时具有腰臀比升高会增加糖尿病肾损害进展的风险,腰臀比>0.99组出现蛋白尿的发生概率明显高于腰臀比>0.87~0.99组和腰臀比≤0.87组,成为主要的影响糖尿病肾损害进展的危险因素。腰臀比>0.87~0.99且血ALB水平≤39.5 g/L组患病概率显著增加。而对CysC>1.13~1.53 mg/L组,分析其高血压病程的影响,病程为≤3年组蛋白尿进展的风险比率高于无高血压组及高血压病程>3年组。

根据CHAID分析结果可知:血CysC升高且存在腰臀比升高的患者患病概率显著增加,是DKD蛋白尿进展的高危人群,而血ALB水平≤39.5 g/L的群体风险也显著增加,高血压病程对病情进展也影响较大。

2.3 模型的结点增益分析

本研究选择蛋白尿的进展为有效的分类。表1显示了各结点的病例的百分数,索引值指目标分类在该结点的响应百分比除以总样本根结点响应百分比。索引值>100%表示各结点的病例组的患病率大于根结点的病例组的患病率。由此,可以看到结点9、10、5的发病率显著高于群体患病率,发病风险更显著,提示这些结点的相关影响因素应重点关注。模型结点增益图显示模型良好,从0%开始向100%突增,到100%结束,如模型未提供任何信息则为对角参考线。见图2。

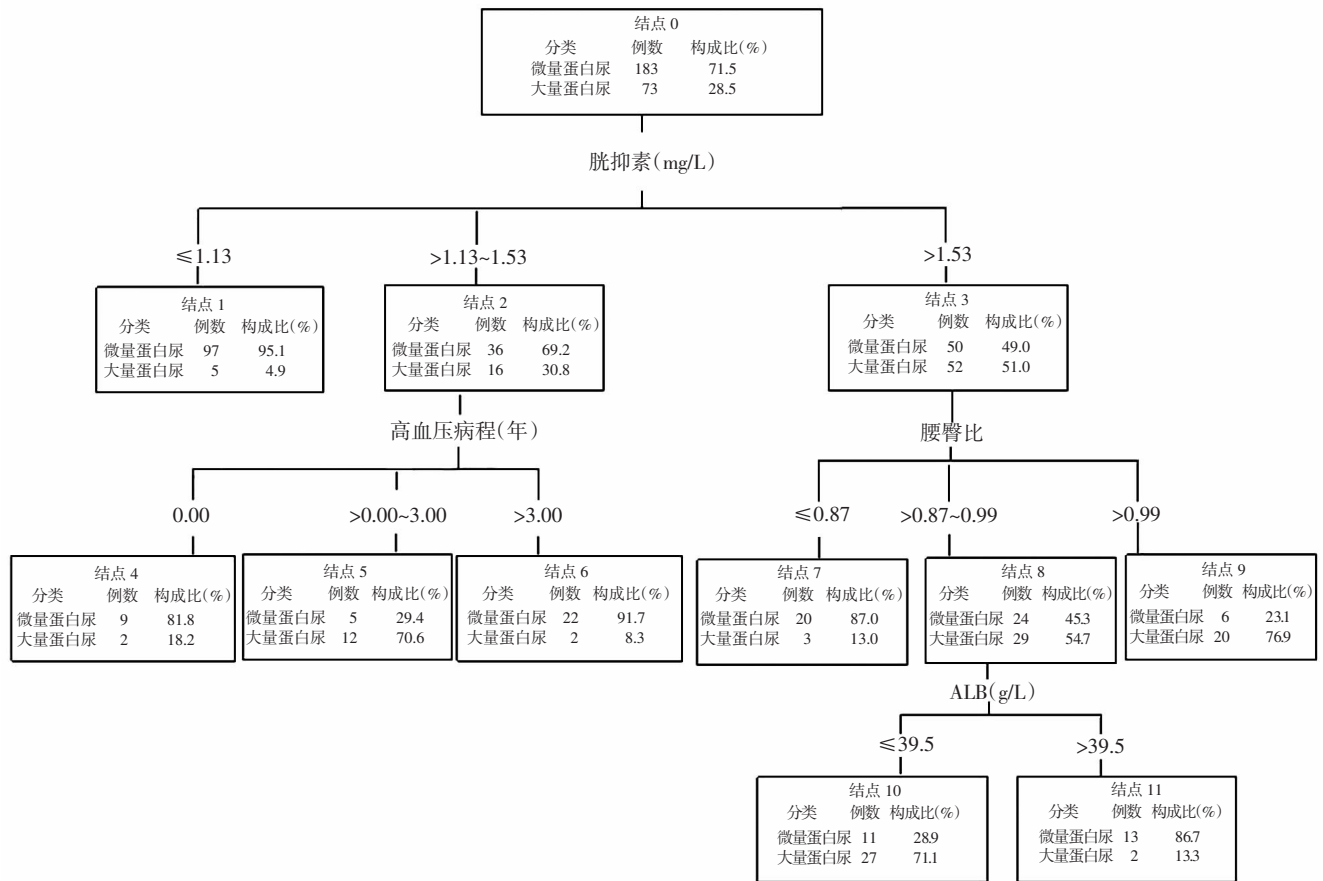


图1 危险因素的分类树状图

表1 分类树模型的结点增益

结点	结点		增益		响应百分比(%)	索引值(%)
	例数	构成比(%)	例数	构成比(%)		
9	26	10.2	20	27.4	76.9	269.8
10	38	14.8	27	37.0	71.1	249.2
5	17	6.6	12	16.4	70.6	247.5
4	11	4.3	2	2.7	18.2	63.8
11	15	5.9	2	2.7	13.3	46.8
7	23	9.0	3	4.1	13.0	45.7
6	24	9.4	2	2.7	8.3	29.2
1	102	39.8	5	6.8	4.9	17.2

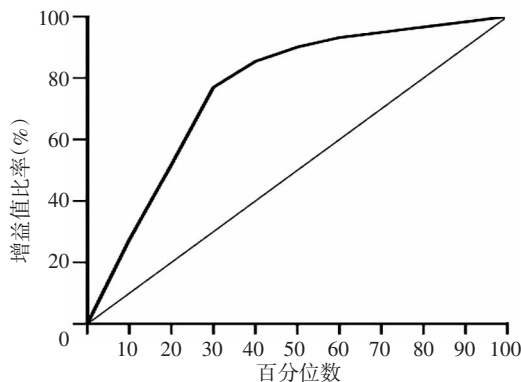


图2 大量蛋白尿结点增益曲线

2.4 分类树模型评价

2.4.1 索引图评价 索引图用来评价建立模型的优劣,索引值应该从100%以上开始,沿近似平行移动一段,再快速向100%方向下降。由图3可知,本研究建立的模型较好,从高于100%开始移动相对平稳的一段后骤降到100%。

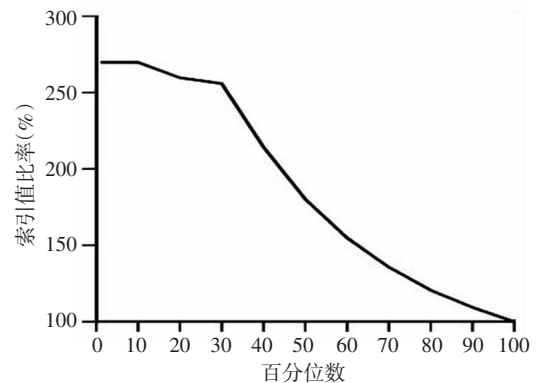


图3 微量蛋白尿累计指数曲线

2.4.2 错分矩阵与 Risk 统计量评价 树型模型 Risk 统计量为0.141,即错估概率为14.1%,表示使用该模型对糖尿病微量蛋白尿向大量蛋白尿进展风险的预测正确率为85.9%,模型拟合效果较好。分类表预测结果与风险表一致,为85.9%,显示该模型依靠危险因素对85.9%出现大量蛋白尿患者进行了正确地分类,而对微量蛋白尿患者分类正确率也较高,为88.0%。

3 讨论

DKD 的发生、发展是多个因素共同作用的结果, NKF-KDOQI 指南^[1]也建议多个因素联合评价。既往对多危险因素影响的疾病研究多使用 Logistic 回归、Cox 回归等方法,但由于各个因素间相互关联,往往存在共线性,会影响统计分析的结果,因此,非参数回归是常用的解决方法,如分类树模型,其对要预测的变量数据也无要求,近年来国外应用较多^[6-7]。分类树分析过程包括“种树”和“剪枝”,经过修剪后成为最优化的模型,并且以树型图的方式展示出分析的过程和相互作用的各个危险因素间的关系,而常用的分析方法,如 Logistic 回归等,则很难做到。

本研究纳入的均为长期随访的患者,糖尿病的治疗按照中国 2 型糖尿病防治指南(2010 版)及美国糖尿病协会 ADA 指南执行,对血糖状况进行管理。患者由专科护士随访,对其进行饮食、运动方面的指导。住院期间患者的饮食由膳食营养科按糖尿病饮食要求配送。本研究对糖尿病肾脏损害蛋白尿进展的主要危险因素使用分类树构建预测模型,可以直接从 34 个候选变量中筛选出 4 个重要解释变量,结果显示血清 CysC 水平、腰臀比、高血压病程、血清白蛋白水平与糖尿病肾损害的蛋白尿进展相关。对于离散化的连续性变量和具有 2 种以上分类的指标,应用 Exhaustive CHAID 法构建的分类树模型会将没有统计学意义的分层类别重组为具有统计学意义的新类别,这比人为地设定分层因素更科学合理^[8],因此,分类树模型在确定危险因素的数值分界点上更为方便。本研究将危险因素(如血清 CysC)重新划分为 3 个分类等级,结果表明, CysC 高的个体糖尿病肾损害发病风险显著增高。该界点的确定不是人为划分,而是依靠统计学显著意义划分,并各自作出深入地分析。有研究也显示在不同的白蛋白尿组 CysC 的水平不同^[9-11],使用 CysC 可以作为评价 eGFR 的指标,反映了 CysC 作用的重要性。从增益表可以看到结点 9、10、5、4 的风险显著高于结点 1 中纳入的群体,蛋白尿进展的可能性更显著。结点 9 表明,糖尿病微量白蛋白尿患者 CysC > 1.53 mg/L 并腰臀比 > 0.99 组是发生病情进展出现大量白蛋白尿高危风险人群;结点 10 表明, CysC > 1.53 mg/L 并伴有 ALB ≤ 39.5 g/L 组病情进展风险也显著增加;结点 5 表明, CysC > 1.13 ~ 1.53 mg/L 并高血压病程 ≤ 3 年组病情进展风险显著增加,高于无高血压组和高血压病程 > 3 年组。考虑原因可能是高血压病程较长组患者对血压重视程度要高于高血压病程 ≤ 3 年组,而且可能发现蛋白尿后服用了血管紧张素转化酶抑制剂(angiotensin-converting enzyme inhibitor, ACEI)或者血管紧张素 II 受体拮抗剂(angiotensin II receptor antago-

nist, ARB) 类的药物可导致血压影响的比例下降,而高血压病程 ≤ 3 年组比例较高,考虑为血压本身对蛋白尿的影响大,患者对自身重视不够、服药不规律及未服用或未规范服用 ACEI 或者 ARB 类的药物所致。采用 Risk 统计量对模型的预测进行评价,结果显示使用该模型对糖尿病微量蛋白尿向大量白蛋白尿进展风险预测的正确率为 85.9%。错分矩阵分类表的预测结果与风险表一致,为 85.9%,该模型依靠危险因素对 85.9% 出现大量白蛋白尿患者进行了正确地分类,而对微量白蛋白尿患者分类正确率也较高(88.0%),说明模型预测效果好。

分类树模型可以用于分析复杂疾病,是多因素同时作用时的有力分析工具,可以应用到 DKD 发病危险因素及病情进展的探讨,为制定预防、治疗方案提供依据。本模型选取住院患者为样本,对住院患者病情进展预测结果较好,但同时存在一定的人群局限性,并且样本量也需进一步增大。

参考文献:

- [1] KDOQI. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Diabetes and Chronic Kidney Disease[J]. Am J Kidney Dis, 2007, 49(2 Suppl 2): S12 - S154.
- [2] 张勇晶, 陈坤, 金明娟, 等. 应用分类树模型筛选恶性肿瘤危险因素的研究[J]. 中华流行病学杂志, 2006, 27(6): 540 - 543.
- [3] 石玲, 王燕. 婴幼儿死亡危险因素的研究-兼论 CHAID 方法的原理及应用[J]. 中国卫生统计, 2002, 19(5): 283 - 285.
- [4] 刘建平, 程锦泉, 张仁利, 等. 应用分类树模型构建缺血性脑卒中发病风险的预测模型[J]. 中国慢性病预防与控制, 2012, 20(3): 254 - 258.
- [5] Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus[J]. Diabetes Care, 2000, 23 Suppl 1: S4 - S19.
- [6] Crocetti E, Mangone L, Lo-Scooco G, et al. Prognostic variables and prognostic groups for malignant melanoma. The information from Cox and Classification And Regression Trees analysis: an Italian population-based study[J]. Melanoma Res, 2006, 16(5): 429 - 433.
- [7] Binongo J N, Taylor A, Hill A N, et al. Use of classification and regression trees in diuresis renography[J]. Acad Radiol, 2007, 14(3): 306 - 311.
- [8] Breiman L, Friedman J H, Olshen R A, et al. Classification and regression trees[M]//Venables W N, Ripley B D, eds. Modern Applied Statistics with S-plus. 2nd ed. California: Wadsworth, 1984: 324 - 326.
- [9] 张红霞, 原杰, 柳洁, 等. 三种肾小球滤过率评估方程对早期糖尿病肾病的诊断价值比较[J]. 中华肾脏病杂志, 2011, 27(5): 341 - 345.
- [10] Grubb A, Nyman U, Bjork J. Improved estimation of glomerular filtration rate (GFR) by comparison of eGFRcystatin C and eGFRcreatinine[J]. Scand J Clin Lab Invest, 2012, 72(1): 73 - 77.
- [11] Cha R H, Lee C S, Lim Y H, et al. Clinical usefulness of serum cystatin C and the pertinent estimation of glomerular filtration rate based on cystatin C[J]. Nephrology (Carlton), 2010, 15(8): 768 - 776.

(收稿:2012-10-30;修回:2012-12-10)

(编辑 龙亮)