

基于中医证候学的糖尿病肾病患者肾小球滤过率评估方法的探索

贾冕 赵进喜 皇甫伟

【摘要】 目的 探索糖尿病肾病患者基于中医证候学评估肾小球滤过率(glomerular filtration rate, GFR)的可能性及其方法。**方法** 基于“十一五”国家科技支撑计划《中医全程干预糖尿病肾病进程综合方案研究》的 1872 例研究数据,采用散点图矩阵、安德鲁斯曲线分析、平行坐标图等计算机可视化技术,探求 GFR 与性别、年龄、身高、体重,以及气虚、血虚、阴虚、阳虚、血瘀、湿浊、痰湿的中医证候积分等 11 个因素间的关系。将病例分为 1400 例的组 1 和 472 例的组 2。基于组 1 的数据,使用两种方法进行 GFR 的估算:(1)对 11 个因素进行线性回归,并根据回归结果进行 GFR 估算。(2)以病例为单位,纳入 11 个因素,使用组 1 病例建立数据库,使用 K 最邻近结点算法(k-nearest neighbor, KNN),进行 GFR 估算。基于组 2 的数据,采用散点图、偏差分析、Bland-Altman 作图法及 ROC 曲线进行验证一致性评价。**结果** (1)GFR 与性别、年龄、身高、体重以及 7 个中医证候等因素之间存在特定联系。(2)散点图显示, KNN 法分布于 $\pm 30\%$ 范围内的点较之回归方程明显增多。两种评估方法偏差的 30%符合率均达到 50%以上。回归方程和 KNN 法偏差的 30%符合率分别达到 58.1%和 69.3%。Bland-Altman 作图显示, KNN 法估算结果的偏差分布较为均匀,数据相对集中,波动范围小于回归方程估算值。用于诊断肾功能不全时, KNN 法的 ROC 曲线下面积达到 0.847。**结论** 基于中医证候学对糖尿病肾病患者的 GFR 进行评估是可行的。KNN 法效果优于回归法,在数据量足够大时,更有利于中医证候学的研究。

【关键词】 糖尿病肾病; 肾小球滤过率; 中医证候; 评估; K 最邻近结点

【中图分类号】 R587.1 **【文献标识码】** A doi:10.3969/j.issn.1674-1749.2016.03.006

Exploration of GFR evaluation method in patients with diabetic nephropathy based on the theory of TCM syndrome JIA Mian, ZHAO Jin-xi, HUANG Fu-Wei. The First Clinical Medical College, Beijing University of Chinese Medicine, Beijing 100700, China

Corresponding author: ZHAO Jin-xi, E-mail: zhaojinximd@126.com

【Abstract】 Objective To explore the possibility of determining glomerular filtration rate (GFR) in diabetic nephropathy (DN) patients by TCM Syndrome, and the corresponding method. **Methods** Visualization technique was applied treating data from previous studies to explore the correlations between GFR and the 11 factors including gender, age, height, weight, and the TCM syndrome score for qi deficiency, blood deficiency, yin deficiency, yang deficiency, blood stasis, dampness, and phlegm dampness syndromes. The patients were divided into 2 groups with 1400 and 472 cases, respectively. 2 methods were demonstrated using group A data to evaluate GFR: (1) Linear regression GFR predictions by the aforementioned 11 factors. (2) K nearest neighbor (KNN) predictions using the group A dataset. The

基金项目: 国家“十一五”科技支撑计划(2006BAI04A03-2)

作者单位: 100700 北京中医药大学第一临床医学院(贾冕);北京中医药大学东直门医院中医内科教研室(赵进喜);北京科技大学计算机与通信工程学院(皇甫伟);北京体育大学博士流动站(贾冕)

作者简介: 贾冕(1987-),女,博士,在站博士后。研究方向:中医药防治糖尿病及其并发症。E-mail: 592857429@qq.com

通讯作者: 赵进喜(1965-),博士,主任医师,教授,博士生导师。研究方向:内分泌疾病及肾脏病的中医药防治。E-mail: zhaojinximd@126.com

prediction effectiveness was evaluated by comparing the predicted results of group B data to the real ratio, and the evaluation is facilitated by scatter diagram, deviation analysis, Bland-Altman method and ROC curve. Results (1) GFR correlated to gender, age, height, weight and 7 of the TCM syndromes. (2) The scatter diagram shows that, in the range of $\pm 30\%$, the points of KNN is significantly increased compared with the regression equation. The deviation of 30% coincidence rate of both two kinds of assessment methods have reached more than 50%. The coincidence rate of deviation of 30% of linear regression and KNN method reached 58.1% and 69.3%. Bland-Altman mapping shows that, the deviation of the estimated results of KNN method is more evenly distributed. When the data is more concentrated, the fluctuation range fell below the estimated value of the regression equation. The area of ROC curve of KNN method had reached 0.847 for the diagnosis of renal insufficiency. **Conclusion** The feasibility of predicting GFR by TCM syndromes is validated. The KNN method produces more robust results compared to linear regression. The amount of data may positively relate to a satisfactory TCM syndrome study.

【Key words】 Diabetic nephropathy; Glomerular filtration rate; TCM syndrome; Evaluation; K nearest neighbor node

糖尿病肾病(diabetic nephropathy, DN)作为一种糖尿病最主要的微血管并发症,是导致肾功能衰竭的主要原因之一。中医药在治疗 DN 方面有其独特优势,越来越多的学者着力于研究 DN 的中医诊疗方法,取得了一定进展。肾小球滤过率(glomerular filtration rate, GFR)是评价肾功能的重要指标,准确了解 GFR,对糖尿病肾病的早期诊断、早期治疗以及病情监测有着重要的作用。肾小球滤过率难以直接测量,目前多以血肌酐、胱抑素 C 等指标进行估算,也没有针对中国糖尿病肾病患者的适用的估算公式。许多研究表明,糖尿病肾病的中医证候积分与部分现代医学指标有着密切的联系。挖掘中医证候学与肾功能指标之间的联系,尝试建立一种基于中医证候学的糖尿病肾病肾功能评价方式,不仅可以更为便捷地评价肾功能,还可以与中医疗效评价体系有机结合,探求中医药治疗糖尿病肾病的原理,进而指导临床诊疗。本研究基于“十一五”国家科技支撑计划《中医全程干预糖尿病肾病进程综合方案研究》,尝试基于中医证候评分的 GFR 评估方法的探索,经验证适用性较好,报道如下。

1 对象与方法

1.1 研究对象

本研究病例来源于由北京中医药大学东直门医院牵头的国家科技部“十一五”科技支撑计划项目:《中医全程干预糖尿病肾病综合方案研究》(项目编号:2006BAI04A03-2,简称“十一五”研究),2008 年 1 月至 2009 年 12 月就诊于北京中医药大学东直门医院、中国中医科学院望京医院、天津中医

药大学第一附属医院、北京大学第一医院、浙江省杭州市红十字会医院、河北省中医院、广州中医药大学第二附属医院等 10 个参研单位的住院及门诊患者。

1.2 诊断标准

糖尿病诊断采用世界卫生组织 1999 年提出的诊断标准,糖尿病肾病依据 2010 年版《中国 2 型糖尿病防治指南》进行诊断及分期^[1]。中医证候诊断标准参考 1992 年中华中医药学会糖尿病分会的《消渴病辨证诊断参考标准》^[2]以及 1987 年中华中医药学会内科肾病专业委员会慢性肾衰中医辨证分型诊断专题讨论会通过的标准^[3]。

1.3 纳入标准

(1)明确诊断的糖尿病肾病 III 期、IV 期患者,年龄 18~70 岁,性别不限;(2)血肌酐在 3 个月内稳定在 $265\ \mu\text{mol/L}$ 以下;(3)血糖平稳,糖化血红蛋白 7.5% 以下,舒张压 90 mmHg 以下;(4)中医辨证具备本虚(气虚、阴虚、阳虚、血虚)表现,或兼血瘀、湿浊、痰湿者。

1.4 排除标准

(1)难治性水肿;(2)尿蛋白大于 10 g/24h,血清白蛋白低于 $25\ \text{g/L}$;(3)肾性高血压,合并充血性心衰 I-IV 级,癌症、妊娠及对治疗方案药物过敏者;(4)六个月内患有糖尿病酮症酸中毒、心肌梗塞、恶性高血压、脑血管意外等危急重症病史者。

1.5 本研究病例选取方法

“十一五”研究纳入病例 320 例,以 3 个月为周期对患者进行了平均 19 个月的访视,监测中医证候积分及相关指标。本研究以诊次为单位,选择中医证候评分及实验室检验资料完整的病例资料,共计

9 次访视,1872 例数据资料,其中男性 1162 人,女性 710 人,平均年龄 (59.59 ± 8.623) 岁,平均身高 (165.66 ± 7.65) cm,平均体重 (70.87 ± 11.51) kg,平均 BMI 值 (25.79 ± 3.68) kg/m²。中医证候总积分平均 21.83 分,其中气虚 4.9 分,阴虚 3.63 分,阳虚 5.66 分,痰湿 3.05 分,血瘀 1.54 分,血虚 1.32 分,湿浊 1.74 分。GFR 平均 79.21 mL/(min · 1.73 m²)。

1.6 病例分组资料

将纳入病例按照 3:1 的比例分为两组:将原始随机编号的顺序,按照“1、1、1、2”的顺序依次重新编号,新编号“1”和“2”分别为组 1 和组 2。经计算,两组当为 1404 例和 468 例。为方便计算,取组 1 为 1400 例,组 2 为 472 例。

1.7 研究方法

(1) 评估方法的建立:使用中国改良简化 MDRD 公式估算肾小球滤过率^[4]。按分组选择病例。基于组 1 的数据,使用多元线性回归建立基于中医证候积分的估算公式;使用 Scipy 科学计算库的 spatial. kdtree 模块,将病例数据中的肾小球滤过率、性别、年龄、身高、体重和气虚、血虚、阴虚、阳虚、痰湿、湿浊、血瘀等 7 个中医证候等信息,以每个病例为单位,进行数据库的建立。(2) 新建方法的验证:基于组 2 的病例,利用新建数据库,采用 K 最近邻结算法(k-nearest neighbor algorithm,KNN,简称“最近邻”),对肾小球滤过率进行评估,并使用新建立的公式对肾小球滤过率进行估算。采用散点图、偏差分析、Bland-Altman 作图法及 ROC 曲线等方法,对两种评估方法进行适用性评价。

1.8 统计方法

使用 Python 2.7.9 数据的可视化辅助处理,散点矩阵图结合应用 pandas 0.15.2,平行坐标图结合应用 pycairo 1.10.0,安德鲁斯曲线结合应用 matplotlib 1.4.3 rc1。使用 medcalc 12.3.0.0 软件绘制 Bland-Altman 图,应用 SPSS 19.0 统计软件进行数据的统计分析。计量指标以均数 ± 标准差($\bar{x} \pm s$)描述。根据正态检验,相关分析采用 Pearson 相关或 Spearman 相关,基线比较采用独立样本 *t* 检验或两独立样本检验。所有假设采用双侧检验, $P < 0.05$ 认为有统计学差异。

2 结果

2.1 病例分组基线比较

对两组病例的数据进行正态检验,两组 eGFR

均服从正态分布, $F = 1.934$,具有方差齐性,组间比较采用独立样本 *t* 检验。其余各项指标均不服从正态分布,采用非参数检验进行两组间比较。两组 eGFR_{中国}、性别、年龄、肌酐、尿素氮及中医证候积分的组间比较显示,两组基线水平相似,不具有统计学差异($P > 0.05$)。

2.2 各因素与肾小球滤过率之间的关系

2.2.1 各因素的散点图矩阵分析 绘制各因素的散点图矩阵,见图 1。由图可以看出:(1)肾小球滤过率与肌酐关系密切;(2)肾小球滤过率与年龄呈负相关,即随年龄增长,肾小球滤过率下降;(3)肾小球滤过率与身高、体重和血虚、阴虚关系不明显;(4)中医证候湿浊、痰湿、血瘀、阳虚和气虚在最右一列图上呈现上尖下宽的“三角形”图形。

2.2.2 平行坐标图 绘制平行坐标系下的调和曲线图,见图 2。图中每条曲线代表 1 个病例的数据,依肾小球滤过率的值赋予不同颜色,按其值从小到大,颜色由红到绿的规律演变。由图可见:(1)肾小球滤过率与肌酐呈现负相关;(2)GFR 与年龄有负相关倾向;(3)中医证候积分与肾小球滤过率的曲线分布关系不显著;(4)趋近于绿色的曲线向中医证候积分的上半部分集中,即 GFR 值较大的病例,中医证候积分分值较小。

2.2.3 安德鲁斯曲线分析 对肾小球滤过率、性别、年龄、身高、体重及气虚、血虚、阴虚、阳虚、痰湿、湿浊、血瘀等 7 个中医证素,共计 12 个因素的相互关系进行降低维度的可视化处理,绘制安德鲁斯曲线(Andrews curve),见图 3。图中每一条曲线代表 1 个病例,坐标数值没有实际意义。由图可以看出,曲线并非杂乱无章,大部分曲线构成了 3 峰 2 谷的类“W”型曲线。

2.3 线性回归方程

基于组 1 数据,将性别赋值“男 1 女 2”,使用性别、年龄、身高、体重、气虚、血虚、阴虚、阳虚、痰湿、湿浊等 11 个因素进行线性回归。回归模型: $eGFR = a_1 \times p_1 + a_2 \times p_2 + \dots + a_n \times p_n$ 。其中, a 为系数, p 为如上 12 个参数, n 取值 1 ~ 12。

回归结果:模型拟合的可决系数调整 $R^2 = 0.155$,方差分析显示, $P < 0.01$,模型是显著的,但拟合优度欠佳。回归方程 1: $eGFR_1 = 304.192 - 3.568 \times \text{性别} - 1.049 \times \text{年龄} - 0.980 \times \text{身高} + 0.214 \times \text{体重} - 0.265 \times \text{气虚} - 0.313 \times \text{血瘀} - 0.152 \times \text{阴虚} - 1.064 \times \text{阳虚} - 0.934 \times \text{血虚} + 0.360 \times \text{痰湿} - 0.677 \times \text{湿浊}$ 。

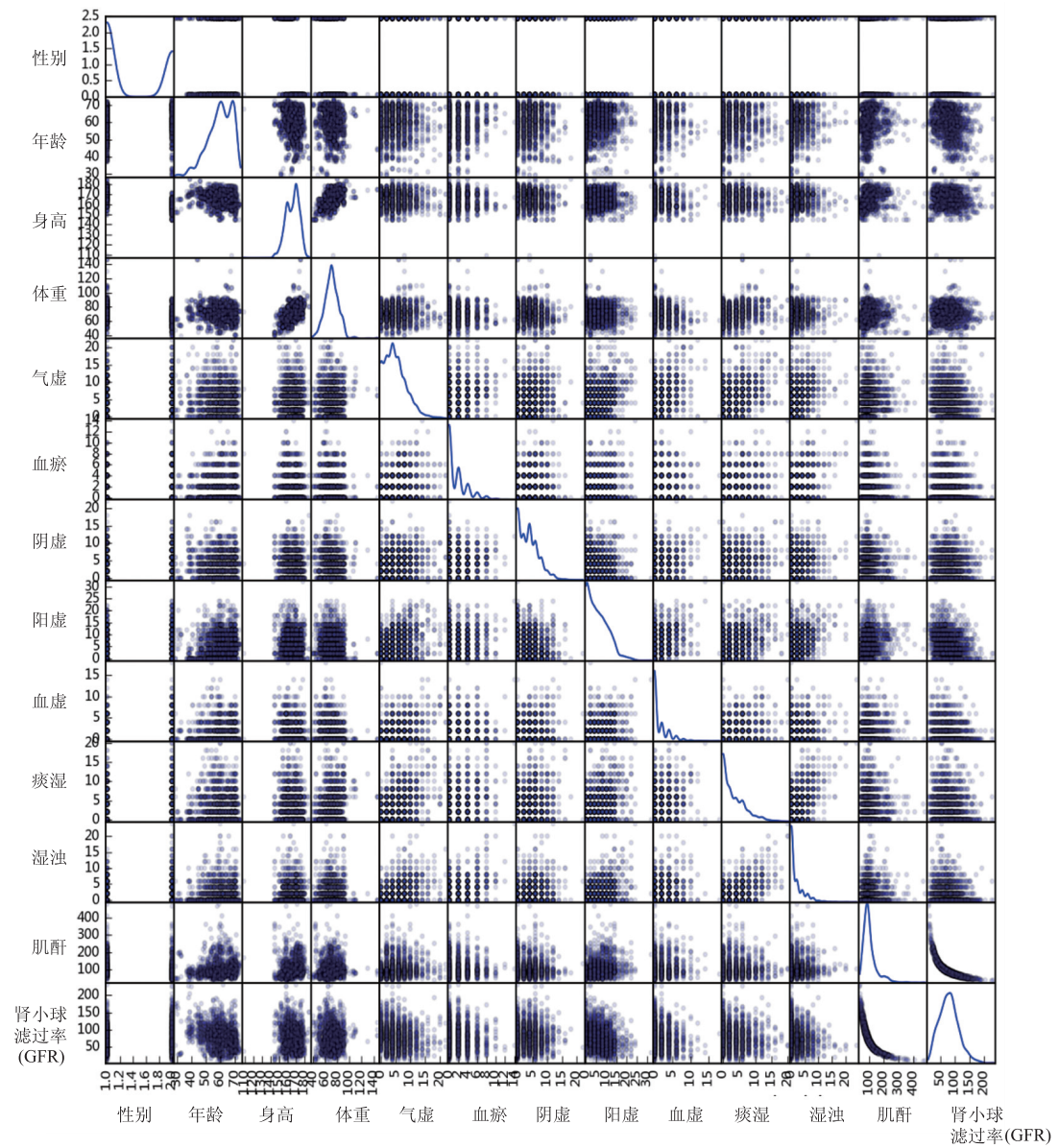


图 1 与肾小球滤过率相关的各因素散点图矩阵

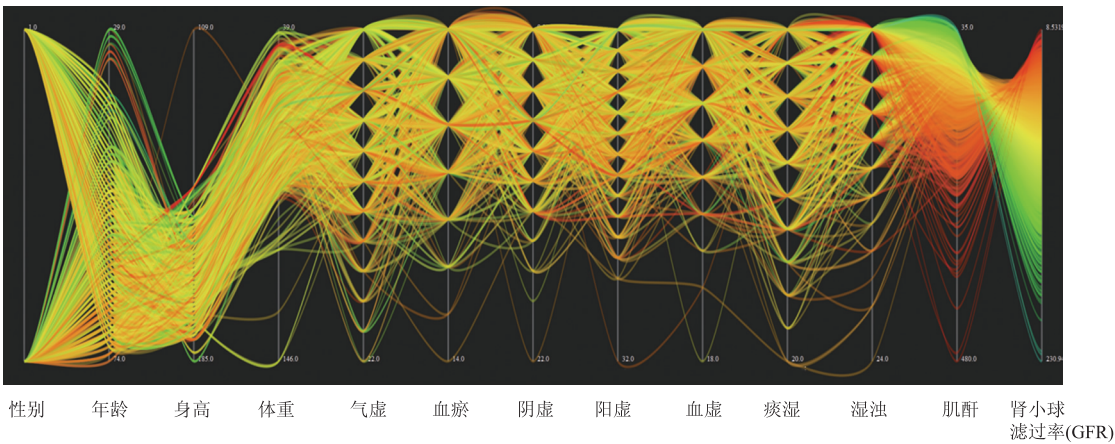


图 2 与肾小球滤过率相关的各因素平行坐标图

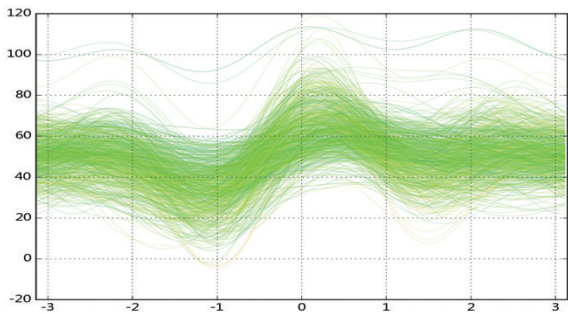


图3 与肾小球滤过率相关的各因素安德鲁斯曲线图

由回归方程可见,性别、年龄、身高、阳虚、血虚、湿浊的系数相对较大,对 GFR 值影响相对较大。回归结果见表 1。

表 1 回归系数及检验结果

模型	非标准化系数		标准系数		t	Sig.
	B	标准误差	试用版			
常量	304.192	32.362		9.400	0.000	
性别	-3.568	2.575	-0.053	-1.385	0.166	
年龄	-1.049	0.096	-0.281	-10.936	0.000	
身高	-0.980	0.183	-0.227	-5.358	0.000	
体重	0.214	0.092	0.074	2.334	0.020	
气虚积分	-0.265	0.247	-0.032	-1.072	0.284	
血瘀积分	-0.313	0.427	-0.021	-0.732	0.464	
阴虚积分	-0.152	0.268	-0.015	-0.568	0.570	
阳虚积分	-1.064	0.199	-0.165	-5.333	0.000	
血虚积分	-0.934	0.413	-0.065	-2.258	0.024	
痰湿积分	0.360	0.315	0.040	1.145	0.252	
湿浊积分	-0.677	0.383	-0.060	-1.768	0.077	

2.4 建立“K 最邻近结点算法”样本库(KNN 样本库)

(1)基于组 1 的数据,纳入性别、年龄、身高、体重、气虚、血瘀、阴虚、阳虚、血虚、痰湿、湿浊共 11 个因素,作为训练数据进行存储。

(2)设定 k 值,本研究取 $k=5$ 计算。

(3)距离计算法

在 11 个因素构建的 11 维样本空间中,任取 1 点,该点与周围 n 个点的直线距离分别为 d_1, d_2, \dots, d_n 。则计该点与周围 n 个点的距离为

$$d = \sqrt{d_1^2 + d_2^2 + \dots + d_n^2}.$$

(4)估算方法

在样本空间中,分别计算待测样本与样本库中 11 个维度的原始样本之间的距离,得到距离 d_1, d_2, \dots, d_n ,比较 $d_1 \sim d_n$,取最小的 k 个样本(本研究 $k=5$),其对应的样本库中的原始样本,即为选定的样本。选定样本的肾小球滤过率的算术平均值,即

为待测样本的估算值。

2.5 新建评估方法的适用性评价

基于组 2 的 472 例数据,分别运用回归方程和 KNN 样本库估算肾小球滤过率,与原值进行对比评价。

2.5.1 散点图分析 分别以原肾小球滤过率为 x 轴,新建方法计算的值为 y 轴,绘制散点图(图 4 和图 5)。由图可见,回归方程的散点分布略呈“水平分布”的状态。KNN 法的估算值呈现 45° 倾斜的趋势,且分布于 $\pm 30\%$ 范围内的点较之回归方程明显增多。

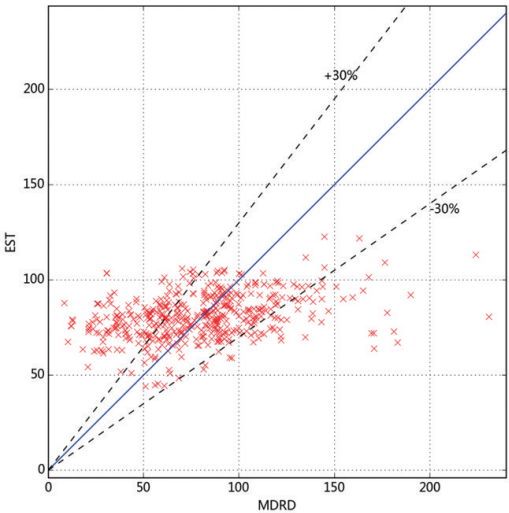


图 4 回归方程的散点图

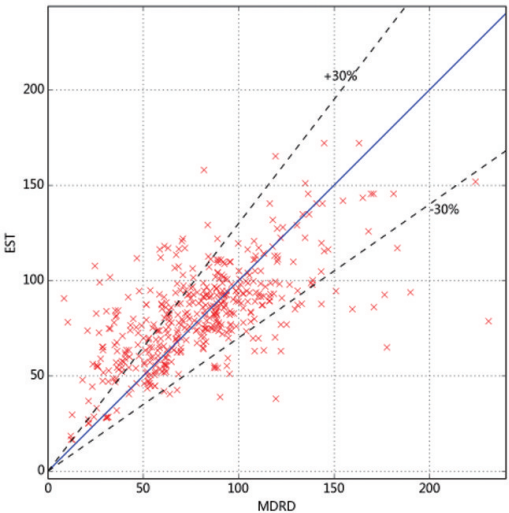


图 5 “最近邻”法的散点图

2.5.2 偏差比较 分别计算回归方程和 KNN 法估算结果的偏差,偏差 = 估算值 - 金标准,偏离百分数 = $[(\text{估算值} - \text{金标准}) \text{的绝对值} / \text{金标准}] \times 100\%$ 。15%、30%、50% 符合率 = 估算值落在金标准 $\pm 15\%$ 、 $\pm 30\%$ 、 $\pm 50\%$ 的病例百分数^[5],详见表 2。由表可

知,两种评估方法 30% 符合率均达到 50% 以上,评估效果较好。KNN 法优于回归方程,其 30% 符合率接近 70%。

表 2 回归方程和 KNN 的偏差对比

公式	偏差	15% 符合率 (%)	30% 符合率 (%)	50% 符合率 (%)
回归方程	1.22±32.6	31.4%	58.1%	77.8%
KNN	2.14±26.55	45.1%	69.3%	84.5%

2.5.3 Bland-Altman 分析 绘制 Bland-Altman 图,添加均值、均值±1.96 标准差的参考线,见图 6 和图 7。由图可知,KNN 法估算结果的偏差分布较为均匀,数据相对集中。回归方程的偏差有较为明显的线性呈现。KNN 法偏差的波动范围小于回归方程估算值。

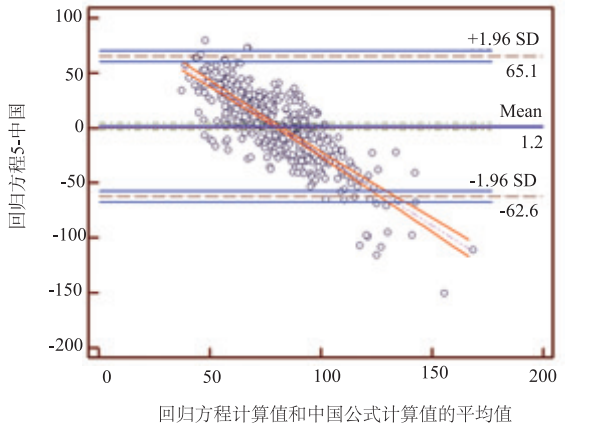


图 6 以中国公式为金标准的回归方程的 Bland-Altman 图

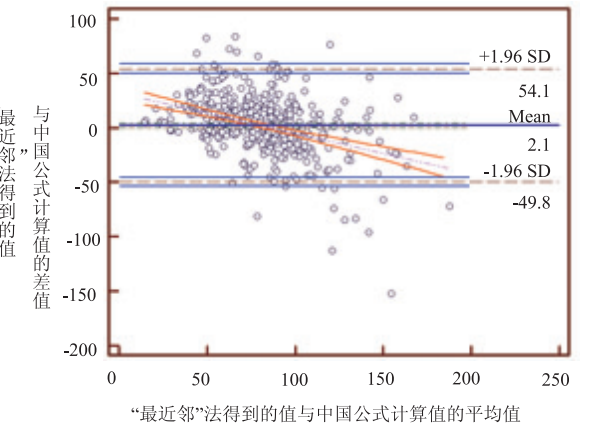


图 7 以中国公式为金标准的“最近邻”法的 Bland-Altman 图

2.5.4 诊断效能 以 $GFR < 60 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$ 作为节点,以原肾小球滤过率的值为金标准,使用 ROC 曲线分析进行诊断效能的评价。结果见表 3、图 8 和图 9。

表 3 验证公式的 ROC 曲线分析

公式	灵敏度	特异度	约登指数	ROC 曲线下面积
回归公式	78.4	50.8	0.2917	0.680
KNN	76.3	80.5	0.5674	0.847

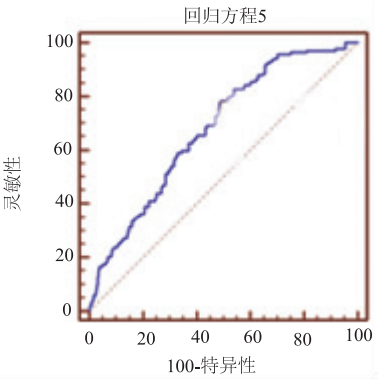


图 8 回归方程的 ROC 曲线分析

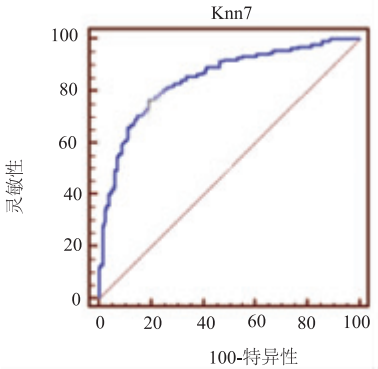


图 9 “最近邻”法的 ROC 曲线分析

3 讨论

本研究为基于中医证候学的糖尿病肾病肾小球滤过率评估方法的建立和验证,其关键点在于探求肾小球滤过率与中医证候之间的关系,并用恰当的方法予以表示。既往公式的建立多以统计学为基础,通过分析其间的相关性。随着科技的发展,计算机技术在众多领域被广泛应用,基于计算机技术的一些新方法也可以在医学领域施展本领,或许将成为未来的发展方向。

3.1 多维空间的可视化处理技术

现在可以理解并直观感受到的仅限于三维以下的空间,当自变量增加到 4 个甚至更多时,很难想象多条相互垂直的线决定的多维空间的具体形象。随着计算机技术的发展,更多的新方法可以用来寻找变量之间的关系。科学计算可视化技术,可以将各种数据进行转化,成为直观的,能够以图形呈现

的信息。常见的有散点矩阵技术、平行坐标可视化技术、安德鲁斯曲线技术、格架图、测量图、放射性可视化技术等。将这些方法运用于医学领域,探索一些重要但难以测量的指标与其他易测量指标之间的关系,利于探索人体机能的奥秘。本研究中利用前三种技术对数据进行处理,可以较为直观地看到数据间的联系。

散点矩阵通过绘制 n 个因素的 n^2 个散点图,展现每两个因素之间的关系。本研究中的散点图矩阵显示出肌酐与肾小球滤过率呈非直线的线性关系;年龄与 GFR 有斜率为负值的线性趋势。中医证候积分与肾小球滤过率之间虽并没有明显的线性关系,但其“三角形”的图形特性提示其对肾小球滤过率的取值存在限制作用:当单个证候积分取值很小时(位于“三角形”尖端),肾小球滤过率取值不会大,即有一定的“最大界限”。因此,当使用多个证候综合评价时,可以较为准确地限制肾小球滤过率的取值范围。平行坐标法相较于安德鲁斯曲线的空间旋转和运算更为便捷。它采用一系列平行的坐标,每个坐标有其独立的量值标准,将高维度空间的“扭曲的线”(或称点集),分别在平行的各个坐标上予以标注,并连接成线。这样,就将高维空间的“扭曲的线”以平面上的折线的形式展现了出来。直接绘制的平行坐标图是由直线构成的,因其拐点可能重合,不易观察,可将其调整为曲线联结,并根据值的大小予以不同色彩的设定,有利于直观地发现因素之间的联系。安德鲁斯曲线技术可以简单理解为,将存在于多维空间的“扭曲的线”进行扭转,寻找某个角度,将其投射在仅有 x 轴和 y 轴的平面上。本研究中的安德鲁斯曲线,即是肾小球滤过率与 11 个自变量间相互关系的曲线的展现,每个数据绘制一条线,1872 个数据共同表现出三峰两谷的类“W”形平滑的曲线,提示肾小球滤过率与 11 个自变量间的确存在一定的内在联系,当数据样本足够充足,能够更为准确地确定相互间的关联,使基于“大数据”的处理技术,从中医证候学角度评估 GFR 成为可能。

3.2 K 最邻近结点算法

KNN 法属于大数据应用的一种。与前几种图像法不同,这种方法不能直观看到数据之间的关系,但可以得到估算的数值,其基本要求是建立足够大量的样本库。当需要估测某个因变量时,通过该指标的自变量的信息,在样本库中筛选 k 个最近

距离的样本,对样本库中这 k 个样本对应的已有因变量的值进行平均,即为估测值。这种算法类似于将相关因素画圆圈起来,再找圆心的过程。该方法的优点在于,所有估测出来的因变量的值均由原始样本库的值平均得到,因此,其范围不会波动过大,即偏差值过大。这就避免了所有基于统计学方法得到的计算公式的弊端:利用一定范围数据建立的公式,在超过数据范围后,往往偏差巨大。比如本研究中 1 例肌酐为 6.2 umol/L 的病例,通过中国公式计算得到 $\text{eGFR} > 800 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$,与实际完全不符,因此予以剔除。同样,这些偏离较远的点在公式推导过程中,往往起到了不小的作用,使线的走向发生明显偏移。KNN 法可以有效避免这种情况的发生,因其选择的是“最小距离”。

KNN 法虽有诸多便利,但其关键因素—— k 值的确定却不十分容易,不同的 k 值结果不同。当样本库足够大时, k 值越大,准确性越高,但运算量呈指数倍增加。确定 k 值一般需要进行测试,由小到大取不同 k 值进行部分数据的运算,将其与标准进行对比,选择与标准最接近的 k 值。本研究中,因样本库存在 11 个维度的信息,仅有 1400 个数据,因此仅进行了少量运算,暂定结果较好的 $k = 5$ 作为标准,以尝试这种方法的可行性。研究结果表明,KNN 法的估算明显优于回归方程的计算,这证明了 KNN 法在中医药领域应用的可能性。

3.3 新建评估方法的评价

回归方程的散点图呈现“水平分布”,结合 Bland-Altman 图分析,其原因在于估算结果在肾小球滤过率较低时高估了真实值,而在肾小球滤过率较高时低估了真实值,随着肾小球滤过率值的降低或升高,偏差逐渐增大。偏差分析显示,其 30% 符合度接近 60%,相较于一般估算公式效果尚佳。以肾小球滤过率 $< 60 \text{ mL}/(\text{min} \cdot 1.73 \text{ m}^2)$ 作为节点诊断肾功能不全时,回归方程诊断的灵敏性一般,特异度较低,ROC 曲线下仅有 0.680,诊断效力不高。KNN 法的散点图呈 45° 倾斜趋势提示估算值与“金标准”符合度高,分布与 $\pm 30\%$ 范围内的点较之回归曲线明显增多,Bland-Altman 图也显示其偏差波动范围相对较小,且数据更为集中,其偏差的 30% 接近 70%,说明该法用于估算肾小球滤过率效果良好。同时,用于诊断肾功能不全时,灵敏度和特异度均 > 0.75 ,ROC 曲线下面积达到 0.847,提示 KNN 法对肾功能不全具有较高的诊断效能。综合评价

两种基于中医证候积分的肾小球滤过率评估方法,二者的估测值与中国改良简化 MDRD 公式计算的肾小球滤过率的值一致性较好,用于评估肾小球滤过率值都是可行的。比较而言,KNN 法在准确性及诊断效能上,均明显优于回归方程。

4 小结

本研究以中国改良简化 MDRD 公式估算的肾小球滤过率值为“金标准”,基于组 1 数据,使用回归分析和 KNN 两种方法进行中医证候学积分与肾小球滤过率之间关联的探索,并在组 2 中进行了验证。从以上数据中,可以得到一些结论:(1)基于中医证候学的肾小球滤过率评估方法的建立是切实可行的。(2)当数据量足够大时,KNN 法进行估算效果明显优于回归方法。这提示,随着“大数据”时代的到来,中医或是医学界采纳更为广泛的科技手段,积累并存储、运用大量数据,或许将成为未来医学发展的方向。(3)从本研究结果来看,既然中医的 7 个证候与肾小球滤过率有密切的吻合度,也就证明了中医证候与各实验室指标之间可能存在着多种联系,蕴藏了大量信息。探索中医证候学与疾病主要疗效指标之间蕴藏的信息,为从中医角度

评价病情,或是选择适宜的治则治法提供了重要的参考依据。

因条件限制,本研究仅能使用现有的中国改良简化 MDRD 公式的估算值作为“金标准”进行探索。如能够将以普遍认可的肾动态显像等方法测得的肾小球滤过率值作为金标准,并将相应的病例信息一同纳入数据库,在汇集更多数据的基础上,采用 KNN 法进行肾小球滤过率的评估,或许将优于各种基于统计学的估算公式。

参 考 文 献

- [1] 中国 2 型糖尿病防治指南(2010 年版)[J]. 中国糖尿病杂志, 2012, (1):81-117.
- [2] 吕仁和,张洁荣,高彦彬. 消渴病(糖尿病)中医分型辨证与疗效评定标准[J]. 中国医药学报,1993,8(3):54-56.
- [3] 刘宏伟. 慢性肾衰中医辨证分型和疗效评定标准-全国中医肾衰研讨会讨论通过[J]. 中医药信息,1991,8(2):27-28.
- [4] 全国 eGFR 课题协作组. MDRD 方程在我国慢性肾脏病患者中的改良和评估[J]. 中华肾脏病杂志,2006,22(10):589-595.
- [5] 史浩,陈楠,张文,等. 简化 MDRD 公式预测慢性肾病患者肾小球滤过率的应用评价及校正[J]. 中国实用内科杂志,2006,36(9):665-669.

(收稿日期:2015-03-29)

(本文编辑:禹佳)