

· 论著 ·

血尿酸与慢性代谢性疾病的连续型和离散型贝叶斯网络效果比较*

山西医科大学卫生统计教研室(030001)

崔宇 宋伟梅 任浩 王旭春 乔宇超 赵执扬 任家辉 仇丽霞[△]

【摘要】目的 建立血尿酸与相关代谢性指标的连续型和离散型贝叶斯网络模型,探寻血尿酸的影响因素,并比较两种网络结果的特点和优劣。**方法** 利用山西省 2015 年慢性病监测的血尿酸及其代谢性疾病的特征指标共 4646 例,采用(improved partial-correlation-based, IPCB)算法建立血尿酸的连续型贝叶斯网络,同时将上述指标离散化,采用(max-min hill-climbing, MMHC)建立高尿酸的离散型贝叶斯网络。**结果** 离散型贝叶斯网络发现 14 条边,其中甘油三酯和舒张压异常与高尿酸直接关系,导致高尿酸的发生;年龄为间接因素;而连续型贝叶斯网络共包含 24 条有向边,年龄、TG、LDL、HDL、SP、DP 与尿酸水平直接相关,随着年龄、TG、LDL 的增大和 HDL 的降低,尿酸水平升高,而尿酸水平升高又导致 SP、DP 升高;TC 与尿酸间接相关。**结论** 两种网络模型适应的资料类型不同,但连续型贝叶斯网络发现的直接相关因素更多,整体解释度更好。

【关键词】 连续型贝叶斯网络 IPCB 算法 血尿酸 代谢性指标 相关因素

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.02.001

Comparison of Continuous and Discrete Bayesian Network Models for Uric Acid and Related Metabolic Indicators

Cui Yu, Song Weimei, Ren Hao, et al (Department of Health Statistics, Shanxi Medical University(030001), Taiyuan)

【Abstract】Objective The continuous and discrete Bayesian network models of serum uric acid and related metabolic indexes were established to explore the influencing factors of serum uric acid and compare the characteristics and advantages of the results of the two networks. **Methods** A total of 4646 patients with serum uric acid and metabolic diseases from chronic disease monitoring in Shanxi Province in 2015 were selected. IPCB algorithm was used to establish a continuous Bayesian network of serum uric acid. Meanwhile, the above indicators were discretized, and MMHC was used to establish a discrete Bayesian network of high uric acid. **Results** The discrete Bayesian network found 14 edges, in which triglyceride and diastolic blood pressure abnormalities were directly related to the occurrence of high uric acid, leading to the occurrence of high uric acid. Age was an indirect factor; Age, TG, LDL, HDL, SP and DP are directly related to uric acid level. With the increase of age, TG and LDL and the decrease of HDL, uric acid level increases, while the increase of uric acid level leads to the increase of SP and DP. TC is indirectly related to uric acid. **Conclusion** The two network models adapt to different data types, but the continuous Bayesian network has more direct correlation factors and better overall explanatory degree.

【Key words】 Continuous Bayesian network; IPCB algorithm; Blood uric acid; Metabolic indicators; Relevant factor

高尿酸血症(hyperuricemia, HUA)是人体氨基酸、核酸分解及嘌呤代谢异常导致尿酸生成增加或排泄减少,从而造成尿酸在体内滞留的一种代谢性疾病^[1]。随着经济的发展和人们饮食生活习惯的改变,我国居民高尿酸血症患病率呈逐年升高趋势,并呈年轻化发展^[2]。研究表明心血管疾病、高血压、高血脂、糖尿病等慢性病与高尿酸血症密切相关^[3],从医学生物效应的角度来看,这些疾病/现象与因素、或因素与因素间可能存在复杂的网络结构关系,这种关系可能是一种整体联动效应,主效应和交互效应都涵盖在网络中,一个可控环节的改变,将会导致整体效应的变化。前期课题组研究的离散型贝叶斯网络(discrete Bayesian networks, DBNs)通过构建有向无环图反映各变量间的潜在的关系,利用条件概率分布表反映关联强度,直

观地描述疾病/现象与因素间复杂的网络风险机制^[4],可以给出与疾病/现象直接相关和间接相关的因素,通过网络很好地反映了疾病与因素之间的整体联动效应,而且对统计假设没有严格要求。血尿酸是一个连续性变量,采用该网络时需要根据医学参考值范围对其离散化,而离散化可能导致信息的丢失^[5]。当研究者希望重点了解血尿酸水平与其他慢性代谢性疾病的整体联动效应,进一步描述与血尿酸水平直接相关和间接相关的因素时,以结构方程模型为数学基础的连续型贝叶斯网络(continuous Bayesian networks, CBNs)是比较理想的一个方法。因此,本文采用改进的偏相关算法(improved partial-correlation-based, IPCB)构建血尿酸水平及相关因素的连续型贝叶斯网络,并根据血尿酸的医学参考值范围,将其分为正常和升高两类,采用最大最小爬山算法(max-min hill-climbing, MMHC)构建离散型贝叶斯网络,探讨两

* 基金项目:国家自然科学基金资助项目(81973155)

[△]通信作者:仇丽霞, E-mail: qlx_1126@163.com

种模型研究血尿酸与其他慢性代谢性疾病关系的合理性,为贝叶斯网络用于因素间关联性研究的模型选择提供思路。

资料与方法

1.资料来源

(1)数据收集

研究样本来自山西省 2015 年慢性病监测数据,共 4846 名研究对象,其中 4646 例数据完整,其中 10 个代谢性疾病的特征指标,分别为:尿酸(uric acid, UA)、年龄、总胆固醇(total cholesterol, TC)、甘油三酯(triglyceride, TG)、低密度脂蛋白胆固醇(low density lipoprotein, LDL)、高密度脂蛋白胆固醇(high density lipoprotein, HDL)、收缩压(systolic pressure, SP)、舒张压(diastolic pressure, DP)、空腹血糖(fasting blood glucose, FBG)、糖化血红蛋白(hemoglobin A1c, HbA1c),均为非高斯分布的连续性变量。

(2)数据离散化

为了满足离散型贝叶斯网络的要求,对 10 个连续性代谢指标进行离散化处理。离散化分组界值采用国际上对高尿酸血症、高胆固醇血症、高低密度脂蛋白血症、高甘油三酯血症、高血压、糖尿病的诊断或判断标准^[6-8],年龄以联合国世界卫生组织对年龄段划分的标准为界值,44 岁及以下为青年人,45 岁至 59 岁为中年人,60 岁及以上为老年人。变量赋值见表 1。

表 1 变量赋值

因素	赋值
年龄(岁)	<45=1; 45~59=2; 60~=3
SP(mmHg)	<140=1, ≥140=2
DP(mmHg)	<90=1, ≥90=2
TC(mmol/L)	<6.2=1, ≥6.2=2
TG(mmol/L)	<2.3=1, ≥2.3=2
LDL(mmol/L)	<4.1=1, ≥4.1=2
HDL(mmol/L)	<1.0=1, ≥1.0=2
UA(μmol/L)	男:≤420=1, >420=2 女:≤360=1, >360=2
FBG(mmol/L)	<7.0=1, ≥7.0=2
HbA1c(%)	<6.3=1, ≥6.3=2

2.研究方法

(1)离散型贝叶斯网络

贝叶斯网络由 Pearl Judea 在 1987 年首次提出,之后逐渐发展至今^[9]。当给定一个随机分类变量集 $X = \{X_1, \dots, X_n\}$,这些变量的联合概率 $P(X_1, \dots, X_n)$ 就可以表示成一个贝叶斯网络 $B = (G, P)$ 的形式。其中 G 是一个有向无环图,由节点和有向边构成,也被称为贝叶斯网络的基本骨架。网络图中的节点是研究的变量,通常由研究者确定,而有向边则反映任意两个变量

之间的概率依赖关系^[10]。对于图中一条从 X_i 指向 X_j 的有向边, X_i 被称为父节点, X_j 被称为子节点,后者发生的概率受前者状态的影响,并且,当父节点 X_i 的值被确定时,子节点 X_j 与其他同父子节点条件独立。 P 代表的是这些变量的条件概率分布表(conditional probability distributions table, CPT),条件概率是一个局部概率,每个子节点的条件概率只与它的父节点有关^[11]。贝叶斯网络用一个联合概率分布表示了变量之间的概率依赖关系,既体现了条件概率的局部性,又体现了概率在变量之间的可传播性,见图 1。

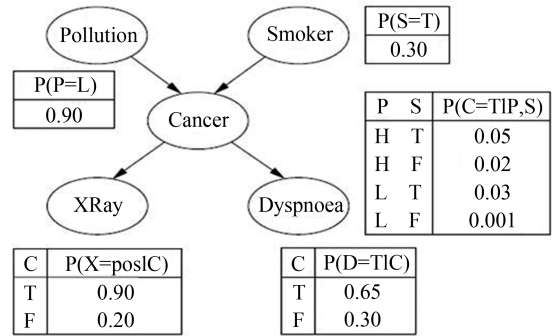


图 1 贝叶斯网络结构示例

离散型贝叶斯网络的构建包括结构学习和参数学习^[12],本研究采用数据驱动的建模方式构建离散型贝叶斯网络,首先由研究者确定网络节点,再采用混合算法 MMHC 进行结构学习^[13],该法首先使用启发式搜索算法 MMPC(max-min parents and children)确定每个变量的候选父子节点集,构建出网络的基础框架。然后以 BDeu 评分为准则通过评分搜索的方法在约束空间执行贪婪搜索,从只有节点存在的无边图开始执行边的增加、删除和转向等操作,最终找到评分最大的网络模型,确定边的方向。MMHC 算法是一种结合了条件独立性测试和评分搜索两种不同理念的混合型结构学习算法,课题组前期已证实其性能优于常用的禁忌搜索算法^[14],第二阶段的搜索被限制在第一阶段 MMPC 算法发现的潜在无向边中,大大降低搜索空间的复杂度。

参数学习是在结构学习获得网络结构的基础上,进一步估计网络中的参数^[15],本研究采用极大似然估计法。

(2)连续贝叶斯网络

当给定一个随机变量集 $X = \{X_1, \dots, X_n\}$ 为连续性定量变量时,传统的结构学习算法如 MMHC 算法等,通常会进行离散化处理,不可避免地造成了信息丢失,影响结果的准确性。由于本研究中尿酸等代谢指标不符合高斯分布的连续性定量变量,因此选择适合于非高斯分布数据的 IPCB 混合算法进行结构学习^[16],建立连续性贝叶斯网络。该算法是传统偏

相关算法的改进,并在判断偏相关时加入了假设检验的原理。IPCB 混合算法由约束阶段和搜索阶段组成,整体更加合理。

约束阶段的输入值为给定数据和预设的阈值参数,该阈值参数就是假设检验中的检验水准。IPCB 算法基于偏相关系数进行条件独立性测试,获得每个节点的候选邻居节点集(possible neighbors, PN)及候选邻居矩阵(possible neighbor matrix, PNM),并将其作为输出值。首先将 PN 初始化为空集,即 PNM 全部为 0。若 $PNM(i, j) = 1$, 则将 x_i 放入 x_j 的候选邻居集;若 $PNM(i, j) = 0$, 则 x_i 不能放入 x_j 的候选邻居集。根据此原则获得每个变量的候选邻居节点集和候选邻居矩阵,建立贝叶斯网络的基本骨架。

IPCB 算法的条件独立性测试基于偏相关系数进行,相对于既往算法的优势是 x_i 和 x_j 的偏相关系数 ρ_{ij} 要经过 t 检验判定是否有统计学意义。

$$t = \frac{\hat{\rho}_{ij} - \rho_{ij}}{SE(\hat{\rho}_{ij})}$$

式中 $\hat{\rho}_{ij}$ 是样本信息提供的偏相关系数估计值, ρ_{ij} 是对总体特征的假定值,本研究假定为 0,即变量间不存在相关性, SE 是 $\hat{\rho}_{ij}$ 的标准误。对于任意变量 x_i 和 x_j ,在给定 Z 时计算其偏相关系数并进行假设检验,若 P 值小于检验水准,则 x_i 和 x_j 的相关性有统计学意义,它们在贝叶斯网络中互为邻居节点,反之便不能纳入候选邻居节点集;直至所有邻居节点和 PNM 被找到,约束算法结束。

搜索阶段采用基于 MDL(G) 评分的爬山法确定节点之间的边及边的方向,构建最终的网络结构。MDL(G) 评分的计算如下:

$$MDL(G) = \sum_{j=1}^n \{NLL[X_j, Pa(X_j), \hat{\Theta}^{mle}_j]\} + \frac{|\hat{\Theta}^{mle}_j|}{2} \log m$$

式中 $\sum_{j=1}^n \{NLL[X_j, Pa(X_j), \hat{\Theta}^{mle}_j]\}$ 表示模型拟合优度,其中, $Pa(X_j)$ 是节点 X_j 的父节点, $NLL[X_j, Pa(X_j), \hat{\Theta}^{mle}_j]$ 是节点及其父节点的 MDL (minimum description length scoring) 评分,也称家族评分或局部评分; $\frac{|\hat{\Theta}^{mle}_j|}{2} \log m$ 表示网络复杂度, m 是样本例数, n 是变量个数; MDL(G) 是网络的总评分,反映网络的整体效应。

搜索阶段的输入包括给定数据集、约束算法获得的 PN 和 PNM 和家族评分最大次数。搜索从空网开始,当家族评分次数小于最大评分次数时,在候选邻居节点间随机对有向边做增加、减少、转向处理,并在此过程中测试 MDL(G) 评分的动态变化,最后得到一个具有最优评分的网络结构。

(3) 统计分析软件

离散型贝叶斯网络使用 R 4.0.5 软件中的贝叶斯网络学习程序包进行结构学习,用 Netica 进行参数学习;连续贝叶斯网络的结构学习用 Matlab R2016a 运行 IPCB 算法加以实现,并用 SPSS 22.0 做变量间的偏相关分析。

结 果

1. 高尿酸影响因素的离散型贝叶斯网络

本研究为现况调查,测量的代谢相关指标均按照医学参考值范围划分为正常、异常两类,描述为某代谢相关指标是否异常,而不定义受试者患有某病;年龄以联合国世界卫生组织年龄段划分标准划分为三个类别,变量赋值见表 1。采用 MMHC 混合算法,条件独立性检验水准设置为 0.1,得到的离散型贝叶斯网络共包含 10 个节点、14 条有向边(见图 2),主要反映了高尿酸与其他指标正常与否的关系。

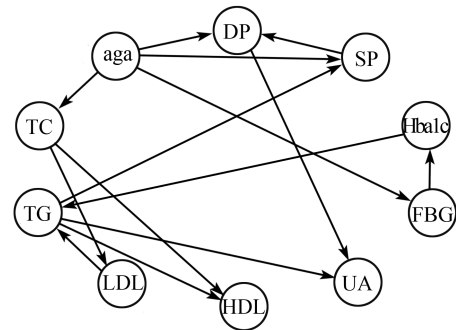


图 2 离散型贝叶斯网络结构

离散型贝叶斯网络提示:与高尿酸直接相关的因素仅有甘油三酯(TG)和舒张压(DP)。根据离散型贝叶斯网络可以给出的尿酸(子节点)与两个父节点(TG、DP)间的条件概率(见表 2)可知: TG 和 DP 均正常时高尿酸的风险为 0.091, TG 正常、DP 升高时高尿酸的风险为 0.137, TG 升高、DP 正常时高尿酸的风险为 0.185, TG、DP 均升高时高尿酸的风险为 0.297;即高尿酸的发生可能与甘油三酯异常、舒张压异常有密切的关系。进一步结合尿酸与其他因素的间接关系分析可知:①甘油三酯升高与高尿酸的关系体现在年龄增大导致空腹血糖升高,从而导致糖化血红蛋白的异常,进而影响了 TG,由于 TG 的升高最终导致了尿酸的升高,由表 3 甘油三酯的各级父节点与其对应的子节点的条件概率表可以清晰地看到这一过程;②舒张压与尿酸间的关系还体现在舒张压与其父节点年龄、收缩压的关系,由表 4 可知:收缩压正常时,随年龄的增大舒张压异常的概率相近;收缩压异常时,舒张压异常的概率随年龄的增大而有减小的趋势,该现象与现有理论相违背,是否由于连续性变量离散化导致的,需要进一步研究。

表 2 高尿酸血症与其父节点关系的条件概率表

TG (mmol/L)	DP (mmHg)	尿酸	
		正常	高尿酸血症
<2.3	<90	0.909	0.091
<2.3	≥90	0.863	0.137
≥2.3	<90	0.815	0.185
≥2.3	≥90	0.703	0.297

离散型贝叶斯网络还提示了其他因素间的关系,如年龄增大导致胆固醇(TC)升高,从而导致低密脂蛋

表 3 甘油三酯与其各级父节点关系的条件概率表

年龄 (岁)	FBG		FBG	HbA1c		HbA1c	LDL	TG	
	正常	升高		正常	升高			正常	升高
<45	0.961	0.039	正常	0.979	0.021	正常	正常	0.838	0.163
45 ~	0.940	0.060	升高	0.444	0.556	正常	升高	0.563	0.437
60 ~	0.884	0.116				升高	正常	0.672	0.328
						升高	升高	0.429	0.571

表 4 舒张压与其父节点关系的条件概率表

年龄 (岁)	SP	DP	
		正常	升高
<45	正常	0.955	0.045
45 ~	正常	0.945	0.055
60 ~	正常	0.965	0.035
<45	升高	0.457	0.543
45 ~	升高	0.483	0.517
60 ~	升高	0.676	0.324

表 5 低密脂蛋白与各级父节点的条件概率表

年龄 (岁)	TC		TC	LDL	
	正常	升高		正常	升高
<45	0.970	0.030	正常	0.970	0.030
45 ~	0.956	0.044	升高	0.086	0.914
60 ~	0.918	0.082			

2. 尿酸代谢水平影响因素的连续型贝叶斯网络

以血尿酸(UA)及其他代谢相关指标的连续性测量值,在检验水准为 0.1 时,采用 IPCB 算法建立连续型贝叶斯网络,如图 3。网络共包含 10 个节点、24 条有向边,主要反映尿酸的水平与其他各代谢指标间的数量关系。

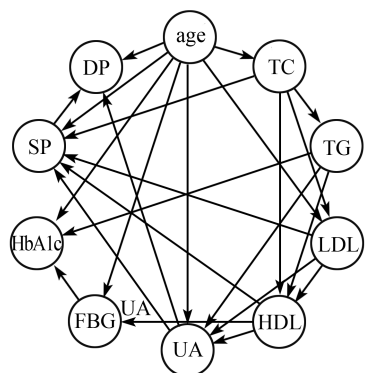


图 3 连续贝叶斯网络结构

图 3 显示与尿酸代谢水平直接相关的因素有 6 个,分别是年龄、甘油三酯(TG)、低密脂蛋白(LDL)、

白升高(见表 5),之后又进入甘油三酯升高而影响高尿酸的途径;概括而言,从与高尿酸相关的间接因素来看,年龄是本质,年龄增大后血压、血脂、血糖异常的风险增加,从而间接影响了尿酸的改变,佐证了年龄是慢性病的危险因素,而且代谢性疾病间都存在一定的相关性;图 2 也可发现,离散型贝叶斯网络也可以描述其他各指标之间的互相关系,如 TC 与 HDL、LDL 的关系, TG 与 SP、HDL 的关系,反映了各类慢性病之间有着复杂的整体效应。

高密度蛋白(HDL)、收缩压(DP)和舒张压(SP),其中,年龄与血脂指标 TG、LDL、HDL 是尿酸的父节点,血压指标 DP、SP 是尿酸的子节点;结合表 6 的偏相关系数可知,年龄的增加、TG 和 LDL 的升高、HDL 的下降均会导致尿酸的水平升高,而尿酸的升高则会导致 SP、DP 的升高。

表 6 尿酸与其直接相关标量的偏相关系数

变量	UA	
	偏相关系数 <i>r</i>	<i>P</i>
年龄	0.026	0.079
TG	0.083	<0.001
LDL	0.028	0.057
HDL	-0.048	0.001
SP	0.084	<0.001
DP	0.147	<0.001

总胆固醇(TC)包括低密度脂蛋白(LDL)、高密度(HDL)和极低密度(VLDL),通过连续贝叶斯网络结构图可知:TC 是尿酸的间接相关因素,结合表 7 可以看出 TC 对尿酸的作用主要体现在 TC 与 TG、LDL、HDL 之间有高度正相关,TC 水平的上升往往伴随着 TG、LDL、HDL 水平的上升,而 TG、LDL、HDL 都是尿酸的直接影响因素,由此可知 TC 间接影响尿酸的水平。

表 7 总胆固醇与其直接相关标量的偏相关系数

变量	TC	
	偏相关系数 <i>r</i>	<i>P</i>
年龄	0.073	<0.001
TG	0.677	<0.001
LDL	0.983	<0.001
HDL	0.881	<0.001
SP	0.040	<0.001

另外,年龄除了直接对尿酸产生作用这一路径外,还可以通过影响 LDL、TC 的水平间接对尿酸的代谢产生影响。

连续贝叶斯网络也可以反映更多其他代谢系统间

的关系。血压 DP、SP 都受到年龄和尿酸的影响,而 SP 除了受年龄和尿酸的影响外,还受到 TC、LDL 和 HDL 的影响,同时 SP 影响着 DP,这些关系反映了年龄、血脂、血尿酸对血压的综合作用;空腹血糖(FBG)受到年龄和 HDL 的影响,FBG 又影响了糖化血红蛋白(HbA1c),而 HbA1c 除了受到 FBG 的影响,也受到年龄和 TG 的影响,提示血糖代谢受年龄和血脂代谢的影响,而与血压、血尿酸之间的关系没有被发现。

以上结果体现了代谢性相关指标间复杂的整体效应,血尿酸的代谢受血脂系统代谢的复杂影响,二

者都受到年龄的影响,年龄还影响了血糖、血压的水平。因此,年龄可能是影响慢性病相关代谢性指标的本质原因。而血脂对血糖、血压、血尿酸都有影响,提示我们血脂代谢在慢性代谢性疾病中占据着非常重要的地位。

3. 高尿酸的离散型贝叶斯网络与尿酸水平的连续型贝叶斯网络比较

针对血尿酸与及其他代谢相关指标间贝叶斯网络的整体效应分析结果,对比离散型与连续型贝叶斯网络的特点,见表 8。

表 8 离散型贝叶斯网络与连续型贝叶斯网络的比较

	离散型贝叶斯网络	连续型贝叶斯网络
目的	研究高尿酸血症的影响因素,探讨高尿酸血症的发病与其他代谢性疾病的关系	研究尿酸水平的影响因素,探讨尿酸水平与相关代谢性指标的关系
建模方法	启发式搜索算法 MMPC 建立网络结构,以 BDeu 评分为准则确定边的方向	基于偏相关的结构学习算法,MDL 评分确定节点之间的边及边的方向
网络属性	反映高尿酸血症的发生与其他疾病或症候类别的关系,属于质性变化关系	反映尿酸水平与相关代谢性指标在数量上的关系,属于量性变化关系
网络复杂度	与高尿酸血症发生的质性关系网络共 14 条边,相对简单	与尿酸水平相关的量性关系网络共 24 条边,相对复杂
直接相关因素	变量的离散化导致信息损失,仅发现 DP 型高血压、TG 型高血脂症直接与高尿酸血症的发生相关,从专业理论的角度看,尚显不足	连续性变量提供更多的信息,发现年龄、TG、LDL、HDL 直接影响了尿酸的水平,而尿酸的水平直接影响了 SP 和 DP 的水平,从专业理论的角度看,合理性更强
与直接相关因素的关联强度	条件概率表描述了高尿酸血症与父节点的关系,TG、DP 均异常患高尿酸血症的风险为 0.297	子节点与父节点间的偏相关系数,描述了与尿酸直接相关的 6 个因素间相关的程度与方向,区分出 4 个影响因素和 2 个结局因素
间接相关因素	年龄与高尿酸血症的多种间接关系:①年龄直接或间接影响 DP 导致高尿酸血症发生;②年龄影响了 TC、LDL、TG 导致高尿酸血症发生;③年龄影响了 FBG、HbA1c 导致高尿酸血症发生	TC 与尿酸是间接关系,主要体现在年龄以不同的方式影响着 TC、TG、LDL、HDL,血脂各指标有着复杂的关系,从而间接影响了尿酸的水平,从专业理论的角度看,合理性更强

讨 论

通过 IPCB 算法构建的连续贝叶斯网络模型可以发现,尿酸(UA)与其他相关代谢指标间存在复杂的网络关系。年龄、甘油三酯(TG)、低密度脂蛋白胆固醇(LDL)、高密度脂蛋白胆固醇(HDL)、舒张压(DP)和收缩压(SP)是尿酸的直接影响因素,而总胆固醇(TC)是尿酸的间接影响因素,它以甘油三酯(TG)、低密度脂蛋白胆固醇(LDL)、高密度脂蛋白胆固醇(HDL)为中间变量对尿酸发挥作用,这是离散型贝叶斯网络模型所不能发现的。以上结果说明了高尿酸血症与高脂血症、高血压之间可能有密切的关系,与现有研究相符^[17-19],提示我们也许可以通过控制血脂水平和血压水平来预防痛风。尽管有研究提出高尿酸血症与糖尿病存在因果关系,但本研究中贝叶斯网络均未发现血糖与尿酸的相关性。

连续型贝叶斯网络模型的分析结果相对于离散型贝叶斯网络模型更具合理性。首先,在尿酸的直接影响因素上,离散型贝叶斯网络只筛选出了两个,即 TG 和 DP,但连续贝叶斯网络又找到了年龄、TC、HDL、LDL 和 DP,揭示了高尿酸血症与高血脂、高血压、年龄之间的关系,显然在发现影响因素的能力上,连续贝叶斯网络相比于离散型贝叶斯网络有明显的优势。其次,连续

贝叶斯网络在各慢性病代谢指标间也比离散贝叶斯网络发现了更多的关系,呈现出了更加具体的,描述多个慢性病指标彼此间关系的整体模型,有利于研究者从更宏观的角度制定高尿酸血症等慢性病的防治策略。

本研究的连续贝叶斯网络模型也存在一定的不足。由于尿酸等指标为连续性变量,连续型贝叶斯网络不能计算类似于离散型贝叶斯网络的条件概率值,因此连续贝叶斯网络只能寻找联系,不能从概率角度量化关联强度。其次,因为本研究在构建连续贝叶斯网络时所使用的结构学习算法 IPCB 是基于偏相关系数的假设检验进行,没有对偏相关系数的强度加以界定,因而发现了一些虽然有统计学意义但相关性较弱的有向边,在临床应用时还需要配合偏相关系数大小合理分析。如何综合考虑偏相关的强度和信度,将是我们接下来对算法加以改进的方向。最后,本次研究使用的横断面调查数据,以数据驱动的方式构建了离散型贝叶斯网络和连续贝叶斯网络,边的方向仅提示可能存在因果关系,其具体意义是否合理还需进一步确定。

参 考 文 献

[1] 关宝生,白雪,王艳秋,等. 痛风/高尿酸血症患者生活习惯的危险因素. 中国老年学杂志, 2014, 34(2):455-457.