

广义相加模型和广义线性模型在糖尿病 相关因素分析中的应用

陈玉柱, 唐振柱, 方志峰, 陆武韬, 李忠友, 周为文, 李晓鹏

广西壮族自治区疾病预防控制中心, 广西 南宁 530028

摘要: **目的** 利用广义相加模型和广义线性模型探讨糖尿病与相关因素的关系。 **方法** 2010–2012 年在广西 5 市/县采用分层整群抽样的方法, 选取 18 岁及以上常住居民作为研究对象。3 827 名被调查者均接受问卷调查, 测量身高、体重、血压、腰围(WC), 检测空腹血糖(FPG)。 **结果** 糖尿病患病率为 9.4%, 男性和女性患病率分别为 10.3%、8.8%, 差异无统计学意义($\chi^2=2.629, P=0.105$)。单因素结果显示年龄、城乡、民族、文化程度、婚姻状况、饮酒、肥胖类型(OBPH) 7 个因素与糖尿病有关($P<0.01$)。多因素 logistic 回归分析结果显示, 农村相对城市有降低患糖尿病的风险($OR=0.633, 95\%CI:0.499\sim0.802, P=0.000$); 60 岁及以上人群与 35 岁以下人群相比患糖尿病风险高($OR=14.037, 95\%CI:6.538\sim30.134, P=0.000$); 中心型肥胖+超重、中心型肥胖+肥胖分别与正常体重比较, 患糖尿病风险高(分别 $OR=2.259, 95\%CI:1.705\sim2.994, P=0.000$; $OR=2.068, 95\%CI:1.368\sim3.125, P=0.001$)。广义相加模型和广义线性模型结果显示饮酒与糖尿病呈现 J 型非线性关系($\chi^2=7.712, P=0.019$), 饮酒 <1 次/周和 ≥ 6 次/周增加患糖尿病的风险; 肥胖类型与糖尿病呈现为平躺 \cap 型曲线关系($\chi^2=13.547, P=0.008$), 中心肥胖和低体重患糖尿病的风险增加。 **结论** 广义相加模型和广义线性模型能直观呈现饮酒、肥胖类型与糖尿病的非线性关系。

关键词: 糖尿病; 广义相加模型; 广义线性模型; 肥胖

中图分类号: R587.1 文献标识码: B 文章编号: 1006-3110(2018)05-0628-04 DOI: 10.3969/j.issn.1006-3110.2018.05.033

Generalized additive and generalized linear models in analysis of diabetes-related factors

CHEN Yu-zhu, TANG Zhen-zhu, FANG Zhi-feng, LU Wu-tao, LI Zhong-you, ZHOU Wei-wen, LI Xiao-peng

Guangxi Autonomous Regional Center for Disease Control and Prevention, Nanning, Guangxi 530028, China

Corresponding author: TANG Zhen-zhu, E-mail: tangzhzh@163.com

Abstract: **Objective** To explore the correlation between diabetes mellitus and its related factors based on generalized additive model and generalized linear model. **Methods** A stratified cluster sampling method was used to select permanent residents aged 18 years and above in 5 cities/counties of Guangxi from 2010 to 2012, and 3,827 selected residents served as the research subjects. A questionnaire survey was conducted, and the residents' height, weight, blood pressure, waist circumference (WC) and fasting plasma glucose (FPG) were measured. **Results** The prevalence rate of diabetes mellitus was 9.4%, and no statistically significant difference was found in the prevalence rate between males and females (10.3% vs. 8.8%, $\chi^2=2.629, P=0.105$). Single factor analysis showed that 7 factors, including age, urban and rural areas, ethnic origin, educational background, marital status, drinking, and obesity phenotypes (OBPH), were associated with diabetes mellitus ($P<0.01$). Multivariate logistic regression analysis indicated that the risk of diabetes mellitus was higher in urban areas than in rural areas ($OR=0.633, 95\%CI:0.499\sim0.802, P=0.000$), higher in residents aged 60 years and above than in ones aged 35 years and below ($OR=14.037, 95\%CI:6.538\sim30.134, P=0.000$) as well as higher in the residents with central obesity & overweight or central obesity & obesity than in ones with normal-weight ($OR=2.259, 95\%CI:1.705\sim2.994, P=0.000$; $OR=2.068, 95\%CI:1.368\sim3.125, P=0.001$). Generalized additive model and generalized linear model revealed that a type-J non-linear relationship was found between drinking and diabetes mellitus ($\chi^2=7.712, P=0.019$), and residents with drinking <1 time per week or ≥ 6 times per week were at an increased risk of developing diabetes mellitus. A type-reversed-tilde linear relationship was found between obesity phenotypes and diabetes mellitus ($\chi^2=13.547, P=0.008$), and residents with central obesity or low weight were at an increased risk of developing diabetes mellitus. **Conclusions** Generalized additive and generalized linear models can directly show the non-linear relationship between diabetes mellitus and alcohol consumption, obesity phenotypes.

Key words: diabetes mellitus; generalized additive model; generalized linear model; obesity

作者简介: 陈玉柱 (1982-), 男, 湖南衡阳人, 硕士研究生, 主管医师, 主要从事营养卫生工作。

通信作者: 唐振柱, E-mail: tangzhzh@163.com。

线性回归、logistic 回归等都是广义线性模型家族的一员,它需要自变量和因变量的关系应当是线性的,否则不满足广义线性模型的应用条件。而当自变量和因变量的关系是非线性时,可以用广义相加模型进行处理^[1-2]。本文根据糖尿病与相关因素的线性和非线性关系,采用相关模型进行分析,为合理利用相关统计学方法探讨相关慢性疾病的影响因素提供借鉴。

1 对象与方法

1.1 调查对象 根据中国居民营养与健康调查 2010-2012 调查方案,于 2010-2012 年在广西南宁市、北海市、凌云、兴安、宾阳 5 市/县采用分层整群抽样的方法,在每个市/县抽取 6 个社区或自然村,再随机抽取 75 户,选取 18 岁及以上常住居民 3 827 人作为为研究对象。

1.2 调查方法与内容

1.2.1 询问调查 采用入户调查的方法,在签署知情同意后,对调查家庭的经济收入、人口进行调查,对个体调查对象的基本情况、一般健康状况、运动、膳食等进行询问调查。

1.2.2 体格检查 采用校正的体重、身高测量仪进行身高和体重测量,站立平稳呼吸时腰围。

1.2.3 指标检测 采集调查对象空腹静脉血,血红蛋白采用氰化高铁法,血糖采用葡萄糖氧化酶法用 721 分光光度计测定。

1.3 诊断标准 (1)肥胖(BMI)和中心型肥胖采用“中国成人超重和肥胖症预防控制指南”推荐的标准,以 BMI≥24 kg/m² 为超重,BMI≥28 kg/m² 为肥胖,以男性腰围≥85 cm,女性腰围≥80 cm 为中心型肥胖。

(2)糖尿病依据为 1999 年 WHO 和国际糖尿病联盟(IDF)糖尿病诊断标准,以空腹血糖(FBG)≥7.0 mmol/L,或有糖尿病史。

1.4 质量控制 (1)人员培训:以调查市或县为单位,对所用调查员和检验员进行统一严格的培训,并考核合格后才进行调查。(2)调查过程的质量控制:成立省级质量控制工作组,市或县调查点安排专人负责质量控制工作,按项目质量控制工作规范及方法,负责调查全过程的质量控制。

1.5 统计分析 数据统计分析采用 SPSS 22.0 对数据进行整理分析。计数资料采用 χ^2 检验,多因素分析采用 logistic 回归分析,检验水准 $\alpha=0.05$ (双侧)。广义相加模型和广义线性模型采用 SAS 9.3 进行统计分析。

2 结果

2.1 一般情况 本次共调查 3 827 人,其中男性 1 611 人,占 44.5%,女性 2 216 人,占 55.5%。男性居民平均年龄为(51.5±15.4)岁,女性平均年龄为(51.5±15.3)岁,差异无统计学意义($t=0.195,P=0.845$)。本次调查居民糖尿病的患病率 9.4%,其中男性为 10.3%,女性为 8.8%,差异无统计学意义($P>0.05$);城镇居民糖尿病的患病率为 13.4%,农村为 5.7%,差异有统计学意义($P<0.01$),城市农村男性相比,差异有统计学意义($P<0.01$),城市农村女性相比,差异有统计学意义($P<0.01$);各年龄段糖尿病患病率差异有统计学意义($P<0.01$),60 岁及以上人群患病率最高。见表 1、表 2。

表 1 2010-2012 年广西不同年龄、性别和城乡糖尿病患病率(n,%)

年龄 (岁)	城市			农村			合计		
	男(n=586)	女(n=994)	小计(n=1 580)	男(n=1 025)	女(n=1 222)	小计(n=2 247)	男(n=1 611)	女(n=2 216)	小计(n=3 827)
18~	82(0.0)	126(1.6)	208(1.0)	188(2.1)	217(0.5)	405(1.2)	270(1.5)	343(0.9)	613(1.1)
35~	61(8.2)	137(2.9)	198(4.5)	229(4.4)	273(4.0)	502(4.2)	290(5.2)	410(3.7)	700(4.3)
45~	193(15.0)	363(10.2)	556(11.9)	338(7.4)	405(5.7)	743(6.5)	531(10.2)	768(7.8)	1 299(8.8)
60~	250(21.6)	368(22.0)	618(21.8)	270(14.4)	327(10.7)	597(12.4)	520(17.9)	695(16.7)	1 215(17.2)
合计	586(15.0)	994(12.5)	1 580(13.4)	1 025(7.6)	1 222(5.7)	2 247(6.6)	1 611(10.3)	2 216(8.8)	3 827(9.4)
χ^2 值	25.203	57.515	80.118	29.423	27.598	56.326	63.343	95.641	157.910
P 值	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

注:男女相比, $\chi^2=2.629,P=0.105$;城市男女相比, $\chi^2=2.051,P=0.152$;农村男女相比, $\chi^2=3.207,P=0.073$;城乡相比, $\chi^2=50.799,P=0.000$;城市农村男性相比, $\chi^2=22.134,P=0.000$;城市农村女性相比, $\chi^2=31.232,P=0.000$ 。

2.2 糖尿病与相关指标单因素分析 单因素结果显示:糖尿病与民族、文化程度、婚姻状况、饮酒、肥胖类型有关($P<0.05$ 或 $P<0.01$);其中饮酒 3~5 次/周患病

率最低;中心型肥胖+超重糖尿病患病率最高。见表 2。

表 2 糖尿病与相关指标单因素分析

项目	调查人数	患病率(%)	χ^2 值	P 值
民族				
汉	2 912	10.3	11.581	0.003
壮族	776	6.7		
其他	139	5.8		
吸烟				
否	3 016	9.8	2.773	0.096
是	811	7.9		
饮酒				
<1 次/周	2 356	9.8	9.624	0.022
1~2 次/周	803	7.6		
3~5 次/周	205	6.3		
几乎每天	463	12.1		
文化程度				
小学及以下	1 632	11.3	18.203	0.000
初中	1 435	7.1		
高中	581	10.7		
大专及以上学历	179	6.7		
肥胖类型(OBPH)				
低体重	316	7.0	90.045	0.000
正常	1 965	6.1		
超重	282	7.4		
中心型肥胖+正常 BMI	271	15.4		
中心型肥胖+超重	722	16.6		
中心型肥胖+肥胖	271	13.3		
家庭经济收入(元)				
<5 000	1 124	8.9	4.617	0.329
5 000~	1 113	8.3		
10 000~	1 065	10.7		
20 000~	373	10.5		
30 000~	152	9.9		
婚姻状况				
未婚	203	1.0	35.830	0.000
在婚	3 239	9.2		
寡居	385	15.8		

2.3 糖尿病与各指标多因素 logistic 分析 在同时纳入城乡、民族、文化程度、年龄、婚姻状况、饮酒、肥胖类型 7 个指标后,城乡、年龄、肥胖类型与糖尿病有关($P<0.01$)。饮酒未进入模型中($P>0.05$)。中心型肥胖+超重、中心型肥胖+肥胖分别与正常体重比较患糖尿病的危险增高($OR=2.259、2.068, 95\%CI: 1.705\sim 2.994、1.368\sim 3.125, P<0.01$)。见表 3。

表 3 糖尿病与各指标多因素 logistic 分析

因素	β	SE	Wald χ^2 值	P 值	OR 值	95.0% CI
城乡	-0.458	0.121	14.307	0.000	0.633	0.499,0.802
年龄(岁)			94.953	0.000		
35~	1.304	0.425	9.409	0.002	3.685	1.601,8.479
45~	1.909	0.395	23.405	0.000	6.749	3.114,14.427
60~	2.642	0.390	45.931	0.000	14.037	6.538,30.134
肥胖类型(OBPH)			40.901	0.000		
低体重	0.155	0.246	0.396	0.529	1.168	0.720,1.892
超重	0.064	0.252	0.065	0.798	1.067	0.651,1.749
中心型肥胖+正常 BMI	0.689	0.197	12.169	0.000	1.992	1.352,2.933
中心型肥胖+超重	0.815	0.144	32.170	0.000	2.259	1.705,2.994
中心型肥胖+肥胖	0.726	0.211	11.878	0.001	2.068	1.368,3.125
常数	-4.385	0.392	124.962	0.000	0.012	

2.4 糖尿病与饮酒频次、肥胖类型的广义相加模型分析 根据单因素和多因素分析结果发现,饮酒频次、肥胖类型与糖尿病为非线性关系。广义相加模型分析显示,表 4 给出以线性参数进入模型的饮酒频次差异无统计学意义($P>0.05$),肥胖类型差异有统计学意义($P<0.01$)。表 5 给出模型非参数部分的光滑成分分析结果。表 6 模型非参数部分的离差分析、图 1 和图 2 显示饮酒频次和肥胖类型对应变量差异均有统计学意义($P<0.05$ 或 $P<0.01$),见表 4~表 6。

表 4 模型参数部分的参数估计

参数	参数估计	标准误	t 值	P 值
截距	-3.026	0.113	26.732	0.000
饮酒	0.077	0.051	1.500	0.133
肥胖类型	0.296	0.035	8.424	0.000

表 5 模型非参数部分的光滑成分分析

成分	光滑参数	自由度	广义交叉确认(GCV)	变量取不同值的数目
饮酒	0.010	1.895	0.144	4
肥胖类型	0.010	3.814	0.063	6

2.5 糖尿病与饮酒频次、肥胖类型的广义线性模型分析结果 广义线性模型分析结果显示,饮酒频次的一次项、二次项,肥胖类型的二次项、三次项均有统计学意义($P<0.05$ 或 0.01)。饮酒频次的一次项为负值,二次项为正值,表明饮酒频次与糖尿病为不规则 J 型曲线,饮酒<1 次/周和 ≥ 6 次/周增加患糖尿病的风险,但以饮酒 ≥ 6 次/周的风险更大。肥胖类型的二次项为正值,三次项为负值,表明肥胖类型与糖尿病为平躺 \cup 型曲线,中心型肥胖+正常 BMI、中心型肥胖+超重、中心型肥胖+肥胖均显著增加患糖尿病的风险,但中心型肥胖+肥胖相比中心型肥胖+超重患糖尿病的风险会降低,低体重患糖尿病的风险有一定程度的增加,见表 7。

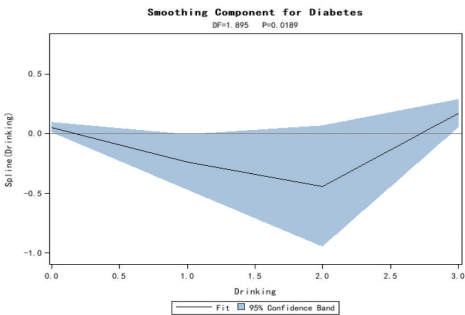


图 1 自变量饮酒的非线性效应图

表 6 模型非参数部分的离差分析

成分	自由度	χ^2 值	P 值
饮酒	1.895	7.712	0.019
肥胖类型	3.814	13.547	0.008

表 7 广义线性模型最大似然比参数估计分析结果

	自由度	参数估计	标准误	95%CI	χ^2 值	P 值
截距	1	-2.601	0.214	-3.020,-2.182	148.273	0.000
饮酒	1	-0.460	0.206	-0.863,-0.057	5.000	0.025
饮酒 * 饮酒	1	0.189	0.071	0.051,0.328	7.186	0.007
肥胖类型	1	-0.597	0.335	-1.253,0.060	3.172	0.075
肥胖类型 * 肥胖类型	1	0.456	0.142	0.178,0.735	10.308	0.001
肥胖类型 * 肥胖类型 * 肥胖类型	1	-0.0616	0.018	-0.097,0.027	11.932	0.001
范围	0	1.000	0.000	1.000,1.000		

3 讨论

本次调查发现,广西部分地区糖尿病患病率为 9.4%,其中男性为 10.3%,女性为 8.8%,与全国平均水平 9.7%基本相同^[3],稍低于杭州萧山区 18 岁及以上居民 10.6%^[4]。本次调查城市居民糖尿病患病率为 13.4%,农村居民患病率为 6.6%。其中城市男性糖尿病患病率最高,达到 15.0%。糖尿病患病率随年龄增加呈上升趋势,以 60 岁及以上人群患病率最高。因此需根据城乡性别、年龄糖尿病患病差异采取有针对性健康教育等干预措施。

本次研究发现饮酒与糖尿病为非线性关系,饮酒 3~5 次/周糖尿病患病率最低。饮酒 <1 次/周和 ≥6 次/周糖尿病患病率均高于其他饮酒频率。国内学者研究发现女性饮酒者糖尿病发病率较不饮酒者低,男性中度饮酒组糖尿病发病率较轻、重度饮酒组低^[5]。Heianza 等^[6]对 1 650 例日本健康男性进行平均 10.2 年的随访,发现每次饮酒量 <23 g 乙醇、饮酒频率 6 次/周,糖尿病的发病率最低,这说明一方面饮酒与糖尿病是非线性关系,在分析时采用广义相加模型是合适的,可以直观呈现它们的关系,另一方面适当饮酒具有降低糖尿病的作用。

BMI 是衡量肥胖的常用指标,而腰围(WC)是定义腹部脂肪分布类型的重要数据,并且不同类型的肥胖与相关疾病的密切程度不同。肥胖与 2 型糖尿病的

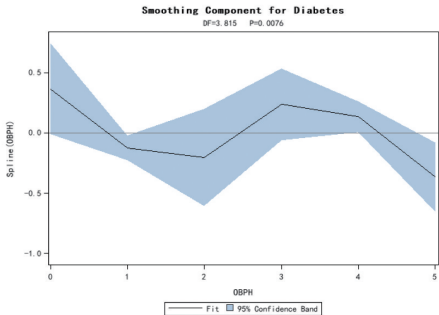


图 2 自变量肥胖类型的非线性效应图

关系密切,肥胖患者发生 2 型糖尿病的风险显著增高。研究发现肥胖和中心型肥胖与血糖水平呈正相关,但中心型肥胖更密切^[7]。杨少玲等^[8]报道,混合型肥胖、腹型肥胖组 2 型糖尿病的患病率显著高于周围肥胖组。孙士杰^[9]研究发现,2 型糖尿病患者超重肥胖病人中,95.8%合并中心型肥胖。本研究综合应用 BMI 和 WC 两指标,把肥胖类型分成 6 个水平,发现一方面合并中心型肥胖,糖尿病患病风险显著增加,这与上述研究结果类似;另一方面合并后的肥胖类型 6 水平与糖尿病不呈直线关系,呈现为平躺∩型曲线,这可能是糖尿病为一种消耗性代谢性疾病,在一些非肥胖的糖尿患者中,长期的糖尿病会导致低体重的增加;而中心型肥胖+正常 BMI,中心型肥胖+超重、中心型肥胖+肥胖糖尿病患病率并没有呈线性增加,这进一步说明中心型肥胖对糖尿病影响的重要性,另一方面中心型肥胖+肥胖人群相对中心型肥胖+超重人群对血糖的代偿机能可能存在差别,具体原因需进一步探讨。

综上所述,城乡、年龄、饮酒和肥胖是糖尿病患病的重要影响因素,针对饮酒、肥胖类型与糖尿病的非线性关系,结合应用广义相加模型和广义线性模型可以准确直观的分析出实际结果。

参考文献

[1] 冯国双,刘德平. 医学研究中的 Logistic 回归分析及 SAS 实现[M]. 第 2 版. 北京:北京大学医学出版社,2015:171-190.

[2] 李丽霞,郜艳晖,周舒冬,等. 广义相加模型及其应用[J]. 中国卫生统计,2007,24(3):243-244.

[3] 国家卫生计生委疾病预防控制中心. 中国居民营养与慢性病状况报告 2015 年[M]. 北京:人民卫生出版社,2015:51-52.

[4] 林君英,蒋园园,葛阳,等. 2014 年杭州市萧山区 18 周岁及以上居民糖尿病患病情况及相关危险因素分析[J]. 实用预防医学,2017,24(2):141-144.

[5] Liu C, Yu Z, Li H, et al. Associations of alcohol consumption with diabetes mellitus and impaired fasting glycemia among middle-aged and elderly Chinese[J]. BMC Public Health, 2010,10(1):1-9.

[6] Heianza Y, Arase Y, Saito K, et al. Role of alcohol drinking pattern in type 2 diabetes in Japanese men: the Toranomon Hospital Health Management Center Study [J]. Am J Clin Nutr, 2013,97(3):561-568.

[7] 赵晓雯,乔虹,袁重胜,等. 哈尔滨市社区成人肥胖状况及其与不同血糖代谢状态的相关性[J]. 中国全科医学,2008,11(6A):969-970.

[8] 杨少玲,朱旅云,胡丽叶,等. 肥胖类型与 2 型糖尿病关系研究[J]. 疑难病杂志,2010,9(3):191-193.

[9] 孙士杰. 2 型糖尿病患者的肥胖流行状况调查[J]. 现代预防医学,2010,37(7):1222-1223.