

# 广义估计方程在糖尿病并发症影响因素分析中的应用\*

吴彬<sup>1</sup> 曹建平<sup>2</sup> 邹煌秀<sup>3</sup> 徐宝川<sup>1</sup> 林佳荣<sup>1</sup>

**【摘要】**目的 探讨广义估计方程在糖尿病并发症影响因素研究中的应用。方法 对 570 例 2 型糖尿病患者的基本情况、生活方式和糖尿病患病情况等进行了问卷调查,采用广义估计方程分析并发症的危险因素。结果 患病时间、血糖水平和体质指数是影响糖尿病并发症发生的主要因素。结论 广义估计方程可较好分析疾病多种结局的影响因素。

**【关键词】** 广义估计方程 糖尿病 并发症 影响因素

## The Application of Generalized Estimating Equations in the Research about Influencing Factors of Complications of Diabetes Mellitus

Wu Bin, Cao Jianping, Zhou Huangxiu, et al (Department of Epidemiology and Health Statistics, Fujian Medical University (350004), Fuzhou)

**【Abstract】 Objective** To investigate the application of generalized estimating equations in the research about influencing factors of complications of diabetes mellitus. **Methods** The basic situation, disease and treatment of 570 diabetes mellitus patients were surveyed and the data were analyzed by generalized estimating equations to select risk factors of diabetes mellitus complications. **Results** Duration of disease, body mass index, blood sugar levels are the main factors affecting DM complications. **Conclusion** Generalized estimating equation can analyze influencing factors of various outcomes of disease.

**【Key words】** Generalized estimating equations; Diabetes mellitus; Complications; Influencing factors

糖尿病最重要的危害就在于其并发症的发生,并发症可导致伤残和早死,严重影响患者的生活质量。分析糖尿病并发症影响因素传统的做法是把并发症作为应变量,各种危险因素作为自变量,采用 logistic 回归模型分析自变量与应变量的关系。一种是设有并发症为 1,无并发症为 0,采用两分类的 logistic 模型进行分析,另一种是把并发症的个数作为应变量,采用有序 logistic 模型来进行统计分析<sup>[1-2]</sup>。但一个糖尿病患者,可能目前没有并发症,也可能伴发一种或多种并发症,不同人并发症的个数相同但种类可能不同,故传统分析把并发症发生情况作为一个应变量不太合适,应该用多个应变量表示并发症发生情况,不同并发症发生可能存在相关性,即多个应变量存在关联。分析此类数据采用上述两种统计方法将丢失大量有用的信息。广义估计方程是 Liang 和 Zeger 在广义线性模型的基础上提出的,它可以分析传统方法无法分析的多个定性应变量且应变量之间存在相关性的资料,得到的参数估计值稳健、可靠<sup>[3]</sup>。本文采用广义估计方程分析糖尿病并发症的危险因素,以期为其他研究者在处理疾病多种结局数据时提供参考。

### 资料与方法

#### 1. 调查对象

选择 2015 年 1 月至 2015 年 2 月到福建省某三甲医院内分泌科就诊的 600 例 2 型糖尿病患者(诊断依据 1999 年 WHO 修订的糖尿病诊断标准<sup>[4]</sup>),纳入标准为病程在 6 个月以上、年龄 18~80 岁、目前无急性并发症的患者,排除文盲病例,糖尿病并发症由门诊医生确诊。

#### 2. 调查内容

采用自行设计调查表进行调查,调查项目包括患者的基本情况:年龄、性别、身高、体重、文化程度、压力状况等;患病及治疗情况:患病年龄、血糖水平、血压水平、用药情况、治疗费用、并发症等;生活方式:吸烟、喝酒、饮食、运动等。血糖和血压水平为离调查时间最近的测定结果。

#### 3. 调查方法

由经过培训的福建医科大学公共卫生学院研究生和本科实习生在门诊医生允许的条件下对就诊的 2 型糖尿病患者进行面对面问卷调查,问卷填写完毕后当场检查回收,收回有效问卷 570 份。

#### 4. 数据统计分析方法和原理

拟合广义估计方程模型<sup>[3,5-7]</sup>。

假设有  $n$  个观察对象,每个观察对象可能发生  $P$  种并发症,第  $i$  个观察对象( $i=1, 2, \dots, n$ )第  $j$ ( $j=1, 2, \dots, p$ )种并发症发生情况记为  $Y_{ij}$ ( $Y_{ij}=0$ ,未患第  $j$ 种

\* 基金项目:全国统计科学研究计划项目(2012LY169);福建医科大学基金项目(2014JY029S)

1. 福建医科大学流行病学与卫生统计系(350001)

2. 福建医科大学卫生管理系

3. 福建省疾控中心健康教育科

并发症,  $Y_{ij} = 1$ , 患第  $j$  种并发症),  $X_{ij} (X_{ij1}, X_{ij2}, \dots, X_{ijm})$  为与  $Y_{ij}$  相对应的自变量向量。不同观察对象之间的观察值相互独立, 同一观察对象的多种并发症发生与否存在相关。广义估计方程的模型结构如下:

(1)  $Y_{ij}$  的期望为  $E(Y_{ij})$ , 有

$$E(Y_{ij}) = \mu_{ij}, g(\mu_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_m x_{ijm}$$

其中  $g(\mu_{ij})$  为联接函数, 可根据数据类型选取合适的联接函数。本次选择 *Binary logistic* 作为联接函数。

(2)  $Y_{ij}$  的方差为  $Var(Y_{ij})$ , 有

$$Var(Y_{ij}) = v(\mu_{ij}) \phi$$

其中  $v(\mu_{ij})$  为已知方差函数,  $\phi$  表示尺度参数, 表示  $Y$  的方差不能被  $v(\mu_{ij})$  解释的部分。

(3) 第  $i$  个观察对象发生  $P$  种并发症存在相关性, 可用  $P \times P$  维“作业相关矩阵” $R_i(\alpha)$  来表示,  $\alpha$  又叫相关参数。

(4)  $R_i(\alpha)$  对应的作业协方差阵为

$$V_i = \phi A_i^{-\frac{1}{2}} R_i(\alpha) A_i^{-\frac{1}{2}}$$

$A_i$  为  $P \times P$  维对角矩阵, 其对角元素为  $v(\mu_{ij})$

广义估计方程为:

$$\sum_{i=1}^n \frac{\partial u_i^T}{\partial \beta} V_i^{-1} (Y_i - \mu_i) = 0$$

$$Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ip})', \mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ip})'$$

模型求解: 假设多个测量值之间无相关性, 按广义线性模型计算  $\beta$  的估计值, 作为迭代运算的  $\beta$  初始值。求出  $\phi, \alpha$  后修正  $\beta$ , 直至收敛。本研究将糖尿病并发症分成视网膜病变、肾脏病变、糖尿病皮肤病、糖尿病足、神经病变、心脑血管病变和其他 7 类。由于无法确定各种并发症的相关结构, 故采用无结构的作业相关矩阵。患者年龄、性别、文化程度、患病时间、吸烟、饮酒、血糖情况等作为自变量拟合广义估计方程模型。应变量和自变量的赋值情况见表 1。所有的数据用 PASW18.0 和 SAS9.0 统计软件分析处理。广义估计方程用 SAS9.0 中的 GENMOD 模块来实现。

### 结 果

1. 一般情况 共发放问卷 600 人, 实际调查 570 人, 有效应答率 95%。其中男性 290 人, 占 50.88%, 女性 280 人, 占 49.12%。年龄 < 40 岁的 31 人, 占 5.44%, 40 ~ 49 岁的 53 人, 占 9.30%, 50 ~ 59 岁的 139 人, 占 24.36%, 60 ~ 69 岁的 189 人, 占 33.16%,  $\geq 70$  岁的 158 人, 占 27.72%。570 名糖尿病患者中, 有 170 人患有并发症, 占总调查患者人数的 29.8%, 其中患有心脑血管和神经病变的并发症人数分别占调查人数的 12.8% 和 12.5%, 各种类型并发症发生情况见表 2。

## 2. 广义估计方程的分析结果

### (1) 单因素分析

以并发症发生情况为**应变量**, 患者的年龄、性别、文化程度、患病时间、吸烟、饮酒、体质量指数(BMI)、规律用药情况、血糖、锻炼情况和压力等为**自变量**拟合广义估计方程, 采用无结构的作业作相关矩阵, 单因素分析发现文化程度、患病时间、吸烟、饮酒、体质量指数、规律用药情况、血糖、锻炼情况、压力和是否将您了解的预防糖尿病并发症知识付之行动对糖尿病并发症发生的影响差异有统计学意义, 详见表 3。

表 1 变量赋值情况

变量	赋值情况
性别	男 = 1, 女 = 2
年龄	岁
文化程度	小学及以下 = 1, 初中 = 2, 中专或高中 = 3, 大专 = 4, 大学本科及以上 = 5
患病时间	< 5 年 = 1, 5 年 ~ = 2, 10 年 ~ = 3, 15 年 ~ = 4, 20 年 ~ = 5
体质量指数	< 18.5 = 1, 18.5 ~ 24.9 = 2, > 24.9 = 3
吸烟情况	不吸烟 = 0, < 10 支/每天 = 1, 10 支 ~ /每天 = 2, 20 支 ~ /每天 = 3, 30 支 ~ /每天 = 4, 40 支 ~ /每天 = 5
饮酒	采用哑变量, 不饮酒作为对比基础 偶尔饮: 是 = 1, 否 = 0 经常饮: 是 = 1, 否 = 0
锻炼情况	经常锻炼 = 1, 很少锻炼 = 0
压力情况	采用哑变量, 压力很小或几乎没有压力作为对比基础 压力比较大: 是 = 1, 否 = 0 压力非常大: 是 = 1, 否 = 0
规律服药情况	规律服药 = 1, 不规律服药 = 0
血压情况	采用哑变量, 大多数时间不正常作为对比基础 几乎正常: 是 = 1, 否 = 0 经常不正常: 是 = 1, 否 = 0
血糖情况	采用哑变量, 大多数时间不正常作为对比基础 几乎正常: 是 = 1, 否 = 0 经常不正常: 是 = 1, 否 = 0
是否将您了解的预防糖尿病并发症知识付之行动	是 = 1, 否 = 0
并发症发生情况	是 = 1, 否 = 0

表 2 570 例糖尿病患者并发症发生情况

并发症的类型	例数	占调查患者总数的比例(%)
视网膜病变	48	8.4
神经病变	71	12.5
糖尿病足	27	4.7
肾脏病变	20	3.5
糖尿病皮肤病	15	2.6
心脑血管病变	73	12.8
其他	2	0.4

### (2) 多因素分析

对单因素分析结果有统计学意义的变量进行共线性诊断, 结果方差扩大因子均小于 10, 条件指数均小于 30, 提示这些自变量不存在共线性。以 7 种并发症发生情况为**应变量**, 患者的文化程度、患病时间、吸烟、饮酒、体质量指数(BMI)、规律服药情况、血糖、锻炼

情况和压力等为自变量拟合广义估计方程,采用无结构的作业相关矩阵,多因素分析结果见表 4。

表 3 单因素广义估计方程分析结果

因素	系数	系数的标准误	Z 值	P 值	OR 值	OR95% 可信区间
文化程度	-0.4762	0.1912	-2.49	0.0142	0.6211	0.4270 ~ 0.9035
患病时间	0.7349	0.1232	5.96	<0.0001	2.0852	1.6379 ~ 2.6548
体质量指数	1.3751	0.4132	3.33	0.0009	3.9554	1.7598 ~ 8.8904
吸烟	0.1776	0.0811	2.19	0.0284	1.1943	1.0188 ~ 1.4001
饮酒						
经常或每天饮:不饮酒	0.7236	0.3515	2.06	0.0391	2.0618	1.0352 ~ 4.1063
偶尔饮:不饮酒	0.3233	0.3372	0.96	0.3373	1.3816	0.7134 ~ 2.6757
规律用药情况	-1.4532	0.5277	-2.75	0.0065	0.2338	0.0831 ~ 0.6577
血糖情况						
几乎正常:大多数时间不正常	-1.7411	0.3367	-5.17	<0.0001	0.1753	0.0906 ~ 0.3392
经常不正常:大多数时间不正常	-0.8374	0.2854	-2.93	0.0033	0.4328	0.2474 ~ 0.7573
锻炼情况	-0.1591	0.0525	-3.03	0.0023	0.8529	0.7695 ~ 0.9453
压力情况	-0.3366	0.1672	-2.01	0.0446	0.7142	0.5146 ~ 0.9911
是否将您了解的预防糖尿病并发症知识付之行动	0.1233	0.0514	2.40	0.0164	1.1312	1.0228 ~ 1.2511

表 4 多因素广义估计方程分析结果

因素	系数	系数的标准误	Z 值	P 值	OR 值	OR95% 可信区间
截距	-5.7855	0.5231	-11.06	<0.0001		
文化程度	-0.3105	0.1833	-1.69	0.0913	0.7331	0.5188 ~ 1.0499
患病时间	0.5680	0.1587	3.58	0.0008	1.7647	1.2930 ~ 2.4086
体质量指数	0.8352	0.4132	2.02	0.0437	2.3052	1.0256 ~ 5.1814
吸烟	0.1087	0.2778	0.39	0.6968	1.1148	0.6467 ~ 1.8804
饮酒						
经常或每天饮:不饮酒	0.4936	0.3294	1.49	0.1365	1.6382	0.8589 ~ 4.8628
偶尔饮:不饮酒	0.2351	0.2976	0.79	0.4293	1.2650	0.7059 ~ 2.2668
规律用药情况	-0.6970	0.4573	-1.52	0.1283	0.4980	0.2779 ~ 1.2205
血糖情况						
几乎正常:大多数时间不正常	-1.3651	0.3453	-3.95	0.0005	0.2553	0.1298 ~ 0.5024
经常不正常:大多数时间不正常	-0.6247	0.2538	-2.46	0.0136	0.5354	0.3256 ~ 0.8805
锻炼情况	-0.1033	0.0637	-1.62	0.1050	0.9018	0.7367 ~ 1.0217
压力情况	-0.1807	0.1672	-1.05	0.2937	0.8346	0.6014 ~ 1.1583
是否将了解的预防糖尿病并发症知识付之行动	0.0956	0.0678	1.41	0.1582	1.1003	0.7928 ~ 1.2567

在对影响因素各参数估计和假设检验中,患病年限长、体质量指数大、血糖不正常的糖尿病患者更容易出现并发症,文化程度、是否锻炼、压力情况、是否将了解的预防糖尿病并发症知识付之行动和规律用药对并发症的影响无统计学意义。

(3) 作业相关矩阵

7 类并发症间的作业相关矩阵见表 5,说明大多数并发症的发生存在一定的正相关关系即发生一种并发症的患者更容易发生其他并发症。

表 5 广义估计方程的作业相关矩阵

变量名	视网膜病变	肾脏病变	糖尿病皮肤病	糖尿病足	神经病变	心脑血管病变	其他
视网膜病变	1.000						
肾脏病变	0.114	1.000					
糖尿病皮肤病	0.147	0.028	1.000				
糖尿病足	0.200	0.092	0.221	1.000			
神经病变	0.230	0.159	0.071	0.166	1.000		
心脑血管病变	0.243	0.127	0.199	0.112	0.142	1.000	
其他	0.018	0.011	0.010	0.013	0.022	0.023	1.000

讨论

在医学研究中,经常会遇见疾病多种结局数据的分析,这类数据与一般资料不同,它的应变量之间存在相关关系,若采用传统的统计分析方法会忽略数据内部的相关性,低估模型参数的标准误,从而高估检验统计量,使得统计结果出现偏差<sup>[3]</sup>。广义估计方程突破了广义线性模型“独立性”条件,解决了数据中应变量相关的问题,能得到稳健的参数估计值。

广义线性模型与经典的线性模型相比虽然不要求应变量服从正态分布、应变量与自变量呈线性关系,但它无法解决多个因变量的问题。混合线性模型虽然能解决多个相关性应变量的问题,但要求应变量服从正态分布且应变量与自变量呈线性关系,当应变量是两分类或多分类定性变量时就不适用了。目前国内对多个相关定性应变量的分析方法为多水平模型和广义估计方程,在弱相关和小样本时,广义估计方程比多水平模型稳定<sup>[5,7-10]</sup>。

一般认为在广义估计方程中只要连接函数选择正确,即使作业相关矩阵的选择不正确,也可以得到稳定的参数估计值<sup>[11]</sup>。然而,Fitzmaurice 的研究结果显示,在模型包含的协变量随时间变化时,若对作业相关矩阵的指定不正确,模型参数估计的效能将降低 40%左右<sup>[12]</sup>。因此,根据实际情况选择合适的作业相关矩阵是比较关键的环节。一般来说,当 1 水平单位数较少且相等的平衡完全设计资料,推荐采用无结构的作业相关矩阵;对重复测量等具有时间顺序的资料,使用与时间相关的作业相关矩阵;对于整群抽样等资料,测量值间无逻辑顺序关系,建议采用等相关结构的相关矩阵;如果难以确定相关结构,可采用无结构的作业工作相关矩阵。本研究数据属最后一种情形,故采用无结构的作业相关矩阵<sup>[3,5]</sup>。

本研究发现血糖异常、体质量指数和患病时间是并发症发生的主要影响因素。这与目前于世鹏等人的研究结果一致<sup>[13-15]</sup>,这提示采用广义估计方程分析多疾病结局资料能取得较为满意的结果。

感谢福建医科大学附属第一医院内分泌科的赵淑好、严孙杰等主任对本次调查给予的指导和帮助。

### 参 考 文 献

[1] 范丽霞. 2 型糖尿病患者慢性并发症特点及影响因素. 中国医院现代远程教育 2012, 10(3): 17-18.  
 [2] 彭宇辉. 二型糖尿病血管并发症的影响因素分析. 福州: 福建医科大学 2007.

大学 2007.  
 [3] 饶克勤主编. 卫生统计方法与应用进展(第二卷). 北京: 人民卫生出版社 2008: 53-170.  
 [4] 钱荣立. 关于糖尿病的新诊断标准及分型. 中国糖尿病杂志, 2008, 8(1): 5-6.  
 [5] 万崇华, 罗家洪主编. 高级医学统计学. 北京: 科学出版社, 2014: 235-369.  
 [6] 洪荣涛, 曹卫华, 欧剑鸣, 等. 多重应答资料统计分析及其 SAS 软件实现过程. 中国卫生统计 2009, 26(6): 595-598.  
 [7] 冯国双, 罗凤基主编. 医学案例统计分析及其 SAS 应用. 北京: 北京大学医学出版社 2011: 222-253.  
 [8] 张华君, 闵捷. 广义估计方程与多水平模型在相关资料中的比较研究. 中国卫生统计 2012, 31(1): 123-125.  
 [9] 柳青. 中华医学统计百科全书多元统计分册. 北京: 中国统计出版社 2013.  
 [10] 毛广运, 沈恬, 陈常中, 等. 广义估计方程在多元统计分析中的运用及检验效率评价. 中国卫生统计 2014, 31(1): 123-125.  
 [11] 张文彤, 田晓燕. 基于广义估计方程的多重应答资料统计分析方法. 中国卫生统计 2004, 21(3): 139-141.  
 [12] Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. Biometrics, 1995, 51(1): 309-317.  
 [13] 于世鹏, 班博, 孙琳, 等. 二级预防对糖尿病高血压患者发生大血管并发症的影响. 中国临床康复 2006, 10(24): 4-6.  
 [14] 周春来, 丘磊. 不同体重指数对初发 2 型糖尿病并发症及治疗方案的影响. 中国医药导报 2014, 11(27): 58-61.  
 [15] 张威. 糖尿病并发心血管疾病的影响因素. 实用心脑血管病杂志 2008, 16(7): 11-12.

(责任编辑: 郭海强)

(上接第 384 页)

[8] 于健, 周玲, 刘春雨, 等. 桂林市部分男性血尿酸水平与代谢综合征相关性研究. 现代预防医学 2012, 39(18): 4903-4905.  
 [9] 王东升, 郑虹, 米树华, 等. 代谢综合征患者的尿酸与血脂脂蛋白相关性研究. 中国综合临床 2006, 22(8): 710-711.  
 [10] 苗志敏, 赵世华, 王颜刚. 山东沿海居民高尿酸血症及痛风的流行病学调查. 中华内分泌代谢杂志 2006, 22(5): 421-425.  
 [11] Lippi G, Monragnana M, Luca Salvagno G, et al. Epidemiological association between uric acid concentration in plasma, lipoprotein(a) and the traditional lipid profile. Clin Cardiol 2010, 33(2): 76-80.  
 [12] Nakanishi N, Tataru K, Nakamura K, et al. Risk factor for the incidence of hyperuricemia: a 6-year longitudinal study of middle aged Japanese men. Int J Epidemiol 1999, 28(5): 888-893.  
 [13] Gonçalves JP, Oliveira A, Severo M, et al. Cross-sectional and longitudinal associations between serum uric acid and metabolic syndrome.

Endocrine 2012, 41(3): 450-457.  
 [14] 马文峰, 陈锦华, 王万山, 等. 广州市某体检人群高尿酸血症患病率及相关危险因素分析. 南方医科大学学报 2012, 32(12): 1812-1815.  
 [15] Ryu S, Chang Y, Zhang Y, et al. A cohort study of hyperuricemia in middle-aged South Korean men. Am J Epidemiol 2012, 175(2): 133-143.  
 [16] Bicek C, Rupprecht HJ, Blankenberg S, et al. Serum uric acid as an independent predictor of mortality in patients with angiographically proven coronary artery disease. Am J Cardiol 2002, 89(8): 12-17.  
 [17] 杨松, 赵彦平, 赵祥海, 等. 农村人群尿酸与血液变化及高血压关系的流行病学分析. 中华疾病控制杂志 2012, 16(7): 564-566.  
 [18] 袁智敏, 张丽崧, 杨丽芳. 广州地区人群高尿酸血症的调查分析. 营养学报 2004, 26(3): 201-203.

(责任编辑: 郭海强)