

# 基于混合线性模型的广西某市县级医院 住院费用影响因素研究<sup>①</sup>

杨博<sup>1</sup>, 郭宇莎<sup>2</sup>, 于雪婷<sup>1</sup>, 何崇伟<sup>1</sup>, 韦征<sup>1</sup>, 黄高明<sup>1②</sup>

(1. 广西医科大学公共卫生学院, 广西 南宁 530021

E-mail: 13878118862@163.com;

2. 广西中医药大学公共卫生与管理学院, 广西 南宁 530200)

**摘要:** **目的** 应用混合线性模型分析广西某市县级医疗机构医疗费用的影响因素, 探索住院医疗费用分析新模式。**方法** 采用 SPSS 23.0 软件对 2016 年 52 112 名在广西某市县级医院住院的患者的住院费用进行描述性分析和多水平混合线性模型的影响因素分析。**结果** 医院地域、经济类型、机构类别以及患者付款方式、性别、年龄、是否有合并症或并发症、是否为手术患者、是否为异地患者、住院时间等与患者医疗费用有关。**结论** 多水平混合线性模型可用于多水平结构数据分析。

**关键词:** 医疗费用; 影响因素; 混合线性模型

中图分类号: R191.1 文献标识码: A 文章编号: 1001-5817(2018)01-0065-04

doi:10.3969/j.issn.1001-5817.2018.01.018

## Research on affecting factors of hospitalization expenses in county hospitals from one city of Guangxi based on mixed linear model

Yang Bo<sup>1</sup>, Guo Yusha<sup>2</sup>, Yu Xueting<sup>1</sup>, He Chongwei<sup>1</sup>, Wei Zheng<sup>1</sup>, Huang Gaoming<sup>1</sup>

(1. School of Public Health of Guangxi Medical University, Nanning 530021

Guangxi, China E-mail: 13878118862@163.com; 2. School of Public Health

and Management of Guangxi University of Chinese

Medicine, Nanning 530200 Guangxi, China)

**Abstract:** **Objective** To analyze the affecting factors of medical expenses in county medical institutions from one city of Guangxi by using mixed linear model and explore the new model for hospitalization medical cost analysis. **Methods** SPSS 23.0 was used to carry out descriptive analysis and multilevel mixed linear model analysis of influence factors of medical costs from 52112 patients who were hospitalized at county hospitals from one city of Guangxi in 2016. **Results** Hospital location, economic types of hospital, hospital category, medical payment methods, gender, age, complications, operations, allopatry, hospital duration were related to patients' medical costs. **Conclusion** The multilevel mixed linear model can be used for the multilevel structure data analysis.

**Key words:** medical costs; affecting factors; mixed linear model

① 基金项目: 广西壮族自治区卫生厅重点科研项目(桂卫重 2011133)

② 通信作者: E-mail: 13878118862@163.com

患者医疗费用资料往往属于层次结构性数据。层次结构数据(hierarchically structured data)广泛存在于医学卫生领域研究中,因为某些特征而具有多个层级或多个水平,也称为多水平数据<sup>[1-2]</sup>。其主要特征是同一层级的观察个体间往往不具备独立性,感兴趣的结局变量在同一单位级别内存在相关性,此时若采用传统统计分析方法,人为地扩大高水平单位特征变量的自由度,或者低估低水平单位间差异,损失大量样本的有效信息且增加犯I类错误的概率<sup>[2]</sup>。混合线性模型(mixed linear model, MLM)虽然保留了一般线性模型的反应变量Y,具有正态性的假定条件,但放弃了独立性和方差齐性的假定,这一模型特别适用于具有多层次结构的资料及重复测量资料<sup>[3-4]</sup>。据此,本文利用混合线性模型分析2016年广西某市县级医疗机构住院患者医疗费用的影响因素,探索混合线性模型在医疗卫生领域层次结构数据中的应用。

## 1 资料与方法

1.1 资料来源 纳入研究范围的标准:2016年1月1日—12月31日在广西某市4个县级行政区划内9家县级医疗机构的全部住院患者。其中,公立医院5家,民营医院4家,综合性医院7家,专科医院2家。收集的数据内容包括医院的行政区划、经济类型、卫生机构类别、医疗付款方式、患者住院次数、性别、年龄、民族、住院时间、主要诊断、其他诊断、是否手术、医疗费用、自付金额等。现住址不全的患者采用户籍地址,利用高德地图API地理信息系统判断患者是否是异地患者,凡患者现住址与就诊医院所在地不同则判为异地患者。除主要诊断外还有其它诊断,无论多少都确定为患者有合并症或并发症。剔除标准:患者实际住院时间小于3d或医疗费用小于1000元的、住址不全者等,最终纳入研究的共52112名患者。

1.2 模型原理 混合线性模型(mixed linear model, MLM)延伸自一般线性模型( $Y = X\beta + \epsilon$ ),引入随机效应部分 $Z\gamma$ ,可表现为: $Y = X\beta + Z\gamma + \epsilon$ 。

式中Y表示反应变量测量值的向量,X为固定效应设计矩阵, $\beta$ 为固定效应参数向量,Z为随机效应设计矩阵, $\gamma$ 为随机效应参数向量,服从均值向量为0、方差/协方差矩阵为G的正态分布,表示为 $\gamma \sim N(0, G)$ , $\epsilon$ 为随机误差向量,服从均值向量为0、方差/协方差矩阵为R的正态分布,即 $\epsilon \sim N(0, R)$ 。

1.3 统计学方法 使用SPSS 23.0软件对资料进行

统计分析,由于原始费用为偏态分布,平均费用采用中位数±四分位数间距(M±Q)表示,组间费用差异采用秩和检验。多因素模型采用混合线性模型,患者医疗费用行对数变换后服从正态分布,作为反应变量Y,自变量中由于行政区划、医院的经济类别和机构类别、付款方式等只有几个固定的取值水平设为固定效应变量,其余变量为随机变量,使用SPSS 23.0软件建立多水平混合线性模型。 $P < 0.05$ 为差异有统计学意义。

## 2 结果

2.1 一般情况 住院患者平均年龄为(43.98±21.89)岁(0~107岁),患者性别比接近1。新农合患者占54.20%、城镇居民医保患者占20.90%,城镇职工医保患者占4.10%,自费患者占20.90%。汉族占97.84%,壮族为0.28%,其他少数民族为1.88%。42.40%的患者有合并症或并发症,19.40%的患者为手术患者,17.20%的患者为异地病源。患者医疗费用中位数为4424.17元(1000.05~234860.98元)。患者住院时间中位数为6d(3~365d)。平均住院次数1.06次,平均自付比40.16%。

2.2 社会因素住院与费用关系 不同区域、医院经济类型、医院类别之间住院费用均存在统计学差异。经两两比较,岑溪市住院费用中位数最高,其次为藤县、苍梧县,蒙山县住院费用中位数最低。公有医院住院费用中位数显著高于私有医院。专科医院住院费用中位数高于综合医院。见表1。

表1 不同社会因素住院费用情况

社会因素	n	M±Q(元)	z	P
区域			886.65	<0.001
岑溪市=1	6342	4843.94±3703.99		
蒙山县=2	8441	3735.19±3946.70		
藤县=3	21497	4640.64±4016.64		
苍梧县=4	15832	4231.9±3834.10		
医院经济类型			16.49	<0.001
公立=1	51165	4450.56±3919.05		
民营=2	947	2775.59±3576.00		
医院类别			5.85	<0.001
综合=1	51550	4419.73±3903.32		
专科=2	562	5807.96±10995.65		

2.3 患者因素与住院费用关系 秩和检验结果显示,各年龄段,不同性别,不同民族,是否有合并症或并发症,是否手术,是否为异地患者,住院时间长短,不同住院次数,以及不同医疗付款方式间医疗费用差异均有统计学意义。住院费用中位数随年龄增大而增大。男

性的住院费用中位数高于女性。壮族的住院费用中位数略高于汉族,两者均显著高于其他少数民族。有合并症或并发症的患者住院费用中位数更高。手术患者住院费用中位数高于非手术患者。本地患者住院费用中位数高于异地患者,住院时间更长,住院费用中位数更高。城镇职工医保患者住院费用中位数最高,其次是新农合患者、自费患者,城镇居民医保患者住院费用中位数最低。见表2。

表2 患者不同影响因素住院费用情况

患者因素	n	M±Q(元)	z	P
年龄组(岁)			7534.778	<0.001
0~17	11129	2878.42±2103.13		
18~39	11088	3947.04±3791.00		
40~59	10502	4947.74±4490.85		
60~107	19393	5299.61±4186.61		
性别			10.653	<0.001
男性	26115	4549.88±4250.39		
女性	25997	4281.09±3663.33		
民族			42.485	<0.001
汉族	50986	3727.43±4460.79		
壮族	145	4496.94±4186.06		
其他民族	981	4435.90±3925.45		
合并症或并发症			49.382	<0.001
是	22093	5201.52±4121.73		
否	30019	3892.11±3547.36		
手术			24.707	<0.001
是	10128	5109.36±4389.86		
否	41984	4277.53±3799.83		
异地患者			3.068	0.002
是	8939	4316.50±4057.59		
否	43173	4442.50±3905.69		
住院时间(d)			126.087	<0.001
<7	32738	3466.72±2482.29		
≥8	19374	7105.54±6010.87		
住院次数			24.455	<0.001
1次	50311	4435.24±3912.60		
2次	1316	4100.00±4911.62		
3次及以上	485	3905.67±4951.93		
付款方式			454.789	<0.001
新农合	28246	4572.32±4114.46		
城镇居民医保	10876	4097.81±3386.62		
城镇职工医保	2119	5371.41±4599.32		
自费	10871	4155.14±3983.36		

2.4 混合线性模型拟合结果 经混合线性模型拟合,结果显示医院行政区划、医院经济类型、医院机构类别、医疗付款方式、患者性别、是否有合并症与并发症、是否为手术患者、是否为异地患者、患者年龄、患者住院时间均为患者医疗费用的影响因素。患者民族、住院次数和自付比对医疗费用没有影响。其中岑溪市和

藤县医疗费用高于苍梧县,蒙山县医疗费用低于苍梧县。新农合患者费用高于自费患者,城镇居民、职工医保患者低于自费患者。公立医院、专科医院、男性、有合并症/并发症、手术、异地患者、年龄和住院天数等均对医疗费用产生正向作用。见表3。

### 3 讨论

由混合线性模型拟合资料结果可知,与苍梧县相比,岑溪市和藤县医疗费用较高,蒙山县医疗费用较低,这可能与区域医疗资源不平衡和区域经济发展水平不同有关。蒙山县距离城市较远,经济水平较低,医疗水平和医疗消费水平相对较低,而岑溪市经济发展水平高,医疗服务水平也较高,有更好的处理重大疾病、疑难杂症的医疗能力。公有医院医疗费用高于民营医院,主要是因为目前民营医院规模小、医疗水平较低,收治的患者病情较轻。专科医院医疗费用更高的原因可能是专科医院就诊的病人多数是恶性肿瘤、心脑血管疾病等重大疾病患者。与自费付款患者相比,新农合患者医疗费用更高,城镇居民和职工医保患者费用更低,这可能是由于农村患者经常拖病拖至危重症才就医的关系。男性患者费用更高,可能与两性健康状况和疾病谱差异有关,南方医科大学郭海燕等<sup>[5]</sup>认为男性慢阻肺和恶性肿瘤的构成比多于女性。异地患者费用更高,可能是因为更加严重的患者才寻求异地就诊。同时,有合并症或并发症的患者和手术患者费用更高可能与病情偏重有关<sup>[6]</sup>,医疗费用与年龄和住院时间成正比,老年人费用更高,住院时间越久费用越高,这些结果均符合常理。

本数据使用混合线性模型拟合效果良好,为结构性数据的层次化分析提供了新的尝试手段。混合线性模型考虑到了结构性卫生数据中可能存在的群体聚集现象,解决了传统线性模型不能适用于测量对象在同一水平存在相关性的问题。在反应变量为正态性连续变量,且数据存在多水平结构时,推荐首选混合线性模型进行拟合。值得注意的是,水平层次和固定效应变量、随机效应变量的选择,应根据具体数据灵活选择。此外重复测量数据和非重复测量数据在模型应用中也有细节上的不同<sup>[7]</sup>。本研究由于原始数据的ICD-10编码不规范,疾病类别没有纳入研究,可能会影响分析结果。

表3 混合线性模型参数估计结果

参数	估计值	标准误	统计量	P	95% CI	
					下限	上限
固定效应						
[岑溪市=1]	0.0938	0.028	3.318	0.023	0.020	0.168
[蒙山县=2]	-0.0690	0.0127	-5.442	<0.001	-0.094	-0.044
[藤县=3]	0.0650	0.0248	2.617	0.036	0.006	0.124
[苍梧县=4](参照)						
[公立医院=1]	0.1119	0.0338	3.312	0.002	0.044	0.180
[民营医院=2](参照)						
[综合医院=1]	-0.2517	0.0309	-8.132	<0.001	-0.312	-0.191
[专科医院=2](参照)						
[新农合=1]	0.0088	0.0029	3.016	0.003	0.003	0.015
[城镇居民医保=2]	-0.0363	0.0043	-8.378	<0.001	-0.044	-0.028
[城镇职工医保=3]	0.0159	0.0052	2.585	0.019	0.015	0.097
[自费=4](参照)						
[男=1]	0.0209	0.0018	11.691	<0.001	0.017	0.024
[女=2](参照)						
[汉族=1]	0.0119	0.0078	1.540	0.124	-0.003	0.027
[壮族=2]	0.0065	0.0169	0.385	0.700	-0.027	0.040
[其他民族=3](参照)						
[有合并症=1]	0.0696	0.0021	33.482	<0.001	0.066	0.074
[无合并症=2](参照)						
[手术=1]	0.1799	0.0025	71.798	<0.001	0.175	0.185
[非手术=2](参照)						
[异地=1]	0.0116	0.0028	4.126	0.000	0.006	0.017
[非异地=2](参照)						
随机效应						
住院次数	0.0007	0.0007	1.092	0.275	0.0001	0.004
年龄	6.53E-06	3.29E-06	1.986	0.047	2.43E-06	1.75E-05
住院天数	0.0003	0.0001	2.120	0.034	0.0001	0.001
自付比例	0.0053	0.0032	1.673	0.094	0.002	0.0172

## 参考文献:

- [1] Veturi Y, Kump K, Walsh E, et al. Multivariate mixed linear model analysis of longitudinal data: an information-rich statistical technique for analyzing plant disease resistance[J]. *Phytopathology*, 2012, 102(11): 1016-1025.
- [2] 张莉, 黄莉, 熊昌辉, 等. 混合线性模型在婴幼儿生长发育研究中的应用[J]. *中国卫生统计*, 2015, 32(1): 10-13.
- [3] 余松林. 混合线性模型的应用[J]. *中国医院统计*, 2006, 13(1): 70-75.
- [4] 钱莎莎, 邢健男, 王璐. 多水平统计模型分析方法及其应用[J]. *中国公共卫生*, 2017, 33(9): 1414-1416.
- [5] 郭海燕, 檀谊洪, 梅婷, 等. 广东省某三甲医院老年人疾病谱及性别差异[J]. *广州医药*, 2016, 47(1): 29-32.
- [6] 董佩, 毛阿燕, 邱五七, 等. 北京市本地与外地肺癌患者就诊及医疗费用分析[J]. *中国病案*, 2017, 18(1): 47-50.
- [7] Gill PS. A robust mixed linear model analysis for longitudinal data[J]. *Stat Med*, 2000, 19(7): 975-987.

收稿日期: 2018-01-10