

Logistic 回归在预约挂号失约行为预测的应用

李静, 宗诚* (江苏省中医院监察室, 江苏 南京 210029)

摘要:目前预约挂号模式在公立三甲医院中已普遍实行,但预约挂号中的失约行为也十分常见。为有效预测预约挂号失约的可能性,维护医院的医疗秩序,从江苏省某三甲中医院门诊预约挂号信息中提取相关数据作为样本,运用卡方检验对可能影响患者失约行为的8个因素进行检验分析,提取出预约挂号失约行为的影响因素;通过 Logistic 回归模型,找出可能存在的规律,构建 Logistic 回归方程,最终达到对未来预约挂号失约行为的有效预测。

关键词:预约挂号;失约行为;Logistic 回归

中图分类号:R197.3 **文献标志码:**A **文章编号:**1009-3222(2017)02-0126-03

依托现代社会传媒的便利性和有效性,目前一般公立三甲医院纷纷建立新的挂号模式——预约挂号,患者足不出户就可以通过电话、网络、手机等方式成功预约门诊号。这种预约模式在方便患者就医、提高患者满意度、提升医院管理水平等方面具有重大意义^[1]。随着预约挂号的普及,预约挂号中的失约行为也十分常见。很多患者成功预约挂号之后,因为各种原因无法按照预约的时间前来就诊,从患者个人角度出发,一次失约行为不可能对患者的生活产生不利影响,但是从整个医院的医疗秩序出发,很多患者的失约行为会造成门诊号源的浪费,很多需要就诊的患者因为挂不到号而耽误诊治。本文通过医院预约挂号的历史数据,运用 Logistic 回归模型进行有效预测,针对可能发生的结果采取有效的预防措施,以期大大减少失约行为造成的医疗资源损失。

1 数据来源与研究方法

1.1 数据来源

本文以江苏省某三甲中医院门诊预约挂号为例,对患者的失约行为进行研究预测。样本数据来源于医院2016年6月1日的全部预约挂号患者信息,当天预约挂号2152人次,其中失约320人次,就诊1832人次。提取包括“性

别”“年龄”“地区”“预约号源顺序”“预约号源时段”“预约途径”“预约医生级别”“预约提前期”等8个可能影响“预约结果”的因素,并制定其量化规则^[2]。

1.2 研究方法

Logistic 回归模型常常应用于医学领域,在流行病学研究中,经常被用来分析疾病与各危险因素之间的定量关系,随着研究的拓展,Logistic 回归模型在患者行为类领域也得到了应用。本文利用 SPSS 软件对预约挂号的历史数据进行汇总,对可能影响患者失约行为的8个因素进行单因素分析,筛选出影响失约行为的自变量,然后运用 Logistic 回归模型建立预测模型,最终达到对未来预约挂号失约行为的有效预测^[3]。

1.3 Logistic 回归模型

考虑具有 n 个独立变量的向量 $x=(x_1, x_2, \dots, x_n)$, 设条件概率 $P(y=1|x)=p$ 为根据观测量相对于某事件 X 发生的概率, Logistic 函数可以表示为:

$$P = \frac{1}{1 + e^{-\epsilon}} \tag{1}$$

其中 $\epsilon = \alpha + \beta_1 x_1 + \dots + \beta_n x_n$ 。Logistic 函数的取值一定是介于 0 到 1 之间的数值,保

收稿日期:2017-03-04;修稿日期:2017-04-02

基金项目:江苏省卫生计生委科研课题(JCZ201520)

作者简介:李静(1987-),女,江苏宜兴人,江苏省中医院助理研究员。*通信作者:jsszjc2004@sina.com

证了 Logistic 回归模型估计的事件概率也是介于 0 到 1 之间,符合随机规律的定义。 p 是事件发生的概率, $1 - p$ 是事件不发生的概率, $\frac{p}{1-p} = e^\epsilon$, 这个比值称为事件发生的发生比 odds, 将公式转化, 得到:

$$\ln\left(\frac{p}{1-p}\right) = \epsilon \quad (2)$$

2 研究结果

2.1 预约挂号信息变量、量化规则及其分析情况

对医院预约挂号数据进行模型重建后,需

要考虑模型的拟合程度。经过卡方检验,在“性别”“预约号源顺序”“预约时段”3 个变量中, P 值均大于 0.05, 患者预约挂号就诊行为与失约行为之间不存在统计学差异, 即这 3 个变量不是患者预约挂号失约行为的影响因素。在“预约途径”“预约医生级别”“地区”“年龄”“预约提前期”5 个变量中, P 值均小于 0.05, 患者预约挂号就诊行为与失约行为之间存在统计学差异, 即这 5 个变量是患者预约挂号失约行为的影响因素, 根据影响患者预约挂号失约行为的 5 个因素建立 Logistic 回归模型。见表 1。

表 1 预约挂号变量、量化规则及其预约患者信息情况

变量	量化规则	失约 = 1	就诊 = 0	合计	χ^2	P
性别	女 = 1	211	1 245	1 456	0.508	0.476
	男 = 2	109	587	696		
年龄	0~20 岁 = 1	34	124	158	14.340	0.002
	20~40 岁 = 2	164	824	988		
	40~60 岁 = 3	76	518	594		
	60 岁以上 = 4	46	366	412		
地区	本地人 = 1	128	891	1 019	25.323	0.000
	外地人 = 2	192	941	1 133		
预约号源顺序	20 号之前 = 1	184	1 118	1 302	2.203	0.332
	20~40 号 = 2	90	499	589		
	40 号之后 = 3	46	215	261		
预约时段	上午 = 1	214	1 114	1 328	0.229	0.632
	下午 = 2	106	718	824		
预约途径	现场预约 = 1	3	30	33	46.767	0.000
	自助机预约 = 2	19	120	139		
	12320 电话预约 = 3	62	244	306		
	12580 电话预约 = 4	68	557	625		
	手机 APP 预约 = 5	102	341	443		
	医院官网预约 = 6	66	540	606		
预约医生级别	副主任(中)医师 = 1	2	16	18	10.258	0.016
	主任(中)医师 = 2	11	120	131		
	专家 = 3	292	1 547	1 839		
	高级专家 = 4	15	149	164		
预约提前期	提前 0~5 天 = 1	182	912	1 094	19.851	0.000
	提前 5~10 天 = 2	66	282	348		
	提前 10~14 天 = 3	72	638	710		
合计		320	1 832	2 152		

2.2 模型系数的混合检验

模型建立后, 需要考虑和比较模型的拟合程度。有三个度量值可作为拟合的判断根据, 主要是针对步骤、块和模型进行的卡方检验, 检

验是否显著相关。样本数据中显著性水平为 0.05, 自由度 $df = 5$, 通过公式可以查出卡方临界值为 11.07, 从表 2 中给出的卡方值 29.710, 大于卡方临界值, 并且 P 值即 Sig. 值小于 0.05,

因此在显著性水平为 0.05 的情况下,样本模型系数的混合检验显著相关。

表 2 模型系数的综合检验

	卡方	df	Sig.
步骤	29.710	5	0.000
块	29.710	5	0.000
模型	29.710	5	0.000

2.3 Logistic 回归拟合方程

表 3 的数据是最终拟合的结果,地区、年龄和预约提前期三个变量入选,如表中的数据所示,地区变量 $P = 0.05 \leq 0.05$,年龄变量 $P = 0.003 < 0.05$,预约提前期变量 $P = 0.002 < 0.05$,拒绝原假设(即原假设为与地区变量无关,与年龄变量无关,与预约提前期变量无关),认为预约挂号失约行为与这三个变量相关。而预约途径变量 $P = 0.163 > 0.05$,预约医生级别变量 $P = 0.994 > 0.05$,接受原假设,认为预约挂号失约行为与预约途径变量和预约医生级别变量无关。可以建立如下线性关系:

$$\epsilon = 0.708 + 0.250 \times \text{地区} - 0.224 \times \text{年龄} - 0.218 \times \text{预约提前期} \quad (3)$$

其中预约提前期分别用数字 1~3 表示 3 种不同的预约时间,数字 1 表示提前 0~5 天预约,数字 2 表示提前 5~10 天预约,数字 3 表示提前 10~14 天预约。

将公式(3)的结果代入公式(1),拟合方程的结果一定是介于 0 到 1 之间的数值,SPSS 软件将因变量“预约结果”的变量转换为 0、1 数值,根据其编码原则,“失约”赋值为 1,“就诊”赋值为 0,该方程可以作为医院预约挂号行为预测,若预测值大于 0.5 说明预约患者有可能会失约,预测值小于 0.5 说明预约患者有可能会不会失约。

在 Logistic 回归模型中,将公式(3)的结果代入公式(2),即:

$$\ln\left(\frac{p}{1-p}\right) = 0.708 + 0.250 \times \text{地区} - 0.224 \times \text{年龄}$$

$$- 0.218 \times \text{预约提前期} \quad (4)$$

以年龄为例,年龄赋值提高一个单位,事件发生比的对数减少 0.224,事件发生比增加 $\exp(-0.224) = 0.799$,因此我们可以说,年龄赋值提高一个单位,事件发生比增加 7.99%。以此类推,地区赋值提高一个单位,事件发生比增加 12.84%;预约提前期赋值提高一个单位,事件发生比增加 8.04%。

表 3 最终方程中的变量

	B	S.E.	Wald	df	Sig.	Exp(B)
预约途径	-0.006	0.005	1.943	1	0.163	0.994
预约医生级别	-.001	0.153	0.000	1	0.994	0.999
地区	0.250	0.128	3.807	1	0.050	1.284
年龄	-0.224	0.074	9.011	1	0.003	0.800
预约提前期	-0.218	0.071	9.412	1	0.002	0.804
常量	0.708	0.660	1.151	1	0.283	2.03

3 结论

本文通过预约挂号历史数据的汇总,SPSS 软件的分析处理,得到了 Logistic 回归拟合方程。构建模型的目的就是通过已知患者的相关信息,合理地预测出患者预约挂号失约行为的可能性,如果满足失约行为的发生概率,医院可以提前采取预防措施。例如提前短信患者就诊时间,或者提醒患者在不能就诊的情况下提前做好预约号源的退号工作,这样医院就能被动变主动,在有限的范围内合理配置医疗资源,避免资源浪费,让更多患者成功就诊。

参考文献:

[1]刘姿,张秀兰,唐泽华,等.预约挂号爽约患者情况调查与对策探讨[J].中国医院管理,2010,30(6):1.
 [2]周颖,罗利,谭明英,等.基于因子聚类分析的门诊患者爽约影响因素研究[J].中国卫生事业管理,2016,28(1):22-24.
 [3]巩亚楠,帕提麦·马秉成,朱登浩,等.随机森林与 Logistic 回归在预约挂号失约影响因素预测中的应用[J].现代预防医学,2014,41(5):769-772.