

本文引用:覃悦,薛花,岳新霞.采用主成分改进的 Logistic 回归分析新疆乌鲁木齐地区学龄前儿童龋齿影响因素的研究[J].新疆医科大学学报,2018,41(5):639—642. doi:10.3969/j.issn.1009—5551.2018.05.025

# 采用主成分改进的 Logistic 回归分析新疆乌鲁木齐地区学龄前儿童龋齿影响因素的研究

覃悦,薛花,岳新霞

(新疆维吾尔自治区人民医院北院护理部,乌鲁木齐 830000)

**摘要:**目的 调查新疆乌鲁木齐地区学龄前儿童龋齿发病情况和相关影响因素。方法 以“区—幼儿园—班级”的技术路线,从新疆乌鲁木齐7区1县随机抽取1所幼儿园,每所幼儿园中随机整群抽取大、中、小各2个班,每班再采取整群抽样,由口腔科医生完成龋齿检查和问卷调查。对原始变量进行共线性诊断,再采用主成分改进 Logistic 回归分析筛选可能的影响因素。结果 共有1366位学龄前儿童纳入研究,有809位儿童患有龋齿,患病率为59.2%。共线性诊断提示各变量间存在共线性问题,采用主成分改进 Logistic 回归分析筛选出7个有意义的因素,开始刷牙年龄、每日刷牙次数、饭后漱口、父母亲文化程度和家庭收入是儿童龋齿发生的保护因素( $OR = 0.657, 0.853, 0.426, 0.675, 0.725, 0.768$ ),吃零食则是龋齿发生的危险因素( $OR = 1.334$ )。结论 新疆乌鲁木齐地区学龄前儿童龋齿发生受多因素影响,通过主成分改进的 Logistic 回归分析可消除原始变量间存在的共线性问题,得到更好的回归模型。

**关键词:**主成分; Logistic 回归; 龋齿; 影响因素

中图分类号:R788.1 文献标识码:A 文章编号:1009—5551(2018)05-0639-04

doi:10.3969/j.issn.1009—5551.2018.05.025

## A study on the influence factors of dental caries of preschool children in Urumqi area of Xinjiang by Logistic regression with improved principal component

QIN Yue, XUE Hua, YUE Xinxia

(Department of Nursing, The North Courtyard Area of the People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830000, China)

**Abstract: Objective** To investigate the incidence and related influencing factors of dental caries among preschool children in Urumqi area of Xinjiang. **Methods** According to the “district-kindergarten-class” technical route, a kindergarten was randomly sampled in every cluster of district/county from 7 Districts and 1 County of Xinjiang Urumqi, and 2 classes were randomly sampled in every cluster of small/medium/large scaled classes from the sampled kindergartens. Everyone out of the sampled class was conducted dental caries examination and questionnaires by the dentist. A collinear diagnosis of the original variables were carried out, and then the possible influencing factors were screened by principal component improved logistic regression analysis. **Results** A total of 1366 preschool children were included in the study. 809 children were suffering from dental caries, the prevalence rate was 59.2%. Collinearity diagnosis indicates the presence of multicollinearity among variables. Improved logistic regression analysis identified 7 signifi-

基金项目:新疆维吾尔自治区自然科学基金(2017D01C135)

作者简介:覃悦(1980—),女,本科,主管护师,研究方向:护理管理。

通信作者:岳新霞,女,公共卫生硕士,主任护师,研究方向:公共卫生,E-mail:874169549@qq.com。

cant impact factors by principal component, and the teeth-brushing starting age, daily brushing frequency, mouth-washing after meals, parents' education level and family income were the protective factors of dental caries in children ( $OR = 0.657, 0.853, 0.426, 0.675, 0.725, 0.768$ ), while snack-eating was the risk factor for dental caries ( $OR = 1.334$ ). **Conclusion** The incidence of dental caries in preschool children in Urumqi area of Xinjiang is affected by many factors. By means of logistic regression analysis of principal components, the collinearity problem of original variables can be eliminated, and a better regression model can be obtained.

**Keywords:** principal component; Logistic regression; dental caries; factor

龋齿是在细菌为主要作用下多因素参与造成的牙齿硬组织发生破坏的常见疾病<sup>[1]</sup>。儿童学龄前时期是恒牙萌出替换乳牙、颌骨和牙弓主要发育阶段。龋齿会影响恒牙正常萌生,损伤口腔黏膜组织,影响儿童咀嚼功能而造成营养摄入障碍,影响发音等<sup>[2]</sup>。如何正确防治学龄前儿童口腔龋齿非常关键。本研究主要调查新疆乌鲁木齐地区学龄前儿童龋齿发病情况和相关影响因素,旨在对正确防治龋齿提供实验依据。

## 1 研究对象与方法

**1.1 研究对象** 按照第三次全国口腔健康流行病学抽样调查的要求<sup>[3]</sup>,采取分层随机整群抽样的方法,以“区—幼儿园—班级”的技术路线,在乌鲁木齐地区逐级抽取样本。首先将乌鲁木齐市区所有教育局注册幼儿园编号,从7区1县随机抽取1所幼儿园,再从每所幼儿园中随机整群抽取大、中、小各2个班,每班再采取整群抽样,共有1366位学龄前儿童纳入研究。年龄段为3~6岁,平均年龄为(4.66±1.05)岁。男童703人(51.5%),女童663人(48.5%)。按照顺序对全部牙齿进行细致检查,检查内容包括龋齿发生个数和龋齿类型(浅层龋齿、中层龋齿、深层龋齿)。诊断标准根据国家龋齿统一标准规定,凡是牙齿表面或齿沟处有色、形、质3方面改变的即诊断为龋齿,乳牙龋、失和补也均作为龋齿统计<sup>[4]</sup>。

**1.2 主成分改进的 Logistic 回归分析** 学龄前儿童龋齿发生影响因素 依据前述检查结果将有龋齿患儿分为实验组,无龋齿患儿设为对照组。通过可能影响龋齿发生的因素自行制作调查问卷,主要参考《第三次全国口腔健康流行病学调查方案》口腔调查问卷标准<sup>[5]</sup>,结合龋病病因学、国内外学者相关研究结果等。调查问卷内容设定共有5大类,即:(1)社会人口学资料:年龄、性别、民族等。(2)生长发育情况:孕期母亲健康状况、是否早产儿等。(3)饮食习惯:婴幼儿时期主要喂养食品、哺乳持续时间、是否

吃零食、吃零食的时间等;(4)口腔卫生情况:开始刷牙的年龄、每天刷牙次数、通常使用的牙膏、饭后漱口情况、幼儿园是否有口腔预防教育等。(5)教育情况:父母、老师教育程度、家庭收入、家长和老师对口腔健康知识的知晓程度等。由经过专业培训的医务人员发送给儿童父母或监护人进行调查。采取主成分改进的 Logistic 回归分析相关影响因素。

**1.3 统计学分析** 采用 SPSS21.0 统计学软件进行单因素、主成份改进 Logistic 回归分析,计量资料采用均数±标准差( $\bar{x} \pm s$ )表示,以  $P < 0.05$  为差异有统计学意义。

## 2 结果

**2.1 一般调查结果** 共1366位儿童参与龋齿调查,发放问卷1366份,回收1366份,回收率为100%。患病有809位儿童,患病率为59.2%,平均年龄为(4.87±0.97)岁,其中男童有413位(51.1%),女童有396位(48.9%)。809位龋齿儿童中,发生1颗龋齿者有436位(53.9%),2颗龋齿者有295位(36.5%),3颗及以上者有78位(9.64%)。龋齿类型上,浅层龋齿有261位(32.3%),中层龋齿有352位(43.5%),深层龋齿有196位(24.2%)。

**2.2 单因素分析结果** 对照组赋值为0,实验组赋值为1,采取卡方检验进行单因素检验,龋齿影响因素赋值见表1。筛选出7个有意义的影响因素,即开始刷牙年龄、每日刷牙次数、饭后漱口、吃零食、家庭收入和父母文化程度。

**2.3 Logistic 回归共线性诊断** 在进入多因素 Logistic 回归分析前,对选取可能影响儿童龋齿的原始16个变量因素做共线性诊断,见表2,结果可示,有4个指标特征根接近0,条件指数有3个>30,且同样3个方差膨胀系数>10,认为原始变量间存在多重共线性问题,不能直接进行 Logistic 回归分析。为解决观察指标间共线性对回归模型的影响,而采用主成分改进的 Logistic 回归分析。

表1 龋齿相关影响因素赋值说明

因素	赋值说明
性别	男=0,女=1
年龄/岁	3~4=0,4~5=1,5~6=2
民族	汉族=0,少数民族=1
孕期母亲健康状况	健康=0,相关疾病=1
早产	是=0,否=1
婴幼儿喂养	母乳=0,奶粉=1,其他=2
吃零食	是=0,否=1
开始刷牙年龄	从不刷牙=0,<3岁=1,≥3岁=2
每天刷牙次数/次	0=0,1=1,≥2=2
饭后漱口	无或偶尔=0,经常=1
幼儿园口腔教育	是=0,否=1
父亲文化程度	初中及以下=0,高中=2,大学及以上=2
母亲文化程度	初中及以下=0,高中=2,大学及以上=2
父母对口腔卫生知晓程度	了解=0不清楚=1
老师对口腔卫生知晓程度	了解=0不清楚=1
家庭收入1万元	<5=0,5~10=1,≥10=2

表2 16个原始变量共线性诊断结果

编号	特征根	条件指数	方差膨胀因子
1	0.865	10.265	3.526
2	2.352	6.574	4.215
3	0.251	14.227	5.362
4	0.042	34.287	18.854
5	0.658	12.425	8.526
6	0.752	10.252	7.962
7	0.684	11.247	8.047
8	1.124	7.536	4.763
9	0.853	9.654	7.344
10	0.027	41.862	28.625
11	0.683	11.862	8.347
12	8.862	1.258	1.478
13	5.324	2.143	2.586
14	0.351	22.536	9.587
15	0.044	35.285	22.853
16	5.698	1.867	2.354

2.4 主成分改进的 Logistic 回归分析 对单因素筛选出有意义的7个影响因素进行主成分分析,见表3,选取特征根接近1和累计>0.85的5个主成分<sup>[6]</sup>,以其主成分替代原变量进行 Logistic 回归分析,最后进入回归模型的因素有开始刷牙年龄、每日刷牙次数、饭后漱口、吃零食、父母亲文化程度和家庭收入等7个影响因素,其OR值分别为0.657、0.853、0.426、1.334、0.675、0.725、0.768,见表4。拟合方程为:Logit(P) = 0.834 7 - 0.627X<sub>1</sub> - 0.806X<sub>2</sub> - 1.165X<sub>3</sub> + 1.268X<sub>4</sub> - 0.743X<sub>5</sub> - 0.722X<sub>6</sub> - 0.925X<sub>7</sub>。

表3 主成分分析结果

主成分	特征根	贡献	累计贡献
1	1.852	0.264	0.264
2	1.524	0.212	0.472
3	1.126	0.155	0.627
4	0.883	0.129	0.756
5	0.808	0.107	0.863
6	0.681	0.089	0.952
7	0.557	0.048	1.000

表4 主成分改进 Logistic 回归分析结果

变量	标准化回归系数β	标准化回归标准差	P值	OR值	95%CI
开始刷牙年龄(≥3岁)	-0.627	0.153	<0.01	0.657	0.426~0.869
每日刷牙次数(≥2次)	-0.806	0.142	<0.01	0.853	0.627~1.142
饭后漱口(经常)	-1.165	0.176	<0.01	0.426	0.0324~0.773
吃零食(是)	1.268	0.244	<0.01	1.334	0.882~1.782
母亲文化程度(大学及以上)	-0.743	0.202	<0.01	0.675	0.417~0.915
父亲文化程度(大学及以上)	-0.722	0.189	0.018	0.725	0.446~1.074
家庭收入(≥1万元)	-0.925	0.143	0.022	0.768	0.512~0.976

### 3 讨论

龋齿又称“虫牙”或“蛀牙”,是口腔牙齿发生细菌性感染的常见疾病。目前多认为细菌感染、口腔卫生、宿主和时间四因素学说是其发病的主要原因<sup>[7]</sup>,可能的机理认为细菌在牙齿表面形成牙菌斑,可发酵口腔内食物残渣中糖而产生酸,继而导致牙齿脱矿,破坏有机质而产生龋洞<sup>[8]</sup>。可以引起牙龈炎和根尖周炎,甚者可出现牙槽骨和颌骨炎症,严重者会对儿童生长发育产生重大影响<sup>[9]</sup>。龋齿发病率高,分布区域广,被世界卫生组织列于肿瘤和心血管

疾病人类三大重点防治疾病<sup>[10]</sup>。因此研究相关区域龋齿的发病情况与相关影响因素是非常必要的,这对于防治龋齿可提供进一步实验依据。

Logistic 回归模型常用于病因分析、生存分析的统计学方法<sup>[11]</sup>。但是如果变量间存在多重共线性问题,采用这样的数据直接使用 Logistic 回归分析,就会增大估计参数的均方误差和标准误<sup>[12]</sup>,使许多有意义的阳性因子被错误地剔除出回归模型,有的甚至使回归系数的方向相反,从而使得原本起到危险作用的危险因素变为保护性因素<sup>[13]</sup>,出现难

以解释的现象,结果很不可信。而经过主成分变换的 Logistic 回归分析方法,通过主成分变换可消除解释变量观察矩阵之间的共线关系<sup>[14]</sup>,满足各自变量彼此相对独立,可得出了更为合理的回归模型<sup>[15]</sup>。

本研究采取了“区—幼儿园—班级”的技术路线,在乌鲁木齐7区1县逐级抽取样本,符合区域调查研究要求<sup>[15]</sup>。共有1366位3~6岁学龄前儿童参与了我们的调查,最终发现809位儿童患有龋齿,发病率为59.2%,这与第三次全国口腔健康流行病学调查的城市龋齿发病率(79.53%)相比有所减少。将龋齿儿童分为实验组,非龋齿儿童列为对照组,分析龋齿发病的影响因素,单因素卡方检验筛选出7个有意义的影响因素,即开始刷牙年龄、每日刷牙次数、饭后漱口、吃零食、家庭收入和父母文化程度。实验中选取特征根接近1和累计贡献>0.85的5个主成分,进入最后回归方程因素的标准误都较小,符合 logistic 回归分析一般要求<sup>[16]</sup>。同时也可以更加合理得出本次调查的结论,开始刷牙年龄、每日刷牙次数、饭后漱口、父母亲文化程度和家庭收入是儿童龋齿发生的保护因素( $OR = 0.657, 0.853, 0.426, 0.675, 0.725, 0.768$ ),而吃零食则是龋齿发生的危险因素( $OR = 1.334$ )。

综上所述,本研究主要调查了新疆乌鲁木齐地区学龄前儿童龋齿患病情况和相关影响因素,通过主成分改进的 Logistic 回归分析来处理原始变量间存在的共线性问题,并且通过分层随机整群抽样获取的大量样本和对照组研究模式进一步分析确认新疆乌鲁木齐地区学龄前儿童龋齿发生受多因素影响<sup>[17]</sup>,主要包括有开始刷牙年龄、吃零食、每日刷牙次数、饭后漱口、幼儿园口腔教育、父母亲文化程度和家庭收入等影响因素。

参考文献:

[1] 刘芳,冯锦虹,仵楠,等.维吾尔族和汉族不同龋敏感儿童远缘

链球菌的荧光定量 PCR 检测[J].新疆医科大学学报,2013,36(4):494-499.

- [2] 袁爽,施乐,吕进.上海市静安区3岁儿童龋齿患病情况调查分析[J].上海口腔医学,2017,26(1):118-120.
- [3] 辜双娇,林居红,王孟宏,等.重庆市4~7岁孤儿口腔健康现状调查分析[J].华西口腔医学杂志,2014,8(4):378-381.
- [4] 陈姬娥,鲍先握.2011—2014年天台县3—6岁儿童龋齿相关因素调查[J].中回乡村医药,2015,22(16):50-52.
- [5] 陈金良,林颖洁,王娜,等.海淀区12岁儿童恒牙患龋与口腔健康状况问卷调查分析[J].河北医药,2017,39(11):1729-1732.
- [6] 宋鑫华,闫鸿浩.基于主成分分析-Logistic模型的边坡稳定性分析[J].科学技术与工程,2017,17(32):224-228.
- [7] 范琳琳,况琴,唐玉娟,等.重庆市渝中区学龄前儿童患龋情况调查分析[J].重庆医学,2015,44(19):2657-2658.
- [8] 王春丽,王涛,夏欣.532名3~6岁幼儿龋齿患病情况调查分析[J].海南医学,2014,9(15):2319-2321.
- [9] 李政,孙磊,王燕玲,等.北京市流动儿童龋齿患病及口腔保健行为分析[J].中国学校卫生,2014,35(5):654-659.
- [10] 徐美玉,赵旭,潘秋华,等.厦门市16所公立幼儿园儿童龋齿流行病学及相关影响因素的研究[J].中国妇幼保健,2017,32(13):3021-3023.
- [11] 张晓敏,朱雪萍.早产低体质量儿长期胃肠外营养相关性胆汁淤积高危因素分析[J].苏州大学学报(医学版),2012,32(1):145-147.
- [12] 苏齐鉴,臧宁,肖信,等.主成分 Logistic 回归模型在消除数据多重共线性中的应用[J].中国卫生统计,2009,26(2):206-208.
- [13] 丁元林,孔丹莉,毛宗福.多重线性回归分析中的常用共线性诊断方法[J].数理医药学杂志,2004,17(4):299-300.
- [14] 周菲,焦桂梅,赵凤兰,等.Logistic回归的主成分改进方法探讨及其医学应用[J].数理医药学杂志,2014,27(1):25-27.
- [15] 蔡永林,郑裕明,汤敏中,等.鼻咽癌非病毒危险因素的主成分改进 Logistic 回归分析[J].现代预防医学,2009,36(13):2416-2419.
- [16] 行岳真,祁瑞,李金梅,等.主成分 Logistic 回归在筛选冠心病危险因素中的应用[J].实用预防医学,2012,19(8):1138-1140.
- [17] 田文玲.“调查研究”的要求[J].中国老区建设,2013,6(5):63.

[收稿日期:2017-10-29]

(本文编辑 杨晨晨)

(上接 638 页)

- [28] 马鸿凯.不同手术方法治疗高血压脑出血患者的效果比较[J].当代医学,2015,21(2):41-42.
- [29] WARLW C P. Epidemiology of stroke[J]. Lancet, 1992, 339(8789):342-344.
- [30] OJEMANN R G, HEROS R C. Spontaneous brain hemorrhage[J]. Stroke, 1983, 14(4):468-475.
- [31] MITCHELL P, GREGSON B A, VINDLACHERUVU R R, et al. Surgical options in ICH including decompressive craniectomy[J]. J Neurol Sci, 2007, 261(1):89-98.
- [32] TAKEUCHI S, TAKASATO Y, MASAOKA H, et al. Decompressive craniectomy with hematoma evacuation for large hemispheric hypertensive intracerebral hemorrhage[J]. Clin

Neurol Neurosurg, 2013, 115(3):317-322.

- [33] ESQUENAZI Y, SAVITZ S I, EL K R, et al. Decompressive hemispheric craniectomy with or without clot evacuation for large spontaneous supratentorial intracerebral hemorrhages [J]. Clin Neurol Neurosurg, 2015, 128(128C):117.
- [34] SHIMAMURA N, MUNAKATA A, NARAOKA M, et al. Decompressive hemi-craniectomy is not necessary to rescue supratentorial hypertensive intracerebral hemorrhage patients: consecutive single-center experience[J]. Acta Neurochirurgica Suppl, 2011, 111(111):415.

[收稿日期:2018-02-16]

(本文编辑 张巧莲)