

基于潜类别回归对特异性过敏患者的分类研究*

王蕊¹, 张茂祥¹, 李萌¹, 刘妍妍¹, 于磊¹, 刘美娜¹

【摘要】 目的 应用潜类别回归分析特异性过敏患者的潜在分类, 为过敏性疾病的早期预防提供依据。**方法** 利用 SAS 软件的 PROC LCA 模块, 对 1124 名特异性过敏患者进行潜类别回归分析研究。**结果** 分析获得特异性过敏患者的 4 个潜在类别及潜在类别的概率。潜在类别命名分别为: 霉菌吸入组, 发生概率是 0.3134; 户外花粉组, 发生概率是 0.3646; 户尘螨组, 发生概率是 0.2799; 综合过敏组, 概率是 0.0422。性别为协变量时, 潜在类别间的差异具有统计学意义 ($P < 0.05$)。**结论** 应用潜类别回归对过敏患者进行分类, 获得的潜类别为过敏原的聚类组合, 过敏原间的关联通过所属潜类别产生, 利用潜类别预防特异性过敏的发生更具实际意义。

【关键词】 特异性过敏 潜类别 潜类别回归分析 协变量

Study on classification of specific allergic patients based on latent class regression analysis

WANG Rui, ZHANG Mao-xiang, LI Meng, LIU Yan-yan, YU Lei, LIU Mei-na
Department of Biostatistics, Harbin Medical University, Harbin, Heilongjiang 150081, China

Corresponding author: LIU Mei-na, E-mail: liumeina369@163.com

Abstract: Objective To analyze the classification of specific allergic patients based on latent class regression analysis for early prevention of allergic diseases.

Methods Using PROC LCA module of SAS software, latent class regression analysis was performed on 1124 specific allergic patients.

*基金项目: 国家重大专项课题 (2011ZX08011-005)

作者单位: 1. 哈尔滨医科大学公共卫生学院卫生统计教研室, 黑龙江 哈尔滨 150086

作者简介: 王蕊 (1988-), 女, 汉族, 黑龙江省齐齐哈尔人, 硕士在读, 研究方向: 过敏反应相关研究 E-mail wangrui0712@126.com tel: 18245110570

通讯作者: 刘美娜, E-mail liumeina369@163.com

Results Four latent class memberships and latent class probabilities were obtained. Latent class memberships were named as mold cockroach inhalation group, outdoor pollen group, dermatophagoides pteronyssinus group and comprehensive allergens group, whose probability of occurring was 0.3134, 0.3646, 0.2799 and 0.0422, respectively. Having sex as covariate, it was found that significant difference existed among four latent class memberships ($P < 0.05$).

Conclusions The classification of specific allergic patients obtained through latent class regression analysis is the combination of allergens, and the association between allergens is produced by the latent class. Using latent class membership to prevent the occurrence of specific allergy is of more practical significance.

Key words: Specific allergy; Latent class membership; Latent class regression analysis; Covariate

随着人类生活的丰富，环境污染日益加重，导致过敏性疾病的发病率呈逐年增长趋势^[1]。特异性过敏是指在过敏患者的皮肤及体液内含有特异性 IgE 抗体，外界存在的过敏原通过不同途径进入机体后与相应的特异性抗体结合，促使机体释放变态反应活性介质产生过敏反应^[2]。特异性过敏的过敏原种类很多，血清检测的常见过敏原有十几种或数十种不等，过敏原间存在复杂的生物学关联即过敏原可同时发生阳性反应^[3]。本研究旨在利用潜类别回归模型对特异性过敏患者的过敏原进行分类，为过敏性疾病的发生和预防提供依据。

1. 对象与方法

1.1 研究对象 2012年4月至2013年11月，选择在哈尔滨医科大学附属第一医院过敏反应科进行血清特异性IgE检测和面对面问卷调查的1124名患者为研究对象。

1.2 方法 采集患者静脉血清标本备检。应用德国MEDIWISS公司生产的“敏筛”定量过敏原体外检测系统，使用AllergyScreen过敏原试剂盒和AllergyScreen过敏原检测仪进行血清特异性IgE检测。试剂盒包括20种过敏原，分别为户尘螨、屋尘、蟑螂、猫毛皮屑、狗毛皮屑、蒿/矮豚草、苋/藜/葎草/蒲公英、桑树、柏榆柳桦栎杨桤榛/白蜡/梧桐、点青/分枝/烟曲/黑曲/交链霉、鸡蛋白、牛奶、牛肉/羊肉、虾、蟹、贝/鱼、腰果/花生/黄豆、小麦、芒果/桃/苹果/樱桃、菠萝过敏原。特异性过敏原检测结果由系统自动生成，抗体浓度由低到高分0-6共7个等级，0等级抗体浓度 $< 0.35 \text{ IU/ml}$ ，判定为阴性反

应；1等级以上抗体浓度 $\geq 0.35\text{IU/ml}$ ，判定为阳性反应。

1.2 统计分析 潜类别回归模型是利用logistic回归模型引入协变量后获得潜在分类的模型。潜类别回归模型的两个基本假设：潜变量为分类变量；局部独立性[4]。

潜类别回归模型最基本公式为：

$$P(Y_i = y / X_i = x) = \sum_{c=1}^{n_c} r_c(x) \prod_{m=1}^M \prod_{k=1}^{K_m} \rho_{mk/c}^{I(y_m=k)}$$

$\rho_{mk/c}$ 表示研究对象在第
c潜类别下，第m外显变

量k水平的条件概率。 $r_c(x)$ 表示x为协变量的模型中，潜在类别c的概率。

基线逻辑回归模型为：

$$r_c(x) = P(L_i = c / X_i = x) = \frac{e^{(\beta_{0c} + x\beta_{1c})}}{\sum_{c=1}^{n_c} e^{(\beta_{0c} + x\beta_{1c})}}$$

潜类别回归模型拟合指标以贝叶斯信息准则（BIC）为标准[5]，利用最大似然估计方法对参数进行估计，迭代的过程则采用EM算法和NR算法。潜类别回归分析利用SAS软件的PROC LCA模块[6]。

2.结果

2.1 调查对象基本情况 1124名特异性过敏患者过敏原反应情况如表1。户尘螨过敏原阳性反应人数最多，阳性构成比为39.77%；其次为蒿/矮豚草过敏原，阳性构成比为39.50%；最少的是小麦过敏原，阳性构成比为6.23%。最常发生阳性反应的9种过敏原分别是户尘螨、蒿/矮豚草、苋/藜/葎草/蒲公英、柏榆柳桦栎杨桤榛/白蜡/梧桐、点青/分枝/烟曲/黑曲/交链霉、蟑螂、芒果/桃/苹果/樱桃、蟹、小麦，91%以上的患者对9种过敏原中至少一种呈阳性反应。

表1 特异性过敏患者过敏原阳性反应情况 (n=1124)

过敏原	阳性例数	阳性构成比 (%)
户尘螨	447	39.77
蒿/矮豚草	444	39.50
苋/藜/葎草/蒲公英	292	25.98
柏榆柳桦栎杨桤榛/白蜡/梧桐	140	12.46

点青/分枝/烟曲/黑曲/交链霉	119	10.59
蟑螂	102	9.07
芒果/桃/苹果/樱桃	88	7.83
蟹	79	7.03
小麦	70	6.23
狗毛皮屑	67	5.96
屋尘	60	5.34
桑树	48	4.27
腰果/花生/黄豆	45	4.00
猫毛皮屑	41	3.65
虾	30	2.67
牛奶	29	2.58
菠萝	28	2.48
牛肉/羊肉	27	2.40
鸡蛋白	20	1.78
贝/鱼	5	0.44
总计*	2181	194.04

*部分患者可同时有两个及两个以上过敏原均为阳性，因此构成比总计大于100%。

2.2 潜类别回归分析 以最常发生阳性反应的9种过敏原作为外显变量，将性别作为协变量进行潜类别回归分析。

2.2.1 模型拟合 模型拟合过程从1个分类的模型开始逐步增加分类数目，利用BIC信息标准进行判断，结果如表2。随着模型分类数目的增加，BIC值先逐渐减小，4分类时降到最小，即最优模型选择4分类的潜类别回归模型。

表2 模型拟合结果

潜在类别数目	BIC	df
1	1809.53	502
2	1021.02	492
3	685.38	482

4	594.13	472
5	597.55	462
6	618.06	452
7	663.27	442
8	714.70	432
9	761.10	422

2.2.2 参数估计 数据分析选择包含4分类的潜类别回归模型，迭代过程采用EM和NR算法，潜类别概率和潜类别下的条件概率的结果如下表3及图1。

表3 潜类别概率和潜类别下条件概率情况

过敏原		潜在类别			
		1	2	3	4
户尘螨	0	0.8698	0.8370	0.0000	0.5969
	1	0.1302	0.1630	1.0000	0.4031
蟑螂	0	0.8409	0.9836	0.9481	0.5147
	1	0.1591	0.0164	0.0519	0.4853
蒿/矮豚草	0	1.0000	0.0311	1.0000	0.0000
	1	0.0000	0.9689	0.0000	1.0000
苋/藜/葎草/蒲公英	0	0.9976	0.4045	1.0000	0.0000
	1	0.0024	0.5955	0.0000	1.0000
柏榆柳桦栎杨桤榛/白蜡/梧桐	0	0.8312	0.9188	1.0000	0.0000
	1	0.1688	0.0812	0.0000	1.0000
点青/分枝/烟曲/黑曲/交链霉	0	0.7581	0.9680	0.9470	0.9342
	1	0.2419	0.0320	0.0530	0.0658
蟹	0	0.8763	0.9606	0.9641	0.8294
	1	0.1237	0.0394	0.0359	0.1706
小麦	0	0.9093	0.9819	1.0000	0.3520
	1	0.0907	0.0181	0.0000	0.6480
芒果/桃/苹果/樱桃	0	0.9153	0.9733	0.9936	0.0441
	1	0.0847	0.0267	0.0064	0.9559

潜类别概率	0.3134	0.3646	0.2799	0.0422
-------	--------	--------	--------	--------

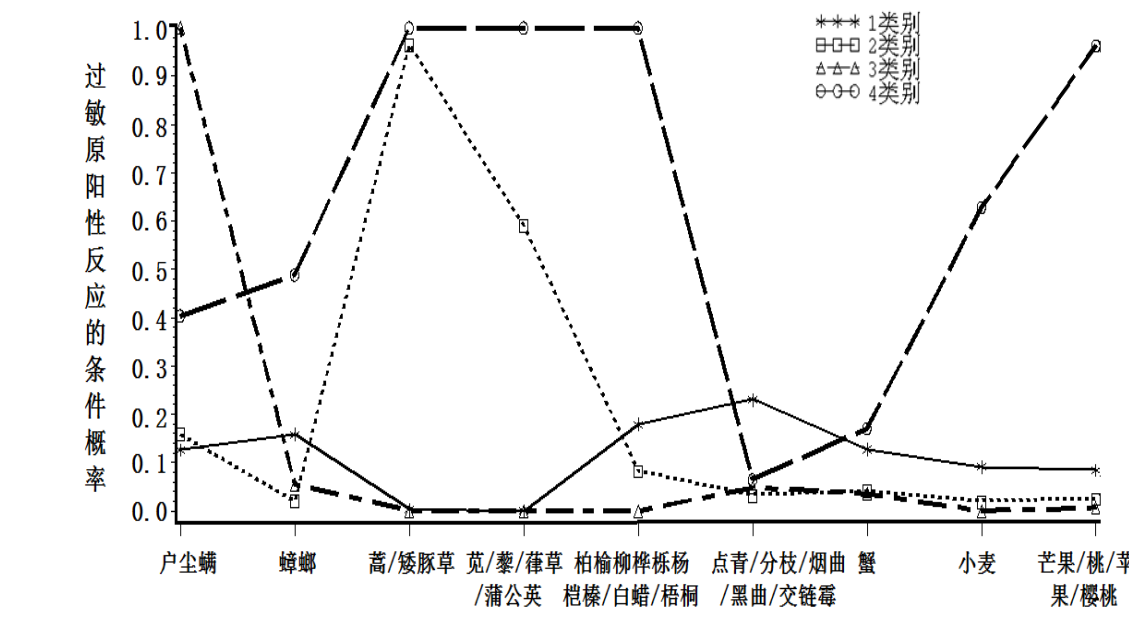


图1 不同潜类别下过敏原阳性反应的条件概率

2.2.3 潜类别的命名 图1所示，第一类别在点青/分枝/烟曲/黑曲/交链霉和蟑螂过敏原处有较高的条件概率，均为吸入性过敏原和共同的繁殖条件，命名为霉菌吸入组；第二类别在蒿/矮豚草、苋/藜/葎草/蒲公英处条件概率明显增高，两者均为春秋季节常见花粉，命名为户外花粉组。第三类别在户尘螨组处条件概率最高，单独作用为主，命名为户尘螨组。第四类别在所有过敏原处条件概率都较高，命名为综合过敏组。

2.2.4 性别协变量的分析 以性别作为协变量，男性作为参照组；以潜在类别作为应变量，霉菌吸入组作为参照组，参照组的选择对结果没有影响，只影响对结果的解释，分析如表4。以霉菌吸入组作为参照组，与男性相比，女性更多分在户外花粉组（OR=0.7163）、户尘螨组（OR=0.8874）和综合过敏组（OR=0.2390）。性别协变量的 $P < 0.05$ ，不同性别的潜在分类结果不完全相同，因此性别对过敏患者的潜在分类存在影响。

表4 性别为协变量的logistic回归分析结果

潜在类别	β	OR
霉菌吸入组	0	1
户外花粉组*	-0.3336	0.7163

户尘螨组*	-0.1194	0.8874
综合过敏组*	-1.4313	0.2390

3.讨论 潜类别回归分析是通过外显变量和潜在变量的关联构建的统计模型，外显变量可以直接通过测量获得，而潜在变量是由模型间接获得，通过获得潜在变量的多个水平对外显特征进行分类的技术。潜类别回归分析中的外显变量和潜在变量都是分类变量。日常生活中接触到的特异性过敏原种类很多，过敏患者往往对多个过敏原同时产生阳性反应，这些过敏原引起过敏可能是单独作用，也可能是联合作用，过敏原间复杂的生物学关联至今并没有明确的解释。本研究通过潜类别回归分析挖掘过敏原的潜在信息，利用过敏原获得潜在分类，类别内的过敏原是相关联的。本研究利用潜类别回归分析获得 4 个潜在分类，按照潜类别概率由大到小分别为户外花粉组、霉菌吸入组、户尘螨组、综合过敏组。户外花粉组的过敏原主要通过花粉形式进入呼吸道，肥大细胞和嗜碱性粒细胞释放致敏活性物质作用于机体产生过敏反应^[7]。户外花粉组为春夏季花粉的组合，该类患者对蒿/矮豚草和苋/藜/葎草/蒲公英过敏原中一种过敏原呈阳性反应，另一过敏原也相对较容易发生过敏，应同时避免另一过敏原的接触。霉菌吸入组的霉菌蟑螂适宜在恒温、密闭、食物丰富和水源充足的条件下繁殖，产生的孢子、代谢物等通过吸入形式进入机体，促使机体产生过敏，应采取加强室内通风、保持环境卫生、减少空调使用等措施阻断霉菌和昆虫类生物的繁殖。该类患者无论是对螨虫或是蟑螂过敏，都应避免该类别内所有过敏原的接触。户尘螨组的螨体分泌物、排泄物都能成为过敏原，主要分布于床铺、被褥、沙发、地毯及尘土里。该类别患者应注重保持环境卫生，采用防螨材料等措施减少过敏原的暴露。综合过敏组为所有过敏原的集合，有两种及以上过敏原呈阳性反应的概率很高，即发生过敏及免疫力低下情况较其他组严重，该类患者除避免接触过敏原外，也要注重增强免疫力。

潜类别回归分析通过引入 Logistic 回归模型，调整协变量对分类结果的影响，能够获得更为准确的分类。不同性别的潜在分类结果存在差异，以往研究认为男性整体的免疫力和健康水平存在差别^[8]，性别是过敏性疾病发生的影响因素，与本研究所得结果一致^[9]。潜类别回归分析将多种过敏原聚类成组合，以过敏原阳性反应组合的形式从环境和免疫力等多方面综合考虑，提出的预防措施更具实际意义。

参考文献

[1] Downs S H, Marks G B, Sporik R, et al. Continued increase in the prevalence of asthma and atopy[J]. Arch Dis Child, 2001, 84(1): 20-23.

- [2] Hans C. Oettgen and Raif S. IgE in asthma and atopy: cellular and molecular connections[J]. *Geha J Clin Invest*, 1999, 104(7):829–835.
- [3] Aalberse RC, Akkerdaas J, Ree RV. Cross - reactivity of IgE antibodies to allergens[J]. *Allergy*, 2001, 56(6): 478-490.
- [4] 邱皓政. 潜在类别模型的原理与技术[M]. 北京:教育科学出版社, 2008.
- [5] Lin TH, Dayton CM. Model selection information criteria for non-nested latent class models[J]. *J Educ Behav Stat*, 1997, 22(3): 249-264.
- [6] Lanza ST, Collins LM, Lemmon DR, et al. PROC LCA: A SAS procedure for latent class analysis[J]. *Structural Equation Modeling*, 2007, 14(4): 671-694.
- [7] Johansson SG, Hourihane JO, Bousquet J, et al. A revised nomenclature for allergy: an EAACI position statement from the EAACI nomenclature task force[J]. *Allergy*, 2001, 56(9): 813-824.
- [8] Klein SL, Roberts CW. Sex Hormones and Immunity to Infection[M]. Springer Berlin Heidelberg, 2010.
- [9] Chen W, Mempel M, Schober W, et al. Gender difference, sex hormones, and immediate type hypersensitivity reactions[J]. *Allergy*, 2008, 63(11):1418–1427.