

大学生 MSM 人群 HIV 感染风险预测模型建立及效果验证研究

黄掣驰^{1,2,3,4}, 呼志丽^{1,2,3}, 胡清海^{1,2,3}, 楚振兴^{1,2,3}, 张晶^{1,2,3}, 徐俊杰^{1,2,3}

1. 中国医科大学附属第一医院检验科, 国家卫生健康委员会艾滋病免疫学重点实验室(中国医科大学), 辽宁 沈阳 110001;
2. 中国医科大学附属第一医院, 辽宁省艾滋病免疫学重点实验室, 辽宁 沈阳 110001;
3. 中国医学科学院艾滋病免疫学重点实验室, 辽宁 沈阳 110001; 4. 辽宁省信息中心, 辽宁 沈阳 110001

摘要: **目的** 建立并验证大学生男男性行为者(men who have sex with men, MSM)的 HIV 感染风险预测模型。 **方法** 利用我国 7 大城市(上海、南京、长沙、郑州、济南、沈阳和昆明)MSM 横断面调查数据库中在校大学生 MSM 相关信息, 通过单因素和多因素 logistic 回归建立模型并采用 Stepwise 逐步回归法筛选模型变量; 采用 C-statistics 检验、Hosmer-Lemeshow 检验($H-L$ 检验)对模型进行判别能力和拟合效果的内部验证和评价; 采用 Bootstrap 重抽样技术进行模型内部验证。

结果 共纳入 742 名大学生 MSM 参加研究, 其 HIV 感染率为 5.9% (44/742)。经 logistic 回归模型拟合并通过 stepwise 法筛选, 以下变量进入最终模型: 有性传播疾病感染史 ($OR=5.4, 95\%CI: 2.5 \sim 11.4, P<0.01$) (5 分); 近半年使用过新型毒品 ($OR=2.2, 95\%CI: 1.2 \sim 4.2, P=0.016$) (2 分); 最近一次男男性行为方式为被插入方或同时为插入方和被插入方 ($OR=2.3, 95\%CI: 1.0 \sim 5.0, P=0.044$) (2 分)。模型的 C-statistic 检验的统计量 AUC 为 0.71, $H-L$ 检验的卡方值为 1.35 ($P=0.717$); 各变量在 1 000 个 Bootstrap 样本中出现的频率均>50%。 **结论** 大学生 MSM 人群 HIV 感染风险预测模型具有较好的判别能力和拟合效果, 有助于大学生 MSM 人群比较客观地评估其自身 HIV 感染风险, 以及促进其参与后续 HIV 检测等防控活动。

关键词: 男男性行为者; 人类免疫缺陷病毒; 大学生; 风险预测模型; Bootstrap

中图分类号: R512.91 文献标识码: A 文章编号: 1006-3110(2020)02-0156-05 DOI: 10.3969/j.issn.1006-3110.2020.02.008

Establishment of a predictive model for HIV infection risk among college student MSM and validation of its performance

HUANG Che-chi^{1,2,3,4}, HU Zhi-li^{1,2,3}, HU Qing-hai^{1,2,3}, CHU Zhen-xing^{1,2,3}, ZHANG Jing^{1,2,3}, XU Jun-jie^{1,2,3}

1. NHC Key Laboratory of AIDS Immunology (China Medical University), Department of Laboratory Medicine, the First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110001, China;
 2. Key Laboratory of AIDS Immunology of Liaoning Province, the First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110001, China;
 3. Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, Liaoning 110001, China;
 4. Information Center of Liaoning Province, Shenyang, Liaoning 110001, China
- Corresponding author: XU Jun-jie, E-mail: xjjcmu@163.com

Abstract: **Objective** To develop and validate a predictive model for HIV infection risk in college student MSM. **Methods**

The information related to college student MSM was collected from the database of MSM cross-sectional survey in 7 cities (including Shanghai, Nanjing, Changsha, Zhengzhou, Jinan, Shenyang and Kunming). Univariate and multivariate logistic regression was used to develop the model, and a stepwise procedure was used to select variables. C-statistics test was employed to assess the model discrimination, and Hosmer - Lemeshow test was used to estimate the model calibration. Bootstrap resampling technology was used for model validation. **Results** A total of 742 college student MSM were enrolled in this study, and the prevalence rate of HIV was 5.9% (44/742). According to logistic regression and stepwise procedure, the variables and their scores in the pred-

基金项目: 中国科技部十三五重大专项课题(2017ZX10201-101-002-007); 中国科技部“一带一路”传染病防控保障关键技术研发(2018ZX10101-001-001-003); 国家自然科学基金面上项目(81872674); 国家自然科学基金青年科学基金项目(81703277)。

作者简介: 黄掣驰(1991-), 女, 辽宁省本溪市人, 硕士, 研究方向: 艾滋病流行病学; 呼志丽(1994-), 女, 山西吕梁人, 硕士在读, 研究方向: 艾滋病流行病学。黄掣驰与呼志丽并列为第一作者。

通信作者: 徐俊杰, E-mail: xjjcmu@163.com。

ictive model for HIV infection risk among college student MSM were having a history of infection with sexual transmitted diseases ($OR=5.4$, $95\%CI:2.5-11.4$, $P<0.01$) (5 points), new-type drug abuse in the past six months ($OR=2.2$, $95\%CI:1.2-4.2$, $P=0.016$) (2 points), and practicing receptive or both insertive and receptive anal intercourse in the most recent homosexual behavior ($OR=2.3$, $95\%CI:1.0-5.0$, $P=0.044$) (2 points). The area under the curve (AUC) of C-statistics was 0.71, and the Chi-square value of Hosmer-Lemeshow test was 1.35 ($P=0.717$). The frequency of each variable in the final model appeared in 1,000 Bootstrap samples were all greater than 50%. **Conclusions** The said predictive model for HIV infection risk among college student MSM has better judgment ability and fitting effect. It is conducive to objectively self-evaluating the risk for HIV infection among college student MSM and promoting them to participate in HIV detection and other prevention and control activities.

Key words: men who have sex with men; human immunodeficiency virus; college students; risk prediction model; Bootstrap

大学生男男性行为者 (men who have sex with men, MSM) 是一个特殊群体,其 HIV 感染情况近年来受到了广泛关注。据估计,我国 18~22 岁大学生 MSM 约有 106 万人^[1-2]。我国在校大学生 MSM 的 HIV 感染率 2003—2006 年为 3.0%,而 2009—2010 年则上升至 6.8%^[3];2016 年平均每天就有 8 个大学生 MSM 感染 HIV^[4]。有研究表明大学生 MSM 近 6 个月发生无保护性行为的比例为 65.2%^[5],高于非大学生 MSM 人群 (45.7%)^[6]。然而该人群 HIV 检测率较低,曾经接受过 HIV 检测咨询的大学生 MSM 比例仅为 1/3 (33.5%)^[7],低于全国普通 MSM 人群的 HIV 检测率 (46.6%)^[8]。同时,接近半数的大学生 MSM 自认为他们感染 HIV 的风险很低或没有感染风险^[9],表明该人群并不能很好地感知自己的 HIV 感染风险。因此,亟待获得一种准确、方便的方式促使大学生 MSM 准确意识到其 HIV 感染的实际风险,并有望借此促进其 HIV 检测行为并降低 HIV 相关高危行为的发生率。疾病风险预测模型通常被用来预测某一个体某种疾病发病或感染状况,目前在预测 HIV 感染及新发感染方面,有 Menza^[10]、HIRI-MSM^[11]、San Diego Early Test (SDET)^[12] 以及广州市普通 MSM 的 HIV 感染风险估计模型^[13]。然而,风险预测模型仅适用于其建模人群^[14],既往国内外尚无针对大学生 MSM 建立的 HIV 感染风险预测模型。因此,本研究拟采用前期对我国 7 个大城市 MSM 人群的大规模流行病学调查中大学生 MSM 亚人群的数据进行拟合分析,旨在建立我国大学生 MSM 人群的 HIV 感染预测模型并对其进行内部验证。

1 对象与方法

1.1 对象 本研究为针对我国上海、南京、长沙、郑州、济南、沈阳和昆明 7 大城市 MSM 人群的现场流行病学调查数据的深入挖掘分析,调查对象招募方式为通过互联网在线招募、基于 MSM 活动场所人群抽样及同伴推动抽样。调查时间集中于 2012 年 6 月—2014

年 12 月。研究对象的纳入标准:1) 自报近 6 个月发生过男男性行为;2) 年龄 ≥ 18 周岁;3) 自报当前的职业为在读大学生;4) 自愿签署知情同意书并能提供血样。前期 MSM 研究对象的招募和入组标准等见发表文献^[15]。

1.2 方法

1.2.1 模型建立及评价 合并数据库中包括的信息有大学生 MSM 的社会人口学信息、高危性行为特征、新型毒品使用情况:本研究中为自报近 6 个月 Rush Poppers (一种可吸入性亚硝酸盐)、冰毒、摇头丸、K 粉 (学名:氯胺酮) 等使用情况、HIV 及其他性传播疾病的感染情况。实验室检测项目包含 HIV 初筛及确认、梅毒螺旋体初筛及确认以及生殖器疱疹病毒 (herpes simplex virus-2, HSV-2) 抗体检测。本研究采用 step-wise 逐步法 (进入水平 $\alpha=0.15$,剔除水平 $\alpha=0.10$) 筛选自变量并以是否发生 HIV 感染为因变量构建 logistic 预测模型。采用 C-statistics 检验进行模型的分度评价;采用 Hosmer-Lemeshow test (H-L test) 对模型进行校准度评价。

1.2.2 模型评分系统的构建及评价 模型中每个变量的评分为其 OR 值的整数部分 (采用四舍五入法保留整数部分),每个个体的评分为模型中各变量评分的总和。利用 AUC (受试者工作曲线下面积) 来判断模型评分系统判别能力的大小。计算评分系统在各切点 (Cut-off) 的敏感度、特异度、阳性预测值及阴性预测值,利用约登指数选择最佳临界点;并通过预测比值比 (diagnostic odds ratio, DOR) 判断评分系统中不同 Cut-off 分值与 HIV 感染的关联程度。

1.2.3 模型内部验证 利用 Bootstrap 重抽样技术对建模数据生成 1 000 个等样本大小的伪样本,若原模型变量出现的频率 $>50\%$,那么这些变量具备可信性并可以继续保留在原模型中^[16]。

1.2.4 实验室检测 收集每一个大学生 MSM 的静脉血 10 ml,用于检测 HIV、梅毒以及 HSV-2。HIV 的初筛和确认按照《全国艾滋病检测技术规范》执行^[17];梅

毒螺旋体的初筛通过快速血浆反应素环片进行检测,对于初筛检测为阳性的样本,再使用梅毒螺旋体明胶凝集试剂来确认,两种试验同时阳性即梅毒现症感染;HSV-2 抗体检测采用 HerpeSelect® 1&2 ELISA IgG 试剂盒测定。

1.3 统计分析 使用卡方检验对比分析感染和未感染 HIV 的大学生 MSM 的社会人口学特征、高危性行为、新型毒品使用、性传播疾病感染情况的分布差异。利用单因素及多因素 logistic 回归建立模型,采用 step-wise 法选择变量。模型和模型评分系统的评价方法分别为 C-statistics、H-L test 和 ROC 曲线下面积 AUC。使用 Mann-Whitney U 检验分析感染与未感染 HIV 的大学生 MSM 在模型评分系统中得分的分布差异。使用 Spearman 秩相关分析评分系统得分与 HIV 感染率是否存在关联性。使用的统计软件为 SAS 9.2、SPSS 22.0,检验显著性水准为双侧 $\alpha=0.05$ 。

2 结果

本研究共有 742 名大学生 MSM 纳入研究,其 HIV 平均感染率为 5.9%(44/742)。

2.1 社会人口学特征 研究对象民族以汉族为主,占 92.2%(684/742);婚姻状况以未婚为主,占 93.3%(692/742);月收入以无收入为主,占 88.4%(656/742);性取向以自报是同性恋(一个人在性爱、心理、情感等方面的兴趣主要对象均为同性别的人)为主,占 69.5%(516/742)。

2.2 模型包含的变量及评分情况 经分析,以下变量进入最终模型:性传播疾病感染 ($OR=5.4, 95\%CI: 2.5\sim 11.4, P<0.01$);近半年使用过新型毒品 ($OR=2.2, 95\%CI: 1.2\sim 4.2, P=0.016$);最近一次男男性行为方式为被插入方或同时为插入方和被插入方 ($OR=2.3, 95\%CI: 1.0\sim 5.0, P=0.044$)。对模型中囊括的每个变量 OR 值通过四舍五入计算后保留其整数部分,以上各变量的分值分别为:5 分、2 分、2 分。见表 1。

表 1 大学生 MSM 的 HIV 感染风险预测模型变量

变量	单因素 logistic 回归			多因素 logistic 回归				得分
	OR	95%CI	P	OR	95%CI	P	B	
性传播疾病感染	5.9	2.8~12.2	<0.01	5.4	2.5~11.4	<0.01	1.68	5
近半年使用过新型毒品	2.4	1.3~4.5	<0.01	2.2	1.2~4.2	0.016	0.78	2
最近一次性角色为被插入方或同时为插入方	2.3	1.1~5.1	0.034	2.3	1.0~5.0	0.044	0.82	2
提供过商业性服务	4.2	1.1~15.4	0.032					

即本研究建立的我国大学生 MSM 人群的 HIV 感染风险预测模型如下:

$$Logit(\pi)=1.68X_1+0.78X_2+0.82X_3-3.87$$

其中: π 为每个大学生 MSM 个体感染 HIV 的概率预测值; X_1 为性传播疾病感染情况,其赋值情况为:1 为实验室诊断为性传播疾病(即感染了梅毒或 HSV-2 中任何一个即定义为性传播疾病感染),0 为未感染任何性传播疾病; X_2 为近半年使用新型毒品情况,其赋值情况为:1 为使用,0 为未使用; X_3 为最近一次性角色:1 为被插入方或同时为被插入方和插入方,0 为插入方。

2.3 模型及评分系统的评价 模型的校准度评价,即 H-L 检验的卡方值为 1.35($P=0.717$);模型的区分度评价,即该模型 C-statistic 检验的统计量 AUC 为 0.71。评分系统的 AUC 为 0.71(95%CI: 0.63~0.80)。在本研究建立的大学生 MSM 人群的 HIV 感染风险预测模型中,得分情况及相应 HIV 感染率,见表 2。HIV 感染率与模型得分呈正相关($r_s=1, P<0.01$)。感染 HIV 的大学生 MSM 在模型评分系统中得分中位数为 4 分($IQR, 2\sim 5$),而未感染 HIV 的研究对象中得分中位数为 2 分($IQR, 2\sim 2$),两者得分差异存在统计学意义($z=-5.2, P<0.01$)。最佳 Cut-off 分值为 ≥ 4 分,即约登指数最大时其敏感度为 56.82%,特异度为 79.94%,阳性预测值为 15.15%,阴性预测值为 96.71%。感染风险评分 ≥ 4 分的大学生 MSM 的 HIV 感染率为 15.2%,是 0~2 分 HIV 感染率(3.3%)的 4.6 倍。具体不同分值的 HIV 感染情况及各 Cut-off 分值处的评价情况,见表 2 和表 3。

表 2 大学生 MSM 的 HIV 感染风险预测模型评分系统不同得分的 HIV 感染情况

分数	总人数(%)	HIV 感染人数(%)	感染率(%)
0	173(23.3)	4(9.1)	2.31
2	404(54.5)	15(34.1)	3.71
4	111(15.0)	13(29.6)	11.71
5	12(1.6)	2(4.5)	16.67
7	27(3.6)	6(13.6)	22.22
9	15(2.0)	4(9.1)	26.67
合计	742(100.0)	44(100.0)	5.93

表 3 大学生 MSM 的 HIV 感染风险预测模型评分系统各 Cut-off 分数的评估情况

Cut-off 分数	敏感度(%)	特异度(%)	阳性预测值(%)	阴性预测值(%)	DOR*
≥ 2	90.91	24.21	7.03	97.69	3.19
≥ 4	56.82	79.94	15.15	96.71	5.24
≥ 5	27.27	93.98	22.22	95.35	5.86
≥ 7	22.73	95.42	23.81	95.14	6.12
≥ 9	9.09	98.42	26.67	94.50	6.25

注:*DOR 即预测比值比(diagnostic odds ratio, DOR),是阳性似然比与阴性似然比的比值。反映预测试验的结果与疾病的联系程度。取值>1 时,其值越大说明该预测试验的判别效果较好。

2.4 模型内部验证 进入最终模型的 3 个变量在

1 000 个 Bootstrap 样本中出现的频率分别为 94.3%、69.3%、51.3%，均大于 50%。表明通过模型的内部验证，这 3 个变量均有较好的可信性。

3 讨论

我国已经将年轻 MSM 作为防控 HIV 的重点干预人群^[18]，而大学生 MSM 是该年龄段人群中的一个重要组成部分，而现有的预测模型并非针对大学生这一特殊群体建立，本研究通过对我国 7 大城市的大学生 MSM 数据进行拟合分析，针对大学生 MSM 这一人群建立 HIV 感染风险预测模型，并利用 Bootstrap 重抽样技术对该模型进行了内部验证。本预测模型共包含 3 个变量，分别为：性传播疾病、近半年新型毒品使用情况、最近一次性角色。这几个变量在之前的研究中均被证实和 HIV 感染有关^[19-23]。

本研究中，性传播性疾病 ($OR = 5.4$) 是预测大学生 MSM 感染 HIV 权重最大的变量，但在 Menza^[10]、SDET^[12] 这两个模型中该变量对于 HIV 感染的预测能力均排在第二位。在本研究中，所有性传播疾病的判定是以实验室检测结果为准，而非自报，结果较客观。提示卫生和教育部门应尽早对在校大学生 MSM 通过学校或同伴教育传输有效的生殖健康知识，提高其对 HIV 及其他性传播疾病的认识，同时也可以采取在校内提供免费安全套等措施，以降低大学生 MSM 人群的 HIV 感染风险。最近一次男男性行为时性角色首次出现在 HIV 感染风险预测模型中，有研究表明对于 MSM，被插入方感染 HIV 的风险是插入方的 12.5 倍^[24]。因此，该变量对大学生 MSM 的 HIV 感染预测能力不容忽视。提示应优先对性角色为被插入方的大学生 MSM 开展干预，鼓励其在进行性行为时能够使用安全套。近半年新型毒品使用情况 ($OR = 2.2$, $95\%CI: 1.2 \sim 4.2$, $P = 0.016$) 为另一个关键预测变量，该变量同时也出现在 Menza 模型^[10]、HIRI-MSM 模型^[11] 中。在本研究中，新型毒品的使用率较高 (25.2%)，我国 MSM 最广泛使用的新型毒品为 Rush^[23]，然而目前 Rush 在我国未被禁用，MSM 可以轻易通过互联网或其他途径购买或获得，提示我国相关部门需考虑制定一些必要措施和方案禁止 Rush 流入非医疗市场。

本研究建立的大学生 MSM 的 HIV 感染风险预测模型 AUC 与其他 HIV 风险预测模型相似^[10-12]，均在 0.67~0.74 的范围内。模型评分系统中最佳 Cut-off 值为 4 分，在此处敏感度为 56.82%，略低于 HIRI-MSM^[11] 和 SDET^[12] 评分系统。此外，Menza^[10] 模型的

研究者认为如果只干预最佳 Cut-off 得分的 MSM，会遗漏一些评分较低的 HIV 感染者，因此有必要对感染风险评分大于零的大学生 MSM 进行干预。对于本模型来说，不仅要积极干预最佳 Cut-off 分值 (≥ 4 分) 的大学生 MSM，同时不能忽略得分为 2 分的大学生 MSM，也应该对其进行干预，促进其进行 HIV 检测。

本研究建立的预测模型安全套使用情况这一变量未进入模型，但此与以往横断面研究有所不同^[25]，这可能是因为本研究采用面对面访谈的方式进行调查，对于敏感性问题的调查可能存在社会期望偏倚，即部分调查对象有意隐瞒其真实安全套使用情况；其次，本模型样本量 742 人，建模数据规模为中等，且其中发生感染 HIV 者为 44 人，相对较少，可能导致检验效能不足，致使有关安全套使用这一变量未进入模型。

本研究模型是首个针对 18~22 岁大学生 MSM 群体建立的 HIV 感染风险评估模型，且数据来自全国多中心，建立的模型代表性相对优于单中心研究数据；大学生 MSM 可利用其客观地自我评估 HIV 感染风险，有利于促使其接受 HIV 检测和提高 HIV 检测频率；另外，本研究也存在一定限制性，样本量有限，用于建模的数据仅为中等规模，且建模数据中感染者也相对较少，这可能会对模型的稳健性造成一定影响，今后仍需收集更大规模的大学生 MSM 相关信息以进一步测试和完善风险评估模型。

参考文献

- [1] 郭静, 黄晓娟, 王秀彬, 等. 利用网络规模迭代法估计北京市大学生男男性行为人群规模[J]. 中华流行病学杂志, 2013, 34(11): 1080-1082.
- [2] 谢沂楠. 各级各类学历教育学生情况 [DB/OL]. (2017-08-24) [2018-01-20]. http://www.moe.gov.cn/s78/A03/moe_560/jytsj_2016/2016_qg/201708/t20170823_311668.html.
- [3] Li Y, Xu J, Reilly KH, et al. Prevalence of HIV and syphilis infection among high school and college student MSM in China: a systematic review and meta-analysis[J]. PLoS One, 2013, 8(7): e69137.
- [4] 尚红. 明确高校艾防主体责任 [DB/OL]. (2017-03-13) [2018-01-15]. http://www.sohu.com/a/128715460_464387.
- [5] Yang Z, Huang Z, Dong Z, et al. Prevalence of high-risky behaviors in transmission of HIV among high school and college student MSM in China: a meta-analysis[J]. BMC Public Health, 2015, 15(1): 1272.
- [6] Shang H, Zhang LQ. MSM and HIV-1 infection in China[J]. Natl Sci Rev, 2015, 2(4): 388-391.
- [7] Xu JJ, Reilly KH, Lu CM, et al. A cross-sectional study of HIV and syphilis infections among male students who have sex with men (MSM) in northeast China: implications for implementing HIV screening and intervention programs[J]. BMC Public Health, 2011, 11(1): 287.
- [8] 李东民, 葛琳, 王岚, 等. 中国 2010—2013 年男男性行为人群艾滋病及相关行为变化趋势分析[J]. 中华流行病学杂志, 2014, 35(5): 542-546.