

# 利用网络规模迭加法估计北京市大学生男男性行为人群规模

郭静 黄晓娟 王秀彬 翁昊艺 郭巍

**【摘要】 目的** 探讨网络规模迭加法估计北京在校大学生社交网络规模和男男性行为人群规模。**方法** 利用网络规模迭加法估计北京市大学生的社交网络规模,并通过反向预测法和态度倾向校正方法分别对社交网络规模C值和目标人群规模进行调整。**结果** 北京市大学生平均社交网络规模为97人,不同专业类型、年级的大学生社交网络规模存在差异;应用网络规模迭加法估计的北京市大学生男男性行为人群规模为16 260人,占在校大学生的3.96%。**结论** 运用网络规模迭加法计算社交网络规模和敏感人群规模具有一定的可信性。

**【关键词】** 男男性行为人群;网络规模迭加法;社交网络规模;大学生

**Estimation on the size of men who have sex with men among college students in Beijing through the Network Scale-Up Method (NSUM)** GUO Jing<sup>1</sup>, HUANG Xiao-juan<sup>1</sup>, WANG Xiu-bin<sup>1</sup>, WENG Hao-yi<sup>1</sup>, GUO Wei<sup>2</sup>. 1 School of Sociology and Population Study, Renmin University of China, Beijing 100872, China; 2 National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention

Corresponding author: GUO Wei, Email: guowei@chinaaids.cn

This work was supported by grants from the Young Scholar Scientific Research Foundation of China CDC (No. 2011A104) and the Fundamental Research Funds for the Central Universities (the Research Funds of Renmin University of China) (No. 13XNC003).

**【Abstract】 Objective** Network Scale-Up Method (NSUM) was used to estimate the Personal Network Size (C) of college students in Beijing and the size of men who have sex with men (MSM). **Methods** A survey was conducted on Beijing college students to estimate the personal network size (C) through NSUM and then using the back-estimating and attitude adjustment to correct C value and the size of the target population. **Results** The mean of C among Beijing college students was 97 and the differences of C between different majors and different grades were both statistically significant. Among Beijing College students, 3.96% of them identified themselves as MSM, with the number as 16 260. **Conclusion** Using the NSUM to estimate C of Beijing college students and size of MSM seemed to be credible which worth popularized.

**【Key words】** Men who have sex with men; Network scale-up method; Social network size; College students

男男性行为人群(MSM)是艾滋病的高危敏感人群。我国哨点监测发现<sup>[1]</sup>,青年学生感染人数上升较快且疫情严重,仅2012年1—10月报告15~24岁青少年HIV感染9514例,较2011年同期增加12.8%。我国一直在加强学生人群艾滋病哨点监测,但仍难以准确估计在校大学生中MSM的规模。传统用于高危

敏感人群规模估计的方法主要包括乘数法、除数法、捕获再捕获法、普查法等<sup>[2-5]</sup>,这些方法一般需要与目标人群直接接触开展独立调查,成本高,难以大规模的推广。本研究旨在运用网络规模迭加法(Network scale-up method)<sup>[6-14]</sup>,在大学生中开展一般人群调查,估计大学生的社会网络规模,进而利用大学生认识“男男性行为”人数,间接估计该人群的规模。

DOI: 10.3760/cma.j.issn.0254-6450.2013.011.008

基金项目:中国疾病预防控制中心2011年青年科研基金(2011A104);中国人民大学科学研究基金(中央高校基本科研业务费专项资金)(13XNC003)

作者单位:100872 北京,中国人民大学社会与人口学院(郭静、黄晓娟、王秀彬、翁昊艺);中国疾病预防控制中心性病艾滋病预防控制中心(郭巍)

通信作者:郭巍, Email: guowei@chinaaids.cn

## 对象与方法

1. 研究对象:从北京市88所高校随机抽取5所,每所高校再随机抽取2个系,每个系二年级及以上随机抽取2个班级(班容量平均约30人),抽中班级中的所有学生均作为调查对象,共2005人。纳入本

研究的调查对象共 1776 人。

2. 概念定义:①已知人群:即已知其规模大小的人群。本研究选取在校大学生中的少数民族、港澳台地区、山东省籍、河南省籍、国外留学生、中共党员和过去 1 年发生意外事故(车祸、溺水等)者,共 7 个已知人群。已知人群规模数据来自北京市教委。②认识的人:为北京市在校大学生,看到对方或对方的名字或绰号时能认出对方,最近 2 年内与其有过接触(包括见面、聚会、电话网络等方式联系既往已经见过面的人,不包括素未谋面的网友)。

3. 分析方法:网络规模迭加法估计人群规模步骤可分为两步:第一步是通过调查数据估计社交网络规模 C 值;第二步是通过已得到的 C 值估算目标(未知)人群规模。

(1)社交网络规模 C 值的估计:运用式(1),通过询问被调查者所认识的属于若干个已知人群的人数,以极大似然法估计被调查者的社交网络规模。

$$\hat{c}_i = \frac{\sum_{j=1}^L m_{ij}}{\sum_{j=1}^L e_j} \cdot N \quad (1)$$

式中  $\hat{c}_i$  是第  $i$  个调查对象社交网络规模的大小,  $m_{ij}$  是第  $i$  个调查对象认识的属于已知人群  $j$  的人数,  $e_j$  是已知人群  $j$  的规模, 已知人群规模的数量有  $L$  个,  $N$  是总人群的规模。实际调查中  $m_{ij}$  需要通过收集数据得到,  $e_j$  和  $N$  则是人口统计的数据。

同理,将  $c$  代入公式可反向计算出未知人群的规模  $e$

$$\hat{e}_j = \frac{\sum_{i=1}^N m_{ij}}{\sum_{i=1}^N c_i} \cdot N \quad (2)$$

(2)社交网络规模 C 值的校正:首先确定已知人群,即采用反向预测法对  $c$  进行校正,依次从已知人群中剔除一个已知人群,利用其他已知人群估计得到 C 值,反推被剔除的已知人群的规模。保留反向预测值与实际值的比值在 0.2~6.0 之间的已知人群作为计算 C 值的已知人群;其次,删除异常值,即根据已知人群调整后 C 值的分布,绘制其频数分布图,剔除异常观察对象;最后进行加权校正,即利用大学生人口学特征(年级和专业类型)对 C 值进行校正。

(3)未知人群规模校正:调查对象对问题的接受程度将对其认识的人群规模造成影响,以不同认可度水平调查对象社交网络规模平均值,除以中等认可度水平社交网络规模平均值,所得到的权重对数据进行校正。权重校正系数公式

$$W_{jk} = \frac{\bar{m}_{k-1-5}}{\bar{m}_{k-3}} \quad (3)$$

式中  $W_{jk}$  是所有被调查者对事件人群  $j$  的认可尊重态度,为  $k(k=1,2,3,4,5)$  的权数校正系数; $\bar{m}_{k=1-5}$  是指所有调查者在第  $k$  个认可水平上认识的事件人群  $j$  人数的平均数; $\bar{m}_{k=3}$  是指所有被调查者在  $k=3$  的认可水平上认识事件人群  $j$  人数的平均数。

4. 统计学分析:利用 EpiData 建立数据库,采用 SAS 统计软件进行分析,统计描述主要采用  $\bar{x} \pm s$ , 不同特征大学生社交网络规模比较采用秩和检验。

### 结 果

1. 基本情况:1761 人中,男生 764 人(43.38%),女生 997 人(56.62%);大学二年级(大二)学生 668 人(37.93%),大三学生 642 人(36.46%),大四学生 451 人(25.61%);医学和理工类专业学生分别占总人数的 21.35% 和 21.24%,艺术类专业 323 人(18.34%),文科类 338 人(19.19%),体育类 350 人(19.88%)。

2. 社交网络规模 C 值的估计和校正:利用式(1)对社交网络规模 C 值进行初步估计,结果显示 C 均值为 98 人,标准差为 109 人,中位数为 57 人。

利用反向预测法进一步剔除对 C 值估计结果造成偏差的已知人群,提高了 C 值估计结果的稳定性和可靠性。不同特征大学生的社交网络规模不完全相同(表 1),根据 6 个已知人群,利用大学生人口学特征校正后的社交网络规模 C 均值为 97 人,标准差为 109 人,中位数为 55 人。其频数分布呈明显的右偏态分布(图 1),对 C 值进行对数转换后  $\lg(C)$  近似呈正态分布(图 2)。

表 1 不同特征大学生社交网络规模估计

变 量	样 本 量	均 数	中 位 数	统 计 量	P 值	
性别	男	764	95	53	0.09	0.76
	女	997	99	57		
年级	大二	668	92	55	30.16	<0.01
	大三	642	113	69		
	大四	451	82	49		
	专业	文科	338	117		
理科	374	81	49			
体育	350	100	55			
医学	376	47	30			
艺术	323	151	101			
总样本	1761	97	55			

3. 目标人群规模的估计与校正:基于计算出的社交网络规模 C,代入式(2)计算北京市在校大学生 MSM 的规模为 15 960 人,占男大学生的比例为 3.89%。本研究中对“有男男性行为”目标人群存在 5 种态度:非常讨厌、比较讨厌、像一般同学对待、比较喜欢、非常喜欢,采用式(3)对该人群规模和比

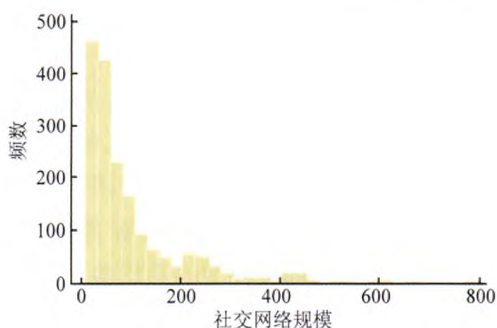


图1 大学生社交网络规模分布

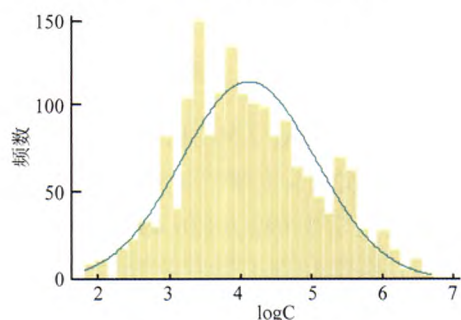


图2 lg(C)的频数分布

例进行校正, 5种态度的校正系数值分别为 0.56、0.52、1.00、2.73、2.04, 基于校正系数代入式(2)重新估计北京市在校大学生 MSM 规模和比例, 结果 MSM 规模为 16 260 人, 占男大学生的比例为 3.96%。

### 讨论

本研究表明, 北京在校大学生社交网络规模 C 值的分布呈右偏态分布, 与国内外相关研究结果一致<sup>[10-14]</sup>。不同特征大学生社交网络规模存在差异, 其分布差异可从其身份特征得到合理解释。本研究医学和理科类专业生比例较高, 而其社交网络规模 C 值较低, 为了排除样本构成对结果的影响, 采用北京市大学生人口学特征进行校正, 结果更为准确和可信。其中在校大学生中 MSM 规模与 2011 年北京市海淀区哨点监测结果接近(3.90%)<sup>[1]</sup>。

本研究引入态度校正系数方法对目标人群规模进行校正, 但不同的态度校正法对目标人群规模估计的影响尚未得到有效验证。因此, 在今后的实践中应继续寻求更为理想的数据校正方法, 使估计的结果更为准确、可信。

已知人群的选择将对网络规模迭加法的运用产生极大影响, 继而影响社交网络规模和未知人群规模的计算。所以在网络规模迭加法的运用中, 要慎重选择已知人群, 对社会交往意愿影响较小的已知人群适宜采用于网络规模的估计。另外, 至少需要 6 个已知人群, 同时不能选择规模太大或太小的人群。

网络规模迭加法是一种以一般人群为调查对象可对敏感人群规模估计的方法, 因不涉及个人隐私而相对更易获得数据, 特别适合对某敏感人群了解较少而不易直接接触情况下开展规模估计, 且可一次对多个人群进行估计, 具有费用低且不易被察觉的优点。McCarty 等发现, 网络规模迭加法存在高估小规模人群 (<200 万) 和低估大规模人群 (>300 万) 的缺陷, 包括屏障效应、传播效应、回忆偏倚和涟漪效应因素影响规模估计的准确性<sup>[6, 14]</sup>。因此, 如何解决各种偏倚效应对调查结果造成的影响还有待于进一步探索。

### 参考文献

- [1] Chinese Ministry of Health, UNAIDS, WHO. 2011 update on the HIV/AIDS epidemic in China. Beijing: Chinese Center for Disease Control and Prevention, 2011. (in Chinese)  
中国卫生部, 联合国艾滋病规划署, 世界卫生组织. 2011 年中国艾滋病疫情估计. 北京: 中国疾病预防控制中心, 2011.
- [2] Liu LR, Liu M. Development of methods for estimating the size of sub-population at high risk for HIV/AIDS. *Epidemiol Lemol Fore Med Sci*, 2005, 32(6): 341-343. (in Chinese)  
刘利容, 刘民. 艾滋病高危人群基数估计方法的研究进展. *国外医学流行病学传染病学分册*, 2005, 32(6): 341-343.
- [3] Lv F, Zhang DP, He X, et al. The progression of methods for estimating the size of populations at risk for HIV/AIDS. *Chin J Epidemiol*, 2003, 24(11): 987-990. (in Chinese)  
吕繁, 张大鹏, 贺雄, 等. 艾滋病高危人群基数估计及其方法. *中华流行病学杂志*, 2003, 24(11): 987-990.
- [4] Wang B, Cheng F, Liang BH, et al. Applications of capture-recapture method for estimating the size of sub-population at high risk for HIV/AIDS. *Mod Prev Med*, 2004, 31(6): 832-834. (in Chinese)  
王斌, 程峰, 梁伯衡, 等. 捕获-再捕获法在艾滋病高危人群基数估计中的运用. *现代预防医学*, 2004, 31(6): 832-834.
- [5] Li X, Cao WH. The progression of methods for estimating the size of populations at risk for acquired immunodeficiency syndrom. *Chin J Epidemiol*, 2007, 28(11): 1138-1140. (in Chinese)  
李希, 曹卫华. 艾滋病高危人群基数估计方法进展. *中华流行病学杂志*, 2007, 28(11): 1138-1140.
- [6] Bernard HR, Hallett T, Iovita A, et al. Counting hard-to-count populations: the network scale-up method for public health. *Sex Transm Infect*, 2010, 86 Suppl 2: S11-15.
- [7] Bernard HR, Johnsen EC, Killworth PD, et al. Estimating the size of an average personal network and of an event subpopulation: some empirical results. *The Small World*, 1991, 20(2): 109-121.
- [8] Johnsen EC, Bernard HR, Killworth PD, et al. A social network approach to corroborating the number of AIDS/HIV+ victims in the U.S. *Soc Net*, 1995, 17(3): 167-187.
- [9] Kadushin C, Killworth PD, Bernard HR, et al. Scale-up methods as applied to estimates of heroin use. *J Drug Issues*, 2006, 36(2): 417-440.
- [10] McCormick T, Salganik MJ, Zheng T. How many people do you know. *Am Stat Assoc*, 2010, 105(489): 59-70.
- [11] Snideroa S, Morrab B, Corradetton R, et al. Use of the scale-up methods in injury prevention research: an empirical assessment to the Case of Choking in Children. *Soc Net*, 2007, 29(4): 527-538.
- [12] Salganik MJ, Fazito D, Bertoni N, et al. Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *Am J Epidemiol*, 2011, 174(10): 1190-1196.
- [13] Bao SL, Wu GH, Zhang W, et al. Application on size estimation through the network scale-up method on men who have sex with men in Chongqing municipality. *Chin J Epidemiol*, 2012, 33(10): 1036-1039. (in Chinese)  
包水莲, 吴国辉, 张维, 等. 利用网络规模迭加法估计重庆市男男性行为人群规模的研究. *中华流行病学杂志*, 2012, 33(10): 1036-1039.
- [14] Hui S, Wang L, Guo W. Estimation on the scale of men who have sex with men using the network scale-up method. *Inter J Epidemiol Infect Dis*, 2010, 37(6): 405-407. (in Chinese)  
惠珊, 王璐, 郭巍. 网络规模迭加法及其在男男性接触者人群规模估计中的应用. *国际流行病学传染病学杂志*, 2010, 37(6): 405-407.

(收稿日期: 2013-06-28)  
(本文编辑: 张林东)