

捕获-再捕获法不同标记物对估计男男性行为人群规模影响的调查

李桂英 卢红艳 孙燕鸣 贺淑芳 马小燕 贺雄

【摘要】 目的 探讨捕获-再捕获方法不同标记物对估计 MSM 规模的影响。方法 2013 年 3—7 月应用捕获-再捕获法,间隔半个月,分别从网络和现场调查前后 2 次捕获 MSM,第一次网络捕获时设置了 2 种不同的标记物,采用 Chapman 等提出的无偏估计公式估计北京地区 MSM 人群规模。结果 使用第一种标记物估计的北京地区 MSM 规模为 239 258 人(95%CI:48 932 ~ 429 584)。使用第二种标记物估计的北京地区 MSM 规模为 115 731 人(95%CI:78 153 ~ 153 309)。结论 标记物不同,估计的北京地区 MSM 人群规模相差很大,使用捕获-再捕获原理进行特定人群规模估计时应慎重选择标记物。

【关键词】 男男性行为者;捕获-再捕获法;规模估计

The impact of different markers regarding the estimation of population size under capture-recapture method on men who have sex with men Li Guiying, Lu Hongyan, Sun Yanming, He Shufang, Ma Xiaoyan, He Xiong. Beijing Center for Disease Control and Prevention, Beijing 100013, China

Corresponding author: Li Guiying, Email: 1176299239@qq.com

【Abstract】 Objective To discuss the impact of different markers on the estimation of MSM using the capture-recapture method. **Methods** From March to July 2013, MSM were captured from their related network to estimate the size of this population, under the capture-recapture method. Two different markers were set in the process of first 'capture'. **Results** The estimated numbers of MSM in Beijing were 239 258 (95% CI: 48 932–429 584) and 115 731 (95% CI: 78 153–153 309) when using the two markers under capture-recapture method. **Conclusion** The estimated size of MSM population varied greatly under different markers which called for the right selection on markers when using the capture-recapture method to estimate the size of specific population.

【Key words】 Men who have sex with men (MSM); Capture-recapture method; Estimate of the population size

MSM 是北京地区 HIV 感染率最高的人群,准确估计该人群的规模是制定综合干预措施的依据。由于 MSM 人群的隐蔽性,应用捕获-再捕获法间接估计该人群规模简单易行。而捕获-再捕获法中,标记物不同会对估计结果产生影响。本研究采用 2 种标记物的捕获-再捕获法估计北京地区 MSM 人群规模^[1,2],比较标记物对估计结果的影响。

对象与方法

1. 研究对象:2013 年 3—7 月选择淡蓝网(MSM 网站)及 blueD 交友软件上现住址在北京市有过口

交或肛交的 MSM 人群。

2. 调查方法:

(1)第一次捕获及 2 种标记物的选择:选择淡蓝网完成第一次捕获,标记物是淡蓝网专为本研究设计的图标——“我在北京我签到”。根据 IP 地址判断有效标记人数,即从 2013 年 3 月 15 日起统计 IP 地址在北京并点击了有效标记物的人数。本次网络标记暂定 4 000 人。淡蓝网进行实时监控,有效标记人数达到要求时停止捕获。第二种标记物是 blueD 交友软件。由淡蓝网统计 2013 年 3 月 15 日至 4 月 15 日 IP 地址在北京并注册过 blueD 的 MSM 人数及其用户名。

(2)第二次捕获:完成第一次网络捕获半个月后,进行第二次捕获和个体识别。第二次捕获结合哨点监测进行。哨点监测采用同伴推动抽样(RDS)

DOI: 10.3760/cma.j.issn.0254-6450.2014.09.017

作者单位:100013 北京市疾病预防控制中心

李桂英、卢红艳同为第一作者

通信作者:李桂英, Email: 1176299239@qq.com

方法招募调查对象。从 MSM 活动场所(包括酒吧、公园、浴池)和网络(如 QQ 和网络聊天室)中招募志愿者作为第一轮种子,种子覆盖城市和农村。第一轮种子不少于 10 人,利用分发招募卡的方法,进一步招募调查对象,直到样本量达到 600 人。招募对象是现住址在北京和经常在此地活动的 MSM。结合哨点监测,对招募来的 MSM 人群开展调查。询问调查对象是否知道淡蓝网“我在北京我签到”的图标,是否进行了点击,点击该图标时是否在北京;询问调查对象是否知道淡蓝网 blueD 交友软件、是否注册过,记录其用户名,由淡蓝网将 3 月 15 日至 4 月 15 日 IP 地址在北京并注册过 blueD 的用户的用户名与现场调查获得的用户名进行比对,获得用户名完全吻合的人数。

结 果

1. 基于第一种标记物的数据:第一次捕获中,2013 年 3 月 15 日至 4 月 15 日共有 3 033 人点击了“我在北京我签到”的图标,其中 IP 地址在北京并点击了有效标记的人数为 2 269 人。第二次捕获中,现场共调查 MSM 人群 526 人,其中知道“我在北京我签到”图标的共 23 人,知道该图标并点击的共 5 人,点击时在北京的共 4 人。

2. 基于第二种标记物的数据:第一次捕获中,2013 年 3 月 15 日至 4 月 15 日共有 73 959 人注册过 blueD 交友软件,其中 IP 地址在北京的共 7 246 人。第二次捕获中,现场共调查 MSM 人群 526 人,其中 IP 地址在北京并注册过 blueD 的共有 70 人,其中与上述 7 246 人的用户名完全匹配的有 32 人。

3. 估计值比较:采用 Corrao 等^[3]提出的无偏估计公式:

$$N = [(M+1)(n+1)/(m+1)] - 1$$

式中 N 的 95% CI : $N \pm 1.96/\sqrt{\text{Var}(N)}$, $\text{Var}(N) = [(M+1)(n+1)(M-m)(n-m)]/[(m+1)^2(m+2)]$

基于第一种标记物估计:

N (“我在北京我签到”) = $(2\ 269 + 1) \times (526 + 1) / (4 + 1) - 1 = 239\ 258$ 人(95% CI : 48 932 ~ 429 584)。

基于第二种标记物估计:

N (blueD 交友软件) = $(7\ 246 + 1) \times (526 + 1) / (32 + 1) - 1 = 115\ 731$ 人(95% CI : 78 153 ~ 153 309)。

讨 论

本研究在同一时间采用同一方法(捕获-再捕获)对同一人群规模进行了估计,结果表明,标记物

不同,估计的北京 MSM 人群规模存在很大差异,提示使用捕获-再捕获原理进行特定人群规模估计时标记物选择比较关键,直接影响到估计结果的真实性。本研究第一种标记物是淡蓝网专为本次 MSM 人群规模估计设计的图标(“我在北京我签到”),为了有一定的影响效应,淡蓝网组织了专项宣传活动;第二种标记物是 blueD 交友软件。比较 2 种标记物发现,对于 MSM 人群来说,blueD 交友软件比“我在北京我签到”图标更有实用价值,本研究调查前了解到 blueD 交友软件是目前 MSM 人群普遍知道并用来寻找圈内朋友的,因此 MSM 人群对注册 blueD 交友软件记忆更清晰。虽然淡蓝网组织了“我在北京我签到”图标的专项宣传活动,但范围有限,不足以改变此标记物的认知度缺陷,也就是说回忆是否点击过“我在北京我签到”图标更容易引起回忆偏倚造成标记的丢失。结合本研究,基于 blueD 交友软件的最终估计结果为 115 731 人,95% CI : 78 153 ~ 153 309。基于“我在北京我签到”图标的最终估计结果为 239 258 人,95% CI : 48 932 ~ 429 584。2 种结果进行比较,可以看出,基于 blueD 交友软件最终估计结果的 95% CI (78 153 ~ 153 309)更合理,而另一种标记物的 95% CI (48 932 ~ 429 584)跨度太大,由此也进一步印证回忆是否点击过“我在北京我签到”图标标记物存在记忆丢失。因此进行 MSM 人群的基数估计时选择合适的标记物是估计成功的关键,标记物一定要有明显的特征以便于接受者留有清晰的印象。

北京市流动人口数目巨大,特别是 MSM 人群流动性更大,为了估计准确,本研究限定了登录地址在北京,点击“我在北京我签到”图标时在北京,现场调查在北京,注册 blueD 交友软件的 IP 地址在北京,因此最大限度地保证了研究样本来自北京,可以认为总体是封闭的。2 种标记物的比较表明,blueD 交友软件作为标记不容易丢失,由淡蓝网将 3 月 15 日至 4 月 15 日 IP 地址在北京并注册过 blueD 的用户名与现场调查获得的用户名进行比对,获得用户名完全吻合的人数,因此个体在捕获和再捕获过程中能够相互匹配。注册 blueD 交友软件的用户名是唯一的,现场调查时使用了指纹仪,个体不会被重复标记。第一次捕获在网上进行,第二次捕获在哨点监测现场进行,两次捕获间隔了 15 d,调查对象是否上网与是否来参加哨点监测不存在关联,所以两个来源是相对独立的。

根据北京市既往 MSM 估计数据^[4,5],结合不

同标记物的比较,结果显示,北京市 MSM 人群规模为 115 731 人(78 153 ~ 153 309)更符合实际情况。

本研究存在不足,由于第一次捕获在网上进行,要求研究对象有上网条件,但北京信息化程度高,在一定程度上弥补了该不足。第二次捕获以现场调查的方式进行,由于 MSM 人群的特点,现场调查时不可能做到完全随机抽样,造成第二次捕获中获得带有第一次捕获标记个体的比例过高过低的可能性都存在,从而使得最终估计结果偏低或偏高,但本研究中现场调查采用了 RDS 方法招募调查对象,从“种子”的选择开始即兼顾到了各个层次的 MSM,而且每个“种子”只发 3 张联系卡,杜绝了对某一支 MSM 人群的过度抽样,在一定程度上满足了随机抽样的要求。因此,对估计结果的影响不大。

参 考 文 献

[1] International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I : History and theoretical development [J]. Am J Epidemiol, 1995,

142(10):1047-1058.
 [2] International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II : Applications in human diseases [J]. Am J Epidemiol, 1995, 142 (10):1059-1068.
 [3] Corrao G, Bagnardi V, Vittadini G, et al. Capture-recapture methods to size alcohol related problems in a population [J]. J Epidemiol Community Health, 2000, 54(8):603-610.
 [4] Wang LY, Xia DY, Wu YH, et al. Application of a multiplier method to estimate the population size of men who have sex with men (MSM) [J]. South Chin J Prev Med, 2006, 32(3):9-15. (in Chinese)
 王丽艳,夏冬艳,吴玉华,等. 乘法估计北京、哈尔滨两市男性同性恋人群规模的研究[J]. 华南预防医学, 2006, 32(3):9-15.
 [5] Xia DY, Lu HY, Zhang DP, et al. Size estimation of men who have sex with men using multiplier method [J]. Chin J AIDS STD, 2006, 12(4):333-335. (in Chinese)
 夏冬艳,卢红艳,张大鹏,等. 利用乘法估计男男性接触者规模的调查研究[J]. 中国艾滋病性病, 2006, 12(4):333-335.

(收稿日期:2014-03-17)

(本文编辑:王岚)

读者·作者·编者

本刊常用医学词汇缩略语

本刊对以下较为熟悉的一些常用医学词汇将允许直接用缩写,即在文章中第一次出现时,可以不标注中文和英文全称。

A 值	吸光度值	HBcAg	乙型肝炎核心抗原	PBS	磷酸盐缓冲液
AIDS	艾滋病	HBsAg	乙型肝炎 e 抗原	PCR	聚合酶链式反应
ALT	丙氨酸氨基转移酶	HBsAg	乙型肝炎表面抗原	RR	相对危险度
AST	天冬氨酸氨基转移酶	Hb	血红蛋白	RT-PCR	反转录聚合酶链式反应
BMI	体重指数	HC	臀围	SARS	严重急性呼吸综合征
CHD	冠心病	HDL-C	高密度脂蛋白胆固醇	SBP	收缩压
CI	可信区间	HFRS	肾综合征出血热	SCr	血清肌酐
COPD	慢性阻塞性肺疾病	HI	血凝抑制试验	T2DM	2 型糖尿病
CT	计算机断层扫描技术	HIV	人类免疫缺陷病毒	TC	总胆固醇
DBP	舒张压	HPV	人乳头瘤病毒	TG	甘油三酯
DNA	脱氧核糖核酸	ICU	重症监护病房	UA	尿酸
ELISA	酶联免疫吸附试验	IDD	碘缺乏病	WBC	白细胞
FPG	空腹血糖	IFG	空腹血糖受损	WC	腰围
GMT	几何平均滴度	LDL-C	低密度脂蛋白胆固醇	WHR	腰臀围比值
HAV	甲型肝炎病毒	M 值	中位数	WHtR	腰围身高比
HBV	乙型肝炎病毒	MRI	磁共振成像	WHO	世界卫生组织
HCV	丙型肝炎病毒	MS	代谢综合征	抗-HBs	乙型肝炎表面抗体
HDV	丁型肝炎病毒	MSM	男男性行为者	抗-HBc	乙型肝炎核心抗体
HEV	戊型肝炎病毒	OR	比值比	抗-HBe	乙型肝炎 e 抗体