

# 单纯随机抽样设计在社区人群调查中的应用

吕筠 刘庆敏 任艳军 王文凤 宫廷 李立明

**【摘要】** 目的 探讨单纯随机抽样设计在社区人群调查中的可行性及调查样本质量。方法 在杭州市下城区和拱墅区根据社区居民电子底册以单纯随机抽样方法抽取居民户,对户内 18~64 岁个体采用 KISH 方法随机抽取一名个体,两区各需完成 500 人的调查。结果 下城区抽取 950 户,完成调查 511 户(53.8%);拱墅区抽取 1380 户,完成调查 506 户(36.7%)。两区因户内不符合年龄要求个体、原户搬迁、社区集体拆迁、底册错误等导致的无应答分别为 38.3% 和 43.5%;各种原因导致的户(或抽中个体)无应答(或拒答)分别为 8.0% 和 19.9%。调查样本与随机抽样户人群的年龄、性别构成无差异。随机抽样户人群与杭州市市区人群性别构成无差异,但年龄结构偏大。结论 在地域相对局限的社区中,基于社区居民电子底册实施单纯随机抽样具备可行性,对调查员的入户时间提出要求,可保证调查样本对抽样框的代表性。

**【关键词】** 单纯随机抽样; 抽样调查; 社区人群

**Application of a simple random sampling method on surveys at the community level** LV Jun<sup>1</sup>, LIU Qing-min<sup>2</sup>, REN Yan-jun<sup>2</sup>, WANG Wen-feng<sup>1</sup>, GONG Ting<sup>1</sup>, LI Li-ming<sup>1</sup>. 1 Department of Epidemiology & Biostatistics, School of Public Health, Peking University Health Science Center, Beijing 100191, China; 2 Hangzhou Center for Disease Control and Prevention

Corresponding author: LI Li-ming, Email: lmlee@pumc.edu.cn

This work was supported by a grant from the Community Intervention for Health (CIH), Organized by Oxford Health Alliance and Funded by Pepsico Foundation

**【Abstract】** Objective To study the feasibility of a simple random sampling on surveys at the community level and to evaluate the quality of samples under survey. Methods A simple random sample of households was taken, based on the electronic listings of community households from Gongshu and Xiacheng districts of Hangzhou city. One of the adults aged 18 to 64 years in the sampled households was identified with KISH method to finish a questionnaire survey. More than 500 people from the sample size was required in each district. Results Of 950 sampled households in Xiacheng district, 511 (53.8%) finished the survey while 506 (36.7%) out of the 1380 sampled households in Gongshu district did. The proportions of non-response due to the following reasons as: none with eligible age in the households, relocation of the original household, mass relocation of the community, and errors in the household listings etc. were 38.3% and 43.5% respectively, in the two districts. Proportions attributed to non-response and refusal to response of sampled household or individual were 8.0% and 19.9% respectively. No statistical significant differences in age and gender were found between the surveyed samples and the population in the sampled households, or in gender between the populations in the sampled households and in Hangzhou city. However, the population in the sampled households showed a more aging population structure than the population in Hangzhou city. Conclusion In a geographically limited area, using a simple random sampling method to do the survey is feasible, based on the electronic listings of household. Enough time spent during the household visit guarantees the interviewers to get a representative sample of the sampling frame. There is an urgent need for the timeliness, completeness and accuracy of electronic household listings to be improved.

**【Key words】** Simple random sampling; Sampling studies; Community surveys

在以社区为基础的人群调查中,研究者由于考

虑组织实施的便利多采取整群随机抽样等设计,较少应用单纯随机抽样设计<sup>[1]</sup>。然而,除单纯随机抽样以外的其他复杂抽样设计,无论是样本量估算、抽样设计还是数据分析都存在一定的技术难度。为了减少工作量选择几个或十几个群进行调查,多数情况下结果无法达到期望的精确度;即使按适当的方

DOI: 10.3760/cma.j.issn.0254-6450.2010.04.014

基金项目:社区健康干预项目

作者单位:100191 北京大学公共卫生学院流行病学与卫生统计学系 教育部流行病学重点实验室(吕筠、王文凤、宫廷、李立明);杭州市疾病预防控制中心(刘庆敏、任艳军)

通信作者:李立明(杭州市钱江特聘专家), Email: lmlee@pumc.edu.cn

法估算样本量,像同质性这样的参数,由于较难在已发表的文献中找到参考值,所以样本量的估算仍然存在很大的不确定性。在分析这类复杂抽样调查数据时,还要考虑是否需要加权,并使用专门的统计方法估计抽样误差和进行统计学检验。目前国内学者对复杂抽样调查的设计和数据分析普遍认识不足和使用不当。本次调查作为社区健康干预 (community interventions for health, CIH) 项目基线调查的一部分,在杭州市下城区和拱墅区的社区人群调查中采取严格的单纯随机抽样设计,为今后进行类似的调查设计提供参考。

### 对象与方法

1. 调查对象:调查现场为杭州市下城区和拱墅区。调查覆盖下城区 6 个街道共 50 个居委会,拱墅区 5 个街道共 33 个居委会。在各区内居住至少满 1 年的为合格调查户 (排除集体户口); 年龄 18 ~ 64 岁,身体或精神状态正常并能回答问题者为调查对象。根据项目设计要求,两区各至少需要 500 名合格个体的调查。在各区内对所有调查社区的居民户电子底册进行累积排序,以单纯随机抽样方法 (固定抽样种子数) 抽取居民户; 对户内 18 ~ 64 岁个体采用 KISH 方法随机抽取一名个体进行调查<sup>[2]</sup>。

2. 调查方法:采取入户调查。如果社区底册中的登记户已更换为其他居民户,新户在本区内居住满 1 年,则以新户替换。如遇以下任何一种情况,放弃抽中的居民户,重新抽样:①户内无 18 ~ 64 岁个体;②登记户已更换为其他居民,新户在本区居住尚不满 1 年;③空户、社区集体拆迁;④调查员在不同日入户 3 次,其中至少 1 次为晚间入户,3 次入户均无人应答;⑤底册信息陈旧,抽中的个体已死亡,或者底册信息错误,无法找到对应的居民户;⑥抽中的个体拒答;⑦抽中的个体因各种原因无应答,如因身体或精神原因无法接受调查、近期出差在外、个体已经移居其他区、集体户口等。调查时间为 2008 年 10 月至 2009 年 2 月。

3. 统计学分析:对调查完成情况及调查样本质量和代表性进行分析评价:①应用 Myers 混合指数 (Myers' blended index) 判断调查样本有无年龄尾数偏好<sup>[3]</sup>;②比较调查样本与随机抽样户内 18 ~ 64 岁人群的年龄和性别构成;③比较随机抽样户与 2007 年末杭州市市区人口的年龄和性别构成<sup>[4]</sup>。年龄、性别构成比较采用频数分布拟合优度的  $\chi^2$  检验。

### 结 果

1. 调查完成情况:扣除集体拆迁户或因社区底册信息不全且入户无应答的居民户,下城区共 853 户有户基本信息 (即户内人口数和成员名单), 合计 2078 人, 户均人数 2.44 人; 拱墅区共 946 户, 合计 2314 人, 户均人数 2.45 人。经检验, 下城区与拱墅区的调查样本及随机抽样户人群年龄和性别分布的差异无统计学意义 (表 1)。

表 1 杭州市下城区和拱墅区社区人群单纯随机抽样调查完成情况

未调查原因	下城区		拱墅区	
	户数	构成比 (%)	户数	构成比 (%)
完成调查	511	53.8	506	36.7
户内无 18 ~ 64 岁个体	109	11.5	118	8.6
户内无人应答	11	1.2	133	9.6
空户、原户搬迁换新户 (新户不合格)	167	17.6	204	14.8
社区集体拆迁	50	5.3	265	19.2
户内抽中个体拒答	46	4.8	92	6.7
户内抽中个体因各种原因无应答	19	2.0	50	3.6
其他 (底册信息陈旧或错误)	37	3.9	12	0.9
合计	950	100.0	1380	100.0

2. 调查样本的质量和代表性评价:

(1) Myers 混合指数: Myers 混合指数男性为 7.547, 女性为 5.466, 总样本为 5.005, 提示调查样本无明显年龄偏好。

(2) 调查样本与随机抽样户人群年龄、性别构成比较: 两区调查样本合计性别比 0.979, 随机抽样户内 18 ~ 64 岁人群的性别比为 0.999, 性别构成的差异无统计学意义 ( $\chi^2=0.105, P=0.746$ )。调查样本与随机抽样户人群的年龄构成差异也无统计学意义 ( $\chi^2=4.984, P=0.289$ )。

(3) 随机抽样户人群与 2007 年末杭州市区人群年龄、性别构成比较: 随机抽样户内全人口性别比为 0.984, 2007 年末杭州市区全人口性别比为 1.018 (18 ~ 64 岁组为 1.028), 无论是全年龄组 ( $\chi^2=1.165, P=0.280$ ) 还是 18 ~ 64 岁组 ( $\chi^2=0.637, P=0.425$ ), 性别构成的差异均无统计学意义。而随机抽样户人群的年龄构成与市区人口比较, 无论是全年龄组 ( $\chi^2=649.115, P<0.001$ ) 还是 18 ~ 64 岁组 ( $\chi^2=362.562, P<0.001$ ), 差异均有统计学意义。随机抽样户人群的年龄结构整体偏大。

### 讨 论

社区居民户底册是开展以社区为基础的卫生工

作调查中重要的基础。本研究的两个项目区均有社区居民户电子底册,可以完成以居民户为单位的单纯随机抽样调查,且根据多数居民户底册信息直接完成 KISH 抽样,减轻了调查员入户后的工作负担。尽管如此,现有的底册中还是存在一些信息陈旧、不全、错误的问题。

空户、原户搬迁换新户是本次调查无应答的主要原因之一。另外,拱墅区因社区集体搬迁导致的无应答高达 20%。城市居民的流动及城市建设伴随的社区拆迁,也给利用居民户底册的调查带来新问题。如果居民户底册更新及时且少有错误,那么,户内没有符合年龄要求的个体、空户、原户搬迁换新户、集体拆迁,这些问题都可在利用电子底册进行抽样的阶段加以解决。扣除前述这部分居民户,入户阶段的应答比例,本文中的下城区高达 87%,而拱墅区约为 65%。

社区入户调查普遍存在的障碍是:如果在工作日的白天入户,较难访问到在校学生或在职成年人,调查结果对这部分人群的代表性较差。但是,如果安排在晚间或周末入户,可实施调查的时间减少,调查进度拖延或者需要更多的调查员。另外,在晚间入户,调查员的安全又是一个问题。本次调查要求调查员对抽中的居民户至少要入户 3 次,其中至少包括一次晚间入户,如果仍然无法完成调查才可放弃。这种做法最大限度地控制了可能存在的应答偏倚,调查样本对随机抽样人群的代表性很好,年龄和性别分布一致。

本次调查选择社区居民户底册作为抽样框,比较随机抽样人群与杭州市区人口的年龄构成,我们发现这样的抽样框,年龄结构整体偏大。思考可能的原因,一方面,两个城区的人口老龄化程度有可能高于整个市区平均水平<sup>[4,5]</sup>。另外,在抽样框架中剔除了集体户口人群,而这部分人群应以中青年为主。尽管如此,我们选择的抽样框应该是适当的,符合本次调查的目标,即捕捉社区常住人口健康相关行为的流行水平及在社区干预活动后人群行为的改变情况,而在职的成年人可另外通过企事业单位中的调查和干预活动实现补充覆盖。

通过本次实践,我们认为在城市地区地域相对局限的社区中开展社区人群调查,基于社区居民户电子底册实施单纯随机抽样设计具备可行性;通过对调查员的入户时间提出要求,可以最大限度地保证应答对象在年龄和性别上的代表性,调查样本质

量较高;后期的数据分析也不存在加权和用特殊的统计方法正确计算抽样误差的复杂技术问题。如果社区居民底册的时效性、信息的完整性及准确性能够得到进一步的完善,将可大大提高单纯随机抽样调查的现场实施效率,即提高入户成功率,减少调查员的工作量,加快调查实施进度。

本文对今后开展类似的城市社区抽样调查的设计和样本量估算具有参考价值。建议今后的研究者在发表流行病学抽样调查的论文时,能够尽量详细地报告抽样设计和实施情况,方便其他研究者在进行调查设计时有更多可供参考的信息,减少因为设计参数的不确定性所要承担的风险,提升我国人群抽样调查的证据质量。

(感谢杭州市、下城区和拱墅区卫生局对项目工作的大力支持,感谢杭州市、下城区和拱墅区疾病预防控制中心的项目人员及全体调查员的辛勤劳动)

#### 参 考 文 献

- [1] Li LM. Epidemiology. 6<sup>th</sup>, ed. Beijing: People's Medical Publishing House, 2007. (in Chinese)  
李立明. 流行病学. 6 版. 北京: 人民卫生出版社, 2007.
- [2] World Health Organization. WHO STEPS surveillance manual. Available at: <http://www.who.int/chp/steps/manual/en/index.html>. Accessible in Aug 2009.
- [3] Siegel JS, Swanson DA. The methods and materials of demography. 2nd ed. Academic Press, 2004.
- [4] Hangzhou Bureau of Statistics, Hangzhou investigation team of National Bureau of Statistics of China, Hangzhou Bureau of Socioeconomic Survey. Hangzhou Statistical Yearbook, 2008. Available at: [http://www.hzstats.gov.cn/web/tjnj/nj2008/2/nj\\_.htm](http://www.hzstats.gov.cn/web/tjnj/nj2008/2/nj_.htm). Accessible in Aug 2009. (in Chinese)  
杭州市统计局, 国家统计局杭州调查队, 杭州市社会经济调查局. 2008 年杭州统计年鉴. Available at: [http://www.hzstats.gov.cn/web/tjnj/nj2008/2/nj\\_.htm](http://www.hzstats.gov.cn/web/tjnj/nj2008/2/nj_.htm). Accessible in Aug 2009.
- [5] Gongshu District Government Office, Gongshu District Information Center. Gongshu district economic and social development profile 2007. Available at: <http://www.gongshu.gov.cn/gongshucms/view/zjgs.jsp? siteuid=CEA370A5CBAB1D0539DEC81C06CF723D&pid=58F5220594AE9A2C1D16DA826984C367&id=E88E174374ED0616BC561E84CABED84>. Accessible in Aug 2009. (in Chinese)  
拱墅区政府办公室, 拱墅区信息中心. 2007 拱墅区经济社会发展概况. 2009. Available at: <http://www.gongshu.gov.cn/gongshucms/view/zjgs.jsp? siteuid=CEA370A5CBAB1D0539DEC81C06CF723D&pid=58F5220594AE9A2C1D16DA826984C367&id=EE88E174374ED0616BC561E84CABED84>. Accessible in Aug 2009.

(收稿日期: 2009-09-07)

(本文编辑: 张林东)