

流行病学研究中的样本代表性问题(一)

潘雄飞 王意 叶依 潘安

430030 武汉,华中科技大学同济医学院公共卫生学院流行病与卫生统计学系

通信作者:潘安, E-mail: panan@hust.edu.cn

DOI:10.16462/j.cnki.zhjbkz.2019.01.001

【摘要】 研究的样本代表性一直是流行病学领域内具有争议性的话题之一。本文首先对流行病学样本代表性进行定义,同时对研究人群、源人群、目标人群、内部真实性、外部真实性等相关流行病学基本概念进行阐述。在此基础上,本文深入分析了横断面研究、队列研究、病例对照研究、干预性研究四种主要流行病学研究设计中样本代表性的价值和实际可行性。总而言之,除针对人群疾病或健康现状的横断面研究外,多数涉及病因或干预效果推断的流行病学研究不应过度强调样本代表性。

【关键词】 流行病学; 样本; 代表性

【中图分类号】 R181 **【文献标识码】** A **【文章编号】** 1674-3679(2019)01-0001-04

Representativeness issues in epidemiological studies PAN Xiong-fei, WANG Yi, YE Yi, PAN An

Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Corresponding author: PAN An, E-mail: panan@hust.edu.cn

【Abstract】 Representativeness has long been a contentious topic in epidemiology. The review starts with defining representativeness in epidemiology and describing fundamental epidemiological concepts including study population, source population, target population, internal validity, and external validity. In addition, it elaborates the role and practicality of representativeness in four major epidemiological study designs including cross-sectional, cohort, case-control, and intervention studies. In summary, representativeness is an important consideration in cross-sectional studies of investigating disease or health status in the target population, but cannot be overstated in epidemiological studies for causal inferences regarding exposure-outcome relationships or intervention effects.

【Key words】 Epidemiology; Sample; Representativeness

(Chin J Dis Control Prev 2019, 23(1):1-4)

样本代表性(representativeness)一直是流行病学领域内颇具争议性的话题,也深深困扰着流行病学界甚至其他的人群研究者^[1-2]。样本代表性相关问题不仅影响着研究设计、实施和数据分析,同时也会影响对研究结果的解读。在精准医学时代,尤其是大队列和大数据研究迅猛发展的背景下,重新讨论该问题具有重要的现实意义。笔者将在两篇系列文章中针对流行病学样本代表性问题进行深入剖析,旨在给广大流行病学和人群研究工作者提供一定的参考意见。作为系列文章的第一篇,本篇将介绍流行病学样本代表性概念,并详述其在四种主要流行病学研究设计中的价值。

1 流行病学中样本代表性的概念

流行病学领域中样本代表性通常是指最终纳入

的研究人群能否代表目标人群。在流行病学研究中,人群可分为目标人群、源人群和研究人群。目标人群为研究结果想要外推适用的人群,源人群则是研究特定问题实际选取样本的人群(也即特定研究假设的总体),而研究人群则为最终具体实施项目时所纳入的人群(也即特定研究的样本)。其中,定义清晰的源人群是开展所有流行病学研究工作的重要基础^[3]。例如,以著名的弗明汉心脏研究(Framingham heart study)为例,其研究的主要目的是探索美国人群中心血管疾病的关键风险因素,因此其目标人群是美国一般居民,源人群则是为了研究这个问题选取的某一地区居民(此处为弗明汉小镇居民),而研究人群则为根据纳入和排除标准最终入组研究的五千多人^[4]。在很多研究中,目标人群和

源人群并未较好区分,而统称为目标人群,这种概念的混淆常会影响研究结果的解读。

样本代表性相关讨论会需要关注流行病学中两个重要概念,即研究结论内部真实性(internal validity)和外部真实性(external validity)^[5]。内部真实性是指研究结果与研究对象真实情况的符合程度,其反映研究人群(样本)到源人群(总体)的统计推断程度,因此,流行病学研究人群通常应能较好的代表源人群。外部真实性,即外推性(generalizability),则是指研究结果与外推对象真实情况的符合程度,其反映的是从源人群到目标人群的推断。通俗而言,流行病学研究样本代表性的主要目标就是确保研究结果具有较好的外部真实性,而研究结果内部真实性是外部真实性的基础。需要指出的是,在统计学中,样本代表性主要是指纳入研究的抽样样本(研究人群)是否可以代表总体(源人群),即样本统计量(如均数、方差等)是否可以代表总体参数,统计学样本代表性常会涉及到抽样和抽样误差等概念和过程,其是统计推断的基础,反映的是研究结论的内部真实性,因此与流行病学研究中代表性概念存在密切联系、相互补充但却有所差异^[6]。

在理想状况下,若研究结果能较好外推至目标人群,则表明研究结论具有较高的应用价值,这对于制订公共卫生政策以及临床实践指南具有重要意义。然而,需要注意的是,流行病学研究中样本代表性并不等同于外部真实性,前者要求在研究对象选择上有周密的抽样计划以确保代表目标人群,这会增加研究实施的困难,可能更易导致应答率偏低或失访,增加选择偏倚和信息偏倚,从而严重影响研究结果的内部真实性^[7]。过度强调样本代表性有时反而会导致无法获得真实的统计推断结果,也即内部真实性较差,更谈不上研究结果的外推性。因此,不同流行病学研究设计中,应对样本代表性问题有不同考虑。

2 什么时候需要考虑样本代表性

在研究疾病及其风险因素分布和疾病负担等方面^[8],选取代表性样本具有重要意义。流行病学研究对样本代表性的重视主要源自于横断面调查研究,通过对目标人群的概率抽样,了解特定时间和特定空间中人群的疾病或健康相关状态的分布情况。设计良好的横断面研究会将目标人群作为源人群,进行概率抽样获得具有代表性的研究人群,因此研究人群常需能较好代表目标人群。全国性疾病调查如成人 2 型糖尿病^[9]、慢性阻塞性肺疾病^[10]、脑卒中^[11]疾病负担调查为典型横断面调查研究,对于国

家和地区卫生决策和卫生资源分配具有重要指导意义,常会需要分阶段开展概率抽样。横断面研究结果通常具有一定时效性、地区性,往往需要周期性更新^[12],从这方面而言样本代表性并非一成不变,其常常可能是一个历史性概念。许多国家都会周期性开展全国性概率抽样调查,例如中国居民营养与健康状况监测^[13]、美国全国健康和营养调查(National Health and Nutrition Examination Survey)^[14]等,以获得最新的具有代表性的结果。在此类研究中,有时为了确保特定群体比如少数族裔能进入研究并获得其疾病或健康相关数据,可采用立意抽样增加特定群体在抽样过程中的概率,以保证特定群体在样本中有足够的样本量。不过,最终在估计整个人群参数的统计分析中会需要考虑立意抽样的影响。需要注意的是,在现实中很多横断面研究常因研究条件限制也会通过非概率抽样方法获取近似具有代表性的研究人群。

3 什么时候不应过分强调样本代表性

与横断面研究中健康事件描述目的不同,队列研究、病例对照研究、干预性研究主要以暴露-结局或干预-效果因果推断为目的,往往首先需要考虑如何提高研究结果在研究人群中的内部真实性,然后再进一步确定其外推的可能性^[6],因此样本的选择并不局限于是否可以完全代表目标人群。此时,源人群常常仅为目标人群的一部分,而研究人群则是源人群中一部分,相对目标人群而言研究人群是否具有代表性则不尽然。

以队列研究为例,其主要是观察不同暴露组人群在随访过程中疾病(或健康)事件的发生率,通过统计分析,从而得出暴露因素是否与疾病(健康)事件有关的结论。队列研究中基线人群的选择主要需要考虑暴露因素是否具有可区分度,当暴露因素在目标人群中比例较低时,通常会考虑增加特定暴露人群的数量以确保随访过程中暴露组有足够的新发病例数,从而确保较好的统计效能。在这种情况下,基线人群暴露因素的分布及其他非暴露因素并不能代表目标人群的分布状况,即样本代表性可能较差。比如 Doll 和 Hill 在利用英国医生研究(British doctors study)确定吸烟与死亡率关系时,其研究的人群是英国医生,该研究群体与一般人群相比在性别、种族、社会阶层等多种因素上均存在显著差异^[15],但这并不会极大的影响研究结果的外推。类似的,美国弗明汉心脏研究队列纳入的是马萨诸塞州弗明汉小镇居民,绝大多数为白种人,收入水平较高,该研究人群并不能代表所有美国人甚至马萨诸塞州人群

的特征,但其在心血管疾病方面的发现如血压和血脂与心血管疾病的关系已在多个其他队列研究中得到验证。其他特定群体中开展的研究比如美国护士健康研究(nurses' health study, NHS)^[16]和医护人员随访研究(health professionals follow-up study, HPFS)^[17],在流行病学研究领域也做出了众多划时代的贡献。

相比于样本代表性,队列研究样本选择时应该重视研究人群的暴露因素是否存在较大异质性,同时混杂因素背景水平需相对较低,且便于随访,基于这些考虑往往会要求研究人群是特定群体,因此常会使用方便抽样或立意抽样,同时采取限制纳入(restriction)、匹配(matching)、分层(stratification)等方法以控制研究中混杂因素,这些处理的目的是确保不同暴露水平的研究对象在除暴露因素外的其他因素上尽量具有可比性,从而获得具有良好内部真实性的效应估计值^[18]。尽管纳入的研究人群本身不具有代表性,但其良好内部真实性结果为外推至目标人群打好了基础^[19]。在此背景下,研究结果的外推性会更多考虑暴露与结局的关联性是否在不同人群中存在生物学机制上的差异,以及研究人群和目标人群在背景危险因素上的差异^[20],而不应局限在研究对象是否能代表目标研究人群^[8]。在某些情况下,这些差异带来的效应修饰(effect modification)可能会导致不同人群中估计的暴露结局关系存在差异。例如,如果吸烟与肺癌的关系在男性和女性中存在异质性^[21],以男性为主的英国医生研究中确定的关联性效应值或许不能直接用于女性比例较高的群体,但在该群体的男性中应该仍然具有可推广性。

病例对照研究是另外一类研究病因的重要观察性研究方法,多用于罕见病或潜伏期较长的疾病研究,或用于探索多种暴露和某一疾病的关系。该研究设计根据疾病状态在源人群中分别选择病例和对照,常常会有研究者认为二者应分别能代表目标人群中的疾病和非疾病人群,但这种观点其实并不完全正确^[7],这主要源于混淆了目标人群和源人群的概念。病例可以是任何特定目标疾病患者,如新发病例或现患病例、轻度或重度病例^[8],因此病例通常容易选定,但是对照选择往往较为棘手,其需要能代表源人群,即产生病例的人群。对照的选择常需要满足两个基本条件:对照应该来自病例同一源人群,同时对照的选择不应依赖于是否存在暴露或暴露水平的高低^[22-23]。正如前文所述,源人群代表性与流行病学研究通常所指的目标人群代表性不是完全等同的概念,前者是指对照能代表研究人群所在

的源人群^[8],更多的是为了改善研究结果内部真实性而非外推性。

理想的病例对照研究设计通常在确定源人群后,再选定相应病例和具有代表性的对照,这样能降低选择偏倚^[7],基于这种考虑,较好的病例对照研究常常在队列人群之中开展,如巢式病例对照或病例队列研究^[24]。相反,很多设计欠佳的病例对照研究可能并未考虑到源人群的选择,病例选择随意,导致源人群范围不明确,此时就无从谈及对照能否代表源人群了。例如,在一些病例对照研究中,病例主要来源于某大型三甲医院,一方面其病人可能较多为严重程度较高的病例,但因其仅为单一医疗机构,病例不能代表某特定地区或特定人群中严重病例状况;另一方面,其病人来源较复杂,病例所在的源人群较难界定^[25],无论从三甲医院非目标疾病人群中选择对照还是从社区人群中选择对照都难以代表病例所在源人群,均会带来不同程度的选择偏倚问题^[26]。通过多中心研究设计在多个地区多个不同等级医院纳入病例的方法,可有助于在理论上界定源人群范围,降低选择偏倚,不过这种研究策略与目标人群代表性问题无关。

干预性研究如随机对照试验(randomized clinical trial)通常存在复杂的纳入和排除标准,入组的研究对象基本很难代表现实中的目标干预对象^[18, 27],因此其外部真实性常常被质疑。干预性研究因其设计优势常可确保研究结论的内部真实性,这为结论外推打好了基础。若在研究对象选择上采用精心设计的概率抽样等方法获得能代表目标人群的样本,则会在可实施性、研究效率、成本方面带来困难,同时可能对应答率、依从性、随访率等均会产生不利影响,在这种情况下干预性研究可能并不能充分发挥控制潜在偏倚和混杂的研究优势,反而降低了内部真实性。因此,在干预性研究中,其设计特征决定了并不能一味追求研究人群对目标应用人群的代表性。在适当增加研究人群多样性基础上,研究结果外推性需要考虑研究人群与目标人群的某些特征差异是否会对研究结果带来效应修饰作用(即存在效应异质性),进而影响相关结论的应用^[28]。例如,某干预研究中主要纳入了中青年研究患者,而现实中干预对象可能会是老年患者,此时需要注意不同年龄患者干预效果是否可能存在异质性,也即年龄对干预效果是否存在潜在效应修饰作用。此外,干预的强度和时长也是外推结果时需要考虑的因素。

4 总结

研究样本的代表性既非科学合理的流行病学研

究设计的充分条件也非必要条件,其价值在不同类型流行病学研究中存在差异。除针对疾病或健康现状的横断面研究外,多数人群研究,特别是以病因或者干预效果推断为目的的研究不应过度强调研究对象对目标人群的代表性。

利益冲突 无

参 考 文 献

- [1] Rothman KJ. Six persistent research misconceptions [J]. *J Gen Intern Med*, 2014, 29(7):1060-1064. DOI: 10.1007/s11606-013-2755-z.
- [2] Nohr EA, Olsen J. Commentary: Epidemiologists have debated representativeness for more than 40 years--has the time come to move on? [J]. *Int J Epidemiol*, 2013, 42(4):1016-1017. DOI: 10.1093/ije/dyt102.
- [3] Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology [J]. *Occup Environ Med*, 2007, 64(9):633-638. DOI:10.1136/oem.2006.029967.
- [4] Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS); overview of milestones in cardiovascular epidemiology [J]. *Int J Epidemiol*, 2015, 44(6):1800-1813. DOI: 10.1093/ije/dyv337.
- [5] Rothman KJ, Greenland S. Validity and generalizability in epidemiologic studies [J]. *Encyclopedia of Biostatistics*, 2005. DOI: 10.1002/0470011815.b2a03129.
- [6] Ponsonby AL, Dwyer T, Couper D. Is this finding relevant? Generalisation and epidemiology [J]. *Aust N Z J Public Health*, 1996, 20(1):54-56.
- [7] Rothman KJ, Greenland S, Lash LL. *Modern Epidemiology* [M]. (Third edition) Australia: Lippincott Williams & Wilkins, 2012.
- [8] Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. [J]. *Int J Epidemiol*, 2013, 42(4):1012-1014. DOI:10.1093/ije/dys223.
- [9] Wang L, Gao P, Zhang M, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013 [J]. *JAMA*, 2017, 317(24):2515-2523. DOI:10.1001/jama.2017.7596.
- [10] Fang L, Gao P, Bao H, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study [J]. *Lancet Respir Med*, 2018, 6(6):421-430. DOI: 10.1016/S2213-2600(18)30103-30106.
- [11] Wang W, Jiang B, Sun H, et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults [J]. *Circulation*, 2017, 135(8):759-771. DOI:10.1161/CIRCULATIONAHA.116.025250.
- [12] Schooling CM, Jones HE. Is representativeness the right question? [J]. *Int J Epidemiol*, 2014, 43(2):631-632. DOI:10.1093/ije/dyt264.
- [13] 中国疾病预防控制中心营养与健康所. 中国居民营养与健康状况监测项目简介 [EB/OL]. (2015-05-05) [2018-09-03] http://www.chinanutri.cn/gzrw_132/zgjmyyjkzjkjcxm/201505/t20150505_114496.html.
- [14] National Center for Health Statistics. National Health and Nutrition Examination Survey. [EB/OL]. (2013-09-30) [2018-09-04] <https://www.cdc.gov/nchs/nhanes/index.htm>.
- [15] Doll R, Hill AB. The mortality of doctors in relation to their smoking habits; a preliminary report [J]. *Br Med J*, 1954, 1(4877):1451-1455. DOI:10.1136/bmj.328.7455.1529.
- [16] Colditz GA, Philpott SE, Hankinson SE. The impact of the Nurses' Health Study on population health; prevention, translation, and control [J]. *Am J Public Health*, 2016, 106(9):1540-1545. DOI:10.2105/AJPH.2016.303343.
- [17] Rimm EB, Stampfer MJ, Colditz GA, et al. Effectiveness of various mailing strategies among nonrespondents in a prospective cohort study [J]. *Am J Epidemiol*, 1990, 131(6):1068-1071.
- [18] Keyes KM, Galea S. *Epidemiology matters: a new introduction to methodological foundations* [M]. USA: Oxford University Press, 2014.
- [19] Pizzi C, De Stavola B, Merletti F, et al. Sample selection and validity of exposure-disease association estimates in cohort studies [J]. *J Epidemiol Community Health*, 2011, 65(5):407-411. DOI:10.1136/jech.2009.107185.
- [20] Szklo M, Nieto FJ, Miller D. *Epidemiology: beyond the basics* [M]. USA: Oxford University Press, 2001.
- [21] Zang EA, Wynder EL. Differences in lung cancer risk between men and women: examination of the evidence [J]. *J Natl Cancer Inst* 1996, 88(3-4):183-192.
- [22] Carneiro I, Howard N. *Introduction to epidemiology* (Second edition) [M]. UK: Open University Press, 2011.
- [23] Schulz KF, Grimes DA. Case-control studies; research in reverse [J]. *Lancet*, 2002, 359(9304):431-434.
- [24] Kopec JA, Esdaile JM. Bias in case-control studies. A review [J]. *J Epidemiol Community Health*, 1990, 44(3):179-186.
- [25] Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles [J]. *Am J Epidemiol*, 1992, 135(9):1019-1028.
- [26] Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls [J]. *Am J Epidemiol*. 1992, 135(9):1029-1041.
- [27] Frieden TR. Evidence for health decision making - beyond randomized, controlled trials [J]. *N Engl J Med*, 2017, 377(5):465-475. DOI:10.1056/NEJMra1614394.
- [28] Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations [J]. *Prev Sci*, 2015, 16(3):475-485. DOI:10.1007/s11121-014-0513-z.

(收稿日期:2018-12-21)

(本文编辑:王军平)