

大型前瞻性人群队列研究进展

李立明 吕筠

【关键词】 队列研究; 前瞻性

Large prospective cohort studies: a review and update Li Liming^{1,2}, Lyu Jun¹. 1 School of Public Health, Peking University, Beijing 100191, China; 2 Chinese Academy of Medical Sciences

Corresponding author: Li Liming, Email: lmlee@pumc.edu.cn

This work was supported by grants from the National Natural Science Foundation of China

(No. 81390541, No. 81390544), Kadoorie Charitable Foundation in Hong Kong and Wellcome Trust in the UK (No. 088158/Z/09/Z, No. 104085/Z/14/Z).

【Key words】 Cohort study; Prospective

队列研究是流行病学最基本的观察性研究设计之一。当以检验病因假说为目的时,队列研究可以探讨有害暴露的致病作用,基于人群构建的队列(population-based cohort)可以用来研究多种暴露因素和多种健康结局的关系^[1],研究结果外推性好;这些都是优于随机临床试验的特点。队列研究中先因后果的时间顺序相对明确,受一些偏倚的影响小,证据强度更高,优于其他观察性研究设计。队列研究尤其适用于某些暴露和疾病结局的研究,如亚临床疾病期长、发病时死亡风险高或病程短的疾病,致病潜隐期较长的暴露因素,会因个体疾病状态、治疗或生活方式改变而变化的暴露因素,容易发生回忆偏倚的暴露因素等^[2]。队列研究也特别适合研究具有共同暴露因素的多种疾病结局。总之,队列研究在病因学研究中具有不可替代的地位和作用。

在传统的流行病学教科书中,队列研究的局限性之一是不适用于发病率很低的疾病的病因研究,因为需要很大数量的研究对象,开展难度很大。然而,近年来,随着人们对复杂性疾病病因研究的不断深入,从单纯的检验环境因素或遗传因素的病因假说,发展到关注基因-基因-基因-环境的复杂交互作用;即使不是罕见病,研究者对样本量的需求也是越来越高。鉴于队列研究在病因学研究中不可替代的地位,且随着各国政府、各种基金组织对科学研究的重视与投入的增加,以及现代信息化管理技术的发展,

如今样本量在20万以上规模的人群队列已不罕见。

1. 国外进展:国外早期最经典的前瞻性人群队列研究当属1948年开始的美国弗明汉心脏研究(Framingham Heart Study, FHS),历经三代人群队列(起始时间分别为1948年、1971年和2002年),确定了心脏病、脑卒中和其他疾病的重要危险因素,带来预防医学的革命,改变了医学界和公众对疾病起源的认识。

自20世纪七八十年代起,一些欧美发达国家陆续开始建立长期随访的人群队列。早期队列的规模相对不大,多数队列中仅采集了少部分队列成员的生物学样本。截至目前为止,这些人群队列多已随访20年以上,为帮助人们了解各种暴露因素(尤其是生活方式和环境暴露)与主要慢性病的长期发生风险间的关联产出了大量有影响力的研究证据。一些人群队列每隔若干年向队列成员邮寄问卷进行随访;如著名的美国护士健康研究(Nurses' Health Study, NHS)及NHS II,每隔两年随访一次,每轮调查的应答率都在90%左右,对了解各种暴露因素的长期变化及对远期慢性病风险的影响有不可替代的作用。另外,随着随访时间的延长,很多人群队列的成员已经陆续步入老年。因此,研究内容上也逐渐从早年的肿瘤、心血管疾病、2型糖尿病等主要慢性病拓展为老年期的神经退行性疾病及衰老问题。

自20世纪90年代以来,各国基于各种研究目的建立的人群队列如雨后春笋。仅国际流行病学杂志(International Journal of Epidemiology, IJE)自2005年起连续刊载的全球不同类型的人群队列简介(cohort profile)就有200余项。新建立的人群队列的规模越来越大。其中,达到50万左右规模的人群队列有四项:欧洲10国的European Prospective

DOI: 10.3760/cma.j.issn.0254-6450.2015.11.001

基金项目:国家自然科学基金(81390541, 81390544); 香港Kadoorie Charitable基金; 英国Wellcome Trust(088158/Z/09/Z, 104085/Z/14/Z)
作者单位:100191 北京大学公共卫生学院(李立明、吕筠); 中国医学科学院(李立明)

通信作者:李立明, Email: lmlee@pumc.edu.cn

Investigation into Cancer and Nutrition (EPIC, 52.1万)、美国的NIH-AARP Diet and Health Study (NIH-AARP, 56.6万)、英国的Million Women Study (MWS, 130.0万)和UK Biobank(UKB, 49.8万)。除此之外,如法国的Nutrinet-Santé Study、瑞典的LifeGene等仍在队列成员募集期,目标规模也是50万人。而美国在2015年初宣布启动的“精准医学行动”(Precision Medicine Initiative)中也计划建立一个100万以上规模的美国人群研究队列。

另有一些前瞻性人群研究是利用国家或地区的人口普查资料、征兵系统、死亡监测系统、医院信息、医疗保险数据等匹配并获取更大数量居民的基本人口社会学信息和健康结局信息,但是缺乏个体生活方式和环境暴露等信息。例如Swiss National Cohort(680万)、Scottish Health and Ethnicity Linkage Study(465万)、1991 Canadian Census Cohort(273万)、台湾省全民健康保险研究资料库(230万)、Kaiser Permanente Southern California (KPSC) Children's Health Study(92万)、Danish Conscript Database(DCD, 72.8万)等。

另外,自20世纪90年代初期以来,“成人疾病胎源说”越来越受到关注,越来越多的出生队列建立起来。英国在早年相隔一段时期就会建立一些规模相对较小的出生队列,如1958 Birth Cohort Study(1.7万出生于1958年者)、1970 British Cohort Study(BCS70, 1.7万出生于1970年者)、Longitudinal Study of Young People in England(LSYPE, 1.6万出生于1989—1990年者)、Millennium Cohort Study(MCS, 1.9万出生于2000—2001年者)。目前,规模比较大的是丹麦的Danish National Birth Cohort(DNBC, 10.1万妊娠早期孕妇及其生产的9.7万婴儿)和挪威的Norwegian Mother and Child Cohort Study(MoBa, 7.2万个家庭,包括孕妇及其孩子和丈夫共27.0万人)两个出生队列。英国学者近期又提出一项更为艰巨的计划,要随访8万名婴儿,实现“从摇篮到坟墓”的终生观察,特别要收集母亲妊娠期和婴儿出生第一年内的详细信息,储存各种组织样本,如尿液、血液、粪便、胎盘组织等^[3]。

随着生命科学与现代信息学技术的迅猛发展,近十余年来建立的人群队列都注意采集和长期保存队列成员的生物学样本。截至目前,国外人群队列中生物学样本库规模较大的是英国的UKB(49.8万份血液和尿液)和欧洲10国的EPIC(38.8万份血液),其次是挪威的MoBa(34.0万份血液)和美国的

Cancer Prevention Study- II Nutrition Cohort-3 (CPS-3, 30.4万份血液)。今后,这些人群队列通过整合基因组学、表观组学、蛋白组学、代谢组学等多个水平上的生物标志物,结合传统流行病学宏观研究的暴露组学,可以更好地理解疾病发生、发展的生物学机制。这就是传统流行病学与现代高通量组学技术结合后的学科发展方向,即系统流行病学(systems epidemiology)^[4]。

在宏观暴露组学的评价上,传统的人群研究受技术和成本的限制,多依靠调查对象自报,能够获得的暴露信息粗略不精准。近年来,智能手机、移动互联网、可穿戴健康设备的快速发展,为医学研究中个体信息的收集提供了前所未有的契机,可以帮助收集个体运动、睡眠、生理指标、社交活动、环境暴露等更为详尽、精确的数据。相比传统的流行病学调查研究,这些数据的获取与整合极大的丰富了可研究的暴露和结局内容,降低了数据采集成本,提高了研究效率。英国的UKB已经开始利用健身腕带对队列成员的每日体力活动、静坐行为和睡眠进行监测,目前已经采集到800万小时的行为数据,数据采集工作仍在进行中。美国国立卫生研究院近年来开始支持针对移动终端、可穿戴健康设备的研发、评价和应用研究,并计划在即将启动的美国百万人群队列研究中,应用这些设备和配套的应用程序采集队列成员的行为和健康信息。

当然,更大规模的人群队列建立时间不久,仍需要一段时间的抚育期。为了满足短期研究的需要,近年来,国外学者利用规模虽然较小、但是已经随访了较长时间的人群队列组建研究协作组(consortium),将不同人群队列的数据汇集,增加统计学效力,加强研究结果的外部真实性。

2. 国内进展:我国的前瞻性人群队列研究起步较晚,数量不多,研究分散,规模多数较小,且缺乏长期、稳定支持,随访时间短。20世纪八九十年代起陆续开展了一些前瞻性队列研究,其中具有一定规模的人群队列有:①重点关注肿瘤:上海纺织女工乳腺癌自检试验队列、林县一般人群试验队列、浙江省饮水与结肠直肠癌研究队列等;②重点关注心脑血管疾病:第三次全国高血压调查随访研究队列、中国多省市心血管病前瞻性队列研究、江西农村中老年人脑血管疾病队列等;③重点关注特殊职业暴露:苯暴露与肿瘤研究队列;④重点关注出生缺陷、儿童早期生长发育:基于“中美预防出生缺陷和残疾合作项目”的三省围产保健监测系统建立的出生队

列。目前,上述队列多数已停止随访。

近十余年来,我国陆续建立起若干前瞻性人群队列,规模更大、研究内容更为广泛。其中具有一定规模且随访一段时间的队列有:①重点关注主要慢性病:上海女性健康研究(7.5万)、上海男性健康研究(6.2万)、中国慢性病前瞻性研究(51.3万)、泰州纵向研究(20万)等;②重点关注老龄化相关问题:中国健康与养老追踪调查(1.7万);③重点关注生命早期环境暴露与出生缺陷、儿童发育和疾病:中国安徽出生队列(1.3万)、湖北同济出生队列(2.4万)等。总的来说,基于中国人群开展的前瞻性研究仍然较少,现阶段随访时间仍然较短,可产出的高质量人群病因学证据仍然有限,大多数国人相关的病因学证据仍然停留在病例对照研究、病例系列研究和横断面研究等因果证据强度较低的设计。近年来,在国际有影响力的学术期刊上发表的人群流行病学研究论文多数是基于前瞻性人群队列;而我国相当一部分研究工作和科研经费的投入仍然处于低水平重复,直接限制了我国的学科发展和研究成果的国际影响力。

3. 中国慢性病前瞻性研究:2004年,中国疾病预防控制中心(项目Ⅱ期改为中国医学科学院)与英国牛津大学合作启动了一项大型的中国慢性病前瞻性研究项目(China Kadoorie Biobank, CKB)^[5-7]。项目目标为调查主要慢性病及相关危险因素状况,通过长期随访,探讨环境、个体生活方式、体格和生化指标、遗传等众多因素对复杂慢性病或症状的发生、发展的影响。

项目现场包括农村和城市各5个地区。2004—2008年完成募集和基线调查,年龄在30~79岁、具有知情同意书和完整基线调查数据的调查对象共512 891人。项目于2008年和2014年对约5%的队列成员各完成一次重复调查,了解结局影响因素的长期变化趋势。针对全部队列成员的长期结局随访监测工作在队列成员参加基线调查后6个月开始^[8];截至2015年,已经累积观察400余万人年。除了利用各地卫生部门现有的常规疾病死亡和发病报告体系及委托社区工作人员主动接触外,项目还通过与医保数据库链接,系统了解特定随访期间内发生的所有住院事件。医保系统信息收集已经成为补充常规发病、死亡监测的重要且有效的途径。为了保证长期结局事件监测工作的质量,项目还陆续启动了主要慢性病病种发病事件的病例复核工作,通过收集病历资料,判断报告和诊断的准确性,了解病例的

真实性。

作为世界上生物学样本存储规模居于领先地位的CKB项目样本库,存储着项目51万余人的基线调查和两次重复调查的全部样本,存储规模达到58.8万份。项目样本库质量安全管理体系统日益完善。建库以来,样本库严格按照样本储存空间、温度、电源、网络、报警、监控等各项标准进行基础部署,实时对每台样本存储设备的温度变化进行记录和监控,建立突发情况应急处理预案,保证样本存储安全始终处于合理可控范围,从而实现样本安全存储和质量管控。项目实现了样本存储管理的全程电子化,通过样本采集、分装、转运跟踪、接收、定位、储存等一系列电子化程序和软件系统对项目所有样本实施全方位监管,从而最大限度减少人工操作产生错误的可能性。

CKB项目具有50余万人的超大规模,建立有存储量世界领先的生物样本库,覆盖我国东北、西北、华东、华南和西南具有不同经济发展水平、社会文化背景以及暴露谱和疾病谱的城乡地区,与当地疾病和死亡监测系统有机整合,极大的提高了疾病监测能力。CKB项目为我国转型期社会人群健康状况的发展和变化、疾病谱的改变及影响因素研究提供了宝贵的人群现场;如果可以得到足够的经费支持、长期维持队列,随着随访时间的延续,必将成为我国人群复杂疾病机制及防治研究不可或缺的重要基础性资源,为制定重大慢性病防治策略和指南提供本土高质量的病因学证据。

参 考 文 献

- [1] Wang JH. Epidemiology (vol. 1) [M]. 3rd ed. Beijing: People's Medical Publishing House, 2015. (in Chinese)
王建华. 流行病学(第一卷) [M]. 3版. 北京:人民卫生出版社, 2015.
- [2] Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies [J]. Nat Rev Genet, 2006, 7(10): 812-820.
- [3] Pearson H. Wanted: 80 000 British babies [J]. Nature, 2015, 518 (7540): 463-464.
- [4] Hu FB. Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology [J]. Clin Chem, 2011, 57 (9): 1224-1226.
- [5] Chen ZM, Lee L, Chen JS, et al. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC) [J]. Int J Epidemiol, 2005, 34(6): 1243-1249.
- [6] Chen ZM, Chen JS, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up [J]. Int J Epidemiol, 2011, 40 (6): 1652-1666.
- [7] Li LM, Lv J, Guo Y, et al. The China Kadoorie Biobank: related methodology and baseline characteristics of the participants [J]. Chin J Epidemiol, 2012, 33(3): 249-255. (in Chinese)
李立明,吕筠,郭彧,等. 中国慢性病前瞻性研究:研究方法和调查对象的基线特征 [J]. 中华流行病学杂志, 2012, 33(3): 249-255.
- [8] Li LM. Follow-up and surveillance techniques appropriate for large prospective cohort study [M]. Beijing: Peking Union Medical College Press, 2015. (in Chinese)
李立明. 大型人群队列随访监测适宜技术 [M]. 北京:中国协和医科大学出版社, 2015.

(收稿日期:2015-09-02)

(本文编辑:王岚)