

· 大型队列项目 ·

京津冀自然人群队列研究的理念与实践

单广良

中国医学科学院基础医学研究所/北京协和医学院基础学院流行病与卫生统计学系,
北京 100005

通信作者:单广良, Email: guangliang_shan@163.com

【摘要】 队列研究是探索病因和疾病防治知识发现的最有效工具之一。队列研究在我国已有 60 余年的工作基础,尤其自 2016 年科技部首次对队列研究专门立项以来,我国队列研究的种类和数量急剧增加。现代队列研究需要在经典的队列研究设计原则基础上,充分应用现代多学科资源和技术方法,以使暴露与结局关联分析和病因推断更加系统和精准。本文介绍了京津冀自然人群队列研究设计的理念、研究进展和挑战,以及在实施中应对的关键问题,可为国内队列研究及其随访机制建设提供参考。

【关键词】 京津冀; 自然人群; 队列研究

基金项目: 国家重点研发计划“精准医学”研究重点专项(2016YFC0900600, 2016YFC0900601); 中国医学科学院医学与健康科技创新工程(2019-I2M-2-007)

Principles and practice on cohort study of general population in Beijing, Tianjin and Hebei province

Shan Guangliang

Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China

Corresponding author: Shan Guangliang, Email: guangliang_shan@163.com

【Abstract】 Cohort study has emerged as an essential tool for gaining knowledge about risk factors and prevention of diseases. The studies related to cohort population have been progressing more than 60 years in our country, while different types and numbers in cohort studies have increased dramatically with the key programs special for the cohort study were approved firstly by the Ministry of Science and Technology of China since 2016. With typical study design, cohort study is currently needed to use the latest multidisciplinary resources and methods in order to be more systematic and accurate on association analysis and causal inference of diseases. For the cohort study of general population in Beijing, Tianjin and Hebei province, the core principles and concepts of study design, progress and the key issues targeting implementation were introduced. In addition, opportunities, bottlenecks and solutions facing us were also explored.

【Key words】 Beijing, Tianjin, Hebei; General population; Cohort study

Fund programs: National Key Research and Development Project of Precision Medicine, (2016YFC0900600, 2016YFC0900601); Innovation Fund for Medical Sciences of Chinese Academy of Medical Sciences (2019-I2M-2-007)

目前,全球存在的主要健康问题除了人口老龄化、慢性病、食品和药品安全等问题外,还面临新发

和突发传染病带来的威胁。鉴于队列研究在探索复合暴露与健康效应关系方面具有明显优势,队列

DOI: 10.3760/cma.j.cn112338-20200615-00843

收稿日期 2020-06-15 本文编辑 李银鸽

引用本文:单广良. 京津冀自然人群队列研究的理念与实践[J]. 中华流行病学杂志, 2021, 42(8): 1493-1497. DOI: 10.3760/cma.j.cn112338-20200615-00843.



研究已成为当今探索复杂病因的实用方法和有利工具之一^[1]。

在我国,队列研究已有 60 余年的工作基础和良好经验。例如,中国双生子队列研究、大庆糖尿病队列研究和阜外医院创建的我国第一个慢性病防治网络首钢模式等均具有明显的中国特色^[2-4],反映了我国处于不同经济发展阶段和疾病防治重点变化的需求,获得的很多原创性研究成果已陆续用于疾病危险因素干预和健康促进的实践,在国际上也产生了一定的影响力。自 2016 年我国首次在国家重点研发计划“精准医学”研究重点专项中发布队列研究专项以来,我国队列研究的种类和数量急剧增加,队列人群规模一次性超过百万。立项的队列研究大致分为自然人群队列、临床队列和专病队列 3 类,同时为这些立项的队列研究配置了标准示范队列研究、数据共享和基因组研究平台等支撑性项目。

目前,随着科技的进步和医学水平的提高,现代队列研究已经将暴露组学、基因组学、代谢组学、计算机、大数据挖掘和人工智能等技术应用到暴露的测量、结局事件发现^[5]以及暴露与疾病的关联分析,使得病因学探讨向着复合暴露与共病等多元化探索方向延伸,不仅扩大了病因发现的多学科融合视野,也在实践着服务于疾病精准预防、诊断和治疗的导向^[6]。在当今的精准医学和大数据时代,队列研究已成为流行病学研究的主旋律之一^[7]。

一、研究理念

1. 队列人群覆盖全年龄人口:研究设计时首先须明确自然人群的界定。人类学、社会学和医学界给出的“自然人群”定义都带有明显的各自学科和专业的局限性^[8]。直到目前尚未见到一个被同行广泛认可的“自然人群”定义。因此,在研究设计之初,考虑到选定人群的人口学、社会经济状况和地域等特点符合京津冀区域的自然特征,故整合了分布于京津冀三地、已有良好研究基础的优质队列资源组成 10 万人群队列。为了使所选队列人群可以反映全生命期中各个阶段的健康与疾病特征,使整合的多个队列人群的年龄段覆盖全年龄人口,并且观察起点不是从出生开始,而是移到从结婚登记开始。每个队列人群处于全生命期的不同阶段,整合队列由 4 部分人群队列组成:①妇幼和儿童青少年队列;②体检人群队列;③职业人群队列;④社区成人队列。

2. 全链条部署一体化实施:立项前京津冀三地

已有的队列研究零散、规模较小、技术标准不统一,研究内容多集中于某个专病,使得各研究间的数据不能简单地整合,难以实现共享和有效利用。在组建京津冀自然人群队列时,本研究特别遴选了具有良好的前期随访工作基础、当地政府可望给予长期支持的优质队列,保留特色,分散布局,但须按统一的研究标准统筹实施融为一体。

以往的一些专病队列或独立队列在病因学研究方面缺少多中心队列人群互比较的基础,可能在复杂病因的探索和推论验证上影响结果的外推。本研究利用多样化队列人群优势,既可以将不同队列人群合并进行整合研究,也可以将某一队列研究结果在其他人群队列中进行比较和验证,提高病因学推断的准确性和论证强度。

3. 建立宏观长远目标,夯实质量根基:依托国家项目建立的队列,虽然只有 4 年的项目执行期限,但是队列研究的价值在于长期随访,所以在京津冀自然人群队列研究设计之初,即考虑了宏观战略和超出项目年限的远期目标。长期随访也有助于识别和控制依时混杂和随访时间偏倚对因果推断准确性的影响^[9]。

本队列研究特别强调以质量为核心。质量控制涵盖了暴露测量与结局事件收集,随访和数据处理等全过程。以往人群健康调查的经验表明,队列研究质量需要有标准化操作规范,但信息化和自动化技术的应用更可为队列研究的科学化和规范化起到有效的支撑和助力作用,同时也能有效地减少人为误差和提高研究质量的保障。

二、一般情况介绍

本项目的主要研究目标:①整合京津冀区域 10 万人的自然人群队列,建立多样化队列长期随访和协作运行机制;②融合现代多学科先进技术,通过对不同队列人群生命全期中各阶段暴露与疾病关系的纵向观察,研究影响疾病发生的主要环境与遗传因素,为重大疾病的病因学和发病机制研究提供科学依据,促进研究成果向疾病早期预防和预测的转化应用;③建立京津冀区域自然人群队列研究数据库和生物样品库,无条件地实现科学数据和样品信息共享。

本项目由 5 个课题组成:①京津冀地区婴幼儿及儿童青少年生长发育队列研究。重点研究生命早期(围孕期)暴露对婴幼儿及儿童青少年生长发育的影响,以及成年人疾病的胎儿起源。本队列> 30 000 人。②京津冀地区生活社区自然人群慢性

病队列研究。分析长期复合暴露蓄积与常见慢性病及老年病的关联,尤其关注大气污染及空气质量改善对健康效应影响的动态变化趋势,建立慢性病风险预测系统。本队列>35 000 人。③京津冀地区体检中心自然人群健康趋势队列研究。应用每年定期的体检记录,分析多时点健康和疾病动态变化及其影响因素,评估暴露及生活方式改变对疾病发生及逆转的影响。本队列>30 000 人。④京津冀地区职业人群健康效应队列研究。重点研究京津冀三地钢铁和煤炭等支柱产业的职业暴露对健康的影响,评估职业环境治理对健康效应改善的影响。本队列>15 000 人。⑤京津冀地区自然人群队列研究协作网络及信息共享平台建设。由项目牵头单位负责,主要负责研究规划制定,组织和协调京津冀三地自然人群队列一体化构建及长效运行机制建设,为其他 4 个课题提供方法学、数据标准化、体检和化验参比体系、数据和样品信息共享平台等。

研究的主要内容包括问卷调查、体格检查、血常规检验、血生化检验和生物标志物筛选等实验室研究。问卷调查严格采取面对面询问方式,严格限制问卷条目数量,避免贪多求全和获取不准确信息。核心问卷内容包括个人一般特征、生活方式、疾病史和疾病家族史等 65 项指标。各队列可根据其人群年龄和职业等特点适当增加少量特色指标。体格检查包括身高、体重、身体成分、血压、握力、骨密度、心电图和肺功能等 31 项指标。血常规检验 22 项指标。血液生化检验包括 FPG、血脂、肝功能、肾功能、血清酶和免疫学指标等 23 项指标。

三、研究进展

1. 分类随访机制和提高随访依从性的措施:队列研究最重要的核心是一个支点和两大要素。一个支点是随访,两大要素是暴露和结局事件。如果失访严重,将无法全面研判暴露与结局事件的关联,队列研究也将失去真正的价值和意义,所以,我们特别强调随访是决定队列研究成败的关键。本项目随访方法首选面对面重复调查。随访问卷的内容和体检指标在与基线内容保持一致的基础上,增加一些可以动态评估和测量的暴露或结局事件信息。

京津冀妇幼队列、儿童青少年队列、体检队列和职业队列人群有定期体检要求,其随访是与定期体检相结合,所以目前的随访率均高于 90%。然而,由于社区队列人群主要来源于生活社区,退休和无固定职业人员相对较多,故在指定的随访期间内全部召回进行面对面随访的难度极大,多个社区

队列随访率均低于 90%。为避免社区队列可能出现严重失访的局面,我们采取电话随访、微信群推送和结局事件被动监测等多种方法作为补充手段来提高随访率。

随访问隔时间则因不同队列人群的年龄、职业特点和生活状态特征而异。妇幼队列按孕期定期保健管理对接随访时间;儿童青少年队列随访与学校安排的学生体质和健康检查对接;体检队列随访与其年度体检时间对接;社区队列随访以社区卫生服务中心为依托选择在温暖季节进行,原则上要求 40 岁以下人群至少每 3 年随访 1 次,40 岁以上人群至少 1~2 年随访一次。

2. 多学科合作和整合研究:随着科学技术的不断进步,新生的医学学科越来越多,专业划分也越来越细,由此也使健康到疾病的系统化研究被割裂。为了弥补任何单一学科研究的短板,本项目汇集了来自基础、临床和预防医学等多学科的队伍,应用暴露组学、基因组学、表型组学和代谢组学方法和技术,从问卷调查、体检、数据录入、样品存储、结果查询到随访和结局事件发现,融入计算机和数据库等技术,以多学科视角和整体观进行病因学探索。例如,以往研究生命早期暴露对儿童生长发育的影响多是从孕期开始观察,但很多夫妻因在备孕期改变了生活方式等暴露状态,所以对研究结局可能会产生一定影响。本研究从结婚登记体检开始收集暴露因素等信息,以期研究父母早期暴露蓄积对子代健康的影响,随后与儿童青少年队列长期随访相衔接,则可系统地研究生命早期健康状况对成年期疾病风险的影响,由此构成一个生命全期的整合研究体系。

3. 研究全程自动化和信息化建设:为了做好调研和随访工作的标准化和质量管理,最大限度地控制人为误差,本项目从研究对象登记、数据录入到体检结果查询等各个环节均研发了对应的信息化系统。例如,问卷扫描自动转换数据库系统,不仅可以将纸质问卷电子化保存,而且明显地减少了即使是双人双录入也难以发现和避免的人为误差。“京津冀健康”APP 用于研究对象免费查询各次体检结果,真正体现科技惠民。在身高、体重、身体成分、血压、骨密度、心电图和肺功能测量等所有体检环节实现了操作自动化和体检数据批量导出功能。

4. 边推进边提高:为了做好项目实施的全程管理和提升研究质量,在项目正式实施前,在雄安队列现场召开了项目启动会,所有课题负责人及骨干

参加,同时在现场观摩标准化操作和质控流程。在完成了全部基线调查后,召开项目推进会暨资源共享协议签署会,明确数据提供方和使用方的责任、利益分配和使用权限。在总结研究经验时,尤其针对发现的问题及时调整布局,预见随访困难并做好应对准备。例如,在基线调查总结中发现个别队列上报的血液样品数量少于基线调查人数,故严格规定各队列实际完成样本数量以同时有问卷基线调查、体检记录并保存血液样品人数为准。

在项目实施第三年,以“继往开来,任重道远”为主题,召开了京津冀队列随访推进会暨“论何观清先生与现场流行病学调查”学术交流会,邀请国内生物样品管理和队列研究专家介绍队列研究成果和经验,同时营造项目文化和学术氛围,增强项目团队凝聚力和发展后劲。在随访进程过半遇到新冠肺炎疫情期间,多数队列随访工作全面停滞,项目组则以视频会议方式再次举办项目随访推进会,敦促随访进度,为课题参加人员提升信心,努力把疫情造成的项目延误和损失降到最低,也为项目验收做好准备。

5. 国内外队列研究经验分享:在队列研究进程中,东北区域自然人群队列研究和西南区域自然人群队列研究项目团队来京津冀队列研究现场交流操作方法和数据共享标准,观摩现场调查技术和流程,分享研究标准、质量控制和随访等经验。

在研究初始阶段,项目组邀请荷兰格罗宁根大学生命全期队列和生物样品库(the Lifeline Cohort and Biobank)项目负责人介绍研究进展和成果,其通过队列研究建立的个体疾病的遗传风险评估系统非常值得借鉴。在项目进展中期,项目组赴日本东北大学交流队列研究技术,参观生物样品库建设的软硬件环境,学习到三代家庭队列的建设和随访路径,以及数据和生物样品信息使用的权限分配和共享方式。

四、挑战

1. 控制失访:失访率高低主要取决于研究对象的依从性,为此我们采取了一些可行有效的措施增加队列人群对项目的信任度、参与感和获得感。在每次现场调查和体检中,为参加者提供健康咨询,当日体检结束时返给一份包含血压、心电图和骨密度等检查的常规体检报告单。血液生化检查和其他所有检查结果将通过手机“京津冀健康”APP 免费查询和打印。此外,根据不同队列人群特点拟逐步采取以下措施提高随访成功率:①制定长期的队

列随访方案,设置固定随访专员;②建立一个京津冀健康公众号或多个队列随访微信群,推送主动健康信息和常见病防治常识,增加队列建设者凝聚力,维护队列稳定性和长期性;③随访前针对不同年龄和职业队列人群特点定向进行不同形式和灵活多样的宣教动员。

2. 结局事件追踪和确认:通过随访可同时获得暴露和结局双重变化的信息,但是单纯地依靠定时定点的随访来获得结局事件是远远不够的。为此我们拟建一个随访和结局事件发现的一体化机制,将不时地抓取结局信息与定期随访相结合,通过互利双赢的原则汇集多部门的健康相关数据。监测内容包括:暴露因素变化的定性和定量测量、健康状况变化和发病结局、死因监测以及研究对象迁移和失访信息。

就目前京津冀可用的结局事件发现途径而言,急需与当地医疗卫生和社会保险等政府部门多方协商,建立多部门合作机制,制定限定队列人群范围的、设定严格使用权限和隐私保护机制的安全可行的结局事件获得方法。京津冀三地虽然在区域上交织接壤,但由于在队列人群的人口结构和当地行政管理等方面仍显现一定差异,所以在随访和结局事件发现机制上,仍需采取分类和多样化原则,对不同区域队列将通过主动和被动两种方式获取结局事件信息。见图 1。虽然被动收集结局事件的某些机制目前尚未实现,但是作为一种未来高效的结局事件获得来源,可望成为长久性队列研究的结局事件发现的重要手段和不可或缺的方法。

3. 数据标准化及有效共享:本项目收集的数据和样品信息面临的突出问题是数据的标准化、共享和有效利用。因为本项目主要是在原有队列基础上延续的,所涉及的队列人群特征多样化,暴露和结局测量的标准和方法等也不尽相同,实施标准和操作方法不统一,使得整合后的各队列仍相对独立,尚不能直接进行数据汇交和共享,导致队列共建的成效和研究效率受限。例如在数据标准化过程中,很难使延续的队列改变以往沿用的数据标准而使用统一规定的标准,所以,在摸清和归纳了各队列数据标准的异同后,在不同队列的异构化和分布式数据基础上,拟在数据汇交端口设计一个针对不同来源和标准数据转换模块,使标准不同的数据转换为统一规定标准和格式的通讯数据。

数据的共享和有效利用需要解决两个层面的问题,即不同课题之间,以及同一课题内部不同参

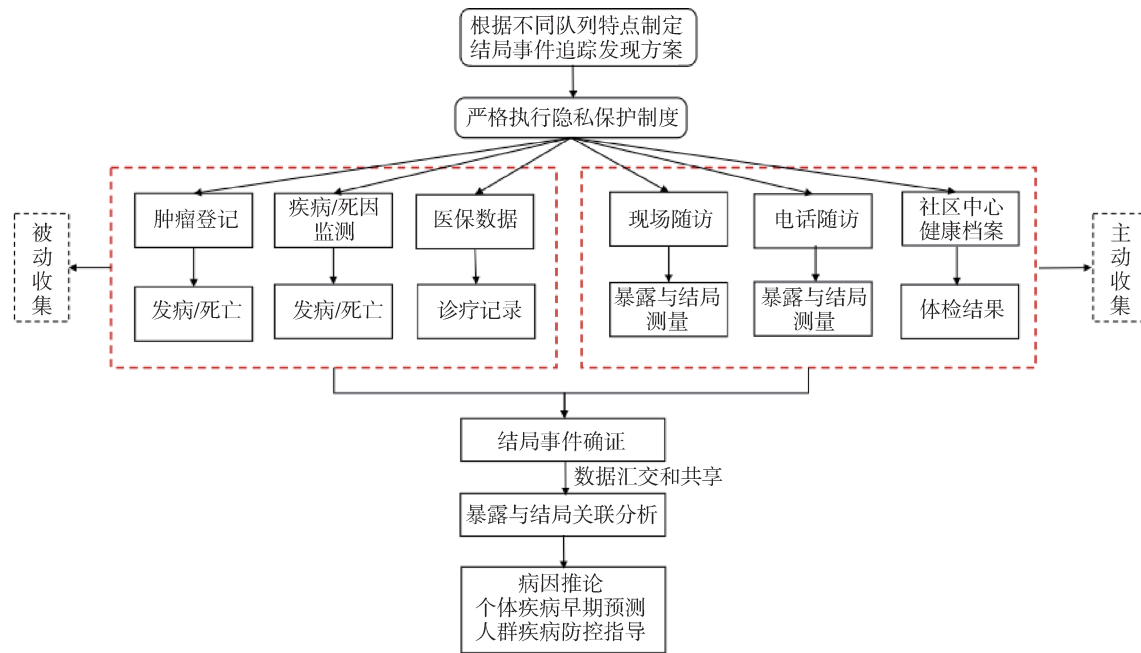


图1 京津冀自然人群队列研究结局事件发现机制

加单位之间的数据利用和共享问题。为此,本项目本着“共生、共建、共享和共赢”原则制定了《京津冀区域自然人群队列研究数据和样品共享协议》,明晰甲乙丙三方的责任、权限和使用要求。甲方(资源管理方)是项目负责单位;乙方是数据资源提交方,包括所有项目参加单位;丙方是数据使用方,目前仅限项目内任一参加单位申请使用。数据和生物样品分布存储在各课题负责单位,但必须遵守项目内无条件共享机制,兼顾各自利益分配但比例清晰。

五、结论

本项目从队列研究目标设定之始到实施过程中,不断遇到操作的可行性、随访的依从性、数据的标准化和各队列研究质量参差不齐等各种各样艰难和复杂的问题,我们也随之不断地梳理核心问题,并找出有效办法以解决阻碍研究顺利实施的关键环节。我们清楚地认识到,不断地解决关键技术和化解不和谐因素,不仅可以推动研究进程和改进质量,还可明显地提升团队合作共赢的信心及增加项目可持续发展的后劲。

京津冀区域自然人群队列研究的目标不是固定不变的,在队列研究进程中,将不断地融入新的技术和方法。随着项目的进展和国家健康需求的发展,将不断补充和完善预定的研究目标,提出新的探索方向,边实施边完善,并注重将队列研究成果转化为人群疾病防治的适宜技术。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] 李立明,吕筠.大型前瞻性人群队列研究进展[J].中华流行病学杂志,2015,36(11):1187-1189. DOI:10.3760/cma.j.issn.0254-6450.2015.11.001.
- [2] Li LM, Lv J. Large prospective cohort studies: a review and update[J]. Chin J Epidemiol, 2015, 36(11): 1187-1189. DOI:10.3760/cma.j.issn.0254-6450.2015.11.001.
- [3] 高文静,李立明.中国双生子队列研究进展[J].中华流行病学杂志,2017,38(6):828-831. DOI:10.3760/cma.j.issn.0254-6450.2017.06.027.
- [4] Gao WJ, Li LM. The Chinese national twin cohort: an update[J]. Chin J Epidemiol, 2017, 38(6): 828-831. DOI:10.3760/cma.j.issn.0254-6450.2017.06.027.
- [5] Gong QH, Zhang P, Wang JP, et al. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing diabetes prevention outcome study[J]. Lancet Diabetes Endocrinol, 2019, 7(6): 452-461. DOI: 10.1016/S2213-8587(19)30093-2.
- [6] 岳寒,顾东风,吴锡桂,等.首都钢铁公司5137名男工心肌梗死发病危险因素的研究[J].中华预防医学杂志,2004,38(1):43-46. DOI:10.3760/j.issn:0253-9624.2004.01.013.
- [7] Yue H, Gu DF, Wu XG, et al. A 20-year prospective study on risk factors for myocardial infarction of 5 137 men in Capital Steel and Iron Company[J]. Chin J Prev Med, 2004, 38(1):43-46. DOI:10.3760/j.issn:0253-9624.2004.01.013.
- [8] 白志鹏,陈莉,韩斌.暴露组学的概念与应用[J].环境与健康杂志,2015,32(1):1-9. DOI:10.16241/j.cnki.1001-5914.2015.01.001.
- [9] Bai ZP, Chen L, Han B. Exposure and expositives: from concepts to application[J]. J Environ Health, 2015, 32(1): 1-9. DOI:10.16241/j.cnki.1001-5914.2015.01.001.
- [10] 王笑峰,金力.大型人群队列研究[J].中国科学:生命科学,2016,46(4):406-412. DOI:10.1360/N052016-00104.
- [11] Wang XF, Jin L. Large population-based cohort studies[J]. Sci Sin Vitae, 2016, 46(4): 406-412. DOI: 10.1360/N052016-00104.
- [12] Khoury MJ. Planning for the future of epidemiology in the era of big data and precision medicine[J]. Am J Epidemiol, 2015, 182(12):977-979. DOI:10.1093/aje/kwv228.
- [13] 苟天来,左停.自然人群的规模及其形成的一些属性——来自皖南山区村落人际关系的社交网络分析[J].中国农业大学学报:社会科学版,2009,26(1):101-107. DOI:10.13240/j.cnki.caujsse.2009.01.017.
- [14] Gou TL, Zuo T. Scale of natural group and some characteristics of one's formalization: a social network analysis in a village community in mountainous western Anhui province[J]. J China Agric Univ: Soc Sci Ed, 2009, 26(1):101-107. DOI:10.13240/j.cnki.caujsse.2009.01.017.
- [15] 刘慧鑫,苏迎盈,李峰,等.队列研究中的依时混杂偏倚和随访时间偏倚[J].中华流行病学杂志,2014,35(10):1169-1171. DOI:10.3760/cma.j.issn.0254-6450.2014.10.021.
- [16] Liu HX, Su YY, Li Z, et al. Time-dependent confounding bias and follow-up duration time bias in cohort studies[J]. Chin J Epidemiol, 2014, 35(10):1169-1171. DOI:10.3760/cma.j.issn.0254-6450.2014.10.021.