

## · 中国大型人群队列研究 ·

## 京津冀自然人群队列研究的经验与展望

何慧婧<sup>1</sup> 潘利<sup>1</sup> 张玲<sup>2</sup> 袁聚祥<sup>3</sup> 刘功姝<sup>4</sup> 张敏英<sup>5</sup> 胡耀达<sup>1</sup> 涂吉<sup>1</sup> 单广良<sup>1</sup>

<sup>1</sup>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学与卫生统计学系, 北京 100005; <sup>2</sup>首都医科大学公共卫生学院流行病学与卫生统计学系, 北京 100069; <sup>3</sup>华北理工大学公共卫生学院预防医学系, 唐山 063210; <sup>4</sup>天津市妇女儿童保健中心科研项目组, 天津 300072; <sup>5</sup>南开大学医学院公共卫生学系, 天津 300071

通信作者: 单广良, Email: guangliang\_shan@163.com

**【摘要】** 京津冀自然人群队列研究是一项大规模前瞻性队列研究, 2017年在北京市、天津市和河北省开展基线调查, 覆盖了从生命早期直至老年的 114 850 名调查对象。截至 2021 年 12 月, 106 031 人已完成至少 1 次随访, 总体随访率为 92.3%。本队列结合三地环境和健康问题特点, 重点围绕儿童青少年生长发育、心血管代谢疾病及其危险因素、老龄化及共病、空气污染的健康效应等主题进行了系统研究。京津冀自然人群队列中全生命期多样化人群的长期随访将为我国健康与疾病研究提供独特和宝贵的科研资源以及丰富的生物样品。本文描述了京津冀自然人群队列的产生背景和战略意义、研究现状和成果、不足和挑战以及未来规划和展望等, 以期为国内外学者开展相关研究提供参考。

**【关键词】** 队列研究; 自然人群

**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, 2020-I2M-2-009)

**General population cohort study in Beijing-Tianjin-Hebei area: past and future**He Huijing<sup>1</sup>, Pan Li<sup>1</sup>, Zhang Ling<sup>2</sup>, Yuan Juxiang<sup>3</sup>, Liu Gongshu<sup>4</sup>, Zhang Minying<sup>5</sup>, Hu Yaoda<sup>1</sup>, Tu Ji<sup>1</sup>, Shan Guangliang<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China; <sup>2</sup>Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing 100069, China; <sup>3</sup>Department of Preventive Medicine, School of Public Health, North China University of Science and Technology, Tangshan 063210, China; <sup>4</sup>Research Program Team, Tianjin Women and Children Health Center, Tianjin 300072, China; <sup>5</sup>Department of Public Health, School of Medicine, Nankai University, Tianjin 300071, China

Corresponding author: Shan Guangliang, Email: guangliang\_shan@163.com

**【Abstract】** The general population cohort study in Beijing-Tianjin-Hebei area is a large-scale prospective longitudinal study conducted since 2017, which covers over 114 850 diverse populations from early life to senior age. Up to December 2021, 106 031 people had completed at least one follow-up, with an overall follow-up rate of 92.3%. Considering of the characteristics of the environmental and health problems in Beijing-Tianjin-Hebei area, in this cohort study we have focused on health problems of children and adolescents' growth and development, cardiometabolic diseases and their risk factors, aging and comorbidity, health hazards caused by air pollution. The long-term follow up of the general population cohort study in Beijing-Tianjin-Hebei area will

DOI: 10.3760/cma.j.cn112338-20220930-00834

收稿日期 2022-09-30 本文编辑 万玉立

引用格式: 何慧婧, 潘利, 张玲, 等. 京津冀自然人群队列研究的经验与展望[J]. 中华流行病学杂志, 2023, 44(1): 7-13.

DOI: 10.3760/cma.j.cn112338-20220930-00834.

He HJ, Pan L, Zhang L, et al. General population cohort study in Beijing-Tianjin-Hebei area: past and future[J]. Chin J Epidemiol, 2023, 44(1):7-13. DOI: 10.3760/cma.j.cn112338-20220930-00834.



contribute to establishing unique and invaluable data and biobank resource for scientific research. This paper aims to comprehensively describe the background, significance, current status and outcomes, limitations and challenges, and future plan and development of general population cohort study in Beijing-Tianjin-Hebei area, thus to provide reference for professionals both at home and abroad to carry out related research.

**【Key words】** Cohort study; General population

**Fund programs:** "Precision Medicine Research" Key Project, National Key Research and Development Program of China (2016YFC0900600, 2016YFC0900601); Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2019-I2M-2-007, 2020-I2M-2-009)

队列研究由于能够同时观察多个暴露因素及其交互作用对健康的影响,且通过长期随访可观察危险因素对健康的长期效应,近年来已成为病因探索、了解疾病发生发展规律的重要研究方法并取得显著成果。京津冀自然人群队列研究是一项大规模前瞻性队列研究,覆盖了从生命早期直至老年的 114 850 名的全生命期多样化人群。在国家重点研发计划“精准医学研究”重点专项支持下,本队列于 2016 年立项,2017 年开始基线调查,截至 2021 年 12 月,106 031 名已完成至少 1 次随访,总体随访率为 92.3%。项目组结合京津冀地区健康问题特点,主要围绕老龄化和城镇化所带来的一系列健康问题,如心血管代谢疾病及其危险因素、老龄化及共病、空气污染的健康效应以及儿童青少年生长发育和成人慢性非传染性疾病(慢性病)的早期防治等内容进行了研究。本文对京津冀自然人群队列研究的背景和理念、现状和成果、存在的不足和挑战以及未来规划和展望等进行系统阐述,以期为国内外学者开展相关研究提供参考。

### 一、京津冀自然人群队列研究的背景

1. 大型队列研究在病因探索中的优势:流行病学有多种研究设计类型,多数慢性病由于发病机制复杂、潜伏期较长等,使传统、小规模队列研究在病因探索中存在较大局限性<sup>[1]</sup>。大规模、前瞻性队列研究由于其大样本、长期随访、暴露因素收集较全、观察结局多样等优点,在复杂疾病病因探索中具有独特优势。目前国内外已开展多项大规模人群队列,例如英国生物银行项目,中国慢性病前瞻性队列研究(China Kadoorie Biobank, CKB)项目,中国国家出生队列等<sup>[2-4]</sup>。2016 年,国家重点研发计划“精准医学研究”重点专项项目正式启动,京津冀自然人群队列与西北、东北、华东、华中、西南和华南等区域的自然人群队列,共同构成百万级前瞻性自然人群队列。自然人群队列相对于专病队列(如乳腺癌队列、罕见病队列等),可考察生活环境暴露、膳

食因素、生活习惯和遗传因素等多种健康问题的影响及其交互作用,有助于开展从健康到疾病的全过程风险管理,其长期随访还可获得疾病的发病趋势等数据。

2. 京津冀地区健康问题的特点与队列研究设计理念:

(1) 京津冀地区健康问题的特点:近几十年来,随着社会经济的快速发展,生活方式的急剧改变以及城镇化和老龄化进程的加速,我国心血管代谢疾病的发病率呈快速上升趋势,慢性病早期风险管理需求迫切<sup>[5-7]</sup>。京津冀地区的社会环境因素使其健康研究具有独特之处:首先,京津冀地区是世界级城市规模群,人口数占全国总人口的 8.1%,该地区面临突出的快速城镇化和老龄化问题,慢性病防控面临严峻挑战<sup>[8]</sup>。结合不同生命时期人群多样化遗传背景和环境暴露信息,探索并早期识别影响健康的危险因素,可促进疾病早期预防和健康老龄化;其次,钢铁、煤炭和石油等重工业是京津冀地区最主要的支柱产业(占三地全部产业的 1/2),加之地理环境等因素的影响,使该地区成为我国大气污染最严重的地区之一<sup>[9]</sup>,亟需了解京津冀地区大气污染的特点和对不同人群差异化的健康效应。此外,作为我国环境治理的重点区域,还可从空气质量改善所引起的疾病归因风险改变,以及带来的其他健康效益等视角开展环境治理效果评估。

(2) 京津冀自然人群队列的设计理念及建设原则:一是队列覆盖全生命期多样化人群;二是在保持原有各队列特色基础上实现全链条部署和一体化实施;三是建立长期随访的机制和遵循质量第一的原则<sup>[10]</sup>。

健康与疾病发育起源学说提示了生命过程中的环境暴露因素对终身健康的影响,是慢性病生命早期起源和防治端口前移的重要理论依据<sup>[11]</sup>。因此,本队列纳入孕产妇人群,将遗传和环境暴露因素的观察提前,以探索生命早期遗传和环境风险对

健康的影响。此外,既往现场调查经验显示,从生活社区招募调查对象往往存在年轻男性较少、中老年女性较多的问题。为减少性别和年龄构成上的不平衡,在设计阶段,项目组增加了以体检中心为基础的调查人群招募,使最终获得的自然人群样本有较均衡的年龄和性别构成。

京津冀自然人群队列启动之前,该地区已有不少在建的队列,但存在研究零散、规模较小、技术标准不统一、研究内容多集中于某个专病等局限性,使得各队列间的数据不能有效整合,难以实现资源共享和利用<sup>[10]</sup>。项目组结合既往开展国民健康调查的实践经验和管理模式,将具备良好工作基础、技术手段成熟且当地政府支持的现有优质队列,按照统一标准和质控要求进行整合<sup>[12]</sup>。目前京津冀自然人群队列共包括 4 个亚队列,分别为婴幼儿及儿童青少年生长发育队列(母亲孕期至 18 岁前儿童青少年)、生活社区自然人群慢性病队列( $\geq 18$  岁)、体检中心自然人群健康趋势队列( $\geq 18$  岁)和职业人群健康效应队列(从事特定职业的在岗成年人群)。队列覆盖了从生命早期直至老年的全生命周期人群,在保持各队列原有特色的基础上,通过制定统一的数据标准化规范、遵循共同的质量控制规则等方式,实现了不同队列全链条管理和一体化实施,为后期建设和长期随访提供了保障。此外,队列建设强调以质量为核心,质控措施覆盖暴露测量、结局事件发现、随访管理和数据处理等全过程。

## 二、队列研究现状

1. 队列建设现状:京津冀自然人群队列由 4 类队列组成,不同队列所收集的资料及收集方式见表 1。4 类队列人群的性别和年龄构成见表 2。部分队列已将研究设计理念和基线特征等内容发表<sup>[13-14]</sup>。

(1) 婴幼儿及儿童青少年生长发育队列:该队列重点研究生命早期风险因素暴露对婴幼儿及儿童青少年生长发育的影响,以及成年期疾病的胎儿起源。共完成 32 712 人的基线调查,其中孕早期孕妇至 2 岁儿童组 5 920 人,3~5 岁组 3 952 人,6~8 岁组 7 312 人,9~11 岁组 3 999 人,12~14 岁组 5 397 人,15~18 岁组 6 132 人。孕产人群主要依托天津市辖社区卫生服务中心,对辖区内孕妇建立系统管理档案,对符合入组条件的孕妇进行告知和招募。之后将自愿入组者转诊到所属区级妇幼保健中心进行后续调查并将所有信息汇总至天津市妇幼保健系统。调查对象年龄为(30.3 $\pm$ 3.9)岁。主

要随访方式为面对面问卷调查和关联妇幼保健系统信息。分别在孕中期、孕晚期、产后 42 d、儿童 6 个月、1 岁和 2 岁时进行随访,随访率为 91.6%。学龄前儿童(3~5 岁)的招募主要集中在天津市,采用整群抽样方式,考虑经济发展水平,在天津市辖 11 个区共选择 42 家幼儿园,对小班儿童进行招募和基线调查。随访问隔为 1 年,分别于中班和大班期间进行随访。随访方式主要为面对面调查和体格检查,随访率为 95.6%。6~18 岁学龄儿童/青少年人群的招募采用多阶段整群抽样方法:首先,在北京市和河北省,根据城镇化和经济发展水平等因素选择区/县,然后在区/县中按照方便抽样的原则选择学校,共纳入 11 所小学,26 所中学。对学校不同年级的班级进行方便抽样,对抽到班级的全部学生进行调查。中小学学生的随访资料通过现场调查、关联学籍信息系统和学校卫生常规体检数据等方式收集,随访问隔为 2 年,随访率为 91.2%。

(2) 生活社区自然人群慢性病队列:该队列主要分析危险因素长期复合暴露与常见慢性病的关联,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。该队列主要在京津冀三座城市和农村的自然生活社区/村招募 $\geq 18$  岁调查对象。调查方式、调查内容和核心团队与项目组前期开展的国民健康调查工作模式一致,具备良好的工作基础和现场协作条件。采用分层整群抽样的原则,考虑城乡和经济水平等因素后进行调查地点的选择。共纳入 12 个区/县作为调查地点,再从每个区/县中选择社区/村,被抽到的社区/村中符合纳入标准的居民均可参加调查。已完成 35 660 人的基线调查,调查对象年龄为(51.59 $\pm$ 14.82)岁,44.18% 为男性。随访问隔为 2~3 年。通过重复测量、关联常规监测系统和电话随访相结合的方式随访,至 2021 年 12 月,已完成 1 次随访,随访率为 90.1%。

(3) 体检中心自然人群健康趋势队列:基于生活社区招募调查对象可能造成人口构成不平衡,而通过体检中心进行队列人群的招募,可以较好地平衡成年自然人群的性别和年龄比例。按照多阶段分层随机整群抽样的原则,该队列在京津冀三地共选择了 6 家体检中心,在每家体检中心根据体检者职业特点对体检单位进行分层随机抽样,最后对抽到的单位中符合条件的 $\geq 18$  岁体检者进行招募。该队列以面对面重复测量和关联每年定期体检记录等方式进行随访,分析多时点危险因素动态变化

表 1 京津冀自然人群队列信息收集方式、内容及数据标准化

队列名称	基线调查		随访调查			数据标准化途径
	调查内容	资料收集方式	调查内容	资料收集方式	随访间隔	
婴幼儿及儿童青少年生长发育队列	①孕期至子代(2岁)队列:孕早期至子代2岁的调查资料。其中孕产妇调查内容包括一般人口学信息;孕前营养、孕期增重、被动吸烟、体力活动等生活方式信息;个人及家族疾病史以及慢性代谢性疾病相关指标(血压、血糖、血脂等)。子代调查内容包括新生儿及幼儿的生长发育情况(身高、体重等)、母乳喂养、亲子活动、新生儿睡眠等营养和生活方式信息等	问卷调查、体格检查、实验室检测、妇幼保健系统数据	同基线调查	面对面重复测量:问卷调查、体格检查和实验室检测;关联妇幼保健系统数据	孕妇及其子代:孕中期、孕晚期、产后42 d、儿童6个月、1岁和2岁时进行随访	按照项目组数据汇交平台提供的数据模板和字段说明,统一变量的命名方式和编码规则
	②3~5岁队列:一般人口学信息;家庭成员疾病史;儿童饮食、运动、睡眠等生活方式情况;生长发育评估(身高、体重、血压、身体成分)和慢性病相关代谢指标情况等	问卷调查、体格检查和实验室检测	同基线调查	面对面重复测量:问卷调查、体格检查和实验室检测	3~5岁:随访间隔为1年,分别于幼儿园中班和大班期间进行随访;6~17岁:随访间隔为2年	将不同年龄段儿童青少年数据按照统一的变量命名方式和编码规则进行整理后,形成儿童青少年生长发育数据库
	③6~17岁队列:一般人口学信息;家庭成员疾病史;儿童饮食、运动、睡眠等生活方式信息;生长发育评估(身高、体重、血压、身体成分)和慢性病相关代谢指标情况等	问卷调查、体格检查和实验室检测	同基线调查	面对面重复测量:问卷调查;关联学籍信息系统和学校卫生常规体检数据		
生活社区自然人群慢性病队列	①社会人口学信息,包括性别、年龄、民族、文化程度、职业、收入等;②个人和家族疾病史及用药史,包括常见慢性病如高血压、糖尿病、心血管疾病和肿瘤等;③生活方式信息,如吸烟、饮酒、体力活动、体育锻炼等;④空气污染暴露相关信息,如室外活动时间、雾霾保护措施、空气净化器使用情况等;⑤心理健康相关信息,如压力、焦虑、抑郁等;⑥体格检查:包括身高、体重、身体成分、血压、心电图、肺功能、骨密度、握力等;⑦肝肾功能、血脂、空腹血糖等血液生化指标	问卷调查、体格检查和实验室检测	同基线调查	面对面重复测量:问卷调查、体格检查和实验室检测;电话随访:问卷调查	随访间隔为2~3年	按照项目组数据汇交平台提供的数据模板和字段说明,将生活社区自然人群慢性病队列和体检中心自然人群健康趋势队列的数据按照统一变量的命名方式和编码规则进行标准化处理,形成≥18岁自然人群队列数据库
体检中心自然人群健康趋势队列	①社会人口学信息,包括性别、年龄、民族、文化程度、职业、收入等;②个人和家族疾病史及用药史,包括常见慢性病如高血压、糖尿病、心血管疾病和肿瘤等;③生活方式信息,如吸烟、饮酒、体育锻炼、静坐时间、睡眠等;④空气污染暴露相关信息,如室外活动时间、雾霾保护措施、空气净化器使用情况等;⑤心理和躯体感受相关信息,如心理健康、疲劳、压力、社会适应等;⑥体格检查:包括身高、体重、血压等;⑦血液生化指标检测:包括FPG、肝肾功能、血脂和血常规等;⑧尿常规检测	问卷调查、体格检查和实验室检测	同基线调查	面对面重复测量:问卷调查、体格检查和实验室检测	随访间隔为1年	
职业人群健康效应队列	①社会人口学信息,包括性别、年龄、民族、文化程度、职业、收入等;②个人和家族疾病史,包括常见慢性病如高血压、糖尿病、高脂血症、脑血管疾病、心血管疾病等;③生活方式信息:吸烟、饮酒、饮茶、膳食和食盐摄入以及睡眠情况等;④职业暴露因素信息:工龄、工种、所在厂区、倒班情况及工作场所有害因素暴露情况;⑤体格检查:包括身高、体重、腰围、臀围、身体成分等、血压和心电图测量、骨密度检测等;⑥血液生化指标检测:包括FPG、肝肾功能、血脂和血常规等;⑦尿常规检测	问卷调查、体格检查和实验室检测	同基线调查	面对面重复测量:问卷调查、体格检查和实验室检测;电话随访:问卷调查	随访间隔为1~2年	按照项目组数据汇交平台提供的数据模板和字段说明,统一变量的命名方式和编码规则

对健康的影响。基线调查 31 310 人,调查对象年龄为(46.06±14.63)岁,51.27%为男性。随访间隔为1年,至2021年12月,已完成至少1次随访,随访率为96.5%。

(4)职业人群健康效应队列:钢铁、石油和煤炭等重工业是京津冀地区的重要产业构成,职业人群

健康和国民经济发展息息相关。因此,职业人群健康效应队列重点研究职业环境中的特殊暴露因素对职业人群健康的损害,并评估职业环境治理效果。该队列采用整群抽样方法,选取了河北省3家钢铁、煤炭和石油公司,对其中在岗并从事生产劳动的≥18岁职业工人进行招募。基线调查 15 168 人,

表 2 不同队列人群的性别和年龄构成

年龄组(岁)	男性		女性		合计	
	人数	构成比(%)	人数	构成比(%)	人数	构成比(%)
婴幼儿及儿童青少年生长发育队列						
0~	3 137	18.77	2 783	17.39	5 920	18.10
3~	2 071	12.39	1 881	11.76	3 952	12.08
6~	3 708	22.19	3 604	22.53	7 312	22.35
9~	2 070	12.38	1 929	12.06	3 999	12.22
12~	2 816	16.85	2 581	16.13	5 397	16.50
15~17	2 911	17.42	3 221	20.13	6 132	18.75
小计	16 713	100.00	15 999	100.00	32 712	100.00
生活社区自然人群慢性病队列						
18~	1 921	12.20	1 419	7.13	3 340	9.36
30~	2 842	18.04	2 623	13.17	5 465	15.33
40~	2 561	16.26	3 627	18.22	6 188	17.35
50~	3 518	22.33	4 859	24.41	8 377	23.49
60~	3 187	20.23	5 245	26.35	8 432	23.65
70~	1 724	10.94	2 134	10.72	3 858	10.82
小计	15 753	100.00	19 907	100.00	35 660	100.00
体检中心自然人群健康趋势队列						
18~	2 124	13.23	2 333	15.29	4 457	14.23
30~	4 082	25.43	4 262	27.93	8 344	26.65
40~	3 362	20.94	3 478	22.80	6 840	21.85
50~	3 348	20.86	2 426	15.90	5 774	18.44
60~	1 893	11.79	1 832	12.01	3 725	11.90
70~	1 244	7.75	926	6.07	2 170	6.93
小计	16 053	100.00	15 257	100.00	31 310	100.00
职业人群健康效应队列						
18~	1 219	9.18	119	6.32	1 338	8.82
30~	4 091	30.79	268	14.24	4 359	28.74
40~	4 746	35.72	1 402	74.50	6 148	40.53
50~	3 167	23.84	93	4.94	3 260	21.49
60~	63	0.47	0	0.00	63	0.42
小计	13 286	100.00	1 882	100.00	15 168	100.00
合计	61 805	100.00	53 045	100.00	114 850	100.00

调查对象年龄为(42.13±8.57)岁,其中87.59%为男性。随访问隔为1~2年,随访方式为面对面重复测量结合电话随访。截至2021年12月,已完成至少1次随访,随访率为91.3%。

## 2. 成果和社会效益:

(1) 队列成果:京津冀自然人群队列主要围绕京津冀地区老龄化和城镇化所带来的一系列健康危害,以及儿童青少年生长发育评价和疾病早期预防等主题进行了系列探索。研究结果以论文、专著和专利等形式发表;通过队列建设培养了一批具有实践经验和多学科背景的科研骨干和研究生;同时,在信息化管理和健康促进等方面也取得了良好的社会效益。

科学研究方面,针对京津冀地区突出的大气污染问题,探索了大气污染物中的多种组分,如PM<sub>2.5</sub>、PM<sub>10</sub>和O<sub>3</sub>等对心血管代谢疾病及其危险因素的健康效应<sup>[9,15-16]</sup>。例如,利用环境监测数据和交通道

路数据,结合个体水平健康相关信息,定量评估和比较了大气污染物中多种组分长期暴露对心电图各项指标的影响,识别出大气污染健康损害的高危人群<sup>[15]</sup>。再如,多数研究仅关注研究对象居住地的大气污染健康效应,而利用体检中心自然人群健康趋势队列中的单位地址信息,探讨了工作场所PM<sub>2.5</sub>暴露对无症状人群空腹血糖代谢异常的影响<sup>[9]</sup>,将慢性病风险识别和预防的端口前移。在慢性病危险因素研究方面,除了传统的心血管代谢因素,队列中还设计了心理健康评价内容,探讨了社区常见慢性病和心理健康,如抑郁、焦虑等之间的关联,体现了健康的综合照护理念<sup>[17]</sup>。煤炭、石油和钢铁工人等暴露于特殊的职业危害因素,如倒班作业、噪声环境等,对这些因素的健康风险进行定量分析,将有助于提出有针对性的职业保护策略和措施,减少职业有害暴露所致的健康损害。例如,基于职业队列的分析发现,倒班作业、噪声环境等

可能增加钢铁工人高同型半胱氨酸血症、血脂代谢异常、肾功能受损等疾病的患病风险<sup>[18-20]</sup>。此外,课题组基于妇幼人群队列,开展了生命早期健康风险识别等方面的探索,如孕期营养及不良生活方式对妊娠结局的影响,以及儿童青少年时期肥胖、高血压等慢性病的患病率、危险因素等<sup>[21-22]</sup>,为成人慢性病的早期防治提供了重要参考。

技术方法和应用层面,在队列实施过程中,项目组设计和研发了用于儿童生长发育和慢性病风险因素评价的应用软件,申请了有关中国儿童肥胖生物标志物筛选的发明专利,促进了科研成果的转化应用。此外,为促进信息化技术在队列建设中的应用,提升研究效率并优化质控措施,项目组研发了用于调查问卷扫描和健康检查结果反馈的应用软件,并比较和借鉴了国外相关研究中信息化技术的应用<sup>[23]</sup>。为使不同来源数据实现有效共享和利用,在梳理各队列数据标准的异同后,项目组设计并研发了数据汇交和共享平台,通过在数据汇交端口设计标准数据转换模块,可使多元异构数据转换为符合统一标准和格式的通用数据<sup>[10]</sup>。数据质量控制和标准化的相关方法及应用案例已作为教材的部分内容发表,可供其他学者参考<sup>[24]</sup>。

(2)社会效益:大规模自然人群队列建设是一项艰巨的长期工程,作为我国百万人群队列的一部分,京津冀自然人群队列不仅可为三地提供优质科研平台和珍贵生物样品资源,也可为国内其他地区学者开展学术合作与交流、开展不同地区人群健康问题差异性研究,以及队列资源整合与拓展等提供独特资源。此外,京津冀自然人群队列中的资料收集方式主要为现场调查,通过面对面询问的方式获取多种健康相关信息。因此,健康宣教和促进等活动可自然地融入整个调查过程,体现医学研究的社会效益。

3. 不足和挑战:队列研究的价值在于长期随访。京津冀自然人群队列于 2017 年启动基线调查,对于暴露因素和结局事件的观察时间尚短,资料收集内容有限,仍不足以分析危险因素长期暴露及其动态变化对健康的影响。目前的科研产出主要基于基线调查数据,未充分发挥队列研究在病因探索、健康和疾病转归,以及发病趋势研究等方面的优势。基于队列纵向数据的高质量科研产出仍需时间积累。然而,长期随访需要有稳定的科研经费支持,而一般科研项目的资助周期往往仅 5 年左右,国内很多队列研究在项目结束后由于缺乏后续

经费支持被迫中断<sup>[25]</sup>。此外,大队列的大数据管理、生物样品库建设、软硬件设备的研发和投入使用等,也需要投入大量人力资源和时间精力。在无法保证稳定支持的情况下,大规模自然人群队列的维持和后续研究都将面临巨大的困难和挑战。

### 三、规划与展望

长期随访与资源共享:长期和高质量随访是队列研究的科学价值所在,也是其实施难点。京津冀自然人群队列在设计之初便特别强调长期随访机制的建立,并根据各队列特色探讨了多种方式相结合的随访管理机制,最大程度降低失访率。京津冀三地虽然在地域上紧密衔接,但队列人群的人口结构和地方管理模式仍存在差异,因此在随访结局事件的发现中需要结合多种手段<sup>[10]</sup>。目前的随访方式包括主动发现和被动监测,前者包括重复测量方式的面对面调查和电话随访,后者包括与疾病监测数据相关联以获取结局信息等。由于长期随访不仅需要人力、物力、财力等方面的大量、稳定投入,还需要当地政府、医疗卫生机构等多部门的全力支持,因此,如前所述,未来队列建设仍面临诸多挑战,包括科研经费的持续支持、人员队伍的相对稳定、技术手段的不断更新、随访方式的不断健全等。此外,面对呼吸道传染病的流行,基于传统现场调查模式的自然人群队列研究,如何平衡科学研究必要性和疫情防控的需求,也是摆在科研人员面前的一道难题。

数据和资源的有效共享是盘活科技资源和产出高水平成果的重要渠道。目前京津冀自然人群队列的基础数据已按照国家统一管理要求,全部汇交至国家人口健康科学数据中心。在项目执行过程中,项目组已按照“共生、共建、共享和共赢”的原则制定了《京津冀区域自然人群队列研究数据和样品共享协议》,为项目组内部各数据和生物样品产生单位之间实现共享提供了机制保障。然而,由于时间关系,目前不同队列的数据尚未完全整合,基于队列整体人群数据的科研工作仍有待进一步梳理。未来,在项目内部,将继续推进不同队列的数据标准化;在项目之外,将按照示范队列提供的标准和技术,与其他区域队列建立更多合作,并与专业的数据管理人员一起,围绕资源共享中的科研成果转化、知识产权保护、信息安全和生物安全等方面存在的问题开展研讨。

### 四、小结

大数据时代的到来使大型队列研究成为流行

病学研究的主旋律。作为精准医学百万人群队列的重要组成部分,京津冀自然人群队列包含了多样化的人群,其长期随访将为我国医学科研工作提供独特和宝贵的资源。未来,项目组将继续结合京津冀地区健康问题特色,持续探索生活环境和遗传因素对健康的影响,特别是当前社会经济发展过程中亟需关注的心血管代谢疾病、老龄化及共病、空气污染的健康效应等健康问题。

京津冀自然人群队列研究项目包括多家合作单位:首都医科大学、华北理工大学、天津市妇女儿童保健中心、南开大学、北京市和平里医院、河北省疾病预防控制中心、北京市朝阳区疾病预防控制中心、首都儿科研究所、北京市体检中心、河北医科大学等。随着队列的长期随访,未来将有更多合作单位的加入。

利益冲突 所有作者声明无利益冲突

作者贡献声明 何慧婧:研究实施、数据整理分析和论文撰写;潘利、张玲、袁聚祥、刘功妹、张敏英:研究设计和实施;胡耀达、涂吉:论文修改;单广良:研究设计和实施、研究指导、经费支持和论文修改

### 参 考 文 献

- [1] 熊玮仪,吕筠,郭或,等.大型前瞻性队列研究实施现况及其特点[J].中华流行病学杂志,2014,35(1):93-96. DOI: 10.3760/cma.j.issn.0254-6450.2014.01.022.
- [2] Xiong WY, Lv J, Guo Y, et al. Overview on the practice and characteristics of large prospective cohort studies[J]. Chin J Epidemiol, 2014, 35(1): 93-96. DOI: 10.3760/cma.j.issn.0254-6450.2014.01.022.
- [3] Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age[J]. PLoS Med, 2015, 12(3):e1001779. DOI:10.1371/journal.pmed.1001779.
- [4] 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. DOI:10.1093/ije/dyr120.
- [5] 胡志斌,杜江波,徐欣,等.中国国家出生队列建设背景和设计简介[J].中华流行病学杂志,2021,42(4):569-574. DOI:10.3760/cma.j.cn112338-20201211-01402.
- [6] Hu ZB, Du JB, Xu X, et al. Profile of China national birth cohort[J]. Chin J Epidemiol, 2021, 42(4): 569-574. DOI: 10.3760/cma.j.cn112338-20201211-01402.
- [7] Zhao D, Liu J, Wang M, et al. Epidemiology of cardiovascular disease in China: current features and implications[J]. Nat Rev Cardiol, 2019, 16(4): 203-212. DOI:10.1038/s41569-018-0119-4.
- [8] Du X, Patel A, Anderson CS, et al. Epidemiology of cardiovascular disease in China and opportunities for improvement: JACC international[J]. J Am Coll Cardiol, 2019, 73(24):3135-3147. DOI:10.1016/j.jacc.2019.04.036.
- [9] Yang W, Wu B, Tan SY, et al. Understanding health and social challenges for aging and long-term care in China[J]. Res Aging, 2021, 43(3/4): 127-135. DOI: 10.1177/0164027520938764.
- [10] He HJ, Guo P, He JS, et al. Prevalence of hyperuricemia and the population attributable fraction of modifiable risk factors: evidence from a general population cohort in China[J]. Front Public Health, 2022, 10: 936717. DOI: 10.3389/fpubh.2022.936717.
- [11] He JS, Hu SH, Xu XM, et al. Association of long-term exposure to PM<sub>2.5</sub> in workplace with fasting plasma glucose among asymptomatic adults: a multicenter study in North China[J]. Environ Int, 2022, 166: 107353. DOI: 10.1016/j.envint.2022.107353.
- [12] 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.
- [13] Shan GL. Principles and practice on cohort study of general population in Beijing, Tianjin and Hebei province[J]. Chin J Epidemiol, 2021, 42(8): 1493-1497. DOI: 10.3760/cma.j.cn112338-20200615-00843.
- [14] Hagemann E, Silva DT, Davis JA, et al. Developmental Origins of Health and Disease (DOHaD): the importance of life-course and transgenerational approaches[J]. Paediatr Respir Rev, 2021, 40:3-9. DOI:10.1016/j.prrv.2021.05.005.
- [15] He HJ, Pan L, Pa LZ, et al. Data resource profile: the China national health survey (CNHS) [J]. Int J Epidemiol, 2018, 47(6):1734-1735f. DOI:10.1093/ije/dyy151.
- [16] Liu K, Cao H, Guo CY, et al. Environmental and genetic determinants of major chronic disease in Beijing-Tianjin-Hebei region: protocol for a community-based cohort study[J]. Front Public Health, 2021, 9: 659701. DOI: 10.3389/fpubh.2021.659701.
- [17] Wang S, Zhang GH, Wang J, et al. Study design and baseline profiles of participants in the Tianjin birth cohort (TJBC) in China[J]. J Epidemiol, 2022, 32(1):44-52. DOI:10.2188/jea.je20200238.
- [18] Cao H, Li BX, Peng WJ, et al. Associations of long-term exposure to ambient air pollution with cardiac conduction abnormalities in Chinese adults: the CHCN-BTH cohort study[J]. Environ Int, 2020, 143: 105981. DOI:10.1016/j.envint.2020.105981.
- [19] Liu K, Cao H, Li BX, et al. Long-term exposure to ambient nitrogen dioxide and ozone modifies systematic low-grade inflammation: the CHCN-BTH study[J]. Int J Hyg Environ Health, 2022, 239: 113875. DOI: 10.1016/j.ijheh.2021.113875.
- [20] Liu XH, Cao H, Zhu HP, et al. Association of chronic diseases with depression, anxiety and stress in Chinese general population: the CHCN-BTH cohort study[J]. J Affect Disord, 2021, 282: 1278-1287. DOI: 10.1016/j.jad.2021.01.040.
- [21] Zhang SK, Wang YB, Li QL, et al. Different exposure metrics of rotating night shift work and hyperhomocysteinaemia among Chinese steelworkers: a cross-sectional study[J]. BMJ Open, 2020, 10(12): e041576. DOI:10.1136/bmjopen-2020-041576.
- [22] Cui SY, Li C, Chen Z, et al. Research on risk prediction of dyslipidemia in steel workers based on recurrent neural network and LSTM neural network[J]. IEEE Access, 2020, 8:34153-34161. DOI:10.1109/ACCESS.2020.2974887.
- [23] Zhang SK, Wang YB, Zhu Y, et al. Rotating night shift work, exposure to light at night, and glomerular filtration rate: baseline results from a Chinese occupational cohort[J]. Int J Environ Res Public Health, 2020, 17(23): 9035. DOI: 10.3390/ijerph17239035.
- [24] Yan YK, Liu JT, Zhao XY, et al. Regional adipose compartments confer different cardiometabolic risk in children and adolescents: the China child and adolescent cardiovascular health study[J]. Mayo Clin Proc, 2019, 94(10):1974-1982. DOI:10.1016/j.mayocp.2019.05.026.
- [25] Wang J, Liu EQ, Wang Y, et al. Association of early pregnancy body mass index and children's birth weight with risk of being overweight in childhood[J]. Am J Hum Biol, 2018, 30(5):e23174. DOI:10.1002/ajhb.23174.
- [26] 涂吉,何慧婧,胡耀达,等.美国国家健康与营养调查中的信息化系统应用及启示[J].中华流行病学杂志,2022,43(7):1127-1133. DOI:10.3760/cma.j.cn112338-20211109-00871.
- [27] Tu J, He HJ, Hu YD, et al. Application and inspiration of information system used in national health and nutrition examination survey of America[J]. Chin J Epidemiol, 2022, 43(7):1127-1133. DOI:10.3760/cma.j.cn112338-20211109-00871.
- [28] 单广良.流行病学与应用多元统计分析[M].北京:中国协和医科大学出版社,2022.
- [29] Shan GL. Epidemiology and applied multivariate statistical analysis[M]. Beijing: Peking Union Medical College Press, 2022.
- [30] 王慧,陈培战,张作文,等.我国人群队列研究的现状、机遇与挑战[J].中华预防医学杂志,2014,48(11):1016-1021. DOI:10.3760/cma.j.issn.0253-9624.2014.11.019.
- [31] Wang H, Chen PZ, Zhang ZW, et al. The current situation, opportunity and challenge of population cohorts in China[J]. Chin J Prev Med, 2014 48(11): 1016-1021. DOI: 10.3760/cma.j.issn.0253-9624.2014.11.019.