

多主体协作的分布式大型队列的组织与管理模式

——以欧洲EPIC队列为例

孙学会¹, 李伟², 周晟², 徐萍², 许丽^{2*}

1. 复旦大学生命科学学院, 上海 200433

2. 中国科学院上海生命科学信息中心, 中国科学院上海营养与健康研究所, 上海 200031

摘要 大型队列已成为生物医药与公共卫生领域重要的开放性科研基础设施和卫生决策支撑平台, 其意义不断凸显, 多国已布局大型队列建设。聚焦典型的多主体协作的联盟式大型队列——欧洲癌症与营养前瞻性调查(EPIC)项目, 分析其组织机制、经费来源、建设模式、标准化方案、资源管理与共享策略等, 剖析其成功实施与运行的关键。分析显示, 多主体协作的联盟分布式大型队列建设需要持续的投入和长效科研机制的保障, 且通过前瞻顶层设计、建立协调统一的管理机制、开展标准全面的数据和生物样本收集、建立完善的政策保障体系, 才可确保其持续、高质量建设。建议中国由政府主导规划并科学设计, 建立一体化的大型队列组织和管理体系; 设立专项进行长期稳定支持, 确保大型队列的持续、高质量建设; 建立标准统一的建设与管理方案, 保证大型队列资源的规范化和系统化; 提高数据资源规模、类型和质量, 建立开放共享的高质量大型队列; 探索建立合理的人类遗传资源保护和知识产权管理制度。

关键词 多主体协作; 大型队列; 建设模式; 管理机制

大型队列 (large cohort) 研究是针对数十万人群的健康及疾病 (尤其是慢性病) 状况进行持续数十年调查与追踪研究的科学方法, 是目前国际公认

研究慢性病病因的首选设计之一, 也是生命组学样本和表型组学数据的重要来源^[1]。大规模人群队列的建立有助于更好地研究暴露与结局之间的关系,

收稿日期: 2023-11-20; 修回日期: 2024-07-01

基金项目: 中国科学院学部局项目(2016-SM-C-01); 国家自然科学基金项目(L1624031); 上海市科委软课题项目(21692191700)

作者简介: 孙学会, 助理研究员, 研究方向为分子流行病学与队列, 电子信箱: xhsun@fudan.edu.cn; 许丽 (通信作者), 副研究馆员, 研究方向为生命健康领域战略, 电子信箱: xuli@sinh.ac.cn

引用格式: 孙学会, 李伟, 周晟, 等. 多主体协作的分布式大型队列的组织与管理模式——以欧洲EPIC队列为例[J]. 科技导报, 2024, 42(22): 84-92; doi: 10.3981/j.issn.1000-7857.2024.02.00300

在帮助理解基因及其与环境因素的相互作用如何影响人群健康和疾病方面,新的大型队列很可能会占据中心舞台^[2]。同时,由于人群特殊性的不可复制、环境特殊性的不可替代,世界各国开始普遍认识到大型队列关乎本国的国民健康、生物资源、国家安全的战略意义,由政府主导建设大型人群队列已成主要趋势^[3]。

目前,大型队列的组织模式主要有2类,一是政府主导的一体化大型队列,如英国生物样本库(UK Biobank)、美国护士健康研究队列、美国百万人队列项目(All of Us 队列项目)^[4]、中国慢性病前瞻性研究项目(CKB)等均采用此模式;二是多主体协作的联盟分布式大型队列,该组织模式广泛应用于欧洲的多个大型研究中,其优势是可在较短时间内形成大样本,充分利用和融合已有基础的队列研究,但队列建设和数据的标准化存在较大挑战^[5]。

欧洲癌症与营养前瞻性调查(European prospective investigation into cancer and nutrition, EPIC)队列是典型的多主体协作的联盟分布式大型队列,也是最早探索建立的大型人群队列研究之一。EPIC 队列通过建立协调统一的管理机制和组织架构,将原本零散、规模较小和对象相对单一的队列整合成为覆盖地域广泛、研究人群及膳食模式多样的大型队列,并通过科学规划和统筹设计保证了数据采集、管理和共享的标准化与规范化等。这种模式也在之后的欧洲甚至跨地区的多项大型队列研究中广泛应用。本文聚焦欧洲EPIC队列的组织与

管理模式,深入剖析多主体协作的联盟分布式大型队列得以成功实施的关键因素等经验。

1 EPIC 队列的发展历程

20世纪80年代,科学界普遍认识到饮食因素与癌症发生相关,但受限于此前的前瞻性队列研究体量较小、参与者来源多样性不足、膳食模式等因素同质性高等问题,早期开展的营养与癌症的流行病学研究较为局限,难以为明确特定膳食因素与癌症发生风险的关联提供有力证据。为克服这些局限,联合国世界卫生组织(WHO)下属国际癌症研究机构(International Agency for Research on Cancer, IARC)于1990年发起了EPIC队列项目,以在欧洲大规模人群中研究膳食模式、生活方式、遗传特征与肿瘤等慢性疾病的关系^[6]。

EPIC队列是一项多中心前瞻性大型队列研究,最早由7个核心国家(法国、德国、希腊、意大利、荷兰、西班牙和英国)的17个研究中心(即子队列,下称EPIC研究中心)启动建设,又陆续吸纳了意大利的那不勒斯研究中心以及新加入的瑞典、丹麦和挪威3个国家的5个研究中心,逐步发展为由欧洲10个国家的23个研究中心组成的大规模人群队列(表1)^[7]。该队列自1992年正式启动,至2000年招募了共521342名健康成年人(男性153432名、女性367910名,年龄多在35~70岁间),并完成基线数据收集,同时,从20世纪90年代中期即启动对癌症发

表1 EPIC队列的主要发展历程

| 时间 | 里程碑事件 |
|------------|--|
| 1990年 | 启动试点研究 WHO下属IARC和EPIC研究中心在欧洲委员会支持下启动EPIC试点工作 欧洲委员会通过“欧洲抗癌计划”正式启动EPIC项目,各国EPIC研究中心陆续加入并开展招募工作:(1) 法国、德国、希腊、意大利、荷兰、西班牙和英国7个核心国家的17个研究中心陆续启动招募工作;(2) 1993年,意大利的那不勒斯研究中心加入;(3) 1993年,瑞典的2个研究中心加入;(4) 1996年,丹麦的2个研究中心加入;(5) 1998年,挪威的1个研究中心加入 |
| 1992年 | 正式启动 |
| 20世纪90年代中期 | 开始随访 正式启动随访工作,由各EPIC研究中心自行开展,每3~4年定期随访一次 |
| 2000年 | 完成招募&基线数据收集 所有EPIC研究中心已完成参与者招募和基线数据收集工作 |

病率和特定原因死亡率的数十年的跟踪和研究,形成了全球开展癌症等疾病病因学前瞻性研究的最大资源。目前,EPIC 队列的数据资源向全球开放,基于该资源发表研究论文量达 1800 余篇,对了解营养与健康的关系以及改善公共卫生现状作出重要贡献。

2 EPIC 队列多主体协作的组织与运行经验

EPIC 队列作为典型的多主体协同建设的大型队列无疑是成功的,经过多年的建设和运行,构建了统一标准的基线数据集和 50 万人级的生物样本库,并已产出多主体协作的重要科研成果。EPIC 队列成功的关键在于其从试点建设开始,就前瞻性设计解决了队列经费筹集、组织管理和协调机制、资源采集流程和标准、资源管理和共享机制等问题,简要归纳总结 EPIC 队列的组织和管理经验如下。

2.1 前期预研构建标准统一、科学可行的研究方案

为构建可保障超大规模队列成功实施和运行的可靠方案,1990—1993 年,EPIC 队列发起初期,牵头机构 IARC 主导在 9 个国家的多个 EPIC 研究中心开展了系列试点研究与验证工作。重点围绕参与者招募方案、问卷调查内容和生物样本采集试点的可行性进行了预研,尤其是重点针对大规模人群的膳食调查方法展开探索,为标准统一、科学可行的研究方案的形成和最终确定提供了宝贵信息。1992 年,根据试点项目的方法学研究结果,欧洲委员会通过其“欧洲抗癌计划”(Europe Against Cancer programme, EAC)正式启动 EPIC 队列研究。

2.2 政府启动和引导支持、多元参与的筹资模式

EPIC 队列的经费来源采用了“项目分担,资金自筹”的模式,由欧盟提供 EPIC 启动建设的种子资金,并支持各 EPIC 研究中心在建设过程中,积极争取所在国家的政府、研究机构和基金会的资金。即欧盟提供部分资金、各 EPIC 研究中心自筹经费支持开展相关队列建设与研究活动。

EPIC 队列是“欧洲抗癌”计划的一部分,欧洲研究委员会(ERC)通过该计划向 EPIC 提供了约

50% 的资金,又通过欧盟第六、第七框架计划提供部分资助,其余部分由世界癌症研究基金会(WCRF)、各参与国政府、研究机构或慈善组织等提供。如英国牛津研究中心的资金主要来源于欧盟、英国医学研究理事会(Medical Research Council, MRC)和英国维康信托基金会(Wellcome Trust)等;意大利米兰研究中心的经费来源包括世界癌症研究基金会,以及意大利卫生部、意大利癌症研究协会、意大利国家研究委员会、西班牙圣保罗基金会等。这种由欧盟作为协调主体启动和引导,吸引各参与主体自行筹措资金来支持队列各项活动的方式,可保障大型队列获得充足持续的经费支持。

2.3 协调统一的管理机制和组织架构

不同参与国家、研究中心间的协调、组织和和管理是多主体协作的联盟式大型队列运行需解决的首要问题,EPIC 队列建立了协调统一的组织管理体系进行统筹管理,即由指导委员会主导、国家协调中心和 EPIC 活动协调中心共同参与的管理与协调机制,以保障该大型队列的稳定有序运行(图 1)。EPIC 队列组建了由发起机构 IARC、EPIC 活动协调中心及 23 个 EPIC 研究中心的代表组成的指导委员会(EPIC Steering Committee, SC)负责项目的总体管理和协调,各国的 EPIC 研究中心根据其队列目标设置自行开展基线调查、重复调查、长期随访监测等活动。同时,EPIC 队列建立了主数据库和中央生物库,由 IARC 负责管理和维护。

具体科研活动上,EPIC 队列针对不同疾病或研究主题(包括肿瘤、传染病、营养与代谢、遗传学等),分别在 IARC、英国帝国理工学院(ICL)及各 EPIC 研究中心成立了 27 个研究工作组,由工作组组长负责指导相关方向的研究、审查相关样本与数据的申请。另外,EPIC 队列依托英国帝国理工学院建立 EPIC 活动协调中心对各工作组的科研活动进行统筹。

2.4 标准规范的资源采集流程

2.4.1 多样化、广覆盖的参与者招募

EPIC 队列的 23 个研究中心广泛分布于西欧 10 个国家的不同地理区域,其招募对象为自然人群,即居住于不同区域、处于不同社会经济层次、具

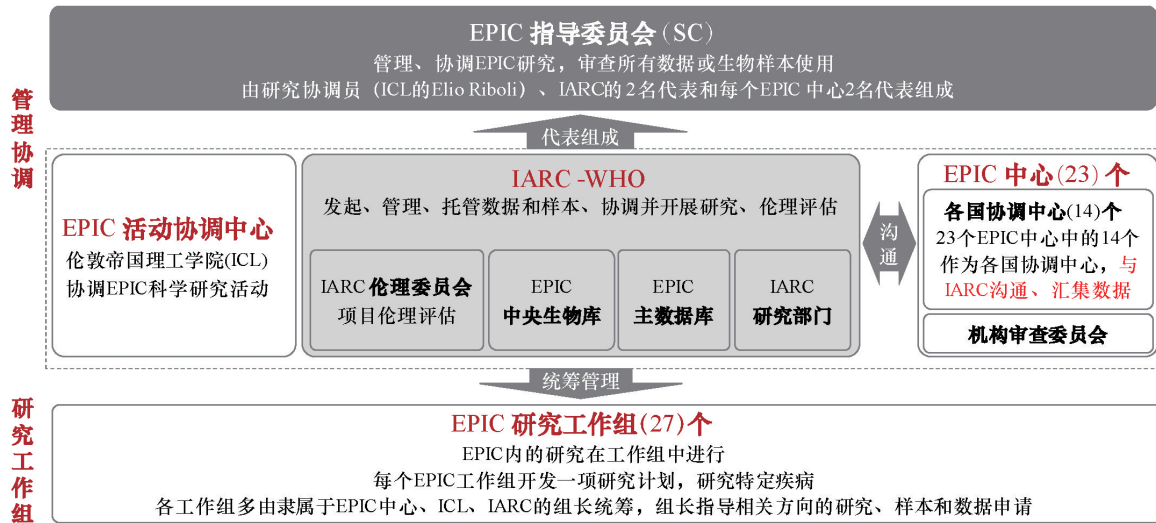


图1 EPIC队列的组织管理架构

不同膳食模式,且不限特定疾病或健康状况的多样化成年人群。

各EPIC研究中心(即子队列)的建设模式包括整合原有队列、从头建设新队列等多种方式,参与者招募方式也不尽相同。项目的研究对象多从居住在特定城镇或地理区域的普通成年人中邀请,有些EPIC研究中心则会依托健康保险、献血者协会、某些项目等进行招募^[8]。如法国的EPIC研究中心以参与健康保险的公立学校职员为邀请对象;意大利、西班牙的部分EPIC研究中心吸纳了当地献血者协会的成员;荷兰乌得勒支EPIC研究中心、意大利佛罗伦萨EPIC研究中心则包括了参加当地乳腺癌筛查项目中的妇女。此外,不同EPIC研究中心涉及的膳食模式也不尽相同,如英国牛津EPIC研究中心的研究对象中包括纯素食者、乳蛋素食者和鱼素食者等不同膳食模式的人群。

2.4.2 标准化的数据与生物样本资源采集

基于EPIC队列在试点阶段探索建立的标准化

工作流程,各EPIC研究中心开展规范化的参与者招募与资源采集工作。根据研究目标设定,EPIC队列基线采集的数据包括膳食摄入数据(即项目的核心数据)、生活方式、身体测量数据,以及血液样本的采集,不同EPIC研究中心采集的资源类型有一定差别(表2)。

1) 膳食摄入数据。膳食数据评估上,不同EPIC中心采用的调查方式有所不同,主要包括定量饮食问卷调查、半定量饮食频率问卷调查(FFQ),以及结合多种调查方式的联合饮食问卷调查(如结合半定量饮食频率问卷调查与7 d饮食记录)3种方式^[8]。其中,定量饮食问卷调查中,EPIC涉及到了260余种食物,并对主餐(早餐、午餐、晚餐、餐间)消费场所及特定食物的食用频率及份量进行调查评估。

2) 生活方式数据。除膳食摄入情况外,EPIC队列还通过问卷调查获取大量生活方式数据,以探究可能影响其营养状况或与癌症发病率相关的风险

表2 EPIC队列的基线采集数据类型

| 资源类型 | | 具体信息 |
|------|-------|--------------------------------------|
| 问卷调查 | 饮食变量 | 定量或半定量饮食情况调查 |
| | 非饮食变量 | 生育史、身体活动状况、吸烟状况、饮酒状况、职业历史、社会经济地位、患病史 |
| 身体测量 | | 身高、体重、腰围、臀围和坐高 |
| 生物样本 | | 血液样本:血浆、血清、血沉棕黄层、红细胞 |

险因素。最早参与EPIC的7个国家商定了一套统一的调查问卷,后加入国家虽然采用了独立的调查问卷,但仍覆盖了该统一问卷所涉及的内容,同时参照EPIC的统一问卷对核心变量进行标准化。

3) 身体测量数据。身体特征与能量摄入、代谢效率、运动情况相关,且多个指标与癌症发病率(如身高、体脂分布)相关。因此,EPIC队列各研究中心使用试点阶段建立的标准化测量方案,或采用参与者自述方法,获取基本的身体特征测量数据,主要包括参与者的身高、体重、腰围、臀围和坐高。

4) 生物样本采集。EPIC队列采集的生物样本主要为血液样本。为确保其血液样本资源的高度标准化,EPIC队列集中向Cryo Bio System公司定制统一的血液样本分装塑料吸管等存储材料和系统,分发给各EPIC研究中心;各EPIC研究中心采用规范的工作流程,采集每个参与者的血液样本,统一处理后进行保存。EPIC队列共采集了38万余名参与者的血液样本。

5) 随访监测数据。除上述基线调查数据资源采集外,自1990年代中期,EPIC队列即通过定期调查、依托医疗信息系统跟踪等方式,每3~4 a定期对研究对象进行随访,调查其生活方式、健康状况的变化,以及重大疾病的患病情况、生命状况和死亡原因等,迄今已开展长达15年的长期随访监测。其中,在癌症发病率和总死亡率监测方面,EPIC队列通过健康保险记录、癌症和病理学注册等渠道,以及随访研究对象或其亲属等方法主动追踪癌症病例发生及死亡情况。

2.5 中心化的资源存储和管理

对于数据和生物样本资源,EPIC队列采用各研究中心独立保存、总协调中心镜像存储或部分储存的方式进行中心化存储和管理。EPIC队列的中央生物库和主数据库(EPIC ORACLE数据库)由IARC负责总体维护和管理,各EPIC研究中心通过本国的国家协调中心(共14个)与IARC沟通,实现数据与生物样本资源的中心化存储(图1)。

1) 数据标准化与集中存储。针对不同EPIC研究中心来源的复杂数据资源,EPIC队列通过建立数据校正机制等进行标准化处理后再集中存储

于其ORACLE主数据库中,以促进数据整合^[9]。同时,为保护参与者隐私与数据安全,EPIC队列中带个人标签信息的数据仅存储在各EPIC研究中心当地,不会转移到项目的ORACLE主数据库。

对于来自各EPIC研究中心的膳食摄入数据,EPIC采用统一格式和分类系统进行处理,并开发了可兼容9种语言的“计算机辅助的24 h饮食回顾数据采集(EPIC-SOFT)”软件^[10],从每个EPIC研究中心抽样选取5%~12%参与者的调查数据进行参考校正,实现对数据的标准化处理,以最大程度地提高不同EPIC研究中心间数据的可比性;另外,项目还通过开发通用食品成分数据库^[11],以利用营养成分对膳食数据进行标准化(图2)^[9]。

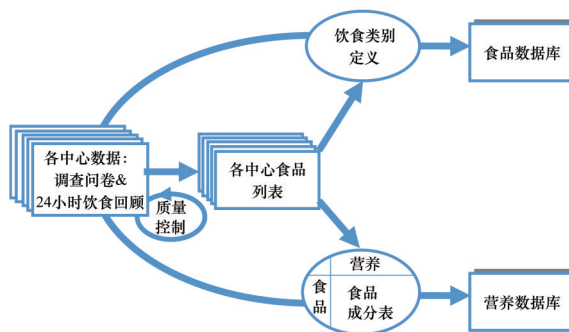


图2 EPIC队列膳食摄入数据存储流程

对于生活方式数据和人体测量数据,EPIC队列也制定了一套全面的重编码方案,将其转化为标准化的通用变量名称和格式,并通过数据质量控制检查后再进行存储(图3)^[9]。

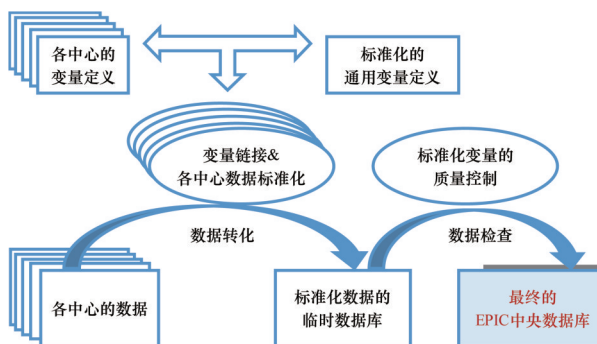


图3 EPIC生活方式数据和人体测量数据存储流程图

2) 生物样本的中心化存储。EPIC 队列生物样本集中存储于由 IARC 生物库(IARC BioBank, IBB)托管的中央生物库中,目前保存了 EPIC 队列的 37 万余名参与者的 400 多万管血液样本^[8]。其中,最早参与的 7 个核心国家的血液样本存储采用镜像储存(各 14 管)的方式,分别存储于当地 EPIC 研究中心的生物库和 EPIC 中央生物库中(图 4);后加入的国家中,则由于存储条件的不同,采用了不同的样本存储方式,如挪威的 EPIC 研究中心所采集的血液样本仅部分保存于 EPIC 中央生物库,瑞典和丹麦的 EPIC 研究中心的血液样本存储于玻璃试管(而非塑料吸管)中,不适合 EPIC 中央生物库的储存条件,因此仅存储在当地生物库中。

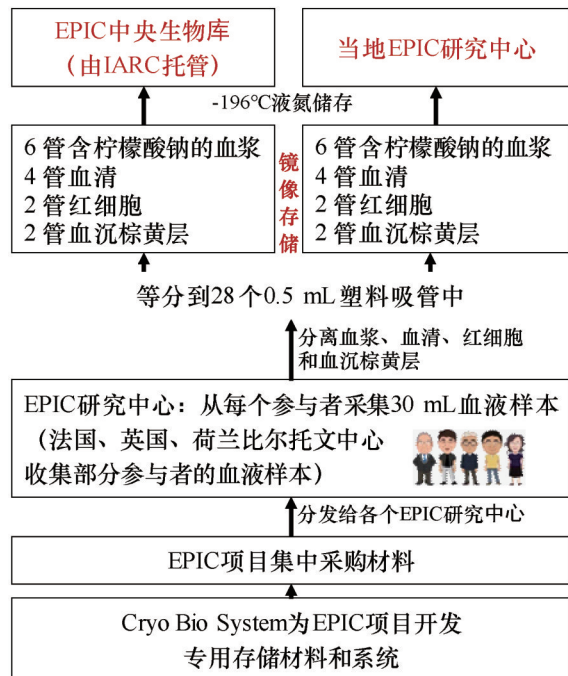


图4 EPIC 的主要研究中心的标准化生物样本采集与存储流程

2.6 统一的资源开放与共享机制

目前,EPIC 队列数据资源向全球科研人员开放与共享,且在必要情况下,生物样本也可供申请使用,具体包括资源使用申请、审批、移交和成果共享等流程。

1) 资源使用申请及审批。EPIC 队列资源的使用申请由指导委员会负责审批,并由相关研究方

向的 EPIC 工作组组长具体指导。研究人员需通过相关 EPIC 工作组组长提交访问申请,也可直接向相应 EPIC 中心首席研究员或 EPIC 指导委员会提交。指导委员会在收到申请后将会对其科学先进性、研究内容、研究实力等进行审核。获批后,申请者需签署相应的数据传输协议(DTA)和材料转让协议(MTA),提交给 EPIC 主数据库、EPIC 中央生物样本库或相应的 EPIC 研究中心以获得相应的数据或生物样本(图 5)。

2) 知识产权与成果管理。大型队列积累了宝贵的数据与生物样本资源,建立合理的知识产权与成果管理机制,是保证各方利益、促进数据持续共享的前提,这对于多主体协同建设的联盟式大型队列尤为重要。EPIC 队列生物样本和数据资源,以及相关衍生数据均归各 EPIC 研究中心所有,但研究人员利用其资源所产生研究成果的知识产权归科研人员所有。

同时,及时公布研究结果、认可资源提供者的贡献是保证数据共享、促进队列资源价值最大化的前提。一方面,EPIC 队列要求资源使用者遵循出版物致谢原则,在相关成果发表时需以口头或书面形式进行说明或致谢,如注明“本研究获得了该生物样本库/数据库提供的支持”;另一方面,EPIC 队列要求研究人员利用该资源研究所获得的衍生数据,需在研究结束 2 年内及时返还,进一步作为其资源积累。

迄今,基于 EPIC 队列资源的研究已产出大量高影响力论文,且持续开展各参与国癌症发生情况监测及预测,为各参与国甚至全球医疗卫生管理提供公共卫生政策参考建议。

3 结论

当前,大型队列不仅是流行病学研究主要方法,还能够汇集海量人群数据和生物样本资源,发挥研究平台功能。因而,大型队列在健康水平提升、新冠疫情控制、科学研究、生物医药产业促进中的意义不断凸显,已成为国民健康水平提升和生物医药创新的关键基石,大型队列研究逐渐受到重

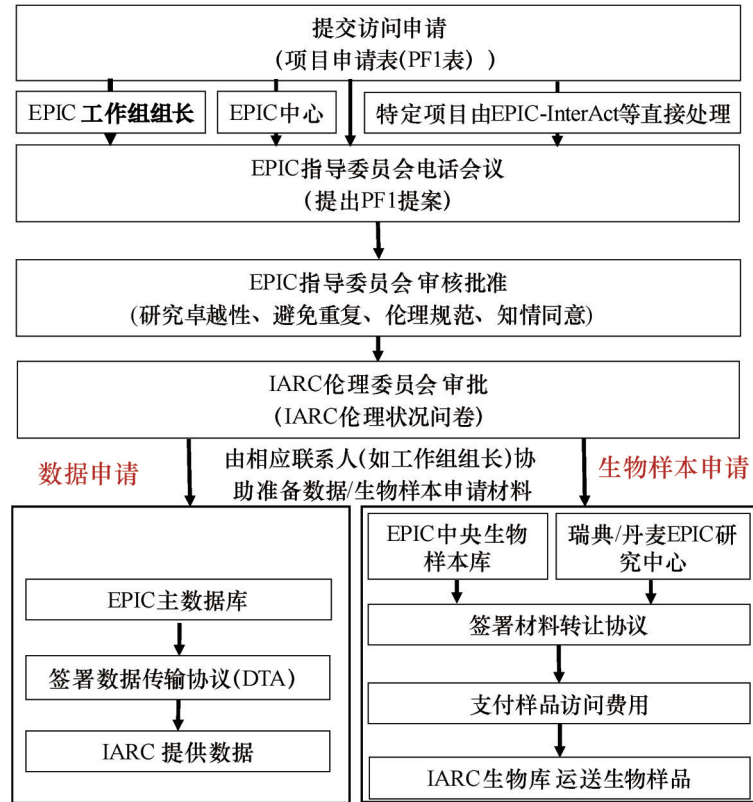


图5 EPIC队列资源访问申请流程

视,全球布局大型队列建设。

数据驱动的科学范式研究和精准医学研究路径的推广,对大型队列建设提出了新的要求,即大型队列建设要覆盖更加广泛的自然人群,数据要整合与共享等。中国目前已建设不同规模、不同类型的队列,覆盖了不同地区、职业、年龄及疾病类型,其中,各地区各单位自发建设的10万人群规模以上的大型队列研究已有10余个,如中国慢性病前瞻性研究项目、中国高血压随访调查队列、泰州人群健康跟踪调查队列等,这些队列为中国相关重大慢病的预防和控制提供了科学基础。并且随着对国家级大型队列的认识逐渐加深,“十三五”期间,国家“精准医学研究”重点专项布局了大型队列建设与研究任务,即希望通过专项的支持和引导,将国内基础较好的队列进行整合,统一队列建设的技术、标准和规范,实现样本和数据统一管理,以及充分共享和应用。但总体来看,中国队列仍面临缺乏顶层设计造成的有效规模、质量控制和利用率不足,

自发建设造成的标准和规范不统一、难以共享利用,执行期短、缺乏稳定持续的经费支持等问题,不能满足生物医学发展新趋势的需要。因此,亟需采取相应举措提升队列数据质量、实现队列间的数据共享和互联互通,推动不同队列之间的真正融合。

基于对欧洲 EPIC 队列建设模式与管理经验的分析,多主体协作的联盟分布式大型队列建设需要持续的投入和长效科研机制的保障,且通过前瞻顶层设计、建立协调统一的管理机制、开展标准全面的数据和生物样本收集、并建立完善的政策保障体系,才可确保其持续、高质量建设。

1) 政府主导规划并科学设计,建立一体化的大型队列组织和管理体系。从政府层面进行科学顶层设计和长远前瞻规划,形成大型队列建设的路线图和实施方案。建设模式上,采取“从头建设”及与已有队列建设相结合方式,探索、持续支持建设规模化、标准化、高质量的大型队列;组织和管理上,建立由专门机构集中管理、协调各类组织机构

共同实施的一体化组织和管理机制,充分调动和集合多方力量共同推动队列建设和运营,提高大型队列建设的工作效率和可操作性。

2) 设立专项进行长期稳定支持,确保大型队列的持续、高质量建设。设立国家级大型队列建设重大专项,整体部署研究,提供稳定的经费支持,并由国家财政保证国家队列的基础支撑和公共服务性质,长期稳定运行。同时,探索多元化资金筹集方式,由政府投入启动经费,完成队列建设,建成后的运行由政府投资与科研项目经费、公益基金和商业资金多元化资金共同支持,形成“政府启动、多元参与”的大型队列建设资助机制。

3) 建立标准统一的建设与管理方案,保证大型队列资源的规范化和系统化。通过试点项目进行工作流程设计、建立标准委员会等方式,制定统一标准方案并进行规范管理,对生物样本与数据资源采集和管理过程进行严格质控,且所涉及的器械和材料均统一采购并集中调试,减少系统偏差,以加强不同队列间数据的稳定性及可比性,保证大型队列资源的标准化、规范化和系统化,提高大型队列资源的有效规模。同时,以标准统一、互联互通、数据共享的形式搭建多点分中心的生物样本和数据资源存储与管理平台,保障大型队列资源的科学管理。

4) 提高数据资源规模、类型和质量,建立开放共享的高质量大型队列。数据密集型生物医学研究范式对大型队列建设提出了新的要求,一方面,要提升大型队列数据规模、类型和精细度,注重人群的多样性和覆盖度,开展覆盖“全生命周期”的多维度、高精度的表型组和暴露组健康数据的收集,并持续扩展大型队列资源的精细化程度、提高队列价值、提升研究能力;另一方面,积极推动大型队列资源的开放共享,并充分利用大数据、区块链、人工智能等前沿新技术建立综合平台和系统,对大型队列海量人群资源的规范管理、常规监测,保障数据的安全开放共享。

5) 探索建立合理的人类遗传资源保护和知识产权管理制度。人类遗传资源管理关乎国家安全,因此,应构建完善的隐私保护和数据安全法规制度

以及保护措施,以在保障参与者隐私与数据安全的基础上,尽可能广泛地实现数据共享;同时,要探索建立合理的知识产权保护政策与成果管理机制,从而保障大型队列各参与主体的权益,以推动大型队列资源的持续共享,提高大型队列的利用率,从而实现大型队列海量数据资源价值的最大化。

参考文献 (References)

- [1] 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.
- [2] 陈柯婷, 於一凡, 刘静, 等. 多中心大型人群队列全生命周期管理理论与实践探索[J]. 现代预防医学, 2022, 49(13): 2317-2319, 2334.
- [3] 陈兴栋, 蒋艳峰, 徐萍, 等. 大型人群队列遗传资源建设与利用[J]. 遗传, 2021, 43(10): 980-987.
- [4] 李伟, 孙学会, 徐萍, 等. 美国 All of US 队列项目建设模式与特点分析[J]. 世界科技研究与发展, 2022, 44(2): 265-274.
- [5] 许丽, 李伟, 孙学会, 等. 大型队列建设模式与运行机制及其启示[J]. 中国卫生资源, 2021, 24(6): 739-743.
- [6] 熊玮仪, 吕筠, 郭彧, 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.
- [7] Riboli E, Kaaks R. The EPIC project: Rationale and study design[J]. International Journal of Epidemiology, 1997, 26 (Suppl 1): 6-14.
- [8] Riboli E. Nutrition and cancer: Background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC) [J]. Annals of Oncology, 1992, 3(10): 783-791.
- [9] Riboli E, Hunt K J, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): Study populations and data collection[J]. Public Health Nutrition, 2002, 5(6b): 1113-1124.
- [10] Ireland J, van Erp-Baart A, Charrondière U R, et al. Selection of a food classification system and a food composition database for future food consumption surveys[J]. European Journal of Clinical Nutrition, 2002, 56(Suppl 2): 33-45.
- [11] Slimani N, Deharveng G, Unwin I, et al. The EPIC nutrient database project (ENDB): A first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study[J]. European Journal of Clinical Nutrition, 2007, 61(9): 1037-1056.

Research on the organization and management mode of distributed large cohort based on multi-agent cooperation: taking EPIC-Europe as an example

SUN Xuehui¹, LI Wei², ZHOU Sheng², XU Ping², XU Li^{2*}

1. School of Life Science, Fudan University, Shanghai 200433, China

2. Shanghai Information Center for Life Science, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China

Abstract Large cohort study has become an important open scientific research infrastructure and health decision-making support platform in the field of biomedicine and public health. With increasingly highlighted significance, large cohort construction has been laid out in many countries. In this paper, by focusing on the organizational mechanism, funding source, construction mode, standardization scheme, resource management and sharing strategy, etc. of the European Prospective Investigation into Cancer and Nutrition (EPIC-Europe), the key to its successful implementation and operation was deeply analyzed, in order to provide experience reference for the construction of a comprehensive, shared and open large cohort with multi-agent collaboration. Results show that continuous input and long-term scientific research mechanism guarantee is indispensable to the construction of multi-agent collaborative alliance distributed large-scale cohort, and its sustainable and high-quality construction also requires forward-looking top-level design, a coordinated and unified management mechanism, standard and comprehensive data and biological sample collection, and a sound policy guarantee system. The following suggestions are made: the government should take a leading role in the planning and scientific design to establish an integrated large-scale cohort organization and management system; special projects with long-term stable support should be set up to ensure the continuous and high-quality construction of large cohorts; a standard and unified construction and management scheme need to be established to ensure the standardization and systematization of large cohort resources; the scale, type and quality of data resources should be improved to establish, open and share high-quality large cohorts; reasonable systems for the protection of human genetic resources and the management of intellectual property rights should be further explored.

Keywords multi-agent collaboration; large cohort; construction mode; management mechanism ●



(责任编辑 徐丽娇)