

科技发展研究

第 18 期

(总第 658 期)

上海科技发展研究中心

2023 年 09 月 20 日

编者按：当前，大数据和人工智能驱动的科学范研究范式不断深化，数据成为推动各领域深刻变革的关键引擎。在生命科学领域，汇集海量数据的大型人群队列研究平台已成为实现医学和健康领域重大突破的关键基础，是医学逐步走向精准化和个性化的必经之路。本期基于中国科学院上海营养与健康研究所（上海生命科学信息中心）的研究成果，对国内外大规模人群队列研究与建设现状、重大计划布局情况进行分析，研究提出有关建议，供参考。

以大型人群队列计划引领上海生命科学创新突破

大规模人群队列研究汇集了海量“全生命周期”健康大数据¹和生物样本资源，是支撑生命科学领域基础研究、医药研发和产业转化的关键基础设施和重大研究资源平台。一方面，基于大型队列进行多组学、环境、表型“大数据”分析是生命科学领域重大突破的关键，大型队列与人工智能技术相结合，可驱动新型药物靶点发现、智能诊疗技术、精准医学的发展；另一方面，大型队列汇集的大规模人群资源可为生物医学转化和产品验证提供关键支撑。国内外高质量大型队列资源平台建设已全面铺开，美欧英日等发达国家（地区）持续部署、

¹ 大规模人群队列包含了基因组学和其他多组学大数据，以及生活方式、环境等数据。

加大大型人群队列建设投入。我国及上海在相关重点专项的支持下，队列建设与研究也加速发展，但在规模化、精细度、标准化、开放性、影响力等方面仍需持续发力。

一、国际大型队列建设现状与做法

当前，政府主导建设国家级大型人群队列是主要发展趋势。英国早在 2006 年就开始实施 50 万人规模的国家级大型队列“英国生物样本库”（UK Biobank）项目，并于 2018 年规划启动 100 万人全基因组测序和 500 万人的大型队列“我们未来的健康”。目前，这三大计划已成为英国对疾病预防和早期诊断开展大规模基础设施建设的核心，并持续面向全球开放其基因组数据等资源。美国于 2016 年启动了“百万人群队列”项目（All of Us Research Program），当前招募规模已达 60 余万人。另外，加拿大“未来健康合作计划”、德国“国家队列”、日本生物样本库、沙特阿拉伯生物样本库等国家级高质量大型队列相继开展建设，部分规模在 50 万人以上。国际典型人群队列建设呈现以下特点：

组织机制方面，强化顶层设计、统一组织管理、各方协同合作。

（1）注重强化顶层设计。如英国 UK Biobank 的规划和设计用时近 7 年（2000-2007 年），明确了队列的研究目标、实施计划等，并通过两次试运行进行预研和评估；美国“百万人群队列”计划实施前，美国国立卫生研究院（NIH）发布了详细的论证报告，并先后于 2016 和 2017 年启动先导性试验和试点项目；多国协作的欧洲癌症与营养前瞻性调查项目（EPIC）正式启动前，国际癌症研究机构（IARC）在多个研究中心开展了系列试点研究，进行可行性分析以及饮食评估方法的探索，才最终确定研究方案。**（2）创新组织机制与运行模式。**大型队列计划的组织机制主要分为**政府主导一体化大型队列**和**多国**

协作联盟式大型队列两类。英国 UK Biobank 是典型的政府主导一体化大型队列，建立董事会集中进行管理和监督，并设立 20 余个分中心（站点）作为实地开展调查场所（图 1）；欧洲 EPIC 项目是联盟式大型队列，涉及欧洲 10 个国家的 23 个研究中心，由发起机构、协调中心、各研究中心代表组成的指导委员会总体协调，各国研究中心根据其目标自行开展调查研究工作（图 2）。

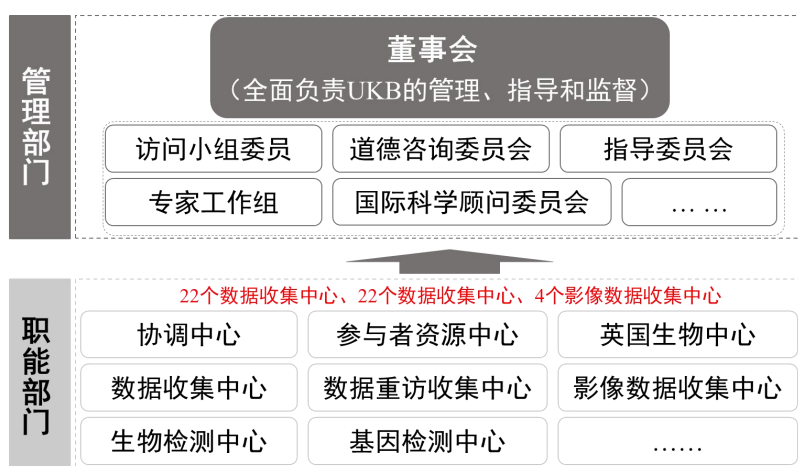


图 1 英国 UK Biobank 的管理模式



图 2 多国协作联盟式大型队列的管理模式（以欧洲 EPIC 项目为例）

投入模式方面，创新多元化资金筹集模式，统筹保障大型人群队列运行。大型队列建设耗资巨大，经费保障是非常重要的环节。（1）**政府稳定持续投入。**美国“百万人群队列”是政府主导建设的大型队

列，项目启动之初即获得了 1.3 亿美元的政府资助（占精准医学计划当年经费的 60.5%），美国政府还通过《21 世纪治愈法案》保障 10 年（FY2017-FY2026）的长期稳定支持，至 2023 财年，美国政府已通过多个渠道向该项目投入 31 亿美元。**（2）多元化筹资模式拓宽资金来源渠道。**英国 UK Biobank 的资金来源渠道更加多元化，主要出资方为英国医学研究理事会（MRC）等政府部门和英国维康信托基金会等慈善机构，美国再生元（Regeneron）、英国葛兰素史克（GSK）等公司也通过资金投入和技术支持等方式为其提供支持。欧洲 EPIC 计划采用了“项目分担，资金自筹”的机制，由“欧洲抗癌计划”（EAC）提供约 50% 的资金，世界癌症研究基金会（WCRF）、各参与国政府、机构或慈善组织等提供其余资金来保障队列运行。

具体建设方面，多手段持续提升队列数据精细度和质量。（1）**队列覆盖人群更广泛、多样，测量表型更丰富、多维、精细。**美国“百万人群队列”项目研究对象覆盖了不同地区、种族和收入水平的人群，采集资源类型除基本健康调查数据外，还开展组学分析，并引入个人智能穿戴设备数据、医疗理赔数据、环境暴露数据等。英国 UK Biobank 在完成基线调查后，又先后开展了针对全部 50 万参与者的全基因组测序、5 万余参与者血浆蛋白质组学分析等，不断提高其数据的规模、类型、质量和精细度。（2）**资源采集和管理更注重标准化、规模化和集约化。**一方面，采用统一流程进行标准化，制定人群招募及资源采集、存储、管理的统一规范和标准。对各环节进行严格质控、清洗和规范化处理，并通过建立数据校正机制等有效保证队列资源的标准化、规范化和系统化。另一方面，对资源进行集中存储与科学管理。英国 UK Biobank、美国“百万人群队列”计划均建立了统一的生物样本库与数据库对资源进行中心化存储和管理。欧洲 EPIC 计划则采用各研究中心独立保存、总协调中心建立中央库进行镜像存储的

方式进行整体维护与协调。

开放共享方面,通过及时全面的数据开放提高影响力和利用率成为共识。英国 UK Biobank、欧洲 EPIC 等的研究数据和生物样本资源均对全球科研人员开放,已支撑了大量研究的开展。美国“百万人群队列”计划虽然尚处于招募阶段,但该队列采取边建设、边分析、边应用的策略,及时对外开放使用,目前已发布了 25 万人的基线调查数据及全基因组测序数据,并依托“百万人群队列”项目开展精准营养研究项目。从影响力来看,基于英国 UK Biobank、欧洲 EPIC、美国“百万人群队列”计划等的数据分析已产出了大量高影响力研究论文,并提出了多个公共卫生政策建议,为本国乃至全球医疗卫生管理提供了重要支撑。

二、国内及上海队列建设现状与问题

我国研究机构自 20 世纪 70 年代开始自主开展人群队列研究,取得的多项原创性成果为我国疾病的预防和控制奠定了基础。

建设进展方面,我国目前已建设 400 余个不同规模、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。“十三五”期间,国家重点研发计划“精准医学研究”专项布局了大型人群队列建设与研究任务,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,为我国大型队列建设搭建框架。

存在问题方面,由于缺乏整体规划和持续支持,我国尚未建成国家级大型队列,已经建设的几个大型队列相较于国际趋势来看,存在如下问题:**(1) 队列建设缺乏顶层科学设计与稳定支持。**相同相似

队列重复建设，造成资源的浪费，各队列研究的开展模式仍以项目支持为主，导致队列研究无法长期持续开展；**（2）人群多样性覆盖不够、采集数据类型相对单一，规范化和开放程序不高。**队列的广度、深度和维度有待进一步提高，不同队列间的联合运行和共享机制不完善，限制了队列资源的共享和利用；**（3）队列研究的成果产出和影响力有待提高，**缺乏具国际影响力的高质量队列。

上海现状方面，队列研究开展较早，具备基础优势，亟待通过**高质量队列建设引领我国相关领域发展。**上海在国内较早开展队列研究，目前已初步建立了多个具有一定影响力的队列，包括样本量 7.5 万例的上海女性健康队列（SWHS）、样本量 6 万例的上海男性健康队列（SMHS），样本量 26 万例的中国糖尿病患者肿瘤发生风险的纵向研究（REACTION）、目标样本量 10 万例的上海老年脑健康队列等专病研究队列，以及目标样本量 10 万人的国内首个大规模出生队列研究“生命早期千天计划”等，集聚了大量健康和各类疾病人群样本和数据资源。同时，上海汇集顶尖的医疗资源，并具有全国领先的技术平台设施储备、医疗卫生体系、医疗卫生信息化水平、信息采集与分析平台和企业，为上海率先建设一体化的高质量大型队列奠定了坚实基础。

三、有关建议

为适应医学研究发展的新需求，结合国际布局及典型大型队列建设模式的经验剖析，建议上海**从政府层面进行科学顶层设计和长远前瞻规划，率先探索、持续支持，建设符合数据密集型生物医学研究范式的规模化、标准化、高质量上海市级大型人群队列，**为上海乃至我国生命科学研究的源头创新与产业发展提供强有力的支撑。

一是政府主导规划、科学设计和统一组织管理大型队列的实施。

从政府层面上进行统筹规划，组织各相关方进行顶层设计，深入论证适合上海的一体化大型队列组织管理模式、标准工作流程与可持续运行机制、以及有效的知识产权保护政策等，形成大型队列建设的路线图和实施方案。**组织和管理上**，实施的一体化组织和管理机制，充分调动和集合全市力量共同推动项目实施，提高大型队列建设的工作效率和可操作性。**建设模式上**，“从头建设”符合生物医学发展需求和目标的高质量大型人群队列，以标准统一、特色保留、促进共享等形式，整合、吸纳现有符合标准的队列协同建设，并遴选高质量的特色队列、深度分析队列等作为子队列重点发展。

二是长期稳定支持，确保大型队列的持续、高质量建设。设立上海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，并由政府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。

三是提高数据规模、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，建议上海一方面**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、高精度的数据收集，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，**要统一标准方案与规范管理**，对生物样本与数据资源采集、存储、管理过程进行严格质控，保证队列资源的标准化、规范化和系统化，提高大型队列资源平台的有效数据规模和质量。同时，**要推进大型队列资源平台的开放共享**，提升队列资源的利用率和影响力，实现其价值的最大化。

执 笔: 许 丽、李 伟、徐 萍

编 辑: 汤天波

责任编辑: 汤天波 编 辑: 姚景怡 联系电话: 53300825 传真: 64315005
地 址: 淮海中路1634号101室 邮政编码: 200031 电子邮件: fzzx@stcsm.sh.gov.cn