

基于自然人群队列的电子数据采集系统构建

杨 豪 王觅也

(1 四川大学华西医院信息中心 成都 610041 2 医疗信息化技术教育部工程研究中心 成都 610041)

〔摘要〕 介绍队列研究概念及趋势,从业务流程、网络架构、功能模块等方面阐述基于自然人群队列的电子数据采集系统设计与构建,分析系统应用成效与不足,为相关研究提供参考。

〔关键词〕 自然人群队列;信息系统;精准医学

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.05.012

Building of Electronic Data Capture System Based on Natural Population Cohort YANG Hao, WANG Miye, 1Information Center, West China Hospital of Sichuan University, Chengdu 610041, 2Engineering Research Center of Medical Information Technology, Chengdu 610041, China

〔Abstract〕 The paper introduces the concept and trend of cohort study, elaborates the design and building of Electronic Data Capture System (EDCS) based on natural population cohort from the aspects of business process, network architecture, functional modules, etc., and analyzes the application effectiveness and deficiencies of the system, so as to provide references for related study.

〔Keywords〕 natural population cohort; information system; precision medicine

1 引言

1.1 研究背景

高血压、糖尿病、肿瘤等慢性非传染性重大疾病已成为我国疾病负担、医疗费用过度上涨和威胁人民生命健康的重要公共卫生问题^[1],尤其在西部地区,老龄人口急剧增多、医疗卫生资源不足、利用效率不佳,加之各地区特殊的自然条件、文化习俗、生活方式等因素,导致一些地方病和重大慢病严重威胁居民健康^[2]。通过自然人群队列项

目建设,有助于了解当地居民健康全貌,带动当地医疗技术发展。采集生物样本,结合四川大学华西医院精准医学研究基础,探索发现重大疾病新的病因和危险因素,为科研团队提供数据支撑,推动医院临床研究能力及医疗水平提升,促进精准医学大数据标准化体系与共享平台、精准医学知识库构建^[3]。然而区域自然人群之间差异性较大、涉及区域范围广、社区采集点较多,加之各医联体医院之间信息系统不统一,数据标准不一致,给后期人员管理、数据采集和分析、随访跟踪带来很大困难^[4]。近年来信息化技术在医疗健康领域得到广泛应用,基于信息技术建立安全、稳定的自然人群队列电子数据采集系统(Electronic Data Capture System, EDCS),成为队列研究必然趋势。

〔收稿日期〕 2020-07-22

〔作者简介〕 杨豪,硕士,助理工程师,发表论文4篇;
通讯作者:王觅也,硕士。

1.2 队列研究概述

队列研究是将人群按是否暴露于某可疑因素及其暴露程度分为不同组,追踪各组结局,比较不同组之间结局差异,从而判定暴露因素与结局之间有无因果关联及关联大小的一种观察性研究方法^[5]。在自然人群中开展队列研究的主要目的是考察普通环境暴露、膳食因素、生活习惯和遗传因素等之间交互作用,通过“健康-疾病-愈后-随访”全生命周期观察,建立标准化、高质量的大型队列研究。目前队列研究趋势是从小样本、单中心的零散研究过渡到大样本、多中心协作研究;从专注于单一研究领域到多学科、多组学交叉融合;从有限时点的生物样本信息采集到全过程、多类型样本信息采集;从单一维度数据处理、存储和分析到海量多维数据交汇、处理和解析。传统人工方式队列成员管理无法支持大样本和长周期队列研究项目。

1.3 EDCS 构建意义

辅助自然人群健康队列研究,收集人群健康基线信息,了解当地居民健康全貌,建立标本库,探索当地重大疾病及罕见病新的病因和危险因素,惠及当地民众。通过提供预防干预或诊疗服务,进一步探索自然人群中重大疾病及罕见病诊疗情况及预后,为基层医疗卫生机构临床决策提供有价值的科

学证据,推动医联体成员单位临床研究能力及医疗水平提升^[6]。

2 系统设计

2.1 业务流程

由项目组在本院医联体周边社区建立自然人群信息采集定点,受试者签署知情同意书后经身份证识别纳入 EDCS,调用医联体医院相关检验信息系统(Laboratory Information System, LIS)生成检验条码。引导受试者定点采集血样、尿样等,在当天将样本送至医联体医院检验科进行检验。待受试者完成采样后在 APP 端填写调查问卷和体格检查信息。医联体医院将各指标检验结果推送至 EDCS,由稽查员审核问卷信息并听取录音核对问卷,具体流程,见图 1。初步计划完成 80 000 人信息采集,采用序贯整群抽样,完成基线调查,包括:问卷调查(一般人口学特征、生活行为方式、饮食习惯、健康状况、既往诊疗信息等);体格检查(身高、体重、腹围、血压、肺功能);实验室检查和标本采集(空腹静脉血和尿液等)。为每个受试者提供免费体检服务,根据检验结果为每个队列人员提供早期干预或诊疗服务。此外为受试者提供 5 年及以上时长随访服务,有助于提升当地群众身体健康水平。

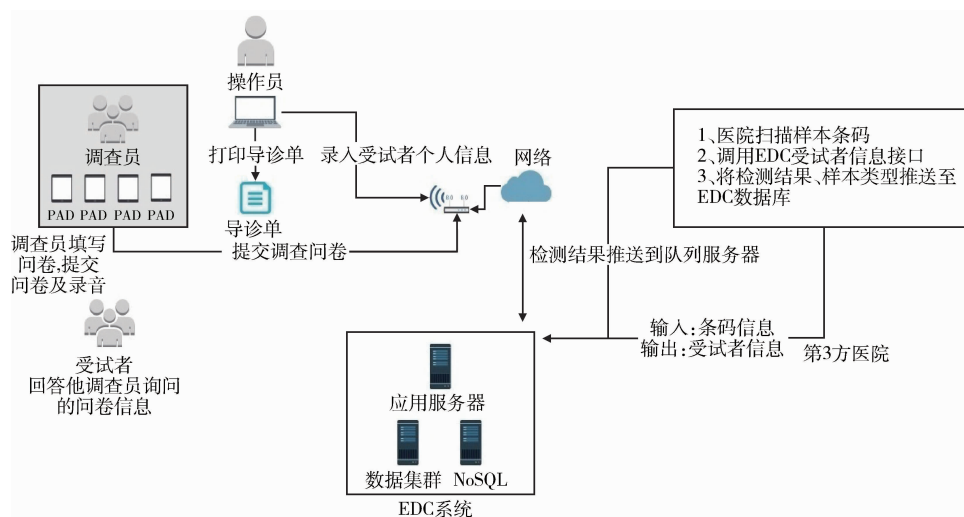


图 1 业务流程

2.2 网络结构

EDCS 主要包括 Web 端和移动端 APP 两部分。系统应用和数据库服务器均部署在院内机房,考虑到系统实用性、易用性和数据服务安全性,主要采

用虚拟专用网 (Virtual Private Network, VPN)^[7],以保障整个系统在较为安全的网络环境内运行。此外本系统涉及与医联体医院 LIS 接口对接和数据回传,其主要通过网络专线直连方式将检验结果回传到院内 EDC 数据库服务器。系统网络结构,见图 2。

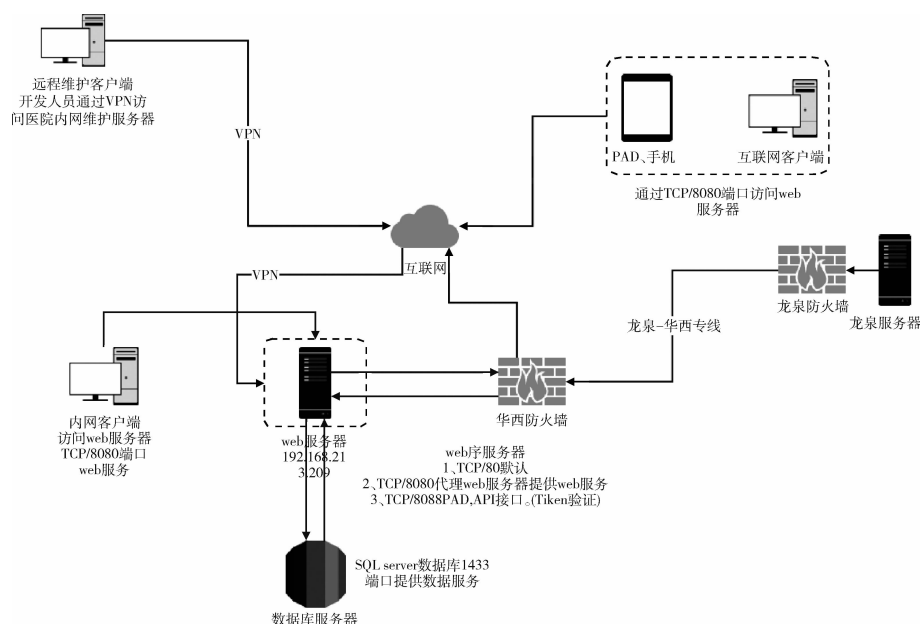


图2 网络结构

2.3 功能

2.3.1 信息管理平台 包括多中心应用、多层次授权、权限管理 3 个部分。多中心应用支持一个队列在多个不同地域中心同时开展调查工作,有助于提高队列研究效率,可在较短时间内搜集足量受试者,同时提升队列丰富性和多样性。多层次授权支持设置医院管理员、中心管理员、社区调查员等多级岗位,各岗位具有系统不同使用权限。权限管理包括页面访问、按钮操作控制等,根据角色岗位进行权限分配,医院管理员可对用户角色进行内部权限设置。

2.3.2 问卷采集 提供多种逻辑校验机制以满足临床业务需求,支持跳题逻辑设定,根据选项配置,在选项后面选择要调整的题目(如性别为男,会跳过月经史等相关问卷);支持勾选逻辑校验题并自定义问卷中的试题逻辑(如是否喝酒,勾选“是”会跳出饮酒时长选项,勾选“否”会自动跳

过)。在答题环节中为帮助受试者理解,题目除文字说明外支持图片说明。支持设置问卷计算模型,如通过受试者填写情况计算焦虑得分、心情指数等。支持题库配置,方便多套问卷重复使用;可设置单选、多选、单选填空、矩阵多选等多种题型。可预先设置问卷模板,1 个问卷模板可对应多份问卷。为保证问卷质量,在受试者信息采集过程中支持录音功能,由中心管理员统一审核问卷。为方便科研和管理需求,问卷支持后台按 ID、填写时间等条件导出。

2.3.3 质量控制 包括答卷汇总、初始答卷、退回答卷、审核答卷等功能。答卷汇总是各种状态答卷集合,可查看答卷基本内容、录音标识、答卷状态和操作权限等信息。中心管理员可审核答卷,对其进行查看、修改、审核、退回和作废等操作。各中心人员可以看到被退回的答卷信息,根据退回时间和原因修改答卷,修改完成后再次提交审核。

2.3.4 随访管理 受试者随访是由项目组成员通

过 PAD 端登录系统对受试者按问卷内容进行访视。整个移动端与 Web 端管理系统实时对接, 根据系统分配的账号登录进入问卷选择界面, 采取调查员 1 对 1 方式实时获取待答问卷。此外可通过切换受试者 ID 号调查其他对象。为保证数据质量, 在随访过程中增加录音和拍照功能, 便于后期管理员进行质控比对。移动端可同步查看调查受试者待答和已答问卷等信息以及受试者检验结果。

2.3.5 统计管理 根据纳入成员地区, 以地图形式进行数据可视化展示, 可直观观察受试者区域分布。主要以折线图方式进行纳入增量统计, 支持按中心、纳入时间等维度展示队列成员增量或总量数据。可按项目组统计受试者不同阶段问卷、病案和临床信息完成情况。此外可根据不同项目组需求采集受试者不同类型样本, 统计样本目标数、完成率等指标。

3 应用效果

项目组在本院医联体医院周边社区开展相关试点, 截至 2020 年 1 月纳入受试者约 5 000 名, 完成受试者基线问卷调查和检验样本采集。在 2021 年启动下个医联体区域队列项目。EDCS 管理平台上存放着大量自然人群健康数据, 包括日常体检、随访问卷等, 通过这些数据可以对自然人群高发病进行研究, 了解其发病机制, 从而做好相关预防措施; 收集自然人群医疗和非医疗数据, 对碎片化信息进行统一管理和全面分析, 实现多平台数据和资源共享, 为疾病精确诊断提供有力支撑。若受试者在入组后产生就诊行为, 可结合 EDCS 数据和医院业务系统就诊数据, 精确查找患者患病原因和治疗靶点, 有针对性地提供治疗方案, 实现疾病快速诊疗。然而现有系统存在一定不足, 目前只对接当地医院检验系统, 实现检验结果实时回传, 但受试者 CT 数据受网络带宽影响, 未能实现实时回传。因涉

及不同医联体医院的对接, LIS 不同, 需开发多套接口, 严重影响项目进度和成本。鉴于数据安全性, 医联体医院数据回传都是以专线方式进行, 但部分医院同本院无网络专线连接, 只能采用离线方式, 无法保证数据质量和效率。

4 结语

通过 EDCS 采集的基线数据和随访问卷信息最终回传到院内大数据集成平台^[8], 整合受试者所有医疗信息, 为其提供全生命周期健康看护, 提升健康水平, 减少无效和过度医疗, 节约医疗费用支出。随着队列研究深入, 数据库规模逐渐扩大, 结合大数据技术建立大规模疾病预警、诊断、治疗与疗效评价的生物标志物、靶标、制剂实验和分析技术体系, 形成重大疾病精准防治方案和临床决策系统, 促进临床研究能力及医疗水平提升。

参考文献

- 1 张祝琴, 刘德培. 国外重大慢性非传染性疾病防治对策研究 [J]. 中国工程科学, 2017, 19 (2): 112-116.
- 2 马璐, 唐萍, 夏煜坤. 西部 12 省医疗卫生资源配置效率分析——基于 DEA—Malmquist 指数分析法 [J]. 行政事业资产与财务, 2019 (8): 41-42.
- 3 孙荣国, 贾晓蓉. 我国精准医学的探讨 [J]. 华西医学, 2016, 31 (12): 2076-2078.
- 4 李婷婷. 基于精准医学的自然人群队列信息管理平台建设 [J]. 现代信息科技, 2020, 4 (6): 124-126.
- 5 杨景丽, 黄文雅, 黄佩瑶, 等. 中国队列研究建立和发展现状 [J]. 中国公共卫生, 2019, 35 (10): 1393-1399.
- 6 范美玉, 陈敏. 基于大数据的精准医疗服务体系研究 [J]. 中国医院管理, 2016, 36 (1): 10-11.
- 7 陈海倩, 张丽娟, 赖宇阳, 等. 基于 SSL VPN 技术的电力安全网关设计与实现 [J]. 电子设计工程, 2020, 28 (13): 97-100, 107.
- 8 王觅也, 郑涛, 李楠, 等. 医疗大数据集成及应用平台体系构建 [J]. 医学信息学杂志, 2019, 40 (8): 37-42.