

· 基础理论与方法 ·

电子数据采集系统 REDCap 在大型自然人群队列研究中的应用

高翔宇¹ 米白冰¹ 武文韬¹ 吴晨璐¹ 李敏敏¹ 刘叶舟¹ 江浩² 王鹏博³ 曾令霞¹
党少农¹ 颜虹¹

¹ 西安交通大学医学部公共卫生学院流行病与卫生统计学系 710061; ² 西安交通大学医学部公共卫生学院 710061; ³ 西安交通大学计算机科学与技术学院 710061

通信作者: 党少农, Email:tjdshn@xjtu.edu.cn

【摘要】 队列研究是流行病学研究的基本方法之一, 随着心血管系统疾病等复杂性疾病病因研究的深入, 大型自然人群队列研究成为医学研究热点。如何高效、安全地采集研究对象的数据是队列建设过程中亟待解决的重要问题。REDCap 是一个开源、免费、高质量的临床研究数据采集与数据库构建管理系统, 在大型人群队列研究中有广泛的应用前景。本文介绍队列研究基线调查及随访数据采集和管理基于该系统的解决方案, 从建立数据工作组、数据采集、队列随访、现场应用等多方面进行梳理总结, 为提高队列研究数据采集和管理效率、助力我国队列研究发展提供参考。

【关键词】 自然人群队列研究; 数据采集; REDCap

基金项目: 国家重点研发计划(2017YFC0907200, 2017YFC0907201); 国家自然科学基金(81230016)

DOI: 10.3760/cma.j.cn112338-20190924-00697

Application of electronic data acquisition system REDCap in large natural population-based cohort studies

Gao Xiangyu¹, Mi Baibing¹, Wu Wentao¹, Wu Chenlu¹, Li Minmin¹, Liu Yezhou¹, Jiang Hao², Wang Pengbo³, Zeng Lingxia¹, Dang Shaonong¹, Yan Hong¹

¹ Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an 710061, China; ² School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an 710061, China; ³ School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710061, China

Corresponding author: Dang Shaonong, Email: tjdshn@xjtu.edu.cn

【Abstract】 Cohort study is one of the basic methods used in epidemiological research. With the development of the etiological analysis of complex diseases such as cardiovascular diseases, large natural population-based cohort study has become a popular topic in medical research. In the process of cohort development, one of the important issues is to ensure the efficiency and safety on data collection. As a database management system, with open source, free clinical research data collection and high quality, REDCap can widely be applied in large population-based cohort studies. This article summarizes the baseline survey and follow-up procedures on cohort studies and introduces a REDCap-system-based solution for data collection and management. Contents on the establishment of data working groups, data collection, cohort follow-up methods and field application are also discussed in this paper, in order to improve the efficiency of data collection and management in cohort study to help the development of cohort study in China.

【Key words】 Natural population-based cohort study; Data collection; REDCap

Fund programs: National Key Research and Development Program of China (2017YFC0907200, 2017YFC0907201); National Natural Science Foundation of China (81230016)

DOI: 10.3760/cma.j.cn112338-20190924-00697

队列研究通过直接观察暴露于某因素不同状况人群的结局来探讨该因素与所观察结局的关系, 由

于研究暴露和结局的时间顺序合理, 其检验病因假设的能力较强, 广泛应用于人群疾病研究中, 是流行

病学领域的一种基本研究方法^[1]。随着人类复杂性疾病(如心血管系统疾病、肿瘤、代谢性疾病等)病因研究的逐渐深入,以往基于单一的环境因素或遗传因素的流行病学病因研究已经不能满足需要^[2]。

大型人群队列研究是通过一定时间跨度内对足够人群进行动态追踪随访,采集反映疾病发生、发展、转归全过程的大规模人群生物学样本和流行病学数据并加以深入分析,识别基因-环境交互作用,可充分揭示疾病发生、发展过程的内在规律,提供疾病风险预测、个性化治疗方案等防治策略,故数以万计、十万计甚至百万计样本量的大型人群队列研究成为医学研究热点,包括我国在内的多个国家已陆续开展一系列的大规模人群队列研究。此类研究需要采集大规模人群的详细暴露情况和多次随访数据,要求大型队列研究的数据容量大、时效性强、质量高、数据信息以及数据结构多样^[3],这对数据的采集、质量控制以及数据库的管理等方面都提出了更高的要求。

队列研究在确定研究对象之后,需要多次采集基线调查和随访监测数据并建立数据库存储。目前,国内的队列现场调查的数据采集方法主要有2种^[4],一种是纸质调查形式,即通过纸质版调查问卷和各种表格采集信息,再进行电子化录入。该方法耗时耗力、效率低且不环保,无法保存工作人员操作变更记录,难以进行质量控制以保证数据安全;另一种方法是电子化调查形式,主要是由队列研究团队定制专用数据采集系统,通过电子终端设备直接采集数据,如结直肠癌流行病学数据采集系统^[5]、老年糖尿病随访调查系统^[6]等。据研究,与纸质数据采集方法相比,电子数据采集方法可节省49%~62%的成本^[7],该方法主要弊端在于不同队列项目的系统设计、架构以及研发水平各异,较难实现数据兼容共享,且各个系统所需的数据采集和管理功能类似,反复开发、分别进行系统维护也造成极大的资源浪费。有鉴于此,范德堡大学(Vanderbilt University)的保罗·哈里斯(Paul Harris)教授领导团队开发基于网络的开源应用系统REDCap(Research Electronic Data Capture),该系统设计完善、功能强大、管理方便,非常适用于高效、安全地采集队列研究的数据^[8]。截至2019年9月12日,该系统已被131个国家、3 667所机构用于开展7.5万余项临床科学研究^[9],其在大型人群队列建设和随访维护过程中的作用和价值值得深入研究。本文将介绍REDCap系统在大型队列研究的数据采集、队列随访、数据库构建方面的应用,

为我国大型自然人群队列研究项目提供方法和具体实践上的参考。

REDCap的通用流程:REDCap系统(最新版本9.5.0)主要应用于临床研究电子化数据采集与数据库构建管理^[10],该系统的开发始于2004年,主要应用于临床研究,经过十余年的完善,现已被广泛应用于多个研究类型,包括横断面调查、临床试验、队列研究等,内容涵盖临床实践质量评估、效果评价、基础科学研究、患者调查、临床决策等^[11]。REDCap的界面交互友好、操作直观且可重复使用,依托REDCap系统进行大型自然人群队列研究的通用流程图见图1。

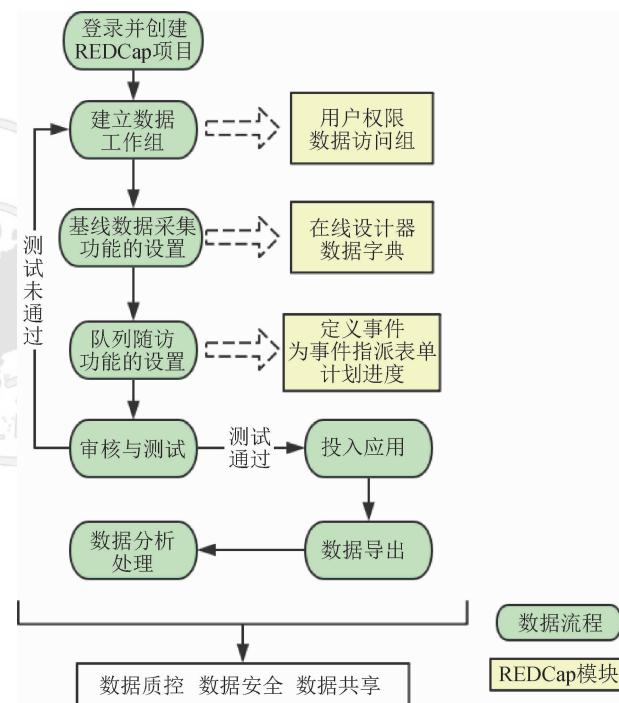


图1 REDCap进行大型自然人群队列研究流程图

大型人群队列的建设,常伴随产生海量的数据。数据种类多种多样,例如问卷数据、标本数据、随访数据以及实验室检测结果等。这些数据的高效采集以及后续的规范化管理都是队列研究需要解决的问题。

基于REDCap的大型自然人群队列研究解决方案:

(1)建立数据工作组:大型自然人群队列研究的样本量大,数据一般不是来源于单一的样本采集点,涉及多中心协同工作的问题。以西北区域自然人群队列(China Northwest Cohort, CNC)为例,由于项目涉及陕西、宁夏、新疆、甘肃和青海5个省份、涵盖12万

多民族人群,且需要完成基线调查、生物样本采集和长期随访,多中心数据采集流程的一致性尤为重要^[12],若一致性控制不好,考虑到队列研究长期随访特点,采集过程中的一些微小错误逐步累积放大,将严重增加数据质量控制的难度,降低后续分析的质量。为更好采集和管理数据,需要将多中心参与数据采集、处理等工作的所有成员纳入系统,即成立“数据管理工作组”,成员应该包括科研人员、统计学家、数据库管理员、样本库管理员、数据采集人员和IT成员等。项目实践过程中,项目组依托 REDCap 提供的“用户权限”和“数据访问组”这两个管理模块建立数据工作组,有效提高数据采集流程一致性。

第一步,CNC 将项目架设搭载在 REDCap 系统,创立账号及项目概况。通过“用户权限”模块将各课题组成员纳入系统,分别对其授予不同访问权限,包括是否允许该用户进行项目设计、是否允许其进行数据导出等权限设置以便于数据管理(图 2),更好地保障数据安全。

第二步,使用“数据访问组”模块将设定好权限的各中心用户按照其所属的课题分配进入不同工作组,完成自上而下的架构(图 3)。在同一个 REDCap 项目中,各个省份不同分中心的数据工作组使用陕西课题组总设计上传的同一套数据采集工具,也可以有力保证数据采集流程的一致性。

(2) 基线数据采集功能的设置:传统的纸质调查在大型自然人群队列研究中显得尤为不便。以 CNC 陕西城乡队列为例,我们在前期的预调查中发现,由于需要采集研究对象“社会人口学特征”“生活行为”“膳食营养与营养素服用情况”等方面信息,调查问卷名目繁多,需要提前打印装订成册、在队列现场填写回收、同时安排专人统计问卷的填写进度并对填写质量进行把关。现场调查完成后,还需建立数据库,进行双录入和逻辑检错。整个过程步骤繁琐,消耗大量人力、财力,数据采集、清洗流程长,质控难度大,从调查完成到数据库建设完成动辄数月,难以适应大型队列的建设要求。采用 REDCap 系统进行基线数据采集,首先要对基线数据采集功能进行设置。

系统内设的在线采集调查问卷设计方法有两种:“在线设计器”与“数据字典”。如果调查问卷较为简单,研究设计者通过浏览器启动联机“在线设计器”进行点选操作,给定每份问卷统一调查编码 record ID 作为唯一标识,然后根据不同调查问题的答案类型选择不同的字段类型,如问题的答案是字符串型变量、短字符串、短文本或者日期变量,选择文本框字段,也可设置是否为必答题目以及对姓名等识别信息作标识符,进行初步的质量控制。对调查问题某些选项进一步调查,可使用逻辑分支按钮进

数据录入权限				
	无访问权限	只读	查看及编辑	编辑调查结果
一、一般情况 (调查问卷)	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
二、饮茶情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
三、饮酒情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
四、吸烟情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
五、饮食情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
六、被动吸烟和室内空气污染	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
七、个人及家庭健康状况 (调查问卷)	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
八、体力活动情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
九、女性生育史情况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
十、睡眠及情绪状况	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
十一、体格检查记录	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	

图 2 “用户权限”模块对用户设置权限

一步设置；如果调查问卷较为复杂，系统提供了更方便操作的“数据字典”用于批量设置，即在 Excel 表格中编写好所有调查问题的变量名、选项等，在完成数据采集工具的设计后即时上传并启用。

在具体实践中将“在线设计器”与“数据字典”两种方法结合，完成基线数据采集功能的设置，为后续电子化采集做准备。两种在线问卷设计方法均可完成文本类、单项/多项选择题在内的各种问题设置，也提供逻辑分支(Branching Logic)进行问题的跳转设置。与传统纸质调查问卷相比，REDCap 系统在调查对象允许的情况下，自动通过移动终端采集被研究对象所处海拔、经纬度等位置信息，这是传统纸质问卷难以实现的。系统中，每个研究对象有唯一标识的 record ID，方便与采集的生物样本编码一一对应。见图 4,5。

(3) 队列随访功能的设置：完成研究对象的确定、招募及基线调查后，即进入长期随访阶段。随访是了解研究对象中结局影响因素暴露的长期变化趋势、确定研究结局状态的重要手段，是本项目能否最终取得成功的关键所在。大型自然人群队列研究的随访包括两种方式^[13]，一是利用当前运行的医保管理系统、妇幼保健信息系统、公安户籍管理系统等各类监测系统、常规工作中形成的资料或数据库，获取全部队列成员的结局(如发病、死亡事件等)信息，即常规监测。但是，这种方法存在一定程度的漏报，可能难以采集全部成员的信息，故需要其他具体到研究对象个体的随访方式来进行补充。二是社区定向监测，将队列成员的名单提供给研究社区街道、居委会或乡镇、村的相关工作人员，定期联系研究对象，主要用来确定迁移失访状况以及常规监测的漏报情



The screenshot shows a table titled '创建新组: 给那些可以分配角色的用户添加新的数据访问组。' (Create New Group: Add new data access groups for users who can be assigned roles). The table lists five data access groups: Ningxia Group, Xinjiang Group, Gansu Group, Shaanxi Group, and Qinghai Group. Each group has a corresponding user listed under '组用户' (Group User): wwt (wentao wu), wuli8503826 (进真 肖), project_manager (baibing mi), and c02958c731 respectively. The table includes columns for '数据访问组' (Data Access Group), '组用户' (Group User), '组中记录数' (Number of records in group), '唯一的组名 (自动产生)' (Unique group name (automatically generated)), '组ID' (Group ID), and '删除组?' (Delete group?).

创建新组: 给那些可以分配角色的用户添加新的数据访问组。					
	+ 输入新的组名	添加组			
分配用户到组: 用户可以被分配到任何的数据访问组。					
分配用户	-- 选择用户 --	至	[无分配]	分配	
数据访问组	组用户	组中记录数	唯一的组名 (自动产生)	组ID	删除组?
宁夏数据组		0	6fc97b701a	34	×
新疆数据组	wwt (wentao wu)	0	de367d3fd3	37	×
甘肃数据组	wuli8503826 (进真 肖)	0	ae9955a18c	36	×
陕西数据组	project_manager (baibing mi)	0	59509b5e56	38	×
青海数据组		0	c02958c731	33	×

图 3 “数据访问组”模块将纳入用户分配进入不同数据工作组



The screenshot shows the 'Online Designer' interface with four survey questions:

- Question 4.7: "你大约从几岁开始养成每天或基本上每天都吸烟的习惯?" (When did you start smoking every day or almost every day?) with a dropdown unit: '岁' (Year).
- Question 4.8: "在刚开始养成吸烟习惯时, 你主要吸那一类烟?" (What type of cigarette did you mainly smoke when you first started smoking?) with options: 香烟 (Cigarette), 其他烟制品 (如旱烟、水烟、烟斗、手卷烟等) (Other tobacco products (such as旱烟, water pipe, cigarette holder, hand-rolled cigarette)), and 混吸香烟和其他烟制品 (Mixed cigarette and other tobacco products).
- Question 4.8a: "自从开始吸烟后, 你是否一直吸烟并且从未有过连续1个月以上的间断时间?" (Since starting smoking, have you ever stopped smoking for more than 1 month at a time?) with options: 是 (Yes) and 否 (No).
- Question 4.9: "你目前或在戒烟前通常吸哪类烟?" (What type of cigarette do you currently or usually smoke before quitting?) with options: 过滤嘴香烟 (Filter-tipped cigarette), 无过滤嘴香烟 (Non-filter cigarette), 手卷烟/旱烟 (Hand-rolled cigarette/旱烟), 烟斗/水烟 (Cigarette holder/water pipe), and 雪茄 (Cigar).

图 4 “在线设计器”的设计页面

况,也可用来获取结局信息。实践过程中,多联合采用上述方式完成队列的随访工作。

CNC 整合已有健康档案、疾病监测网络、医院健康信息系统、社会保障与服务机构信息系统的全联接健康管理体系对研究对象进行长期、连续、动态的随访监测,随访主要采集常见慢性病、肿瘤、死亡以及迁移和失访等资料,这些资料后续可以通过“添加/编辑记录”模块以及“数据导入工具”模块导入 REDCap 系统中。具体随访功能设置如下。

第一步,研究设计者使用“定义事件”模块。将项目每次随访计划定义为一次“事件”,如在“第四年随访”是基线数据采集完成 2018 年 9 月 1 日算起 4 年后要进行的事件,在“定义事件”模块建立该事件“第四次随访”,相距时间天数填入“1460”,同时系统允许设置机动的偏移天数。见图 6。

第二步,设置“为事件指派表单”模块。因为同一队列项目的多次随访用到的调查问卷可能存在差异,我们可以通过该模块在系统中为该次定义的“事件”,如“第四次随访”指派需要填写的调查问卷,即“为事件指派表单”。

第三步,建立“进度计划”模块。以 CNC 为例,“起始日期”即基线数据完成的具体时间

2018-09-01,根据“定义事件”模块中设置完毕的基线采集完成与第四次随访的时间间隔可计算事件发生时间,时间范围即是“定义事件”模块中设置的机动天数。进度计划的具体时间点可以加入系统提供的“日历”模块,该日历可被项目用户共享,这些都有助于研究设计者更有规划地安排研究进程。见图 7。

(4) 现场应用:第一步,对整个数据采集流程进行测试和审核。为便于全流程的测试和审核工作,REDCap 系统设计为在未正式应用之前就可以生成调查链接进行预调查,对问卷的逻辑分支、采集到的数据格式等方面进行测试和审核,审核通过的项目才能“转入生产状态”,即进入基线数据现场采集阶段。

第二步,采集数据,问卷分发。调查人员使用在线问卷通过智能设备进行面对面督导填写,系统也提供了“调查分发工具”模块作为补充,该模块可直接将设计好的随访问卷生成网络链接或二维码,研究对象可通过多种设备实时填写相关问卷。工作人员可在系统后台实时查看填写进度,通过电话、微信等形式督促未填写、未完成的研究对象填写问卷,有效减轻其工作负担,提高效率。同时,考虑到交通不便的情况,将调查链接发送到具体邮箱或短信通知发送给村医等合作单位人员通过网络统一填写采集

B	C	D	E	F	G	H	I	J	K	L	M	N	O
Form Name	Section	Field Type	Field Label	Choices, Calculations	Field Note	Text Vali	Text Vali	Text Vali	Identifie	Branching	Log Required	Custom Alques	
1 dcdc_2d07		radio	4.1 你今天抽过烟吗?0, 是 1, 否								LH		
2 dcdc_2d07		radio	4.2 你现在多长时间抽0, 不吸烟 1, 偶尔吸 2, 选择后两项, 请转问题4.7								LH		
3 dcdc_2d07		radio	4.3 在以前, 你是否曾0, 从不吸 1, 偶尔吸 2, 选择后两项, 请转问题4.5								LH		
4 dcdc_2d07		radio	4.4 到目前为止, 你一0, 是 1, 否		选否, 直接跳转问卷第五部分						LH		
5 dcdc_2d07		text	4.5 你过去曾吸烟现在已不吸, 那你戒烟有多久	单位: 年	number	0	99				LH		
6 dcdc_2d07		text		单位: 月	number						LH		
7 dcdc_2d07		radio	4.6 促使你戒烟的一个0, 因为已经患有的疾病(请选择一项最主要的回答)								LH		
8 dcdc_2d07		text	4.7 你大约从几岁开始养成每天或基本上每天都	单位: 岁	number	0	99				LH		
9 dcdc_2d07		radio	4.8 在刚开始养成吸0, 香烟 1, 其他烟制品(如旱烟、水烟、烟斗、手卷烟等) 2, 混吸香烟和其他烟制品								LH		
10 dcdc_2d07		radio	4.8a 自从开始吸烟后0, 是 1, 否								[q4_08] = '0'		LH
11 dcdc_2d07		checkbox	4.9 你目前或在戒烟前0, 过滤嘴香烟 1, 无滤嘴如习惯混吸, 最多可以同时选择三项最主要的回答								LH		
12 dcdc_2d07													

图 5 “数据字典”的设计页面

分组 1: 前三次随访
分组 2: 第四次随访
+添加新的分组

分组名称: 第四次随访		重命名分组 2 删除分组 2			
事件 #	相距时间 天数	偏移范围 最小/最大	事件名称	自定义事件 标签 (可选)	唯一事件名称 (自动产生)
1	1460	-10/+10	第四次随访	fourth	8883e8bbe1_arm_2
<input type="button" value="增加新事件"/>		<input type="text" value="天"/> + <input type="text" value="0"/> <small>从其他单位转换</small>	<input type="text" value="此事件的描述性名称"/>	<input type="text"/> <small>自定义事件标签 (可选)</small> <small>例如:</small> <small>[visit_date], [weight] kg</small>	

图 6 “定义事件”模块的设置页面

图 7 建立进度计划的页面

数据,更加方便快捷。

第三步,数据整理与分析。与传统纸质调查问卷收集数据再进行电子化录入的方式不同,使用该系统在线问卷直接可收集到电子化数据,该系统也支持实时将研究数据导出为 CSV、SPSS、SAS、Stata 等多种软件适用格式用于分析。此外,该系统允许外部程序通过编程接口(API)远程连接到 REDCap,用于程序化地查询或修改数据,自动化执行指定项目的数据导入导出,马璐等^[14]已测试应用 R 统计软件直接调用 API,便捷地进行数据分析。此外,值得关注的一些重要问题如数据质控、数据安全与数据共享^[4,13],REDCap 也给出了相应的解决方案。

4. 应用优势:

(1)数据质控:与传统纸质调查面对面访谈、调查员询问的方式不同,采用一般电子数据采集系统开展基线调查,被调查者可以在任何地点填写在线问卷,在带来便利的同时,也使得调查过程不可控,导致答卷过程的规范性缺乏有效监控^[15],故大型人群队列研究的数据质量控制需重点关注,而 REDCap 系统针对数据质量控制提供了覆盖采集工具设计到项目正式运行的全程质控,主要分 3 个阶段实现。第一阶段,数据采集工具设计过程进行质控。该过程中,除了能设置问卷变量是否为必填字段、对输入时间格式进行严格规定以外,还可通过嵌入式功能规定录入数字的上下限用来检查数据范围

(如对成年人身高录入范围预设)^[10],当被调查者录入出错时进行提醒;可设定逻辑跳转规则,以保证在线问卷跳转正确。第二阶段,系统流程测试和审核环节进行质控。研究设计者在项目正式启用前,系统流程要求生成链接用于预调查试填,未通过审批的项目需进一步完善,这是设计阶段的最后一步质控。第三阶段,现场调查实时质量控制。REDCap 提供了多个模块用于数据完整性检查。如“记录状态在线统计表”模块用 6 种颜色的标识区分不同的问卷填写状况,如“不完整”“部分反馈的调查问卷”“完整反馈的调查问卷”等,便于对填写进度较慢的研究对象精准督促,对已完成问卷实时验收。“数据导出、报表及统计”模块用于调查期间日常审核,轻松查看数据报表和统计图表,包括研究对象的基本信息、调查问卷的填写完成情况以及对缺失数据的统计信息等,有效实现问卷质量和进度把控。此外,REDCap 系统内设了 8 种“数据质量规则”用于数据质控即“数据质量”模块,它也支持研究设计者自定义“数据质量规则”,对所有不符合规则的数据进行统计整理,有助于研究者进行数据质控。

(2)数据安全:根据大型队列研究数据安全技术规范,队列研究的数据安全从现场调查、数据处理到数据库安全等方面均有要求^[4],数据隐私保护问题越来越受到重视。由于纸质版资料保密性差,从问卷填写到录入再到保存,经手人员较多、较难留存在

线操作记录,且问卷需要库房储存专人看管,均不利于问卷的信息保密;而定制数据采集工具存在一些缺陷,按照安全规范的要求,开发的专用数据采集工具需要配备专门的安全管理员来保障数据的信息安全,但因第三方研发团队多为计算机领域人员,对医药领域数据安全规范少有全盘考虑,难以持续保证其安全性。

针对数据安全的要求,REDCap系统提供“日志”模块用以记录所有工作人员的线上操作记录。该系统从设计阶段已引入HIPAA等标准全面提高安全性,相比EpiData、Microsoft Excel或Microsoft Access等数据采集、管理工具更加安全,可以通过安全地网络协议([https](https://))从任何具有Internet连接和Web浏览器的设备访问,提供审计跟踪等功能^[8],数据安全可以得到有效保障。

(3)数据共享:由于大型队列建设周期长、难度大、耗资大,如何充分利用大型队列数据和生物样本以支持我国医学研究是队列建设需要考虑的问题。大数据时代,数据共享可将同样的研究问题放在一起比较,充分挖掘,实现数据有效利用有力支撑精准医学发展。由于各种原因(如不同项目之间的数据采集变量不同、采集的数据格式有差异等),已有研究数据难以共享。而REDCap提供的数据工具共享库(The Shared Data Instrument Library)有不同研究的数据字典^[16],研究者可以“自下而上”地上传自己的数据字典,也可下载其他项目的数据字典并应用于研究设计阶段。采用已经过实践验证的数据字典,如统一采用RAND生活质量量表(SF-36)的数据字典建立调查项目^[17],后续可一键实现对生活质量量表数据的共享。此外,采用统一数据字典采集的数据质量更有保障,也可促进不同站点之间的数据共享。

5. 总结与展望:REDCap系统操作界面友好,仅需简单的编程知识即可操作,项目生成周期快,适合多个项目同时开展,可以采集定量资料、定性资料、文本资料甚至图片音像等不同数据类型,其移动客户端可以将采集好的数据保存在终端设备,待网络环境方便以后,同步到服务器云端保存。因此,REDCap系统在大型自然人群队列项目的数据采集、管理和分析工作中具有较高的应用价值和前景。将传统纸质问卷、项目定制的数据采集系统与CNC应用的REDCap系统各方面比较见表1。

可通过申请获得REDCap开放源码使用许可,以此系统为核心架构,结合应用过程中发现的实际

表1 不同数据采集方式对比

类别	纸质问卷	项目定制采集系统	REDCap
问卷成本	有	无	无
问卷设计	不易调整	便于调整更改,跳转灵活,但无模板参考	便于调整更改,跳转灵活,且有模板参考
问卷分发与填写	分发运送及填写方式传统,仅适用于现场开展,大量印刷,成本较大且不环保	现场或远程填写一般通过本地客户端,系统跳转需要设计调试	调查者现场及远程填写方便,提供在线网页链接和本地客户端填写,跳转灵活
项目进度及填写质量	需人工统计并审核	取决于系统的搭建情况,水平参差不齐	专门模块查看项目进度及把控数据质量控制
数据安全性	数据易篡改或复制,问卷数据保存仍有问题,安全性差	取决于系统维护情况,需投入专业人员保证数据安全	数据访问权限便于设置,专业数据库服务器远端储存,数据安全性高
其他	数据采集完成,需进一步数据录入	不同项目的系统功能类似,反复开发,分别维护,资源浪费	功能多样,包括临床设计随机化、队列项目随访功能等需要专门培训学习

问题,可进一步开发符合我国具体情况的特色模块。如针对我国人口老龄化问题展开研究时,可以将智能语音阅读插件嵌入网站链接,可以便捷地采集阅读不便的老年人的信息,更人性化、高效化。REDCap系统虽然便捷,但研究设计者在实际使用时仍然需要进行专门系统学习。

随着“互联网+”和“大数据”应用快速深入人们生活的方方面面,医学研究领域越来越广泛地应用互联网技术。大型自然人群队列研究的无纸化问卷调查就是这样的时代趋势。本文基于国外临床研究数据采集与数据库构建管理系统REDCap,结合CNC队列研究实例,从建立数据工作组、基线数据采集、队列随访、应用优势等多个角度进行阐述,提出了大型自然人群队列研究中数据采集无纸化解决方案和全流程的质量控制措施,有效提高我国队列研究数据采集、管理及分析的质量,助力我国精准医学实践。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] 詹思延. 流行病学[M]. 北京:人民卫生出版社,2012:60-79.
Zhan SY. Epidemiology [M]. Beijing: People's Health Press, 2012:60-79.
- [2] 孙点剑一,吕筠,李立明. 流行病学超大规模队列研究——开启

- 21世纪人类复杂性疾病病因研究的钥匙[J]. 中华疾病控制杂志, 2013, 17(1):66-71.
- Sun DJY, Lyu J, Li LM. Mega cohort: a powerful tool for etiologic research on complex human diseases in 21st century [J]. Chin J Dis Control Prev, 2013, 17(1):66-71.
- [3] 余灿清,李立明. 大型队列研究中的数据科学[J]. 中华流行病学杂志, 2019, 40(1):1-4. DOI: 10.3760/cma.j.issn.0254-6450.2019.01.001.
- Yu CQ, Li LM. Data science in large cohort studies [J]. Chin J Epidemiol, 2019, 40(1):1-4. DOI: 10.3760/cma.j.issn.0254-6450.2019.01.001.
- [4] 中华预防医学会. 大型人群队列研究数据安全技术规范(T/CPMA 002-2018)[J]. 中华流行病学杂志, 2019, 40(1):12-16. DOI: 10.3760/cma.j.issn.0254-6450.2019.01.004.
- Chinese Preventive Medicine Association. Technical specification of data security for large population-based cohort study (T/CPMA 002-2018) [J]. Chin J Epidemiol, 2019, 40(1):12-16. DOI: 10.3760/cma.j.issn.0254-6450.2019.01.004.
- [5] 刘芳,邓宁,郑玲,等. 基于HTML5的结直肠癌流行病学数据采集系统[J]. 中国生物医学工程学报, 2013, 32(4):464-470. DOI: 10.3969/j.issn.0258-8021.2013.04.012.
- Liu F, Deng N, Zheng L, et al. An HTML5-based epidemiology data acquisition system for colorectal cancer [J]. Chin J Biomed Eng, 2013, 32 (4) : 464-470. DOI: 10.3969/j.issn.0258-8021.2013.04.012.
- [6] 吴嘉琦. 基于openEHR的科研数据管理系统设计与开发[D]. 杭州:浙江大学, 2016.
- Wu JQ. Design and development of an openEHR-based scientific research data management System [D]. Hangzhou: Zhejiang University, 2016.
- [7] Pavlović I, Kern T, Miklavčič D. Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study [J]. Contempo Clin Trials, 2009, 30(4):300-316. DOI: 10.1016/j.cct.2009.03.008.
- [8] Patridge EF, Bardyn TP. Research electronic data capture (REDCap) [J]. J Med Libr Assoc, 2018, 106(1):142-144. DOI: 10.5195/jmla.2018.319.
- [9] Vanderbilt. REDCap [EB/OL]. [2019-09-12]. <https://projectredcap.org/>.
- [10] Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap) — a metadata-driven methodology and workflow process for providing translational research informatics support [J]. J Biomed Inform, 2009, 42 (2) : 377-381. DOI: 10.1016/j.jbi.2008.08.010.
- [11] Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners [J]. J Biomed Inform, 2019, 95: 103208. DOI: 10.1016/j.jbi.2019.103208.
- [12] 吴美琴,吴宇航,赵丽,等. 多中心间的协同性决定队列生物样本的一致性[J]. 中国医药生物技术, 2015, 10(6): 489-493. DOI: 10.3969/j.issn.1673-713X.2015.06.003.
- Wu MQ, Wu YH, Zhao L, et al. Multi-center synergy determines consistency of cohort biological samples [J]. Chin Med Biotechnol, 2015, 10 (6) : 489-493. DOI: 10.3969/j.issn.1673-713X.2015.06.003.
- [13] 中华预防医学会. 大型人群队列终点事件长期随访技术规范 (T/CPMA 002-2019) [J]. 中华流行病学杂志, 2019, 40(7): 748-752. DOI: 10.3760/cma.j.issn.0254-6450.2019.07.002.
- Chinese Preventive Medicine Association. Technical specification of long-term follow-up for end point in large population-based cohort study (T/CPMA 002-2019) [J]. Chin J Epidemiol, 2019, 40 (7) : 748-752. DOI: 10.3760/cma.j.issn.0254-6450.2019.07.002.
- [14] 马璐,耿辉,贺海蓉,等. REDCap系统中文版的设置与实现 [J]. 中国循证心血管医学杂志, 2019, 11(2): 138-140, 144. DOI: 10.3969/j.issn.1674-4055.2019.02.03.
- Ma L, Geng H, He HR, et al. Setting and implementation of Chinese version of REDCap system [J]. Chin J Evid Based Cardiovasc Med, 2019, 11 (2) : 138-140, 144. DOI: 10.3969/j.issn.1674-4055.2019.02.03.
- [15] 杜江波,陆群,靳光付,等. 人群队列研究的数据管理与质量控制策略[J]. 中华预防医学杂志, 2018, 52 (10) : 1078-1081. DOI: 10.3760/cma.j.issn.0253-9624.2018.10.022.
- Du JB, Lu Q, Jin GF, et al. Data management and quality control strategies for population based cohort study [J]. Chin J Prev Med, 2018, 52 (10) : 1078-1081. DOI: 10.3760/cma.j.issn.0253-9624.2018.10.022.
- [16] Obeid JS, McGraw CA, Minor BL, et al. Procurement of shared data instruments for Research Electronic Data Capture (REDCap) [J]. J Biomed Inform, 2013, 46 (2) : 259-265. DOI: 10.1016/j.jbi.2012.10.006.
- [17] Hays RD, Morales LS. The RAND-36 measure of health-related quality of life [J]. Ann Med, 2001, 33 (5) : 350-357. DOI: 10.3109/07853890109002089.

(收稿日期:2019-09-24)

(本文编辑:万玉立)