

利用 REDCap 系统进行数据双录入及质量控制的方法

闫淼佳 赵芑 吴立晨 徐坤 颜虹 曾令霞 米白冰 党少农

西安交通大学公共卫生学院流行病与卫生统计学系 710061

通信作者:米白冰,Email:xjtu.mi@xjtu.edu.cn

【摘要】 医学研究中数据质量是决定研究成败的关键,因此做好数据的质量控制工作是确保研究成功的重要一环。REDCap 系统是医学研究领域一种新兴的数据采集系统,在国内外研究中逐渐得到应用。如何应用 REDCap 系统实现医学研究数据的双录入及质量控制,是研究者在研究中应用该系统时关心的热点问题。本文将从研究项目的创建、数据采集工具的设计、数据的双录入、数据核查和导出等方面,系统全面地介绍如何利用 REDCap 系统进行数据双录入及质量控制。

【关键词】 REDCap; 双录入; 质量控制; 医学研究

基金项目:国家重点研发计划(2017YFC0907200, 2017YFC0907201);中央高校基本科研业务费专项(xzy032020033);陕西省科学技术厅自然科学基金基础研究计划一般项目(2020JQ-090)

Method of double data entry and quality control by REDCap system

Yan Miaoja, Zhao Peng, Wu Lichen, Xu Kun, Yan Hong, Zeng Lingxia, Mi Baibing, Dang Shaonong
Department of Epidemiology and Health Statistics, School of Public Health, Xi'an Jiaotong University,
Xi'an 710061, China

Corresponding author: Mi Baibing, Email: xjtu.mi@xjtu.edu.cn

【Abstract】 In medical research, the quality of data is the key to success. Thus, data quality control becomes an important part of ensuring the research's high quality. REDCap system is an emerging data acquisition system in medical research, which is gradually applied in research at home and abroad. It is a hot issue to realize double data entry and data quality control in using the REDCap system, which researchers are concerned about when this system is supposed to apply. This article will systematically introduce how to use the REDCap system for double data entry and quality control from the aspects of research project creation, data collection tool design, double data entry, data checking and exporting.

【Key words】 REDCap; Double data entry; Quality control; Medical research

Fund programs: National Key Research and Development Program of China (2017YFC0907200, 2017YFC0907201); Fundamental Research Funds for the Central Universities (xzy032020033); Natural Science Basic Research Plan of the Shaanxi Province (2020JQ-090)

医学领域中大部分研究都需要收集大量数据,并依赖于对所收集数据的分析而展开^[1]。研究过程中的数据质量控制非常重要,决定了研究的质量和成败^[2-3]。以往,医学研究常采用纸质问卷收集数据,并应用 EpiData 软件进行数据双录入和质量

控制^[4]。EpiData 软件录入快捷、功能完备,是医学研究中广泛采用的数据管理和质量控制工具^[5],但该软件为单机版程序,仅可进行单中心数据质量控制,没有联网核查功能,不适用于多中心研究对数据质量控制工作发展的要求。越来越多的医学研

DOI: 10.3760/cma.j.cn112338-20200415-00574

收稿日期 2020-04-15 本文编辑 万玉立

引用本文:闫淼佳,赵芑,吴立晨,等.利用 REDCap 系统进行数据双录入及质量控制的方法[J].中华流行病学杂志,2021,42(5):918-922. DOI: 10.3760/cma.j.cn112338-20200415-00574.



究采用电子数据收集系统(Electronic Data Capture, EDC)收集研究数据,以保证数据质量和过程管理的规范^[6]。但商业 EDC 价格高昂、配置复杂、学习和使用成本高,难以在研究(尤其是探索性研究)中广泛应用^[7]。广大临床医生及科研人员急需一种低成本、方便易用的电子数据收集系统。REDCap (Research Electronic Data Capture)就是满足这一要求的工具,该系统由范德堡大学的 Paul Harris 教授团队研发^[8-9],具有易上手、便捷、安全、稳定等多种优点,近年来在国内外医学研究中得到广泛应用^[10]。

之前已有学者详细介绍了 REDCap 系统的自主架设方法^[11],但未介绍其数据录入方法。REDCap 系统数据录入方式多样,既可在已有纸质问卷的基础上,进行数据的电子化双录入及质量控制,也可开放给调查对象直接在线填写一次问卷,再由研究人员对问卷进行审核。且该系统应用范围广,通过不同的功能模块设置可实现单次研究数据、长期纵向数据等多种类型数据的录入。

目前医学研究中,收集纸质问卷后再进行数据电子化录入的方式应用较多,考虑到数据质量控制的重要性,本文将依托西北自然人群队列研究中膳食成分数据库建立过程中双录入的实践经验,针对性地介绍在已有纸质问卷基础上,利用 REDCap 系统实现研究数据双录入及质量控制的方法。

1. 研究平台:基于西安交通大学开放研究数据平台(<https://www.wrcnet.cn/redcap/>)上运行的中文界面 REDCap 系统进行,该系统由“西北区域自然人群队列研究”项目团队基于 REDCap 9.7.7 平台源代码搭建并汉化^[8-9]。为评估西北自然人群队列研究中研究对象的营养膳食摄入情况,将《中国食物成分表标准版(第六版)》^[12]作为原始问卷模板与数据来源,以其电子化录入为例,进行 REDCap 系统双录入及质量控制功能的具体实施过程演示。

2. 整体质量控制框架:由 2 位数据录入人员分别进行同一数据的录入,双录入过程在项目管理员全程监控管理下进行,录入完成后依托 REDCap 系统进行数据对比与核查,参照核查结果和原始数据进行修改,确认无误后导出数据以便后续研究。整体研究的质量控制框架见图 1。

3. 具体实施过程:

(1)项目设置阶段的质量控制:REDCap 系统无须安装,通过本研究团队搭建的西安交通大学开

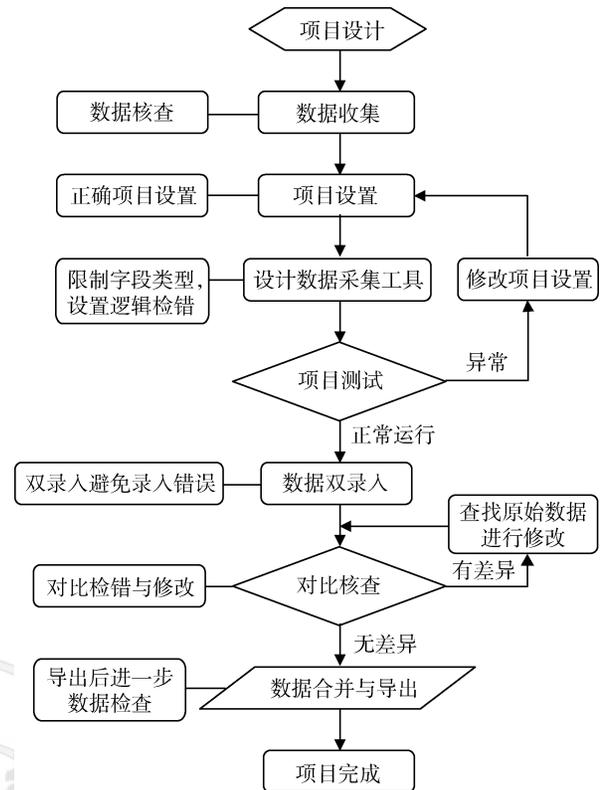


图 1 利用 REDCap 系统进行数据双录入及质量控制项目流程

放研究数据平台网页端(<https://www.wrcnet.cn/redcap/>)即可进入 REDCap 系统进行研究。该平台已完成汉化,界面简单,直观易懂。

在进行数据录入之前,首先需要进行研究项目的创建。进入 REDCap 系统,点击左上角的“+新项目”按钮,即可进入创建项目页面。可设置项目名称、项目用途、项目说明、是否使用模板。这些功能设置有助于研究者在项目设计阶段对项目的研究目的及内容有整体了解,建立项目的整体质量控制框架。

值得一提的是,基于 REDCap 系统特有的“纵向数据收集”功能,可以实现对长期纵向研究数据收集的双录入质量控制。若研究数据为纵向数据,通过在项目设置中开启“使用定义事件的纵向数据收集”功能,可实现纵向数据的统一收集与管理。开启此功能后,REDCap 系统将根据研究人员的设置,为该项目创建项目进度计划,为每一个时点分配相应的问卷,并生成进度统计表。REDCap 系统支持在每次随访中使用不同的问卷,也可使用相同的问卷进行多次录入,将每次随访收集到的数据统一管理。在进度表中,每行代表同一个对象在不同时点上的数据,每列代表每次随访,可直观查看数

据收集的进度和质量,以“妊娠期高血压危险因素监测”项目为例,项目进度表见图2。对比EpiData, REDCap 系统将多次随访数据整合为一个统一的数据库,可实现联网修改、核查与下载,对于需要多次随访的队列研究、大型多中心的医学研究等非常实用。

在 REDCap 系统中,纵向数据与单次研究数据的双录入与质量控制操作方法是相同的,本文将进行单次研究数据双录入及质量控制功能的演示,根据这些功能即可对多种类型的研究数据实现良好的质量控制。

创建项目后,需要通过“在线设计”功能设计数据采集工具也就是问卷模板进行数据的收集录入。工具的设置界面见图3,简洁直观,便于操作。在这一阶段,通过规范字段命名、限制字段类型、选择验证选项、设定是否必须、是否标识符、设置字段跳转等,实现数据录入阶段的逻辑检错,防止误输入,达到数据录入阶段的初步质量控制。当输入信息与设置的验证选项不符时,系统会弹出提示框,修改错误信息后才可继续录入数据。本文以研究团队2019年《中国食物成分表(第六版)》的膳食数据库电子化录入为例,其中各种营养成分含量数据均

为数字,则可在验证选项中选择“数字”,并规定其取值范围。

完成一系列的设置后,可选择将项目转到生产状态,就可进行实际数据的收集。项目转入生产状态后将无法随意修改设置,因此在投入生产之前需要保证各项设置正确无误。为此,可先进行“项目测试”,在测试阶段根据出现的问题进行调整。正式投入运行前的反复测试是正确创建及设置研究项目的重要前提,也是项目实施流程中良好质量控制的组成部分。

REDCap 系统在测试阶段可以随时修改采集工具,并且无须生成多个不同格式的文件,生产状态下也可在系统管理员的同意下进行修改,修改后不会丢失之前录入的数据,相较 EpiData 软件进步明显。此外,只有拥有修改权限的人员才可修改数据采集工具,所有修改过程均须通过管理员审核,且全部操作均在日志中保存相关记录。上述措施保证了数据采集工具的稳定性,也是重要的质量控制措施。

(2)数据双录入:是医学研究中广泛使用的质量控制方法,通过对同一数据的2次录入和对比,可以快捷发现录入过程中的错漏并修改。利用 REDCap 中的数据双录入模块(double data entry module, DDE)可以实现数据的双录入以及核查功能。

系统管理员需在“用户权限”设置中,将2个用户分别设定为数据录入人员#1和数据录入人员#2,并设置相应自定义权限。如在数据双录入中,可限制数据录入人员的权限,使他们无权访问项目设计、数据导出等功能,防止误操作。两个数据录入人员使用相同的ID对同一条数据进行录入,录入的数据联网储存于同一项目下,录入完成后即可实现相同ID记录的对比。

以数据录入人员#1为例,在项目页面的左侧菜单栏中,选择“添加/编辑记录”,即转到图4界面,可添加新记录或修改现有记录。Record ID 是 REDCap 中非常重要的唯一识别码,创建记录后

Record ID	Enrollment			Antenatal_Care_01	Antenatal_Care_02	Antenatal_Care_03
	01 知情同意书	02 研究基本情况	03 首次产检基本情况	03 首次产检基本情况	04 历次产检基本情况 第二次产检	04 历次产检基本情况 第三次产检
7	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

图2 纵向数据收集的进度统计表

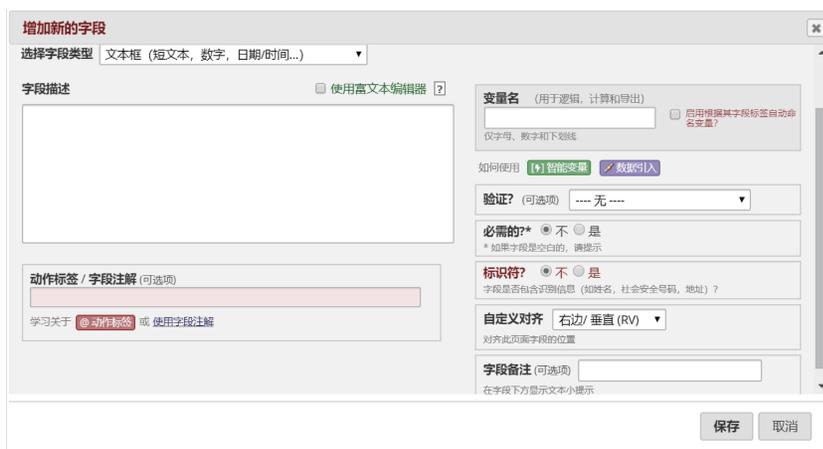


图3 REDCap 系统在线设计数据采集工具的添加字段界面

无法修改或删除,后续的数据对比核查是通过 Record ID 将 2 个录入记录链接在一起进行对比,因此务必保证 ID 正确无误。

录入完成后,需选择该条记录的状态,包括“未完成”“未验证”“完成”3 个状态,该记录缺少某些条目时可选择“未完成”,数据完整但未经过对比核查时可选择“未验证”,数据完整无误则可选择“完成”,然后保存记录。在左侧菜单栏中单击“记录状态进度统计表”,可看到所有录入完成的记录。每个记录后有一有色按钮,该按钮的颜色表示该条记录的状态,“完成”状态的记录显示为绿色,“未验证”状态的记录显示为黄色,“未完成”状态的记录显示为红色。在该页面单击记录 ID 后的有色按钮可进入该条记录的具体页面,可快捷查看和修改数据。也可通过搜索功能寻找所需的记录。

记录状态可作为该条数据质量的直观体现,灵活运用其设置,可为后续核查与修改提供便利。

(3)数据的对比与核查:完成双录入后,可以通过 REDCap 系统中的“数据比较工具”模块进行双录入数据的对比。在该页面下拉框中选择某一 Record ID,单击“比较选定的记录”则可对两个数据录入人员分别录入的同一个 ID 记录进行对比,也可直接单击“对比所有记录”对所有相同 ID 的记录进行对比。

对比后将生成对比报告,对同一个 ID 的记录,数据录入人员 #1 录入的记录用 ID-1 表示,数据录入人员 #2 录入的数据用 ID-2 表示,若两个记录对比后无差异,则可合并文件,合并文件命名为 ID 号。若有差异,在对比报告页面直接单击红色数字,则可进入该记录的修改页面,根据原始数据进行核实后,可快捷修改和保存该条记录。修改后再次进行对比,确认无误后可进行记录的合并。

在双录入对比核查过程中,REDCap 系统的明显优势就是可快捷实现多来源数据的联网核查与修改,避免了文件的互相传输所带来的不便。

(4)数据导出:在“数据导出、报表及统计”模块中,可实现所有

数据的查看和数据的导出。在“记录状态进度统计表”中只能看到单条记录的具体数据,而该模块中的“查看报表”功能,可将所有记录整合成为一个表单,快捷查看所有录入的记录。

REDCap 系统将录入的研究数据储存于网络,有权限的研究人员可利用“数据导出”功能,将录入数据导出至本地。系统支持导出为各种主流统计软件的格式,包括 Excel、SPSS、SAS、R 软件、Stata 软件(图 5)。根据后续分析计划安排,可选择相应的导出格式,并进行其他导出设置。将数据导出为所需的特定格式,可进一步在其他统计软件中对数据进行检查与整理,保证数据质量。

讨 论

本文主要介绍了利用 REDCap 系统进行数据双录入及核查的过程以及质量控制措施,为研究者使用 REDCap 系统进行数据录入和管理提供参考,有助于提高医学研究效率。

REDCap 系统作为一种新兴的临床数据采集系统,对比现在主流的录入软件和线上问卷系统,有着简单、方便、高效、稳定等多重优势。数据双录入是目前普遍使用的一种质量控制方法,利用

表 1 能量和食物一般营养成分

图 4 REDCap 数据录入界面

图 5 数据导出界面

REDCap 系统进行双录入,不仅能快捷实现数据对比核查,并且从项目创建到数据导出的每个阶段均有其相应质量控制措施,如逻辑检错、权限设置、数据质量可视化及快捷修改等,路线清晰、操作简单且效果良好。每个参与项目的研究者有其个人用户和密码,通过项目管理员的权限设置及系统的全程日志监控,有效保证研究数据的安全准确。

与 EpiData 软件相比,REDCap 系统不仅可以通
过本文介绍的各种功能实现良好的质量控制,还具有无须安装、界面简单、集问卷调查与数据录入于一体、无须多个格式文件、能够实现多来源数据的联网核查与合并、数据不易丢失、可实现纵向数据长期收集等多种优点,在灵活性、效率及质量上优势明显。但也存在一些缺点,如在数据输入时记录 ID 无法修改、无法删除记录、数据输入完成后需要 Tab 键手动切换字段等。但瑕不掩瑜,REDCap 系统的独特优点是其他数据录入和管理方式无法替代的,并且系统也在不断地完善与升级中,因此值得在医学研究中,尤其是多中心临床研究、队列研究等大型研究中推广应用。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Faizi N, Kumar AMV, Kazmi S. Omission of quality assurance during data entry in public health research from India: Is there an elephant in the room? [J]. Indian J Public Health, 2018, 62(2): 150-152. DOI: 10.4103/ijph.IJPH_386_16.
- [2] 张路霞, 王海波, 李全政, 等. 中国的大数据与医学研究 [J]. 英国医学杂志: 中文版, 2018, 21(4): 203-206. DOI: 10.1136/bmj.j5910.
Zhang LX, Wang HB, Li QZ, et al. Big data and medical research in China [J]. BMJ (Chinese), 2018, 21(4): 203-206. DOI: 10.1136/bmj.j5910.
- [3] 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.
- [4] 孙玉环. 基于 EpiData 与 SAS 系统的纸版问卷数据录入质量控制技巧 [J]. 中国卫生统计, 2012, 29(4): 607-608, 611. DOI: 10.3969/j.issn.1002-3674.2012.04.051.
Sun YH. Quality Control skills of Paper questionnaire data input based on EpiData and SAS system [J]. Chin J Health Statist, 2012, 29(4): 607-608, 611. DOI: 10.3969/j.issn.1002-3674.2012.04.051.
- [5] 郑文新, 汪文新. EpiData 软件在流行病学调查中建立数据库的应用 [J]. 数理医药学杂志, 2005, 18(5): 486-487. DOI: 10.3969/j.issn.1004-4337.2005.05.043.
Zheng WX, Wang WX. Application of EpiData software in establishing database in epidemiological investigation [J]. J Mathemat Med, 2005, 18(5): 486-487. DOI: 10.3969/j.issn.1004-4337.2005.05.043.
- [6] 李庚, 李晓彦, 温泽淮. 国际临床数据交换标准协会标准在电子数据采集系统中的应用研究 [J]. 世界科学技术-中医药现代化, 2017, 19(2): 332-338. DOI: 10.11842/wst.2017.02.023.
Li G, Li XY, Wen ZH. Application of clinical data interchange standards consortium (CDISC) standards to electronic data capture system [J]. World Sci Technol Modernizat Tradit Chin Med, 2017, 19(2): 332-338. DOI: 10.11842/wst.2017.02.023.
- [7] 谢高强, 李英山, 姚晨. 电子数据采集对我国临床研究的机遇和挑战 [J]. 中国新药杂志, 2013, 22(6): 620-623.
Xie GQ, Li YS, Yao C. Opportunity and challenge of electronic data capture on clinical researches in China [J]. Chin J New Drugs, 2013, 22(6): 620-623.
- [8] 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.
- [9] 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.
- [10] Tomko RL, Gray KM, Oppenheimer SR, et al. Using REDCap for ambulatory assessment: Implementation in a clinical trial for smoking cessation to augment in-person data collection [J]. Am J Drug Alcohol Abuse, 2019, 45(1): 26-41. DOI: 10.1080/00952990.2018.1437445.
- [11] 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.
- [12] 杨月欣. 中国食物成分表: 标准版 [M]. 北京: 北京大学医学出版社, 2018.
Yang YX. China food composition tables: standard edition [M]. Beijing: Peking University Medical Press, 2018.