Managing REDCap Data: The R package REDCapDM

João Carmezim[1], Pau Satorra[1], Judith Peñafiel[1], Esther García-Lerma[1], Natalia Pallarés[1], Naiara Santos[1], Cristian Tebé[1]

[1] jcarmezim@idibell.cat, psatorra@idibell.cat, jpenafiel@idibell.cat, egarcia@idibell.cat, npallares@idibell.cat, ctebe@idibell.cat, Biostatistics Unit, Bellvitge Biomedical Research Institute (IDIBELL)

BACKGROUND: REDCap (Research Electronic Data CAPture) is a web application for creating and managing online surveys and databases. Clinical data management is essential before performing any statistical analysis to ensure the quality and reliability of study data.

OBJECTIVES: The aim of the R package "REDCapDM" is to process REDCap data and provide useful tools to perform all tasks involved in the data cleansing process prior to statistical analysis.

METHODS: The 'REDCapDM' package is organized into four dimensions, each serving a specific purpose. First, reading and processing raw data from REDCap or through a REDCap API connection in R. Second, performing data transformation and data organization. Third, it identifies queries, specifically missing values, values that fall outside the lower and upper limit of a variable and other types of inconsistencies in data from REDCap in R. Fourth, it performs an automatic control of queries that have been resolved or are pending resolution.

RESULTS: A total of 5 functions have been included in the package to speed up and automate the process of data managing of REDCap data in R. The "redcap_data" function is used to read and prepare the data set originated from REDCap. The "rd_transform" function is used to perform several data transformations. The "rd_query" and "rd_event" functions allow the identification of different types of data issues and generate a query report. This query report is consists of two elements: a data frame containing the information needed to identify each query in the REDCap project and a summary of the number of queries per variable/event and query type. Finally, we have the "check_queries" function, which is used to compare a previous dataset of queries with a new one and check which queries are new, which are pending resolution, which are miscorrected and which are solved.

CONCLUSIONS: The R package "REDCapDM" contains a collection of useful functions to import, organize, and perform a quality-checking of data from REDCap. This package fills a gap in the available tools for managing REDCap data, making it an invaluable asset for researchers. The "REDCapDM" package is freely available from the CRAN library (https://cran.r-project.org/web/packages/REDCapDM/index.html) and is updated regularly.

Keywords: REDCap, clinical data management, query.