

# Information Platform to Improve Research Utility of Clinical Study Data

Simone Andersson

## ABSTRACT

The value of the results in every clinical study depends on the quality of the captured data and the analysis process. The goal is to discover and prove a scientific matter, to do this the data has to be valid, easy to access and understand.

To capture data is often a complex, time consuming process that could go on for multiple years, and this process is done for all studies. Imagine the possibility to reuse the captured data in similar future projects, it will reduce the need for capturing new data and give researchers access to more data to use in their analyzes.

This article introduces a platform constructed for a newly formed research group, the platform is for capturing and storing data that will ease the workload of the researchers. The system will provide the feasibility to reuse their captured data in future projects and improve quality in the captured data to improve the analysis process.

**Keywords:** REDCap, EDC, Proteomics, Reuse data

## 1. INTRODUCTION

Often when a research project is completed and the researchers have published the data it is open for others to use. However, it is not always that easy to access others data, a video (NYU Health Sciences Library, 2012) is a very good example of how it may work today when it comes to using others data. It reveals problems with documentation, storage and format that could occur while sharing data.

Considering how much data that already exists, on ClinicalTrials.org there are 2040 studies registered in the area of COPD, chronic obstructive pulmonary disease, of these 1151 are completed (U.S.

National Institutes of Health, 2014).

This indicates that there is a lot of data about patients with COPD that is probably no longer used.

If all this data would have been properly documented and/or captured following some standards it would result in a large data set that could be reused in other studies regarding COPD, giving the researchers a large data set to start with and easing the work of capturing new data for them. This is a long-term goal and it will take a lot of time before data will be shared and reused between research groups but an intermediate goal is for the research groups to start reusing their own captured data and use standards to make the data readable for others.

To make this possible a tool for capturing and storing data needs to fulfill the following data requirements.

- Digitalize data collection, data is often captured on paper, which increases the risk of wrongful data in the system. By digitalizing the data collection human errors will be minimized and ease the workload of the research group (Weber, Yarandi, Rowe, & Weber, 2005).

- Fast and easily searchable database, the database has to be well thought out and structured, this to improve and facilitate the analytic phase but also to give an overview of what data exists and how it is stored in the database.

- Standardize data, is to capture data according to a standard to make it understandable and easy to reuse for both the group and others in the future. To accomplish this data and all used processes are to be well documented and the

standard to be followed in future studies.

## 2. METHODS

A newly started research group intends to improve their way of handling data and prepare the data for reuse in future studies. To start this process a tool that follow the previous data requirements is installed and used for one study, the following sections will introduce this study, but the goal is to use it for all appropriate studies.

### Study

KOL-Örestad is a longitudinal clinical study created primary objective of identifying and validating new biomarkers in the different stages of COPD. The study is registered at ClinicalTrials.gov under the official title *Biomarkers of early COPD and Cardiovascular Diseases in Smokers/Ex-smokers Attending Primary Care* –

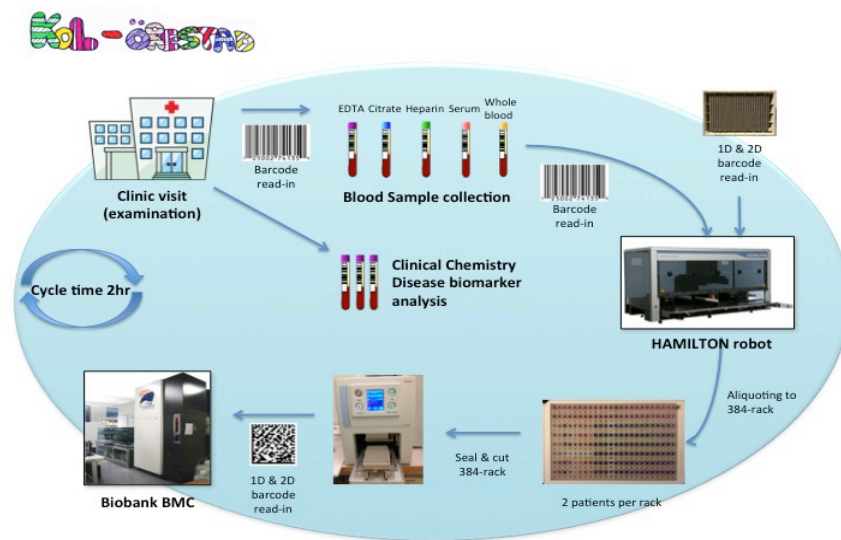


Figure 1: KOL-Örestads workflow

*Longitudinal Study* (U.S. National Institutes of Health, 2014). It will involve approximately 300 patients in the ages between 35 and 80, 200 diagnosed with COPD from all four stages of GOLD classifications (Global Initiative for Chronic Obstructive Lung Disease, 2014), and 100 healthy patients (for control) both smokers and non-smokers.

Each patient will do visits at the clinic every six months for five years, at each visit spirometry test and physical examination are performed, the patient fills out a questionnaire and blood samples are taken. Three 5mL tubes of blood are sent for direct analyses to Clinical Chemistry and five are sent to BMC, Biomedical Center, in Lund where the research group is located. At BMC the samples are divided into small aliquots (70 $\mu$ L) on 384 racks, and stored in a biobank at minus 80 degree Celsius for future analysis. Figure 1 illustrates the workflow of the study.

It is a multi-site study that will collect data at three different sites, *Örestadskliniken*, *Clinical Chemistry* and *BMC*. All sites capture multiple data variables that are all to be stored at BMC where the analysis process of the captured data will be performed.

To link the collected data and samples to a specific patient the patient gets a patient ID at the clinic, the ID links to personal data, name and social security number, and to the study data. These two types of data have to be separated, the setup of the study will not allow personal data to be used outside of the clinic,

however, all data will be accessible at the clinic. This is a requirement base the inform consent for the study that has to be accomplished by the installation.

### **REDCap**

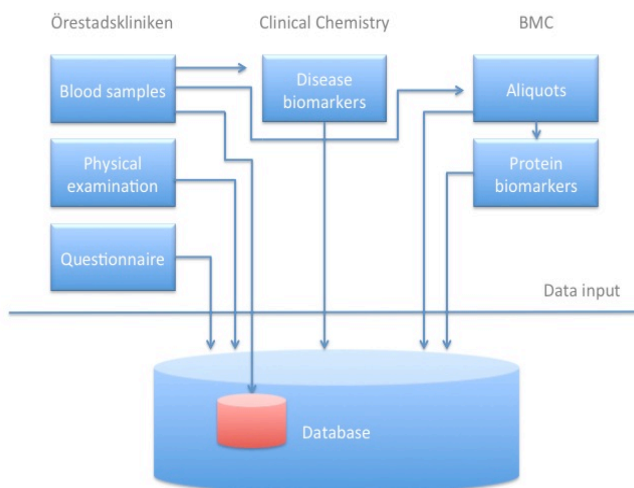
To reach the requirements REDCap, Research Electronic Data Capture, was used. REDCap is an EDC, Electronic Data Capture, tool developed to fit a broader purpose, standardization for users. REDCap realized that a critical factor for a successful EDC tool was to create a simple workflow in which the user easily can setup and modify metadata (Harris, Taylor, Thielke, Payne, Gonzalez, & Conde, 2009).

This software is used all over the world for studies and has at the time of this writing REDCap has 1,100 active institutions with a total of 142,000 users and 110,000 projects running (REDCap).

### **Accomplish requirements**

REDCap fulfills the data storage requirement provided with a feature called identifier, by using this the personal data can be marked as identifiers, this means that this data will only be accessible for users with specific permissions. This results in one database for all captured data, figure 2.

Digitalize data collection, this is what REDCap was developed for, create a simple tool for collecting data electronically. This solution eliminates, or close to eliminates, the paper trail in the study, which results in a higher quality of the data in the database with less human errors. Using REDCap also resulted



**Figure 2: Input data sources**

in a shorter and more user specific questionnaire were only relevant questions where asked.

Fast and easily searchable database is the second requirement and is fulfilled by the software through a relational database, the user can export all data or specify what variables of data to export, analyze collected data in real time in the application graphical view, creating reports, etc. This is both fast and easy to use. It is also possible to export data in different formats to fit the requirements for different analytics programs, which will ease the analytic process. The software also provides a readable overview of all variables in a project.

Standardize data can be collected by using the shared library or create new ones that will be used in the future in similar studies. The shared library is a repository of forms that can be used again by the same group or shared between institutions, this is a great way to get standards within and between institutions, it can also

reduce the work needed to setup a new study by using and modifying existing forms used in previous studies. Standardize data also includes documentation about everything regarding the study, REDCap has a file repository where all files for a project can be uploaded. This repository will include documentation about the collected data to allow others to use and understand the database.

### 3. RESULTS

The implementation resulted in three forms that are now used in the study, all forms fulfills the requirements defined and are using features provided by REDCap to optimize the data capturing process.

### 4. DISCUSSION

REDCap is a great tool to use for setting up new studies, although the goals were reached there are still more that can be done in REDCap to improve and ease the work in this and future studies.

Features like spellcheck of text field to minimize spelling errors, default values to use the value that was entered on the previous visit as default for some questions, auto-completion of medicine and disease by checking them against known databases such as MedDRA (Medical Dictionary for Regulatory Activities) and Fass (Läkemedelsindustriföreningens) to

get a standard of how they are stored. Features like this would improve the quality of the data stored in the system.

## 5. Conclusion

The appearance of the study changed through out the whole process due to that new needs were found and old needs changed during the implementation of the EDC. REDCap was an excellent tool to use in this process, the workflow of the environment lead to that all test resulted in new improved forms that finally could be used in the study.

The KOL-Örestad study started using the system in June 2014, the employees at the clinic and the researchers where satisfied with the setup and excited to start use it.

## REFERENCES

- Global Initiative for Chronic Obstructive Lung Disease. (2014). *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease*. Global Initiative for Chronic Obstructive Lung Disease.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 2 (42), 377-381.
- NYU Health Sciences Library. (2012, August 3). *Data Sharing and Management Snafu in 3 Short Acts*. Retrieved February 12, 2014, from <http://www.youtube.com/attributions?v=N2zK3sAtr-4>
- REDCap. (n.d.). <http://project-redcap.org/>. Retrieved 05 06, 2014, from <http://project-redcap.org/>: <http://project-redcap.org/>
- U.S. National Institutes of Health. (2014, Feb). Retrieved May 2014, from ClinicalTrials.gov: <http://clinicaltrials.gov/ct2/show/NCT02076061?term=KOL+001&rank=1>
- Weber, B. A., Yarandi, H., Rowe, M. A., & Weber, J. P. (2005). *A comparison study: paper-based versus web-based data collection and management*. *Applied Nursing Research*, 18, 182-185.