Inefficiencies Assessment and Data Mapping for Clinical Trial Data Transfer

Jad Elgaali and Troy Zhang

Abstract:

Introduction: The Center for Research Informatics (CRI) and AbbVie are collaborating to build data transfer pipelines between each institution's database using REDCap as the intermediary between each organization. The end-goal is to automate aspects of clinical trials to make the process more efficient for the stakeholders.

Methods: The project was divided into two phases, each of which was conducted over the course of five weeks. The first phase we met with with Research Coordinator (RC) and Data Manager (DM) to understand the workflow of the stakeholders. During the second phase we met with a CRI analyst for data mapping.

Results: The RC meets with patients, collects their data and enters it into the Electronic Data Capture (EDC) system. The DM transfers data to the EDC, but doesn't meet patients. Most pain points involve queries, most significantly that stakeholders spend 3⁄4's of their time resolving queries. For the data mapping, vitals and demographic data were mapped.

Conclusion: There are inefficiencies in the RC and DM's workflows. Some are due to departmentmental workflow decisions. This includes insistence on the use of paper documentation, which is harder to build automation pipelines for. The end-goals of this project can remove some stakeholder problems.

Introduction:

Clinical trials are a critical part of medicine for the development of new medications and treatments. The probability of success that drugs make it all the way to market has recently been shown to be around 14%, much higher than the previously believed rate between 3-9% (5). However, clinical trials have been conducted the same way for decades for a variety of reasons ranging from federal regulations (1) to medicine being notorious for lagging behind other industries in the adoption of new technology (2). For example, Research Coordinators in clinical trials still use paper source documents which have to be manually created in lieu of electronic documentation.

There are various stakeholders involved in clinical trials, which typically includes a Principle Investigator (usually the physician), the nursing staff, Research Coordinators and Data Managers. The Research Coordinator and Data Manager are responsible for the collection and entry of clinical trial data into the various databases involved in the study. These databases typically include the hospital's electronic medical records system (EMR) and the study sponsor's electronic data capture system (EDC).

According to the stakeholders, hospitals are efficient at collecting clinical data from patients; however, the process of collecting research data in clinical trials is where many inefficiencies lie. The process of having multiple databases to enter data into and patient forms to fill out creates redundancies and inefficiencies in the workflow of the Research Coordinator and Data Manager in clinical trials. Additionally, lack of standardization and interoperability between healthcare systems is an issue that negatively impacts the stakeholder's workflow (3).

AbbVie and the University of Chicago's Center for Research Informatics (CRI) are collaborating to automate the data transfer of clinical trial data from the University of Chicago's Clinical Research Data Warehouse (CRDW) to AbbVie's Electronic Data Capture system (EDC), Medidata Rave. In this project, REDCap, a research data warehouse, will be used as the intermediary between the CRDW and the EDC (4).

The problem we are addressing is to understand the workflow of the stakeholders in clinical trials and to identify their pain points. Furthermore, we had a data mapping phase in which we identified the specific data elements in clinical trials could be feasibly automated at this time. If successful, this project can remove some of the inefficiencies in stakeholder workflow and improve quality control and patient care.

Methods:

The project was divided into two phases, each of which was conducted over the course of five weeks. During the first phase, we gathered information to gain an understanding of how patient data in clinical trials goes from collection to entry in the EMR, to being transferred into the EDC. Our goal was to understand the workflow of the stakeholders involved in this process, specifically the Research Coordinator from the Gastroenterology (GI) Department and Data Manager from the Hematology-Oncology Department. Additionally, we sought to identify the pain points in their workflow.

During the second phase of the project, we met with an analyst from the CRI to process data elements from AbbVie's EDC, Medidata Rave. We then identified which data elements can potentially be automated and mapped them. We were also tasked to evaluate the viability of REDCap as a platform for this project.

Phase 1 - Interviews

We met with the Research Coordinator from the GI Department and Data Manager from the Hematology-Oncology Department every week for five weeks. The meetings were held back-to-back, first meeting the Research Coordinator and then the Data Manager. Each interview lasted for 2 hours for a total of 10 hours of interviews with each stakeholder, or 20 hours total. We used the questions from Table 1 to structure our interviews. Every week the answers were documented for review and used to refine our questions for the following week. Through repeated questioning, we were able to learn the stakeholders workflows and identify any problems they faced.

Table 1: Questions asked to the stakeholders

Q1	What clinical trials are you working on?
Q2	What is your typical workflow?
Q3	What are the pain points in your workflow?
Q4	Are their redundancies you feel could be streamlined?
Q5	How is AbbVie's system to use (what is the ease of use)?
Q6	Why does your department still use paper for source documentation?
Q7	How often do you get queried?
Q8	What percentage of your queries result in a value-change?
Q9	How do you deal with queries and what is your query workflow?

Phase 2A - Data Mapping

We met with an analyst from the CRI to process and map the data elements which we believed could be automated. This was done through two meetings over the course of two weeks. Each meeting lasted two hours. The analyst assisted in the processing phase as he knew from his experience with data analysis which types of elements could be feasibly automated. When an element which could potentially be automated from the EDC to the CRDW was identified, the automation pathway was mapped onto a separate document.

Phase 2B - REDCap

We also looked into REDCap in order to evaluate its viability as a platform for the CRI and AbbVie to build the data transfer pipelines. We wanted to understand how widely used the platform is and its potential for scalability. After the organizations build the pipelines, future analyses will be better able to assess scalability. We collected and documented the relevant data from the information directly available on REDCap's official website (4).

Results:

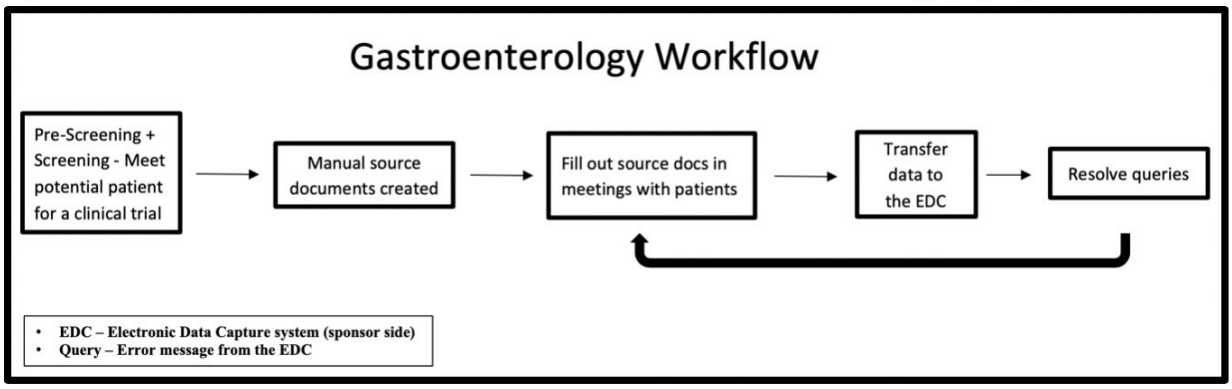
Phase 1: Interviews

Throughout the interviews, we were able to understand the stakeholders workflows and pain points. Additionally, we identified two categories of pain points: those shared and those unique to each stakeholder.

Research Coordinator Workflow:

The workflow for the Research Coordinator began with the pre-screening and screening phases which involved meeting patients who may be eligible for clinical trial enrollment. This initial phase took about 35 days. After patients were screened and approved for the clinical trial, the Research Coordinator created the paper source documents based on the trial's protocol from the sponsor. The EDC contained all data points of interest to the sponsor for clinical research. After the data was entered into the EDC, the Research Coordinator may be queried if some of the data was incomplete, incorrect, or if the sponsor wanted clarity on a particular data point. A query was an error message that appears to the stakeholder in the EDC. The Research Coordinator attempted to resolve these queries, and when resolved, the query was closed. The flowchart from Figure 1 demonstrates this workflow, and the last three steps are continuously repeated throughout the duration of the clinical trial until its completion.

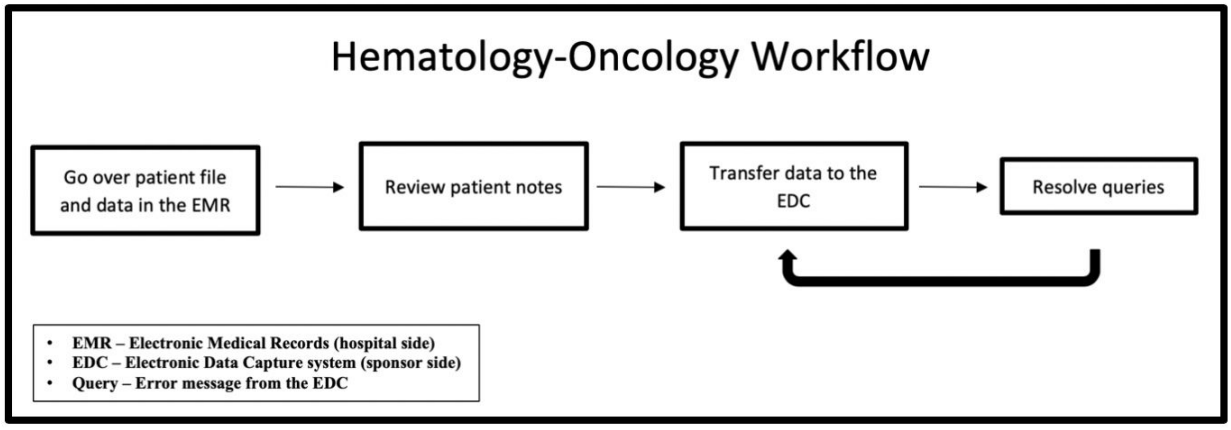
Figure 1: Clinical Trial Workflow of the Research Coordinator from the Gastroenterology (GI) Department



Data Manager Workflow:

The Data Manager, unlike the Research Coordinator, did not meet with patients or make paper source documents. The Data Manager began his workflow by reviewing the patient file and data in the EMR, as well as the electronic patient notes. The Hematology-Oncology Department, unlike the GI Department, did not use paper source documents. The Data Manager also received queries, and like the Research Coordinator, had to clarify data points or fix any mistakes until the sponsor was satisfied, after which the query was closed.

Figure 2: Clinical Trial Workflow of the Data Manager from the Hematology-Oncology Department



Shared Pain Points:

There were four shared pain points between the Research Coordinator and Data Manager. These were:

- *Non-resolvable queries* - these were queries (usually auto-generated in the EDC) where the stakeholder receives a query without a text box to prompt an answer.
- *Time in the EDC resolving queries* - the stakeholders both stated they spend about three-fourths of their time in the EDC resolving queries.
- *No "Save and Continue" button* - after completing each form in the EDC, the stakeholders had to manually click "Save" and move their cursor to the next form.

- *Time to close query* - the Research Coordinator and Data Manager told us that it occasionally took longer than 48 hours for a query to be closed out after they addressed it in the EDC.

Individual Pain Points:
An individual issue the Research Coordinator faced was that the EDC prompted her for a diagnosis for every symptom she input. Also, she occasionally received queries stating that the patient should still be on a medication when they should not be.

The Data Manager, who began working in July 2018, still had unresolved queries from 2016. He also occasionally had to enter the normal ranges of lab results in the EDC, even if the normal ranges from the lab and sponsor are identical. Lastly, the Data Manager noted that sometimes the units prompted in the EDC were different from the units given from the lab, and that he had to convert from what the lab gave to what the sponsor is asking for. The Hematology-Oncology Department had a special spreadsheet they created to convert units in these situations.

Query Resolution Workflow:
Each stakeholder had a different workflow to specifically deal with queries. The Research Coordinator answered queries as they came up and prioritized queries near a data cut event. A data cut event was a deadline set by the sponsor by which the stakeholder had to have all outstanding queries resolved so the data science team on the sponsor's side could begin cleaning and prepping the clinical trial data for research. Additionally, when asked in further detail about what percentage of her queries result in a value-change, the Research Coordinator estimated that about 20-30% of her queries result in a value change. A value-change meant that when queried, the actual data input was incorrect and had to be corrected.

The Data Manager was more systematic with his query resolution workflow. He only worked on queries on days where there were no patient visits. The Data Manager separated queries by clinical trial and began with the studies with the oldest queries and worked his towards the studies with the most recent queries. Additionally, he estimated that about 50-60% of the queries he received resulted in a value-change.

Phase 2A - Data Mapping
There were 750 data elements to process and when an element which could potentially be automated was identified, the automation pathway was mapped. The pathway included the field name from the EDC, the data source, the field name in the CRDW, and the discrete pseudocode for mapping each data element. Table 2 represents the vitals we were able to map.

After processing, we were able to map most of the demographic information and vital signs. A notable demographic data point, the patient's age, could be mapped and must be pulled manually. The demographics we were able to map included the birth date, gender and race/ethnicity. Each racial/ethnic group had to be individually mapped. This included Hispanic, White, Black, Asian, Native American, and Native Hawaiian/Pacific Islanders. The CRDW specified that those of Indian and Middle Eastern descent are classified as Asian. The vital signs were redundant in the EDC, having five separate sections asking for the same data.

Table 2: Mapped data elements from Medidata Rave (left column) to the Clinical Research Data Warehouse

Rave Variable Name	Vital Signs	Data Source	CRDW Variable	Discrete pseudocode in CRDW
VS_BODY_POSITION	Body Position	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = Position for BP
VS_BP_SYSTOLIC	Systolic Blood Pressure	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = BP written as number, script to separate
VS_BP_DIASTOLIC	Diastolic Blood Pressure	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = BP written as number, script to separate
VS_BP_U	Blood Pressure Units	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = BP
VS_PULSE	Pulse	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = Pulse
VS_PULSE_U	Pulse Units	units field not populated, auto fill in bpm		
VS_RESP_RATE	Respiration Rate	CRDW	MEAS_VALUE	MEAS_VALUE
VS_RESP_RATE_U	Respiration Rate Units	units field not populated, auto fill in bpm		
VS_TEMPERATURE	Temperature	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = Temp
VS_WEIGHT	Weight	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = Weight
VS_HEIGHT	Height	CRDW	MEAS_VALUE	MEAS_VALUE where FLOW-MEASURE_NAME = Height

- CRDW – Clinical Research Data Warehouse

Phase 2B - REDCap

We found that REDCap was used by 3601 institutions across the globe with over one million users in 131 countries as of August, 2019 (4). Additionally, there were over 740,000 projects utilizing REDCap (4).

Discussion:

Phase 1 - Interviews

While the general workflow of the stakeholders was similar once they reached the data entry stage, the query workflow for query resolution was quite different between the two stakeholders. The differences in query resolution workflow was due to each stakeholder's workflow decision and what they believed to be the most efficient way to resolve queries.

The percentage of queries that the Research Coordinator and Data Manager said resulted in a value-change (20-30% and 50-60%, respectively) was the stakeholder's estimate. The large disparity makes sense when considering the type of data each stakeholder worked with. The Research Coordinator worked in the GI Department and subsequently mostly dealt with qualitative data, using text to describe the patient's condition. The Data Manager from the Hematology-Oncology Department mainly worked with numerical data. The chances of

inputting the wrong condition would be less likely than mistakenly entering an incorrect number, especially when there were large quantities of numerical data points to enter.

As for the pain points, the Data Manager believed that having to enter the normal range in the EDC if it was the same from the sponsor and lab was not the most efficient use of his time and that it should auto-populate. The Research Coordinator said they often do not have a diagnosis immediately and the sponsor can overwhelm her with queries.

Phase 2A - Data Mapping

There were redundant sections asking for the vital signs. The reason for this was to collect the vitals during each patient visit and record it in the EDC. The number of data elements which could be mapped between the EDC and CRDW was limited at this stage to the vitals and demographics. This was due to the inherent difference between EDC's, each of which was custom built for each clinical study by the sponsor. We identified what we believed to be the constants between all clinical trials which we may be able to automate. However, we will not know if the mapping was successful until the automation pipelines are built, which will be the primary goal of the upcoming phase of the project. Lastly, the pseudocode from the data mapping document will eventually become a SQL script the analyst will write to extract the data elements from the CRDW.

Phase 2B - REDCap

REDCap's use across the world as a research data warehouse makes it viable as a platform for this project. Scalability can better be addressed after the automation pipelines are built. However, in the long-term, the focus should be on using Fast Healthcare Interoperability Resources (FHIR) to build such pipelines.

Conclusion:

Throughout the two phases, we identified the workflow and pain points of the Research Coordinator and the Data Manager and mapped data elements which could potentially be automated between the EDC and CRDW. Finally, we evaluated the viability of REDCap as a platform for this collaboration.

There were inefficiencies in the workflows of the Research Coordinator and Data Manager, some of which the end-goals of this project may address. Some of the inefficiencies were the result of workflow decisions made by the department. An example was the insistence on continued use of paper source documents, which will be harder to build automation pipelines for.

We identified shared, as well as individual pain points for each of the stakeholders. Most of these problems involved queries. An example is that both stakeholders spend about three-fourths of

their time in the EDC is spent resolving queries. As for individual pain points, the Research Coordinator would be asked for a diagnosis for every symptom input and the Data Manager still has unresolved queries from 2016. These are a few of the inefficiencies in the workflows of the Research Coordinator and Data Manager we identified, some of which the end-goals of this project can address. The goal of building semi-automated data transfer pipelines between Medidata Rave and the CRDW is to make the process of clinical trials more efficient for the stakeholders involved so they can focus on patient care, which is a major reason for the existence of these clinical trials.

Next Steps:

The next phase of the project is to begin to build the automation data pipelines between the CRDW and REDCap and AbbVie and REDCap. REDCap will be used as the interface between the two platforms such that the data stored in REDCap from the CRDW and AbbVie will be able to access and extract the data. This should be done thoughtfully with regard to the stakeholders whose jobs this will effect. The end goal is to try and remove some of these redundancies and inefficiencies in their workflow and streamline and make their jobs easier overall.

References

1. DerSimonian, Rebecca, and Nan Laird. "Meta-analysis in clinical trials." *Controlled clinical trials* 7, no. 3 (1986): 177-188.

2. DerSimonian, Rebecca, and Nan Laird. "Meta-analysis in clinical trials revisited." *Contemporary clinical trials* 45 (2015): 139-145.

3. Mead, Charles N. "Data interchange standards in healthcare it-computable semantic interoperability: Now possible but still difficult. do we really need a better mousetrap?." *Journal of Healthcare Information Management* 20, no. 1 (2006): 71.

4. "REDCap." REDCap. Accessed July 12, 2019. https://www.project-redcap.org/.

5. Wong, Chi Heem, Kien Wei Siah, and Andrew W. Lo. "Estimation of clinical trial success rates and related parameters." *Biostatistics* 20, no. 2 (2019): 273-286.