**Personal Statement:**

This is Troy Zhongyi Zhang. The Clinical Trial Data Transfer project is one of my top choices all over the capstone projects list since this is a collaborative project between AbbVie and UChicago CRI. I would have the opportunity to work and communicate with people on both sides. There is always a gap between schools and companies, but what takeaways I am looking for is to participate in the joint project between two research institutes and find the correct linkage for me to connect the knowledge I obtained from school and the real-world datasets analysis. The clinical informatics has great significance since people will be touching electronic medical record (EMR) every day, and more and more people around the world will abandon the hand-written notes and proceed to electronic record method. Different institutes might have their unique statistical method and format standard to count and record the data. The analysis group will always be helpful to predict potential risks and watch the trend of a specific disease. The different biomedical data analysis system and the transferring interface are still overwhelming challenges that consume huge time and money costs for researchers to work on. Due to the complexity of the human body and healthcare medical indexes, the first-hand raw data that researchers captured directly in the hospital can be extraordinarily disordered. These data will not be able to input into a data mining or machine learning model. Meanwhile, the storage for these plentiful data could be expensive. The data mapping is an indispensable procedure to clean the data and filter the useful variables for the next-step analysis. The University of Chicago is a prestigious research institute with plenty of databases. The CRDW databases under the Biological Science division is the one that I am most curious. Personally, I would also love to work with the databases from the University of Chicago. Read through the description of Capstone Project list, I found there might be an on-site working opportunity at AbbVie. I cherish and will appreciate the opportunity since I enjoy face-to-face communications more than remote internet or phone-call method of exchanging ideas. I believe that I will be capable of learning a lot from the on-site work.

**Description of the project and anticipated results:**

UChicago BSD developed a system, which is called CRDW database that has exceptional depth and broadness. The CRDW mainly draws data from Epic EMR, Centricity billing system, Cancer Registry, National Death Registry, REDCap, etc. REDCap is a web application for building and managing online surveys and databases. Any types of data capture systems for research studies and operations are HIPAA compliant. The database was initiated since 2016. The AbbVie’s data science team is looking for the data transfer interface that can be used in their data analysis work. The research coordinator from UChicago Hospital transcribed data manually into Electronic data collection system previously. The manual Data entry could cause delays and errors. They need to query and monitor the follow-up procedures of the data entry process. The project contains a multi-step iterative process with piloting automation clinical trial data transfer to utilize the procedures between CRDW and REDCap. The automated data transfer pipelines are used to data mapping and curations. The next phase is the method we choose for data flowing from CRDW to REDCap databases, and this is the general job description from the CRI side. There would be a variable sensor to transfer and write the data into the database. Then we map the data from eCRF (electronic case report form) into REDCap to associate the field in Medidata RAVE. The Medidata RAVE capture data electronically for NCI studies. The phase III is we are going to do a comparative analysis between auto extracted patient level CRDW data and the manually curated information data. The primary task for me is to mapping source-level data to multiple common data model types. There could be a testing phase for our interface achievement. From the CRI’s perspective, we need to manage the complex health information, and then we are responsible for the development and maintenance of data models. However, the primary task for us is to map the source level data to multiple common data model types. Finally, we could contribute data and insight into AbbVie Data Transfer auto efforts.

Our primary task is to map the data we acquired from the UChicago hospital. The primary tools for us to use will be EXCEL. We will need to filter the transferable variables among different formats from hundreds of variables. Mostly, from my hypothesis, the numerical values can easily to be transferred, but the categorical variables can be complicated. The different system has an independent method of recording the categorical meanings, which can be very tough for us to translate. This could be one big challenge that we are going to face. The final result is we are going to map and transfer all the data from UChicago hospital data capture system, which is ODM format, into the CRDW readable format. We will talk with our research coordinator about how to map the raw data. We will also have further meetings with the CRI team to study the characteristics of CRDW to a deeper extent. The anticipated results are the automation clinical trial data transfer system. CRI team may test the accuracy and efficiency of the new system with FHIR interface. We don’t want to miss any necessary variables, and we wish all the categorical variables with different formats and column names can be delivered into the right spot in CRDW. The new interface we created could successfully reduce the delay and errors compared with the manual-input data method.