## Research Data Management Teaching Case: Studying Vitamin D as an Augmentation of Treatment for Bipolar Depression

A PI is planning to investigate the use of vitamin D as an augmentation of treatment for bipolar depression. There have been some studies that suggest a low vitamin D level may be associated with unipolar depression, what one could consider the classic form of depression. Bipolar depression is particularly difficult to treat, and there is a need for alternative ways to help support mood in people with bipolar disorders such as depression. The PI has received funding from a foundation to proceed with her study.

The project is going to be a double blind, placebo-controlled trial using 5000 IU of vitamin D versus placebo for twelve weeks. The research team will measure subjects’ vitamin D levels at the beginning and at end of the study. In addition, the team will use a number of validated quantitative instruments to evaluate subjects’ mood periodically throughout the study.

The most critical member on the research team is the Research Coordinator. She will be advertising, recruiting, screening, handling the consent forms, and administering a number of the mood evaluations. She is also a phlebotomist, so she’ll be taking the participants’ blood draws. There is a Co-I, whose only role is to hold the key to the study’s randomization. This way the PI, Research Coordinator and the study’s subjects will be blinded to whether the subjects are receiving vitamin D or placebo.

To conduct the study, the PI will have a number of medical students/volunteers working on her project. They will be responsible for helping to set up the data collection spreadsheet and entering the project’s data using *REDCap*, a secure web-based, HIPAA-validated software platform available only to members of an academic consortium housed at Vanderbilt University. It is widely used for data capture in biomedical and CTSA research. Once the data from the paper-based forms are entered into *REDCap*, the PI will export these data files into *SAS*, and then these files will be handed to the staff statisticians in the Medical School’s Department of Quantitative Health Sciences, also known as the Quantitative Methods (QM) Core, for statistical analyses.

After getting IRB and FDA approvals, the project will begin advertising for prospective participants. Candidates will contact the Research Coordinator to set up a screening and baseline appointment. During this appointment, they will complete a consent form and screening forms, which will be used to help determine if they think they have bipolar disorder, if they are depressed, and if they feel they are likely to have low vitamin D levels. The team is seeking participants with low levels of vitamin D so they will be appropriate for supplementation.

Candidates that pass the initial screening will then have a baseline clinical examination during the same visit. The team will perform a SCID, or Structured Clinical Interview (DSM) and a Clinical Global Impression (CGI) screening, to verify that these subjects do have bipolar depression. Subjects will fill out the Montgomery Åsberg Depression Rating Scale (MADRS), which quantifies the severity of their depression. Candidates scoring over more than ‘mildly depressed’ will be allowed into the study if they also meet the appropriate vitamin D levels and do not meet any exclusionary conditions. Subjects need a vitamin D level under 30ng/ml to be considered appropriate for supplementation.

If the subjects meet these requirements, then they will fill out other baseline mood questionnaire assessments, including the Young Mania Rating Scale (YMRS). This scale is used to provide data on how elevated their mood is. They will also provide data for the Hamilton Anxiety Rating Scale (HARS). Subjects must also list their medications. This will provide the PI with a baseline for the medications that subjects are taking to treat their mood disorder. The subjects will also fill out a demographics survey. This includes data on their race, household income, education, etc. In addition to measuring their vitamin D, the team will also look at other lab values from the initial blood draw to assess their metabolic function. This screening is to exclude candidates with parathyroid disorders and a number of other exclusionary criteria that may disrupt vitamin D metabolism. Thus, the team will collect parathyroid, calcium and phosphorus clinical values as well as subjects’ vitamin D levels.

Over the following twelve weeks, the subjects will be expected to come into the clinic for their biweekly appointments. During these visits the team will collect data from the subjects using the several mood assessments, and surveys on their medication compliance and any changes in their bipolar medications, their compliance with the vitamin D versus placebo regimen, and their ratings of any side effects, severity and well-being.

At the 12th week visit, the subjects will have their final evaluations; they will provide their final mood ratings data, and have the final blood draw to provide data on changes in their vitamin D level, if any, to compare with their initial baseline levels. Lastly, they will provide their final data on their medications and vitamin D versus placebo compliance, and side effects, severity and well-being ratings.

As mentioned, the study relies on the use of paper-based instruments and forms for initial data capture. The main clinical data repository will be the *REDCap* spreadsheet. There are also *Excel* and paper files maintained by the Research Coordinator. These data keep track of the phone calls to the subjects, any participant dropouts, any delays attending biweekly appointments, and, if so, information on why they missed appointments.

For each of the questions on the MADRS there is a 0-4 scale for quantifying mood severity, so for each answer the team can conduct a sub-analysis of symptoms, and then an overall analysis of the total score. Similarly, the HARS and YMRS for anxiety and mood elevation use multiple questions with a numerical scale for severity. The CGI assessment uses a scale of 1-7, and the subject demographic questionnaire also uses an ordinal scale.

The team’s *REDCap* and *Excel* spreadsheets have not been created yet. Since all data are initially on paper, there will be a lot of filing and manual entering of these datasets into the *REDCap* software, which team members will access through a password-protected website on the Medical School’s server. The PI is concerned about illegible handwriting; there will be many paper documents coming in. She also tends to make notes during her clinical consultations. These might contain a qualifying comment that she jots down while sitting with the patient to help her to contextualize the subject’s data, but recognizes that these notes do not usually get entered into the database because they are difficult to analyze. The PI is interested in maintaining the confidentiality of her subjects’ and their data, and will keep the paper forms in a locked cabinet, in a locked room, in a locked building.

The PI is concerned about ensuring that the medical students set up the *REDCap* database correctly: spreadsheet fields will need to be correctly labeled and the appropriate value limits will need to be programmed correctly. In her last study, she had encountered a major hitch when transferring her *REDCap* data into the *SAS* software for analysis. Specifically her file date formats were not transferable. If a patient had come into the study at 01-01-12, for example, she would have collected patient history data, so all these previous file dates would be labeled as 2011 or earlier, but when she imported these data into the *SAS* software, the dates were reformatted to 2013.

The Research Coordinator will manage *Excel* files with data recording the number of candidates screened, how many subjects came in for a visit, and how many called on the phone. In addition, she will collect subject attrition data; for example, if the team screens 200 candidates, they collect data on the 50 who get to the next screening, and, out of those 50, they collect data on the forty who qualify, and then, out of those forty, they collect data from the qualified 30 subjects that actually participate.

The PI is concerned about a major staff change that will occur within her research team in the middle stages of her project: her Research Coordinator will be leaving the University. The PI feels her Research Coordinator is very organized and expects a smooth transition, but she knows they will have to be prepared. In her past projects, the naming conventions for files and folders and their storage locations were arbitrarily chosen either by the Research Coordinator or the medical students, so she will have to spend some time with the new Research Coordinator or any new students to help them locate and make sense of the data.

The PI is interested in working with the University’s data management liaison librarians to help draft a plan for managing her project’s data to help with this transition. Although the PI does not yet have an archival and curation plan for maintaining the posterity of her data post-project, she is also willing to work with the library to preserve and share her project data and make these datasets available in a repository or collection; she feels the more that can be gleaned out of them the better.