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

Andrew T Creamer, MAEd MSLIS
Wendy K Marsh, MD
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, HIPPA-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 wellbeing.

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 wellbeing 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.

**Discussion Questions**

1. **Types of data**
   
   a. What types of data are being collected for this study?

   b. How could the PI ensure all research assistants/medical students/volunteers use the same data sources and data definitions?

2. **Contextual details**

   a. What file formats and naming conventions could be used for the separate data sources and for the integrated file used for analysis?

   b. What impact would the naming conventions, proprietary software, and software updates have on later data access?

   c. What other contextual details would the PI specifically need to document to make the dataset meaningful to others?

3. **Data Storage, Backup, Security**

   a. Where and on what media could the data from each data source be stored?

   b. How, how often, and where could the data from each source be backed up?
c. How could the PI manage data security across research assistants/volunteers on the study for each data source?

d. How long following the completion of the study could the PI store the data?

4. Data protection/privacy

   a. How could the PI address any ethical or privacy issues?

5. Policies for reuse of data

   a. How could the PI create a de-identified copy of the data?

6. Policies for access and sharing

   a. Could the project’s data ownership or access be restricted, or could these data be made available in an Open Access Repository or library collection?

   b. Who would own any copyright?

7. Archiving and preservation

   a. What could be a long-term strategy for maintaining, curating and archiving these data?

   b. What data could be included in an archive?

   c. Where and how could it be archived?

   d. What other contextual data or other related data could be included in the archive?

   e. How long should the data be kept beyond the life of the project?
Summary of Teaching Points

Module 1: Overview of Research Data Management

Evolving data needs for the 3-stages of a double blind, experimental drug versus placebo study with multiple points and sources of data capture

Module 2: Types, Formats & Stages of Data

Issues with formatting dates on the files exported into SAS (e.g. writing the date using the format recommended by ISO 8601: YYYY-MM-DD or else the date stored with a file created on one computer will be changed if the file is moved to another computer.)

Issues when a Research Coordinator leaves in the middle of the project

Module 3: Contextual Details

Illustrates need for a data dictionary for spreadsheet field names and limit values


http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_049331.hscp?dDocName=bok1_049331

Illustrates need for a protocol for naming files and folders

http://researchdata.wisc.edu/manage-your-data/file-naming-and-versioning/

Module 4: Plan for Data Storage, Backup and Security

Illustrates need for a protocol for file and folder locations

Illustrates protocol for backup e.g. scanning of paper documents, use of external drives, or use of non-proprietary software formats for making backup

http://researchdata.wisc.edu/manage-your-data/data-backup-and-integrity/

Module 5: Legal and Ethical Issues

Illustrates HIPAA, IRB, FDA considerations

Module 6: Plan for Data Sharing and Re-Use
Illustrates need for a plan to share data and making post-project datasets available for re-use

Illustrates need for stripping patient identifiers before depositing in a repository

Module 7: Plan for Archiving and Preservation of Data

Illustrates need for preservation, archival and future access plans

Illustrates need for creating a dataset in an open source software format for preservation

http://www.openoffice.org/product/calc.html

Illustrate need for a metadata standard for description of a medical dataset

Stages of the Project

Stage 1: Pre-Study

1. Names and contact information for potential subjects
2. Consent and screening forms
3. Mood, Mania and Anxiety Evaluation Data
4. Blood values and lab measurements (e.g. Ca+, Phosphorus, vitamin D, etc.) and exclusion lab values
5. Clinical Global Impression assessment data
6. List of medications
7. Filing, scanning, data entry

Stage 2: Biweekly Evaluation

8. Bi-weekly date of visit/phone call, mood, mania and anxiety values, compliance values, wellbeing, side effects and severity screening, changes in medications
9. Data on appointments, delays, if any, or disruptions, dropouts, attrition, etc.
10. Filing, scanning, data entry

Stage 3: Final Evaluation

11. 12th week ultimate mood, mania and anxiety values, blood labs, compliance, Clinical Global Impression data, side effects and severity data screening
12. Filing, scanning, data entry
13. Data cleaning and preparation for export and analysis

Stage 4: Export data from REDCap to SAS for analysis

Stage 6: Abstracts and Publication

Stage 7: Preserving Datasets in an Open Repository and in the library’s collection and linking the publications with the datasets


**VitD and Bipolar Initial Visit Algorithm**

- Telephone screen
  - BD, feeling down, potential low vit D
    - N - > X

  Y
  ↓

1st appt
- Consent
- Draw VitD (gold sst) plus others (2\textsuperscript{nd} gold sst) - > VitD 25,OH to lab to process (2hr avg until in meditech)

  Hold other labs
  Continue w study Screen:
  - Demographic Questionnaire

  ↓

  - SCID, Bipolar
    - Y
    ↓

  - MADRS, >7
    - N - > X

  ↓

Completion of Baseline:
- Demographic Questionnaire
- HAM-A
- Medication Record
- CGI-S and I
- Ht and Wt for BMI

- Check meditech for VitD level, <30ng/ml
  - Y

  ↓

- Randomize to VitD vs Plcb (Co-PI, key master)
- Schedule f/up in 2 wks
- Send to Pharmacy to pick up Rxn
- Send rest of labs (2\textsuperscript{nd} gold sst) to be analyzed
Teaching Case Details

1. Types of data
   
a. What types of data will the PI create or capture?

   **Quantitative experimental data: clinical lab values and quantitative mood, anxiety, and mania ratings**

   b. How will the PI capture, create, and/or process the data?

   **Paper forms, Excel spreadsheets, REDCap spreadsheets for data capture, SAS for analysis**

2. Contextual Details (Metadata) Needed to Make Data Meaningful to others
   
a. What type of file formats and naming conventions will the PI use?

   **There is no protocol set for file formats and file and folder naming conventions.**

3. Storage, Backup and Security
   
a. Where and on what media will the PI store the data?

   **Paper forms will be locked in a cabinet. Project data will be kept only on the secured REDCap files on the Medical School’s server. REDCap is not typical open source (has some restrictions, not for commercial use, available to members of a consortium); the Research Coordinator will maintain paper forms and Excel files for managing the subjects’ visits. There is no protocol for the Excel file and folder locations or for scanning paper forms as a backup.**

   b. What is the backup plan for the data?

   **There is no protocol set for backing up the paper-based or electronic data.**

   c. How will you manage data security?

   **The REDCap software is accessed through a password secured website and data sit on the Medical School’s server; REDCap is HIPAA validated.**

4. Provisions for Protection/Privacy

   **HIPAA, IRB, FDA**
a. How is the PI addressing any ethical or privacy issues (IRB, anonymization of data)?

A Co-I holds the key to the study’s randomization. All research team members and medical school students complete CITI and IRB training. The study seeks IRB and FDA approval since it uses an investigational drug.

b. Who will own any copyright or intellectual property rights to the data?

The PI

5. Policies for re-use

a. What restrictions need to be placed on re-use of the project’s data?

The data must be stripped of patient identifiers; Consult with IRB regarding patient consent, institutional and funder policies regarding restrictions for data sharing

6. Policies for access and sharing

The PI is interested in sharing her data and working with the library to preserve and make the data available in a digital repository, but she has no protocol for the data’s preservation, archiving and/or future access.

Summary of Data Management Challenges

This case illustrates a project that does not yet have a data dictionary for the REDCap files or protocols for file and folder naming conventions (e.g. ISO 8601 date format), and data storage locations. The team needs to format the dates of their REDCap fields and files correctly. There is also no protocol established yet for backing up the data or for preserving the paper forms (e.g. pdf scans); clinical data are only in REDCap and not saved on any external drives or servers. The new Research Coordinator will need to know how to locate any subject files and forms he or she is expected to maintain. Similarly, he or she and any new students on the project team would need to know how to make sense of the project’s data within the REDCap and Excel spreadsheets because he or she will be unfamiliar with what the field names mean, what evaluation and clinical assessments they represent, and their maximum and minimum values. The case illustrates how the project could benefit from making a copy of the project’s data using an open source spreadsheet (e.g. ASCII) for backup and for later preservation and access purposes.