Katherine Webster

Grand Valley State University Department of Biostatistics

Dr. Robert Downer

Forma Therapeutics – Biometrics Intern

Summer 2022

Table of Contents

Section **Page**

Learning Objectives/Internship Objectives

- Learn biopharmaceutical industry knowledge and standards
- Learn to organize, alter, and analyze data in a clinical trial setting
- Clean raw datasets and code them into SDTM datasets following CDISC standards
- Create analysis datasets using CDISC ADaM standards
- Create tables and listings and figures for studies
- Write and review statistical analysis plans and perform analyses

Introduction

Forma Therapeutics is a clinical-stage biopharmaceutical company headquartered in Watertown, Massachusetts. The company focuses mainly on developing therapeutic drugs for blood-borne diseases such as sickle cell anemia and thalassemia. This internship was completed remotely in Grand Rapids, Michigan from June through August of 2022 with an extension into October. Projects worked on as the Biometrics Intern included: data standardization via SAS programming, statistical analysis plan (SAP) authorship, data management user acceptance testing (UAT), clinical data analyses, and report generation for use in Clinical Study Reports (CSRs). Further details on these projects and the role played in them will be discussed in the description of work section. Industry and professional skills within the biopharmaceutical industry were attained during this period and will be delineated within the discussion section of this report.

Description of Work

The first main area of work I focused on in this internship was learning how to standardize clinical trial datasets within a structured programming environment. This includes taking raw data collected from the clinical trials and transforming them into a standard set of variables that are defined by the Clinical Data Interchange Standards Consortium (CDISC). I began this process by reviewing the CDISC Implementation Guides for Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets. Next, I was able to apply these standards to unfinished, pre-existing SAS programs where I quality-control (QC) checked them against a main/production program and final datasets via PROC COMPARE. Generally, hundreds of variables are mapped from the raw, clinical database to the SDTM datasets that are in turn utilized in the ADaM datasets that contain the relevant analysis derivations for the

creation of clinically meaningful reports. Following review and QC, this is then presented to the company and eventually the FDA for drug approval. This work is critical to the clinical trial, the company, and most importantly, the patients who could benefit from an approved therapeutic drug.

Another aspect of my work included learning what a Statistical Analysis Plan (SAP) entails and how they are written and reviewed. The SAP follows the trial protocol closely and details how the collected data will be processed and analyzed with specifics on types of statistical tests and analyses. I had the opportunity to work closely and co-author one of Forma's Phase 1/2 trials working with patients who have sickle-cell anemia (SCD). This process takes many months and rounds of review before it is finally approved to be published and circulated. A good understanding of statistical tests, how they apply in a clinical trial setting, and how the clinical endpoints are defined and operationalized are crucial for understanding how to write a SAP. Additionally, mock versions of tables, listings, and figures (TLFs) must be drafted to show how the data will be presented in a physical manner. This can be done with a template within Microsoft Word or can be coded in SAS (where they are generated with collected data) and be exported to a Word document. I was able to help draft some of the tables and listings which included brainstorming how summary statistics and by-patient data would be presented.

Lastly, I was able to contribute on a couple of miscellaneous projects in the biometrics department. These included User-Acceptance Testing (UAT), and Sponsor Review of clinical data within data management. UAT involved using an Electronic Data Capture (EDC) system via Medidata Rave EDC. This is an online database where all patient data is kept confidential and can be used to review the accuracy of the collection of the raw data. This system can also be used for UAT before raw data is uploaded. A skeleton of the database is created to ensure accurate

raw data capture. Data managers then do tests of the system by "creating" new patients with various demographics and information. I was able to partake in UAT as an intern when we had our weekly Biometrics Exposure Program which aimed at showing interns how the biometrics department functioned.

Sponsor review of datapoints is crucial to the accuracy of data collection as sometimes human error is introduced, such as adding an extra zero. I partook in this with one of the clinical scientists where he helped me understand and validate certain values and ranges for lab tests and vital signs. For example, if a heart rate read to be 1200 beats/min, this would be flagged for review since this is understood as physically impossible. For example, the site may have meant to record 120 beats/min but inadvertently added an extra zero. This speaks to the important and necessary nature of doing these data review checks.

Internship Discussion

The learning objectives for this internship were almost completely achieved. The main objective was to perform programming responsibilities and this consumed the majority of my time as an intern. I successfully learned how to code in a clinical trials setting and will use this knowledge in my future positions. One objective I wish I had more time with was statistical analysis of the collected data. I didn't get to see the SAP I was working on be finished and enacted after data was collected. This would have helped me understand how the ideas are applied to a physical dataset.

I gained many useful skills from this internship experience in both scientific and professional areas. Scientific skills that I gained include knowledge of CDISC standards, coding for clinical trials datasets, writing of SAPs, database management and review, and general knowledge of sickle cell anemia and how it impacts people's everyday lives. Professional skills I

developed include learning how to work for different bosses, communicating to a whole team, relaying and admitting when mistakes occur, and carrying myself through adversity.

In my experience as a SAS programming intern, I was adequately prepared for the scientific portion since over the course of my degrees I have had many years of SAS coding experience. This is the strongest skill I had going into the internship and helped me greatly. One area I believe that could be touched on more in the biostatistics PSM program is the focus on clinical trials. We had one class on clinical trials but most of the knowledge I learned there was not applied to my position at Forma. While learning how a clinical trial is conducted and viewed in an ethical perspective is important, it was not a focus that this internship tasked me with. Learning more of the statistical and programming part of clinical trials in that class would have been helpful. I also think it would be helpful to have more guest speakers who work as biostatisticians or clinical programmers to give presentations and talk about their work. I found the presentation by Rhonda Pardue at the end of the semester to be insightful into what its like to work within industry.

I also felt I was adequately prepared for the professional aspect of the internship. Specifically, PSM 650 (Ethics and Professionalism) and PSM 662 (Seminar in Professional Science) were extremely helpful in giving us an insight into what the workplace is like within these scientific sectors. I felt like I had an appropriate professional demeanor even when a couple of bumps in the road arose. The skills I learned in those classes made me feel ready to enter the work world and to handle any problems or communication issues within it.

Throughout this internship I encountered a few challenges, that now looking back, gave me vital knowledge of how to act and perform in a professional setting. The learning curve for the first couple weeks and even months was quite daunting at first. I quickly learned I could not

simply “Google” most of what was being required of me to learn like I could in college with most things. These things required actual experience in doing the tasks rather than being able to mindlessly search for the answers and get them in seconds. Another challenge was learning how to communicate with different types of people with different personalities. I learned that what you interpret can be understood very differently by other people. I believe this is something that everyone runs into at every job, and it was nice to have such a positive environment at Forma to learn these lessons. My supervisors were very understanding and worked through any issues that came up with me, and I felt like they really, truly cared about me and my well-being.

This internship experience overall was a great steppingstone between being a master’s student and being a professional within industry. I believe it should be continued to be a requirement within the biostatistics master’s program since I gained invaluable knowledge and experience from it. The only thing I thought was difficult was finding and securing the internship. Surely this has many variables such as timing and availability, but I felt I could have used more help from the Biostatistics/PSM department in the search for internships. If I could do this all over, I would have tried to ask for more help and at an earlier date. I’m still very grateful for the internship experience I had and believe it is actively setting me up for success within the professional world.