Internship Report

(PSM 691 – Internship)

Submitted by:

Biostatistician Intern, Corewell Health

M.S. Data Science and Analytics

Grand Valley State University

Internship Supervisor:

Jessi Parker

Manager, SSAS

Corewell Health

Date: December 1, 2023

Table of Contents

Contents

Internship Objectives

1. Apply learned statistical techniques from academic settings to analyze actual healthcare data at Corewell Health, utilizing SAS and R proficiently.
2. Develop the ability to present complex statistical findings in a clear and understandable manner to facilitate decision-making processes.
3. Ensure compliance with privacy regulations while supporting the accessibility of data for research initiatives.
4. Collaborate effectively with interdisciplinary teams, healthcare professionals, and principal investigators to provide statistical support for research projects.
5. Improve analytical capabilities through hands-on experience with diverse healthcare datasets.
6. Gain a comprehensive understanding of healthcare data security protocols and privacy regulations.
7. Engage in continuous learning and professional growth within the field of data science and analytics.

Introduction

Corewell Health is a not-for-profit health care system that was created in 2022 with the merger of Beaumont Health and Spectrum health. Headquartered in Grand Rapids, Michigan, it is the largest health care organization in West Michigan with over 60,000 employees. ("Corewell Health," 2023)

My internship was under the Scholarly Activity & Scientific Support (SSAS) team, which is managed by Jessi Parker, who was also my supervisor during the internship. The team I worked with worked remotely, hence making my work remote as well. SSAS is further divided into 5 parts which includes Scientist, Biostatistician, Epidemiologists, Contractor, and Data Scientist. So, more specifically, I worked with the Biostatisticians. Alicia Stowe and Cuyler Huffman acted as my mentors helping me through the internship responsibilities. The SSAS team helps provide statistical, epidemiological, data extraction and scientific support to the research teams throughout their project. Working closely with the Principal Investigators and the research teams on their statistical and data requirements, we continually needed to be in communication and touch with them.

My role in the team focused on providing de-identified, limited and identifiable datasets and statistical analysis to researchers. To fulfil these roles, I worked with SQL to transform and pull the requested data and used SAS and/or R to perform statistical analysis, visualization, and reporting.

<u>Description of Work</u>

As a Biostatistician Intern at Corewell Health, my role primarily focused on leveraging statistical analysis tools and data management and acquisition techniques to support healthcare research initiatives. The internship provided an immersive experience in applying data science methodologies within a healthcare setting.

Honest Broker Program

The majority of projects I have worked on during the internship so far involved taking care of the honest broker requests (Level 1), that were assigned to me by the team. These requests were made by the Principal Investigator (PI) with their research team who are responsible for a research study. During this role, I acted as a neutral third party and provided the requested data to the research team in compliance with the Honest Broker Policies and Procedure ensuring data accuracy and safety.

Based on where the research process for a research team is the dataset that is requested varies. It could be one of four, Prep to Research, Identifiable, Limited, or De-identified. On a Prep to Research request I reported the population count that we have in our databases to let the research team come to a decision if it would be enough for their research study. This was only a high-level summary of the requested population and not a dataset. Other three are datasets that as per request were either de-identified or limited to some HIPPA Identifiers or identifiable.

Once the PI fills out and submits the Honest Broker Form and the team assigns me a project, I looked at the data request form in REDCap (Research Electronic Data Capture) to see what kind of dataset is requested for the project, if there is any inclusion or exclusion criteria like a timeframe, location, demographics, diagnosis etc. and what variables are expected to be in the final dataset. I also contacted the team or the PI if I needed some more information on the request. And since the data is stored in a Relational Data Model, I used the company standard SQL Server to pull all the requested data. By writing SQL queries with complex joins the requested data was pulled and once pulled, I sent the SQL codes and the final report/dataset to the team for review. Once it was reviewed, I sent the reports and/or datasets to the PI using proper channel (either in cloud or REDCap import).

Data Cleaning, Statistical Analysis and Visualization

During my internship, I successfully completed several projects that focused on data cleaning, statistical analysis, and the creation of visualizations. A major aspect of importing data into REDCap required attention to the structure of the dataset and attributes of variables. Raw data was cleaned thoroughly to comply with REDCap standards. This entailed adjusting variables, both addition and removal, to align with the platform's requirements. Throughout these projects, I relied extensively on RStudio to efficiently handle data processing tasks.

One specific project centered on conducting statistical analysis to assess the significance of differences in variable distributions pre- and post-intervention. Employing SAS, I performed a series of tests tailored to the nature of the data. Continuous variables underwent a normality check, and if they exhibited normal distribution, a two-sample independent t-test was applied. Conversely, non-normally distributed data underwent a Wilcoxon Rank Sum test. Categorical variables were subjected to analysis using the Chi-square test. Furthermore, in instances where more than 20% of expected counts in a cell were below 5, a Fisher's Exact test was executed. The outcomes of these analyses were compiled into a comprehensive report using the PROC REPORT procedure, formatted appropriately for each variable. Subsequently, I submitted the finalized report to the team for review and verification.

Also, I worked on a project focusing on visualizing survey data. The survey required separate reports for each question in the questionnaire. To start, I made a detailed SQL query to get a dataset for analysis. Once I had the data, I cleaned it up to make sure it was good for visualization. Then, I made a special SAS Macro to make frequency tables. This Macro could handle each question and its variables. It made tables where the different values of variables were shown as columns and the various answer choices were shown as rows. Along with tables, the Macro could also make stacked bar charts to show how the data was spread out.

Time Tracking

The team I worked with required time tracking as a fundamental practice. It required meticulously recording the amount of time allotted to each project component. Every project had thorough records kept that detailed the amount of time spent on various phases. This technique captured the precise timeframes for tasks such as data gathering, analysis, report generation, and other project-related activities, covering every aspect of the projects to which I contributed. We were able to evaluate project progress, distribute resources effectively, and draw lessons from the past for future improvements or related initiatives because of the methodical tracking that promoted transparency.

Internship Discussion

The coursework in the PSM program significantly helped me in improving my professional skills during my internship at Corewell Health. While the program primarily focused on enhancing behavioral aspects, such as developing professional communications like emails, resumes, and cover letters, it proved very helpful in me securing the internship opportunity. It also provided invaluable insights into understanding organizational structures and operations. Although the program didn't directly address technical aspects related to the work, it played a crucial role in preparing me for the real-world business environment and understanding roles, responsibilities and about accountability within the context of an analyst. This allowed me to effectively achieve the internship objectives and further develop my personal and professional skills during my time at Corewell Health.

With two semesters of internship experience I'm pleased to report that the internship allowed me to fulfill most of the set objectives. I utilized statistical techniques learned in my school to analyze large and complex healthcare datasets using SAS and R proficiently and querying large and complex databases using SQL Server. I helped manage requests for healthcare data, making sure the information was accurate and secure while following the rules. Moreover, ensuring data compliance with privacy regulations while supporting data accessibility for research initiatives became an integral part of my workflow and collaborating effectively with interdisciplinary teams and healthcare professionals improved my teamwork and communication skills. This internship also helped improve my skills in analyzing data using programs like RStudio and SAS, where I looked at how information was distributed, tested ideas, and made detailed reports including tables and charts.

This internship opportunity helped me get a diverse skill set on both scientific and professional aspects. I improved my proficiency in statistical analysis using SAS and R, specifically in assessing variable distributions and conducting a wide range of statistical tests. I developed skills in data cleaning, visualization, and the creation of detailed reports using tools like RStudio and SAS PROC REPORT. Professionally, I enhanced my abilities in managing honest broker requests, navigating REDCap for data acquisition, and following the data compliance regulations. As most of the projects I was involved in focused on handling data acquisition requests, I was able to improve my SQL skills by writing optimized SQL queries. The experience also helped improve my time management and communication skills, especially in presenting complex statistical findings.

One unexpected challenge during this internship was the remote work setup, which affected the usual in-person interaction and connectivity within the team. Working remotely, especially as a biostatistician intern and working with healthcare data, initially presented some difficulties. Understanding the medical terminology used in the requests from research teams was particularly challenging at the start. I was worried that this would hamper my work and limit me in some way. However, the timely assistance from my mentors Alicia and Cuyler proved very helpful. Their support and guidance were vital in navigating through the complexities of projects, understanding technical requirements, and overcoming initial hurdles. Their mentorship not only alleviated my concerns but also enabled me to adapt and excel in this unique work environment.

I would like to extend my heartfelt gratitude to my mentors ███████████, for their consistent support and guidance throughout my internship. Their prompt assistance in understanding project requirements and technical aspects significantly contributed to my learning and professional growth. I also want to express my appreciation to my supervisor, Jessi for her invaluable mentorship and continuous encouragement. Their collective guidance played a vital role in helping me navigate and succeed in this internship at Corewell Health.

References

Corewell Health. (2023, November 27). In Wikipedia. https://en.wikipedia.org/wiki/Corewell_Health