Top of Form

|  |  |  |
| --- | --- | --- |
| **Client:** | Elina Dawoodani | **File:** 24-148 |
| **Dept:** | Nutrition Science | **Faculty:**  **Student:** |
| **Date:** | 3/5/25 | **Initial Meeting:**  **Follow-up:** |
| **Consultant and Attendees:** Elina Dawoodani, Sumeeth Guda, Dr. Bruce Craig, Dr. Qingland Ding | | |
| **Statement of Problem:**  How does iron deficiency contribute to the shrinkage of the hippocampus and subsequently how does this contribute to the increased risk of obesity? | | |
| **Goal of this Project:** PhD Dissertation | | |
| **Background:**  The client is a PhD student in Nutrition Sciences who used the consulting service before in the previous semester. The main goal of her project is to analyze the NIH ABCD (Adolescent Brain Cognitive Development) dataset for her PhD dissertation to accomplish and investigate the following 4 tasks:   1. Investigate the relationship between iron levels and the volumes of the left and right hippocampus in adolescents. 2. Analyze how hippocampal atrophy is associated with adiposity, focusing on whether structural brain changes contribute to increased BMI and waist circumference. 3. Examine the relationship between iron status and adiposity and determine if hippocampal volume mediates the connection between iron status and both BMI and waist circumference.   She came to the SCS this semester because she has a grant application that she needs to file for and she wanted to know if the analysis methods she proposed in the grant proposal are statistically sound. | | |
| **Progress of project at this time:** Analysis (All data collected) | | |
| **Relevant information presented at the meeting:**  Dataset Description:   * 11,500 participants in the baseline cohort. * The dataset has several variables, but the client is focusing on iron biomarkers, hippocampal variables and variables/covariates associated with weight (BMI/waist circumference).   Meeting Notes:  At the start of the meeting Elina discussed and shared the physiological and statistical aspects of a study specifically by sharing how iron deficiency impacts the hippocampus by shrinking it and in turn how this influences the risk obesity. The focus is on understanding the relationship between low iron levels, hippocampal shrinkage, and increased body fat. She is working on a new latent effect model to show how these factors are interconnected. The study will involve children and adolescents, with data sourced from NIH’s ABCD dataset, which includes children aged 9-14 years. The variables of interest include ferritin, hemoglobin, hippocampal volume, weight, height, waist circumference, and various confounders (age, stress, intracranial volume, pubertal status, activity level, and stress level). The data has been collected longitudinally, with measurements taken every 1-2 years, and includes a total of 11,500 patients.  However, one issue is that the dataset is not yet cleaned, and there are unknowns regarding which patients have data across multiple time points. Preliminary exploratory data analysis (EDA) has been performed, but further cleaning and organization are needed. Dr. Craig noted that, depending on the model, gaps in data could still be analyzed, but minimizing gaps would be ideal.  The client originally had four research aims but has decided to focus on three key objectives:   1. Investigating the relationship between iron status and hippocampal sub-volumes. 2. Examining the association between iron levels and BMI. 3. Analyzing how iron deficiency and BMI are mediated by hippocampal shrinkage.   This model followed this DAG structure:    The study will explore these aims both at the individual level and across groups. While the research seems to resemble a regression problem, it may align more closely with a mixed model/repeated measures approach, given the longitudinal nature of the data. Elina suggested initially separating participants into groups based on their iron status (sufficient or deficient) and examining individual and group differences. Dr. Craig emphasized that iron status can fluctuate, and individual participants may experience varying levels of iron deficiency over time. Elina noted that iron deficiency often begins early in life, making it difficult to track its onset, but it can be crucial for understanding long-term hippocampal changes. Once hippocampal damage occurs due to iron deficiency, it is permanent, even with iron supplementation.  Lastly, Dr. Craig proposed examining the dataset in terms of how iron deficiency at different time points (1 year, 2 years, or 3 years) impacts hippocampal development.  Dr. Ding then steered the conversation back to their original reason for coming to the SCS to see if the proposed analysis methods were valid and statistically sound for the data.  One of the issues that we had at the start was that we didn’t look at the data, and moreover Elina had not processed or done the exploratory data analysis on the data yet. Dr. Craig emphasized that the quality and cleanliness of the data would largely determine the analysis method. Since Elina has not yet processed or cleaned the data, it is premature to suggest an analysis technique. Once the data is cleaned and the EDA is complete, a clearer recommendation can be made. Proper data cleaning will help us understand what variables are available and how to handle missing data, which is crucial for the analysis.  Additionally, it was discussed whether to include only participants with complete data (i.e., no missing values for key variables) or whether to include participants with missing values in their longitudinal data. Dr. Craig recommended including as much data as possible, even if there are gaps. By identifying patterns in the available data, it may be possible to use participants with missing values in the analysis. This was especially important for the ferritin data, because Dr. Craig suggested comparing variables across different years to explore associations, even in the presence of missing data. He emphasized that prediction can be complicated with missing data, but data cleaning and exploration would help clarify how to handle it.  Because of the longitudinal nature of the data, there was a time factor which wasn’t explicitly stated on the DAG. Dr. Craig suggested that incorporating time into the analysis could help with understanding longitudinal trends. Using the first year of data as a baseline would be helpful for comparison purposes. The time factor would also guide whether to focus on individuals with complete longitudinal data or subsets of participants who may have dropped out over time.  To handle the missing values within the iron level values, Elina considered using imputation to estimate missing iron level values, but Dr. Craig recommended first extracting and cleaning as much data as possible before deciding on imputation techniques. Imputation could be considered if the missing data is significant, but this should be done cautiously. If she uses any imputation on the baseline, then essentially that would nullify any interpretations of the longitudinal trends using that estimated baseline. | | |
| **Recommendations for Design and/or Analysis:**  What needs to be completed is that the client should currently focus on cleaning and processing the data to understand its structure and missingness before we can help her select an appropriate analysis method. Doing the exploratory data analysis and plotting the trends of individuals could be useful in determining the ideal test. But the main focus needs to be that she extract as much data as possible from the ABCD dataset. Once the data is ready, a better recommendation can be made, and techniques like mixed models or longitudinal analysis could be used, depending on the dataset's completeness.  If she needs help from the consultant, even though we can’t look at the dataset, she can give a snapshot of the dataset or a toy example based on the dataset to Sumeeth, and he can do a procedure in R to reorganize the data for her intended analysis purpose. And work with him to get the datasets in order and make sure her understanding of the patterns are valid and sound. Depending on the timing, another meeting can be set up to discuss the ideal analysis methods for the datasets. | | |
| **Who will carry out these actions?**  Client:   * Process and clean the dataset and do basic exploratory data analysis on the data itself. * Give an example dataset based on the master ABCD dataset to the consultant or transform it in a way that ensures the privacy of the data.   Consultant:   * Work with the client and give her advice on the EDA process and give her example R code using the toy data as an example to process and visualize the data. | | |
| **Status:** Follow Up needed. | | |

Bottom of Form