Top of Form

|  |  |  |
| --- | --- | --- |
| **Client:** | Sarwat Amin | **File:** 25-023 |
| **Dept:** | Biomedical Engineering | **Faculty:**  **Student:** |
| **Date:** | 2/18/25 | **Initial Meeting:**  **Follow-up:** |
| **Consultant and Attendees:** Sarwat Amin, Sumeeth Guda, Dr. Matthew Ward, Dr. Chong Gu | | |
| **Statement of Problem:**  How early can we detect a deviation of physiological data from pre-vaccine baseline? | | |
| **Goal of this Project:** Journal Article / PhD Dissertation | | |
| **Background:**  The client is a PhD student in the BME department who is investigating the long-term effects of the COVID-19 vaccine amidst the electrocardiogram (ECG) data from 84 people. They are estimating heart rate variability (HRV) metrics from the ECG sensors and using the metrics along with the biological data (skin temperature, respiration, etc), they are discovering which metrics are the earliest to deviate / change from the pre vaccine baseline. They collected this data through an observational study where they collected data from a pre and post vaccine time period.  The client and her advisor came to the SCS because they needed help analyzing the pre-vaccine data since it violated their initial assumption that the collected data would be normally distributed. Additionally, she designed an 81 by 22 treatment matrix and fitted a stepwise logistic regression model to analyze the results. They wanted validation that their analysis methods were valid and statistically sound. | | |
| **Progress of project at this time:** Analysis (All Data Collected) | | |
| **Relevant information presented at the meeting:**  At the beginning of the meeting, Dr. Ward explained that their lab’s main focus is to analyze data with heart rate physiology sensors, and with a special focus of measuring the time dependent data with sensors. Within Sarwat’s project, using a wearable heart rate sensor, they are computing the hourly means of different physiological and HRV metrics. The timing of this study is split up between 4 days pre vaccine where they collect the baseline data, they apply the vaccine, and then collect 5 days of post vaccine data. 84 participants were used to collect the data from in this study.  With respect to the deviation and noise from the sensor measurements. They mentioned that since the heart rate is affected by daily routine and circadian rhythm, the data and heart rate value would be different than the heart rate in the evening, they wanted to account for this and account from the deviation of the pre-vaccine values and how early they can detect the deviation. They did this by taking hourly statistics from the pre-vaccine baseline data for each individual in the study then compared per hour the pre-vaccine value to each and every post vaccine values to find out the difference between the pre and post values. They came to the SCS because they would like to understand how to deal with time dependent differences between pre and post baseline data with the vaccine. With respect to modelling, they wanted to know if a stepwise logistic regression model is useful in characterizing the metrics and testing the difference between the pre and post. Logistic regression was their rationale, since they collect the symptoms and side effects through a self-reporting survey with mostly (yes/no) questions.  Right off the bat, Dr. Gu and Sumeeth recognized that while the sensor data and vaccine data would be precise, ultimately by having a survey instrument to collect the symptoms ultimately this would generate a lot of noise within the model. Additionally, since Circadian rhythm and activity level are being tracked, this could lead to fairly inconsistent data due to every individual having different levels and patterns.  Addressing the client’s main question, they collected data for 22 HRV metrics. They mentioned that they wanted to calculate the distance between the pre and post vaccine datasets, and compare the detection of a significant response (When the data deviates from the natural circadian rhythm / pre vaccine baseline). Dr. Gu explained that the client would need to establish a differencing norm in a way to register the time points. (i.e: say you have 5 days continuous measurements of pre vaccine data, then in the 5 days they might not be a day for day matching between the datasets). The main takeaway was to make the time segments comparable to each other, since activity levels fluctuate, and the baseline comparison could be off. This was stressed in the meeting, since this study has a lot of background noise within the data collection process and while in general, circadian rhythm is fairly consistent, the individual’s activity level skews the baseline and noise.  Dr. Gu mentioned that to establish a consistent baseline he would do the following:  - Assume circadian rhythm follows a clock and activity level fluctuates.  - From 12 AM to 12 AM (next day) have a hard cut point with the time and take the 5 days pre and instead of taking average from 9 AM to 10 PM across 5 days, do a daily adjustment by comparing the average and subtract the whole day from that average. Standardize the data by day, but establish the time period.  The main rationale behind Dr. Gu’s suggestion was because the day to day the baseline could change, but normalizing the data can make it easier to compare baselines. Adjust for the daily span, look at the difference between the highest and lowest values. The total variation could be dramatically different between the two values. Whether this method is reasonable or relevant depends on the client's subject matter knowledge in this domain, the key is to apply their domain knowledge by determining what norm level is significant the need to apply what they feel is deviation to the norm. Before they do this, they need to define what is the baseline and what is significant deviation, but don't use standard statistics (mean, variance, etc.) to define the deviation. Domain knowledge is more practical for this observational study. Quantitative measures could be too harsh with detection as the CI could be too wide. Empirical measures could be too lax. Domain knowledge is a middle ground approach. | | |
| **Recommendations for Analysis:**   1. Change the hard cut point from 9AM – 10PM for the 5 days post vaccine to a 12 AM – 12 AM period. By doing this the baselines from the pre and post can be easier to compare. To minimize the noise, normalize the data collected. 2. With regards to interpreting the data and imposing a significance threshold on the difference, apply domain knowledge to the analysis since the noise from the data collection will make the traditional difference tools (TV distance, Confidence intervals, etc), too harsh and conservative. Look at what similar studies did to detect the difference and what they deemed to be significant. | | |
| **Who will carry out these actions?**  Client:   * Follow the instructions Dr. Gu recommended regarding changing the data analysis period from 9 AM – 10 PM to 12 AM -12 AM to make the pre and post datasets more comparable. * Use domain knowledge to determine the significance threshold.   Consultant:   * Answer any clarifying questions that the client has regarding the analysis of the data. | | |
| **Status:** Continuing | | |

Bottom of Form