Tristan Macy

12/11/2020

**SAPS/SAPS-PD Background:**

For any disease or condition, developing appropriate biomarkers is an essential step in the process of creating a treatment or cure. Developing biomarkers for neurological diseases is particularly difficult. The scale for the assessment of positive symptoms (SAPS) is a rating scale developed to measure positive symptoms in schizophrenia. The scale is comprised of four domains: hallucinations, delusions, bizarre behavior, and positive formal thought disorder. Since its creation, the SAPS rating scale has been modified for use in other neurological conditions. One recent adaptation applies to psychosis in Parkinson’s disease (PD) patients and is called SAPS-PD.

**Dataset**

The SAPS-PD is comprised of components from both the hallucinations and the delusions domains from the original SAPS. The hallucinations domain contains 5 questions and the delusions domain contains 4 questions. Each question is judged on a Likert scale from 0 (None or No Symptom; Good) to 5 (Severe Symptom; Bad). The final score is the mean of the 9 Likert scale questions.

In a preliminary study of the natural history of PD-related psychosis in a group of PD patients an investigator enrolled patients from two hospitals and followed them for six months. Each patient was to fill out the SAPS-PD at baseline, after three months, and again after six months. Data is split between hospital center A and hospital center B.

**Data Cleaning:**

Due to the nature of the study, the patients within the datasets are difficult to follow-up with, thus both Center A and Center B have a variety of missing data that needs to be addressed. Any SAPS-PD score that is left blank will be filled with a. and a new variable center will be made. Center A will be referred to as Center 1 and Center B will be referred to as Center B.

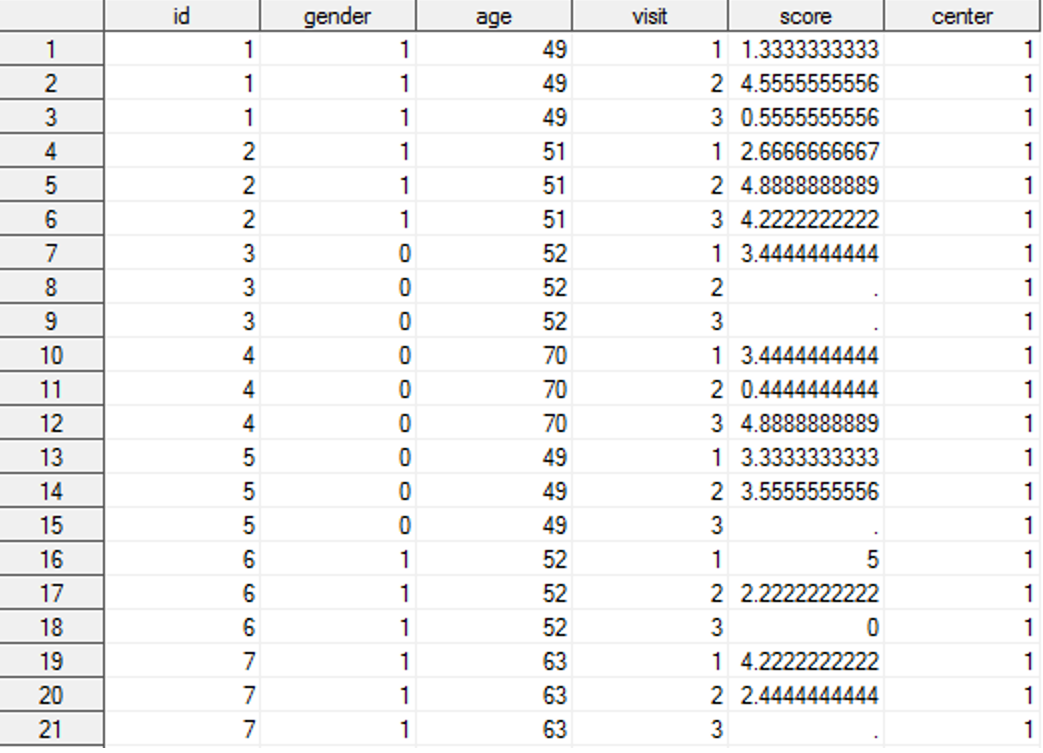


Figure : Center A Revised

**Figure 1: Center B Revised
**

Figure : Center B Revised

The two datasets are also in differing formats, in order to fix this, we will convert Center B from its wide format to match that of Center A’s long Format and then concatenating the two datasets together into one dataset.

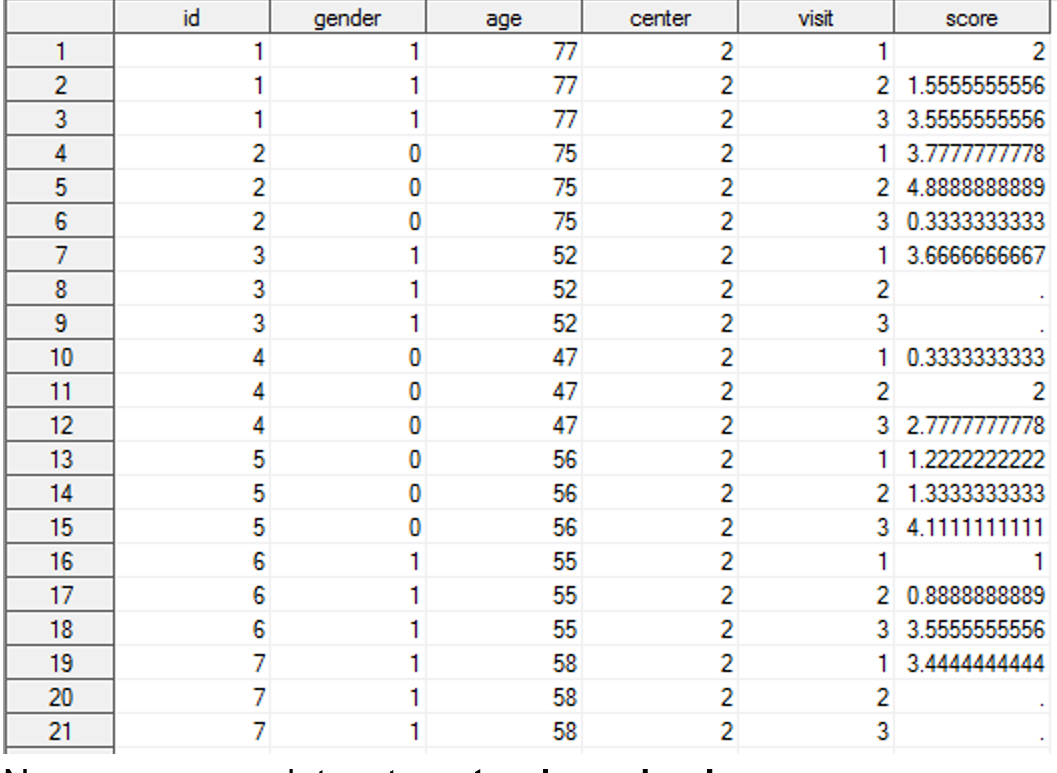


Figure : Center B Long Format

**Exploration:**

With the two datasets cleaned and merged into one, we are now able to generate meaningful observations from the data.

We will first generate a report using a PROC REPORT statement to provide us with a summary of key demographic variables. We will first look at the Patients Age Group with Adult (0-64), Senior (65-74) and Elderly (75+).

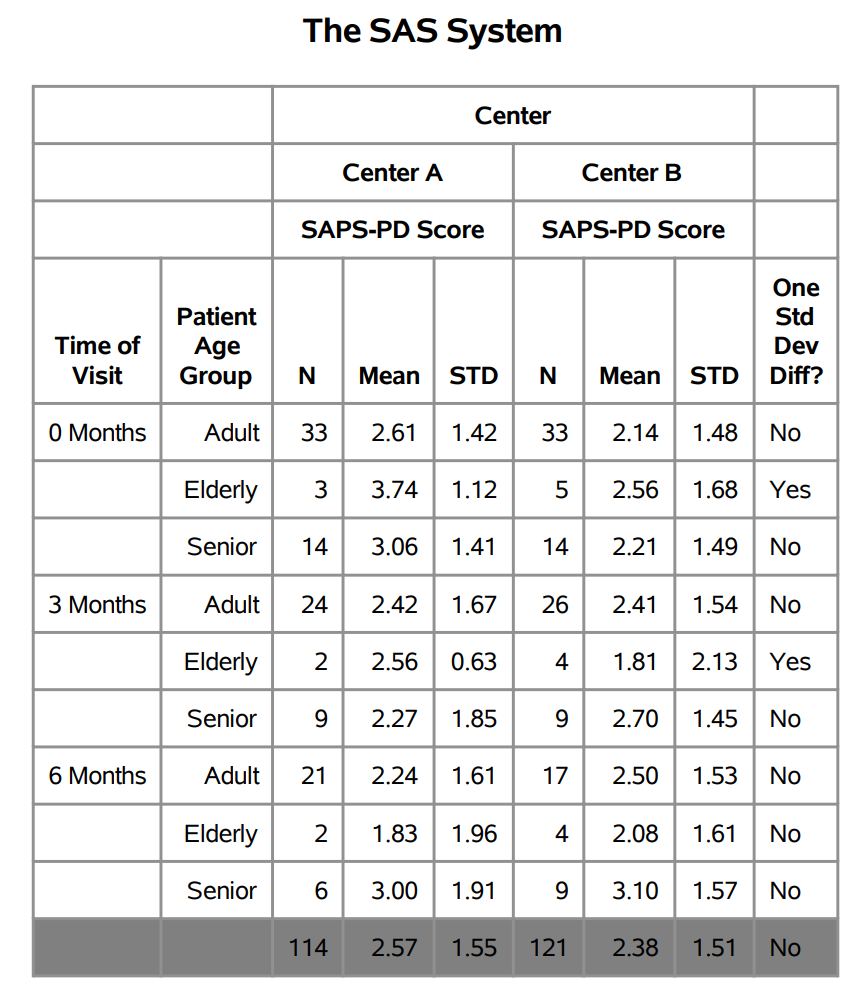


Figure : Report (Age Group)

The patient makeup of this dataset appears to be a majority of adults with the second largest being seniors leaving only a few Elderly. With the addition of the “One Std Dev Difference” category, we are able to see that the elderly patients in the 0- and 3-Month’s category are experiencing positive symptoms at a significantly higher rate than those of the other groups at 0 and significantly lower at 3. The elderly branch however has the fewest people, which may contribute to the higher score.

Next, we will generate a report with a focus upon the patient’s gender. In addition, rather than looking at whether there is a standard deviation difference, we will generate a t-statistic score assuming that the two centers have unequal variances. We will also add a summation at the bottom of each Time of Visit rather than overall.

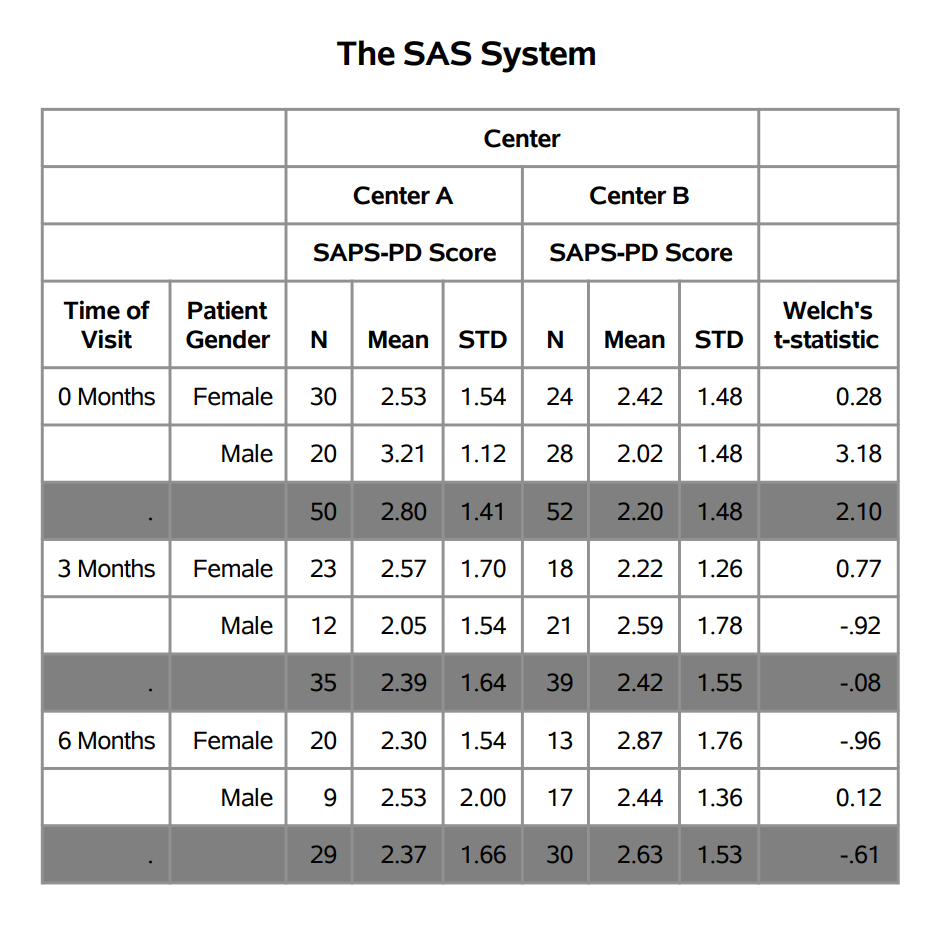
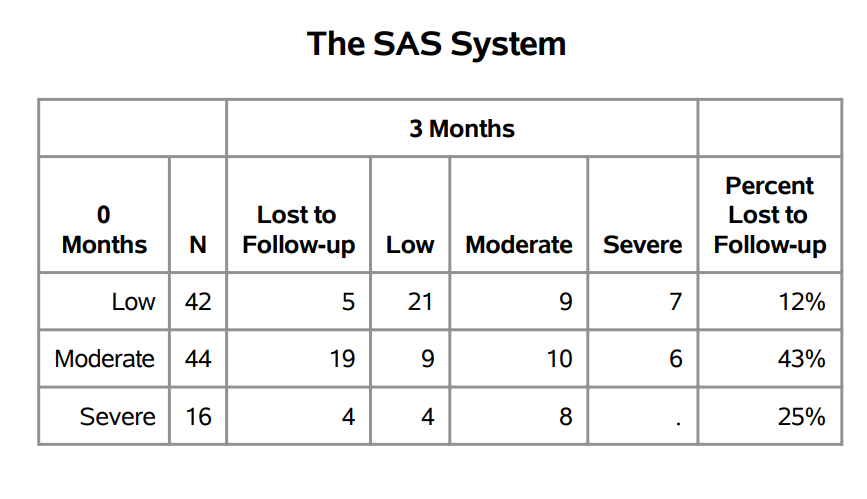
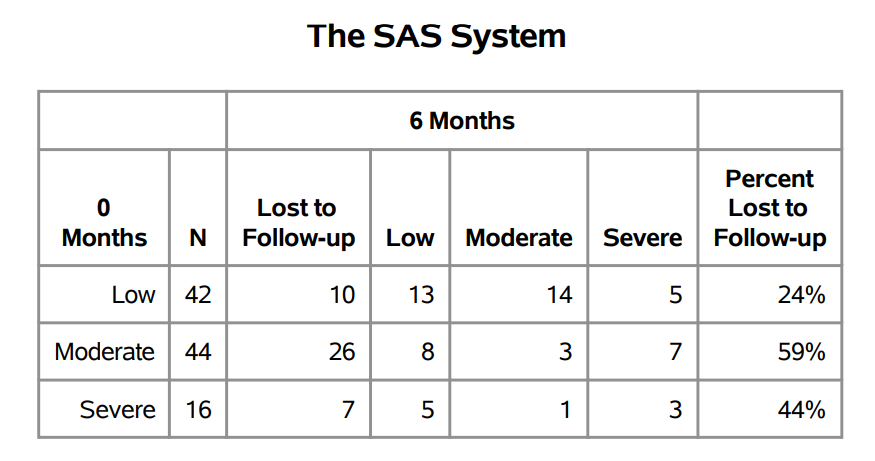


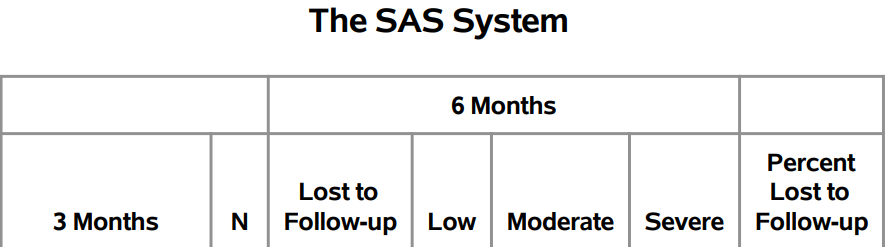
Figure : Report (Gender)

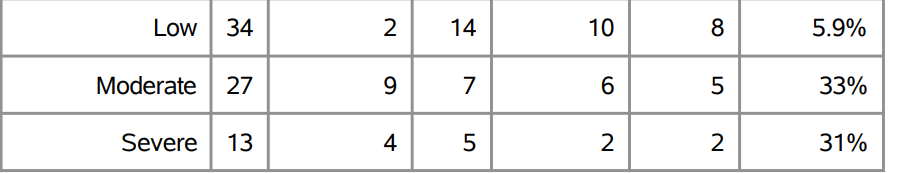
From the report, we can see that the majority of patients are women and that the men have higher t-statistics within the 0 and 6 month categories with women having the higher in the 3-month category. From the t-statistics, we can see a dramatic decrease from the time of visit with the scores dipping into the negatives within 3 months and 6 months.

Given that losing patients to follow-up is prevalent within datasets such as these, the last reports we will do will determine whether the severity of a patient’s psychosis will impact whether they are lost to follow-up. The severity of psychosis will be measured with their SAPS-PD Scores with 0-2 being Low, 2-4 being Moderate and 4+ being Severe.









Looking at the 0–6-month loss, there had been a 60% loss in Moderate and 44% in Severe demonstrating a major problem in the ability for patients to follow-up in the clinical study. Even the low cases have a 24% which is a large amount of loss of data within the dataset. Looking at the two windows of 0-3 and 3-6 we can see that the majority of the loss happened within the 0-3 Month period. The concern with SAPS-PD studies being difficult to conduct are thus well founded with the loss of patients severely affecting the data.