



Client 24-148 data cleaning issues

From Guda, Sumeeth Krishna <sguda@purdue.edu>

Date Thu 3/13/2025 4:57 PM

To Craig, Bruce A. <bacraig@purdue.edu>

 1 attachment (629 KB)

Results_Basic_FeAT.docx;

Hello Dr. Craig,

Elina emailed me regarding her dataset. To recap what she said, she removed the missing values from her dataset and created a subset with the variables of interest. However, she only has 22 participants (from year 2, year 4) in this cleaned dataset which is lower than her proposed 208 she computer during the power analysis.

She recapped what she is investigating:

1. Looking at the longitudinal effect of iron biomarkers on BMI, via the mediation of hippocampal volume deficits.
2. The study didn't measure any iron biomarkers (or any blood variables, actually) until year 2 follow-up, hence year 2 becomes my baseline for this study.
3. Even after year 2, they measure iron biomarkers of select individuals only.

Then her cleaning results:

644 participants from 2_year_follow_up_y_arm_1 and

182 participants from 4_year_follow_up_y_arm_1 (chose year 4 because MRI valuables are measured once in two years)

She looked for individuals in both years 2 and 4 but she was left with only 22.

The reason I am contacting you is regarding what advice should I give Elina regarding her dataset. It seems like she is only looking at complete case data hence why she has a low number of participants. However, I know in the IM meeting it was established that if she doesn't have any null values in the baseline, ultimately, we can analyze the trends within the years. So, what should I recommend Elina do regarding her small dataset. The only way she can increase her participant size is if she incorporates the data with some missing values (as long as it's not in the baseline). However, I am not sure if this would cause issues with her analysis.

She attached a document containing some rough visualizations of the 22 participants. It seems like she has normality within the hippocampal volume, and Poisson distribution ferritin and WBC (fat).

However, to answer her original question should she have more samples or should she only look at participants from the 2 years.

Any clarification would be greatly appreciated.

Best,

Sumeeth

Get [Outlook for Android](#)