Answers to some common questions I get about the pipeline:

Why didn’t you functionalize more of this?

I have explored functionalizing parts of the pipeline. Other people have created packages intended for working with UKB data (Hanscombe et al., 2019). I found these challenging to use because of small differences in how my UKB data was formatted which broke some functions, along with the need to make adjustments to satisfy specific concerns of our research group. Each distinct analysis has required enough small adjustments that attempting to create functions which could adapt to the differences proved challenging. So I’m leaving this in the form that worked best for myself and fellow grad students and hope it is something you can iterate off of as well.

Why are things breaking when I upload my data?

As discussed above, check that your data is formatted in the way the pipeline assumes it will be. I’ve tried to provide examples of what the format should be, but due to the protected nature I can’t share any examples of real UKB data.

What is the meaning of this random comment buried somewhere in the middle of the script?

I love writing comments! I have gone over these scripts and tried to ensure the comments are relevant and up to date, but some are really just little notes to myself and may have become out of date as I developed the pipeline over several years. Please feel free to reach out to [vthornton@wustl.edu](mailto:vthornton@wustl.edu) to ask for clarification and let me know if something is no longer useful and needs to be fixed.

Why is the processing of alcohol and smoking variables so different?

Partly just historical, I wrote the scripts at different times. Some of the challenges converting the raw UKB data into something useful are different for these variables also. In smoking, a lot of participants with limited smoking history never got asked some questions so you have to distinguish not asked but can be inferred from true missing. For alcohol, there is a lot of summing across drink types and handling weekly vs. monthly.