Statement of the problem:

1) Can we develop cut points for amyloid positive based on plasma if we condition on CSF levels based on a population of patients who have both CSF and plasma data.
   1a) Fit a bivariate model and then use cut points to develop amyloid label in study people in your study. The statistical goal would be to study this approach in small to large sample sizes. The goal would be to understand how big of a study with CSF and Plasma do you need to estimate good cut points. Could we do this approach on smaller sample sizes with informative priors based on biology or other studies.
   1b) Integrating Bayesian model.
2) If we want to classify individuals into amyloid positive in a future study, how many patients do we need CSF data on to but a good cut point model using plasma alone.
3) If we have a set of cut points from historical studies can we use those to inform the priors when we develop a new study specific cut point? [Maybe needs some refining]

Details for #1.

Simulate mixture data like your one dataset for N=100, 250, 500. Set probability of amyloid positive group to 1/3 and use the means and variances from the real data. Simulate 1000 datasets for each sample size. Set the seed for each simulation so that we can rerun any simulation.

For each dataset fit the joint model and make graphs and tables of results.

Research papers on how to develop cut points from a model where we want three categories. Yes, no, indeterminate. Alex to send 1-2 papers for you to start.