Hello ALA Mode!

I think this is a very interesting topic, and I look forward to seeing how your study develops over the semester. Although I don't have any clinical experience in cervical cancer screening, I am broadly interested in women's health and disparities and I think your findings would be beneficial in improving health care delivery and ensuring equitable access to critical screening services.

Praise
First, fabulous choice in team name! It looks like you've been able to identify a rich dataset, with various lifestyle and demographic factors that could potentially impact quality of health. It's also great that the MEPS survey data is available at an individual level, and is not aggregated to population level statistics. This should give you a lot of power in your research endeavors, and the opportunity to perform a nuanced analysis.

You've also done a great job in narrowing the focus of your research question to the specific screening class of cervical cancer, which seems like a reasonable scope for the final project.

Advice
While it's great to have a lot of data available to work with, and as it sounds like you have already begun to consider, working with such a huge volume of covariates can be challenging. In addition to leveraging some of the automated covariate selection methods we discussed in class (e.g. forward/backward selection, elastic net, etc.), it could be helpful to start with a literature review to better understand what factors drive seeking preventive medical care in order to narrow down your potential pool of covariates based on subject matter knowledge.

One factor that I have seen health care activists routinely cite as critical in accessing preventative care, particularly in minority communities, is the language in which health care service is delivered, as well as the language in which health information is disseminated in the local community. I noticed that there is a language field included in the MEPS survey, and there is some geographical information available as well. It might be worth exploring whether it's possible to identify if health care is available in an accessible language in different communities. The Medicaid database on data.gov also has publicly available information on cervical cancer screening in various geographies, as well as some information of services offered in other languages (although this data is aggregated at county and state population levels, and not at the individual level).

Additionally, it could be interesting to see what the screening rate is amongst those who are at higher risk for cervical cancer (e.g. those who have a more children, smoke) compared to those who are at lower risks.

Concern
Handling the skip pattern methodology of questioning in the survey will be difficult, and it might be worth exploring various methods to fill in the missing data through interpolation

techniques. I am concerned that you might lose too much survey data, and be limited in how you can stratify or categorize covariates otherwise.

Additionally, there are some potential confounding variables (e.g. family history of cancer) that are not included in the survey (at least not that I saw at first glance), which could bias your results.

Hope you find this feedback helpful, and I'd be glad to chat about any specific questions you might have as well. Best of luck with your research! Looking forward to seeing the results.