Hi Anja, Anna, and Ligia,

I hope you are all well! I really enjoyed reviewing your post and can't wait to see where your project goes. Here were my thoughts as I reviewed:

I LOVE this project idea! This is such a cool topic and the prediction tool you come up with is going to be so interesting. I also think it's super important that you called out the need to check for demographic variables as potential confounders/effect measure modifiers right away, because I think that's definitely something to look out for when it comes to access to healthcare. Nice job!! Go ALA Mode (I also love your group name)!!!

My questions from your post:

- What individual-level factors are included in your dataset and how are they measured? How are you measuring accessibility to healthcare?
- Somewhat confused about what you mean by "recode variables," is that just referring to possibly making continuous variables categorical, making sure categorical variables are being treated as factors, etc.?
- What are you including in "expenditure variables"?
- How will you be selecting variables? LASSO?
- Will you be considering GAMs if they make sense for the data?

My possible suggestion:

- Because you have so many observations, you could consider training on 75%/80% (or even less) and testing on more so that you can have a more accurate understanding of how well your model predicts! Also, how are you planning to choose what observations will be trained/tested on?

No concerns! Like I said, I love this! I think you guys are going to do great!

Please let me know if you are confused by anything I have said here and want clarification. Good luck and enjoy your project! ☺

Kirstie Turnbull