1.

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GIPGXH
https://academic-oup-com.ezaccess.libraries.psu.edu/poq/article/87/2/415/7147284

Oceno, Marzia and Wei-Ting Yen. 2022. "The Impact of Racial Descriptive Norms on Vaccination against COVID-19." Public Opinion Quarterly.

In this article, the authors question if racial descriptive norms can be used as a tool to increase confidence and reduce hesitancy within the US public. Using an experiment, the authors varied whether information about uptake intent by race was provided, and what racial group was reported to be more likely to get a vaccine if one were available to them today. Their results show that the tendency to conform to one's racial ingroup can play a key role in improving vaccination attitudes across races.

2.

The aim of their module is explanatory, as they are trying to explain which factors most predict someone getting a vaccine.

3.

In their model, the authors use rural residence as a control. I argue that for predictive modeling, this variable should be absent from the regression. The effect of rural residence is likely small, and already partly explained by other factors such as income, Christianity, race, and education. By removing rural residence, I hypothesize that the predictive power of the model improves, as rural residence likely has low significance.

I will be modeling this model:

vaccinationIntent = Christianity + ruralID + sex + race + education

4. Following a predictive test and assessing with MSE, I found that the full model had an MSE with .1423. The shortened model, without the rural identifier, had a MSE of .1415. Therefore, removing the rural identifier is beneficial for the predictive power of the model.

5. My replication provides evidence to suggest that the predictive power of models is better when covariates with small significance/no significance are removed from the model.