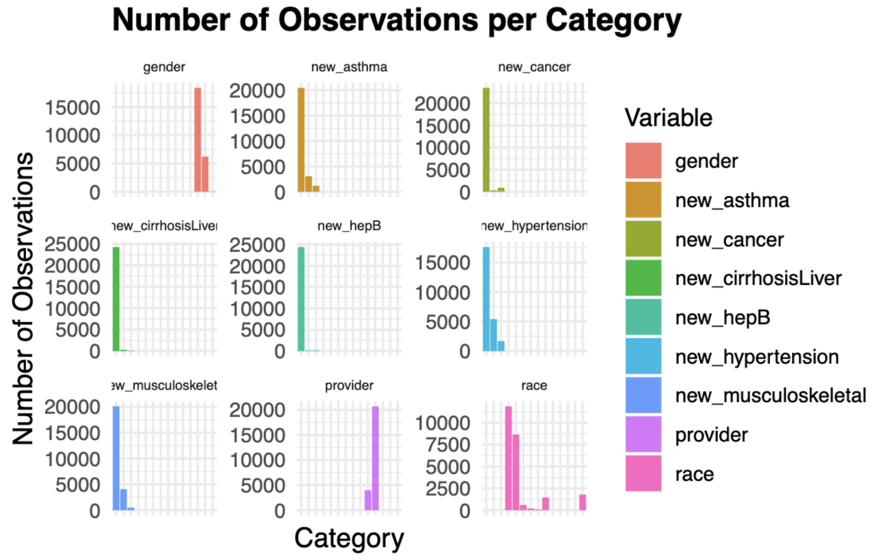
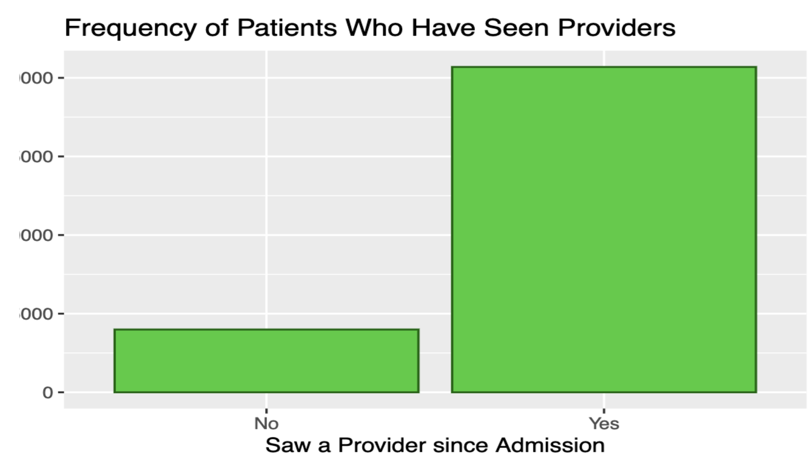
PREDICTING PROVIDER VISITS SINCE PRISON ADMISSION: THE ROLE OF MEDICAL HISTORY

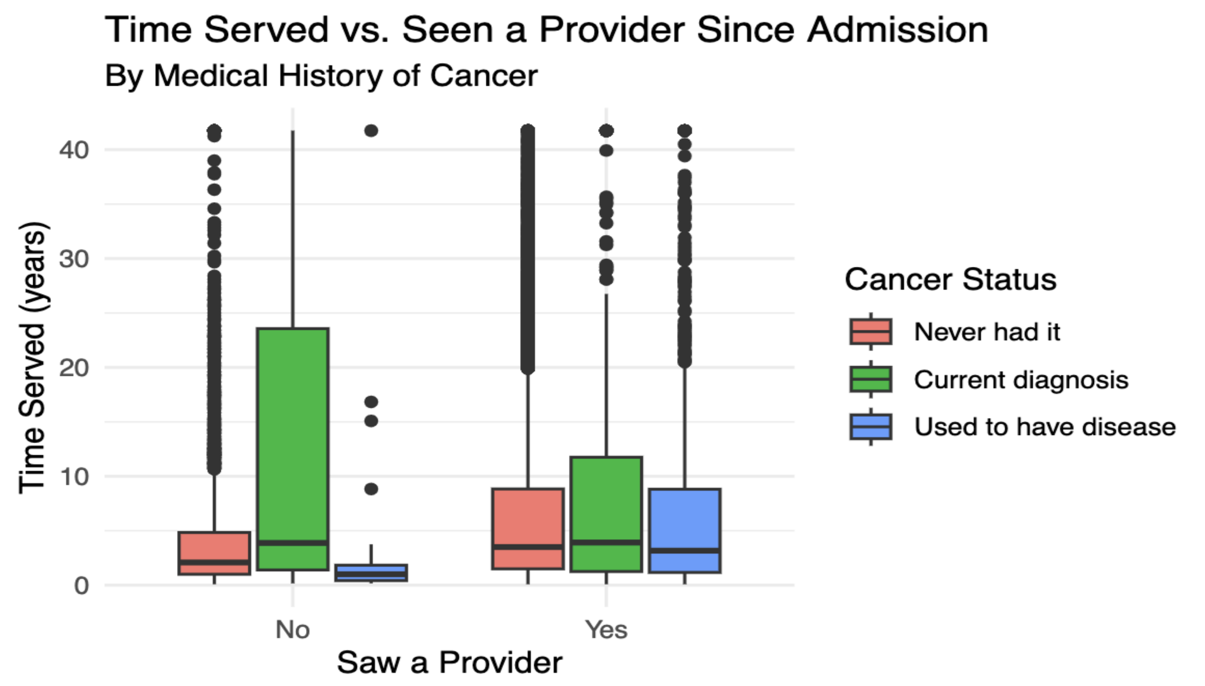
Sonya Eason

How much insight can we actually get into the likelihood of inmates accessing providers from medical history? Is information about the less common medical conditions also informative?

INTRODUCTION



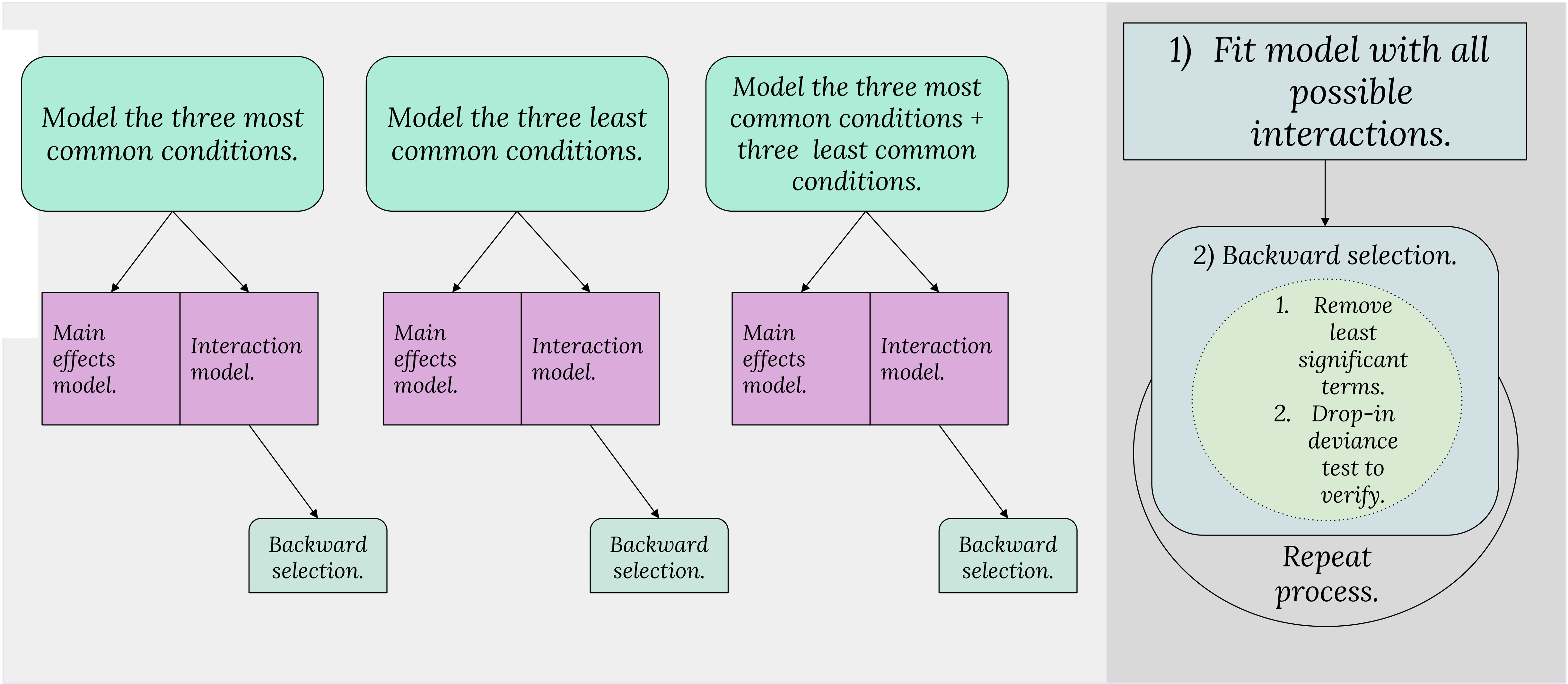




METHODS

1. Split data into testing and training sets.
2. Fit models to training set.
   a. Three categories of models were built: those that incorporated the three most common medical conditions, those that incorporated the three least common medical history, and those that incorporated both.
3. Check model conditions.
4. Perform cross validation, collect AIC metric on assessment sets.
5. Evaluate testing set performance with ROC_AUC.

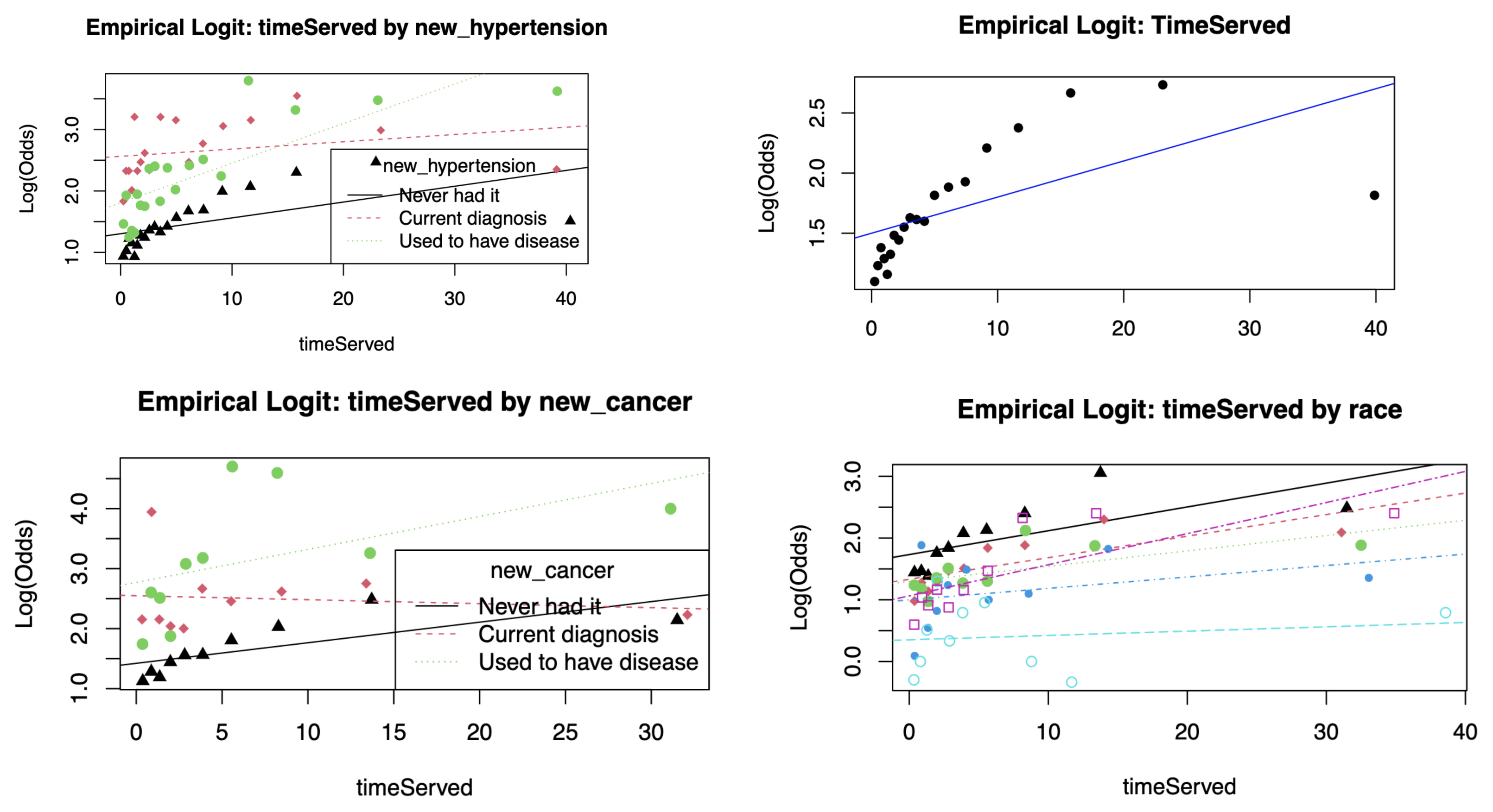
FIGURE 1: Model Building



The figure on the left displays how models were built. The figure on right elaborates on how the interaction models were built.

FIGURE 2: Model Conditions

- *Linearity was checked for quantitative predictors and interactions and was satisfied.*
- *Randomness was also satisfied due to methodology of data collection.*
- *There was no evidence that independence was violated. The location or the specific prison could have influenced provider access as well, but that information was suppressed in public use data file.*



RESULTS

model	AIC	ROC_AUC
Common	12420	0.685
Common with Sig Intx	12376	0.684
Less Common	12888	0.635
Less Common with Sig Intx	12867	0.636
All Conditions	12409	0.690
All Conditions with Sig Intx	12360	0.690

- *AIC values were collected based on the assessments sets of the training set from cross validation.*
- *ROC_AUC was collected based on testing set.*
- *The ideal model minimizes AIC and maximizes ROC_AUC.*
  - *Based on this criteria, "All Conditions with Sig Intx" performed the best out of the six models.*

CONCLUSIONS

$$\log\left(\frac{P(\text{provider}=1)}{1-P(\text{provider}=1)}\right) = \beta_0 + \beta_1(\text{race}_1) + \beta_2(\text{race}_2) + \beta_3(\text{race}_3)$$
$$+\beta_4(\text{race}_4) + \beta_5(\text{race}_5) + \beta_6(\text{gender}) + \beta_7(\text{timeServed})$$
$$+\beta_8(\text{new\_musculoskeletal}_1) + \beta_9(\text{new\_musculoskeletal}_2) + \beta_{10}(\text{new\_hypertension}_1)$$
$$+\beta_{11}(\text{new\_hypertension}_2) + \beta_{12}(\text{new\_asthma}_1) + \beta_{13}(\text{new\_asthma}_2)$$
$$+\beta_{14}(\text{new\_cancer}_1) + \beta_{15}(\text{new\_cancer}_2) + \beta_{16}(\text{new\_hepB}_1)$$
$$+\beta_{17}(\text{new\_hepB}_2) + \beta_{18}(\text{new\_cirrhosisLiver}_1) + \beta_{19}(\text{new\_cirrhosisLiver}_2)$$
$$+\beta_{20}(\text{new\_hypertension} \times \text{timeServed}) + \beta_{21}(\text{new\_hypertension} \times \text{new\_musculoskeletal})$$
$$+\beta_{22}(\text{new\_cancer} \times \text{timeServed}) + \beta_{23}(\text{new\_cancer} \times \text{gender})$$

A model with more medical history data with interactions performed the best in predicting provider visits. With a goal of parsimony, a model with more medical history predictors still proved effective.