

PIPELINE STEPS (PT 2)

3. For the training data split within each of the 11 cohorts, fit the AI/ML model to predict the chosen endpoint
4. a. Predict the endpoint using the *internal test data* for the 11 cohorts. Note that the proportion of the minority class will be nearly identical for the training and testing data in each of the 11 cohorts
- b. Predict the endpoint using the single *external validation data*. The proportion of the minority class in this external data will be fixed to the observed proportion in the real world