Using Machine Learning Forecasts to Improve Evaluation

Geoffrey C. Barnes, Ph.D.

Western Australia Police

University of Cambridge

Global Evidence and Implementation Summit

22nd October 2018

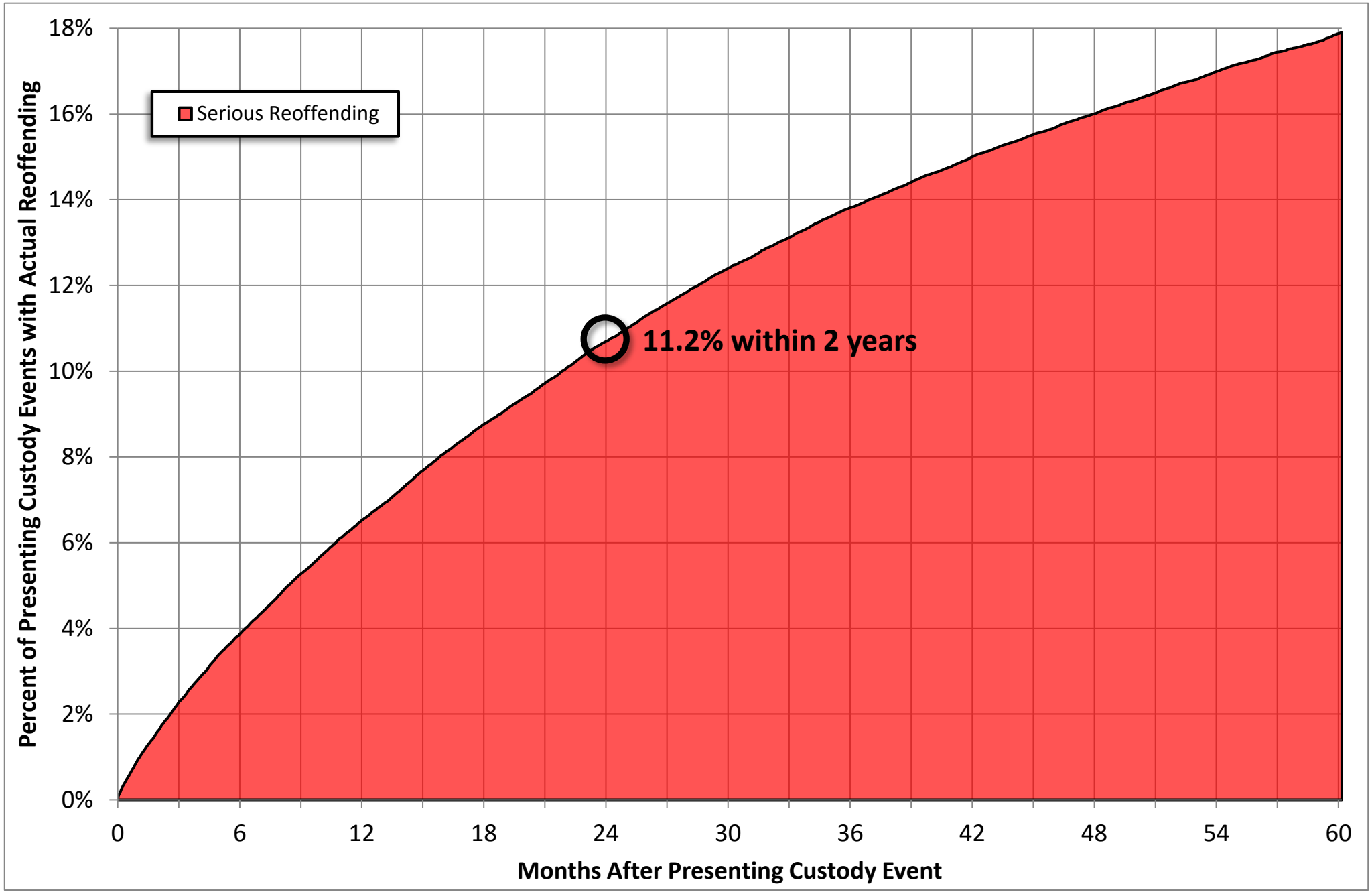
Durham Checkpoint Evaluation

What Outcomes Are Predicted by HART (Harm Assessment Risk Tool)?

Three different outcome categories

- Combines the prevalence of offending (i.e., yes/no) with the seriousness
- High Risk: A new serious offence occurring within 2 years
  - Murder
  - Attempted murder
  - Aggravated violence (i.e., GBH)
  - Robbery
  - Any sexual offence
  - Any firearm offence

What Outcomes Are Predicted by HART?

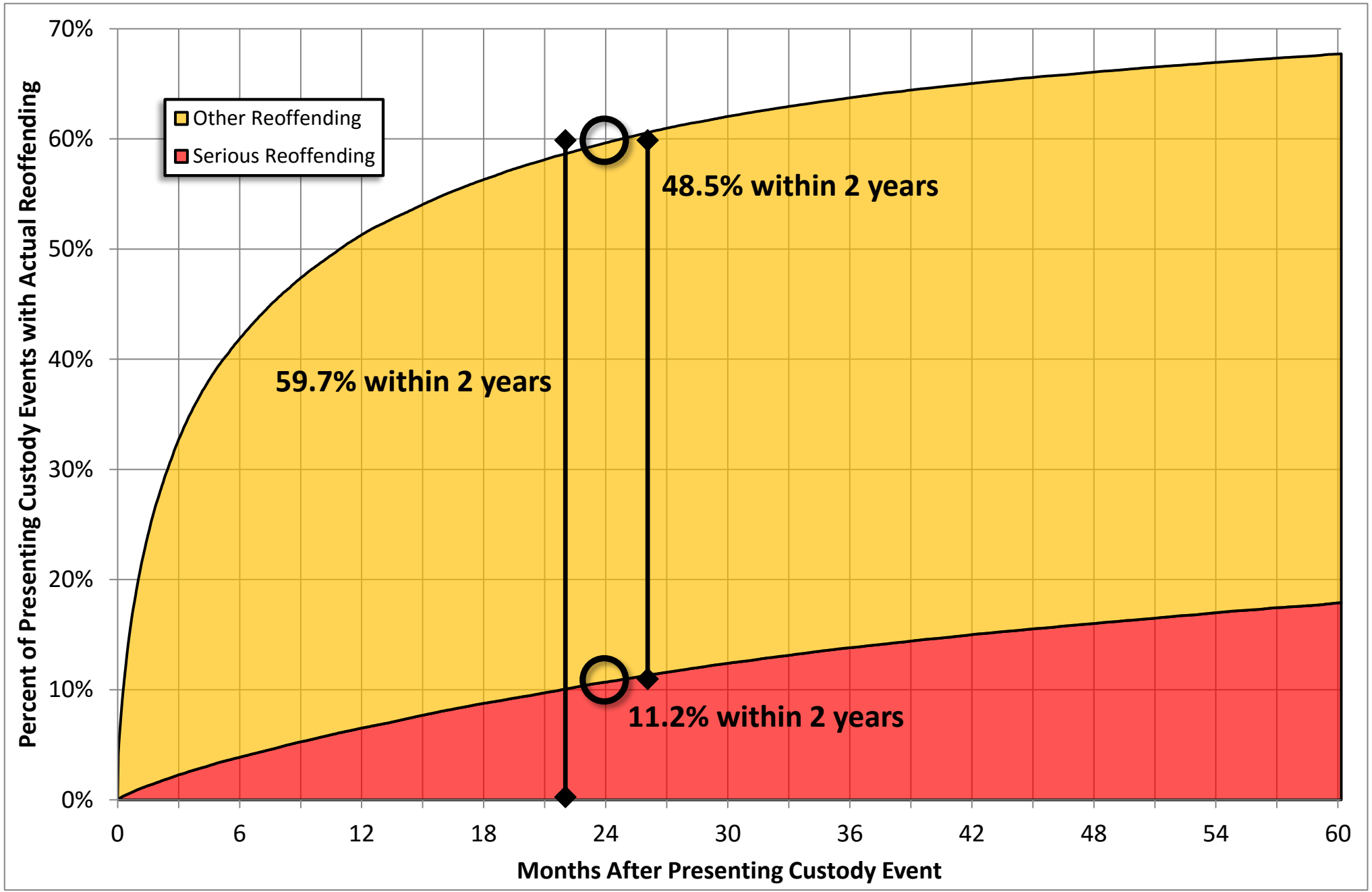


What Outcomes Are Predicted by HART?

Three different outcome categories

- Combines the prevalence of offending (i.e., yes/no) with the seriousness
- High Risk: A new serious offence occurring within 2 years
  - Murder
  - Attempted murder
  - Aggravated violence (i.e., GBH)
  - Robbery
  - Any sexual offence
  - Any firearm offence
- Moderate Risk: Any new non-serious offence within 2 years

What Outcomes Are Predicted by HART?

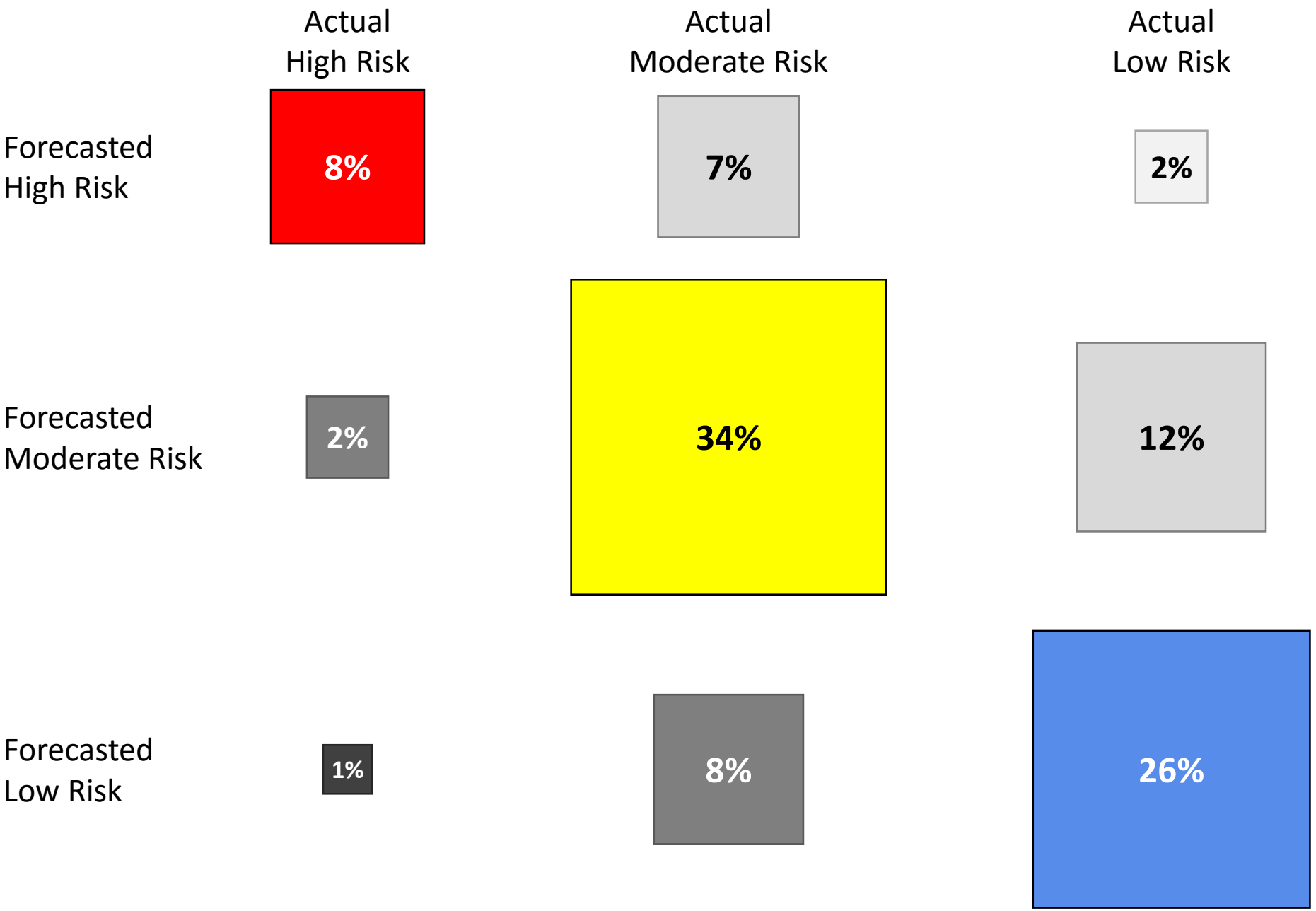


What Outcomes Are Predicted by HART?

Three different outcome categories

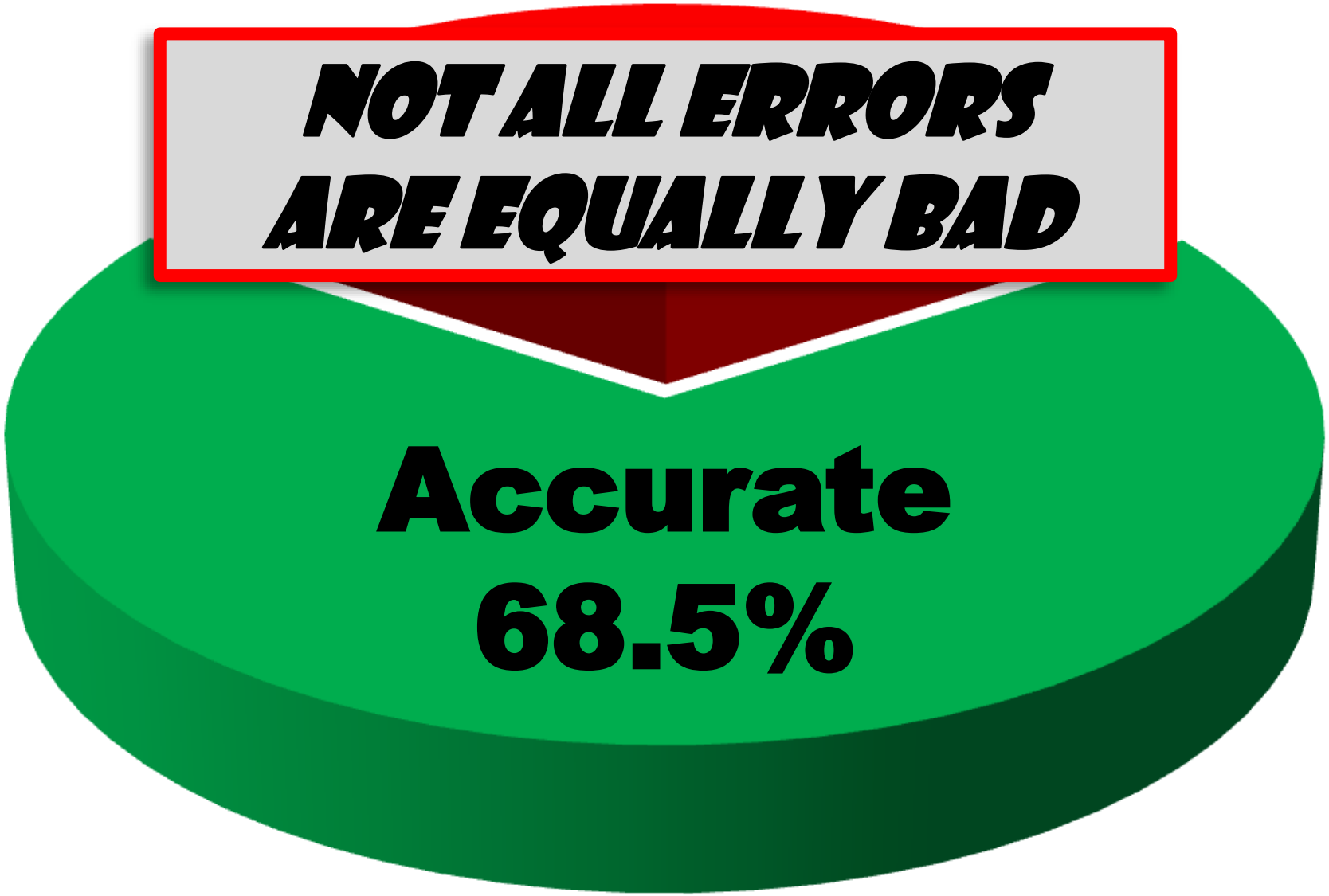
- Combines the prevalence of offending (i.e., yes/no) with the seriousness
- High Risk: A new serious offence occurring within 2 years
  - Murder
  - Attempted murder
  - Aggravated violence (i.e., GBH)
  - Robbery
  - Any sexual offence
  - Any firearm offence
- Moderate Risk: Any new non-serious offence within 2 years
- Low Risk: No new offences of any kind within 2 years

Estimated Accuracy and Error - Durham Custody Model:

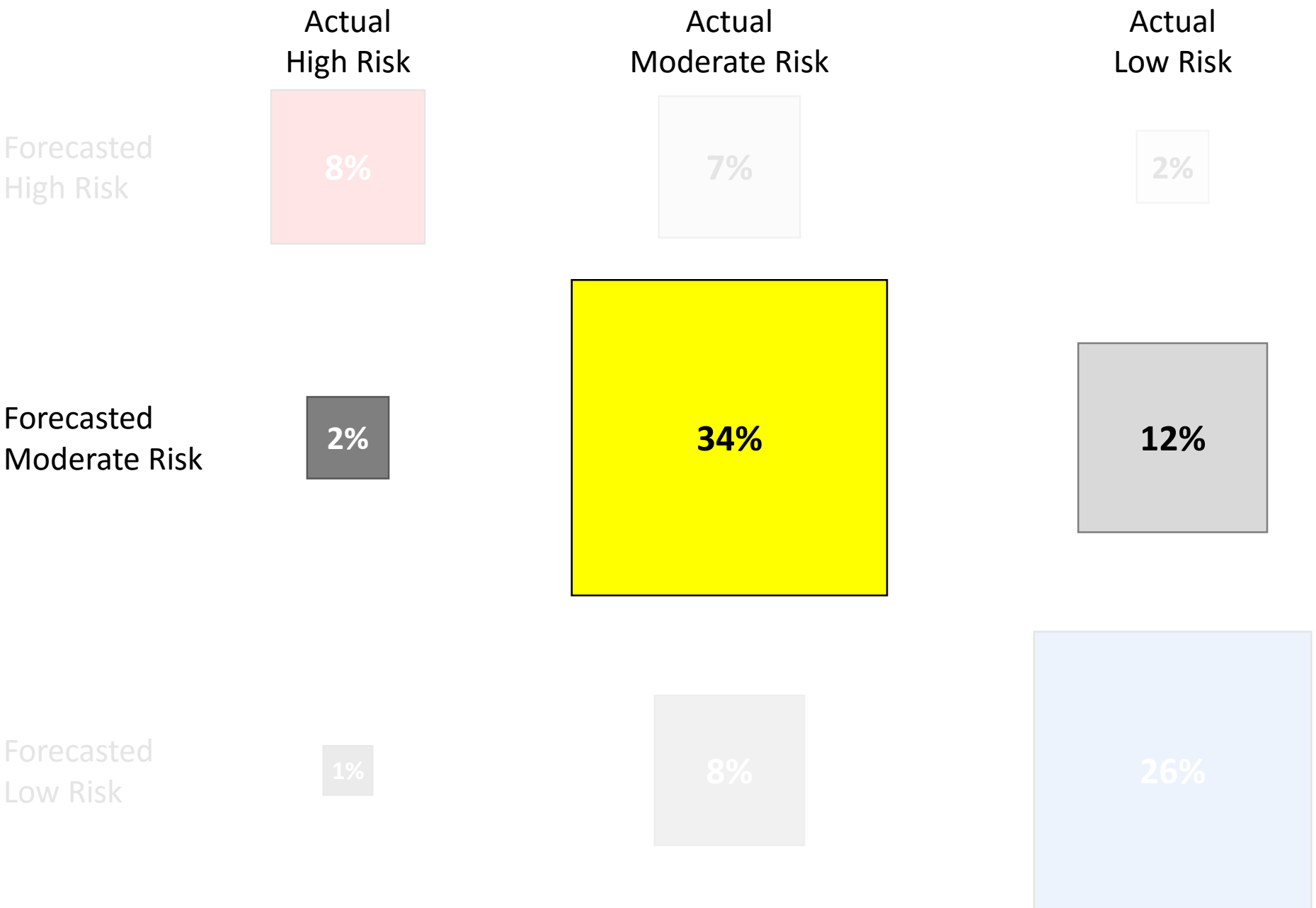


Estimated Accuracy and Error - Durham Custody Model:

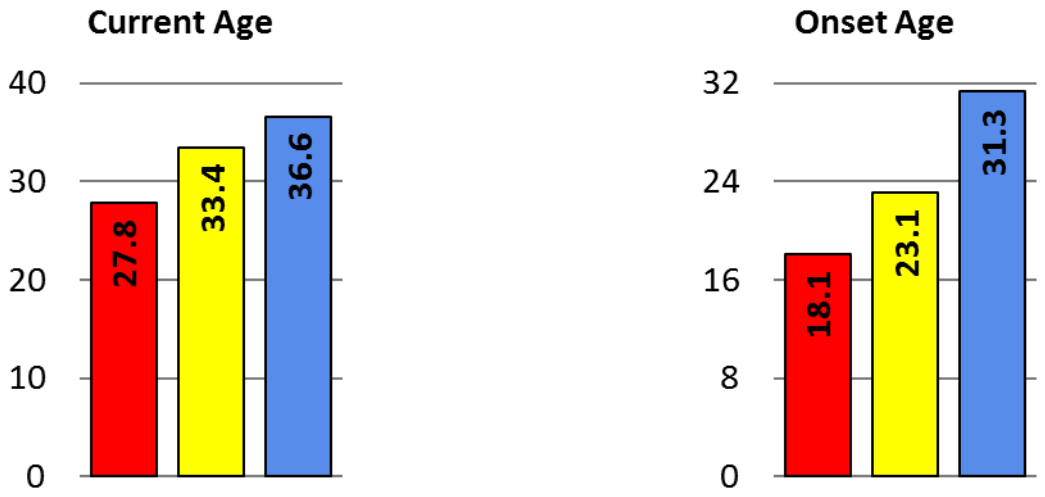
Estimated Accuracy and Error - Durham Custody Model:



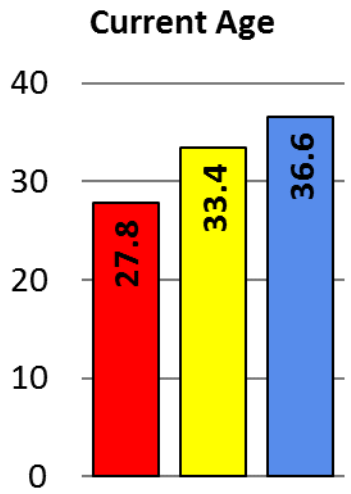
Estimated Accuracy and Error - Durham Custody Model:

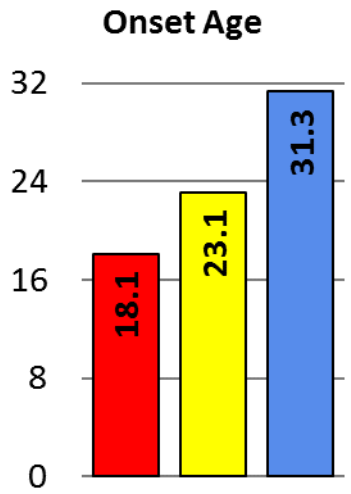


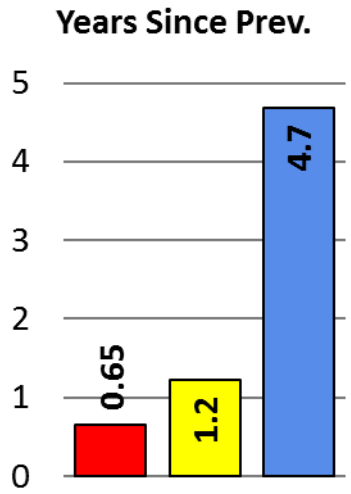
Mean Predictor Values across Forecasted Risk:

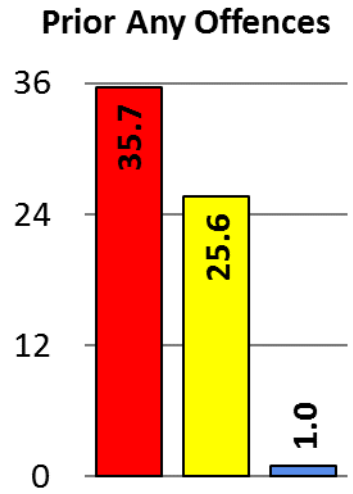


Mean Predictor Values across Forecasted Risk:

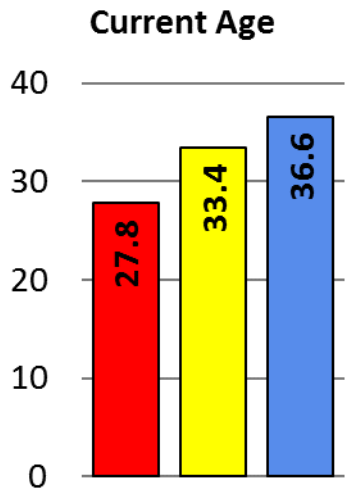


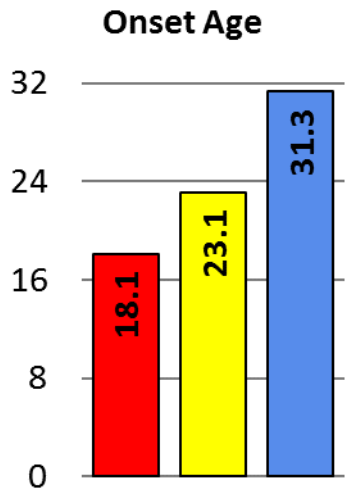


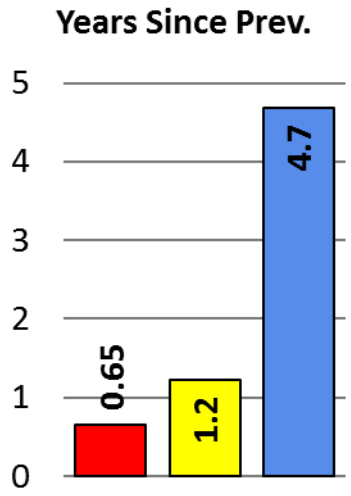


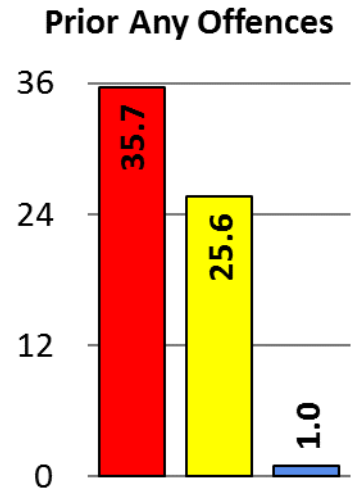


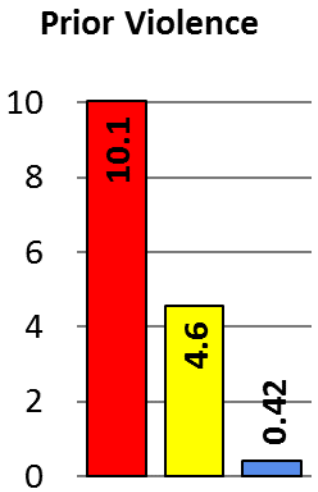
Mean Predictor Values across Forecasted Risk:



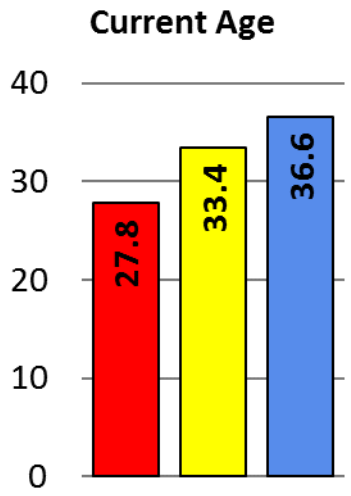


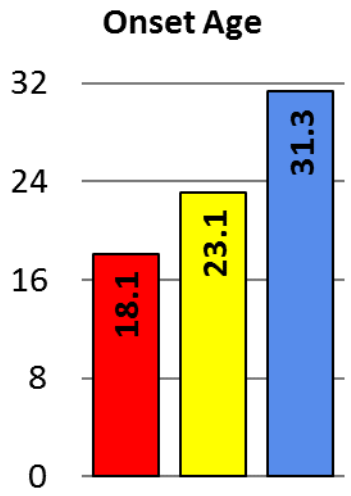


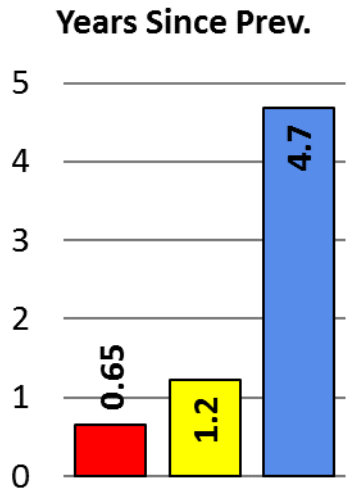


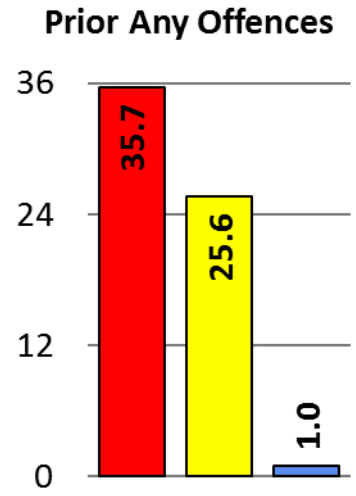


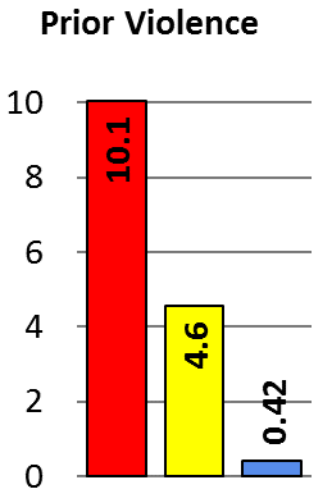
Mean Predictor Values across Forecasted Risk:

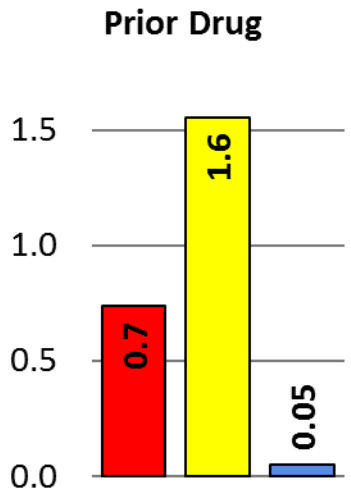




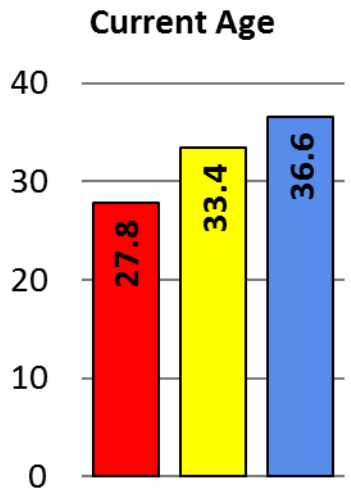


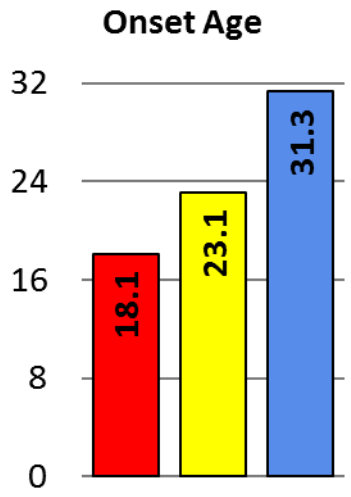


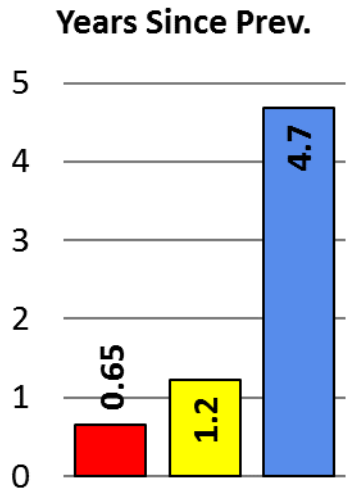


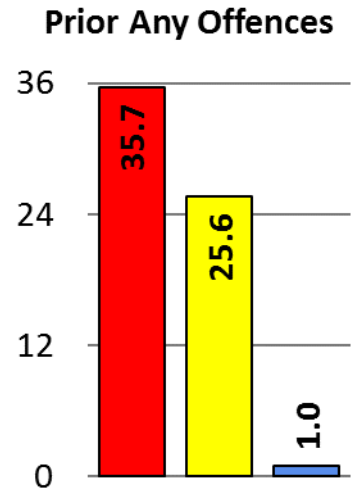


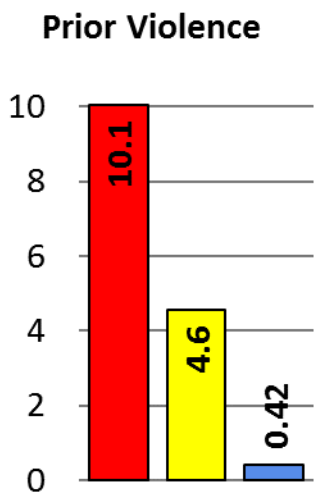
Mean Predictor Values across Forecasted Risk:

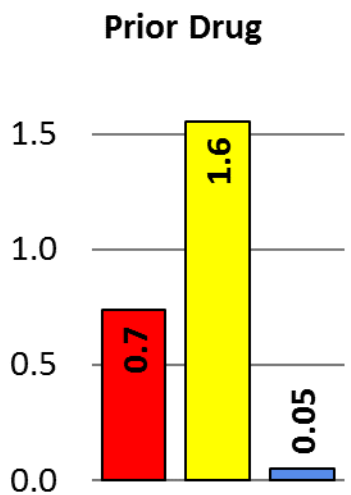


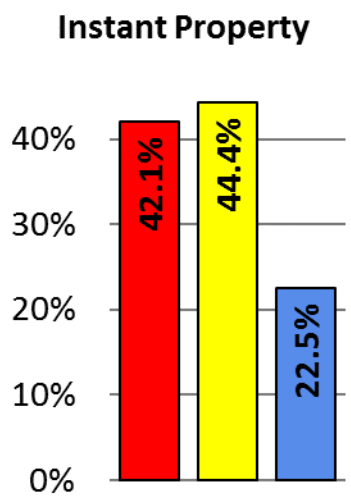


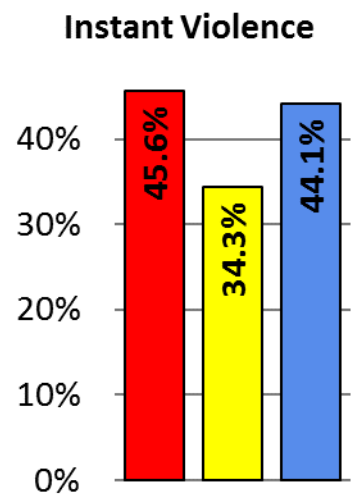






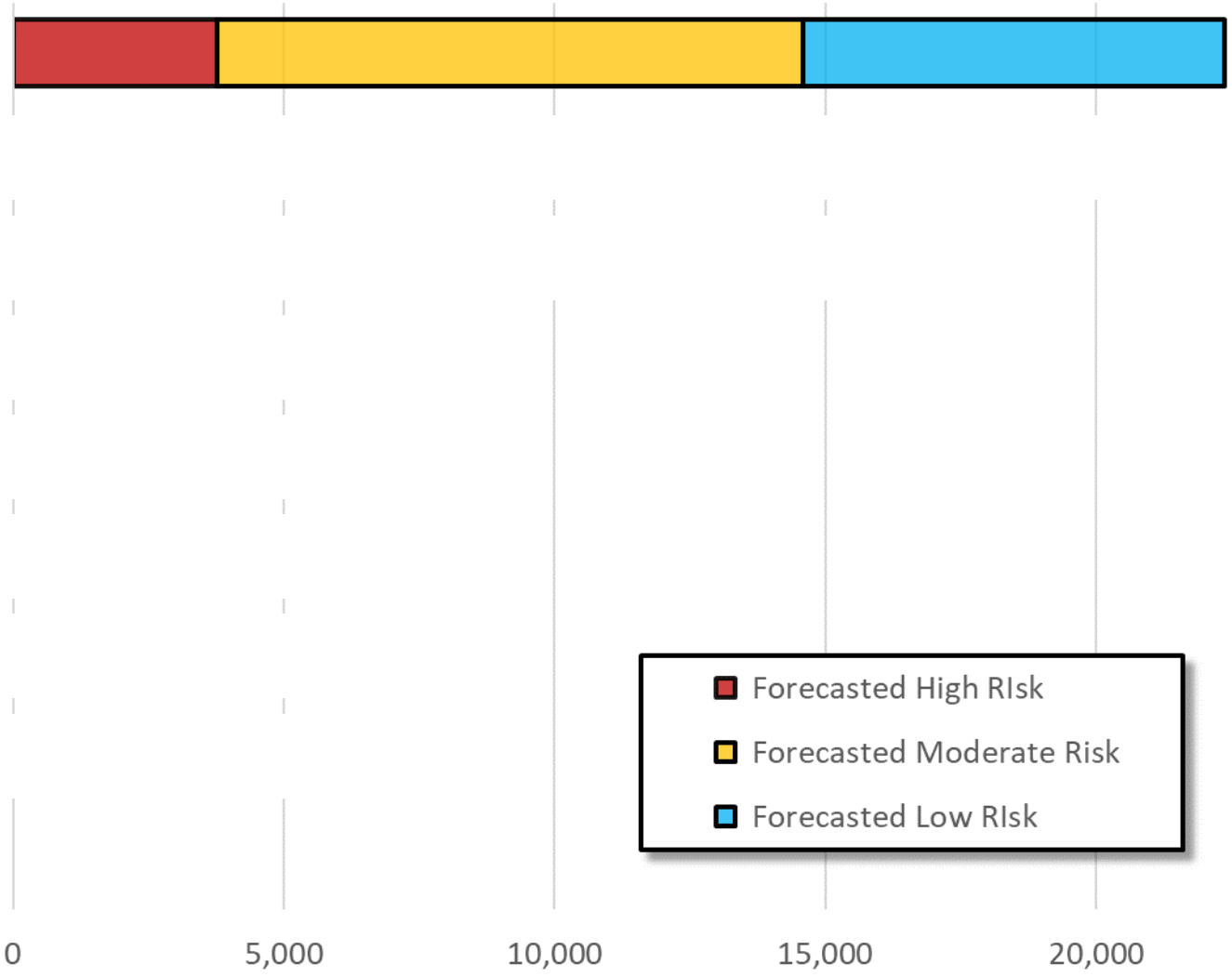






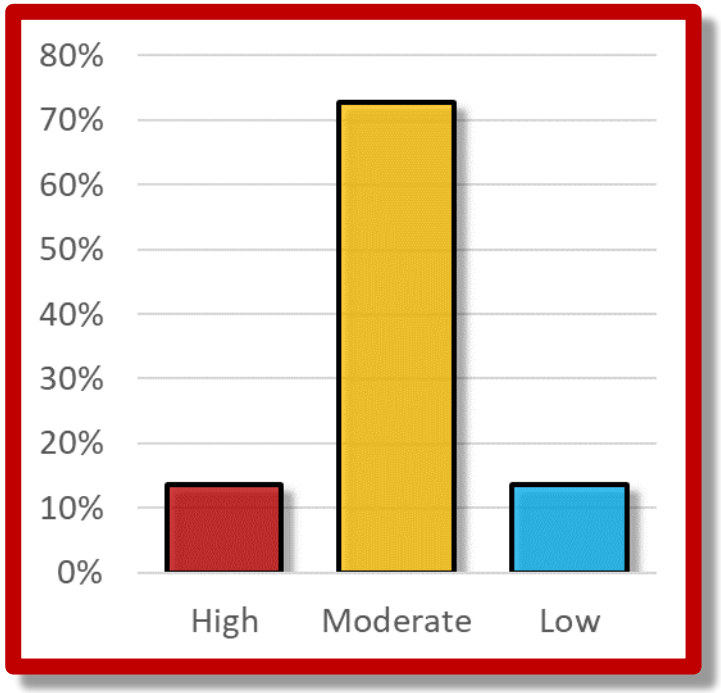
Typical Enrollment Attrition in Criminal Justice Experiments:

Annual supply of arrests .....

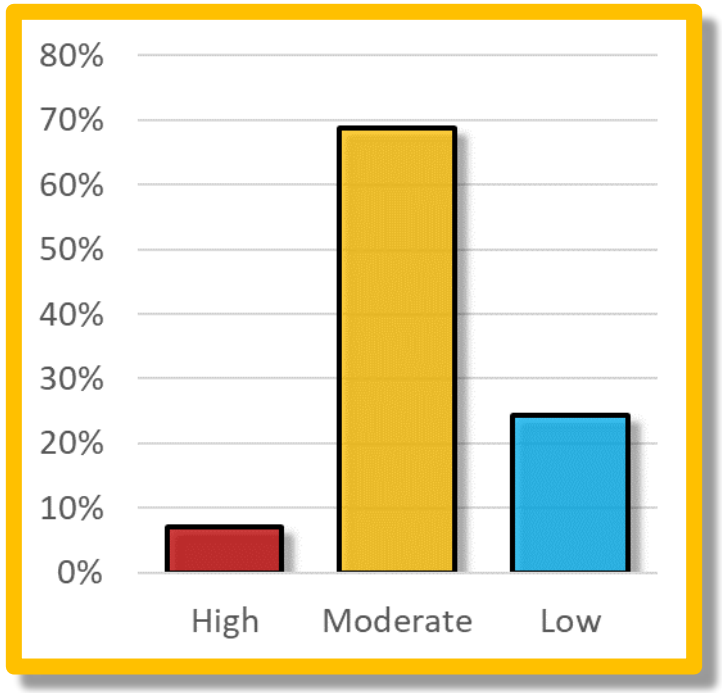


What Do Durham Custody Sergeants Say In Their Own Assessments of Risk?

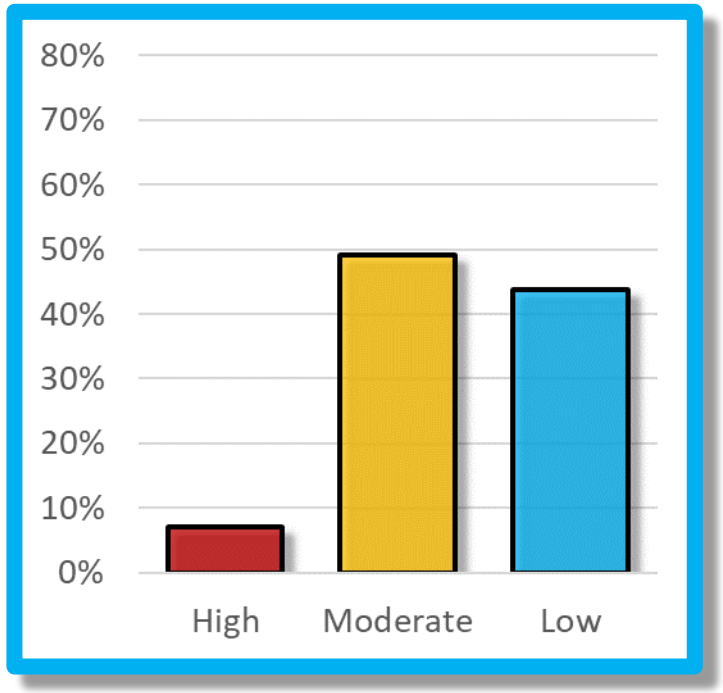
When the model's forecast is High Risk…



When the model's forecast is Moderate Risk…



When the model's forecast is Low Risk…

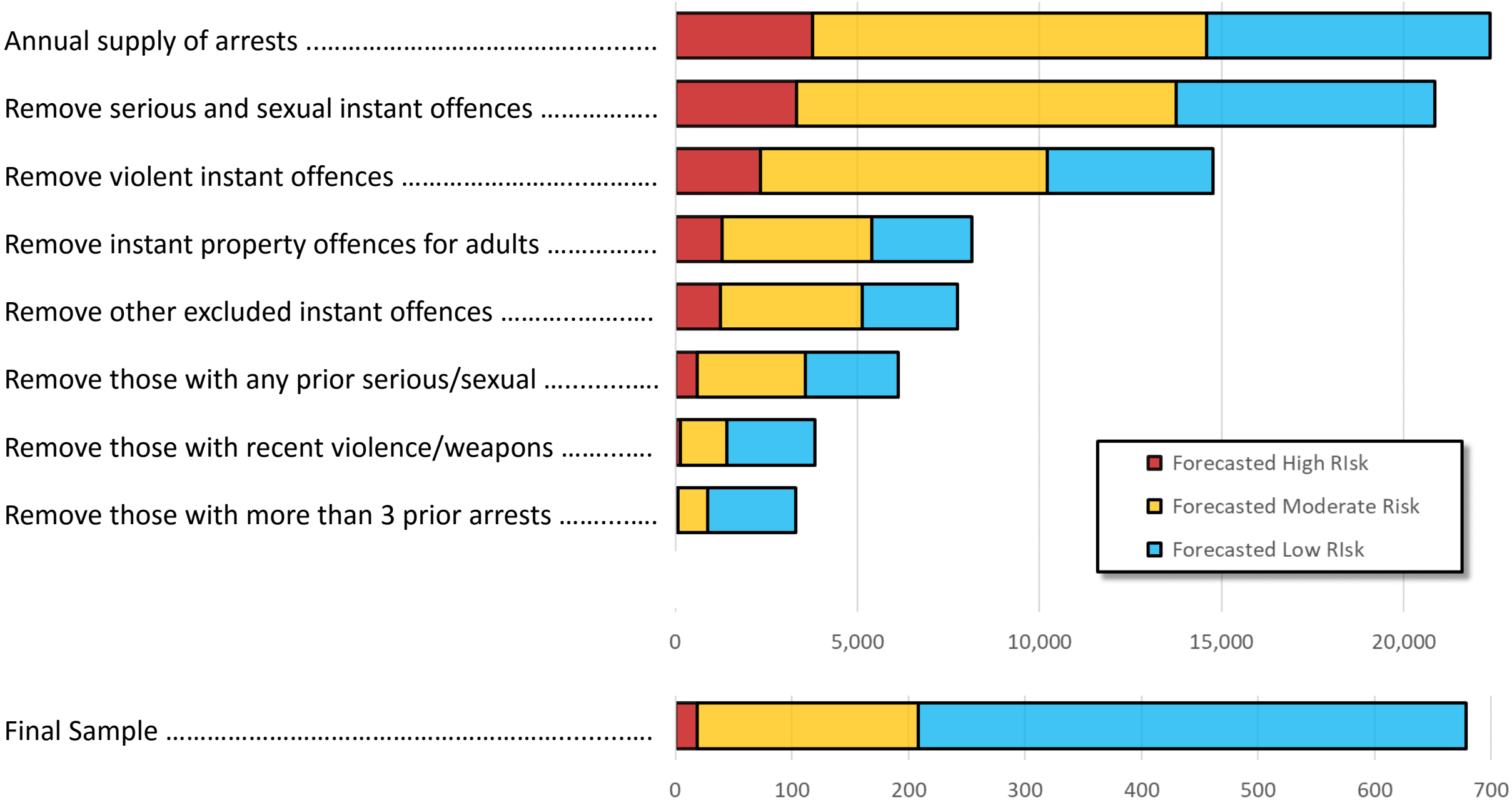


When the Sergeant thinks somebody is High Risk… they are the least likely to make a refer for random assignment (5%)

When the Sergeant thinks somebody is Moderate Risk… they will refer for random assignment when they remember (30%)

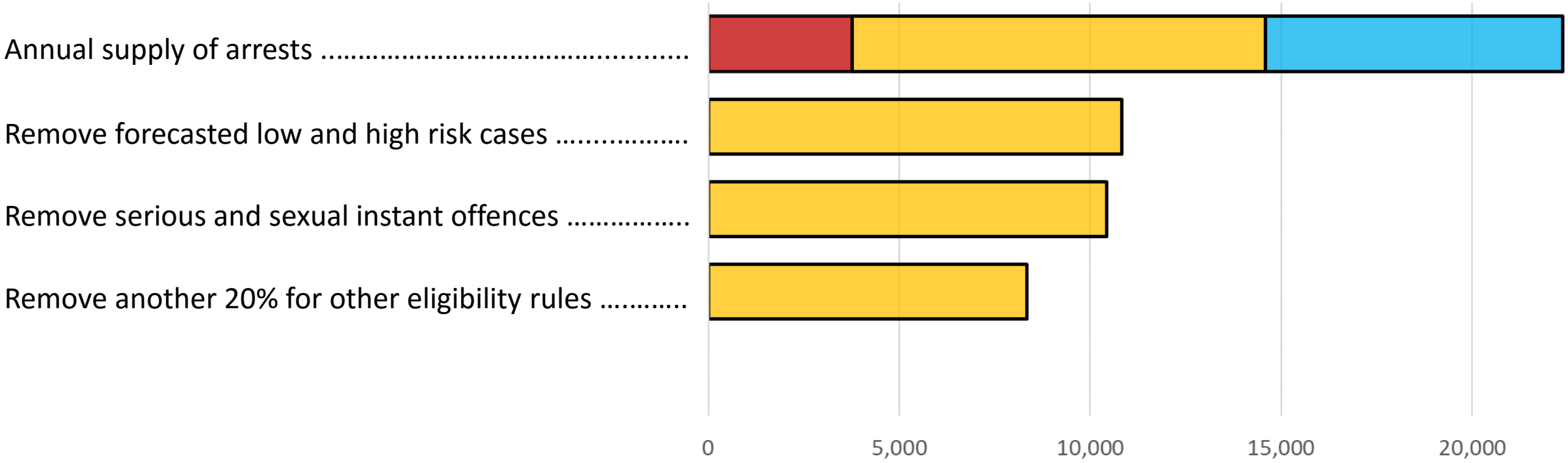
When the Sergeant thinks somebody is Low Risk… they will resist referral to avoid the risk of control assignment (10%)

Typical Enrollment Attrition in Criminal Justice Experiments:

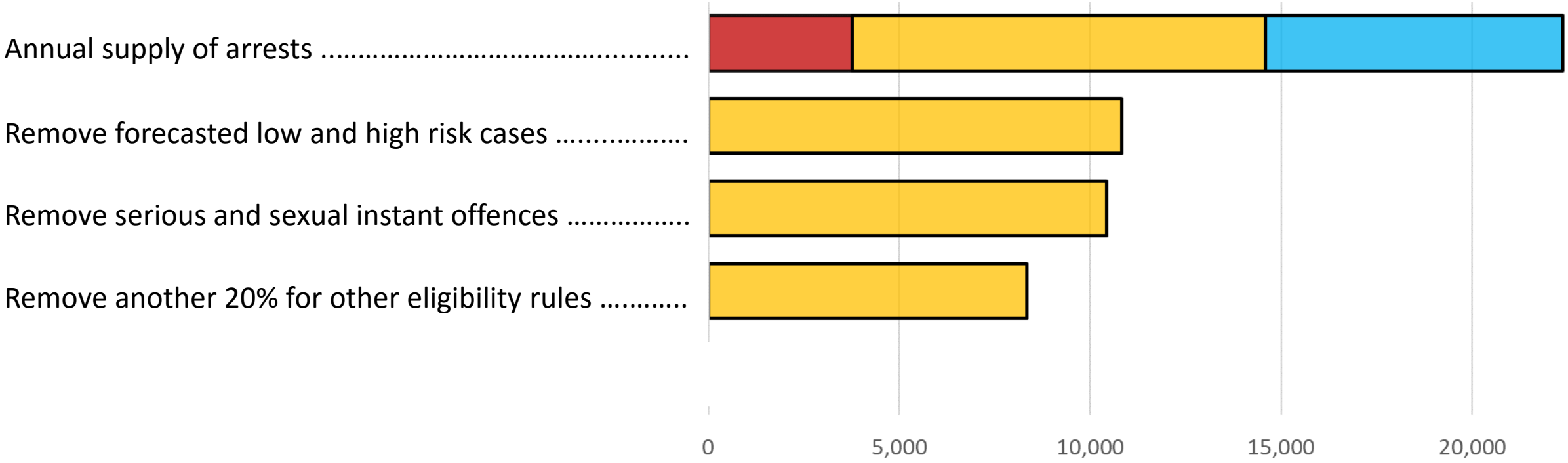



So what happens when a random forest is used to govern eligibility?

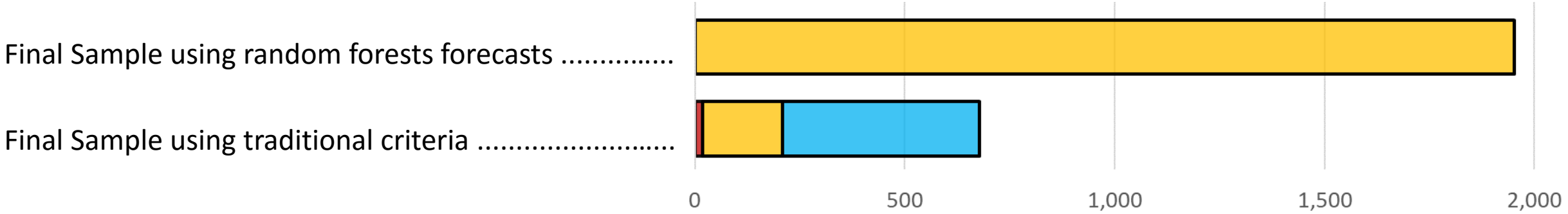
Better Experimental Enrollment Using Random Forest Forecasts:



Better Experimental Enrollment Using Random Forest Forecasts:

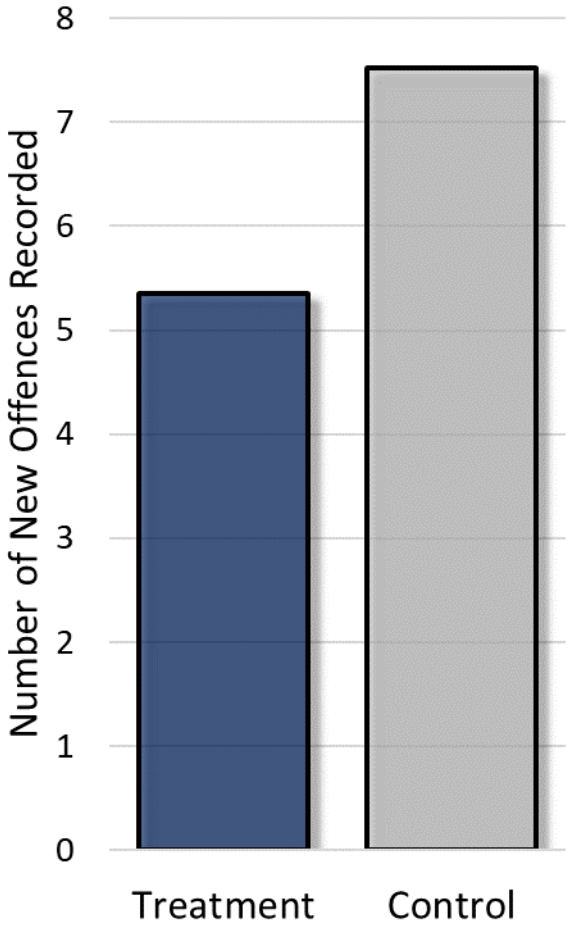


Comparison of the two samples:

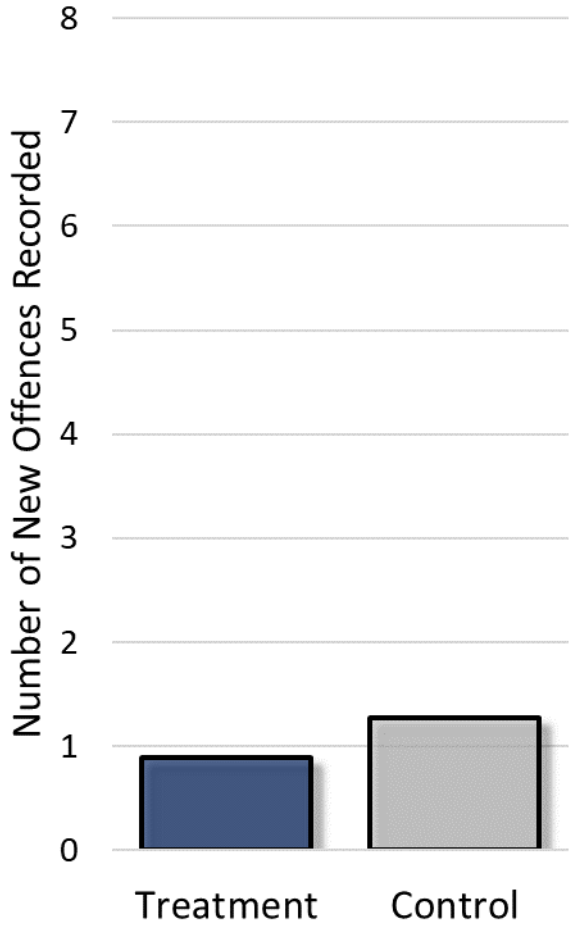


Monte Carlo Simulation of Different Eligibility Rules:

When the random forest model is used to determine eligibility…



When traditional offence-based criteria are used to determine eligibility…



<u>Difference Between Treatment and Control:</u>

Random Forest:	-2.17
Traditional:	-0.38

<u>Percentage Decrease in Offending:</u>

Random Forest:	28.9%
Traditional:	29.9%

<u>Standarised Difference in Means (*t*):</u>

Random Forest:	7.94
Traditional:	1.84

<u>Average Significance Result (*p*):</u>

Random Forest:	1.35×10^{-9}
Traditional:	0.161

<u>Statistical Power (Probability of $p < .05$):</u>

Random Forest:	100%
Traditional:	30%


Two Things that Cops Hate...