



**NIJ** National Institute  
of Justice

STRENGTHEN SCIENCE. ADVANCE JUSTICE.

[Home](#) / [Funding & Awards](#)

# Recidivism Forecasting Challenge

[Add Your Submission](#)

The National Institute of Justice's (NIJ) "Recidivism Forecasting Challenge" (the Challenge) aims to increase public safety and improve the fair administration of justice across the United States. In accordance with priorities set by the DOJ, NIJ supports the research, development, and evaluation of strategies to reduce violent crime, and to protect police and other public safety personnel by reducing recidivism. Results from the Challenge will provide critical information to community corrections departments that may help facilitate more successful reintegration into society for previously incarcerated persons and persons on parole.

As the research, development, and evaluation agency of the U.S. Department of Justice, NIJ invests in scientific research across diverse disciplines to serve the needs of the criminal justice community. NIJ seeks to use and distribute rigorous evidence to inform practice and policy, often relying on data analytic methods to do so. The Challenge aims to improve the ability to forecast recidivism using person- and place-based variables with the goal of improving outcomes for those serving a community supervision sentence. In addition to the

## Webinar [↗](#)

NIJ hosted a webinar discussing the Challenge. [Read the transcript and review the slides.](#)

## Access the Data [↗](#)

[Accept terms](#) to access data provided by State of Georgia.

## Submit Your Entry [↗](#)

When your entry is ready, [use the form](#) to submit.

[Download a sample submission.](#)

Challenge data provided, NIJ encourages contestants to consider a wide range of potential supplemental data sources that are available to community corrections agencies to enhance risk determinations, including the incorporation of dynamic place-based factors, along with the common static and dynamic risk factors.

The Challenge will have three categories of contestants: students; individuals/small teams/businesses; and large businesses. NIJ will evaluate all entries on how accurately they forecast the outcome of recidivism. Recidivism is defined in this Challenge as an arrest for a new crime. To receive prize money, (114 total prizes available, up to 15 per contestant/team) winning applicants must provide a comprehensive document detailing the lessons learned about what variables did and did not matter to their final forecasting model and, when applicable, what type of models outperformed other models. Contestants are encouraged to provide additional intellectual property regarding specific techniques, weighting, or other sensitive decisions.

The Challenge uses data from the State of Georgia about persons released from prison to parole supervision for the period January 1, 2013 through December 31, 2015. Contestants will submit forecasts (percent likelihoods) of whether individuals in the dataset recidivated within one year, two years, or three years after release.

NIJ expects that new and more nuanced information will be gained from the Challenge and help address high recidivism among persons under community supervision. Findings could directly impact the types of factors considered when evaluating risk of recidivism and highlight the need to support people in specific areas related to reincarceration. Additionally, the Challenge could provide guidance on gender specific considerations and strategies to account for racial bias during risk assessment.

## Download and Print

Download printer-friendly versions of:

[- The Challenge \(pdf, 28 pages\)](#).

[- The Codebook](#)

## I. Challenge Overview

The National Institute of Justice's (NIJ) "Recidivism Forecasting Challenge" (the Challenge) aims to increase public safety and improve the fair administration of justice across the United States. In accordance with priorities set by the DOJ, NIJ supports the research, development, and evaluation of strategies to reduce violent crime, and to

protect police and other public safety personnel by reducing recidivism. Results from the Challenge will provide critical information to community corrections departments that may help facilitate more successful reintegration into society for people previously incarcerated and on parole.

The Challenge aims to improve the ability to forecast recidivism using person- and place-based variables with the goal of improving outcomes for those serving a community supervision sentence. We hope through the Challenge to encourage discussion on the topics of reentry, bias/fairness, measurement, and algorithm advancement. In addition to the Challenge data provided, NIJ encourages contestants to consider a wide range of potential supplemental data sources that are available to community corrections agencies to enhance risk determinations, including the incorporation of dynamic place-based factors along with the common static and dynamic risk factors. NIJ is interested in models that accurately identify risk for all individuals on community supervision. In order to do this, contestants will need to present risk models that recognize gender specific differences and do not exacerbate racial bias that may exist. This document will outline the procedures, rules, and prize categories for the Challenge and provide relevant background information, the rationale for the Challenge, and potential implications of results from the Challenge.

This Challenge is issued pursuant to 28 U.S.C. 530C.

## II. Background

As the research, development, and evaluation agency of the U.S. Department of Justice, NIJ invests in scientific research across diverse disciplines to serve the needs of the criminal justice community. NIJ seeks to use and distribute rigorous evidence to inform practice and policy, often relying on data analytic methods to do so. The field of data analytics is rapidly evolving in industries such as consumer behavior prediction, medical/anomaly detection, and informatics. Many new data analytic methods could be further applied to the criminal justice system, specifically recidivism forecasting algorithms, which could foster rapid innovation in this field. Since practitioners began to systematically assess risk, risk assessment tools have been constantly refined with new factors and types of data, including dynamic factors that capture current information regarding persons on probation and parole. The incorporation of new datasets could help drive the fourth generation of risk assessment tools.

To provide a comprehensive comparative analysis with an aim of minimizing racial bias, NIJ is releasing the “Recidivism Forecasting Challenge.” With this Challenge, NIJ aims to: 1) encourage “non-criminal justice” forecasting researchers to compete against more “traditional” criminal justice forecasting researchers, building upon the current knowledge base while infusing innovative, new perspectives; and 2) compare available forecasting methods in an effort to improve person-based and place-based recidivism forecasting. Accordingly, the Challenge will have three categories of contestants: students; individuals/small teams/businesses; and large businesses ([see Section V](#)). NIJ will evaluate all entries on how accurately they forecast the outcome of recidivism ([see Section VIII](#)). Recidivism is defined in this Challenge as an arrest for a new crime. To receive prize money, winning applicants must provide a comprehensive document detailing the lessons learned about what variables did and did not matter to their final forecasting model and, when applicable, what type of models outperformed other models.<sup>[1]</sup> Within this document, contestants are encouraged to provide potential other measures that could be used in the future, and any additional intellectual property such as specific techniques, weighting, other sensitive decisions, and, if possible, uploading their code to an open source platform (e.g., Github) ([see Section IX](#)).

The Challenge [uses data from the State of Georgia](#) about persons released from prison to parole supervision for the period January 1, 2013, through December 31, 2015 ([see Section XII](#)). This data went through a process to reduce the risk of reidentification. This process included dropping variables, aggregating variable attributes, and aggregating variables. This data then went through a disclosure risk analysis to ensure these precautions were acceptable. See [Appendix 1](#) for more information on the preparation of the dataset.

This dataset is split into two sets, training and test. We used a 70/30 split, indicating that 70% of the data is in the training dataset and 30% in the test dataset. The training dataset includes the four dichotomous dependent variables measuring if an individual recidivated in the three-year follow-up period (yes/no), as well as recidivated by time period (year 1, year 2, or year 3). Recidivism is measured as an arrest for a new felony or misdemeanor crime within three years of the supervision start date. The test dataset does not include the four dependent variables. The initial test dataset will include all individuals selected in the 30% test dataset. After the first Challenge period (forecasting the probability individuals recidivated year 1) concludes, a second test dataset will be released containing only those individuals that did not recidivate year 1. The same will be done after the second Challenge period. It should also be noted that the test dataset

will contain variables that describe supervision activities, such as drug testing and employment. These data will not appear in the test dataset until the second Challenge period (i.e., year 2 dataset). We believe this is more reflective of practice where activities must accrue and correctional officers must become aware prior to a recidivism event. The additional data released at the second Challenge period will not change at the third Challenge period release (i.e., year 3 dataset); they are measures of supervision activities during the entire time people were under supervision or until the date of recidivism, for those arrested. The only thing that changes with the third Challenge period release is the removal of those individuals that did recidivate in year 2.

At the end of each submission period, contestants will submit a data file that contains two variables: the original ID field and a probability field containing the percent likelihood of recidivating, for each individual in the test dataset ([see Section V](#)). Contestants are encouraged to provide an estimate (with uncertainty) of how well they expect their algorithm to perform. After all submission periods have ended, NIJ will compare the Challenge entries to the actual data for each recidivism period, and will determine the winners among the different categories of contestants. The most accurate forecasts for males, females, and a combined score for each time period will be awarded prizes. Additionally, the forecasts with the least amount of racial bias for the male and female categories for each year will qualify for an additional prize ([see Section VIII](#)).<sup>[2]</sup> There are 114 prizes. Each contestant/team may win up to 15 prizes. NIJ's total prize allocation across all contestants is \$723,000 (see Sections VIII and IX).

NIJ is providing a codebook that details the available variables in the provided dataset (see [Section XII](#) & [Appendix 2](#)). Contestants should be aware that other entities may make non-administrative data available through free or fee-based services (e.g., cloud and data sharing sites) that may or may not also be useful in developing their algorithms. Contestants are permitted, but not required, to use any other datasets or services (see [Section IV](#) for potential sources of additional data).

### III. Rationale for Challenge —

Community correction departments divert millions of people from incarceration to the community. At yearend 2018, 4.39 million adults were on probation or parole. This equates to approximately 1 in 58 adults in the U.S. being under community supervision. <sup>[3]</sup> Unfortunately for many people on probation and parole, the community sentence fails

to reduce additional system involvement. A recent examination of state prison admissions indicated that 45% of admissions were the result of probation and parole violations. These “revolving door” admissions cost an estimated 9.3 billion dollars annually.[4]

Community corrections officers fulfill a complex dual function of surveilling persons on probation and parole to promote community public safety, while simultaneously providing rehabilitative programming to promote individual desistance from crime. Decisions regarding the appropriate balance of surveillance and support are intricate and significant. If the community corrections agency does too much or too little, the reentry plan will likely fail to prevent recidivism, which could be harmful to the individual and their community.[5] Predicting recidivism risk is critical to these decisions.

Community corrections officers’ caseloads are often large. In many states, officers supervise hundreds of individuals on their caseload.[6] Furthermore, recent policy decisions are placing individuals under supervision who would have previously been incarcerated.[7] Noting that optimal caseloads may vary, the American Probation and Parole Association offers some guidelines on caseload size: 20:1 for intensive supervision; 50:1 for moderate to high risk individuals; and 200:1 for low risk individuals.[8] In Georgia, the average caseload size is 123 when excluding specialized caseloads, with specialized caseloads capped at 50.[9] Most risk assessment instruments have broad risk categories. Although these categories likely indicate particular types of programming and monitoring strategies, practitioners could benefit from a more nuanced view of the characteristics of persons most at risk for recidivating among all persons on supervision classified as “high risk.”

## **Mitigating Potential Bias**

Community supervision directly impacts individuals, families, communities, public safety stakeholders, and others. Some groups are more likely than others to be involved in correctional systems. Incarceration and the subsequent release of persons is concentrated in neighborhoods of disadvantaged populations and communities of color.[10] Racial and ethnic minorities continue to be over represented in community supervision compared to their representation in the U.S. population. At the end of 2018, 46% of individuals on parole identified as white, 37% as Black, 15% as Hispanic, 1% as American Indian/Alaskan Native, 1% as Asian, less than 1% as Native Hawaiian/Other Pacific Islander, and less than 1% as two or more races.[11] The racial disproportionality



is particularly evident for Black individuals who make up 13% of the nation's population and 37% of persons on parole supervision.<sup>[12]</sup> Research has also indicated that employment challenges and rates of poverty are more severe for Blacks and Hispanics upon release compared to Whites.<sup>[13]</sup> In exploring ways to help individuals desist from criminal activity, it remains important to examine the connection between recidivism and race and ethnicity.

Within the criminal justice system, bias can be introduced prior to or at arrest, and compounded along each stage of the criminal justice system process. Since institutional and community corrections are at the end of the system, their risk assessments are implemented at a point where Blacks, Latinos, and American Indians and Alaska Natives are overrepresented among those who have been arrested and incarcerated.<sup>[14]</sup> Risk assessments, in general, tend to poorly predict risk among people of color.<sup>[15]</sup> This results in false-positives, in other words, when a risk assessment indicates that an individual is at high-risk for recidivating when they are not. Risk assessments over-predict risk for recidivism among people of color, and suggest they pose a greater risk to public safety than they actually do.<sup>[16]</sup> The over-prediction of risk typically results in an individual being subject to greater surveillance and more stringent conditions than necessary. Research shows a mismatch between risk and supervision conditions may actually increase the likelihood for recidivism.<sup>[17]</sup> The inclusion of more detailed, untapped information with common risk assessment protocols may provide greater accuracy and mitigate bias, thus producing fairer outcomes that facilitate successful re-entry into the community and reduce recidivism. This Challenge will help achieve that goal.

## **Understanding Gender-Based Differences**

At the end of 2019, males accounted for 92% of the nation's prison population.<sup>[18]</sup> Similar proportions are observed for those under community supervision, with females accounting for 25% of those on probation and 13% of those on parole.<sup>[19]</sup> These data support the notion that the pathways to criminality and institutional custody are different for men and women.<sup>[20]</sup> Many common risk assessments employed today tend to underperform when predicting risk among women probationers and parolees.<sup>[21]</sup> This lack of accuracy increases the likelihood of recidivism and may undermine the ability of practitioners to meet women's unique criminogenic risks and needs. The Challenge seeks to increase knowledge in these areas to better support and address the needs of men and women.

## **Challenge Implications**

Under this Challenge, NIJ is providing a large sample accompanied with rich data amendable for additional data to be paired with it. NIJ expects that new and more nuanced information will be gained from the Challenge and help address high recidivism among persons under community supervision. Findings could directly impact the types of factors considered when evaluating risk of recidivism and highlight the need to support people in specific areas related to reincarceration. Additionally, the Challenge could provide guidance on gender specific considerations and strategies to account for racial bias during risk assessment.

#### IV. Potential Sources for Additional Information and Data

Contestants can view the following and other sites/resources for reference or to obtain data to merge with the Challenge dataset provided:

- [BJS Publications on Probation and Parole Populations](#)
- [Georgia Department of Community Supervision Dashboard](#)
- [Georgia Department of Community Supervision](#)
- [Georgia Council on Criminal Justice Reform](#)
- [Georgia Census Page](#)
- [Georgia Crime Information Center](#)
- [Georgia State Board of Pardons and Paroles](#)
- [Georgia Emergency Management and Homeland Security Agency](#)
- [The Census Bureau's American Community Survey Public Use Microdata](#)
- [Accessing PUMs Data](#)
- [Homeland Infrastructure Foundation-Level Data \(HIFLD\)](#)
- [National Archive of Criminal Justice Data](#)
- [TIGER Data Products Guide](#)

#### V. How to Enter

Entries must be submitted through the [Challenge submission page](#). Registration and entry are free. Contestants must submit their entry via the National Institute of Justice's "Recidivism Forecasting Challenge" announcement. The Challenge data will be released in multiple stages, including training data, the initial data for the 1-year forecast, and then updated data for the 2-year and 3-year forecasts. Consequently, the Challenge will have 3 separate submission deadlines (see [Sections VI](#) and [X](#)). Entries must consist of a



team roster ([see Section IX](#)) and a data file containing the results from their analysis, as described later in this section. The winners of the Challenge will be required to submit a research paper to receive their prize money ([see Section IX](#)).

Contestants must submit entries under the appropriate contestant category (individuals must be a U.S. resident and companies must have an office with a U.S. business license):

**Student:** Enrolled as a full-time student in high school or as a full-time, degree-seeking student in an undergraduate program (associate or bachelor’s degree). Graduate students, professors, and other individuals not described in the student category must enter in the Small Team/Business category or the Large Business category.

**Small Team/Business:** A team comprised of 1-10 individuals (e.g., graduate student(s), professor(s), or other non-student individuals), or a small business with less than 11 employees. Teams should enter using a team name and provide a list of all individuals on the team. A small business should enter using the name of its business.

**Large Business:** A business with 11 or more employees. Large businesses should enter using a team name (the name of its business) and list of all individuals who worked on the Challenge on the team roster.

A Student may enter into the Small Team/Business category or the Large Business category if they choose, and a Small Team/Business contestant may enter the Large Business category if desired; however, large businesses may *not* enter the Student or Small Team/Business categories.

Individuals may participate in only one Challenge entry for each Challenge time period (i.e., the 1-year forecast, 2-year forecast, and 3-year forecast). A Challenge entry is defined as forecasts for one or more of the Challenge groups of interest (i.e., males, females, and combined forecasts).[\[22\]](#)

Entries must meet these requirements:

Requirement	Description of requirement
File formats (must be machine readable)	txt, rtf, pdf, doc, docx, odt, ppt, pptx, odp, xls, xlsx, ods Any file that contains macros or similar executable code will be disqualified.

Required variables	Maintain Original ID field provided (Named ID) Probability Field, between zero and one with up to four decimal places (Named Probability)
--------------------	---

Entries should **NOT** be submitted in a zipped folder. Contestant’s data file should contain two data fields/variables; the original ID field and the probability field (between 0 and 1, with up to 4 decimal places). The individual files should be uploaded with the following naming convention: “TEAMNAME\_1YearForecast,” “TEAMNAME\_2YearForecast,” and “TEAMNAME\_3YearForecast.” Entries that do not conform to this naming convention **will be** disqualified. If contestants run separate models for males and females the results should be combined into one data file for submission. Any entry that NIJ deems to contain any potentially malicious file will be disqualified.

[Download a sample submission.](#)

## VI. Important Dates

- On April 30, 2021, Initial release of training data.
- On April 30, 2021, Initial release of test data.
- May 31, 2021, End of submission period 1.
- June 1, 2021, Release of updated test data.
- June 15, 2021, End of submission period 2.
- June 16, 2021, Release of final test date.
- June 30, 2021, End of submission period 3.
- By August 16, 2021, winners will be notified and Challenge website updated with an announcement of the winners.
- By September 17, 2021, winners are to submit paper outlining the variables that were tested, indicating which were of statistical significance and which were not.
- Payment will occur as timely as possible after NIJ’s acceptance of the submitted paper.

## VII. Judges

NIJ staff (in consultation with BJS and BJA staff) have selected algorithms described in the following section ([see Section VIII](#)) to judge the accuracy and bias of the forecasts. The Director of NIJ or their designee will make the final award determination. If the Director of NIJ or their designee determines that no entry is deserving of an award, no prizes will be awarded.

## VIII. Judging Criteria

It is important to note at the onset of this section that NIJ is not implying that these are the best measures for understanding the accuracy of risk predictions or measuring fairness. However, with any competition a transparent metric is necessary so that all contestants understand how they will be scored.

NIJ will use the Brier score to measure the accuracy of Challenge entries:

$$\text{Brier Score} = \frac{1}{n} \sum_{t=1}^n (f_t - A_t)^2$$

Where n is the count of individuals in the test dataset,  $f_t$  is the forecasted probability of recidivism for individual t, and  $A_t$  is the actual outcome (0,1) for individual t. Since the Brier score is a measure of error, applicants should look to minimize this metric (lowest score wins).

In order to measure algorithm racial fairness (Black and White), NIJ will use a rubric that will use the Brier Score calculated above as the base score and penalize it based on the conditional procedure accuracy. To do this, NIJ will use a 0.5 threshold to convert the predicted probabilities into binary (yes/no) predictions. From there, a confusion matrix will be constructed in the following format:

Confusion Matrix			
	$\hat{Y}$ – Fail	$\hat{Y}$ – Succeed	Conditional Procedure Error
Y – Fail	A True Positive	B False Negative	B/(A+B) False Negative Rate
Y – Succeed	C False Positive	D True Negative	C/(C+D) False Positive Rate

Conditional User Error	$C/(A+C)$ Failure Prediction Error	$B/(D+B)$ Success Prediction Error	$(C+B)/(A+B+C+D)$ Overall Procedure Error
------------------------	---------------------------------------	---------------------------------------	--

Here  $Y$  is the actual outcome and  $\hat{Y}$  is the forecasted outcome. A “fail” is considered a positive outcome (the variable of interest occurred, i.e., the individual did recidivate); and, a “succeed” is considered a negative outcome (the variable of interest did not occur, i.e., the individual did not recidivate).

Based on this confusion matrix, NIJ opted for a measure of fairness that considers the conditional procedure accuracy of a forecast. As such, we based this measure on the false positive (i.e., where an individual is forecasted to recidivate but does not) equality of an algorithm between races; in other words, we look at the difference in false positive rates between Whites and Blacks. NIJ notes that we could have selected a false negative rate; however, only one of these could be used and NIJ selected false positive due to the meaning of this error. Therefore, the fairness penalty (FP) function is the difference in false positive rates:

$$FP = 1 - |FP_{Black} - FP_{White}|$$

Since the Brier Score is a measure of error on predictions bound between 0 and 1, the error should also be bound between 0 and 1. As such we consider one minus the Brier Score a metric of correctness. The final index NIJ will use to calculate which algorithms are the most accurate while accounting for bias is:

$$Fair\ and\ Accurate = (1 - BS)(FP)$$

Since this is now a measure of fairness and accuracy, applicants should look to maximize this metric (highest score wins).

If an applicant fails to provide a probability for an individual in the test dataset it will be considered an incorrect forecast (i.e., if the individual did recidivate, it will be scored as though you forecasted a 0 probability; and, if the individual did not recidivate, it will be scored as though you forecasted a 1 probability).

## IX. Prizes

A total of \$723,000 is available in prizes and will be divided among the three contestant categories and accounting for racial bias as follows:

## Student Category [🔗](#)

Student Category: Male Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$5,000	\$5,000	\$5,000
2nd Place	\$3,000	\$3,000	\$3,000
3rd Place	\$1,000	\$1,000	\$1,000

Student Category: Female Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$5,000	\$5,000	\$5,000
2nd Place	\$3,000	\$3,000	\$3,000
3rd Place	\$1,000	\$1,000	\$1,000

Student Category: Average			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$5,000	\$5,000	\$5,000
2nd Place	\$3,000	\$3,000	\$3,000
3rd Place	\$1,000	\$1,000	\$1,000

## Small Team Category [🔗](#)

Small Team Category: Male Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$10,000	\$10,000	\$10,000

2nd Place	\$6,000	\$6,000	\$6,000
3rd Place	\$2,000	\$2,000	\$2,000
4th Place	\$1,000	\$1,000	\$1,000

<b>Small Team Category: Female Parolees</b>			
<b>Place</b>	<b>1 Year Forecast</b>	<b>2 Year Forecast</b>	<b>3 Year Forecast</b>
1st Place	\$10,000	\$10,000	\$10,000
2nd Place	\$6,000	\$6,000	\$6,000
3rd Place	\$2,000	\$2,000	\$2,000
4th Place	\$1,000	\$1,000	\$1,000

<b>Small Team Category: Average</b>			
<b>Place</b>	<b>1 Year Forecast</b>	<b>2 Year Forecast</b>	<b>3 Year Forecast</b>
1st Place	\$10,000	\$10,000	\$10,000
2nd Place	\$6,000	\$6,000	\$6,000
3rd Place	\$2,000	\$2,000	\$2,000
4th Place	\$1,000	\$1,000	\$1,000

## Large Team Category [↗](#)

<b>Large Team Category: Male Parolees</b>			
<b>Place</b>	<b>1 Year Forecast</b>	<b>2 Year Forecast</b>	<b>3 Year Forecast</b>
1st Place	\$25,000	\$25,000	\$25,000
2nd Place	\$12,000	\$12,000	\$12,000
3rd Place	\$4,000	\$4,000	\$4,000
4th Place	\$2,000	\$2,000	\$2,000



5th Place	\$1,000	\$1,000	\$1,000
-----------	---------	---------	---------

Large Team Category: Female Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$25,000	\$25,000	\$25,000
2nd Place	\$12,000	\$12,000	\$12,000
3rd Place	\$4,000	\$4,000	\$4,000
4th Place	\$2,000	\$2,000	\$2,000
5th Place	\$1,000	\$1,000	\$1,000

Large Team Category: Average			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$25,000	\$25,000	\$25,000
2nd Place	\$12,000	\$12,000	\$12,000
3rd Place	\$4,000	\$4,000	\$4,000
4th Place	\$2,000	\$2,000	\$2,000
5th Place	\$1,000	\$1,000	\$1,000

## Accounting for Racial Bias Category [🔗](#)

Accounting for Racial Bias: Male Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast
1st Place	\$12,500	\$12,500	\$12,500
2nd-5th place in this prize category will be recognized on the Challenge website.			

Accounting for Racial Bias: Female Parolees			
Place	1 Year Forecast	2 Year Forecast	3 Year Forecast

1st Place	\$12,500	\$12,500	\$12,500
2nd-5th place in this prize category will be recognized on the Challenge website.			

*Note that there is no "Average" category for the "Accounting for Racial Bias" category.*

Prizes will be awarded after winners have submitted all required documents including a research paper, proof of student status (if relevant), and the necessary financial information.<sup>[23]</sup> The research paper, based on the prize(s) won, should describe which variables did and did not matter to the final forecasting model, and when applicable, what type of models outperformed other models. Contestants should also discuss false positive and false negative results and how they attempted to minimize racial bias. Additionally, contestants should discuss how the methods used in their final projections could be used in practice. Contestants are encouraged to provide additional intellectual property such as techniques, weighting, other sensitive materials, and, if possible, uploading their code to an open source platform (e.g., Github). Contestants should also consider alternative measures that could be used in the future. The research papers will be made available for public use after they have been reviewed by NIJ.

Contestants who win multiple prizes will receive one prize payment for the total amount won. If and when travel is permitted, NIJ may invite a member of the winning teams to Washington, D.C., to participate in a round table discussion.

Contestants entering as a team must submit a "Team Roster" that lists the individual team members. The Team Roster must indicate what category you are applying to and the proportional amount of any prize to be paid to each member. Each team member must sign the Team Roster next to his or her name. For example:

John Q. Public      20%      \_\_\_\_\_(Signature)\_\_\_\_\_.

This allows teams to designate proportional prizes to their members. Team entries that do not submit a Team Roster signed by each member of the team will be disqualified.

All prizes are subject to the availability of appropriated funds and will be awarded at the discretion of the NIJ Director.

## X. Other Rules and Conditions

## A. Submission Periods [↗](#)

The 1-year forecast must be submitted by 11:59 p.m. (ET) May 31, 2021, the 2-year forecast by 11:59 p.m. (ET) June 15, 2021, and the 3-year forecast by 11:59 p.m. (ET) June 31, 2021. Entries submitted after the designated Submission Periods will be disqualified and will not be reviewed.

## B. Eligibility [↗](#)

The Challenge is open to: (1) individual residents of the 50 United States, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, Guam, and American Samoa who are at least 13 years old at the time of entry; (2) teams of eligible individuals; and (3) corporations or other legal entities (e.g., partnerships or nonprofit organizations) that are domiciled in any jurisdiction specified in (1). Entries by contestants under the age of 18 must include the co-signature of the contestant's parent or legal guardian.

Contestants may submit or participate in the submission of only one entry. Employees of NIJ and individuals or entities listed on the Federal Excluded Parties list (available from SAM.gov) are not eligible to participate. Employees of the Federal Government should consult with the Ethics Officer at their place of employment prior to submitting an entry for this Challenge. The Challenge is subject to all applicable federal laws and regulations. Submission of an entry constitutes a contestant's full and unconditional agreement to all applicable rules and conditions. Eligibility for the prize award(s) is contingent upon fulfilling all requirements set forth herein.

## C. General Warranties and Conditions [↗](#)

- **Release of Liability:** By entering the Challenge, each contestant agrees to: (a) comply with and be bound by all applicable rules and conditions, and the decisions of NIJ, which are binding and final in all matters relating to this Challenge; (b) release and hold harmless NIJ and any other organizations responsible for sponsoring, fulfilling, administering, advertising or promoting the Challenge, and all of their respective past and present officers, directors, employees, agents and representatives (collectively, the "Released Parties") from and against any and all claims, expenses, and liability arising out of or relating to the contestant's entry or participation in the Challenge, and/or the contestant's acceptance, use, or misuse of the prize or recognition.

The Released Parties are not responsible for: (a) any incorrect or inaccurate information, whether caused by contestants, printing errors, or by any of the equipment or programming associated with, or used in, the Challenge; (b) technical failures of any kind, including, but not limited to malfunctions, interruptions, or disconnections in phone lines or network hardware or software; (c) unauthorized human intervention in any part of the entry process or the Challenge; (d) technical or human error which may occur in the administration of the Challenge or the processing of entries; or (e) any injury or damage to persons or property which may be caused, directly or indirectly, in whole or in part, from each contestant's participation in the Challenge or receipt or use or misuse of any prize. If, for any reason, a contestant's entry is confirmed to have been erroneously deleted, lost, or otherwise destroyed or corrupted, contestant's sole remedy is to submit another entry in the Challenge so long as the resubmitted entry is prior to the deadline for submissions.

- **Termination and Disqualification:** NIJ reserves the authority to cancel, suspend, and/or modify the Challenge, or any part of it, if any fraud, technical failures, or any other factor beyond NIJ's reasonable control impairs the integrity or proper functioning of the Challenge, as determined by NIJ in its sole discretion. NIJ reserves the authority to disqualify any contestant it believes to be tampering with the entry process or the operation of the Challenge, or to be acting in violation of any applicable rule or condition. Any attempt by any person to undermine the legitimate operation of the Challenge may be a violation of criminal and civil law, and, should such an attempt be made, NIJ reserves the authority to seek damages from any such person to the fullest extent permitted by law. NIJ's failure to enforce any term of any applicable rule or condition shall not constitute a waiver of that term.
- **Intellectual Property:** By entering the Challenge, each contestant warrants that (a) he or she is the author and/or authorized owner of the entry; (b) that the entry is wholly original with the contestant (or is an improved version of an existing solution that the contestant is legally authorized to enter in the Challenge); (c) that the submitted entry does not infringe any copyright, patent, or any other rights of any third party; and (d) that the contestant has the legal authority to assign and transfer to NIJ all necessary rights and interest (past, present, and future) under copyright and other intellectual property law, for all material included in the Challenge proposal that may be held by the contestant and/or the legal holder of those rights. Each contestant agrees to hold the Released Parties harmless for any infringement

of copyright, trademark, patent, and/or other real or intellectual property right, which may be caused, directly or indirectly, in whole or in part, from contestant's participation in the Challenge.

- **Publicity:** By entering the Challenge, each contestant consents, as applicable, to NIJ's use of his/her/its/their name, likeness, photograph, voice, and/or opinions, and disclosure of his/her/its/their hometown and State for promotional purposes in any media, worldwide, without further payment or consideration.
- **Privacy:** Personal and contact information submitted through <https://nij.ojp.gov> is not collected for commercial or marketing purposes. Information submitted throughout the Challenge will be used only to communicate with contestants regarding entries and/or the Challenge.
- **Compliance with Law:** By entering the Challenge, each contestant guarantees that the entry complies with all federal and state laws and regulations.

## XI. Prize Disbursement and Challenge Winners

### A. Prize Disbursement and Requirements

Prize winners must comport with all applicable laws and regulations regarding prize receipt and disbursement. For example, NIJ is not responsible for withholding any applicable taxes from the award.

### B. Specific Disqualification Rule

If the announced winner(s) of the Challenge prize is found to be ineligible or is disqualified for any reason listed under "Other Rules and Conditions: Eligibility," NIJ may make the award to the next runner(s) up, as previously determined by the NIJ Director.

### C. Rights Retained by Contestants and Challenge Winners

1. All legal rights in any materials or products submitted in entering the Challenge are retained by the contestant and/or the legal holder of those rights. Entry in the Challenge constitutes express authorization for NIJ staff to review and analyze any and all aspects of submitted entries.
2. Upon acceptance of any Challenge prizes, the winning contestant(s) agrees to allow NIJ to post on NIJ's website all associated files of the winning

submission. This will allow others to verify the scores of winning submissions.

## XII. Codebook

[Appendix 2](#) contains the codebook for the Challenge dataset. The codebook will provide a description of the variables within the dataset and other relevant information regarding the provided data.

Download a pdf version of the codebook.

## XIII. Contact Information

**For substantive questions about the Challenge**, e-mail NIJ at [NIJRecidivismChallenge@usdoj.gov](mailto:NIJRecidivismChallenge@usdoj.gov)

**For technical questions about the application process**, first read through How to Enter (Section V). If you still have questions, contact the OJP Grants Management System Help Desk – open 7 a.m. to 9 p.m. ET – at 1-888-549-9901 or by email at [OJP.ITServiceDesk@usdoj.gov](mailto:OJP.ITServiceDesk@usdoj.gov).

# Appendices

## Appendix 1 – Initial Data Preparation

The NIJ Recidivism Forecasting Challenge uses data from the State of Georgia about persons released from Georgia prisons on discretionary parole to the custody of the Georgia Department of Community Supervision (DCS) for the purpose of post-incarceration supervision between January 1, 2013 and December 31, 2015.

DCS provided data on supervised persons to include demographics, prison and parole case information, prior community supervision history, conditions of supervision as articulated by the Board of Pardons and Paroles, and supervision activities (violations,



drug tests, program attendance, employment, residential moves, and accumulation of delinquency reports for violating conditions of parole).

The Georgia Bureau of Investigation provided data from the Georgia Crime Information Center (GCIC) statewide criminal history records repository. The GCIC data provides the Georgia prior criminal history measures, to include arrest and conviction episodes prior to prison entry. GCIC “rap sheet” data captures all charges at an arrest episode, defined as a custodial arrest where a person is fingerprinted by law enforcement. Arrest episodes with multiple charges are described in this data by the most serious charge. The exception is criminal history domestic violence and gun charges, which count all charges across all episodes. GCIC data also provides the recidivism measure, defined as a new felony or misdemeanor arrest episode within three years of parole supervision start date.

Roughly 8% of the original population were excluded from the Challenge data for the following reasons (in order of frequency): lack of unique identifier to link DCS to GCIC, invalid Georgia zip code, transferred to another state for supervision, invalid birth date, expired by death. Nine youths (under the age of 18 at prison release) were excluded. The remaining population contained less than 500 people identified as Hispanic, and less than 100 individuals in each of the following categories: Asian, Native American, other, and unknown. To prevent potential deductive disclosure of individuals identities, the Challenge data only contains the racial categories of Black and White.

Creation of the Challenge dataset met all Office of Justice Programs Institutional Review Board requirements to ensure human subject privacy and protection. The person-level data are de-identified, stripped of person, address, and agency identifiers. The final Challenge dataset includes over 25,000 records with data fields amendable for pairing to additional data sources, such as the person’s residential U.S. Census Bureau Public Use Microdata Area (PUMA). Each person’s residential address at prison release was mapped to a PUMA, then geographically neighboring PUMAs were collapsed into 25 unique spatial units (defined in [Appendix 3 – Combined PUMA Definitions](#)). The following table defines the data fields. The level of missing data varies across data field and is indicated by a blank.

# Recidivism Forecasting Challenge Database - Fields Defined [🔗](#)

3/29/2021

Download a pdf version of the codebook.

Position	Variable	Definition
1	ID	Unique Person ID

## Supervision Case Information [🔗](#)

Position	Variable	Definition
2	Gender	Gender (M=Male/F=Female)
3	Race	Race (Black or White)
4	Age_at_Release	Age Group at Time of Prison Release (18-22, 23-27, 28-32, 33-37, 38-42, 43-47, 48+)
5	Residence_PUMA*	Residence US Census Bureau <a href="#">PUMA Group*</a> at Prison Release
6	Gang_Affiliated	Verified by Investigation as Gang Affiliated
7	Supervision_Risk_Score_First	First Parole Supervision Risk Assessment Score (1-10, where 1=lowest risk)
8	Supervision_Level_First	First Parole Supervision Level Assignment (Standard, High, Specialized)

## Prison Case Information [🔗](#)

**Captured at prison admission with the exception of prison days, which is calculated as (prison release date - prison admit date).**

Position	Variable	Definition
9	Education_Level	Education Grade Level at Prison Entry (<high school, High School diploma, at least some college)

10	Dependents	# Dependents at Prison Entry (0, 1, 2, 3+)
11	Prison_Offense	Primary Prison Conviction Offense Group (Violent/Sex, Violent/Non-Sex, Property, Drug, Other)
12	Prison_Years	Years in Prison Prior to Parole Release (<1, 1-2, 2-3, 3+)

### Prior Georgia Criminal History [🔗](#)

**Georgia Crime Information Center (GCIC) fingerprintable arrest and conviction history prior to prison entry. Arrest and conviction episodes may contain multiple charges.**

Position	Variable	Definition
13	Prior_Arrest_Episodes_Felony	# Prior GCIC Arrests with Most Serious Charge=Felony
14	Prior_Arrest_Episodes_Misdemeanor	# Prior GCIC Arrests with Most Serious Charge=Misdemeanor
15	Prior_Arrest_Episodes_Violent	# Prior GCIC Arrests with Most Serious Charge=Violent
16	Prior_Arrest_Episodes_Property	# Prior GCIC Arrests with Most Serious Charge=Property
17	Prior_Arrest_Episodes_Drug	# Prior GCIC Arrests with Most Serious Charge=Drug
18	Prior_Arrest_Episodes_PPViolationCharges	# Prior GCIC Arrests with Probation/Parole Violation Charges
19	Prior_Arrest_Episodes_DomesticViolenceCharges	Any Prior GCIC Arrests with Domestic Violence Charges
20	Prior_Arrest_Episodes_GunCharges	Any Prior GCIC Arrests with Gun Charges

21	Prior_Conviction_Episodes_Felony	# Prior GCIC Felony Convictions with Most Serious Charge=Felony
22	Prior_Conviction_Episodes_Misdemeanor	# Prior GCIC Convictions with Most Serious Charge=Misdemeanor
23	Prior_Conviction_Episodes_Violent	Any Prior GCIC Convictions with Most Serious Charge=Violent
24	Prior_Conviction_Episodes_Property	# Prior GCIC Convictions with Most Serious Charge=Property
25	Prior_Conviction_Episodes_Drug	# Prior GCIC Convictions with Most Serious Charge=Drug
26	Prior_Conviction_Episodes_PPViolationCharges	Any Prior GCIC Convictions with Probation/Parole Violation Charges
27	Prior_Conviction_Episodes_DomesticViolenceCharges	Any Prior GCIC Convictions with Domestic Violence Charges
28	Prior_Conviction_Episodes_GunCharges	Any Prior GCIC Convictions with Gun Charges

### Prior Georgia Community Supervision History [🔗](#)

**Georgia community supervision revocations prior to prison entry. A revocation is an unsuccessful termination of supervision. A parole revocation reflects a return to prison to serve the remaining sentence.**

Position	Variable	Definition
29	Prior_Revocations_Parole	Any Prior Parole Revocations

30	Prior_Revocations_Probation	Any Prior Probation Revocations
----	-----------------------------	---------------------------------

## Georgia Board of Pardons and Paroles Conditions of Supervision [🔗](#)

**Discretionary parole release requires people to abide by conditions set by the Board; those affecting programming or notification are included here.**

Position	Variable	Definition
31	Condition_MH_SA	Parole Release Condition = Mental Health or Substance Abuse Programming
32	Condition_Cog_Ed	Parole Release Condition = Cognitive Skills or Education Programming
33	Condition_Other	Parole Release Condition = No Victim Contact or Electronic Monitoring or Restitution or Sex Offender Registration/Program

## Supervision Activities [🔗](#)

**Includes violations, drug tests, delinquency reports for violating conditions, program attendances, residence changes, and employment during the parole supervision episode.**

Position	Variable	Definition
34	Violations_ElectronicMonitoring	Any Violation for Electronic Monitoring
35	Violations_InstructionsNotFollowed	Any Violation for Not Following Instructions
36	Violations_FailToReport	Any Violation for Failure to Report
37	Violations_MoveWithoutPermission	Any Violation for Moving Without Permission
38	Delinquency_Reports	# Parole Delinquency Reports
39	Program_Attendances	# Program Attendances
40	Program_UnexcusedAbsences	# Program Unexcused Absences
41	Residence_Changes	# Residence Changes/Moves (new

		zip codes) During Parole
42	Avg_Days_per_DrugTest	Average Days on Parole Between Drug Tests
43	DrugTests_THC_Positive	% Drug Tests Positive for THC/Marijuana
44	DrugTests_Cocaine_Positive	% Drug Tests Positive for Cocaine
45	DrugTests_Meth_Positive	% Drug Tests Positive for Methamphetamine
46	DrugTests_Other_Positive	% Drug Tests Positive for Other Drug
47	Percent_Days_Employed	% Days Employed While on Parole
48	Jobs_Per_Year	Jobs Per Year While on Parole
49	Employment_Exempt	Employment is Not Required (Exempted)

### Recidivism Measures [↗](#)

**Recidivism is measured as a Georgia fingerprintable arrest recorded in GCIC for a new felony or misdemeanor crime within 3 years of parole supervision start date.**

Position	Variable	Definition
50	Recidivism_Within_3years	New Felony/Mis Crime Arrest within 3 Years of Supervision Start
51	Recidivism_Arrest_Year1	Recidivism Arrest Occurred in Year 1
52	Recidivism_Arrest_Year2	Recidivism Arrest Occurred in Year 2
53	Recidivism_Arrest_Year3	Recidivism Arrest Occurred in Year 3

## Appendix 3 - Combined PUMA Definitions

**Residence at Prison Release by US Census Bureau Public Use Microdata Area (PUMA)**



Group Defined	
Code	PUMAs included
1	1003, 4400
2	1008, 4300
3	1200, 1300
4	1400, 1500, 1600
5	1700, 1800
6	2001, 2002, 2003, 4005
7	100, 200, 500
8	4000, 4100, 4200
9	5001, 6001, 6002
10	2400, 5002
11	1001, 3004, 4600
12	1002, 1005, 3300, 3400, 4001, 4002, 4006
13	3101, 3102
14	1900, 3900, 4003, 4004
15	3001, 3002, 3003, 3005
16	2500, 4500
17	2800, 2900, 3200, 3500
18	600, 700, 800
19	900, 1100
20	300, 401, 402
21	1004, 2100
22	2200, 2300

23	1006, 1007, 2004
24	2600, 2700
25	3600, 3700, 3800

## Notes

[note 1] Upon review by NIJ, the winner's documents will be made available to the public on the Challenge website <https://nij.ojp.gov/funding/recidivism-forecasting-challenge/>

[note 2] The Challenge data only contains individuals with the racial categories of Black and White to reduce the possibility of deductive disclosure of individuals identities. There were less than 500 individuals who were identified as Hispanic, and less than 100 individuals in each of the following categories Asian, Native American, other, and unknown; these cases were dropped from the sample to prevent inadvertent disclosure of personal Identifying information.

[note 3] Kaeble, D. & M. Alper (2020). *Probation and Parole in the United States, 2017-2018*. Washington, DC: Bureau of Justice Statistics.

[note 4] Council of State Governments (2019). *Confined and Costly: How Supervision Violations are Filling Prisons and Burdening Budgets*. Washington, DC: Council of State Governments.

[note 5] Lowenkamp, C., E Latesa. & A. Holsinger (2006). "The Risk Principle in Action: What have we Learned from 13,676 Offenders and 97 Correctional Programs?" *Crime & Delinquency* 52:77-93

[note 6] [Pew Charitable Trust \(2020\). \*Probation and Parole Systems Marked by High Stakes, Missed Opportunities\*. Washington, DC: Pew Charitable Trust.](#)

[note 7] Bird, M. & R. Grattet (2018). *Evaluating the Effects of Realignment Practices on Recidivism*. Washington, DC: National Institute of Justice.

[note 8] American Probation and Parole Association (2006). *Caseload Standards for Probation and Parole*. [www.appa-net.org/eweb/docs/APPA/stances/ip\\_CSPP.pdf](http://www.appa-net.org/eweb/docs/APPA/stances/ip_CSPP.pdf).

[note 9] Georgia Department of Community Supervisions, personal communication, January 11, 2021

[note 10] See Clear, T. (2007). *Studies in Crime and Public Policy. Imprisoning Communities: How Mass Incarceration Makes Disadvantaged Neighborhoods Worse*. London: Oxford University Press; National Research Council (2014). *The Growth of Incarceration in the United States: Exploring Causes and Consequences*. Washington, D.C.: National Academies Press; Sampson, R. & C. Loeffler (2010). "Punishment's Place: The Local Concentration of Mass Incarceration." *Daedalus* 139: 20-31.

[note 11] Kaeble, D. & M. Alper (2020). *Probation and Parole in the United States, 2017-2018*. Washington, DC: Bureau of Justice Statistics.

[note 12] Bradner, K. V. Schiraldi, N. Mejia, & E. Lopoo (2020). *More Work To Do: Analysis of Probation and Parole in the United States, 2017-2018*. New York: Columbia University Justice Lab.

[note 13] Western, B. & C. Sirois (2019). "Radicalized Re-entry: Labor Market Inequality After Incarceration." *Social Forces*: 97:1517-1542.

[note 14] See Berk, R. & A. Elzarka (2020). "Almost Politically Acceptable Criminal Justice Risk Assessment." *Criminology & Public Policy* 19:1231-1257.

[note 15] Desmarais, S., K. Johnson, & J. Singh (2016). "Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings." *Psychological Services*: 13:206-222.

[note 16] See Berk, R. & A. Elzarka (2020). "Almost Politically Acceptable Criminal Justice Risk Assessment." *Criminology & Public Policy* 19:1231-1257.

[note 17] Lowenkamp, C., E. Latesa, & A. Holsinger (2006). "The Risk Principle in Action: What have we Learned from 13,676 Offenders and 97 Correctional Programs?" *Crime & Delinquency* 52:77-93

[note 18] Carson, E. (2020). *Prisoners in 2019*. Washington, DC: Bureau of Justice Statistics.

[note 19] Kaeble, D. & M. Alper (2020). *Probation and Parole in the United States, 2017-2018*. Washington, DC: Bureau of Justice Statistics.

[note 20] National Institute of Justice (1999). *Research on Women and Girls in the Criminal Justice System*. Washington, DC: National Institute of Justice; Ramirez, R. (2016). *Reentry Considerations for Justice Involved Women*. Kensington, MD: Center for Effective Public Policy.

[note 21] Desmarais, S., K. Johnson, & J. Singh (2016). "Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings." *Psychological Services*: 13:206-222.

[note 22] A contestant is not required to provide a forecast for every Challenge submission time period. Additionally, team rosters may fluctuate depending on the submission period (i.e., gain or lose a member) with the requirement that contestants are only listed on one team roster for each Challenge submission time period (i.e., 1-year, 2-year, and 3-year forecast).

[note 23] The individual or team representative will be contacted by NIJ staff requesting contact information for all listed on the team roster, if not already submitted. Each person on the winning entries will then receive a form that asks for banking information that is sent directly to the Office of the Chief Financial Officer (OCFO); OCFO will then disperse the awards accordingly.



U.S. DEPARTMENT OF JUSTICE  
OFFICE OF JUSTICE PROGRAMS