The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Predicting Recidivism in Georgia Using Lasso Regression Models with Several New Constructs
Author(s): Duddon Evidence to Policy Research Team
Document Number: 305049
Date Received: July 2022
Award Number: NIJ Recidivism Forecasting Challenge Winning Paper

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs’ National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.
Predicting recidivism in Georgia using lasso regression models

with several new constructs

Report from the Duddon Evidence to Policy Research Team
to the NIJ Recidivism Forecasting Challenge Team

I. Introduction

Probation and parole have been the primary tools that U.S. criminal justice officials use to supervise convicted offenders within communities for more than 100 years, with the goal of reducing recidivism among offenders\(^1\) and making communities safer (Bonta et al., 2008). Yet the effectiveness of these tools, particularly in reducing recidivism, is still being questioned. Further, criminal justice systems in the U.S. have higher rates of community supervision than their European counterparts and U.S. parole and probation practices still demonstrate significant racial disparities (Bradner, 2020). As one researcher wrote, “While probation and post-release supervision serve important purposes in many cases, they are often imposed on the wrong people and executed in ways that predictably lead to revocation” (Klingele, 2013, p. 1015). Given the significant public resources that are spent on community supervision, researchers and government officials are continuing to assess the practice and its impact on reducing recidivism.

In this report, I summarize the factors associated with recidivism, prior research on the role of these factors in predicting recidivism, and the research design I undertook to develop recidivism predictions as part of the National Institute of Justice’s (NIJ) Recidivism Forecasting

---

\(^1\) For the sake of brevity in this report, I use the term “offender” to mean an individual who has been convicted of a misdemeanor or felony crime and who is under some form of community supervision within a criminal justice system. However, I recognize that the term “offender” may not be preferred or appropriate in many contexts.
Challenge. I submitted an entry with predictions for the third period of the challenge only; the entry placed fourth in the Male Parolees category for that period.

A. Factors influencing recidivism

Prior research has focused primarily on the individual characteristics of offenders that affect their probability of recidivism. Individual-level factors that influence recidivism include age, gender, race and ethnicity, levels of education and employment, past criminal behavior, family and housing characteristics, substance abuse and mental health history, and behaviors during treatment programs (e.g. Bonta & Andrews, 2016; Huebner, DeJong, & Cobbina, 2010; Mulder et al., 2011; Tillyer & Vose, 2011).

Other influential factors, however, are not characteristics of offenders themselves but rather of the people who supervise offenders following their prison release. For example, an extensive body of research indicates that human service interventions, delivered in the communities in which released offenders live, can be effective in reducing recidivism. In particular, interventions appear to be most effective when they adhere to a set of principles, first articulated in 1990, that fall under the headings of Risk, Need, and Responsivity (RNR). The RNR principles dictate that the level of community supervision should be matched to the risk level of the offender, that any accompanying interventions should be directly related to the offending itself, and that interventions should be delivered in a manner accessible to the offender (Andrews, Bonta, & Hoge, 1990).

More recent work has emphasized a range of additional problem-solving, interpersonal skills that should form the foundation for training of community supervision officers (Bonta & Andrews, 2016). Studies indicate that when offenders are assigned to probation officers who received training in these skills and principles, the odds of recidivism tend to be lower than the
odds of offenders assigned to officers who did not receive the training (Chadwick, Dewolf, & Serin, 2015). Thus, the supervision officer’s skills and training, and the strength of the officer’s relationship with the offender, may help reduce recidivism.

Subsequent studies moved beyond individual-level factors to examine contextual and environmental characteristics that influence recidivism. For example, researchers have found that offenders who return to neighborhoods with high levels of poverty and other measures of inequality and disadvantage recidivate at higher rates than those who return to more advantaged communities, even after controlling for individual-level factors (Kubrin & Stewart, 2006). Additional work, however, examined whether environmental factors moderated individual-level characteristics, finding that individual-level factors appear to be more influential (Tillyer & Vose, 2011).

**B. Predicting recidivism**

Research on the predictive value of these individual- and contextual-level characteristics has been ongoing for decades. Early studies led to the development of a group of assessment tools in the 1980s that incorporated both individual and contextual characteristics to appraise recidivism risk. In succeeding years, these tools, generally referred to as Level of Service (LS) scales, have been revised and improved to become the most frequently used risk assessment tools both in the U.S. and abroad (Olver, Stockdale, & Wormith, 2014). Regardless, LS scales have been criticized for not capturing gender-based variation and for being poorly predictive for female offenders (Gobeil, Blanchette, & Stewart, 2016). Similar arguments have been made with respect to racial and ethnic minorities, particularly minorities in the U.S. (Olver, Stockdale, & Wormith, 2014).
In general, the wide variation in practices and training among local criminal justice officials in the U.S. presents challenges for researchers in evaluating recidivism prediction tools (Phelps, 2020). At the same time, a growing body of research raises concerns about the fairness and accuracy of these tools (Klingele, 2020), although many researchers argue that computer-based prediction algorithms outperform human predictions (Slobogin, 2020; Zhiyuan et al., 2020).

Newer techniques for analyzing data, along with refinements of classical statistical approaches, could improve the effectiveness of recidivism prediction (Hester, 2019; Tollenaar & Van der Heijden, 2019) and, by extension, of community supervision itself. Some studies have estimated the risk of recidivism using traditional statistical approaches. These approaches include logistic regression models that incorporate various psychometric scales (Bernman et al., 2019), survival models (Hester, 2019), and factor analyses of various criminal risk assessment instruments, including one of the LS scales (Kroner, Mills, & Reddon, 2005).

In contrast to the goals of these and other approaches within inferential statistics, other methods, such as those using machine learning and data mining, are purely predictive (see Zeng, Ustun, & Rudin, 2017). For the goal of predicting recidivism, it is not yet clear that one approach is better than another (Tollenaar & Van der Heijden, 2013; 2019). Some recidivism prediction experts recommend the use of both approaches to obtain an optimal model, noting that even small improvements in predictive accuracy can translate into meaningful reductions in crime in some communities (Tollenaar & Van der Heijden, 2019).

Investigative journalists have argued, however, that some computer algorithm models using machine learning methods are inherently racially biased, stemming from systemic biases in the judicial systems that generate the underlying data and possibly in the algorithms themselves.
These arguments have not yet been validated (Wang & Han, et al. 2020). Nevertheless, efforts are ongoing to develop machine learning techniques that outperform both inferential statistics and humans.

Machine learning techniques used in recidivism prediction include use of increasingly complex decision tree models incorporating, for example, random forests, as well as logistic regression models with penalty terms added to prevent over-fitting (Duwe & Kim, 2017; Wang & Han, et al. 2020). In evaluating these options, at least one study indicated that random forest modeling may not be living up to its promise as a broadly-applicable predictive tool and that penalized logistic regression models may slightly outperform other approaches for predicting some types of recidivism (Tollenaar & Van der Heijden, 2019).

This prior research shows that traditional statistical approaches and more recent machine learning techniques may both be useful in predicting recidivism. Thus, I decided to attempt at least one approach in each category as part of my entry in the Recidivism Forecasting Challenge.

C. Team information

I entered the Challenge in my capacity as the Executive Director of Duddon Evidence to Policy Research, a law and policy research and consulting business that I operate as a sole proprietorship. In developing my models and subsequent predictions, I consulted with an experienced quantitative psychologist and psychometrician and with a Georgia attorney who had previously served as a public defender for DeKalb County. Additional information on these individuals is at the end of this report.

Time and financial constraints limited my entry to predicting recidivism for the third Challenge period only. Thus, the descriptions regarding the variables and models I used in my
forecasts apply to the Challenge training dataset released at the third period only. I used the Stata Statistical Software: Release 16 for all data management tasks and analyses.

**II. Variables**

The dataset the NIJ prepared for the Challenge provides information on most of the individual and contextual characteristics that prior research has found to be predictive of recidivism. This individual-level information includes both static attributes that do not change or that change in only one direction (e.g. age at first conviction) and dynamic characteristics that are subject to change over time.

However, although prior research has demonstrated that the training and skills of community supervision officer influence recidivism rates among offenders, a direct measure of these skills is not available in the Challenge data. Some of the variation in recidivism rates that may be associated with officer training and skills could be captured by the dataset’s location variables or by one or more of the variables that measure community supervision activities. Unfortunately, testing this possibility is beyond the scope of the Challenge.

**A. Location variables**

In addition to the commonly-evaluated individual characteristics, the NIJ encouraged Challenge participants to consider incorporating additional dynamic place-based factors. Such factors may be important in some types of prediction models in order to control for contextual factors that influence recidivism, particularly in models using data from widely varying geographic locations. As Wang and Han et al. (2020) found, “recidivism prediction models that are constructed using data from one location do not tend to perform as well when they are used to predict recidivism in another location” (p. 2).
However, to ensure human subject privacy and protection, the NIJ aggregated location information for individuals in the Challenge dataset into one of 25 broader geographic areas within Georgia. Each of these broader regions includes counties that tend to have similar geographic and socio-economic characteristics. Nevertheless, some variation in important place-based characteristics could exist within each aggregated region. Moreover, recidivism may occur in a location – including a different state - outside of the location in which an offender was released (Durose & Antenangeli, 2021).

Given the (necessarily) imprecise nature of an individual’s location information in the Challenge dataset and the possibility that recidivism rates may be influenced by an individual’s presence in another location, I chose not to add additional place-based factors to the existing Challenge dataset. Instead, I created an indicator variable measuring which of the 25 NIJ-defined geographic areas encompasses an individual’s address at prison release. This indicator variable could help control for place-based factors that influence recidivism but that are not directly measured in the dataset.

**B. Constructed variable: Age while incarcerated**

Prior research shows that an offender’s age influences the probability of recidivism in a variety of ways, most commonly by decreasing the recidivism probability as age increases, particularly in middle age and beyond. In addition to the existing age information in the Challenge dataset, I attempted to capture more nuanced age-related variation in recidivism rates that may result from an offender’s age, not at entry or release from prison, but rather, the age while incarcerated. For example, individuals who are incarcerated throughout their peak earning years may have difficulty finding and maintaining employment, particularly if they did not receive additional education or job training while in prison. Likewise, the impact on recidivism
of being incarcerated at younger ages may be different in direction or magnitude compared to being incarcerated at an older age.

Because no single variable in the Challenge dataset directly measures age while incarcerated, I constructed variables that may serve as proxy measures in the following manner: I created 28 dummy variables that measure both an individual’s age at parole release from prison (one of seven categories in the Challenge data) and the years spent in prison prior to release (one of four categories). These dummy variables could help differentiate between younger and older individuals who spent varying lengths of time in prison, particularly during their peak earning years, and thus, may help control for any variation in the associated recidivism rates.

**C. Constructed variable: Possible plea agreements**

The Challenge dataset provides sufficient information for creating variables for most of the individual-level factors shown by prior research to influence recidivism. However, one potentially predictive factor that has not been fully explored in previous studies is plea bargaining. Plea bargaining data could be helpful in predicting recidivism if plea agreements tend to result in criminal justice officials under-estimating an offender’s recidivism risk.

For example, an individual’s felony record may not accurately measure the association between the commission of felonies and the probability of recidivism if the number of felony convictions is relatively low as the result of plea bargaining. A criminal record that includes a high number of felony arrests but a relatively low number of felony convictions could be reflective of plea bargaining, with the discrepancy between numbers of arrests and convictions representing pleas to lesser charges or pleas involving dropped charges. Controlling for the possibility of plea agreements could improve recidivism prediction by reducing the number of false negatives, or cases in which the recidivism risk is under-estimated.
Much of the reason for the lack of research on the association between plea agreements and recidivism is likely related to a corresponding lack of local-level data on plea bargaining that is available to researchers. Efforts are underway to gather data for researchers on plea dispositions in criminal cases (Wilson Center, 2021), but these efforts are still new.

I attempted to create a rough proxy measure of the possible presence of plea agreements by constructing an additional variable in the Challenge dataset that measures the difference between the number of prior felony arrest episodes and the number of prior felony conviction episodes. I also created a similar variable with respect to misdemeanor offenses. My theory, albeit untestable within the confines of the Challenge, is that the greater the positive numerical difference between arrest and conviction episodes, the greater the possibility of the presence of plea agreements. Similarly, negative values for the numerical difference could indicate complex plea bargain exchanges among both misdemeanor and felony charges.

However, my use of these variables as an appropriate proxy for plea agreements is limited, primarily because I cannot ascertain from the Challenge dataset whether the number of arrests and convictions arose from the same criminal episode or from multiple and possibly unrelated episodes over time. If the arrests and convictions arise from one or more unrelated episodes, the difference between the two numbers may simply be random.

D. Missing data

Several continuous and categorical variables in the Challenge dataset are missing data in percentages ranging from about 1.5% to almost 22%. Multiple imputation techniques are often used to address problems associated with missing data in statistical models. However, attempting to use multiple imputation in the different types of models I examined seemed to add a potentially unmanageable level of complexity to the process.
Instead, in order to preserve as much information from the data as possible and to preclude the possibility of listwise deletion of observations with missing data, I flagged the missing values in the following manner: For categorical variables, I created an additional category that indicated whether the observation was missing a value for that particular variable. For continuous variables with missing data, I first generated categorical variables from the continuous measures by dividing them into quantiles, with each quantile serving as a category. I then added a final category indicating whether the observation was missing a value for the underlying continuous measure.

**E. Outcome variable**

The Challenge asked contestants to generate predicated probabilities of recidivism for individuals in the test datasets. Thus, an appropriate outcome measure to use in the training dataset is a dummy variable, coded as one if the individual recidivated and zero if not. Recidivism as a binary construct may be problematic in some contexts (see Klingele, 2019), but it is appropriate for the purposes of the Challenge.

**III. Models**

I chose to estimate two types of models for predicting recidivism: exploratory factor analysis and lasso regression.

**A. Exploratory factor analysis? No.**

Because we cannot directly measure an individual’s risk of recidivism, that risk – the propensity to recidivate – could be viewed as a latent construct comprised of many facets, some of which are measurable. Under this view, exploratory factor analysis (EFA) would be an appropriate approach for understanding the underlying structure of variables potentially related
to recidivism. EFA could also be useful in reducing a relatively large dataset such as the Challenge dataset into a more manageable group of variables (Rohe & Zeng, 2020).

Undeterred by previous research indicating that factor analysis may not be useful in predicting future offending (Kroner, Mills, & Reddon, 2005), I proceeded to explore the application of EFA. I theorized that the dataset’s recidivism measure may be part of a distinct cluster composed of the recidivism variable and other variables in the dataset. If so, that cluster - or factor – could provide information on the different variables that relate to an underlying propensity to recidivate. I therefore conducted a preliminary EFA on all the variables in the Challenge dataset, using maximum likelihood estimation. Although the EFA obtained a solution, the recidivism variable did not load on any of the resulting factors at standard criterion levels. So I set aside classical statistical methods and moved on to machine learning approaches, using the lasso estimator.

**B. Lasso for model selection? Yes.**

The lasso is an estimator of coefficients in a regression model that includes a penalty term to address over-fitting. The lasso is particularly useful in selecting relatively few out of many possible variables that affect an outcome. It may be used for both prediction and model selection (Drukker & Liu, 2019). When the goal is model selection, the lasso selects variables in one dataset, such as a training dataset, and then fits a model using those same variables in another dataset, such as a test dataset. Because the selected model may be used to make predictions in the test dataset, the lasso approach is particularly appropriate for the purposes of the Challenge. The lasso may also be used for inference, which presents an additional level of complexity that, thankfully, is beyond the scope of the Challenge.
Although the Challenge outcome measure of recidivism is dichotomous, I initially fit linear lasso models under the theory that a linear probability model (a linear regression model with a dichotomous dependent variable) might fit the data well. However, the non-linear models easily outperformed the linear models, so I abandoned the latter approach.

In using the lasso for non-linear model selection, I took the following steps: I first divided the Challenge’s training dataset still further into two samples for training and testing. I then fit both logit and probit models on the training sample using the lasso and the varying options that Stata 16 provides for selecting a parameter of the penalty function. In an effort to better control for any potentially unique variation associated with gender and race, I fit these lasso models not only using the full training sample but also using subsamples of Black and White parolees and male and female parolees. To determine which of these lasso models might perform better for out-of-sample prediction, I compared the deviance and deviance ratios among the models. None performed particularly well for out-of-sample predictions, possibly as a result of over-fitting in the training sample models or problems relating to data sparseness among the subsamples.

A closer examination of goodness-of-fit statistics led me to select separate models for the race and gender groups. I therefore used three different lasso models to estimate recidivism probabilities for Black male and White male parolees and for female parolees: I selected lasso probit (using cross-validation) models for Black male and White male parolees and the multistep adaptive lasso probit (also using cross-validation) for all female parolees. In general, with these data, probit models slightly outperformed their logit counterparts.

C. Interactions? No.

I also fit lasso models with multiple interactions terms. However, because many of the interaction terms had insufficient observations or sparse data, the models did not perform better
than models without the interactions. As Tollenaar & Van der Heijden (2013) noted, “If variables are suitably transformed and included in the model, there seems to be no additional predictive performance by searching for intricate interactions and/or non-linear relationships” (p. 582)

2 The authors softened their stance somewhat in a later study, allowing that some machine learning approaches may be helpful in determining whether important interactions may be missing from a given model (Tollenaar & Van der Heijden, 2019).

D. Lasso-selected variables

The half-joking warnings in Stata 16’s Lasso Reference Manual caution against placing importance on the lasso’s selection or lack of selection of any one variable. The lasso selects variables that either belong in the “true” model or that are correlated with variables that belong in the “true” model. Thus, the manual cautions, the researcher should not be overly concerned with the exact variables selected, since it is the group of variables selected as a whole, and their predictive power, that is important.

Despite these warnings, Stata 16 offers commands for displaying the lasso’s selected variables and their penalized coefficients. Tables 1-3 show these results for the lasso models I used in my recidivism forecasts. Table 1 shows the lasso-selected variables for the best-fitting models for Black male parolees, Table 2 for White male parolees, and Table 3 for female parolees, each estimated using the respective models described previously. Although my predicted probabilities for female parolees did not place among the top four entries in the Challenge, I include the information for females in Table 3 for comparison purposes. As the tables show, the best-fitting model for predicting recidivism among Black male parolees selected 103 variables, while only 60 variables were selected for White male parolees and 12 for females under the more parsimonious adaptive lasso approach.
The purpose of presenting these tables is not to impose on readers the task of examining the 175 variable descriptions and coefficients. Rather, the tables are useful in showing how the associations indicated by most of the variable coefficients comport with prior research on factors influencing recidivism risk. Positive coefficients indicate that the variable may be associated with an increase in the odds of recidivism, while negative coefficients indicate the opposite result. As the tables show, variables with positive coefficients tend to be those that measure younger age groups, lower education levels, less active employment histories, and greater prior involvement in serious crimes. Variables with negative coefficients tends to be those measuring older age groups, adherence to the conditions of probation or parole, and lower prior criminal involvement. Location also appears to be influential.

In truth, these associations do not reveal insights beyond what prior research has shown. Additional insights, however, may be possible from the results showing that the lasso selected many of the additional variables I constructed, indicating that some of these measures may be useful in predicting recidivism. However, because the purpose of the lasso regression used in this context is to build good models for prediction, no assumptions can be made about the statistical significance of any associations between the constructs I added and the recidivism outcome.

IV. Future considerations

The primary contribution, if any, of my Challenge entry to the research on predicting recidivism may be the finding indicating that additional data on plea bargaining could be helpful in accounting for some of the variation in recidivism rates. Plea bargaining may distort existing prediction models that use data on prior criminal involvement because plea agreements may lead to outcomes that do not accurately reflect the nature and level of that involvement. Data on plea agreements are difficult to obtain, however, though efforts are underway to address the lack of
data. In the meantime, proxy measures could be developed and tested for their usefulness, particularly proxy measures that are more refined than the ones I used here.

In addition, several of my measures of age during incarceration appear to show some promise in accounting for variation in recidivism outcomes. This information is probably already available within most criminal justice systems and could therefore be refined into an appropriate metric for testing, if not already in use.

Overall, the results from these lasso models suggest that future significant developments in recidivism prediction may not come from improvements in methodologies but rather from improvements in the breadth and reliability of the data on which the predictions are made. Unfortunately, systemic racial biases and a lack of transparency in criminal justice processes tend to produce unreliable data and faulty predictions that even complex statistical and machine learning methods cannot overcome.
Duddon Research Team information

1) Team member:
Shana M. Judge
Executive Director, Duddon Evidence to Policy Research
shanajudge@dudonresearch.org

2) Team consultant:
M. Muffy Blue
Owner, The Blue Law Firm
muffy@thebluelawfirm.com

3) Team consultant:
John Sideris
Professor of Research
University of Southern California, Chan Division of Occupational Science and Occupational Therapy
sideris@chan.usc.edu
References


Table 1. Variables and corresponding penalized coefficients for lasso probit model used to predict recidivism probabilities for Black male parolees.

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>% days employed - 4th decile</td>
<td>0.0793</td>
</tr>
<tr>
<td>age at prison release = 18-22</td>
<td>0.0527</td>
</tr>
<tr>
<td>no college education</td>
<td>0.0356</td>
</tr>
<tr>
<td># of prior felony arrests = 10 or more</td>
<td>0.0339</td>
</tr>
<tr>
<td># of prior misdemeanor arrests = 6 or more</td>
<td>0.0319</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 8</td>
<td>0.0305</td>
</tr>
<tr>
<td>% days employed - 3rd decile</td>
<td>0.0290</td>
</tr>
<tr>
<td>age at prison release = 23-27</td>
<td>0.0286</td>
</tr>
<tr>
<td># of prior misdemeanor convictions = 4 or more</td>
<td>0.0284</td>
</tr>
<tr>
<td># felony arrests - convict &lt; 0</td>
<td>0.0273</td>
</tr>
<tr>
<td># of prior arrests for property crimes = 4</td>
<td>0.0270</td>
</tr>
<tr>
<td># of prior arrests for property crimes = 5 or more</td>
<td>0.0258</td>
</tr>
<tr>
<td>first parole supervision risk score = 8</td>
<td>0.0235</td>
</tr>
<tr>
<td># unexcused absences from program = 1</td>
<td>0.0218</td>
</tr>
<tr>
<td>primary prison conviction = violent non-sex offense</td>
<td>0.0217</td>
</tr>
<tr>
<td># parole delinquency reports = 1</td>
<td>0.0201</td>
</tr>
<tr>
<td>not in (age=43-47 &amp; prison &gt; 3yr) category</td>
<td>0.0199</td>
</tr>
<tr>
<td>years in prison prior to release &lt; 1</td>
<td>0.0195</td>
</tr>
<tr>
<td># felony arrests - convict = 7</td>
<td>0.0190</td>
</tr>
<tr>
<td>first parole supervision risk score = 9</td>
<td>0.0186</td>
</tr>
<tr>
<td># felony arrests - convict = 2</td>
<td>0.0183</td>
</tr>
<tr>
<td># of prior arrests for a probation or parole violation = 5 or more</td>
<td>0.0182</td>
</tr>
<tr>
<td>not in (age=43-47 &amp; prison=2-3yr) category</td>
<td>0.0180</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 5</td>
<td>0.0176</td>
</tr>
<tr>
<td>first parole supervision risk score = 6</td>
<td>0.0175</td>
</tr>
<tr>
<td>age at prison release = 28-32</td>
<td>0.0150</td>
</tr>
<tr>
<td># of program attendances = 5</td>
<td>0.0140</td>
</tr>
<tr>
<td>not in (age=38-42 &amp; prison &gt; 3yr) category</td>
<td>0.0135</td>
</tr>
<tr>
<td>residence changes during parole = 3 or more</td>
<td>0.0132</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 18</td>
<td>0.0131</td>
</tr>
<tr>
<td># of prior arrests on violent charge = 2</td>
<td>0.0126</td>
</tr>
<tr>
<td>no prior felony arrest</td>
<td>0.0113</td>
</tr>
<tr>
<td># felony arrests - convict = 8</td>
<td>0.0112</td>
</tr>
<tr>
<td>jobs per year while on parole = 2</td>
<td>0.0112</td>
</tr>
<tr>
<td>jobs per year while on parole = 4</td>
<td>0.0095</td>
</tr>
<tr>
<td>avg # of days on parole btw drug tests = 7th quantile</td>
<td>0.0072</td>
</tr>
<tr>
<td># of prior arrests for property crimes = 2</td>
<td>0.0062</td>
</tr>
<tr>
<td># of prior arrests on a drug charge = 3</td>
<td>0.0060</td>
</tr>
<tr>
<td># misdemeanor arrests - convict = 1</td>
<td>0.0057</td>
</tr>
<tr>
<td># of parole delinquency reports = 3</td>
<td>0.0048</td>
</tr>
</tbody>
</table>
missing data on first parole supervision risk score 0.0040
avg # of days on parole btwn drug tests = 15th quantile 0.0039
# unexcused absences from program = 3 or more 0.0036
# of prior felony arrests = 6 0.0035
avg # of days on parole btwn drug tests = 16th quantile 0.0028
avg # of days on parole btwn drug tests = 20th quantile 0.0018
# of prior misdemeanor arrests = 5 0.0015
avg # of days on parole btwn drug tests = 3 0.0007
% drug tests positive for THC = 8th decile 0.0007
# of program attendances = 7 0.0007

Variables with negative coefficients (associated with decreased odds of recidivism)

- constant -0.9486
- missing data on jobs per year -0.3091
- no gang affiliation -0.0832
- age at prison release = 48 or older -0.0754
- no prior arrest for a probation or parole violation -0.0658
- missing data on days btwn drug tests -0.0484
- % days employed - 8th decile -0.0436
- no prior misdemeanor convictions -0.0359
- first parole supervision risk score = 2 -0.0316
- not in (age=23-27 & prison 1-2 yrs) category -0.0316
- % drug tests positive for meth - 1st decile -0.0304
- # felony arrests - convict = 1 -0.0302
- no violations for not following instructions -0.0286
- no prior felony arrests -0.0284
- no parole release condition for mental health or substance abuse program -0.0269
- no violations for moving without permission -0.0240
- no prior parole revocation -0.0213
- first parole supervision level not assigned as high -0.0205
- no prior misdemeanor convictions -0.0203
- % drug tests positive for THC = 1st decile -0.0196
- age at prison release = 43-47 -0.0195
- no dependents -0.0190
- not in NIJ combined PUMA group 20 -0.0189
- % days employed -7th decile -0.0180
- no prior arrest on a gun charge -0.0171
- no violations for electronic monitoring -0.0152
- residence changes during parole = 1 -0.0137
- first parole supervision risk score = 3 -0.0122
- not in NIJ combined PUMA group 17 -0.0102
- no prior arrests for property crimes -0.0099
- # of prior arrests on a drug charge = 1 -0.0093
- # of prior convictions on a drug charge = 0 -0.0091
- % drug tests positive for other drug - 10th decile -0.0091
avg # of days on parole btwn drug tests - 11th quantile  -0.0090
not in NIJ combined PUMA group 22  -0.0075
avg # of days on parole btwn drug tests = 10th quantile  -0.0068
not in (age=38-42 & prison=2-3yr) category  -0.0057
# of program attendances = 10 or more  -0.0043
not in NIJ combined PUMA group 11  -0.0043
primary prison conviction = violent sex offense  -0.0043
not in (age=18-22 & prison < 1yr) category  -0.0035
# of prior felony arrests = 5  -0.0030
not in NIJ combined PUMA group 10  -0.0028
not in (age=33-37 & prison < 1yr) category  -0.0027
# of prior arrests for property crimes = 4  -0.0027
# of prior felony arrests = 10 or more  -0.00217
# of program attendances = 5  -0.00195
# of program attendances = 5  -0.00191
# of program attendances = 5  -0.00186
not in NIJ combined PUMA group 5  -0.00178
primary prison conviction not a violent sex offense  -0.00178
first parole supervision risk score = 6  -0.00173
not in NIJ combined PUMA group 25  -0.00171
avg # of days on parole btwn drug tests - 15th quantile  -0.00166
% drug tests positive for meth - 10th decile  -0.00164
# of prior arrests for a probation or parole violation = 5 or more  -0.00155
# of prior arrests on a drug charge = 2  -0.00126
# of prior misdemeanor arrests = 6 or more  -0.00112
no college education  -0.00101
# of dependents = 1  -0.00093
% days employed - 3rd decile  -0.00092

Table 2. Variables and corresponding penalized coefficients for lasso probit model used to predict recidivism probabilities for White male parolees.

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables with positive coefficients (associated with increased odds of recidivism)</strong></td>
<td></td>
</tr>
<tr>
<td>% days employed - 4th decile</td>
<td>0.0818</td>
</tr>
<tr>
<td># felony arrests - convict = 7</td>
<td>0.0393</td>
</tr>
<tr>
<td>not in (age=43-47 &amp; prison &gt; 3yr) category</td>
<td>0.0306</td>
</tr>
<tr>
<td># of prior arrests for property crimes = 4</td>
<td>0.0297</td>
</tr>
<tr>
<td>first parole supervision risk score = 9</td>
<td>0.0283</td>
</tr>
<tr>
<td>years in prison prior to release &lt; 1</td>
<td>0.0276</td>
</tr>
<tr>
<td>not in (age=48+ &amp; prison 2-3 yrs) category</td>
<td>0.0233</td>
</tr>
<tr>
<td># of prior felony arrests = 10 or more</td>
<td>0.0217</td>
</tr>
<tr>
<td># of program attendances = 5</td>
<td>0.0195</td>
</tr>
<tr>
<td># of program attendances = 5</td>
<td>0.0191</td>
</tr>
<tr>
<td>age at prison release = 23-27</td>
<td>0.0186</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 5</td>
<td>0.0178</td>
</tr>
<tr>
<td>primary prison conviction not a violent sex offense</td>
<td>0.0178</td>
</tr>
<tr>
<td>first parole supervision risk score = 6</td>
<td>0.0173</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 25</td>
<td>0.0171</td>
</tr>
<tr>
<td>avg # of days on parole btwn drug tests - 15th quantile</td>
<td>0.0166</td>
</tr>
<tr>
<td>% drug tests positive for meth - 10th decile</td>
<td>0.0164</td>
</tr>
<tr>
<td># of prior arrests for a probation or parole violation = 5 or more</td>
<td>0.0155</td>
</tr>
<tr>
<td># of prior arrests on a drug charge = 2</td>
<td>0.0126</td>
</tr>
<tr>
<td># of prior misdemeanor arrests = 6 or more</td>
<td>0.0112</td>
</tr>
<tr>
<td>no college education</td>
<td>0.0101</td>
</tr>
<tr>
<td># of dependents = 1</td>
<td>0.0093</td>
</tr>
<tr>
<td>% days employed - 3rd decile</td>
<td>0.0092</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td># of prior arrests on a drug charge = 3</td>
<td>0.0062</td>
</tr>
<tr>
<td>first parole supervision risk score = 8</td>
<td>0.0060</td>
</tr>
<tr>
<td># felony arrests - convict = 2</td>
<td>0.0056</td>
</tr>
<tr>
<td># misdemeanor arrests - convict = 3</td>
<td>0.0046</td>
</tr>
<tr>
<td>avg # of days on parole btwn drug tests - 8th quantile</td>
<td>0.0046</td>
</tr>
<tr>
<td>not in NIJ combined PUMA group 18</td>
<td>0.0044</td>
</tr>
<tr>
<td># of program attendances = 8</td>
<td>0.0040</td>
</tr>
<tr>
<td># of parole delinquency reports = 3</td>
<td>0.0039</td>
</tr>
<tr>
<td># unexcused absences from program = 1</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

**Variables with negative coefficients (associated with decreased odds of recidivism)**

- constant: -0.9469
- missing data on jobs per year: -0.1963
- no prior arrest for a probation or parole violation: -0.0929
- age at prison release = 48 or older: -0.0797
- no gang affiliation: -0.0780
- no prior misdemeanor arrests: -0.0462
- no violations for moving without permission: -0.0455
- no prior misdemeanor convictions: -0.0429
- no violations for not following instructions: -0.0381
- primary prison conviction = violent sex offense: -0.0364
- # felony arrests - convict = 1: -0.0261
- no unexcused absences from program: -0.0226
- age at prison release = 43-48: -0.0219
- % days employed - 8th decile: -0.0202
- jobs per year while on parole = 5: -0.0175
- not in (age=28-32 & prison < 1yr) category: -0.0157
- avg # of days on parole btwn drug tests - 11th quantile: -0.0141
- # of prior felony arrests = 1: -0.0118
- % drug tests positive for meth - 1st decile: -0.0106
- residence changes during parole = 1: -0.0106
- # of prior felony arrests = 7: -0.0105
- first parole supervision risk score = 2: -0.0101
- % drug tests positive for THC = 1st decile: -0.0080
- no electronic monitoring violations: -0.0065
- # of prior arrests on a drug charge = 1: -0.0036
- not in NIJ combined PUMA group 11: -0.0032
- not in NIJ combined PUMA group 10: -0.0031
- years in prison prior to release > 3: -0.0029
- no parole release condition for mental health or substance abuse program: -0.0012
Table 3. Variables and corresponding penalized coefficients for adaptive lasso probit model used to predict recidivism probabilities for female parolees.

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables with positive coefficients (associated with increased odds of recidivism)</strong></td>
<td></td>
</tr>
<tr>
<td># of program attendances = 5</td>
<td>0.1096</td>
</tr>
<tr>
<td># of prior felony arrests = 10 or more</td>
<td>0.0875</td>
</tr>
<tr>
<td># of parole delinquency reports = 4 or more</td>
<td>0.0807</td>
</tr>
<tr>
<td># of prior arrests for a probation or parole violation = 5 or more</td>
<td>0.0395</td>
</tr>
<tr>
<td><strong>Variables with negative coefficients (associated with decreased odds of recidivism)</strong></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-1.2771</td>
</tr>
<tr>
<td>missing data on jobs per year</td>
<td>-0.4320</td>
</tr>
<tr>
<td># misdemeanor arrests - convict = 0</td>
<td>-0.1647</td>
</tr>
<tr>
<td>years in prison prior to release &gt; 2-3</td>
<td>-0.1333</td>
</tr>
<tr>
<td>first parole supervision level = standard</td>
<td>-0.1140</td>
</tr>
<tr>
<td>no prior misdemeanor convictions</td>
<td>-0.0934</td>
</tr>
<tr>
<td>no prior arrests for property crimes</td>
<td>-0.0840</td>
</tr>
<tr>
<td>no prior misdemeanor arrests</td>
<td>-0.0838</td>
</tr>
<tr>
<td>first parole supervision level not assigned as high</td>
<td>-0.0485</td>
</tr>
</tbody>
</table>
Appendix: Answers to specific questions from the NIJ Challenge Team.

- Were variables added to the data set? If so, detail the variables.

  No. See the rationale in Section II.A of the report.

- What variables were constructed? How were the variables constructed?

  Yes, I constructed other variables. See the discussion in Sections II.B and II.C of the report.

- Which variables were statistically significant?

  See the discussion in Section III.D of the report.

- What variables were not statistically significant? How was this handled? For example, were they dropped from the overall model?

  See the discussion in Section III.D of the report.

- What type of model was used?

  See the discussion in Section III of the report.

- Did you try other models? Were they close in performance? Not at all close?

  Yes, I tried other models. See the discussion in Section III.A of the report.

- What other evaluation metrics should have been considered/used for this Challenge? For example, using false negatives in the penalty function.

  I appreciated both the straightforward nature of the Challenge’s existing evaluation metric and the logic behind using false positives in the penalty function. However, I would be interested in hearing arguments supporting the use of false negatives.

- Did the 0.5 threshold affect anything? Would your team recommend a different threshold?

  No, the threshold did not affect my analyses. See the discussion in Section III.D of the report.

- Did the fact that the fairness penalty only considered false positives affect your submission?

  No. I considered examining the false positive results in my predictions prior to submitting my entry. But this examination would have involved a manual review of each false positive observation, which runs counter to the Challenge goal of improving recidivism prediction models. Also, I ran out of time.
• Are there practical/applied findings that could help the field based on your work? If yes, what are they?

Maybe. I hope. See the discussion in Section IV of the report.

• What should NIJ have considered changing (other than metrics) to improve this Challenge?

My consultant colleagues and I thought the Challenge was very well-designed – my colleagues commented to me on this point several times. I appreciated the focus on a discrete, non-subjective outcome and the opportunity it provided to individuals and small businesses like mine to test our research mettle.

I will note that I did not receive information on the Challenge until the relative last minute and thus, was only able to submit an entry for the third year period. I subscribe to a range of DoJ email lists (news from BJS, BJA, etc.) and follow NIJ on Twitter. But I missed announcements relating to the Challenge and, in fact, only heard about it through an email notification from the Association of Public Policy Analysis and Management. Thus, I was wondering whether notifications regarding the Challenge were widely disseminated.

Also, as I noted in an email to the Challenge team, it would have been helpful to know how my Brier scores compared to other scores in categories in which I did not place in the top four. I could use this information to further identify weaknesses in my analysis and better understand why my approach performed relatively well in one category but not in others. Could future Challenges post additional top scores without identifying the associated teams?

• For future Challenges, what should NIJ consider changing to improve Challenges? For example, more/less time, different topic, or data issues (missing data)?

By its very nature, academic and scholarly work has to build on previous academic and scholarly research. But that process tends to inhibit unconventional approaches and inventiveness, and often stifles research with practical implications. That’s why the idea of allowing anyone to enter the Challenge, even high school students and those who work outside of academia, is so appealing. I hope NIJ continues to provide these types of opportunities that allow interested people from across the country – including those who work in small businesses and who are not directly affiliated with an academic institution – to participate.

I think future challenges should maintain the category for high school students and consider ways to encourage students to participate (I realize that NIJ is already working on this issue). Could future challenges somehow encourage more established researchers to partner with high school classes or individual students, with a category for teams comprised in that manner?