

Online Appendix: Algorithmic Risk Assessment in the Hands of Humans

Megan T. Stevenson & Jennifer L. Doleac

August 25, 2023

A

A.1 Use of risk assessment at sentencing

The following states use some form of risk assessment at sentencing:

- AL: ALA. CODE §12-25-32 (1975)(Definition) and §12-25-33(6)(1975)(guidelines for use).
- CO: COLO. REV. STAT. Title 16-11-102 (1)(b)(II)(specific to sex offenders).
- ID, NE, OR: Jennifer K. Elek, Roger K. Warren, & Pamela M. Casey, Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions. National Center for State Courts (2015), https://www.ncsc.org/data/assets/pdf_file/0016/26251/final-pew-report-updated-10-5-15.pdf.
- HI, IL: Tammy Howell, LSI-R, LS/RNR and LS/CMI Documentation, Public Safety Division, Multi-Health Systems, Inc., <https://www.scstatehouse.gov/Archives/CitizensInterestPage/SentencingReformCommission/Miscellaneous/exhibittoDanford100809presentationsLSIdocumentation.doc>.
- KS: KS R. SUP. OFFENDERS RULE 1802, Court Services Officer Assessment of Offenders (July 9,2021).
- CA, FL, WI: Danielle Kehl, Priscilla Guo, & Samuel Kessler, Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing (2017), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>.
- PA: 42 PA. CONS. STAT. §2154.7. See also Ryan S. Meyers, Sentencing Policy Specialist, Pennsylvania Commission on Sentencing, Introducing Risk Assessment at Sentencing in Pennsylvania, presentation at the Criminal Justice Advisory Board Annual Conference (Apr. 4, 2018), <https://www.pccd.pa.gov/training/Documents/Conferences%20and%20Training/Sentence%20Risk%20Assessment.pdf>.
- ND: N. D. CENT. CODE §12.1-32-15 (12) and N.D. R. CRIM. P. 32(c)(4)(C).
- NY: OFFICE OF PROBATION AND CORRECTIONAL ALTERNATIVES, NEW YORK CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS (NYCOMPAS) RISK AND NEEDS ASSESSMENT INSTRUMENT: PRACTITIONER GUIDANCE FOR PROBATION AND COMMUNITY CORRECTIONS

AGENCIES (2015) <http://www.criminaljustice.ny.gov/opca/pdfs/2015-5-NYCOMPAS-Guidance-August-4-2015.pdf>(Use of Risk and Needs Assessment in sentencing is not mandated per se, but is encouraged for a variety of sentencing procedures).

- IA – While not mandated legislatively, the use of risk assessments in sentencing has been permitted by the Iowa Supreme Court. *State v. Headley*, 926 N.W.2d 545, 551 (Iowa 2019)(holding that risk assessments “provide pertinent information that a sentencing judge may consider.”).

- AZ, IN, KY, MI, MO, OH, OK, UT, VA, WA, WV: Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 *Stan. L. Rev.* 803, 872 n.11 (2014).

- VT: STATE OF VERMONT AGENCY OF HUMAN SERVICES DEPARTMENT OF CORRECTIONS, PRE-SENTENCE INVESTIGATION (PSI) REPORTS - DIRECTIVE #342.01 (Dec. 1, 2021), <https://acrobat.adobe.com/link/track?uri=urn%3Aaaid%3Ascds%3AUS%3Ac5cd33aa-f75e-4c13-b447-71c44f3e0332#pageNum=1>.

There are 6 more states in which at least one county either uses or permits the use of risk assessments in sentencing:

- NC: Tammy Howell, LSI-R, LS/RNR and LS/CMI Documentation, Public Safety Division, Multi-Health Systems, Inc., <https://www.scstatehouse.gov/Archives/CitizensInterestPage/SentencingReformCommission/Miscellaneous/exhibittoDanford100809presentationsLSIdocumentation.doc>.

- LA: LA STAT. ANN. §15:326(A).

- MN: MINNESOTA DEPARTMENT OF CORRECTIONS, *STUDY OF EVIDENCE-BASED PRACTICES IN MINNESOTA: 2011 REPORT TO THE LEGISLATURE 3-4* (2011), <https://mn.gov/doc/assets/12-10EBPreport`tcm1089-271698.pdf>.

- ME: Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 *Stan. L. Rev.* 803, 872 n.11 (2014).

- TX: Jennifer K. Elek, Roger K. Warren, & Pamela M. Casey, *Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions*. *National Center for State Courts* 35, 41-43, A-132 (2015), <https://www.ncsc.org/`data/assets/pdf file/0016/26251/final->

[pew-report-updated-10-5-15.pdf](https://www.ncsc.org/`data/assets/pdf file/0016/26251/final-pew-report-updated-10-5-15.pdf). See also Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 *Stan. L. Rev.* 803, 872 n.11 (2014).

- AR: Roger K. Warren, *State Judicial Branch Leadership In Sentencing and Corrections Reforms*, Center for Sentencing Initiatives, Research Division, *National Center for State Courts* (2013), <https://www.ncsc.org/`data/assets/pdf file/0018/26217/state-judicial-branch-leadership-brief-csi.pdf>.

A.2 Benchmarking Virginia’s risk assessment against a random forest risk assessment

How accurate is Virginia’s nonviolent risk assessment? A common method of evaluating predictive accuracy is to measure the area under the ROC curve (AUC). The AUC, which

ranges from 0 to 1, is the probability that a randomly selected person who recidivates has a higher risk score than a randomly selected person who did not recidivate. Statistically, the AUC is equivalent to the Wilcoxon rank-sum test (Mason and Graham, 2002).

We begin by building our own risk assessment as an accuracy benchmark. Like the real risk assessment, our alternative risk score is trained to predict the likelihood of a new felony conviction within three years of release. Our risk score uses a random forest algorithm trained on all defendants who received a risk score in fiscal years 2003-2013. We terminate in 2013 because the next year Virginia switched to a different risk assessment. The training data includes all of the inputs that were used to calculate Virginia’s risk score, as well as a wide variety of other data points, including age in years (Virginia’s risk score only includes age brackets), the total score for each of the three sentence-guidelines worksheets, and all the criminal history variables from the first page of the worksheets: additional current offenses, prior against-person felonies, prior drug felonies, prior property felonies, prior convictions, prior incarcerations, recent legal restraints, prior juvenile incarcerations, prior misdemeanor convictions, and prior weapons offenses. We also include race as a predictor, both to maximize accuracy and because we wanted to use a race-weighted recidivism predictor as a control in Subsection IV.B.

We use the `randomForest` package in R, with 1000 trees and node size of 10. We generate out-of-bag predictions (i.e. the prediction for individual i comes from a model in which i ’s data was not used) for all defendants in the sample. We then calculate the AUC for our self-built risk assessment. Given the expanded set of inputs, a prediction algorithm that can take advantage of nonlinear and interacted effects, and the fact that we are testing predictive accuracy within the training sample, we consider this to be close to an upper bound on the AUC within this context.

The AUC for our self-built predictive tool is 0.591 [0.581-0.609].¹ The AUC for Virginia’s risk assessment is 0.579 [0.569-0.589]. The two numbers are qualitatively quite similar with overlapping confidence intervals. Virginia’s risk assessment appears sound.

¹The reported AUC for other risk assessments is frequently higher than this. We expect that this is because the Virginia risk assessment applies only to a relatively homogenous sample: those convicted of low-level nonviolent offenses. The discriminatory potential of algorithms depends a lot on the heterogeneity of the underlying sample. By focusing only on similar cases, you miss many of the ‘easy’ comparisons that can increase an AUC. (E.g. someone released at age 70 after 40 years of incarceration on a murder charge will have a lower recidivism risk than a 19-year-old convicted on a drug charge.)

Table A1: Regression discontinuity around the low-risk cut-off– robustness to different transforms of the sentence length

	Sentence (arcsinh)	Sentence (top coded)
RD_Estimate	0.235 (0.111)	1.750 (0.809)
Mean	2.291	10.745
N	9404	9404

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table uses regression discontinuity to test for changes around the low-risk cutoff. The risk score is the running variable. The outcomes are the sentence length with an arcsinh transform and the sentence length top coded at the 99th percentile (84 months). Means of the dependent variable are shown at the bottom. Optimal bandwidths are used. The sample includes all individuals who received a risk score in fiscal years 2003-2004. Covariates are included in all regressions.

Table A2: Event study test for increased correlation between sentence and predicted risk score – robustness to different transforms of the sentence length

	Sentence (arcsinh) (1)	Sentence (top coded) (2)
Predicted risk x post	0.238 (0.063)	1.441 (0.574)
Predicted risk score	0.808 (0.113)	5.928 (1.051)
N	24040.000	24040.000
R ²	0.351	0.482

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows estimates of the extent to which the correlation between sentencing and defendant risk increased after risk assessment adoption. The outcomes are the sentence length with an arcsinh transform and the sentence top coded at the 99th percentile (84 months). Standard errors are clustered at the judge level. The sample includes all risk-assessment-eligible defendants sentenced between 2001-2004. Covariates are included in all regressions.

Table A3: Testing for changes in case composition before/after risk assessment adoption

Panel A: Testing for pre-/post changes in case composition for eligible cases (event study)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Prison	Female	Black	Young	G.L.	Prior	Risk	Drug	Larc.	Fraud	Freq.
Post	-0.0119 (0.00816)	0.0113 (0.00721)	-0.0255 (0.0202)	-0.0159 (0.00571)	-0.0619 (0.284)	0.000531 (0.00497)	-0.00732 (0.0127)	0.00314 (0.00427)	0.00869 (0.00851)	-0.00137 (0.00804)	9.346 (5.513)
Obs.	24040	22449	22445	23730	24040	24040	24040	24040	24040	24040	208
Mean	0.558	0.237	0.597	0.268	12.46	0.134	0.532	0.0982	0.292	0.176	147.9

Panel B: Testing for differential changes in case composition (difference-in-differences)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prison	Female	Black	Young	G.L.	Prior	Freq.
Eligible x post	-0.0168 (0.0109)	0.0114 (0.00751)	-0.00159 (0.0109)	0.000267 (0.00702)	0.337 (0.857)	-0.000148 (0.00528)	-1.033 (12.93)
Obs.	64675	60194	60186	63853	64675	64675	349
Mean	0.558	0.237	0.597	0.268	12.46	0.134	224.0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: Panel A shows event study tests for changes in case characteristics pre/post risk assessment adoption. The samples include all risk assessment eligible individuals sentenced in fiscal years 2001-2004, with the exception of those for whom the outcome variable is missing. The estimates are from regressions of each covariate on an indicator for being post-risk-assessment adoption. Column 11 tests for changes in the weekly number of risk-assessment-eligible cases. The data is collapsed to the weekly level and we regress the weekly number of cases on a post-risk-assessment indicator. Panel B tests for differential changes in case characteristics after risk assessment adoption. The samples include all individuals sentenced in fiscal years 2001-2004, with the exception of those for whom the outcome variable is missing. The coefficients come from a regression of each covariate on an indicator for being convicted of a risk-assessment-eligible offense, an indicator for being post-risk-assessment adoption, and the interaction of the two. Column 7 tests for differential changes in the weekly number of cases. The data is collapsed to the weekly level for eligible and ineligible cases and the same difference-in-differences specification is run. With the exception of the tests for changes in case frequency, standard errors are clustered at the judge level. G.L. stands for the guidelines-recommended sentence.

Table A4: Risk assessment’s net impact on sentencing and recidivism – clustering on judicial circuit

	Pr(Incarceration)		Log sentence		Recidivism (3yr)	
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible x post	0.00449 (0.0101)	0.0100 (0.00864)	0.0198 (0.0334)	0.0410 (0.0225)	0.00774 (0.00604)	0.00801 (0.00508)
Observations	64594	64594	64594	64594	64594	64594
R ²	0.0465	0.432	0.00280	0.625	0.00588	0.0438
Mean DV	0.791	0.791	1.399	1.399	0.192	0.192
Covariates	N	Y	N	Y	N	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents difference-in-differences estimates in which outcomes are compared across eligible/ineligible cases before/after risk assessment is adopted. The difference between this table and the main difference-in-differences results are that standard errors are clustered at the judicial circuit level. The outcomes are the probability of incarceration, the log sentence (bottom coded at 2 weeks) and the likelihood of being convicted of a new felony within 3 years. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted in fiscal years 2001-2004.

Table A5: Robustness tests for main results: varying time windows

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pr(incarceration)		Months sentence		Recidivism(3yr)				
	2yr	6yr	8yr	2yr	6yr	8yr	2yr	6yr	8yr
Eligible x post	0.00918 (0.00891)	0.00438 (0.00761)	0.00295 (0.00750)	0.0199 (0.0246)	0.0311 (0.0198)	0.0291 (0.0181)	0.00722 (0.00809)	0.0106 (0.00496)	0.00970 (0.00461)
Observations	34126	94362	125676	34126	94362	125676	34126	94362	125676
R ²	0.425	0.440	0.438	0.621	0.630	0.631	0.0482	0.0437	0.0416
Mean DV	0.809	0.809	0.811	1.519	1.496	1.520	0.171	0.181	0.178

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the main difference-in-differences results shown in Table ???. The specification is the same as that shown in Equation ?? but instead of using four years of data we use two, six, and eight years, centered around the date of risk assessment adoption. The outcomes are the probability of incarceration, the log sentence (bottom coded at 2 weeks) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. Covariates are included in all regressions.

Table A6: Robustness tests for main results: varying control groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Only DLF offenses		No DLF-ineligible offenses		No violent offenses				
	Pr(inc.)	Sent.	Recid.	Pr(inc.)	Sent.	Recid.	Pr(inc.)	Sent.	Recid.
Eligible x post	0.0143 (0.0118)	0.0600 (0.0275)	0.0129 (0.00682)	0.00501 (0.00727)	0.0205 (0.0264)	0.00126 (0.00609)	0.00800 (0.00961)	0.0295 (0.0247)	0.00944 (0.00588)
Observations	43977	43977	43977	44738	44738	44738	57545	57545	57545
R ²	0.432	0.584	0.0432	0.171	0.502	0.0523	0.429	0.585	0.0400
Mean DV	0.808	1.498	0.174	0.808	1.498	0.174	0.808	1.498	0.174

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the main difference-in-differences results shown in Table ???. The specification is the same as that shown in Equation ?? but our control groups vary. The control group in Columns 1-3 consists of those convicted of a drug, larceny, or fraud offense but who were risk-assessment ineligible due to one of the various restrictions. The control group in Columns 4-6 consists of all defendants who were risk-assessment-ineligible EXCEPT for those convicted of drug, larceny, and fraud offenses. In Columns 7-9, the control group consists of anyone convicted of some sort of nonviolent offense. The outcomes are the probability of incarceration, the log sentence (bottom coded at 2 weeks) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted of a felony within fiscal years 2001-2004. Covariates are included in all regressions.

Table A7: Robustness tests: difference-in-differences with various simulated sentences as the outcome

	Log sentence				Pr(incarceration)			
	Alt 1	Alt 2	Alt 3	Alt 4	Alt 1	Alt 2	Alt 3	Alt 4
Eligible x post	-0.509 (0.0317)	-0.127 (0.0232)	-0.217 (0.0287)	-0.709 (0.0365)	-0.193 (0.0108)	-0.0728 (0.00846)	0.0473 (0.00999)	-0.0730 (0.0122)
Observations	64675	64675	61572	61572	64675	64675	64675	64675
R ²	0.599	0.628	0.655	0.630	0.403	0.423	0.452	0.412
Mean DV	1.399	1.399	1.399	1.399	0.761	0.761	0.761	0.761

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table provides robustness tests to the simulations of sentencing by algorithm. It shows estimates from difference-in-differences specifications in which the outcomes are a variety of alternative sentence simulations. In Alt 1, all low-risk defendants are assigned a sentence of length 0. In Alt 2, low-risk defendants whose guidelines-recommended sentence is jail are assigned a sentence of length 0 and low-risk defendants with the guidelines-recommended sentence of prison are assigned a 12 month jail sentence. Alt 3 is the same as Alt 2, except high-risk defendants get the average sentence given by judges to people with the same risk score and guidelines-recommended sentence. Alt 4, is the same as Alt 3, except low-risk individuals with a prison sentence get assigned a sentence of length 0. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted of a felony within fiscal years 2001-2004. Covariates are included in all regressions.

Table A8: Risk assessment's net impact on the sentence length using alternative transforms

	Log sentence (1)	Sentence (top coded) (2)
Eligible x post	0.0406 (0.0229)	0.197 (0.211)
Observations	64675	64675
R ²	0.625	0.684
Mean DV	2.117	9.338
Covariates	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents difference-in-differences estimates in which outcomes are compared across eligible/ineligible cases before/after risk assessment is adopted. The outcomes are the sentence length with an arcsinh transform and the sentence in months top coded at the 99th percentile of positive sentences for eligible defendants (84 months). The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted in fiscal years 2001-2004. Covariates are included in all regressions.

Table A9: What factors predict deviation from the risk score?

	(1) Log sentence	(2) Log sentence	(3) Log sentence	(4) Log sentence
Black	0.240**** (0.031)	0.141**** (0.026)	0.115**** (0.029)	0.118**** (0.029)
Married	-0.088** (0.039)	-0.061* (0.033)	-0.075** (0.033)	-0.079** (0.034)
Female	-0.182**** (0.039)	-0.120**** (0.034)	-0.131**** (0.034)	-0.132**** (0.034)
Unemployed	-0.040 (0.038)	-0.093*** (0.032)	-0.072** (0.033)	-0.073** (0.033)
Young	-0.294**** (0.038)	-0.269**** (0.032)	-0.262**** (0.032)	-0.258**** (0.032)
R^2	0.071	0.354	0.392	0.392
Risk score	Y	Y	Y	Y
Guidelines-recommended sentence	N	Y	Y	Y
Judge fixed effects	N	N	Y	Y
Alternative risk score	N	N	N	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows estimates from regressions of log sentence (bottom coded at 2 weeks) on demographic factors. The first column controls only for the risk score. The second column adds controls for the exact guidelines-recommended sentence; the third adds judge fixed effects; and the fourth adds controls for an alternative risk score that includes race as a predictor. The sample includes all individuals who received a risk score in fiscal years 2003-2004, with the exception of those missing the demographic information listed.

Table A10: Robustness tests for racial impact triple-differences results: varying time windows

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(incarceration)			Log sentence		
	2yr	6yr	7yr	2yr	6yr	7yr
Eligible x post x Black	-0.0279 (0.0188)	-0.00183 (0.0143)	-0.00749 (0.0137)	-0.0276 (0.0590)	0.0565 (0.0405)	0.0415 (0.0381)
Observations	31824	79845	96436	31824	79845	96436
R ²	0.429	0.438	0.437	0.623	0.627	0.629
Mean DV	0.819	0.819	0.819	1.552	1.552	1.552

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the triple-differences racial impact results shown in Table ???. The specification is the same as that shown in Equation ?? but instead of using four years of data we use two, six, and seven years, centered around the date of risk assessment adoption. (Race data is not available for fiscal year 1999.) The outcomes are the probability of incarceration and the log sentence (bottom coded at 2 weeks). Standard errors are clustered at the judge level. The mean dependent variables for eligible cases with Black defendants during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants sentenced within the fiscal year range as shown except for those from Alexandria and Fairfax, where race information is not available. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being Black, as well as interactions between all of the above.

Table A11: Robustness tests for age disparity triple-differences results: varying time windows

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(incarceration)			Log sentence		
	2yr	6yr	8yr	2yr	6yr	8yr
Eligible x post x young	0.0413 (0.0212)	0.0467 (0.0141)	0.0492 (0.0123)	0.0952 (0.0651)	0.134 (0.0465)	0.150 (0.0418)
Observations	33710	93283	124009	33710	93283	124009
R ²	0.425	0.440	0.438	0.621	0.630	0.631
Mean DV	0.731	0.731	0.731	1.264	1.264	1.264

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the triple-differences age disparity estimates shown in Table ???. The specification is the same as that shown in Equation ?? but instead of using four years of data we use two, six, and eight years, centered around the date of risk assessment adoption. The outcomes are the probability of incarceration and the log sentence (bottom coded at 2 weeks). Standard errors are clustered at the judge level. The mean dependent variables for eligible cases with young defendants during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants sentenced within the fiscal year ranges as shown. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being young, as well as interactions between all of the above.

Table A12: Robustness tests for triple-differences racial impact results: varying control groups

	(1)	(2)	(3)	(4)	(5)	(6)
	Only DLF offenses Pr(inc.)	Only DLF offenses Sent.	No DLF-ineligible offenses Pr(inc.)	No DLF-ineligible offenses Sent.	No violent offenses Pr(inc.)	No violent offenses Sent.
Eligible x post x Black	-0.0123 (0.0193)	0.0331 (0.0493)	-0.0120 (0.0168)	0.0525 (0.0548)	-0.0178 (0.0173)	0.0245 (0.0469)
Observations	40673	40673	41958	41958	53503	53503
R ²	0.434	0.584	0.175	0.505	0.432	0.586
Mean DV	0.819	1.552	0.819	1.552	0.819	1.552

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the triple-differences racial impact estimates shown in Table ??.

The specification is the same as that shown in Equation ?? but our control groups vary. The control group in Columns 1-2 consists of those convicted of a drug, larceny, or fraud offense but who were risk-assessment ineligible due to one of the various restrictions. The control group in Columns 3-4 consists of all defendants who were risk-assessment-ineligible EXCEPT for those convicted of drug, larceny, and fraud offenses. In Columns 5-6, the control group consists of anyone convicted of some sort of nonviolent offense. The outcomes are the probability of incarceration and the log sentence (bottom coded at 2 weeks). Standard errors are clustered at the judge level. The mean dependent variables for eligible cases for Black defendants during the pre-risk assessment period are shown in the bottom row. The sample covers fiscal years 2003-2004; data from Alexandria and Fairfax counties are omitted because race information was not available. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being Black, as well as interactions between all of the above.

Table A13: Robustness tests for triple-differences age disparity results: varying control groups

	(1)	(2)	(3)	(4)	(5)	(6)
	Only DLF offenses Pr(inc.)	Only DLF offenses Sent.	No DLF-ineligible offenses Pr(inc.)	No DLF-ineligible offenses Sent.	No violent offenses Pr(inc.)	No violent offenses Sent.
Eligible x post x young	0.0458 (0.0189)	0.110 (0.0541)	0.0388 (0.0184)	0.119 (0.0607)	0.0400 (0.0180)	0.102 (0.0525)
Observations	43413	43413	44170	44170	56823	56823
R ²	0.432	0.584	0.171	0.501	0.429	0.585
Mean DV	0.731	1.264	0.731	1.264	0.731	1.264

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the triple-differences age disparity estimates shown in Table ???. The specification is the same as that shown in Equation ?? but our control groups vary. The control group in Columns 1-2 consists of those convicted of a drug, larceny, or fraud offense but who were risk-assessment ineligible due to one of the various restrictions. The control group in Columns 3-4 consists of all defendants who were risk-assessment-ineligible EXCEPT for those convicted of drug, larceny, and fraud offenses. In Columns 5-6, the control group consists of anyone convicted of some sort of nonviolent offense. The outcomes are the probability of incarceration and the sentence length (with an arcsinh transform.) Standard errors are clustered at the judge level. The mean dependent variables for eligible cases for young defendants during the pre-risk assessment period are shown in the bottom row. The samples cover fiscal years 2003-2004. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being young, as well as interactions between all of the above.

Table A14: Risk assessment’s impact by age (under 30): **simulated** versus **actual**

	<u>Pr(Incarceration)</u>		<u>Log sentence</u>	
	Simulated	Actual	Simulated	Actual
	(1)	(2)	(3)	(4)
Post x eligible x under 30	0.0937 (0.0148)	0.0350 (0.0137)	0.286 (0.0424)	0.108 (0.0424)
Observations	63853	63853	63853	63853
R ²	0.423	0.432	0.627	0.625
Mean DV, under 30	0.741	0.768	1.382	1.382
Covariates	Y	Y	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This figure shows triple difference estimates of risk assessment’s impact by age. Here, young age is defined as being under 30 (it’s defined as being under 23 in the main specification). The outcomes are the probability of incarceration and the log sentence (bottom coded at 2 weeks), as well as simulations thereof. Simulated sentences (odd-numbered columns) entail automatic diversion for low-risk individuals after risk assessment was adopted. Standard errors are clustered at the judge level. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all individuals sentenced in fiscal years 2001-2004, with the exception of those for whom age information is missing. Covariates are included in all regressions.

Table A15: Robustness tests for simulations: testing for race disparities with various simulated sentences as the outcome

	Log sentence				Pr(incarceration)			
	Alt 1	Alt 2	Alt 3	Alt 4	Alt 1	Alt 2	Alt 3	Alt 4
Eligible x post x Black	0.165 (0.0558)	0.0516 (0.0460)	-0.0735 (0.0457)	0.0546 (0.0580)	0.0338 (0.0181)	-0.00182 (0.0161)	-0.0233 (0.0165)	0.0124 (0.0191)
Observations	60186	60186	57324	57324	60186	60186	60186	60186
R ²	0.602	0.630	0.656	0.632	0.406	0.425	0.454	0.415
Mean DV	1.552	1.552	1.552	1.552	0.798	0.798	0.798	0.798

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table provides robustness tests to the simulations of sentencing by algorithm. It shows estimates from triple-differences racial disparities specifications in which the outcomes are a variety of alternative sentence simulations. In Alt 1, all low-risk defendants are assigned a sentence of length 0. In Alt 2, low-risk defendants whose guidelines-recommended sentence is jail are assigned a sentence of length 0 and low-risk defendants with the guidelines-recommended sentence of prison are assigned a 12 month jail sentence. Alt 3 is the same as Alt 2, except high-risk defendants get the average sentence given by judges to people with the same risk score and guidelines-recommended sentence. Alt 4, is the same as Alt 3, except low-risk individuals with a prison sentence get assigned a sentence of length 0. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted of a felony within fiscal years 2001-2004, except for those with missing race information. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being Black, as well as interactions between all of the above.

Table A16: Robustness tests for simulations: testing for age disparities with various simulated sentences as the outcome

	Log sentence				Pr(incarceration)			
	Alt 1	Alt 2	Alt 3	Alt 4	Alt 1	Alt 2	Alt 3	Alt 4
Eligible x post x young	0.348 (0.0525)	0.207 (0.0529)	0.351 (0.0515)	0.552 (0.0525)	0.121 (0.0177)	0.0764 (0.0181)	0.114 (0.0165)	0.158 (0.0167)
Observations	63853	63853	60773	60773	63853	63853	63853	63853
R ²	0.599	0.628	0.655	0.631	0.403	0.423	0.453	0.415
Mean DV	1.264	1.264	1.264	1.264	0.705	0.705	0.705	0.705

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table provides robustness tests to the simulations of sentencing by algorithm. It shows estimates from triple-differences age disparities specifications in which the outcomes are a variety of alternative sentence simulations. In Alt 1, all low-risk defendants are assigned a sentence of length 0. In Alt 2, low-risk defendants whose guidelines-recommended sentence is jail are assigned a sentence of length 0 and low-risk defendants with the guidelines-recommended sentence of prison are assigned a 12 month jail sentence. Alt 3 is the same as Alt 2, except high-risk defendants get the average sentence given by judges to people with the same risk score and guidelines-recommended sentence. Alt 4, is the same as Alt 3, except low-risk individuals with a prison sentence get assigned a sentence of length 0. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all defendants convicted of a felony within fiscal years 2001-2004, except for those with missing age information. In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being young, as well as interactions between all of the above.

Table A17: Risk assessment's impact by race and age: robustness to alternative transforms of the sentence length

	Sentence (arcsinh) (1)	Sentence (top coded) (2)	Sentence (arcsinh) (3)	Sentence (top coded) (4)
Eligible x post x Black	0.0383 (0.0469)	0.578 (0.404)		
Eligible x post	0.0213 (0.0331)	-0.106 (0.279)	0.0122 (0.0226)	-0.0123 (0.225)
Eligible x post x young			0.121 (0.0516)	0.918 (0.427)
Observations	60186	60186	63853	63853
R ²	0.627	0.686	0.625	0.684
Mean DV, black	2.266	10.63		
Mean DV, young			1.978	8.699
Covariates	Y	Y	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This figure shows triple difference estimates of risk assessment's impact by race and age. The outcomes are the sentence length with an arcsinh transform and the sentence length top coded at the 99th percentile of positive sentences for eligible defendants (84 months). Standard errors are clustered at the judge level. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes all individuals sentenced in fiscal years 2001-2004, with the exception of those for whom race/age information is missing (for the race/age specifications respectively). In addition to the standard set of covariates, each specification includes an indicator for being post-2003, for being risk-assessment-eligible, for being Black/young (respectively) as well as interactions between all of the above.

Table A18: Varying measures of recidivism

Inc-arcerated	Log Sentence	Recidivism					3yr post-release
		6mo	1yr	3yr	5yr	7yr	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Main sample; recidivism=new felony conviction							
Eligible x post	0.0100 (0.00845)	-0.00219 (0.00272)	-0.00109 (0.00346)	0.00784 (0.00546)	0.00959 (0.00639)	0.0132 (0.00708)	0.0158 (0.00578)
Observations	64675	64675	64675	64675	64675	64675	64675
R ²	0.432	0.626	0.0189	0.0271	0.0439	0.0666	0.0393
Mean DV	0.601	1.324	0.0205	0.0400	0.119	0.245	0.148
Panel B: Subsample that matches to court data; recidivism=new felony conviction							
Eligible x post	0.0104 (0.00886)	-0.00269 (0.00296)	-0.00208 (0.00364)	0.00568 (0.00555)	0.00831 (0.00672)	0.0114 (0.00746)	0.0129 (0.00611)
Observations	59208	59208	59208	59208	59208	59208	59208
R ²	0.435	0.627	0.0191	0.0280	0.0455	0.0677	0.0401
Mean DV	0.603	1.338	0.0202	0.0398	0.121	0.249	0.151
Panel C: Subsample that matches to court data; recidivism=new charge							
Eligible x post	0.0104 (0.00886)	0.0369 (0.0245)	0.00600 (0.00497)	0.00368 (0.00595)	0.00571 (0.00820)	0.0126 (0.00816)	0.00787 (0.00825)
Observations	59208	59208	59208	59208	59208	59208	59208
R ²	0.435	0.627	0.0490	0.0846	0.115	0.112	0.0887
Mean DV	0.603	1.338	0.0937	0.159	0.334	0.434	0.379

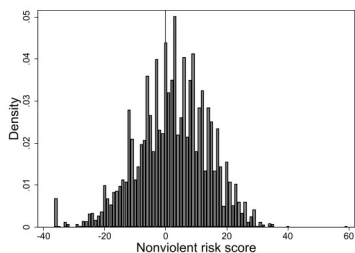
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

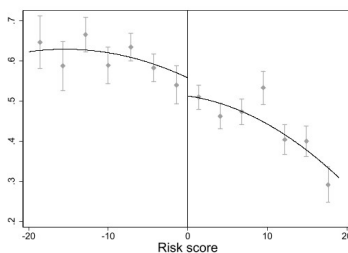
Note: This table shows how risk assessment use affects recidivism, using a variety of different recidivism measures. In Panel A, the recidivism measure is the likelihood of receiving a new felony conviction within varying time windows: 6 months, or 1, 3, 5 and 7 years after sentencing, and, in the final column, 3 years post-release. In Panel B, we use the same recidivism measures but restrict our sample to cases that we are able to match to court data; Alexandria and Fairfax are dropped. In Panel C, our recidivism measure is the likelihood of receiving new felony charges within varying time windows; again, Alexandria and Fairfax are dropped since information on new felony charges is unavailable. The first two columns show how risk assessment affects the probability of incarceration and the log sentence (bottom coded at 2 weeks) in these various samples. Standard errors are clustered at the judge level. The mean dependent variables for eligible cases during the pre-risk assessment period are shown in the bottom row. The sample includes defendants convicted within the fiscal years 2001-2004. Covariates are included in all regressions.

Figure A.1: Covariate balance across risk score cutoffs

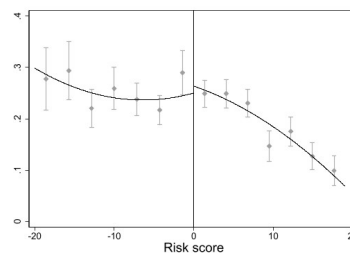
(a) Distribution of nonviolent risk score



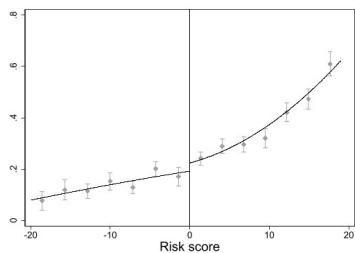
(b) Drug



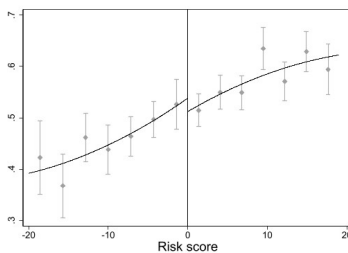
(c) Fraud



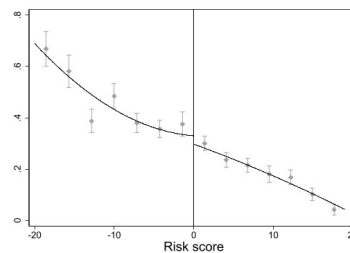
(d) Larceny



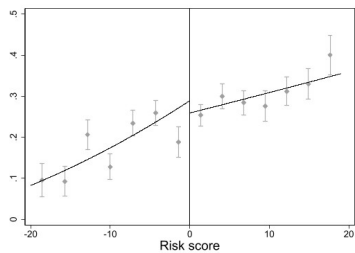
(e) Black



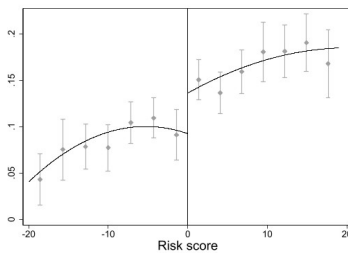
(f) Female



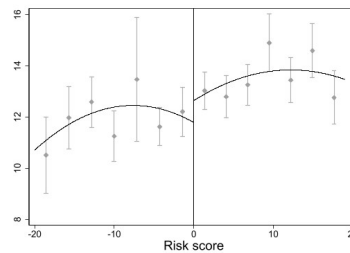
(g) Young



(h) Prior convictions

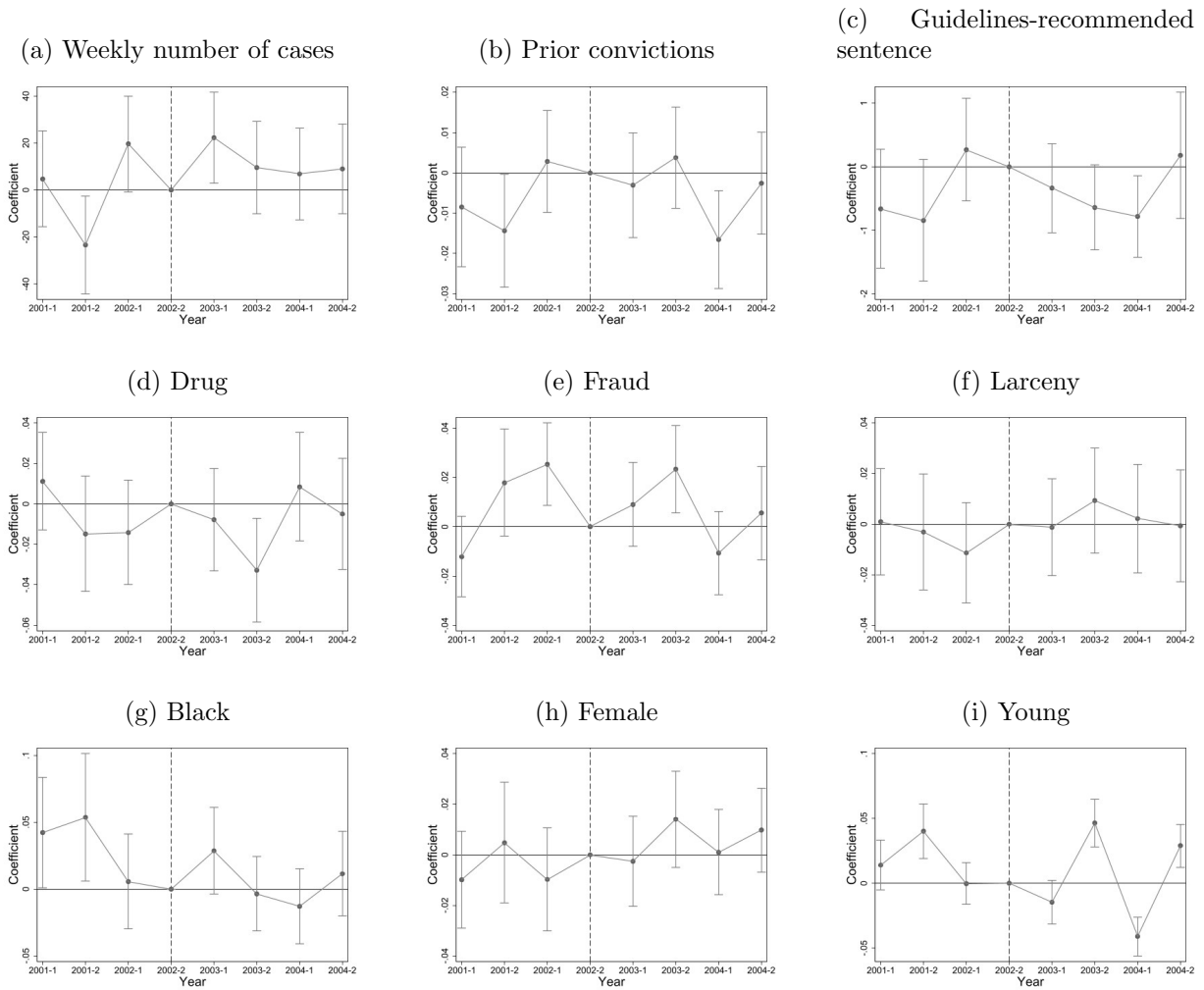


(i) Guidelines-recommended sentence



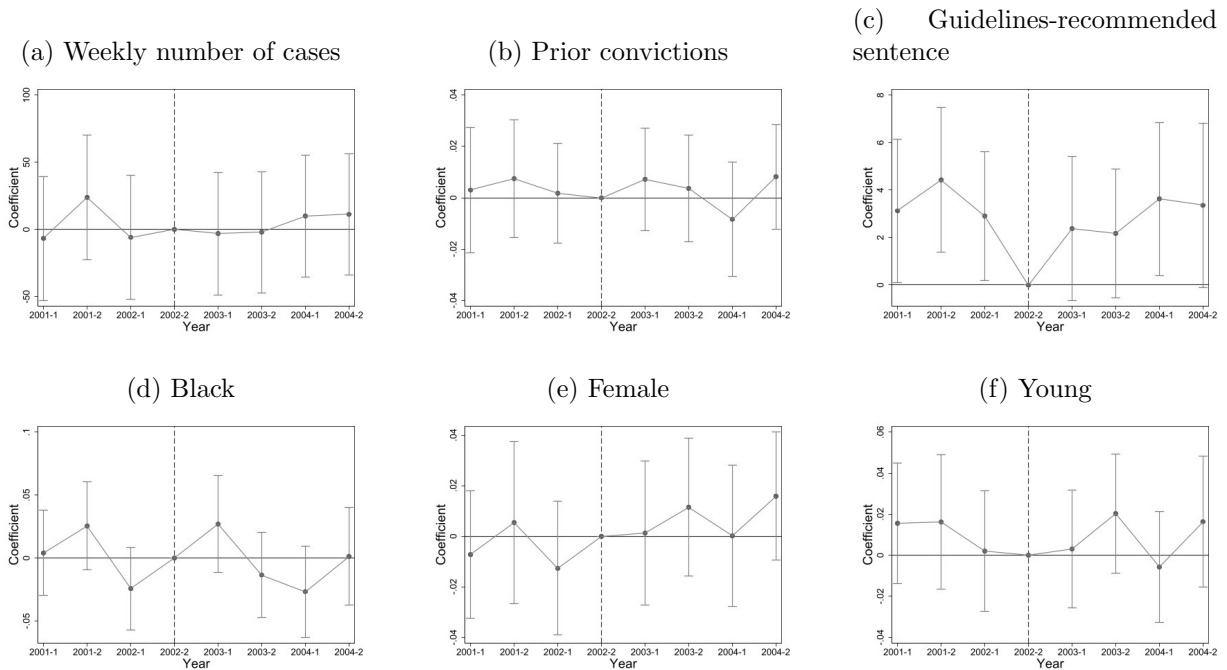
Note: The first sub-figure shows the distribution of the risk assessment with a vertical line at the low-risk cutoff. The remaining sub-figures test covariate balance across the low-risk cutoff. The vertical axes of these figures include an indicator for having a recent prior conviction, the guidelines-recommended sentence (which is not altered by the risk score), a dummy for being convicted of a drug, fraud, or larceny offense, and indicators for being Black, female, or under the age of 23. The horizontal axes show the risk score normalized so that scores below 0 received a diversion recommendation. Each dot shows the mean for a bin of three risk scores, and the whiskers show the 95% confidence interval for that mean. The lines represent fitted polynomial trends of degree 2.

Figure A.2: Dynamic event study: Testing for changes in case composition after risk assessment adoption



Note: This figure shows dynamic event study estimates from regressions of various covariates on dummies for being sentenced during the first or second part of each fiscal year (months July-December and months December-June), with the year prior to risk assessment adoption omitted as a reference. No other controls are included. The sample includes only risk-assessment-eligible defendants sentenced in fiscal years 2001-2004. The data is collapsed to the weekly level for the chart that shows trends in the weekly number of cases. Standard errors are clustered at the judge level (with the exception of the specification showing weekly number of cases) and 95% confidence intervals are shown.

Figure A.3: Dynamic difference-in-differences: Testing for changes in case composition after risk assessment adoption



Note: This figure shows dynamic difference-in-differences estimates with various covariates as outcomes, with the year prior to risk assessment adoption omitted as a reference. No other controls are included. The sample includes all defendants sentenced in fiscal years 2001-2004. The data is collapsed to the weekly level for the chart that shows trends in the weekly number of cases. Standard errors are clustered at the judge level (with the exception of the specification showing weekly number of cases) and 95% confidence intervals are shown.



Sentencing Guidelines Cover Sheet

Complete this form ONLY for applicable felonies sentenced on or after July 1, 2004.

Clear Form

◆ OFFENDER

First: _____ Middle: _____
 Last: _____ Suffix: _____
 Date of Birth: _____ / _____ / _____ Social Security Number: _____
Month Day Year
 CCRE: V A _____ PSI Number: _____
For Use by Probation Officer

◆ COURT

Judicial Circuit: _____ City/County: _____ FIPS Code: _____
 Judge's Name: _____ Office Use Only
 Preparer Name: _____ Preparer Title: Commonwealth's Attorney Probation Officer
 Prosecuting Commonwealth's Attorney: _____ Defense Attorney: _____

◆ CONVICTIONS

Offense	Counts	VCC	Offense Date
Primary Offense: _____	<input type="checkbox"/>	____-____-____	____ / ____ / ____ <small>Month Day Year</small>
Additional Offenses: _____	<input type="checkbox"/>	____-____-____	____ / ____ / ____
_____	<input type="checkbox"/>	____-____-____	____ / ____ / ____

Primary Offense Code Section: § _____ Docket Number: _____

◆ METHOD OF ADJUDICATION

Jury Trial → Sentence Set by Jury: Life Sentence _____ _____ _____
Enter Sentence Years Months Days
 Bench Trial Guilty Plea Alford Plea/Nolo contendere

◆ SENTENCING GUIDELINES RECOMMENDATIONS

Section B

Probation / No Incarceration
 Incarceration 1 Day to 3 Months
 Incarceration 1 Day to 6 Months
 Incarceration 3 to 6 Months
 Probation / No Incarceration or Incarceration to 6 Months

Mandatory Minimum _____

Section C

Life Sentence
 Incarceration *(Enter Midpoint and Range Below)*

Range Midpoint _____ _____
Years Months

Sentence Range _____ _____ TO _____ _____
Years Months Years Months

Recommendation Adjusted for Mandatory Minimum

◆ NONVIOLENT RISK ASSESSMENT Section D of Drug, Fraud, and Larceny Worksheets

Recommended for Alternative Punishment Not Applicable
 NOT Recommended for Alternative Punishment

Nonviolent Risk Assessment Section D

Offender Name:

◆ Ineligibility Conditions

- A. Was the offender recommended for Probation/No Incarceration on Section B? Yes No
- B. Do any of the offenses at sentencing involve the sale, distribution, or possession with intent, etc. of cocaine of a combined quantity of 28.35 grams (1 ounce) or more? Yes No
- C. Are any prior record offenses violent (Category I/II listed in Table A of the Guidelines Manual)? Yes No
- D. Are any of the offenses at sentencing violent (Category I/II listed in Table A of the Guidelines Manual)? Yes No

If answered YES to ANY, go to "Nonviolent Risk Assessment Recommendations" on cover sheet and check Not Applicable. If answered NO to ALL, complete remainder of Section D worksheet.

◆ Offense Type *Select the type of primary offense*

- | | | |
|---------------|----|--|
| Drug | 3 | |
| Fraud | 3 | |
| Larceny | 11 | |

◆ Additional Offense(s) If YES, add 5 →

0

◆ Offender *Score factors A to D and enter the total score*

- | | | | | | |
|--|----|--|---|--|----------------------|
| A. Offender is a male | 8 | | + | | |
| B. Offender's age at time of offense | | | | | |
| Younger than 30 years | 13 | | | | |
| 30 - 40 years | 8 | | | | |
| 41 - 46 years | 1 | | | | |
| Older than 46 years | 0 | | + | | |
| C. Offender not regularly employed | 9 | | | | |
| D. Offender at least 26 years of age & never married | 6 | | | | |
| | | | | | = Enter A to D Total |
| | | | | | 0 |

◆ Arrest or Confinement Within Past 18 Months *(prior to instant offenses)* If YES, add 6 →

0

◆ Prior Felony Convictions and Adjudications *Select the combination of adult and juvenile felony convictions/adjudications that characterizes the offender's prior record.*

- | | | |
|--|---|--|
| Adult felony convictions only | 3 | |
| Juvenile felony convictions or adjudications only | 6 | |
| Both adult and juvenile felony convictions/adjudications | 9 | |

◆ Prior Adult Incarcerations

- | | | | |
|---------|-----------------|---|--|
| Number: | 1 - 2 | 3 | |
| | 3 - 4 | 6 | |
| | 5 or more | 9 | |

Total Score

0

- 38 or less, check Recommended for Alternative Punishment.
- 39 or more, check NOT Recommended for Alternative Punishment.

Go to Cover Sheet and fill out Nonviolent Risk Assessment Recommendations.

Go to Cover Sheet

Clear Form

References

Mason, S. J. and N. E. Graham, “Areas Beneath the Relative Operating Characteristics (ROC) and Relative Operating Levels (ROL) Curves: Statistical Significance and Interpretation,” *Quarterly Journal of the Royal Meteorological Society*, 2002, 128 (584), 2145–2166.