

FEDERAL DEFENDER FACT SHEET: *Flawed U.S. Sentencing Commission Report Misstates Current Knowledge*

Summary

In April 2020, the U. S. Sentencing Commission issued a report entitled “Length of Incarceration and Recidivism.”¹ In its report, the Commission claimed that “incarceration lengths of more than 120 months had a deterrent effect.”² No effect was found for sentences 60 months or less, while sentences between 60 and 120 months yielded inconsistent results.³

None of the findings in this report should be used by judges, legislators, or the Commission to make decisions of any kind. The report badly misrepresents the research literature (Section [I](#)), uses a weak methodology for inferring causation (Section [II](#)), and fails properly to control for defendants’ criminal history (Section [IV](#)). The report states its findings in a misleading form prone to misinterpretation and exaggeration (Section [III](#)). The anomalous pattern of findings fits no theory of deterrence (Section [VI](#)), and no previous study has found the same pattern. Further, it is unlikely the report’s findings would replicate or withstand tests for robustness, but because the Commission will not release data underlying the report, independent evaluation is impossible (Section [IX](#)).

As a bipartisan agency, charged with being a “clearinghouse” for information on the effectiveness of sentencing practices,⁴ the Commission should issue accurate reports on the current state of knowledge regarding important policy questions. This report fails to meet that standard.

¹ See USSC, *Length of Incarceration and Recidivism* (Apr. 29, 2020), https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2020/20200429_Recidivism-SentLength.pdf (“USSC Report”).

² USSC Report, at 4.

³ See *id.*

⁴ 28 U.S.C. § 995(a)(12)(A).

I. The Commission’s report badly misrepresents prior research on the effects of sentence length on recidivism.

The Commission’s review of prior literature on the effects of sentence length on recidivism is misleading and incomplete. In just two paragraphs, the report dismisses prior research as “limited and insufficient.”⁵ Only four studies are cited in endnotes, several which are decades old, and none of which studied federal sentences.⁶ The Commission highlights a study of individuals under post-release state supervision in Ohio as “[t]he most recent study utilizing a rigorous research design” and claims that it “found a deterrent effect.”⁷ But the study was unable to control properly for age, prompting its authors to note that “our finding could, in large part, be attributed to the process of incapacitation and maturation”—not deterrence.⁸

The Commission cites a major review of research published in 2009 for the proposition that existing research is “insufficient for developing federal sentencing policy.”⁹ But the Commission omits the review’s key finding: that the great majority of studies find that imprisonment has a null or criminogenic effect.¹⁰

The Commission’s report ignores a comprehensive 2014 review led by the National Research Council, which concluded, contrary to the Commission’s key findings, that any deterrent effects are modest at best, and diminish, rather than increase, as sentence lengths increase.¹¹

⁵ USSC Report, at 3.

⁶ *See id.* and accompanying endnotes 11-16.

⁷ *Id.* at 3.

⁸ *See* Benjamin Meade et al., *Estimating a Dose-Response Relationship Between Time Served in Prison and Recidivism*, 50 J. Res. Crime & Delinq. 525, 542 (2013). The study was unable to control for defendant age at release from imprisonment, instead using defendant age at commitment.

⁹ USSC Report, at 3, n. 11.

¹⁰ *See* Daniel S. Nagin et al., *Imprisonment and Reoffending*, 38 Crime & Just. 115, 178 (2009). (“Thus, a key finding of our review is that the great majority of studies point to a null or criminogenic effect of the prison experience on subsequent offending.”).

¹¹ *See* Nat’l Res. Council, *The Growth of Incarceration in the United States: Exploring Causes and Consequences*, 154 (Jeremy Travis et al., eds. 2014). Citing previous NRC studies, the report also noted that “insufficient evidence exists to justify predicating policy choices on the general assumption that harsher punishments yield measurable deterrent effects.” And “[n]early every leading survey of the deterrence literature in the past three decades has reached the same conclusion” *Id.* at 90.

Remarkably, the Commission’s report completely ignores a very recent, highly sophisticated study of over 300,000 individuals sentenced under the federal sentencing guidelines, which was co-authored by a team including former Research Directors of both the U.S. Sentencing Commission and the Federal Bureau of Prisons. This 2018 study found that a 28 percent increase in prison time would reduce recidivism from a base rate of 20 percent to 19 percent—a 1 percent drop. “[O]ur sense is that the effect is not large, and that small differences in prison length of stay have a neutral effect on recidivism.”¹²

II. The report used a relatively weak methodology for identifying causal effects, such as the deterrent effect of lengthier incarceration.

Determining whether differences in incarceration lengths is the *cause* of any differences in recidivism rates is a very difficult methodological problem. Defendants receiving different sentences differ in many ways that can affect recidivism, other than incarceration lengths alone. For this reason, experiments using randomly created study and comparison groups are recognized throughout science as the gold standard for inferring causation. The deterrent effect of imprisonment has been studied using “natural experiments” (where randomly assigned judges impose different sentences), but this has been useful largely for studying the effects of differences among relatively short sentences, or between incarceration and non-custodial alternatives.¹³ The 2018 study of federally sentenced individuals not mentioned by the Commission used the logic of both regression discontinuity and instrumental variables to infer causation, which are considered by many methodologists as the next best thing to randomized experiments.¹⁴

The Commission used a different matching methodology in combination with statistical control. Study and comparison groups were created that differed in their length of incarceration,¹⁵ but were as similar as possible on many factors known to affect

¹² William Rhodes, et al., Relationship Between Prison Length of Stay and Recidivism: A Study Using Regression Discontinuity and Instrumental Variables With Multiple Break Points, 17 *Criminology & Pub Pol’y* 731, 754 (2018).

¹³ See, e.g., David Weisburd et al., Specific Deterrence In a Sample of Offenders Convicted of White-Collar Crimes, 33 *Criminology* 587 (1995) (using random assignment of judges and variation in sentencing to find no deterrent effect of imprisonment versus non-custodial sanctions among a sample of pre-guideline federal white-collar defendants).

¹⁴ See Rhodes, at 733.

¹⁵ Study groups were formed from defendants sentenced within five different time intervals: two to three years, three to four years, four to five years, five to ten years, and ten years or more. Matched comparison groups were then formed from defendants sentenced to shorter terms, with the degree of difference varying substantially among the groups and models. See USSC Report, at 6.

recidivism.¹⁶ Any difference in recidivism between the groups was then attributed to a deterrent effect of the difference in sentence lengths. Both the matching and multiple logistic regression used in the Commission study are limited by the availability and reliability of data on other causes of recidivism. The Commission did not match or control for potentially significant differences between the groups, such as the portions sentenced under mandatory minimum statutes, or receiving departures or variances other than for substantial assistance, or the types of drugs involved in drug offenses. Due to a trade-off between the precision of matches and the size of sample available, the Commission used three different designs, with models 1A and 1B using more precise matches but smaller samples than model 2.¹⁷

Further, matching cannot control for unknown factors that might explain any differences in recidivism rates, for example, defendant attitudes or problems perceived by the judge for which no data are available. Study and comparison groups could therefore differ in unknown and uncontrolled ways. Statistical control through multiple regression is also limited by the ability of models to properly specify relationships among the variables. And while the Commission also used weighting and distance matching to attempt to control for differences among the groups, this further complicated the statistical model, making it more vulnerable to statistical artifacts of various kinds.¹⁸

III. The report presents results in a way known to mislead decision makers and exaggerate findings and does not provide crucial data needed for interpretation.

The Commission presents its results as “odds ratios”—a non-intuitive statistic known to exaggerate findings and mislead non-statisticians. Changes in odds are commonly (and erroneously), interpreted as changes in probability, relative risk, or frequency.¹⁹ For

¹⁶ The Commission matched the groups on: age-at-release, high school completion, gender, race, criminal history category (CHC), primary sentencing guideline, whether the instant offense was violent or involved a weapon, and whether the defendant received a substantial assistance departure or safety valve adjustment. *See* USSC Report, at 9-10, tbl. 2. The three different models included in the report varied in the number and precision of the matches. *Id.* at 10-13, 16-17, 20-22, 24-25.

¹⁷ *See* USSC Report, 10-13.

¹⁸ For a review of the variety of problems that have plagued research in this area, *see* David Roodman, *Deterrence is De Minimis*, Open Philanthropy Project (Sept. 24, 2017), <https://www.openphilanthropy.org/blog/deterrence-de-minimis>.

¹⁹ *See, e.g.*, Keith McNulty, *Are You Mixing Up Odds with Probability?*, Medium.com (Jan. 21, 2019), towardsdatascience.com/are-you-mixing-up-odds-with-probability-5f2d385fa890 (last visited May 26, 2020); *see also* Jason W. Osborne, *Bringing Balance and Technical Accuracy to Reporting Odds Ratios and the Results of Logistic Regression Analyses*, 11 *Prac. Assessment, Res., & Evaluation*, Article 7 (2006).

example, the Commission presents the results for Study 1A as “individuals incarcerated for more than 120 months were approximately 45 percent less likely to recidivate” compared to defendants receiving sentences averaging 24 months shorter. But this does *not* mean that the recidivism rate for the group with longer incarceration was 45 percent less. The difference in actual recidivism rates is far less than that.²⁰ Since the report does not provide base rates or the actual rates of recidivism for any of the groups in the study, it is impossible to translate from the reported odds ratios to meaningful real-world differences in recidivism rates.

IV. The variable used to match and control for differences in criminal history was contaminated by the inclusion of career offenders and armed career criminals.

Along with age, criminal history is a major factor associated with recidivism.²¹ Therefore, it is especially important that study and comparison groups be properly matched on criminal history, so that the effects of criminal history are not mistaken for a deterrent effect of lengthier incarceration. Because Congress mandated very long sentences for certain defendants with specified prior offenses, the guidelines contain special rules for so-called “career offenders” and “armed career criminals.”²² In addition to increases to their offense levels, these defendants are automatically placed in Criminal History Category VI (CHC VI), even though most do not have a sufficiently extensive prior record to otherwise be placed in that category. Research has repeatedly shown that these defendants have a significantly *lower* recidivism rate than other CHC VI defendants.²³

The Commission used CHC as the primary matching and control variable for criminal history, despite this known flaw with CHC as a recidivism risk scale. It did not ensure the study and comparison groups contained the same number of the career offenders and armed career criminals. Because these defendants receive very long sentences, it is likely that the study groups receiving the longest sentences contained a larger number of these defendants than the comparisons groups.²⁴

²⁰ Assuming a recidivism rate of 50 percent in a comparison group, a 15 percent lower rate in a study group would result in an odds ratio of .55, which the Commission describes as “45 percent less likely” to recidivate.

²¹ See USSC, *Recidivism Among Federal Offenders: A Comprehensive Overview* 5 (2016) (“*Recidivism Among Federal Offenders*”).

²² See USSG §§ 4B1.1, 4B1.4.

²³ See *e.g.*, *Recidivism Among Federal Offenders*, at 19, figs. 7B & 7A (reporting a recidivism rate of 69.5 percent for career offenders and armed career criminals, lower than the rates for CHCs IV, V or VI).

²⁴ Upon request from the Federal Defenders, the Commission refused to reveal the portions of career offenders and armed career criminals in the study and comparison groups. The Commission claimed it conducted a “confirmatory analysis” without career offenders and armed

As a result of this contamination, 1) the study group may have had a *lower* risk of recidivism than the comparison group, regardless of the difference in length of incarceration, and 2) the regression model predicted (i.e. controlled for) a greater risk of recidivism in the study group than was in fact the case. *It is impossible to assess without more information from the Commission whether the difference in recidivism that the report attributes to the deterrent effect of lengthier incarceration may, in fact, have been due to this uncontrolled pre-existing difference between the groups.*

V. The matching methodology and “doubly robust estimation” touted by the Commission cannot overcome missing data or flaws in the control variables.

The Commission used a “two-stage process” that it argued was “particularly powerful in that only one of the two models needs to be correctly specified to obtain unbiased estimates.”²⁵ It first used matching and weighting to create roughly comparable groups; it then used multiple logistic regression to estimate the effects of differences in incarceration lengths after accounting for the effects of the control variables (the same variables used for matching). But this procedure “is robust to misspecification of one (*but not both*) of these models.”²⁶ In other words, the procedure cannot correct for missing variables or flaws in the controls. Moreover, when both the matching/weighting model and the multiple logistic regression models are misspecified, “the resulting effect estimate may be more biased than that of a single, misspecified maximum likelihood model.”²⁷ For example, if the problem with the CHC control variable described above led to overestimation of the risk of recidivism for the longer-incarcerated study groups, the double estimation could not correct it and could actually compound the problem.

VI. The pattern of major findings in the report is inconsistent with the hypothesis that larger “doses” of incarceration cause greater reductions in recidivism.

The hypothesized deterrent effect of imprisonment is often conceived as a “dose-response relationship”—the longer the “dose,” the greater the reduction of recidivism. In Study 1A, where defendants in the longest incarceration groups differed by an average of just 24

career criminals that yielded substantially similar results as the published findings. This analysis is not mentioned in the report and the Commission has refused to release this analysis and other underlying data necessary for outside researchers to conduct replication or review of the report’s findings. *See infra*, note 39.

²⁵ USSC Report, at 8.

²⁶ Michele Jonsson Funk et al., *Doubly Robust Estimation of Causal Effects*, 173 Am. J. Epidemiology 761, 761 (2011) (emphasis supplied).

²⁷ *Id.* 766.

months, the Commission reported defendants receiving longer sentences were “45 percent less likely to recidivate.” However, in Studies 1B and 2, the effect was smaller, just 30 and 29 percent, respectively, *even though incarceration differences averaged 98 and 100 months*.²⁸ The report offers no explanation for this counter-intuitive result, which raises questions about the amount of noise in the data and statistical artifacts produced by the models. Likewise, the inconsistent findings among the studies regarding defendants incarcerated for 60-120 months (see Figure 17) raise questions about whether the findings are robust to different specifications of the model.²⁹

VII. The report does not test whether *reductions* in sentences already imposed and partially served would increase recidivism among the reductions’ beneficiaries.

This new report evaluates only differences in recidivism among defendants who received different sentence lengths. Previous Commission research—not cited in this report—found that post-sentencing reductions and early release did *not* lead to increased recidivism among beneficiaries of retroactive application of the Fair Sentencing Act.³⁰ The new report does not attempt to reconcile these findings, nor does it caution against improperly generalizing its findings to early release decisions.

VIII. The analyses performed and statements of findings suggest a greater interest in justifying longer sentences than in demonstrating their ineffectiveness.

The Commission claims interest in three potential relationships between length of incarceration and recidivism—deterrent, criminogenic, or no effect.³¹ But the analyses performed, and the way findings are reported, is not so balanced. Results from all three models showed that defendants in the 24- to 36-months group who received longer sentences were *more* likely to recidivate—a criminogenic effect consistent with some prior research.³² But the Commission dismissed these results because 1) they were not statistically significant, and 2) the estimates varied among the models. However, for defendants receiving longer sentences in the 36 to 48 and 48 to 60 month groups, all models consistently found no deterrent or criminogenic effect. But the Commission did not assess the power of this null finding in order to conclude that these sentences had

²⁸ Compare USSC Report, at 16, fig. 6; 19, fig. 8, with USSC Report, at 20, fig. 10; 23, fig. 12; 24, fig. 14; 27, fig. 16.

²⁹ See USSC Report, at 29, fig. 17.

³⁰ See USSC, *Recidivism Among Federal Offenders Receiving Retroactive Sentencing Reductions: The 2011 Fair Sentencing Act Guideline Amendment 3* (2018).

³¹ See USSC Report, at 3, tbl. 1.

³² See, e.g., USSC Report, at 29, fig. 17.

neither a criminogenic nor deterrent effect.³³ The Commission’s efforts to establish confidence in its results appear greater in the direction of justifying longer sentences.³⁴ Moreover, while the formation of groups and statement of key findings appear designed to give the impression that the results have relevance to the effectiveness of mandatory minimum sentences, the study was not in fact designed for this purpose.³⁵

The Commission declares it “considers findings that do not achieve a p-value less than .05 to be unreliable for policy making,”³⁶ which implies it considers findings achieving that level suitable for policymaking. But it should be noted that the Commission’s tests of statistical significance assess only one potential source of unreliability—random error such as that which arises from random sampling or assignment. The significance tests do not assess *systematic* sources of error, bias, unreliability, or invalidity, of the kind that can arise from missing control variables or model misspecification (such as the criminal history contamination explained in Section [IV](#)). Given that the Commission’s findings are not consistent with the larger body of research on this topic; are based on a relatively weak quasi-experimental method and statistical model with specification problems; are not consistent with a dose-response theory of deterrence; and have not been replicated by independent researchers nor tested for robustness; the findings in this report should not be used for policymaking or decisions of any kind.

IX. The data underlying the report should be released immediately.

Public release of data, particularly on important matters of public policy, is a hallmark of science and an important check on the validity of findings.³⁷ The Commission has previously committed to releasing data underlying its reports but has largely not done

³³ See USSC Report, at 14. The report includes no statistical power analyses that might have helped evaluate the reliability of the finding of no effect.

³⁴ In addition, the report repeatedly notes that average time to re-arrest was longer for longer-sentence study groups—a result consistent with a deterrent effect. Only in footnotes is it revealed that these results did not reach statistical significance. See, e.g., USSC Report, at 26, n.51.

³⁵ The decision to use five-to-ten and greater than ten years as boundaries “was to align with five-year clustering often seen in federal mandatory minimum sentences.” USSC Report, at 6. Creating groups and casting findings in these terms invites the false inference that the study was designed to assess the deterrent effects of mandatory minimum statutes. But this would require a very different research design. For example, it appears defendants subject to mandatory minimums were contained in both study and comparison groups in unknown portions.

³⁶ USSC Report, at 14. The p-level measures the likelihood that an apparent difference between groups is actually due to a random source of error.

³⁷ See American Statistical Association *Ethical Guidelines for Statistical Practice* (Apr. 2018), <https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx> (last visited May 26, 2020).

so.³⁸ Upon inquiry by the Federal Defenders, the Commission indicated it does not expect to release the datasets used in this report to outside researchers or the public.³⁹ The Commission's unwillingness to release this data (with appropriate safeguards to protect individual confidentiality) is particularly concerning given the profound policy issues involved, and the serious questions about the report's findings and methodology.

³⁸ Former Commission Vice Chair, William B. Carr announced at a National Seminar on Sentencing Guidelines that the Commission would begin releasing datafiles underlying Commission reports at the time the reports were published. In an April 24, 2014 letter from Glenn Schmitt, Director of the Commission's Office of Research and Data stated "in the next few months we will be posting the datasets used in recent Commission publications." (on file with Sentencing Resource Counsel). Despite these statements, no such consistent practice has been established.

³⁹ Personal communication between Paul J. Hofer, Sentencing Resource Counsel Policy Analyst, and Ryan Cotter, Deputy Dir., Off. of Res. & Data and Glenn Schmitt, Dir., Off of Res. & Data.