Risk Assessment as Policy in Immigration Detention Decisions

David K. Hausman[1]

Abstract

A large literature examines the effects of algorithmic risk assessments on judges' bail decisions in criminal cases. This article examines these effects in the immigration detention context. In 2017, U.S. Immigration and Customs Enforcement changed its risk assessment tool. Before the change, the tool could recommend detention, release or referral to a supervisor. After the change, the tool stopped recommending release—ever. Taking advantage of the suddenness of this change, I show that the removal of the release recommendation reduced actual release decisions by about half, from around 10% to around 5% of all decisions. Officers continued to follow the tool's detention recommendations at only a slightly lower rate after the change, and when officers did deviate from the tool's recommendation to order release, supervisors became more likely to overrule their decisions.

[1]Assistant Professor, University of California - Berkeley School of Law. Special thanks to former colleagues at the ACLU Immigrants' Rights Project, who helped me obtain this dataset under the Freedom of Information Act. Thanks also to Daniel Ho, Nicolas Idrobo, Dorothy Kronick, Sandra Mayson, Justin Simard, Jeremy Weinstein and participants in the Law and Society Association pre-conference on Citizenship and Migration and the Stanford RegLab working group for helpful comments and suggestions. Finally, particular thanks for helpful comments and suggestions from two anonymous reviewers. Disclosure: I served as a consulting expert in litigation in New York City challenging the policy discussed in this article, and I was an attorney at the deposition, in separate litigation, that is cited in the paper. I am no longer serving in either role, although I continue to consult occasionally for immigrants' rights organizations. I did not receive payment for any part of the analysis in this paper. All errors are mine.

1 Introduction

How do judges and other decisionmakers incorporate the recommendations of algorithmic tools when making bail decisions? A large and fast-growing literature examines this question in the criminal justice context, where judges are the decisionmakers. This article considers the same question in the immigration detention context, Immigration and Customs Enforcement ("ICE") officers' decisions carry stakes similar to those in the criminal bail context. Although ICE officers lack the prestige and independence of judges, they, like judges making bail decisions, consider flight risk and danger to decide whether to detain or release someone.

On June 5, 2017, United States Immigration and Customs Enforcement ("ICE") altered its risk assessment algorithm. Before the change, the tool could recommend release, detention, or referral to a supervisor. After the change, the tool could only recommend detention or referral to a supervisor; release was no longer a possible recommendation. When ICE removed the release possibility, it also changed the tool to refer cases to a supervisor two-thirds less often. After these changes, the tool began recommending detention in more than nine out of ten cases. In other words, ICE made a policy decision to increase the use of detention, and it implemented that policy change through a change to its risk assessment software.

I find, using a regression discontinuity in time design, that ICE's change to its risk tool decreased noncitizens' chance of release (including both release outright and release on bond) by about half, from around ten percent to around five percent. Officers and supervisors continued to rely on the tool's recommendations even after the tool stopped recommending release: the probability that an officer would override the tool's recommendation of detention only increased slightly on June 5, 2017, even as detention became a much more common recommendation. And supervisors counteracted officers' slightly increased chance of disagreeing with the tool, becoming more likely to overrule officers' release decisions.

These findings advance the growing literature on human-algorithmic decisionmaking [Green and Chen, 2019] in the context of government detention. A key problem identified by the human-machine decisionmaking literature is that decisionmakers often discount algorithms' predictions [e.g. Hoffman et al., 2017] and, in the bail context, choose to set a high bond (or no bond) even where the algorithm predicts little risk—particularly when the defendant is Black [Main, 2016, Albright, 2019]. This problem persists despite evidence that algorithms may outperform human judges' predictive judgments in many contexts [e.g. Hoffman et al., 2017], including predictions about flight risk and recidivism in the bail context [e.g. Kleinberg et al., 2017].[1]

These results, by contrast, suggest that lower-level officials, at least, may do little to calibrate their reliance on algorithms' predictions and recommendations. ICE officers and supervisors continued to follow the algorithm's recommendations even as release recommendations ceased to exist. These results match those of Albright [2023], who finds that humans respond strongly to changed algorithmic recommendations.

More broadly, however, ICE officials were far more likely to follow algorithmic recommendations—when those recommendations indicated detention—than judges considering criminal bail. ICE officers overrode the release recommendation (before it was eliminated) well over half the time, but virtually never overrode the RCA's detention recommendations. By contrast, while judges' risk assessment override rates in the criminal context have varied—for example, judges overrode 12% of release recommendations and over half of detention recommendations in Angelova et al. [2023, 12], and judges overrode 57% of recommendations for diversion and 27% of recommendations against diversion in Stevenson and Doleac [2022, 2]—I am not aware of any context in which judges virtually never overrode a risk assessment tool's detention recommendation. ICE officers' near-complete unwillingness to override the RCA recommendation in favor of release is therefore notable. That strong pattern might reflect

[1] Assuming that risk assessment algorithms do often outperform human risk judgments (which, as Stevenson [2018] explains, is likely but not certain), those tools may fall prey to the biases, racial and otherwise, that lurk in the data on which they are trained [Mayson, 2019]. This possibility, together with the promise of these tools for reducing the use of detention, has led to an outpouring of interest in ways that pretrial risk assessment algorithms might be tweaked to counteract racial bias or bias in favor of detention [e.g. Corbett-Davies et al., 2017, Kleinberg et al., 2018, Mayson, 2019, Huq, 2018-2019, Kleinberg et al., 2019, Yang and Dobbie, 2020-2021]. Unfortunately, the ICE dataset does not contain race information (or other information, such as names or nationality, that might allow inferences about race), so I am unable to evaluate the role of racial bias in this context.

ICE officers' lack of decisional independence.

These results also illustrate that, for the same reasons that algorithmic recommendations may help influence decisionmakers to improve predictions, such recommendations may accomplish nontransparent administrative policy change [Potter, 2019]. The effectiveness of policy change by algorithm fits the growing body of evidence demonstrating that line-level law enforcement officers are sensitive to the incentives set for them by their supervisors. Mummolo [2017] shows that a procedural change in the New York City police department—requiring additional documentation for street stops—decreased the number of those stops and increased the rate at which searches yielded contraband; Ba and Rivera [2019] show that police union memos in Chicago reduced complaints against the police; Mas [2006] finds that when police pay declines as a result of a union arbitration loss, arrest rates fall and crime rates rise. Like police supervisors, ICE management was able to change officers' behavior: decisionmakers did not "calibrate their reliance on the risk assessment based on the risk assessment's performance" [Green and Chen, 2019].

Finally, these findings also add to the small existing empirical literature on algorithmic decisionmaking at ICE. That literature has not examined the causal effect of the 2017 shift. Koulish [2017] offers the first quantitative overview of the risk assessment tool, and Noferi and Koulish [2014] examine an earlier version of the risk classification assessment data used in this article and conclude that ICE engages in significant overdetention. Koulish and Calvo [2021] examine the contextual determinants of ICE officers' decisions to override RCA recommendations. Koulish [2016] determines that noncitizens who are mandatorily detained are no more likely to pose significant risk than those whom ICE has discretion to release. Evans and Koulish [2020] and Koulish and Evans [2021], finally, document the many versions of the ICE risk tool and the determinants of its risk predictions. I add to this body of work by evaluating, for the first time, the consequences on release rates of the 2017 change to the risk tool.

As scholars and policymakers work to improve algorithmic decisionmaking tools, they should bear in mind the likely effects of nontransparent changes.

2 Background

ICE's Risk Classification Assessment algorithm was introduced in mid-2012 as part of an effort to rationalize immigration detention procedures [Schriro, 2009]. The tool used sparse information about immigration detainees' cases—most prominently, information about criminal convictions or their absence—to make predictions about flight risk and danger to the community [Evans and Koulish, 2020, 804-16]. Based on those predictions, the tool made recommendations about whether ICE should detain the noncitizen, release the noncitizen on bond, release the noncitizen on community supervision (without a bond), or an ICE supervisor should make the decision in the first instance.[2] The RCA procedure required two levels of review of the tool's recommendations: first, an ICE officer could agree or disagree with the tool's recommendations, and second, an ICE supervisor could agree or disagree with the ICE officer's recommendation.

Unlike criminal bail decisions, which are typically made by judges, the human decisions to accept or override the algorithm's recommendations were made by ICE Enforcement and Removal Officers, civil servants within the agency [Evans and Koulish, 2020, 802]. The risk assessment tool fit into the already-existing procedure for ICE officers' decisions about whether to detain noncitizens. Section 236(a) of the Immigration and Nationality Act, 8 U.S.C. § 1226(a), provides that, while certain noncitizens are contesting their deportation, the government may decide whether to detain or release them. The implementing regulations require this decision to take place in two steps, one in the Department of Homeland Security (where the RCA tool is used) and the other in the Department of Justice (where an Immigration Judge makes a decision). In the first step, which is the subject of this article, an ICE officer makes a decision about detention within 48 hours of arrest. 8 C.F.R. § 287.3(d).

[2] In addition, and not relevant here, the software evaluated what level of custody risk each person posed once in detention.

That decision is discretionary: the officer "may, in the officer's discretion, release an alien . . . provided that the alien must demonstrate to the satisfaction of the officer that such release would not pose a danger to property or persons, and that the alien is likely to appear for any future proceeding." 8 C.F.R. § 1236.1(c)(8).

In the years after its implementation, the RCA tool met with resistance and criticism. A report from the Office of the Inspector General concluded that the tool was "time consuming, resource intensive, and not effective in determining which aliens to release or under what conditions" [OIG, 2015, 11]. The Inspector General emphasized the administrative burden of the many questions asked by the tool, along the fact that ICE had never conducted any testing of whether the tool correctly predicted risk [OIG, 2015, 12]. The Inspector General also relied on what it considered a high override rate: in the first year and a half of the RCA tool's use, ICE officers overrode its recommendations more than twenty percent of time [OIG, 2015, 11-12].

ICE responded with a series of changes to make the tool's recommendations more frequently match ICE officers' intuitions by more frequently recommending detention [Koulish and Calvo, 2021]. This article examines the most direct of these changes: the removal, on June 5, 2017, of the possibility that the tool would recommend release. In a deposition,[3] the unit chief of ICE's Information Technology Management Unit explained that this change was intended to implement President Trump's interior enforcement priorities, which were announced months earlier: "The goal would be a lower override rate because that means that you've built a tool that is mimicking the decisions that the officer would normally make. So, for instance, when the [February 20, 2017 enforcement priorities] memo came out and officers were directed to take certain actions, that memo came out, they started taking those actions, then the RCA was changed – RCA was changed to mimic those actions they were already taking"[Wilson, 2019].

[3]Disclosure: I conducted part of this deposition, which took place while I was an attorney at the ACLU.

3 Data

Data come from a Freedom of Information Act Request from the American Civil Liberties Union to Immigration and Customs Enforcement in 2019 (2019-ICFO-10844). ICE produced the data after the ACLU filed a complaint in the District Court for the Southern District of New York. As part of its initial response, ICE produced individual-level data on all Risk Classification Assessment decisions from mid-July 2012 to the end of August 2019. A similar dataset is also available as replication data for Evans and Koulish [2020]

The dataset includes 143,498 detain/release decisions in 2017. The record of each decision includes the RCA tool's estimate of danger and flight risk (low, medium or high) and the tool's recommended action in each case: detain, release, or referral to a supervisor. The dataset also includes the final decision: detain, release or bond. The dataset also records which version of the RCA software was used in each case; the 2017 change involved an update from version 6.3 to 6.4.

Where the tool refers the case to an officer in the first instance, a supervisor reviews the officer's decision, and the dataset includes both the officer's initial decision and the supervisor's agreement or disagreement. In 2017, officers made the initial determination in about three quarters of all cases; the rest were referred directly to a supervisor. Where officers made a first decision, supervisors agreed 92% of the time. Unless otherwise noted, all results in this article concern final outcomes after supervisors' approval or reversal.[4]

4 The June 5, 2017 Changes

The RCA tool changed in two key ways on June, 5, 2017: its flight risk predictions became more pessimistic, and it became more likely to recommend detention, particularly of

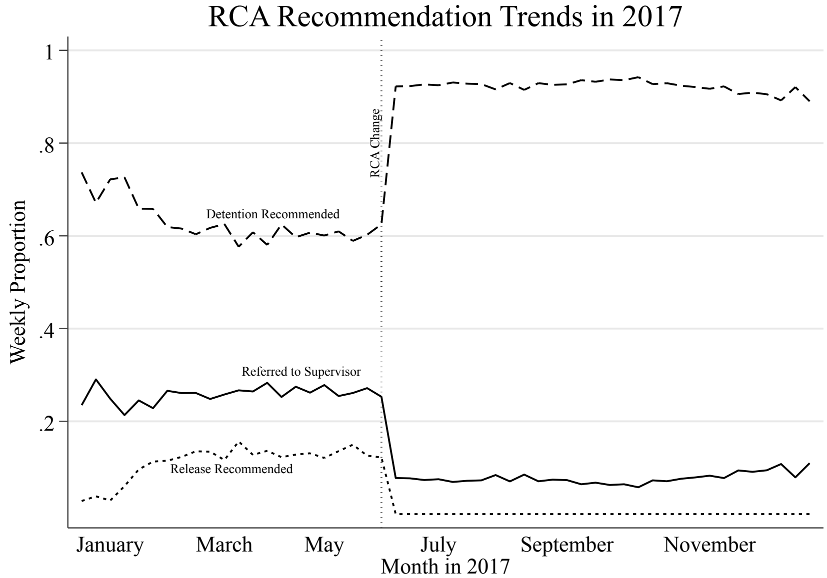
[4]Nearly a third of all RCA decisions (384,519 of 1,348,363 decisions) occurred without any RCA risk prediction. These missing recommendations reflect expedited removal cases (cases involving people apprehended at the border or near the border shortly after entry). The missing predictions begin in August 2013, and ICE noted in its response to the Office of the Inspector General that it, in that month, "streamlined the RCA by generating an automatic detain decision in expedited removal cases, allowing field offices to skip the submission/approval steps otherwise required" [OIG, 2015, 13].

noncitizens for whom it estimated a low or medium risk to the community.

Figure 1 shows how RCA detention recommendations and eventual decisions changed June 5, 2017. The top panel shows RCA recommendations over time in 2017. On June 5, detention recommendations jumped, and release recommendations fell sharply, as did referrals to supervisors.

Officers and supervisors' reliance on RCA recommendations changed little in response to these large changes in recommendations. The bottom panel of Figure 1 shows that the chance that officers and supervisors would agree with a recommendation to detain fell only slightly as such recommendations became much more frequent. The bottom panel of Figure 1 also highlights a key difference between this context and that of judges using risk assessment tools in the bail or sentencing context: ICE officers' highly asymmetric override rate, in which overrides of release recommendations were common but overrides of detention recommendations were rare.

At the same time as the tool's recommendations changed, its risk predictions also changed. Figure 2 shows the effect of the change on flight risk and danger predictions. Danger predictions changed only slightly, but flight risk predictions became drastically more pessimistic.



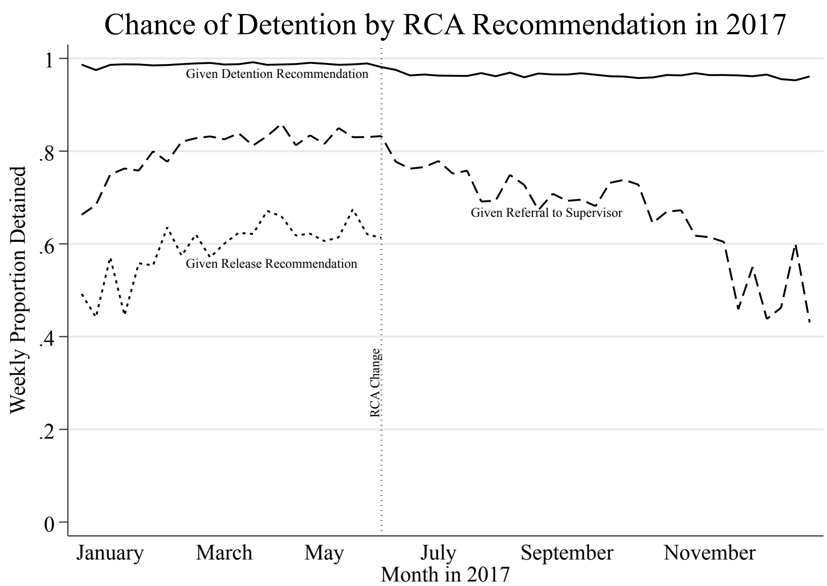


Figure 1: The top panel shows trends in RCA recommendations in 2017. The proportion of cases in which the RCA recommended detention jumped on June 5, 2017 as the proportion of cases fell in which the tool recommended release or referred the decision to a supervisor. The bottom panel shows actual detention decisions—after review by a supervisor—conditional on the RCA's recommendations. The dotted line shows that, before June 5, 2017, when the release recommendation still existed, ICE officers overruled it and ordered detention in well over half of all cases. The dashed line shows that, in cases referred directly to a supervisor, detention became less common over the course of the second half of 2017; perhaps supervisors gradually adapted to a change in case composition as the RCA referred fewer cases to them. Finally, the solid line shows that officers and supervisors became only slightly more likely to override detention recommendations as those recommendations suddenly became more frequent.

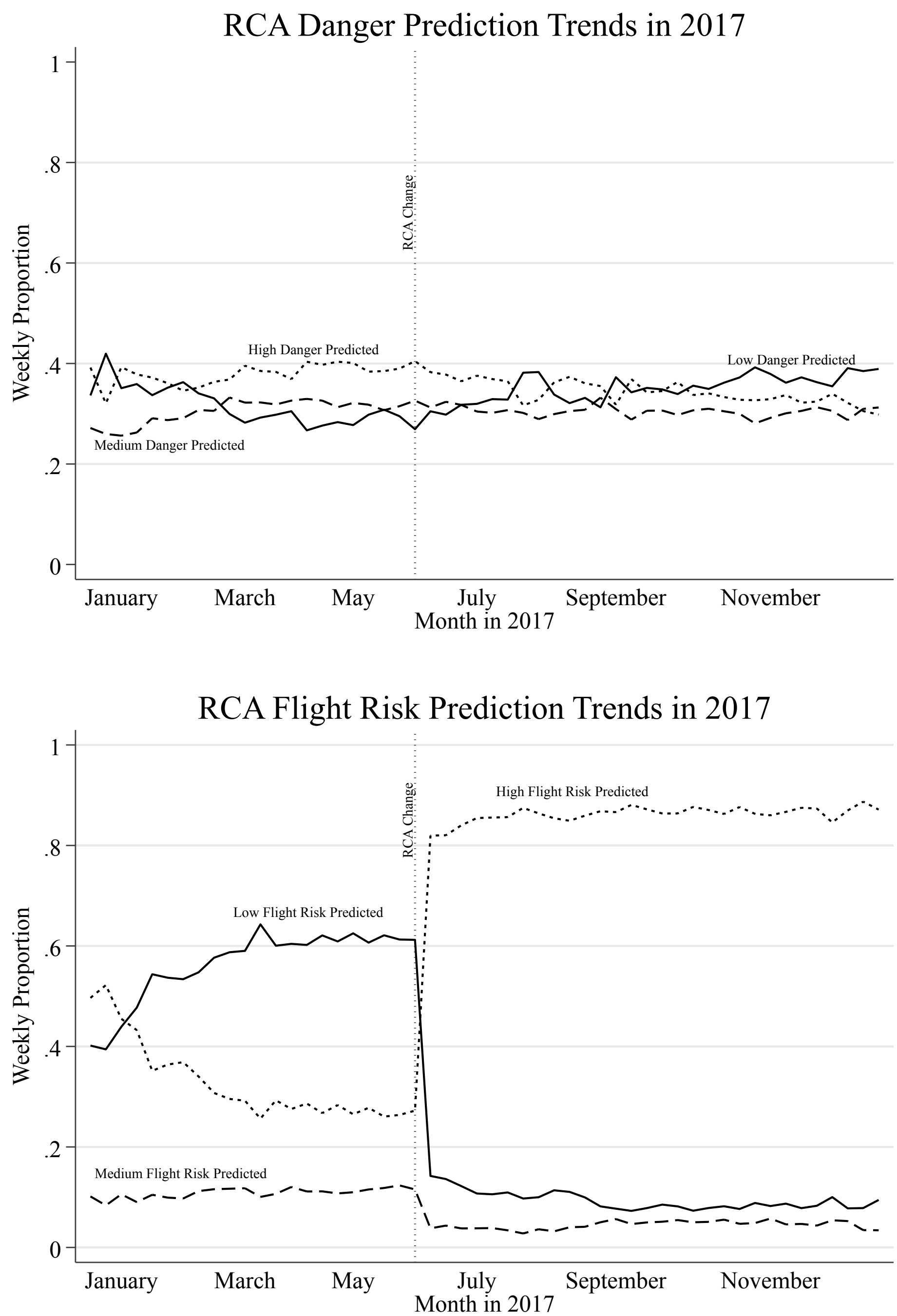


Figure 2: Effect of June 5, 2017 Change on Risk Predictions. The top panel shows RCA danger predictions over time in 2017; these predictions remained fairly steady when the RCA software changed on June 5, 2017. By contrast, RCA flight risk predictions, shown in the lower panel, suddenly became more pessimistic.

5 Research Design

Because the change to the RCA took place on a single day—June 5, 2017—I use a regression discontinuity in time design (in other words, an interrupted time series design) to evaluate its effect.[5] This context is well suited to such a design for two reasons. First, the dataset includes high-frequency individual-level data, which reduces the risk of confounding from unobserved variables. Second, there was a clear moment of intervention when the software was changed.

In order for the regression discontinuity design to identify a causal effect, I must assume that no other change occurred at the same time as the changes to the RCA tool. That assumption is plausible here: no other policy changes occurred on June 5, 2017. The ICE official's deposition indicates that the software was changed to match a policy change—the Trump administration's new enforcement priorities memo—that occurred months before, and another ICE official indicated that the software change on this date was unannounced and unaccompanied by any other policy change.[6]

To implement the regression discontinuity design, I use the robust nonparametric treatment effect estimator and bandwidth selection algorithm developed by Calonico et al. [2017, 2014]. I also show results with bandwidths varying across the range of bandwidths suggested by optimal bandwidth selection methods (as recommended by Arai and Ichimura [2018]). With smaller bandwidths it is safer to assume that the effect is not driven by events other than the software change; larger bandwidths offer more precision.[7]

Table 1 shows descriptive statistics from 2017 before and after the change. The detention recommendation rate is 30 percentage points higher in the post-period than in the pre-period, and the detention decision rate is four percentage points higher in the post-period than in

[5]On June 5 itself, officers in some cases used the old version of the tool and in other cases used the new version, perhaps because the new version was implemented after the work day had started. Because I lack time stamps, and the running variable (days) does not precisely predict the treatment (software version) on that first day, I omit that day in the main results. In the Appendix, I address this problem with a fuzzy regression discontinuity design; the results are nearly identical.

[6]Personal communication, anonymous ICE official, May 25, 2023.

[7]For falsification and density checks, see Appendix A.

the pre-period. These simple differences are similar in magnitude to the estimates from the regression discontinuity design.

	Recommendation			Decision		
	Detain	Release	Supervisor	Detain	Release	Bond
Pre-Change	0.63	0.11	0.26	0.90	0.05	0.06
Post-Change	0.92	0.00	0.08	0.94	0.02	0.04
Total	0.80	0.05	0.15	0.92	0.03	0.04

Table 1: Summary Statistics. All figures are proportions, where the outcome equals 1. The pre-change period is from January 1 to June 4, 2017, and the post-change period is from June 5, 2017 to Dec 31, 2017.

6 Results

Did the 2017 change—with its removal of the release recommendation and decrease in supervisor referrals—cause ICE to order detention more often? Figure 3 shows the main results graphically, binning the data and fitting a line before and after the cutoff [Calonico et al., 2017, 2014]. The figure suggests that the release rate (the rate at which ICE granted either outright release or release on bond) dropped by about half, from around 10% to around 5%.

Figure 4 then plots regression coefficients for the same results. The top panel of Figure 4 shows that the results are robust to the order of the local polynomial and different estimation techniques; the bottom panel shows that the estimates are robust to many bandwidths, including those recommended by four optimal bandwidth selectors. The various estimates are all similar; the change to the tool reduced the release rate by slightly more than five percentage points, cutting the previous release rate roughly in half. Finally, the appendix shows fuzzy regression discontinuity results, where the first stage predicts which software version is used; the results are nearly identical.

The 2017 change cut the release rate in half. What were the relative contributions of

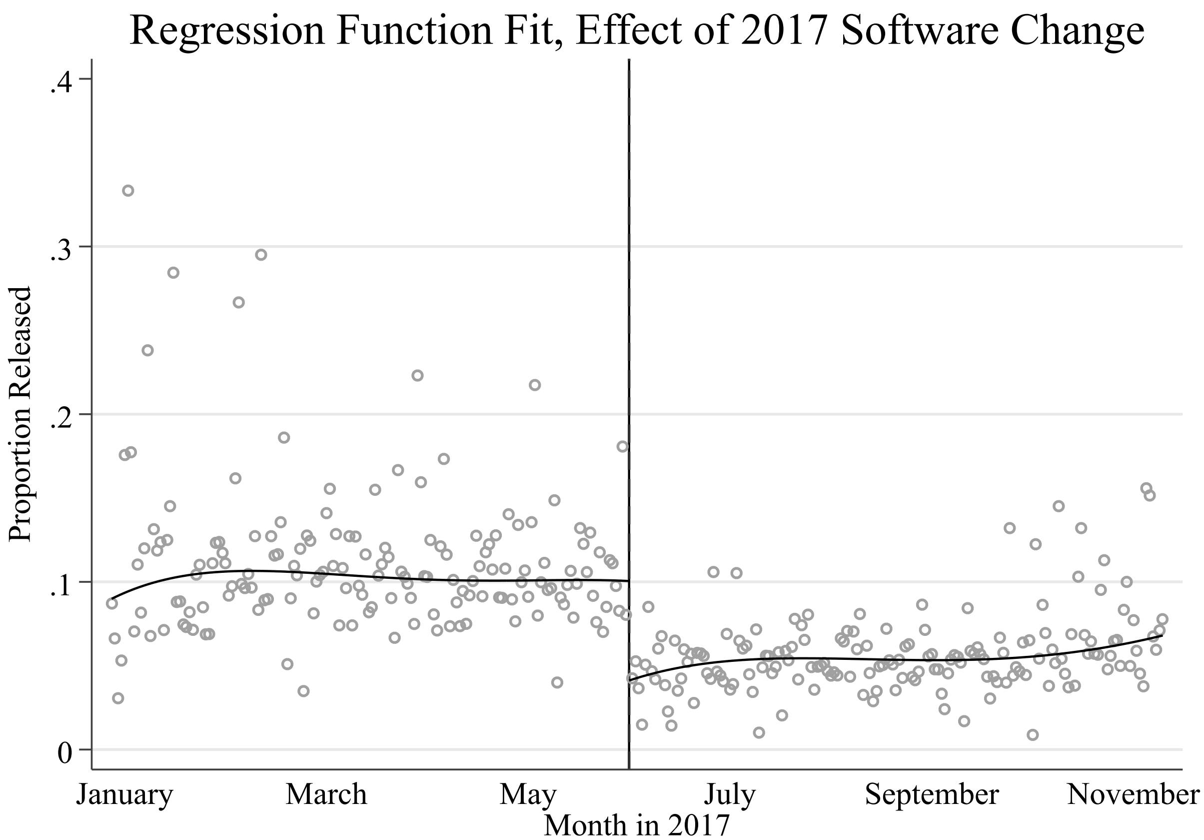
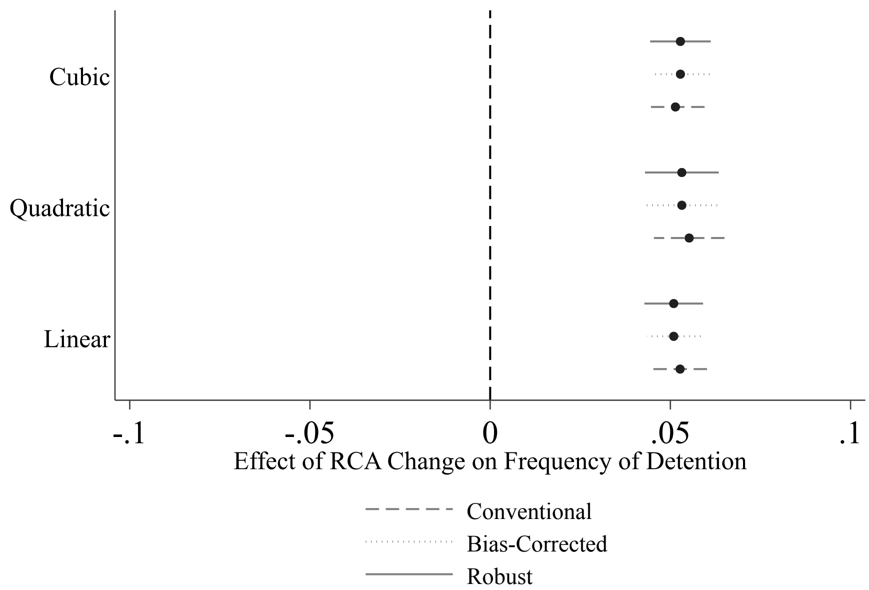


Figure 3: Regression Discontinuity Plot of Release Rate Over Time. On June 5, 2017 (marked by the vertical line) ICE changed its software to remove the release recommendation and to make referrals to a supervisor less common. The figure shows a plot of release rates over time, along with a polynomial fit, using the standard binning and polynomial fit methods proposed by Calonico et al. [2017, 2014].



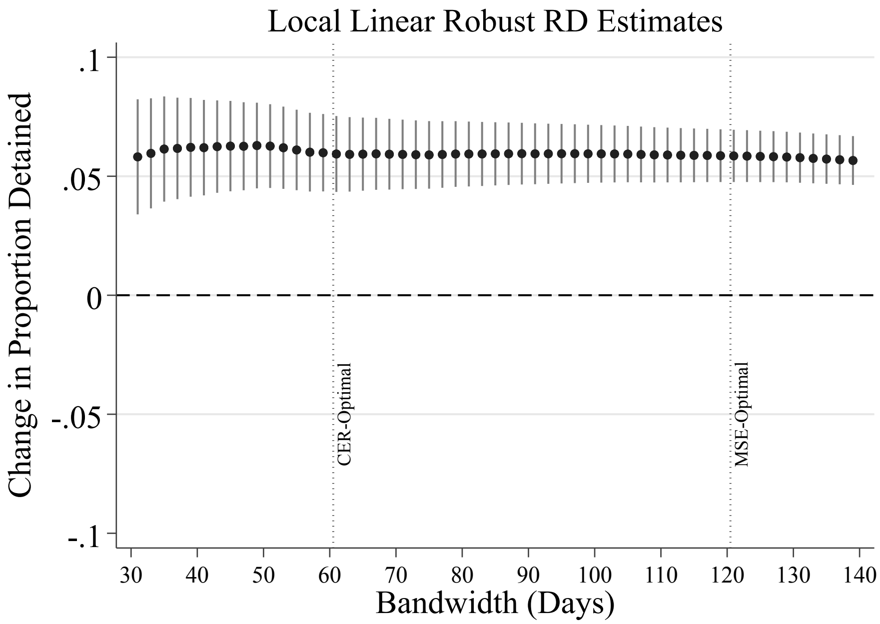


Figure 4: Effect of Software Change on Detention Rate: Regression Discontinuity Estimates. Both the top and bottom panels show estimates, using the nonparametric estimation technique from Calonico et al. [2017, 2014], of the effect of ICE's June 5, 2017 software change on the likelihood that noncitizens were detained. The day of the change is omitted because some cases on that day involved the new software version and others did not. The top panel shows how the estimates vary with the order of polynomial and the estimation technique; the bandwidths are 120, 152, and 387 days for for the linear, quadratic, cubic variants respectively. The bottom panel shows how the estimates vary with various bandwidths. The estimates consistently imply that the change, which removed the release recommendation and made referrals to a supervisor less common, made detention 5-6 percentage points more likely and therefore cut the release rate in half, from about 10% to about 5%. Dots show point estimates, and bars indicate 95% confidence intervals.

the changes in predictions and recommendations to the eventual changes in ICE decisions? Unfortunately, this question is unanswerable because the changes to flight risk predictions, danger predictions, and detention recommendations all occurred simultaneously, and the dataset does not contain enough information to estimate counterfactual predictions or recommendations.

It is possible, however, to examine the relative roles of supervisors and officers. Recall that where an officer made a decision in the first instance (i.e. where the tool did not refer the decision directly to a supervisor), supervisors then agreed or disagreed with those decisions. Supervisors' disagreement was relatively rare: it occurred in about 8% of cases in 2017. Despite that low baseline disagreement rate, supervisors' reaction to the June 5, 2017 change was noticeably different from that of officers. Figure 5 shows that officers became slightly more likely to disagree with recommendations to detain when those recommendations became more frequent. In other words, officers partly compensated for the change to the tool, disagreeing with detention recommendations more often to order release. But when officers disagreed in this way, by ordering release, supervisors became more likely to overrule the officers' decisions and order detention.

Figure 5 shows these differential effects on officer and supervisor behavior. The bottom panel includes all cases in which the tool recommended detention before and after the software change. Officers became less likely to agree with that recommendation and more likely to order release. The top panel shows what supervisors did in the subset of cases in which the tool recommended detention but officers ordered release: after the software change, supervisors became more likely to overrule those officers' release decisions. In other words, supervisors played a role in making the change in RCA recommendations translate into a change in decisions. Both the change in the software and supervisors' response to that change encouraged officers to order detention more often.

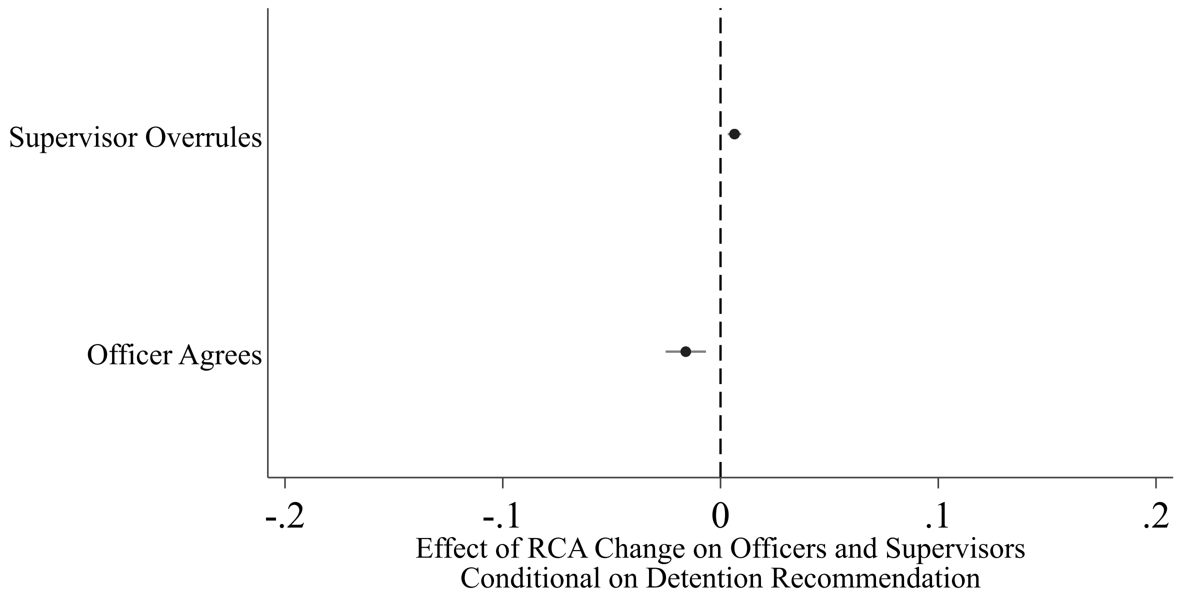


Figure 5: Effect of Software Change on Officer and Supervisor Decisions. The figure shows local linear estimates, using the estimation and MSE-optimal bandwidth selection technique from Calonico et al. [2017, 2014], of the effect of ICE's June 5, 2017 software change on officer and supervisor behavior. The lower coefficient shows the chance, in cases where the tool recommended detention, that the ICE officer agreed with that recommendation and actually ordered detention. Officers became slightly more likely to disagree with the recommendation and to order relief. When officers did so, supervisors became slightly more likely to overrule their decision and to order detention nonetheless (see the top coefficient).

7 Discussion

As scholarly interest in risk assessment tools in the criminal bail and sentencing contexts has grown, the similar but distinct immigration detention context has received less attention. When ICE changed its risk assessment tool for detention decisions on June 5, 2017, it entirely removed the possibility that the tool would recommend release. After the change, the tool could only recommend detention or refer cases to a supervisor. ICE officers became only slightly less like to follow the recommendation, and detention therefore became more common. In other words, the software change accomplished a policy change.

The ICE detention decisionmaking context differs from the typical criminal bail context: independent judges set criminal bail, whereas ICE officers, who answer to their supervisors, make detention decisions for noncitizens. The hierarchical ICE context might have lent the tool's changing recommendations additional weight. But these results are consistent with those of Albright [2023], who finds that changes to algorithmic recommendations can lead to large real-world changes in the criminal bail context. One possible normative implication is that such policy changes, like non-algorithmic policy changes, should be made transparently [e.g. Carlson, 2017-2018].

These results also have limitations. Because the ICE dataset lacks demographic information, I am unable to examine the relative effects of the change by nationality, race, or demographic characteristics. And I am also unable to measure the relative importance of the changes to the flight risk scores and the changes to the detention recommendations themselves. This article does, however, present strong evidence that the 2017 changes to ICE risk assessment software had real-world effects, leading officers to detain immigrants more often.

References

Alex Albright. If you give a judge a risk score: Evidence from kentucky bail decisions. *Harvard John M. Olin Fellows Discussion Paper*, 2019.

Alex Albright. The hidden effects of algorithmic recommendations. *Preprint. Last accessed March*, 28:2023, 2023.

Victoria Angelova, Will S Dobbie, and Crystal Yang. Algorithmic recommendations and human discretion. Technical report, National Bureau of Economic Research, 2023.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434): 444–455, 1996. doi: 10.1080/01621459.1996.10476902. URL `https://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476902`.

Yoichi Arai and Hidehiko Ichimura. Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics*, 9(1):441–482, 2018. doi: https://doi.org/10.3982/QE590. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/QE590`.

Bocar Ba and Roman Rivera. The effect of police oversight on crime and allegations of misconduct: Evidence from chicago. *Working Paper*, 2019.

Sebastian Calonico, Matias D. Cattaneo, and Rocío Titiunik. Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4):909–946, 2014. doi: 10.1177/1536867X1401400413. URL `https://doi.org/10.1177/1536867X1401400413`.

Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404, 2017. doi: 10.1177/1536867X1701700208. URL `https://doi.org/10.1177/1536867X1701700208`.

Alyssa M. Carlson. The need for transparency in the age of predictive sentencing algorithms notes. *Iowa Law Review*, 103(1):303–330, 2017-2018. URL `https://heinonline.org/HOL/P?h=hein.journals/ilr103&i=311`.

Matias D. Cattaneo, Michael Jansson, and Xinwei Ma. Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261, 2018. doi: 10.1177/1536867X1801800115. URL `https://doi.org/10.1177/1536867X1801800115`.

Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press, 2020. doi: 10.1017/9781108684606.

Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 797–806. Association for Computing Machinery, 2017.

Kate Evans and Robert Koulish. Manipulating risk: Immigration detention through automation. *Lewis & Clark Law Review*, 24:789, 2020.

Ben Green and Yiling Chen. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, nov 2019. doi: 10.1145/3359152. URL `https://doi.org/10.1145%2F3359152`.

Mitchell Hoffman, Lisa B Kahn, and Danielle Li. Discretion in Hiring*. *The Quarterly Journal of Economics*, 133(2):765–800, 10 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx042. URL `https://doi.org/10.1093/qje/qjx042`.

Aziz Z. Huq. Racial equity in algorithmic criminal justice. *Duke Law Journal*, 68(6):1043–1134, 2018-2019. URL `https://heinonline.org/HOL/P?h=hein.journals/duklr68&i=1067`.

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human Decisions and Machine Predictions*. *The Quarterly Journal of Economics*, 133(1):237–293, 08 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx032. URL https://doi.org/10.1093/qje/qjx032.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. *AEA Papers and Proceedings*, 108:22–27, may 2018. doi: 10.1257/pandp.20181018. URL https://doi.org/10.1257%2Fpandp.20181018.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein. Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, 10:113–174, 04 2019. ISSN 2161-7201. doi: 10.1093/jla/laz001. URL https://doi.org/10.1093/jla/laz001.

Robert Koulish. Using risk to assess the legal violence of mandatory detention. *Laws*, 5(3): 30, jul 2016. doi: 10.3390/laws5030030.

Robert Koulish. Immigration detention in the risk classification assessment era. *Conn. Pub. Int. LJ*, 16:1, 2017.

Robert Koulish and Ernesto Calvo. The human factor: Algorithms, dissenters, and detention in immigration enforcement. *Social Science Quarterly*, 102(4):1761–1786, 2021.

Robert Koulish and Kate Evans. Punishing with impunity: The legacy of risk classification assessment in immigration detention. *Geo. Immigr. LJ*, 36:1, 2021.

Frank Main. Cook county judges not following bail recommendations: Study. *Chicago Sun-Times*, 2016.

Alexandre Mas. Pay, Reference Points, and Police Performance*. *The Quarterly Journal of Economics*, 121(3):783–821, 08 2006. ISSN 0033-5533. doi: 10.1162/qjec.121.3.783. URL https://doi.org/10.1162/qjec.121.3.783.

Sandra G. Mayson. Bias in, bias out. *Yale Law Journal*, 128:2218, 2019. URL https://heinonline.org/HOL/Page?handle=hein.journals/ylr128&id=2309&div=&collection=.

Jonathan Mummolo. Modern police tactics, police-citizen interactions and the prospects for reform. *The Journal of Politics*, 2017.

Mark Noferi and Robert Koulish. The immigration detention risk assessment. *Georgetown Immigration Law Journal*, 29:45, 2014. URL https://heinonline.org/HOL/Page?handle=hein.journals/geoimlj29&id=49&div=&collection=.

OIG. U.s. immigration and customs enforcement's alternatives to detention. Technical report, 2015.

Rachel Augustine Potter. *Bending the Rules*. University of Chicago Press, 2019. doi: doi:10.7208/9780226621883. URL https://doi.org/10.7208/9780226621883.

Dora Schriro. Immigration detention: Overview and recommendations. Technical report, Immigration and Customs Enforcement, 2009.

Megan Stevenson. Assessing risk assessment in action. *Minnesota Law Review*, 103:303, 2018. URL https://heinonline.org/HOL/Page?handle=hein.journals/mnlr103&id=313&div=&collection=.

Megan T Stevenson and Jennifer L Doleac. Algorithmic risk assessment in the hands of humans. *Available at SSRN 3489440*, 2022.

Jeffrey Wilson. Deposition in hernandez v. barr. on file with author, March 2019.

Crystal S. Yang and Will Dobbie. Equal protection under algorithms: A new statistical and legal framework. *Michigan Law Review*, 119(2):291–396, 2020-2021. URL https://heinonline.org/HOL/P?h=hein.journals/mlr119&i=309.

A Appendix: Covariate and Density Test

For the regression discontinuity design to yield the causal effect of ICE software change, we must assume that no other systematic changes occurred at the same time as the update from version 6.3 to 6.4. As a falsification check, I test whether the change in software versions affected the likelihood that ICE officers would record a special vulnerability (often a disability of some kind) in the dataset. Unfortunately, the dataset is quite sparse, and this is the only covariate that I am confident should not have been affected by the software change. I code this variable as 1 if the officer records any special vulnerability and zero otherwise. Table A1 shows the results; as expected, there is no evidence of any effect.

Table A1: Fuzzy Regression Discontinuity Estimates of Effect of RCA Software Change

	(1) Linear	(2) Quadratic	(3) Cubic
	(1) Linear	(2) Quadratic	(3) Cubic
Conventional	-0.00104	-0.00370	-0.00454
	(0.00289)	(0.00358)	(0.00351)
Bias-corrected	-0.00289	-0.00486	-0.00596
	(0.00289)	(0.00358)	(0.00351)
Robust	-0.00289	-0.00486	-0.00596
	(0.00312)	(0.00390)	(0.00359)
N	74475	106574	192931

Estimates use the nonparametric estimation technique from Calonico et al. [2017, 2014]
Standard errors in parentheses
Bandwidths are 94, 135, and 245 days, respectively
Bandwidths selected using MSE-optimal bandwidth selector
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In addition, I perform a density check. The intuition behind this test is that there should not be bunching of observations around the cutoff. Figure A1 shows the distribution of daily case counts before and after the cutoff. Note that the treatment was introduced on a Monday, so it is unsurprising that case counts were higher than on the previous Saturday and Sunday; that pattern matches other weeks. More formally, the manipulation test proposed by Cattaneo et al. [2018] yields a p-value of .14. That test uses a triangular kernel function, a combination of MSE-optimal bandwidth selectors, and a jackknife variance-covariance matrix estimator.

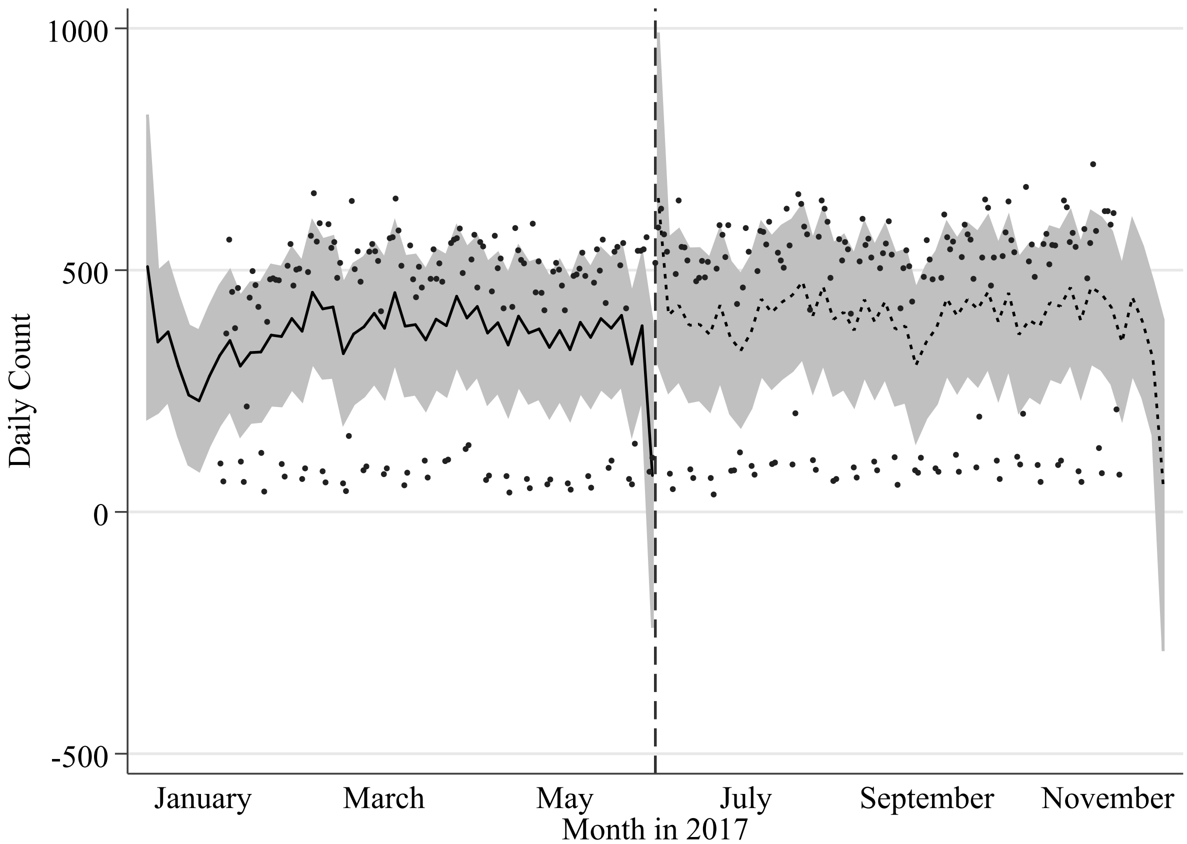


Figure A1: Plot of Case Counts Over Time. On June 5, 2017 (marked by the vertical dashed line) ICE changed its software to remove the release recommendation and to make referrals to a supervisor less common. The dots are case counts by day (the smallest time unit in the dataset), and the line shows a local linear polynomial fit of daily counts over time, with a bandwidth of two days, fit separately on either side of the cutoff. Note that the treatment occurred on a Monday, and case counts are systematically lower on weekends.

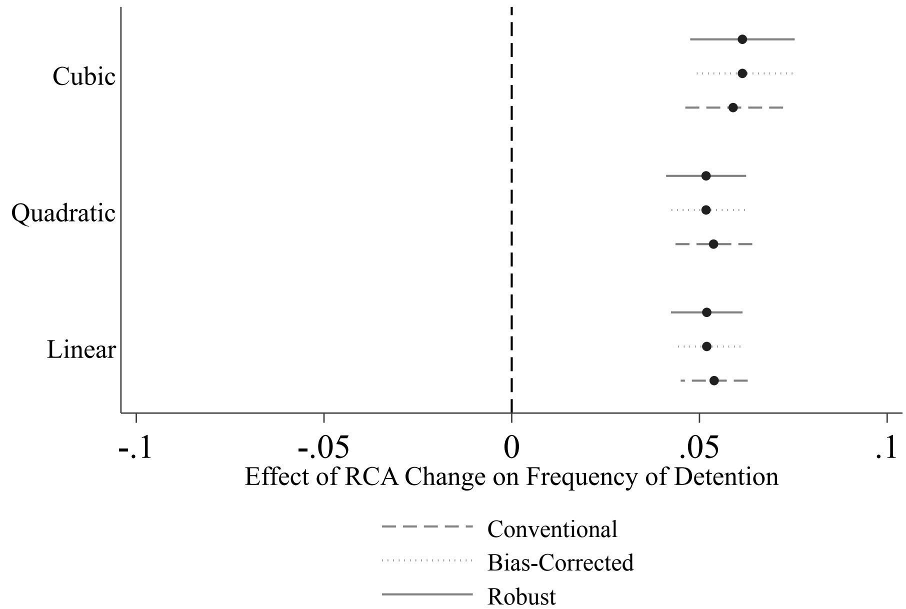
B Fuzzy Regression Discontinuity Results

This section presents fuzzy regression discontinuity estimates, which are very similar to the simpler estimates presented in the main text. A fuzzy RDD is appropriate here because the running variable does not perfectly predict treatment assignment (although the problem is limited to a single day, which I simply omit in the main results).

The fuzzy RDD requires that I assume that the arrival of June 5, 2017 affected ICE decisions only through the software change (the exclusion restriction) and that the arrival of June 5, 2017 did not cause any ICE officers to revert to the use of the previous version of the software (no defiers) [Angrist et al., 1996]. Both assumptions are plausible. First, there were no other policy changes on that day, and second, as of the day after the change, the compliance rate reached 100% as ICE used the new software in every case. The variation on the day of the change is not driven by differences across regions; all but one of the 24 field offices had a mix of the two software versions on that day, and the one exception saw only two cases that day. Unfortunately, the data does not include time stamps, but one likely possibility is that the software version changed sometime in the course of the relevant day.

In the first stage of the fuzzy regression discontinuity design, the score is the day, the cutoff is June 5, 2017, and the treatment is the change to the software [Cattaneo et al., 2020]. The second stage estimates the effect of the change in software versions on the probability of release. As in the simple RDD, to implement the fuzzy RDD, I use the robust nonparametric treatment effect estimator and bandwidth selection algorithm developed by Calonico et al. [2017, 2014]. Again, I show results with bandwidths varying across the range of bandwidths suggested by other optimal bandwidth selection methods (as recommended by Arai and Ichimura [2018]).

Table B2 shows nonparametric regression results [Calonico et al., 2017, 2014], and Figure B2 shows the same results graphically. The top panel of Figure B2 shows that the results are robust to the order of the local polynomial and different estimation techniques; the bottom panel shows that the estimates are robust to many bandwidths. The various estimates are all similar; the change to the software reduced the release rate by slightly more than five percentage points, cutting the previous release rate roughly in half.



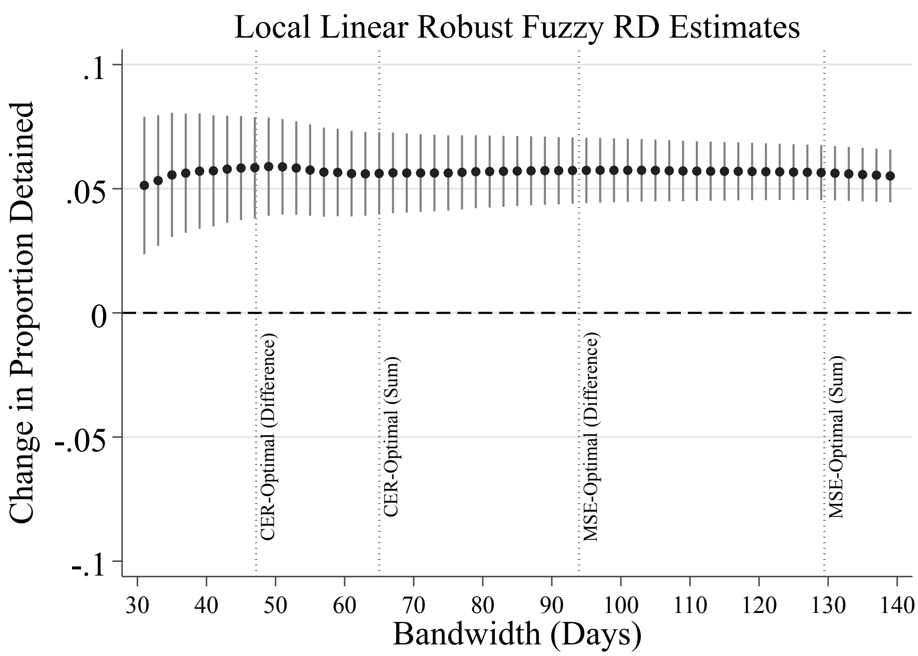


Figure B2: Effect of Software Change on Detention Rate: Fuzzy Regression Discontinuity Estimates. Both figures show estimates, using the nonparametric estimation technique from Calonico et al. [2017, 2014], of the effect of ICE's June 5, 2017 software change on the likelihood that noncitizens would be detained. The top panel shows how the estimates vary with the order of polynomial and the estimation technique; the bottom panel shows how the estimates vary with various bandwidths. The estimates consistently imply that the software change, which removed the release recommendation and made referrals to a supervisor less common, made detention 5-6 percentage points more likely and therefore cut the release rate in half, from about 10% to about 5%. Dots show point estimates, and bars indicate 95% confidence intervals.

Table B2: Fuzzy Regression Discontinuity Estimates of Effect of RCA Software Change

	(1) Local Linear	(2) Local Quadratic	(3) Local Cubic
Conventional	0.0539***	0.0538***	0.0540***
	(0.00456)	(0.00526)	(0.00420)
Bias-corrected	0.0520***	0.0518***	0.0555***
	(0.00456)	(0.00526)	(0.00420)
Robust	0.0520***	0.0518***	0.0555***
	(0.00486)	(0.00544)	(0.00439)
N	73538	122427	320015

Estimates use the nonparametric estimation technique from Calonico et al. [2017, 2014]

Standard errors in parentheses

Bandwidths are 94, 156, and 405 days, respectively

Bandwidths selected using MSE-optimal bandwidth selector

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