January 2019

Managing Pretrial Misconduct: An Experimental Evaluation of HOPE Pretrial

Janet Davidson
Chaminade University
jdavidso@chaminade.edu

George King
Hawaii State Department of Public Safety
George.R.King@hawaii.gov

Jens Ludwig
University of Chicago
jludwig@uchicago.edu

Steven Raphael
Goldman School of Public Policy
University of California, Berkeley
stevenraphael@berkeley.edu

We thank Patricio Dominguez-River, Claire Riggan, and Caitlin Martinez for their excellent research assistance. We also wish to thank Joshua Christopher and Sergey Shevtchenko for their technical assistance with the rolling randomization process. This evaluation was funded by a grant from the Laura and John Arnold Foundation.

EXECUTIVE SUMMARY

Hawaii's Opportunity Probation with Enforcement (HOPE) strategy provides one blueprint for managing those convicted of a criminal offense in a community correction setting. The HOPE strategy in Honolulu has evolved considerably over the past fourteen years, from a small-scale experiment involving the highest risk probation cases in Honolulu to a key element of the continuum of probation supervision and services in Honolulu County. In its current form, HOPE serves high risk probationers who have not performed well at lower levels of supervision yet are not deemed in need of an intensive drug-court intervention or who cannot access the smaller drug court program due to resource constraints (Institute for Behavior and Health 2015, Alm 2016, Hawken 2016). HOPE probation involves three components: (1) a probation officer committed to rehabilitation, (2) the active involvement of a patient judge who is knowledgeable about addiction and addiction treatment, and (3) swift, certain, consistent and proportionate sanctions for technical violations such as failed drug tests, missed appointments, and failure to attend or complete treatment. From the start, probationers are encouraged to succeed, expectations are clearly articulated by a presiding judge, and subsequent failures as well as successes are promptly addressed in a swift and consistent manner in open court. The program also offers drug treatment for whoever requests it (and in some instances mandated treatment for frequent violations), cognitive behavioral therapy, and other targeted interventions for those who need them.

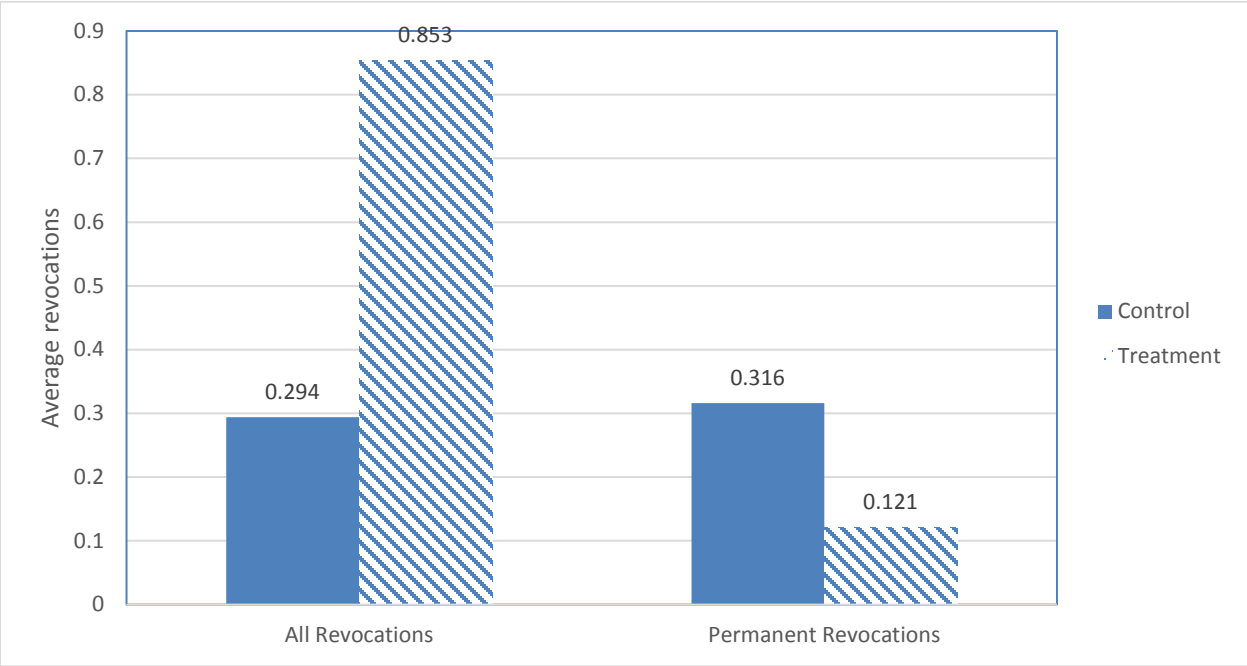
To date, HOPE has been targeted primarily at individuals who have been convicted of an offense and who are serving probation or parole terms. With the exception of the program evaluated here, we are unaware of an attempt to apply HOPE principals to individuals conditionally released pretrial. Defendants are detained pretrial to minimize the likelihood of pretrial misconduct, to ensure attendance at trial, and in some instances due to public safety concerns. Pretrial misconduct may include new criminal offending, substance abuse, failure to report to a pretrial service staff, or failing to appear for court hearings and/or trial. While the use of commercial bail bonds provides an incentive for those who bail out to refrain from such misconduct, heavy reliance on commercial bail puts low-income/low-asset defendants from low-income/low-asset families at a serious disadvantage. Hence, there is demand for alternative non-incarceration forms of pretrial monitoring that may be effective in preventing pretrial misconduct for those unable to make bail.

In this project we evaluate the application of the case management and treatment delivery practices developed under the HOPE probation strategy to pretrial individuals who are conditionally released from jail subject to criminal justice supervision. In the jurisdiction we study (Honolulu, Hawaii), defendants on supervised release are typically monitored by pretrial officers located at the county jail. The revocation of supervised release occurs once a defendant has failed to comply several times with a set of pre-specified conditions, including but not necessarily limited to refraining from drug use and additional criminal activity, maintaining contact with the assigned pretrial officers, and making all scheduled court dates. The intervention we evaluate applies random drug testing in conjunction with swift, certain, consistent, and proportionate sanctions to pretrial misconduct. That is to say, misconduct is met with quickly administered arrest and re-incarceration, yet subsequent jail spells are proportionate to the seriousness of the violation. The intervention also includes drug treatment interventions for those who repeatedly fail drug tests (or who request treatment services) and direct interaction following each violation

with the presiding judge of a court devoted to HOPE probation as well as HOPE pretrial defendants.

Between September 2014 and August 2016, felony defendants who failed to make bail and who were granted supervised release were randomly assigned to either status-quo pretrial services or to the HOPE pretrial treatment group. We use administrative data on drug tests, revocations, supervised release case dispositions, and criminal history records to assess whether applying HOPE to individuals on pretrial supervised release impacts various measures of pretrial misconduct, criminal case disposition, and post-disposition arrests. Our findings are the following:

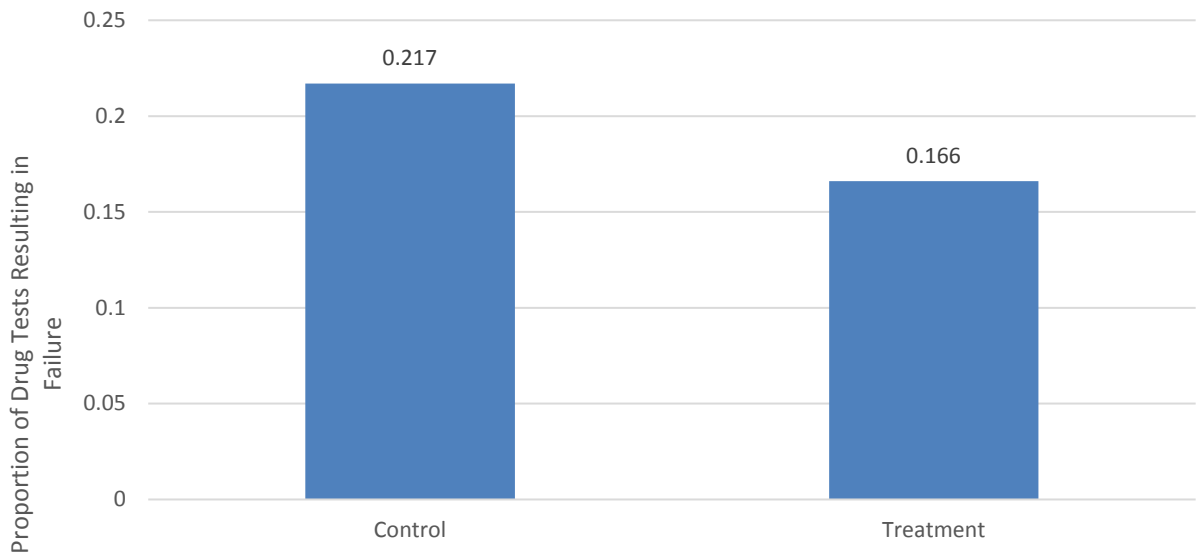
Figure ES1: The Average Number of Revocation Motions Filed and the Average Number of Permanent Revocations that End the Supervised Release Period



FINDING #1: HOPE treatment group members experience more pretrial supervised release revocations most of which are better characterized as modifications but fewer permanent revocations ending the supervised release term relative to control group members. Individuals assigned to the HOPE treatment group were significantly and substantially more likely to experience a revocation of their supervised release status. However, these revocations involved temporary HOPE revocations, referred to in HOPE court as modifications. Despite the higher number of average revocations/modifications, the average number of warrants issued per study subject was only marginally higher for treatment group members (with the difference relative to the control group not statistically significant). This likely reflects the fact that violations for HOPE group members often involve on-the-spot arrests for failing a drug test. Treatment group members were significantly and substantially less likely to experience a terminating revocation

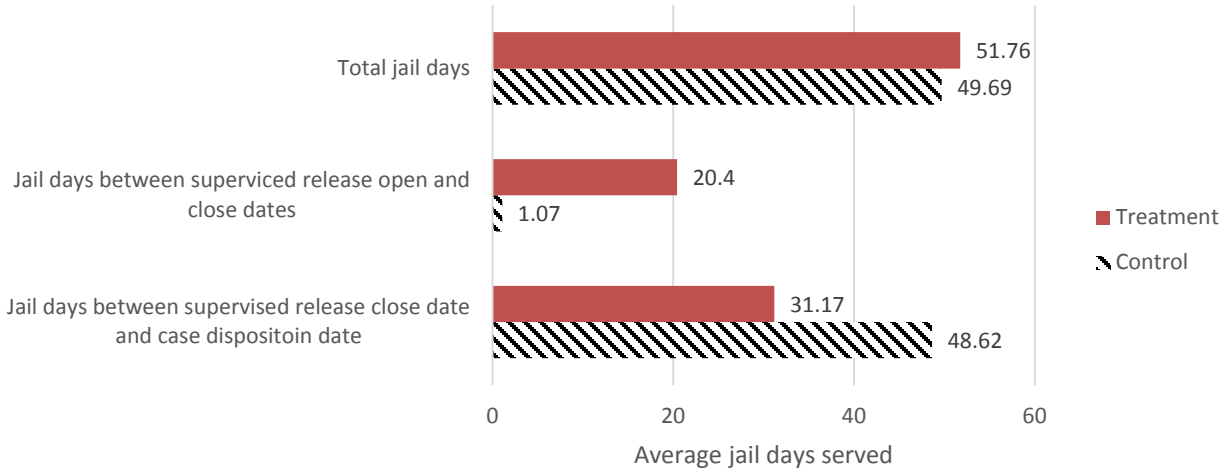
of their supervised release case (with the estimated reduction relative to the control group of roughly 60 percent). Total time on supervised release among treatment group members was 42 days longer on average relative to control group member (a difference of roughly 45 percent).

Figure ES2: The Proportion of Drug Tests Resulting in Failure for HOPE Treatment and Control Group Members



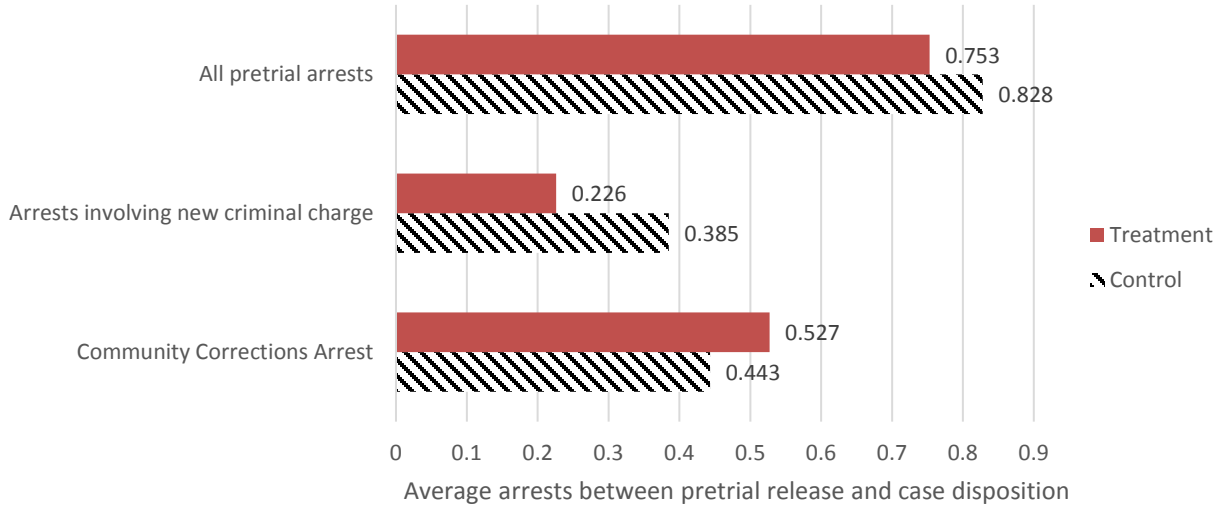
FINDING #2: Treatment under HOPE pretrial reduced the proportion of drug tests resulting in failure. The drug test failure rate for treatment group members was roughly 21 to 30 percent lower than the comparable failure rate observed for the control group with the difference statistically significant. Since random drug testing as well as the use of a call-in hotline for determining whether one had to report for a drug test on a given day was applied to members of both the treatment and control group, the lower failure rate for the treatment group likely reflects general deterrence associated with the swift and certain sanctioning policies of HOPE.

Figure ES3: Average Jail Days Served for the Entire Post-Supervised Release Period and for Sub-Periods



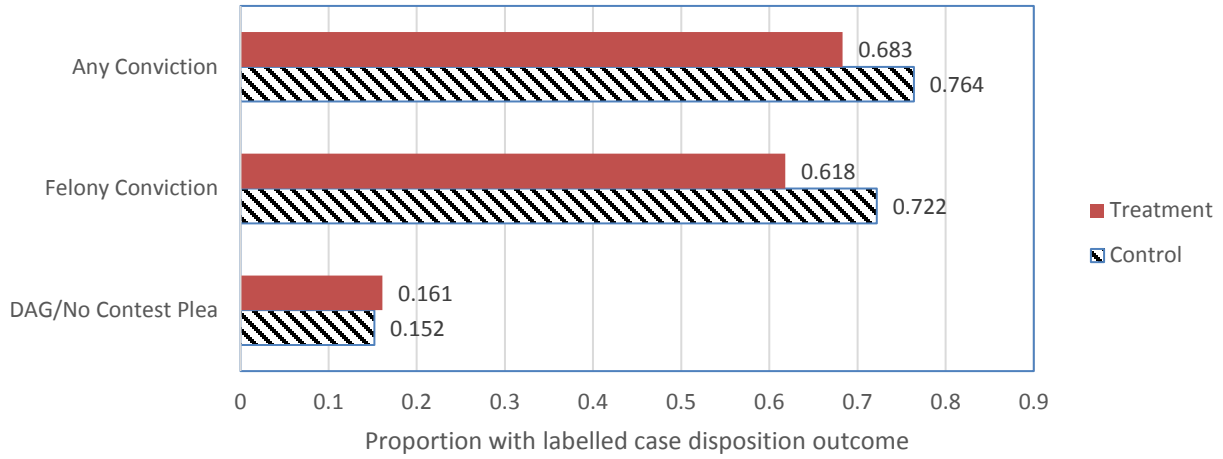
FINDING #3: HOPE treatment did not impact total jail days served between the supervised release date and the disposition date for the criminal case. However, treatment group members serve jail days earlier in their supervised release term while control group members serve more jail days later. We find no significant effect of the HOPE treatment on total jail days served between the supervised release date and the case disposition date. However, we do observe a significant and substantial increase in jail days served for the treatment group between the supervised release open date and supervise release close date, and an opposing reduction in average jail days served between the supervised release close date and the case disposition date. In other words, HOPE shifts jail incarceration days to earlier in the pretrial supervision period but has no overall effect on average jail days served pretrial.

Figure ES4: Average Arrests, Total, Arrests Involving a New Charge, and Arrests Involving Community Corrections Violations or Contempt of Court Charges Only



FINDING #4: Average total pretrial arrests occurring after supervised release does not differ significantly between the treatment and control group. However, treatment group members are significantly and substantially less likely to be arrested with a new criminal charge. We find no effect of treatment on total new arrests during the pretrial period. The null effect on total arrests however masks important compositional differences by arrest type. HOPE treatment group members are arrested for a new offense at a rate that is roughly 40 percent lower relative to the control group. There is a statistically significant decline in pretrial arrests where the most serious charge is a felony. Regarding the nature of new arrests, we observe statistically significant declines in property offenses, and in other offenses (e.g., non-person, non-drug, non-property offenses). We observe an offsetting increase in arrests for treatment group members where arrest charges are limited to community corrections violations (such as getting arrested on-the-spot for failing a drug test) and contempt-of-court charges.

Figure ES5: Criminal Case Disposition Outcomes



FINDING #5: Treatment group members are less likely to be convicted and less likely to be convicted for a felony. Several recent high quality quasi-experimental studies have found a causal effect of pretrial detention on the likelihood of conviction (Dobbie, Golden, and Yang 2016; Heaton, Mayson, and Stevenson 2017). This may occur through several channels. Defendants may be better able to mount a defense while not incarcerated. Alternatively, doing well on supervised release may lead to leniency by the prosecutor, the presiding judge, or both. Treatment group members survive considerably longer on supervised release (by roughly 42 days relative to a control group average of approximately 100) and are less likely to experience a permanent closure of their supervised release case, and in turn be incarcerated on the case disposition date. These factors may impact disposition outcomes. Indeed, treatment group members were significantly less likely to be convicted (with an absolute difference relative to the control group of 8.1 percentage points, and a proportionate difference of 11 percent) and less likely to be convicted of a felony (with an absolute difference relative to the control group of 10.4 percentage points and a proportionate difference of 14.4 percent). We find no significant difference in the likelihood that the case results in a deferred acceptance of guilt (DAG)/ no contest plea.

FINDING #6: We do not find statistically significant effects of treatment on post-disposition arrest outcomes. At present, the post disposition observation period for our study subjects is at least 180 days for 92 percent of our subjects and at least 365 days for 84 percent of our sample. We use these subsets to test for an effect of HOPE pretrial on post-disposition arrests occurring within the state of Hawaii. While our point estimates suggest declines in arrests of all severity and nature for treatment group members relative to the control group, very few of the estimates are statistically significant.

1. Introduction

The United States is currently experiencing a sea change in its approach to criminal justice policy. Several states as well as the federal government have made deliberate efforts to reduce the use of incarceration and to enhance the use of alternative sanctions often managed through probation and parole. Effective management of individuals on some form of community corrections is key to reducing correctional populations. Probation and parole violations are quantitatively important sources of admissions into the nation's prisons and jails, with community corrections violations accounting for roughly 41 percent of prison admissions in 2016 (Carson 2018).

Criminal justice policy makers, practitioners, and researchers are also paying increasing attention to the issue of pretrial detention and whether current practices for determining who is and who is not detained during case processing could be improved upon. Approximately 65 percent of jail inmates in the United States are un-convicted, pretrial detainees (Zeng 2018) most of whom were either unable to make bail or engaged in one of various forms of pretrial misconduct resulting in the revocation of a conditional release. Roughly 60 percent of felony defendants are released pretrial with half of releases occurring through some sort of bonding and the other half being released on personal recognizance or non-financial conditional release (Cohen and Reaves 2007). Of the 38 percent detained through a case's conclusion, the overwhelming majority (32 of the 38 percent) is detained due to the inability to make bail.

For those held on bail there is mounting quasi-experimental evidence that pretrial detention may result in worse adjudication outcomes for criminal defendants. For example, Dobbie, Golden and Yang (2016) exploit cross-magistrate variation in pretrial detention practices

to estimate the effect of pretrial detention on various outcomes. The authors find significant and substantial increases in the likelihood of a conviction operating primarily through an increase in the likelihood of a guilty plea. The authors also find that pretrial detention causes significant reductions in current and future employment and earnings. Heaton, Mayson, and Stevenson (2017) analyze the effects of pretrial detention for misdemeanor offenses on case disposition outcomes, jail sentences, and future offending using a quasi-experimental research design. The authors find a substantial increase in the likelihood of a guilty plea, in the likelihood of a jail sentence, in the length of the sentence, and in the likelihood of future arrests associated with being detained pretrial. MacDonald and Raphael (2017) find that a California sentencing reform that narrowed racial disparities in pretrial detention in the City and County of San Francisco consequently narrowed racial disparities in conviction rates and average sentence length.

There are multiple possible strategies for reducing community corrections failures. These strategies commonly involve various combinations of enhanced social services, targeted enhanced monitoring and in some instances targeted reduced monitoring (as evidenced by the example of New York City probation provided in Schiraldi 2016), alternative sanctions strategies for less and more serious violations, and various degrees of direct involvement of judges in case management. By necessity, resource constraints require that more intensive intervention be restricted to the highest-needs and highest-risk cases. Hence, probation, parole, and pretrial monitoring involves some form of triage and an implicit tailoring of interventions to either specific portions of a probation/parole/pretrial service officer's caseload or tailoring to the specific needs and circumstances of an individual case.

The Hawaii Opportunity Probation with Enforcement (HOPE) strategy provides one blueprint for managing those convicted of a criminal offense in a community correction setting. The HOPE strategy in Honolulu has evolved considerably over the past fourteen years, from a small-scale experiment involving the highest risk probation cases in Honolulu to a key element of the continuum of probation supervision and services in Honolulu County. In its current form, HOPE serves high risk probationers who have not performed well at lower levels of supervision yet are not deemed in need of an intensive drug-court intervention or who cannot access the smaller drug court program due to resource constraints (Institute for Behavior and Health 2015, Alm 2016, Hawken 2016). HOPE probation involves three components: (1) a probation officer committed to rehabilitation and the use of evidence-based practices, (2) the active involvement of a patient judge who is knowledgeable about addiction and addiction treatment, and (3) swift, certain, consistent, and proportionate sanctions for technical violations such as failed drug tests, missed appointments, and failure to attend or complete treatment. From the start, probationers are encouraged to succeed, expectations are clearly articulated by a presiding judge, and subsequent failures as well as successes are promptly addressed in a swift and consistent manner in open court. The program also offers drug treatment for whoever requests it (and in some instances mandated treatment for frequent violations), cognitive behavioral therapy, and other targeted interventions for those who need them.

An experimental evaluation of the HOPE program in Hawaii found significant reductions in drug abuse (as measured by failed drug tests), new arrests, and the likelihood that probation ended with a prison admission (Hawken and Kleiman 2009). A 76-month follow up study found persistent reductions in future arrests, prison admissions, as well as prison time served, though

the effects were smaller relative to the earlier estimates for the first post-randomization year (Hawken et. al. 2016). Many localities and at least one state have implemented elements of the original HOPE strategy, with a host of non-experimental studies finding impacts similar to those from the original HOPE evaluation. However, two experimental evaluations of swift, certain, and fair demonstration programs (one multi-site demonstration and one single-site program) failed to find impacts of the intervention relative to probation as usual as practiced in the intervention localities (Lattimore et. al. 2016; O'Connell, Brent, and Visher 2016).

To date, HOPE-style demonstrations have been targeted primarily at individuals who have been convicted of an offense and who are serving probation or parole terms. With the exception of the program evaluated here, we are unaware of an attempt to apply HOPE principals to individuals conditionally released pretrial. Defendants are detained pretrial to minimize the likelihood of pretrial misconduct, ensure attendance at trial, and in some instances due to public safety concerns. Pretrial misconduct may include new criminal offending, substance abuse, failure to report to a pretrial service staff, or failing to appear for court dates and/or trial. While the use of commercial bail bonds provides an incentive for those who bail out to refrain from such misconduct, heavy reliance on commercial bail puts low-income/low-asset defendants from low-income/low-asset families at a serious disadvantage. Hence, there is demand for alternative non-incarceration forms of pretrial monitoring that may be effective in preventing pretrial misconduct for those unable to make bail.

In this project we evaluate the application of the case management and treatment delivery practices developed under the HOPE probation program to pretrial individuals who are conditionally released from jail subject to criminal justice supervision. The intervention applies

the active involvement of the HOPE judge, the HOPE warning (to be discussed in detail below), treatment resources, and swift, certain, consistent and proportionate sanctions to the management of pre-trial releases. In the jurisdiction we study (Honolulu, Hawaii), defendants on supervised release are typically monitored by pretrial officers located at the county jail. The revocation of supervised release typically occurs once a defendant has failed to comply several times with a set of pre-specified conditions, including but not necessarily limited to refraining from drug use and additional criminal activity, maintaining contact with the assigned pretrial officers, and making all court dates. The intervention we evaluate applies random drug testing in conjunction with swift, certain, consistent yet proportionate sanctions to pretrial misconduct. That is to say, misconduct is met with quickly administered arrest and re-incarceration, yet subsequent jail spells are proportionate to the seriousness of the violation. The intervention also includes drug treatment interventions for those who repeatedly fail drug tests (or who request treatment services) and direct interaction following each violation with the presiding judge of a court devoted to HOPE probation as well as HOPE pretrial defendants. The HOPE court typically monitors roughly 2,000 of the county's approximately 3,800 probationers on active felony supervision. Hence, the experiment studied here, with 190 pre-trial defendants assigned to HOPE treatment over two years and a steady-state caseload of roughly 50 treatment group members at any given time, grafted a small increase in cases onto the much larger caseload managed by the HOPE court.

Between September 2014 and August 2016, felony defendants who failed to make bail and who were granted supervised release were randomly assigned to either status-quo pretrial services or to the HOPE pretrial treatment group. We use administrative data on drug tests,

revocations, supervised release case dispositions, and criminal history records to assess whether applying HOPE to individuals on pretrial supervised release impacts various measures of pretrial misconduct, criminal case disposition, and post-disposition arrests. The following bullet points summarize our key findings.

- Individuals assigned to the HOPE treatment group were significantly and substantially more likely to experience a revocation of their supervised release status, most of which are better characterized as temporary modifications involving a short jail stay. Despite the higher number of average revocations, the average number of warrants issued per study subject was only marginally higher for treatment group members (with the difference relative to the control group not statistically significant). This likely reflects the fact that revocations for HOPE group members often involve on-the-spot arrests for failing a drug test that do not require the serving of a warrant.
- Treatment group members were significantly and substantially less likely to experience a terminating revocation of their supervised release case (with the estimated reduction relative to the control group of roughly 60 percent). Time on supervised release among treatment group members was 42 days longer on average relative to control group members (a difference of roughly 45 percent).
- The drug test failure rate for treatment group members was roughly 21 to 30 percent lower than the comparable failure rate observed for the control group. Since random drug testing as well as the use of a call-in hotline for determining whether one had to report for a drug test on a given day was applied to members of both the treatment and control group, the lower failure rate for the treatment group likely reflects general deterrence associated with the swift and certain sanctioning policies of HOPE.
- We find no significant effect of the HOPE treatment on total jail days served between the supervised release date and the case disposition date. However, we do observe a significant and substantial increase in jail days served for the treatment group between the date of supervised release and supervise release close date, and an opposing reduction in average jail days served between the closing of a supervised release case and the case disposition date. In other words, HOPE shifts jail incarceration days to earlier in the pretrial supervision period but has no overall effect on average jail days served pretrial.
- We find no effect of treatment on total new arrests during the pretrial period. The null effect on total arrests however masks important compositional differences by arrest type. HOPE treatment group members are arrested for a new offense at a rate that is roughly 40 percent lower relative to the control group. There is a statistically significant decline in pretrial arrests where the most serious charge is a felony. Regarding the nature of new arrests, we observe statistically significant declines in property offenses, and in other offenses (e.g., non-person, non-drug, non-property offenses). We observe an offsetting increase in arrests for treatment group members where arrest charges are limited to

community corrections violations (such as getting arrested on-the-spot for failing a drug test) and contempt-of-court charges.

- Treatment group members are approximately 10 percent less likely to be convicted and 14 percent likely to be convicted of a felony.
- At present, the post disposition observation period for our study subjects is at least 180 days for 92 percent of our subjects and at least 365 days for 84 percent of our sample. We use these subsets to test for an effect of HOPE pretrial on post-disposition arrests occurring within the state of Hawaii. While our point estimates suggest declines in arrests of all severity and nature for treatment group members relative to the control group, few of the estimates are statistically significant.

2. Mechanisms of HOPE that May Impact Outcomes and a Review of Relevant Research

HOPE may impact pretrial misconduct through a number of channels. First, swift and consistent sanctions may create general deterrence effects. Criminologists and other social scientists who study criminal offending commonly distinguish between the certainty of a sanction, the celerity or swiftness with which a sanction is administered, and the severity of a sanction (see for example the literature reviews of deterrence by Chalfin and McCrary (2017) and Nagin (2013)). Rational choice theory tends to emphasize the expected value of punishment (the product of the likelihood of being sanctioned multiplied by some operationalization of the sanction cost) and conceptualizes offenders as calculating decision-makers who weigh the relative costs and benefits of their actions.[1] In this theoretical framework, severe penalties and the apprehension probability are effectively interchangeable, with perhaps some allowance for discounting associated with the timing of the rewards vs. the potential costs. Swiftness matters only insofar as it alters the timing of punishment and thus the discounted present value of the penalty. While HOPE sanctions are modest, random drug testing and the uniform and consistent

[1] See for example the seminal modeling of criminal offending by Becker (1968) and the extension of the Becker model to incorporate temporal discounting presented in Polinsky and Shavell (1984).

response to violations clearly increase the likelihood of a sanction and thus the expected value of punishment.

Alternative behavioral models incorporate the effect of extreme present orientation caused by time inconsistent preferences (Lee and McCrary 2009, Cook 2016), visceral stimulation associated with anger or intoxication (Nagin and Pogarsky 2004), or responsiveness among potential offenders to the situational availability of criminal opportunities that may induce impulsive, spur-of-the-moment criminal offending (Clarke 1995, Dominguez-Rivera 2018). In these models, the certainty and celerity of HOPE sanctions may be particularly impactful as an immediate consequence may be sufficient to overcome myopia and counter impulsivity. Moreover, the reading of the initial HOPE warning and the observation in open court of consistent and swift application of sanctions (and in some instances rewards) in the modification hearings of others may augment the salience of the potential sanction, and perhaps yield greater compliance with one's release conditions.

Aside from the impact of sanctioning, several elements of HOPE are designed to reduce drug use. Random testing, swift and certain responses to a failed test, and targeted drug treatment, to the extent that they reduce drug use and that drug use is in and of itself criminogenic, may increase compliance with supervised release conditions and reduce subsequent criminal offending.

Finally, some argue that sanctioning regimes that place such heavy emphasis on drug testing may be inadvertently punishing individuals in community corrections for what is essentially an illness over which they have little control and volition. In this view, zero-tolerance for technical violations may set unrealistic expectations and perhaps undermine the credibility of

a local community corrections system (Cullen, Pratt, and Turanovic 2016). To be sure, HOPE as implemented in Hawaii is much more than swift and certain sanctions. The HOPE model as articulated in Alm (2016) places great emphasis on the agency of the HOPE clientele, procedural justice, drug treatment, cognitive behavioral therapy, and other specialized therapeutic interventions targeted towards individual needs. Moreover, even when sanctioned for relapses, HOPE participants are able to influence the severity of the sanction through their personal actions (owning up to the violation, showing up following a missed appointment etc.), and sanctions for drug use do not graduate and do not prompt permanent revocations. Nonetheless, in community corrections settings that tend more towards services and alternative sanctions, and that are decisively moving away from intensive supervision and incarceration, a HOPE style intervention may be perceived by probationers as well as probation officers as unduly harsh and counter-productive (Schiraldi 2016).

The broader body of research on criminal deterrence tends to find significant effects on criminal offending of interventions that increase the certainty of a sanction but not necessarily the severity of a sanction. For example, the quasi-experimental research testing for an impact of police staffing levels on crime rates based on city-level panel data regressions (Chalfin and McCrary 2018), incident-induced surges in policing levels (Di Tella and Schargrodsky 2004, Klick and Tabarrok 2005), and experimental variation in policing levels targeted at crime hot-spot (National Academies of Science, Engineering, and Medicine 2018) all find significant and substantial effects of police presence (and, by extension the likelihood of apprehension) on crime rates. On the other hand, reviews of the empirical research on the effects of the death penalty (Chalfin, Haviland, and Raphael 2012, Nagin and Pepper 2012, Donohue and Wolfers 2010),

reaching the age at which one is automatically tried as an adult (Lee and McCrary 2005, Hjalmarsson 2009), and sentencing enhancements (Raphael and Ludwig 2003, Helland and Tabarrok 2007), tend to find that enhancing sentencing severity has small and often unmeasurable effects. These findings may be due to incorrect beliefs or imperfect information regarding sanction severity (Hjalmarsson 2009), an extreme present orientation (Lee and McCrary 2009), or some combination thereof.

There is a growing body of research evaluating the impacts of the original HOPE intervention as well as a slate of HOPE-inspired programs that have been implemented throughout the United States. All of the studies evaluate the total effect of the various practices associated with a given intervention rather than the individual impacts of specific practices (that is to say, the evaluations cannot distinguish the separate effect of say random drug testing from swift sanctions). Nonetheless, there is variation in implementation details across localities that hint at which mechanisms are particularly impactful and important to the success of an intervention.

Three of the evaluations commonly discussed together randomize probationers to either a HOPE or swift-and-certain treatment group or to a group subjected to standard probation practices in the jurisdiction under study. The remaining studies employ a host of quasi-experimental estimation strategies. These strategies range from simple pre-post comparisons for individuals who started out in regular probation and moved into a HOPE-type intervention, to propensity score matching of treated probationers to individuals from the general probation caseload, to panel data studies that exploit the staggered roll-out of an intervention across counties located within a given state. The non-experimental studies are more consistent in

finding significant and substantial treatment effects, though the original randomized control trial (RCT) evaluation of HOPE as well as the results from one of the replications sites also yield significant results.

The tendency in empirical literature reviews is to place greater weight on the findings from randomized control trials, either implicitly in the narrative of the review or in some instances formally in meta-analyses of extant evaluations that deliberately down-weight the findings from quasi-experimental research. However, there are many examples of high-quality non-experimental evaluations as well as examples of RCT evaluations that either encounter implementation problems or where the intervention may differ significantly from the policy model motivating the replication effort. Below, we present independent discussions of these two bodies of research and try to articulate the relative strengths and weaknesses of each in contributing to our understanding of the effectiveness of HOPE and HOPE-inspired interventions.

Hawken and Kleiman (2009) present a quasi-experimental evaluation of the original HOPE intervention pilot in Honolulu. The study also evaluates a subsequent randomized control trial that drew study subjects from the general probation caseload. The initial quasi-experiment involved non-random assignment of the highest risk probationers serviced by the Oahu Integrated Community Sanctions Section (ICSS). Probationers were ranked according to risk measures with relatively high-risk probationers assigned to the original HOPE intervention and the comparison group assigned to services as usual in ICSS. HOPE probationers, while having higher drug test failure rates relative to the comparison group during the three months preceding the intervention, had test failure rates less than one-third the rate for comparison group members three months post intervention and less than one fourth the rate for comparison group

members at the six-month, post-intervention mark. The HOPE probationers were also substantially and statistically significantly less likely to miss appointments and to have probation permanently revoked. The evaluation does not report the number of probation modifications, the term used for the temporary revocations deployed for treatment group members. While there was no difference in jail days served, there were substantial difference in ultimate prison sentences, with HOPE group members on average sentenced to roughly 190 fewer prison days (a 47 percent reduction relative to the comparison group).

The randomized control trial took place in 2007 and drew from the active caseload under community supervision by the Adult Client Services Probation unit in Honolulu. The study randomized 493 current probationers into either the HOPE treatment group or a probation-as-usual control group. Confirming the findings from the non-experimental study, the treatment group members missed appointments at less than half the rate of control group members, failed drug tests at a rate slightly greater than one-third the rate for the control group, and were roughly half as likely to be rearrested and to have their probation revoked. The treatment group also experienced less incarceration, receiving average incarceration sentences equal to roughly half that of the control group (with the difference in average sentenced prison days of 129 relative to the control group average of 267).

In a follow-up evaluation, Hawken et. al. (2016) analyze the long-term outcomes for the original quasi-experimental HOPE study subjects as well as the subjects from the RCT. The authors follow the non-experimental subjects for roughly ten years and the randomly-assigned subjects for 76 months. For the non-experimental subjects, those assigned to HOPE are less likely to acquire new criminal charges relative to the comparison group (0.19 charges for the HOPE

subjects compared with 0.78 among comparison subjects, with the difference statistically significant at the one percent level). HOPE subjects were also incarcerated 148 fewer days relative to the comparison group, though this difference is not significant. Regarding the 76-month results for RCT study subjects, HOPE treatment group members were half as likely to acquire new drug-related charges (with the difference significant at the one percent level). However, there were no significant difference in the likelihood of new charges for property or violent offenses. Finally, over the 76 months follow-up period, HOPE treatment group members were slightly less than half as likely to be returned to prison.

There are a number of non-experimental studies of programs that were either inspired by HOPE or that evolved independently but with elements quite similar to HOPE. For example, the Special Sanctions Court created by the Fort Bend County judiciary in Texas combined active case management by a sanctions judge, frequent court hearing especially early in the probation term, and sanctions that were delivered swiftly and that intensified or moderated with the behavior of the probationer. Snell (2007) conducted an empirical outcomes evaluation of the Special Sanctions Court comparing program participants to all individuals on probation during a two-year period preceding the creation of the court. While the evaluation is non-experimental, the treatment and non-experimental comparison group are generally balanced on observable characteristics at baseline. This was true for demographic characteristics, the nature of the controlling offense, and criminal history. Program participants exhibited relatively lower rates of probation violations, with the differences relative to the comparison group substantial and statistically significant. Moreover, program participants committed fewer new offenses and were less likely to acquire a new conviction relative to the comparison group. Treatment group

members also completed more community service hours and survived longer time periods on probation without a violation.

The state of Virginia authorized and funded a pilot swift-and-certain probation program called the Immediate Sanctions Probation Program (ISPP). Pilot programs were conducted in four counties with programs commencing between November 2012 and January 2014. The Virginia Criminal Sentencing Commission (2016) conducted a quasi-experimental evaluation of the effort using statistical matching methods to select a comparison sample of 63 study subjects against which they compared key case outcomes for those who participated in the ISPP. The findings from this evaluation are mixed. While the ISPP study subjects were arrested and convicted for new felony offenses at a significantly lower rate than comparison subjects, ISPP participants were more likely to have probation revoked and be sent to prison. In short, the closer scrutiny of ISPP participants generated more revocations to prison despite greater compliance. The evaluation concludes that the ISPP program was more expensive per participant relative to probation as usual, primarily due to the greater likelihood of a subsequent prison sentence.

Kentucky launched the Supervision, Monitoring, Accountability, Responsibility and Treatment (SMART) Probation program in 2012. The demonstration applied many of the elements of HOPE to high risk probationers in seven counties, including frequent random drug testing, swift sanctions for technical violations determined during court hearings immediately following the violation, active involvement of judges, and provisions for graduating sanctions and mitigation of the intensiveness of supervision dependent on the degree of compliance and progress made by the probationer. A non-experimental evaluation presented by Shannon (2013) found that SMART treatment group members failed drug tests at a rate that was 40 percent that

of a matched comparison sample of non-SMART probationers. The study also found significantly lower rates of violations among SMART treatment group probationers for substance use, technical violations of probation, and the acquisition of a new charge. Notably, treatment group members score significantly higher on each of the eight level-of-service risk score/needs assessment dimensions at baseline. Hence, the HOPE treatment group members appear to have been non-randomly selected in a manner that increased the likelihood that high-risk clients were allocated to HOPE. This non-random selection should have biased the findings towards finding no effect or even a normatively negative effect of the intervention.

DeVall, Lanier, and Hartman (2014) present a non-experimental analysis of Michigan's Swift & Sure Sanctions and Probation Program. The pilot program was implemented in seven Michigan counties between October 2012 and September 2013. The Michigan demonstration incorporates key elements of the original HOPE program including intensive supervision, sanctions for each violation administered swiftly and determined through court hearings, graduated sanctions, targeted treatment for repeat violations, and uniform and consistent treatment of all probationers. For the outcomes evaluations, the study selects a comparison group based on propensity score matching on the precipitating offense and COMPAS risk scores. The matching procedure achieves balance along these dimensions between the treatment and control group. The outcome evaluation focuses principally on recidivism and revocations to jail and prison sentences. The authors find a 34 percent reduction in recidivism of any type for the treatment group relative to the matched comparison sample. In addition, 13.7 percent of the treatment group members receive a jail sentence as consequence of recidivism. This is substantially lower than the 21.6 percent of comparison group members that are sentenced to

jail. The percentage being revoked to prison are comparable across groups (4 percent for Swift & Sure treatment group members compared with 5.3 percent for the matched comparison sample).

Further non-experimental evidence regarding swift and certain sanctions comes from several evaluations of the South Dakota 24/7 Sobriety Project and its reproduction in Montana. Started in 2004, the 24/7 project requires those convicted of DUI offenses as well as those for whom drug and alcohol abuse are considered to be contributing factors to future recidivism risk to undergo either twice daily breathalyzer exams in the presence of law enforcement, continual monitoring of blood alcohol levels through the wearing of a specialized ankle bracelet, the wearing of a patch that can detect narcotic use, or frequent urinalysis testing. The twice-daily breathalyzer tests condition is the most frequently ordered testing regime. Violations due to alcohol/drug use or missing an appointment are met with immediate arrest and a short jail stay. Loudenburg, Drube, and Leonardson (2011) present a quasi-experimental evaluation where those ordered to comply with the 24/7 testing standards are compared to other individual with DUI convictions. The overwhelming majority of tests administered under the 24/7 program are negative (over 99 percent of tests), indicating a very high degree of compliance. The authors find that relative to the comparison group, those ordered onto 24/7 are significantly less likely to experience a subsequent DUI and experience longer duration on average until the next DUI arrest.

Kilmer et. al. (2013) exploit variation in the roll out of 24/7 across South Dakota's 66 counties occurring over the period 2004 through 2010, effectively using late-adopting counties as a control group for early-adopting counties to estimate the effect of the intervention on a host

of outcomes. This is a particularly strong methodological strategy that has been successfully deployed to study many social phenomena including the long-run effects on children of access to the social safety net (Hoynes, Whitmore-Schanzenbach, and Almond 2016) and the long-run effects of school desegregation on the adult socioeconomic outcomes of African-Americans (Johnson 2011). Kilmer and his colleagues employ several alternative thresholds for an indicator variable measuring substantive use of the 24/7 program in a given county based on the proportion of DUI arrest resulting in 24/7 testing conditions for conditional release or the allowance of work permits. Using county-level panel data regression models inclusive of fixed effects for county and year, the authors test for an effect of 24/7's introduction in a given county on repeat DUI's, domestic violence incidents, the number of traffic crashes, and the number of traffic crashes involving men 18 to 40 years of age. The authors find substantial and significant reduction in repeat DUIs, domestic violence incidents, and accidents involving young men when 24/7 becomes operational in a given county. The results are robust to alternative measures of gauging local use of the program. Midgette and Kilmer (2015) provide an analysis of microdata to evaluate the introduction of 24/7 testing practices for chronic DUI offenders in Montana. Analyzing data for a twelve-month follow up period, the authors find a reduction in the likelihood of a third DUI arrest on the order of 45 percent relative to the comparison group.

A final notable non-experimental study evaluates a statewide swift and certain probation program. Hamilton et. al. (2016) present a quasi-experimental evaluation of the statewide introduction of a swift and certain probation sanctions regime in Washington State. Motivated in part by the earlier HOPE evaluation results as well as a pilot intensive supervision program carried out within the state, the state of Washington passed legislation that altered policy for

monitoring and sanctioning those on probation in a manner patterned after HOPE. Specifically, low-level violations are given a pass the first time though the probationer must formally acknowledge the consequences for subsequent violations. Subsequent low-level violations are punished by three-day jail spells through five such violations. Thereafter, all violations are considered high-level and punishable by up to thirty days in jail. Low-level confinements can be initiated by probation officers subject to a supervisor's approval and do not require a sanctioning hearing.

The study defines three non-experimental groups: a treatment group comprised of those entering probation immediately following the change in policy up to the date when the violation of absconding is redefined as a high-level violation, a treatment group for subsequent probation entries with the stricter definition applied to absconding, and a non-experimental comparison group comprised of those entering probation in 2010 and 2011 (years preceding the implementation of the program) for which there is a full year follow-up period preceding the change in policy. Cook (2016) and Nagin (2016) raise concerns about the coincidence of the timing of the legalization of recreational marijuana use in the state with the statewide implementation of the change in the sanctioning regime. In particular, the comparison group observation period corresponds to a time period prior to legalization while the observation period for the two treatment group occurs after legalization. This may bias the results towards finding better outcomes for the treatment group to the extent that marijuana use was an important source of revocations in the past. Information on the relative importance of marijuana violations during the earlier time period is not clear from the results presented in the study.

The authors use the elements of Washington State's Static Risk Offender Needs Guide-Revised to match non-experimental comparison group members from the pre-reform period to the two treatment groups defined for the post period. Study subjects who enter probation under the swift and certain regime are less likely to acquire a violation resulting in confinement (28 percent for the comparison group compared to 24 percent for the treatment group), experience fewer jail incarceration days and fewer prison incarceration days during the twelve-month follow up period, and are significantly less likely to be convicted of a felony offense. The recidivism effect primarily reflects a reduction in violent felony and property felony convictions.

As the review thus far indicates, the non-experimental research evaluating various implementations of swift and certain probation programs for the most part tends to reinforce the initial findings from Hawaii. The combination of swift yet proportionate sanctions, the active involvement of judges, random drug testing, and targeted therapeutic interventions consistently reduce the incidence of failed drug tests and recidivism. The evidence pertaining to incarceration days is more mixed as is the effect of these interventions on the number of revocations. In some evaluations (the Virginia demonstration in particular), the intensive supervision yielded both greater compliance among probationers as well as higher incarceration levels. This suggests that closer scrutiny and sanctioning may improve behavior yet create more opportunities to revoke, suggesting that avoiding such an outcome requires commitment from the presiding judge to be restrained in the use of permanent revocations.

As a result of this body of work, several research teams launched rigorous RCT evaluations of swift and certain interventions, one of which made explicit efforts to reproduce the structure and sanctions of HOPE across four sites while the other would better be described as

experimenting with swift and certain sanctions implemented entirely by probation officers. Beginning with the latter, O'Connell, Brent, and Visher (2016) evaluate the impacts of a swift and certain probation intervention implemented in a mid-sized city in Delaware. Unlike the original HOPE program, the Decide Your Time (DYT) intervention was contained entirely within the probation department. Upon assignment to the treatment group, new subjects were informed of the interventions elements and given two weeks to prepare a sobriety plan. They were then subject to frequent random drug testing that declined in frequency with 90 days of compliance but stepped up in frequency for non-compliance. Non-compliance was also met with a four-day incarceration spell and mandatory drug treatment. Further non-compliance was sanctioned by an imposed curfew and additional and longer incarceration spells. Consistent compliance was rewarded regardless of past non-compliance.

The DYT treatment group members missed significantly more appointments than control group members (0.99 appointments for treatment group members compared with 0.47 for the control group). However, they were also subjected to many more drug tests on averages (11.2 for treatment compared to 1.53 for the control), and hence were exposed to many more opportunities to miss an appointment. Treatment group members racked up more failed drug tests, an unsurprising finding given the relatively high frequency of testing for this group. However, the proportion of tests failed by treatment group members was significantly and considerably lower (0.50 for the treatment group relative to 0.86 for the control group). There is no statistically significant difference in the likelihood of successfully completing probation

The authors test for an effect on various recidivism outcomes including whether the individual is arrested regardless of reason, being arrested for a new crime, acquiring a probation

violation, acquiring a probation violation without committing a new crime, and whether the subject experiences any incarceration. It's not clear from the discussion whether incarceration as measured by the authors reflects sanctions served in the probation violation center for technical violations or sanctions associated with the permanent revocation of probation and subsequent lengthier prison and/or jail terms. All outcomes are measured at six months, twelve months, and eighteen months following random assignment. There is no evidence of a significant treatment effect in any comparisons, with control and treatment group means quite close to one another.

There are of course several important differences between HOPE as implemented in Hawaii and the DYT demonstration. The most obvious difference is the lack of involvement of the judiciary and the fact that the sanctions were administered at the initiation and by the power of probation officers. In a reaction essay to this study and the study we will discuss next, Judge Steven Alm (2016) noted that repeated technical violations beyond a fixed number would automatically trigger permanent probation revocations in the locality studied. This fact in conjunction with the omission of modification hearings and frequent interactions with the presiding judge may reflect a lack of buy-in from the judiciary. These program implementation aspects may also have undermined the sense of fairness and perhaps the degree of procedural justice perceived by probationers assigned to DYT.

Lattimore et. al. (2016) evaluate a randomized control trial of an intervention designed to reproduce the original HOPE program in four mainland counties: Saline County Arkansas, Essex County Massachusetts, Clackamas County Oregon, and Tarrant County Texas. The NIJ-sponsored replication was dubbed the Honest Opportunity Probation with Enforcement Demonstration

Field Experiment (HOPE-DFE). The intervention involved random assignment of new probation cases, those on probation for less than six months, as well as those with at least one-year left on their probation term to either a HOPE-like treatment group or a probation-as-usual group. Treatment group members were read a HOPE warning following assignment, were subject to frequent random drug testing with step-downs and step-ups in frequency dependent on compliance, and swift and graduated sanctions associated with non-compliance determined via violation hearings occurring shortly after the violation.

The study finds no statistically significant difference in the number of overall arrests between treatment and control group members over an observation period slightly under two years in length. However, arrests involving a property crime decline by a statistically significant five percentage points (relative to a control group mean of 0.20) while arrests resulting in a drug charge decline by a statistically significant three percentage points (relative to a control group mean of 0.15). Interestingly, the lower arrests rates for property crimes among treatment group members is similar across sites, although the difference for any one site is never quite statistically significant. The significant effect on drug arrests is driven entirely by a large treatment effect on drug arrests in Texas (9 percent for the treatment group relative to 17 percent for the control group).

There are no significant differences in the likelihood of a probation revocation or in various gauges of new recidivism convictions. In two of the study sites, HOPE subjects experienced probation revocations at a rate more than double that of control group members. The study does not specify whether the measure of revocation includes revocations that were swift and certain responses to technical violations as opposed to permanent revocations to

prison. Aside from recidivism outcomes, the authors also estimate duration models to test for an impact of treatment on time to first arrest, revocation, and conviction. In the two sites where treatment causes an increase in revocations (Arkansas and Oregon), assignment to treatment also shortens the time to the first revocation and the first conviction.

The results from the HOPE-DFE are certainly disappointing. Nonetheless, several have raised issues that may in part explain the null findings from the replication. Alm (2016) notes that there was some resistance to the HOPE pilot by probation officers in one of the four counties (Clackamas County Oregon) that may have impacted the effectiveness of the demonstration. Schiraldi (2016) notes similar resistance in New York City to his effort as the NYC Commissioner of Adult Probation to implement HOPE that ultimately undermined his effort (though in retrospect Schiraldi sides with his probation officers due to his belief that HOPE would have undermined ongoing reform efforts in the department).

Hawken (2016) offers several alternative possible explanations for the differences between the HOPE-DFE findings and the original HOPE evaluations from Hawaii. First, the probation departments that selected into the HOPE study may have been a non-random set, with Hawken suggesting that the four departments studied were well run and likely among the more innovation and best-practices oriented in the nation. In other words, to the extent that the best-run probation department selected into the study via the application process, and to the extent that the potential effect of HOPE are heterogeneous and larger in less innovative departments, site selection bias may tend towards smaller or no impacts of HOPE.

Second, the local criminal justice systems in the HOPE-DFE may differ from that of Honolulu's in a manner that may limit the effects of the HOPE intervention as implemented in

Hawaii. For example, Hawken notes that in Hawaii repeated punishments for minor technical violations are rarely met with permanent revocation of probation, and that HOPE in Hawaii limits the application of a revocation and subsequent open prison terms to those deemed to pose a public safety threat (for absconding, for example). In the four areas studied, there was certainly room for judges frustrated with repeat violations to revoke parole as a consequence and perhaps even local formalized practice that mandates revocation following a documented number of violations.

This discussion elucidates the tension associated with enhanced monitoring, such as more frequent random drug tests. As shown by Petersilia and Turner (1993), closer scrutiny and more intensive supervision is likely to uncover more violations and may lead to more revocations even when intensified supervision and targeted services create the intended rehabilitative and deterrent effects. Hence, absent careful application of sanctions, members of HOPE may be simultaneously exhibiting greater compliance on average while experiencing more incarceration and a higher likelihood of a revocation.

Finally, Hawken raises concerns regarding how the consequences of failure on probation vary considerably from locality to locality and that this may ultimately impact how HOPE impacts local criminal justice outcomes. For example, a probation failure in Hawaii would result in the imposition of an open prison term, sometimes decades in length. Hence, the consequences of permanently ending probation were quite severe. It's not clear that revocation in the reproduction sites necessarily resulted in lengthy prison sentences, or that they would in less punitive settings such as that described by Schiraldi (2016) during his time as commissioner of probation in New York City.

3. Description of HOPE Pretrial and the Status Quo

The HOPE pretrial pilot program builds on the lessons learned from HOPE probation in Honolulu to test an alternative model for pretrial supervision of felony defendants. This pilot effort involves a collaboration between the First Circuit of the Hawaii Judiciary, the Honolulu branch of Hawaii's Office of Pretrial Services, the Honolulu Prosecuting Attorney, and the Honolulu Public Defender. The target population includes all individuals charged with a felony who fail to make bail and are then subsequently released via the Honolulu supervised released program. In this section, we describe the pretrial supervised release process as well as the manner in which treatment under HOPE pretrial differs from standard practice in the jurisdiction studied.

A. Current Standard Practices Pertaining to Supervised Release

Following a felony arrest and booking into the Oahu Community Correctional Center (OCCC, the Honolulu jail), the defendant's first court appearance occurs at arraignment before an administrative judge. In addition to the formal filing of charges, the administrative judge sets a bail amount and, in some instances, specifies conditions that must be abided by if one makes bail. Those who do not make bail are held in OCCC until their first hearing with the judge assigned to the case. At this initial hearing (as well as in subsequent hearings), the defense may file a motion for pretrial supervised release that the presiding judge may either grant or deny. If granted, the judge will often specify conditions that must be met, such as identification of a

sponsor, showing up for all court dates, refraining from illegal drug use, and refraining from activity likely to generate a new arrest and/or criminal charge.

Individuals placed on supervised release meet with a pretrial officer located at OCCC. During this meeting, the pretrial officer informs the defendant about expectations, conditions that must be abided by, and notes that they may be called in for drug testing and other types of monitoring. Supervised releases are also informed that non-compliance will eventually result in a revocation and admission to jail where the defendant will remain until the criminal case reaches a disposition. The pretrial services officer is in charge of monitoring the behavior of the defendant and ultimately preparing the affidavit for a motion to revoke for defendants that are repeatedly in violation. Importantly, discretion regarding whether and when to prepare such an affidavit lies with the pretrial officer in charge of a given case. The actual motion to revoke however is filed by the prosecutor, a factor that adds delay to the time between the decision of the pretrial officer to pursue a revocation and the actual filing of the revocation in court. Interviews with staff members at OCCC suggest that revocations usually occur following several failures to comply.

B. *How HOPE Pretrial Differs from Standard Practice*

Treatment under the HOPE pretrial pilot commences after the point that a motion for supervised release is granted. Individuals are still released following a consultation with a pretrial officer. However, those randomly assigned to HOPE must report to the courtroom of the first Circuit judge in charge of HOPE for a scheduled hearing to receive the “HOPE Warning.” Judge Steven Alm served as the HOPE court judge through most of our evaluation period through his

retirement in August 2016. Afterwards, the HOPE probation and pretrial program were taken over by Judge William Domingo.

The HOPE Warning involves a description delivered by the judge in open court of what is expected of pretrial supervised releases and what will happen if they test positive for drugs, miss an appointment or a scheduled court date, or engage in some other form of pretrial misconduct.[2]

---

[2] Hawken and Kleiman (2009) include the text of a sample warning used in HOPE probation. The following text, taken from Hawken and Kleiman, accords with the observations of HOPE warning observed by the PIs in several days of court observations. The warning for HOPE pretrial is similar though modification are made for the difference in sanctioning.

"Good morning. I am Judge _____. You are here because we believe that you can be successful on probation, rather than being incarcerated at Halawa or in Mississippi or wherever they are now sending folks. But you are also here because you haven't been doing your part and following the rules of probation, and the probation officer thinks you are headed for a revocation. I hope you do succeed on probation. So does your lawyer, your probation officer, and your family. I think you can succeed on probation, or I wouldn't have put you on probation to begin with. But to do so, you must act responsibly. You are the one responsible for making sure that you comply with your conditions of probation. When you are on probation rather than being sent to prison, you are making a deal with me to follow the rules. Hopefully, you will learn that the more responsible you are, the more freedom you will have. The less responsible you are, the less freedom you will have."

"The 3 things I am immediately concerned about for you on probation are illegal drug use, meeting regularly with your probation officer, and complying with the other conditions of your probation, like going to drug treatment, etc.[If drugs are an issue] You have to call the drug test hotline every weekday morning [441-8962]. If you miss a drug test a warrant will be issued immediately. If you are using drugs, you are breaking the law, you are violating your probation and, if you are in treatment, it's not working. Unless you recently found some cash on the sidewalk or inherited some money, continuing to test positive for drugs also means that crimes are probably being committed by someone in order to get the drugs. If you miss an appointment with your probation officer, it tells me one of 3 things:"

"1) you know you will test positive;
2) you are doing something you shouldn't be doing; or
3) you are blowing off the probation officer."

"All 3 are bad. You are being brought here to court today so I can clearly spell out what the consequences will be if you don't follow the rules of probation. From now on, if any of these things happen -- if you fail a drug test, if you fail to meet with your probation officer when you are supposed to, or you fail with other terms of your probation, such as not getting an assessment, not going to treatment, etc. -- you will go to jail. If you test positive, you will be arrested on the spot, held in custody, and we will have a hearing two days later. If you used drugs, you will go to jail. If you missed a drug test or a scheduled appointment or don't comply with other conditions of probation, I will issue a bench warrant for your arrest immediately, and HPD's SSD (SWAT Team) or Crime Reduction Unit (CRU) officers will arrest you. They have agreed to serve the warrants for me. They won't come alone. They will arrest you at work or home or wherever. That would be embarrassing and folks may get hurt. It is better to just come in even if you violated. If you did, you will go to jail but not for as long as you will if we have to find and arrest you. I understand that things happen in life. If your car breaks down on the way to the probation office, push it to the

The warning involves a description of a system of proportionate and immediate sanctions, the services available to the defendant, and the stated desire on the part of the judge that the person succeed on HOPE. For HOPE pretrial, the potential sanctions include a few hours in the court cellblock, a week in jail, two weeks in jail, and revocation of pretrial supervised release and readmissions to jail for an incarceration spell to last at least through the disposition of the criminal case. Sanctions are more severe for individuals who fail to take responsibility for their violations. For example, individuals on HOPE whose urine samples test positive on the spot face one sanction if they admit to the violation, yet a stiffer sanction if they insist that the sample be

---

side of the road, call your probation officer, tell her/him that you will be late, and get on the bus. If you or your child is at the Emergency Room, call your probation officer to reschedule your appointment and be ready to bring proof of the medical treatment when you come for that appointment. But apart from that type of thing, if you try to reschedule your appointment, I am expecting the probation officer to say 'no' and to notify me immediately if you miss the appointment so I can issue the warrant."

"All of your actions in life have consequences, good or bad. If you confront your problems and learn to change your thinking and your behavior, you will be able to follow the rules of probation and be able to remain free in society. Remember, responsibility brings more freedom."

"On the other hand, if you violate the rules, there will be consequences, and they will happen right away. But it's all about choices."

"You may now have daily responsibilities like a job, or a class, or you may have a special event coming up -- baby's first luau, son's football game, daughter's graduation, whatever. If you test positive, miss an appointment or otherwise don't comply, you will go to jail right away. And you may get in trouble with your job or miss that special event. But that will really be <u>your</u> choice. You didn't care about your job or that luau or that graduation when you got high or missed the appointment or didn't go to treatment. It wasn't important enough to you then. You made a <u>choice</u> to use. You are not 13 years old. You are an adult. You can make a choice <u>not</u> to use and to be responsible about seeing your probation officer when you are supposed to and complying with the other terms of your probation. Remember, these violations will <u>modify</u> your probation and will send you to jail. If the probation officer sees that you <u>continue</u> to violate the terms of your probation and that you are no longer suitable for being on probation at all, they may file a motion to <u>revoke</u> your probation. If I see such a <u>revocation</u> motion in front of me and the violations are proven, I may well give you the 5, 10 or 20 years in prison. The probation officers are my eyes and ears in working with you and supervising you outside of court. I rely on them and their judgment. If you are unable or unwilling to comply with the terms of probation, the place for you is prison. That will give the probation officers more time to work with those folks who <u>want</u> to be on probation, who want to change their thinking and change their behavior and learn to be responsible."

"Do you understand everything I just said? Do you have any questions for me? Good. I wish you luck and success on probation and hope I don't see you back here."

sent to a lab and if the laboratory results confirm the initial finding. Similarly, the jail sanction for someone who misses an appointment or court date but who self-reports within a day or two will be shorter than the sanction imposed on someone for whom a warrant is issued and served by law enforcement.

The warning is delivered and discussed with each new member of the HOPE caseload, with the judge requiring that each person receiving the warning acknowledge what is expected and how the program works, usually by repeating expectations back to the judge. In addition, individuals who comply, who have their drug testing frequency reduced, or who are released from probation as a result of progress made on HOPE are acknowledged in open court.[3]

New cases/clients are calendared to appear at a set time, but usually must sit through an hour or so of HOPE probation and pretrial modification hearings for existing cases/clients calendared at the same time before their names are called and the warning delivered. Hence, for many the warning follows observation of the outcomes of several hearings where sanctions are administered in a consistent manner across cases. A key aspect of the HOPE treatment concerns the swiftness with which sanctions are delivered and the transference of discretion from the pretrial officer to the presiding judge. A pretrial officer still manages the day-to-day monitoring of supervised releases assigned to HOPE. However, pretrial misconduct is met with immediate arrest and a subsequent modification hearing in open court.

An additional key aspect of HOPE involves random drug testing on short notice. Those on HOPE pretrial (as well as those on HOPE probation) must call a hotline number each morning to

[3] For the court sessions observed by the PIs, such successfully cases were usually accompanied by applause from family members and other individuals on HOPE sitting in court awaiting their hearings.

find out if they must come in for a drug test. A positive test results in their immediate arrest and booking into the local jail. If arrested on the spot they remain in jail until a hearing can be scheduled in HOPE court. An important deviation from HOPE probation is that HOPE pre-trial hearings are scheduled for one week following the arrest rather than two days. This difference is the result of the fact that the prosecutor must file the revocation motion prepared by the pretrial officer, creating a bottleneck and delay that does not exist in HOPE probation.[4]

In its initial incarnation, HOPE employed graduated sanctions for repeat violations. First-time violators received relatively light sanctions (a day or two in jail) while those who repeatedly violated received longer jail stays. However, the strategy soon evolved to maintain moderate sanctions for minor technical violations. According to Alm (2016), this change reflects reconceptualization of the underlying nature of technical violations, especially failed drug tests, as being indicative of a problem in need of treatment that is not necessarily responsive to stiffer sanctions. In addition, HOPE evolved to incorporate sanctions less severe than jail spells (for example, several hours in the court cell block).[5]

HOPE pretrial and HOPE probation cases are for the most part treated similarly with regard to sanctioning severity (with a key difference being the fact that hearings are scheduled seven rather than two days later for on-the-spot arrests for drug test failures). However, the consequences of an ultimate failure (revocation of pretrial release as opposed to being sentenced to an indeterminate term in state prison) differ across the two groups. Both types of cases are

---

[4] Probation officers can directly file motions to modify with the court and thus HOPE hearings are typically scheduled for two days following the violation and arrest.

[5] During two days of observing modification hearings, the evaluators observed instances where individuals who missed appointments yet turned themselves in right away were sanctioned with a few hours in a holding cell located near the judge's chambers.

calendared together, with probation cases observing pretrial HOPE hearings and pretrial cases observing HOPE probation hearings.

The final key aspect of HOPE pretrial concerns the availability of treatment resources for those who repeatedly fail drug tests or for those who request such services. Individuals who repeatedly test positive for proscribed substances may be compelled into a drug treatment program prior to a permanent revocation of supervised release. In essence, compliance with the randomized testing regiment is used to triage cases and ration treatment resources accordingly. This is a sharp contrast to drug court where all cases receive intensive treatment services. Probation terms are considerably longer than terms on pretrial supervised released (with probation terms ranging from two to five years dependent on probationer outcomes as compared to an average of 100 days on pretrial supervised release). Hence, one might expect the service delivery dose associated with HOPE as applied to pretrial to be lesser than the comparable dose applied to those on probation supervision.

C. *Modification to Standard Practice Coinciding with the Launch of HOPE Pretrial*

The initial demonstration intended to contrast pretrial supervised releases managed under the HOPE model to the standard practices employed in Honolulu. Prior to HOPE pretrial, individuals on supervised release were not subject to random drug testing based on calling into a hotline. It was decided to implement comparable random drug testing and the use of the drug-testing hotline for all supervised release cases with the launch of the pretrial demonstration. Hence, the treatment-control contrasts presented below reflect impacts of the intervention that are beyond any impact of random testing. Note, the sanctions structure for the control group

did not change relative to past practices. That is to say, individuals on supervised release were afforded multiple chances to comply before a permanent revocation was initiated.

We should also reiterate that minimum sanctions for an on-the-spot arrest is one week under HOPE pretrial as opposed to two days under HOPE probation. This difference results from the logistic complication created by the fact that while the pretrial officer prepares the affidavit for a revocation/modification the prosecutor files the actual motion with the court. This design difference should be kept in mind especially when interpreting the estimated impacts of the intervention on average jail days served.

4. Design of the Randomized Control Trial Evaluation of HOPE Pretrial

Our evaluation of HOPE pretrial employs a randomized control trial design. Study subjects were randomized into the treatment or control group on a rolling basis from late August 2014 through the end of August 2016. The point of randomization occurred following approval of a motion for supervised release filed by the defense. The research team in collaboration with the criminal justice practitioners in Honolulu collectively decided to randomize following the presiding judge's decision pertaining to supervised release. This choice was intended to avoid selection driven by any expectations the presiding judge might have regarding the effectiveness of the standard pretrial supervision practices relative to the HOPE pretrial experiment. Before discharge, pretrial service officers entered the names of each supervised release into a web-based application hosted by the Goldman School of Public Policy at the University of California, Berkeley. The application was designed to return a random assignment to either the treatment or control group and digitally record the assignment. The application also permitted the pretrial

officers that used the system to download a spreadsheet of all assignments to date and to display on screen the most recent assignments.

It was determined by intake services staff prior to the commencement of the trial that the pretrial officer assigned to HOPE could manage a caseload of 50 open cases at any given time. Given initial estimates of the monthly flow of supervised releases and the average duration on supervised release, the initial randomization rate to the HOPE treatment group was set to 17 percent of all cases. Over the first seven months of randomization, the total flow of pretrial supervised releases was much lower than originally estimated. Between September 2014 and March 2015 there were only 209 supervised releases, 38 of which were randomized to the treatment group. We proposed, and the Hawaii team agreed to, an increase in the randomization rate to 50 percent in early April of 2015. The randomization rate employed by the web-based application was thus adjusted accordingly on April 3, 2015 through the end of the randomization period on August 31, 2016. All HOPE cases were assigned to a newly-hired pretrial officer while the incumbent staff managed the cases of control group subjects.

In total 519 individuals were randomized to either the treatment or control group. Figure 1 displays cumulative assignments of unique individuals by month. A small number of individuals (19) were randomly assigned twice. The pretrial supervising officer investigated each case and informed the research team that for each instance the person in question failed on pretrial supervised release, was readmitted to jail, and then was subsequently released once again to a supervised release term by the judge presiding over the criminal case (hence, these observations had two supervised release motions granted on the same felony case). With the new term, the staff entered these subjects into the randomization app once again. In twelve of the cases the

individuals were assigned to the same study group for both randomizations. For seven however, the subjects were assigned once to the treatment group and once to the control group. Since our outcome domains include criminal case disposition and post-disposition recidivism, one cannot accurately define these seven observations as either members of the treatment or control group. Hence, we drop these seven subjects from the analysis.

In addition, to these problematic cases, there were three subjects that were not supervised released cases that were mistakenly entered into the randomization app. These subjects were dropped. In addition, there were three subjects for whom the supervising pretrial officer overrode the randomization to the treatment group. One was reversed due to concerns that the defendant in question was involved in a criminal case involving a death that received much media attention. There was concern that the new officer hired to oversee HOPE cases did not have sufficient experience to manage this case. The assignment of an additional subject to the treatment group was overridden due to the fact that the conditions of release placed on the individual by the presiding judge were less stringent than what was imposed under HOPE pretrial. The final subject involved an individual who was severely mentally ill who the pretrial officer deemed incapable of complying with HOPE. We drop these three cases from the study. Thus, our final sample consists of 506 observations (190 in the treatment group and 316 in the control group).

Table 1 compares average demographic characteristics for the treatment and control group members (we discuss the data sourced behind these tabulations shortly). The two groups have comparable proportions male have similar distributions cross racial and ethnic groups and are balanced with regards to the proportion veteran and the proportion that are U.S. citizens.

While the treatment group is slightly older on average (by 1.76 years with the difference statistically significant at the ten percent level of confidence), an overall test for balance on the demographic characteristics listed in Table 1 fails to reject the hypothesis that the two group are drawn from the same underlying population.

Table 2 compares the criminal histories of treatment and control group members as well as the nature of the current felony case generating the supervised release. For criminal history, we use automated criminal history records to tabulate the number of prior felony arrests, the number of prior arrests for a community corrections violation, the number of prior misdemeanor arrests, the number of prior other arrests, the number of prior felony convictions, the number of prior misdemeanor convictions, and the age at first arrest within the state of Hawaii.[6] To characterize the nature of the controlling offense, we classify each case according to the most serious charge. Each charge is classified in the data as a felony class A, felony class B, felony class, C, misdemeanor, petty misdemeanor, severity unrecognized, or a violation. We use the penal code literals to classify each charge as either a person charge, a property charge, a drug charge, an "other offense" charge, or a community corrections charge. The table displays the nature of the most serious charge across the categories created by crossing the severity dimension (felony A, felony B, etc.) with the nature of the offense (person, property, drug, other). In addition to these variable, the table also displays the average days between arrest and pretrial supervised release for both groups.

The table reveals general balance on these criminal history and current offense severity variables. While the treatment group members are somewhat less likely to have a current

[6] Note, the rap sheet data that we analyze does not contain arrested prior to the age of 18.

offense classified as Felony B: person and more likely to have the most serious charge be Felony B: drug, the current offense distributions are quite similar.[7] A test of the overall balance of the treatment and control group for the criminal history and current offense variables listed in Table 2 fails to reject the hypothesis that the two samples are drawn from the same underlying population.[8]

5. Description of the Data and Estimation Methodology

We test for an effect of assignment to the treatment group on the following outcomes: revocations by type of pretrial supervised release, the outcomes of drug tests, jail days served, pretrial arrests for community corrections violations and for new offenses by type, case disposition outcomes, and post-disposition arrests for new offenses by type and for community corrections violations. Our analysis is based upon administrative data collected by the Hawaii Department of Public Safety and automated criminal history data maintained by Hawaii's Criminal Justice Information Services Division within the Office of the Attorney General. In this section, we describe the administrative data in detail, the manner in which we measure pretrial misconduct outcomes, case disposition outcomes, and post-disposition outcomes, and detail the exact estimators used to estimate the effects of the intervention.

Before discussing the data sources, we must describe several key dates that are key to our outcome definitions. As already noted, a supervised release case is opened once a presiding

[7] For a small number of observations, we could not match the subject to a felony arrest that preceded the supervised release date but that had a case disposition date following the close of the supervised release case. This will further restrict our sample size for some of the variables below. We will discuss this issue in detail in the data description section to follow.

[8] To be specific, a regression of the treatment indicator variable on all of the variables listed in Table 2 yields an overall F-statistic of 1.2 with a p-value of 0.25.

judge grants a motion for supervised release filed on behalf of a defendant. The supervised release open date is recorded in the administrative records and provides the start date for our pretrial observation period.[9]

The supervised release case can close through several channels. First, the supervised release case may close due to the disposition of the criminal case. In these instances, the supervised release close date corresponds either exactly with the case disposition date or may differ from the case disposition date by a few days. For 57 percent of our study subjects, the recorded supervised release close date corresponds to the criminal case disposition date. The overwhelming balance of subjects have criminal case disposition dates that follow the close of the supervised release case.[10] Alternatively, the supervised release case may close once a motion to permanently revoke supervised release prepared by a pretrial officer and filed by the prosecutor is granted by the judge managing the pretrial supervision defendant. Such a close may or may not occur while the defendant is in custody. If an arrest for a new offense triggers the filing of the motion, the individual's pretrial supervision case will officially close when the motion is granted. The revocation will occur without the issuing of a warrant as the individual will already be in custody. On the other hand, if the pretrial officer prepares a motion to revoke due to the failure of the defendant to report to the officer, due to absconding, a failure to appear for a court date, or due to some other form of misconduct that has yet to result in an arrest, the motion if granted will generate an arrest warrant. The supervised release case close date

[9] In our data, for 97 percent of the study subjects the supervised release date corresponds with the randomization date time stamp recorded by the randomization app. A few observation are randomized a few days after a supervised release cases is officially opened, nearly all within one week of the supervised release open date. There is one observation entered into the randomization app 13 days later and another 14 days later.

[10] There are six observations with case disposition dates that preceded the supervised release close date by a few days.

corresponds to the date that the warrant is issued. Apprehending the individual following the issuance of warrant can take anywhere from a few days to a year.

The final key date for our analysis is the criminal case disposition date. The case disposition date is recorded for all charges associated with an arrest in the state criminal history records. The latest observed date for the charged filed for a given incident marks the end point of the pretrial period.

We define the overall pretrial observation period as the time period between the supervised release open date and the case disposition date. We define the supervised release period as the period between the supervised release open date and the supervised release close date. For some outcomes (jail days in particular), we separately analyze the effects of treatment for the overall pretrial period as well as the period of supervised release.

A. Description of Data on Supervised Release Revocations and Warrants

We use two data sources to measure pretrial misconduct that generates a revocation of pretrial release. First, we analyze data for each revocation initiated and issued during the pretrial period for each of our study subjects. The revocation records include information on the date the revocation was initiated, a primary and secondary reason for the revocation (e.g., failure to report to the pretrial officer, absconding, failed drug test, failure to comply with drug treatment, curfew, unauthorized travel etc.), and for those revocations involving a warrant, the warrant initiation date. We use these data to tabulate all revocations issued against each study subject during the pretrial period as well as the number of revocations involving failure to report to pretrial services, failure to appear in court, absconding, failed drug tests, revocations due to a new arrest, revocations due to a failure to comply with a mandated treatment, and revocations

for other reasons. Note, the sub-categories of revocations by type sum to a total that exceeds the number of revocations per person, since the revocation records list up to two reasons for each revocation. We also test for an effect of treatment on the number of warrants issued.

The revocations data include both revocations driven by swift, certain, yet proportionate sanctions as well as permanent revocations of pretrial supervised release. In HOPE probation, the temporary revocations are actually referred to as motions to modify rather than revoke. One would expect more of these temporary revocations for HOPE treatment group members as this is an explicit feature of the intervention. Permanent, case-closing revocations, however should not mechanically increase with assignment to the treatment group.

To measure permanent revocations, we use data on the disposition of each supervised release case recorded when the supervised release case closes. The disposition codes separately identify cases that close due to revocation. The data elements are less detailed than the reasons codes employed in the revocations data. Permanent closures are coded as being due to “drug activity”, “failure to appear”, “new charges”, and a catch all “other” category. The “other” category is the most frequent code used by staff. Nonetheless, the disposition codes allow us to unambiguously identify cases where pretrial misconduct permanently ends the pretrial supervised release case.

We are able to approximately measure the reasons for a permanent revocation using the more detailed reasons field recorded in the revocations data by taking into account the timing of a revocation relative to the supervised release close date. Figure 2 presents the empirical distribution of the difference between the supervised release close date and the revocation initiation date for revocations filed against members of the treatment group and those files

against members of the control group. Note temporary modifications are not used for control group members. For the control group, revocations are prepared by the pretrial officer with the intention of closing the supervised release case. The distributions reveal that 86 percent of revocations filed against members of the control group occur within the thirty-day window preceding the supervised release close date. Moreover, among members of the control group, 90 percent of those against whom a motion to revoke is filed have recorded supervised release case dispositions indicating a permanent revocation.[11] As a supplement to our main results, we report in the appendix treatment effect estimates for revocations filed within thirty days of the supervised release close date using the more detailed reason-for-revocations variable available in the revocations administrative data.

B. Description of Drug Test Data

We were provided an initial extract of administrative data on drug tests roughly half way through the roll out period. A preliminary analysis of the drug test data revealed that many of the study subjects did not have a single recorded test. Based on this assessment, we physically reviewed the paper records for each drug test kept in the supervised release case files for all study subjects and created a parallel enumeration based on these tests. In conjunction with our final pull from the administrative data, we observe 1,870 drug tests administered between the supervised release open date and the supervised release close date. Of this total, the paper file review unearthed 222 tests not included in the administrative records (approximately 12 percent of the total tests observed).

[11] The fact that it is not 100 percent reflects the fact that revocation motions are sometimes denied by the presiding judge.

Despite this additional data collection effort, we still do not observe test results for all study subjects. To be specific, we observe at least one test outcome for 80 percent of treatment group subjects and 61 percent of control group subjects. Among those tested at least once, we observe 6.5 tests on average for treatment group members and 4.6 tests on average for control group members. If we incorporate study subjects where test results are not observed, the average number of tests for treatment group subjects is 5.2 while the average for control subjects is 2.8.

The administrative records indicate whether the test was positive, whether the subject refused the test, tampered with the test, or did not show up for a scheduled test. Our chief outcome codes tests where drugs are detected, where the individual refuses or tampers with the test, or where the individual doesn't show up as a test failure.

Key concerns regarding the drug test data are the facts that we do not observe data for all study subjects and that the treatment group members were tested more extensively and intensively than control group subjects. This seems to be the case as well in some of the evaluations we reviewed in section 2 (note for example, the large differences in average tests in the DYT experiment), and not explicitly addressed in others. These patterns may be due to several factors. First, there may be instances where the individual does not show up for the test yet the no-show is not recorded in their paper records or in the administrative data base. Second, subjects may be revoked or have their cases disposed prior to their first drug test. Third, the treatment group members have substantially longer supervised release periods than control

group members, with an average difference between the supervised released close and open dates of 105 days for the control group and 147 days for the treatment group.[12]

In our analysis of drug tests, we make efforts to address non-random selection by presenting along with a simple comparison of means, estimates that statistically adjust for pre-determined characteristics of the individual (criminal history, pre-determined demographic characteristics) and the underlying criminal case (nature of the controlling offense). Of course, we cannot control for time on supervised release as this is an outcome that is endogenously being determined by treatment. We also present separate estimates for sub-groups meant to increases comparability (for example, the first test for individuals tested at least twice). We discuss this issue in greater detail below with the presentation of the results.

C. Description of Jail Admissions Data

The Department of Public Safety provided administrative records for all admissions into and releases from a Hawaii correctional institution (inclusive of local jails and state prisons) from January 2013 through April 2018. For each of the study subjects we identify each admission occurring between the supervised release open date and the case disposition date. We use the beginning and end dates of each jail spell in conjunction with the supervised release date and the case disposition date to calculate total days served during the pretrial period.

Given the nature of HOPE sanctions, where short jail stays are used to punish technical violations, and the fact that revocations filed against control group members are generally permanent and usually result in a readmission to jail for the remainder of the pretrial period, we expect the treatment effect to be heterogeneous over different portions of the pretrial period.

[12] The difference between these two means is statistically significant at the one percent level of confidence.

Hence, we test for separate effects of treatment on jail days served for the supervised release period (the time period between the supervised release open and close dates) and the period between the supervised release close date and the case disposition date. Note for subjects where the case disposition date and supervised release close date coincide, this latter time period will have a length of zero days.

D. Description of Data Used to Measure Criminal History, Pretrial Arrests, Case Dispositions, and Post-Trial Recidivism

The CJIS criminal history data records the list of arrest charges associated with each arrest incident, the filed charge if any associated with each arrest charge, the disposition of each charge, and the disposition date. The data also provides information on the severity of each charge (e.g., felony A, felony B, felony C, misdemeanor etc.) the specific penal code violated, and a literal description for each charge. We provided CJIS with a list of the standard criminal justice IDs for each of our study subjects. In turn, CJIS provided us with these criminal history records for all arrests for these individuals occurring within the state of Hawaii through April 18, 2018.

We use the charge literal to classify each charge as a person, property, drug, or other offense, and use the broad categorization in conjunction with the severity variable to identify the most serious charge associated with an arrest incident. We classify higher severity charges as more serious (i.e., A felony is more serious than B felony, which is more serious than C felony, which is more serious than misdemeanor, and so on). Within severity levels, we specify person offenses as more serious than property offenses, followed by drug offenses, and other offenses. We also use the disposition codes to identify arrests that result in a conviction and the severity level of the conviction.

We use these coded data to characterize each individual's criminal history, the nature of the current offense, pretrial arrests, case dispositions, and post-trial recidivism. We have already seen the criminal history variables in our discussion of covariate balance between the control and treatment groups, yet it is worthwhile to revisit our method for tabulating these variables. The felony arrest immediately preceding the supervised release open date provides the characteristics of the current criminal case. Using all arrests preceding this arrest, we tabulate the total number of arrests where the most serious charge is a felony, the total number of arrests where the most serious charge is a misdemeanor, and the total number of arrests where the most serious charge is of lesser severity. We also tabulate prior arrests involving a community corrections violation charge or contempt of court charge only (i.e. there are no criminal charges other than those associated with bail, a probation or parole violation, a revocation of supervised release, or a charge of contempt of court). We also tabulate the number of prior felony convictions and misdemeanor convictions as well as the age of each individual at their first observed Hawaii arrest. We do not observe arrests occurring prior to the age of 18.

For arrests following the arrest associated with the current offense, we tabulate the number of arrest incidents by type occurring between key dates. To gauge pretrial arrests, for each individual we tabulate the total number of arrests occurring between the supervised release open date and the case disposition date. Aside from overall arrests, we tabulate pretrial arrests for a series of sub-totals distinguished by our dimensions of offense severity. Specifically, we tabulate arrests for each individual where there is a new criminal charge, arrests that only involve community-corrections related charges, arrests where the most serious charge is a felony, where the most serious charge is a misdemeanor, and where the most serious charge is of lesser

severity, arrests where the most serious charge is a person offense, a property offense, a drug offense, an "other-crime" offense, and a community corrections offense. To measure post-disposition recidivism, we tabulate similar arrest totals for each study subjects for the six-month period following case disposition and the twelve-month period following case disposition. Since we observe arrests through mid-April 2018, we restrict the post-disposition analysis to subjects with criminal case disposition dates that occur six months prior (for the six-month analysis) or twelve months prior (for the twelve-month analysis). These restrictions reduce the size of the sample used to estimate treatment effects for these outcomes. We discuss this issue further with the presentation of the results below.

We also use the criminal history data to characterize the disposition of each criminal case. In particular, we measure whether the case results in a conviction, whether the case results in a felony conviction, or whether the case results in a deferred acceptance of guilt (DAG) plea. Note, conviction is measured based on being convicted for any of the charges filed against the defendant while felony convictions are based on convictions for any of the felony charges recorded at arrest and filed against the defendant. We classify a case as resulting in a DAG plea if this is the disposition for at least one charge and the individual is not convicted on any of the other arrest charges.

E. Estimation of Treatment Effects

Given that study subjects are randomly assigned to either the treatment or control group, simple differences in mean outcomes can be interpreted as causal effects of the intervention. As our discussion of the patterns in Tables 1 and 2 revealed, the treatment and control groups have similar average demographic and criminal history characteristics and are similar in terms of the

nature of their current offense. The groups however are not completely balanced on the timing of randomization as we increased the randomization rate into the treatment group eight months into the study period. Moreover, to the extent that pre-determined covariates such as gender, race, prior criminal history, or characteristics of the current arrest predict variation in the outcomes that we study, estimating treatment effects using multivariate regression may improve the precision of our estimates and thus increase statistical power.

Hence, we estimate the causal effects of being assigned to the HOPE treatment group in the following manner. First, we present simple comparisons of means and calculate the difference in means for each of the outcomes discussed above. Second, we present alternative estimates based on a regression of each outcome on a treatment group indicator and a dummy variable indicating that the subject was randomized during the period with the higher treatment group randomization rate. Finally, we re-estimate the model adding controls for demographic characteristics, characteristics of the current offense, and criminal history. We base statistical inference on two-tailed hypothesis tests based on Huber-White robust standard errors for all estimates. For our analysis of drug test outcomes, we also cluster our standard errors by study subject in the multivariate regression models. We discuss this issue further with the presentation of the results below.

6. Empirical Results

In this section we present estimates of the treatment effects of HOPE pretrial on the outcomes discussed above. For each set of outcomes, we first present a table of average values for the control and treatment group as well as the difference in means between the two groups

for each outcome. We then present comparable estimates from multivariate models from two specifications. The first model regresses the outcome on a treatment indicator variable and a dummy variable indicting being randomized into the study on or after April 3, 2015. The second model adds to the first specification the demographic variables listed in Table 1 and the criminal history and characteristics of the controlling offense listed in Table2. We omit the variable measuring the time between arrest and supervised release, though including this variable does not alter the results. In addition, we control for current age and age squared as well as age at first arrest and its square in the complete model.

A. Revocations

Table 3 presents average values for revocation outcomes. Panel A presents average values for all revocations, revocations for absconding, failure to appear in court, and failure to report to the pretrial officer (grouped together and individually), revocations for failed drug tests, revocations for failing to comply with drug treatment conditions, revocations for a new arrest, and revocations for other reasons. Average revocations are substantially higher for treatment group members. Control group subjects experienced 0.294 revocations on average compared with 0.853 revocations for the treatment group, giving a statistically significant difference in average revocations during the supervised release period of 0.558 (statistically significant at the one percent level of confidence). This is driven primarily by large differences in revocations for failure to report to the pretrial officer (a statistically significant difference of 0.286) and failed drug tests (a statistically significant difference of 0.274), and smaller differences for other reasons (with a difference of 0.101) and failure to comply with drug treatment (0.059). Interestingly,

failure to appear in court is a rare outcome for this population, with no control group members having a revocation filed against them and an average value for the treatment group of 0.021.

Despite the substantially higher average revocations for the treatment group, the average number of warrants issued is only slightly higher with the difference not quite statistically significant (0.272 warrants on average for control group members compared with 0.347 for treatment group members). This small difference in warrant reflects an important difference in the nature of revocations involving treatment group members and control group members. Many of the arrests for treatment group members involve on-the-spot arrests made by the pretrial officer (for example, for failing a drug test). In fact, the intervention was designed to generate such temporary revocations for technical violations.

Panel B presents average values for permanent revocations. Again, we measure a permanent revocation using the supervised release disposition code recorded at the close of each case. The disposition variable codes permanent revocations as due to a failure to appear, new drug activity, new charges, and other. Roughly 32 percent of control group members experience a permanent revocation compared with only 12 percent of treatment group members. The difference of approximately 20 percentage points is statistically significant at the one percent level of confidence. Permanent revocations are lower by 3.1 percentage points for new drug activity (significant at the five percent level of confidence) and by 1.7 percentage points for new arrest charges (with the difference significant at the 10 percent level of confidence). The biggest difference is observed for the catchall “other category.”

The reasons for revocations indicated in the revocation records do not match up well with the reason given for a permanent revocation indicated in the supervised release disposition

codes. This may be due to permanent revocation being due to multiple factors. To probe more thoroughly the reasons for a permanent revocation, appendix Table A1 presents a comparable analysis for revocations filed within 30 days of the supervised release close date. Recall from our discussion of figure 2 that revocations filed against control group subjects tend to occur during the thirty-day period preceding the supervised release close date. Focusing on revocations occurring close to the end of the supervised release period permits us to exploit the more detailed information on the reason for the revocations included in the revocations records.

Average revocations within thirty days are generally lower for all categories for the treatment group, and in most instances the differences are statistically significant. The average control group member experiences 0.253 revocations within 30 days of the supervised release close date compared with an average for treatment group members of 0.137 (with the difference statistically significant at the one percent level of confidence). While revocations within 30 days of the supervised release close date are lower for the treatment group in all categories, we observe statistically significant lower average values for the treatment group for revocations for absconding, drug treatment revocations, and revocations for new arrests.

Table 4 presents comparable multivariate estimates of treatment effects for all revocations and permanent revocations. The table presents the control group average in the first column. The second column presents the treatment effect after adjusting for being randomized during the period where the randomization rate is 50/50. The final column presents treatment effect estimates adjusting for demographic covariates, criminal history, and the nature

of the current criminal offense.[13] Heteroscedastic robust standard errors are presented in parentheses alongside each estimate. The estimates in Table 4 correspond quite closely to the mean difference presented in Table 3. Treatment group members are substantially more likely to acquire revocations, though are considerably less likely to experience permanent revocations. Adding covariates does little to the standard errors of the estimates, in some instances causing slight increases and in others slight decreases. Parallel multivariate estimates for revocations occurring within thirty days also align with the basic mean differences in these outcomes. These additional results are presented in appendix Table A2.

B. Drug Test

Table 5 contrasts the proportion of drug tests that results in a failure (test positive, tamper, refuse, or recorded no-show) for tests administered to the treatment and control group. The table presents several contrasts. First, we compare average failure rates for all 1,870 tests. Next, we look at the first five tests administered to study subjects, calculating average failure rates for all tests combined and for the first through fifth test. Note, the number of tests diminishes as test order increases as fewer and fewer subjects are tested at higher cumulative levels.

Treatment group members fail their dug tests at a rate that is five percentage points lower than the control group. Given the control group average of 0.217, this amounts to a 23 percent reduction in drug test failures. Note, unlike previous HOPE interventions, both the treatment and control groups were subjected to random drug testing and required to call into the hotline.

[13] The estimates in the second and third column are the coefficients on the treatment indicator from a regression of the outcome on the treatment indicator and the remaining predetermined variables in the model specification.

Hence, the effect estimated here likely reflect the isolated impact of the difference in sanctioning for the HOPE treatment group. When we focus on the first five tests only, treatment group members fail drug tests at a rate that is roughly four percentage points lower than the control group (a 19 percent reduction relative to the control group average). While the point estimates for the difference in failure rates between the treatment and control group averages for the individual tests by order are all negative (with the exception of the point estimate for test #5), the standard errors are quite large and only the estimate for test #2 is statistically significant.

Table 6 presents comparable results from multivariate models where we first adjust for differences in the timing of randomization and then adjust for predetermined demographic and criminal history variables. Again, we present the control group average and the two treatment effect estimates. To subject our inference to a higher standard, here we calculate standard errors that cluster by study subject, effectively allowing for intra-subject correlation in the error components of the regression model. This basically increases the standard errors relative to model estimates that do not allow for clustering. For all tests combined, the point estimates indicate a treatment effect on failed drug tests ranging from a reduction of 6.1 to 6.7 percentage points (amounting to a 29 to 31 percent reduction in failures relative to the control group). The empirical estimates for the models focusing on the first five tests and for the individual tests by order are similar to the raw difference in average presented in Table 5.

Recall from our discussion of the drug test data that treatment group members are tested more times and are more likely to be tested at all relative to control group members. To reiterate, we observe at least one drug test for 80 percent of the treatment group but only 61 percent of the control group. We verified with the pretrial staff that both groups were subject

to similar drug testing requirements in terms of random testing, calling into the hotline and frequency. Hence, differences in observed testing must reflect either quicker case resolution for control members (either through permanent revocation or an earlier disposition of the criminal case), a differential propensity to show up for drug tests (although technically this is recorded in the administrative records), or a differential propensity to record the results of tests. Note, in addition to the administrative records, we reviewed the paper files of each subject and thus believe we have captured all administered/scheduled tests for which a paper record exists in the individual's file. It is indeed the case that treatment group members are on supervised release for longer periods on average relative to control group members, with the average time between the open and close dates for treatment group members equal to roughly 1.4 times the comparable average for control group members. Interestingly the proportion tested at all for the treatment group is 1.3 times that for the control group. Hence, differential exposure time must account in part for this difference. However, the average number of tests administered (inclusive of individuals for whom we have no tests) for treatment group members was roughly 1.9 times that of the control group. Hence, differential time on supervision cannot fully explain the difference in observed testing.

Given that both were subject to similar random testing regimes, our prior would be that those for whom we do not observe drug tests are likely avoiding pretrial service officers due to a high likelihood of failing the test. Indeed, among members of the control group, those for whom we observe at least one drug test are 14 percentage points less likely to have the supervised release case permanently close due to revocation and survive on supervised release 34 days longer than control group members with no test results. Hence, non-random selection of control

group members into testing may be leading us to under-estimate the effect of treatment on failed drug tests.

To partially address this selection issue, we present additional estimates where we focus on study subjects that experienced similar quantities of testing. To be specific, we test for a treatment effect on failure rates for the first two tests for subjects tested at least twice, for the first three tests for subjects tested at least three times, for the first four tests for subjects tested at least four times, and for the first five tests for subjects tested at least five times. Table 7 presents control group averages and multivariate treatment effect estimates for these sub-groups for all of the tests pooled and for individual test by sequence order. Again, we cluster the standard errors by study subject for models that pool more than one test per subject. The results from this exercise generally find treatment effects that are somewhat larger than the estimates that simply compare all test results for treatment and control group members. Among subjects tested at least twice, treatment reduces the failure rate by 6.2 to 8.2 percentage points, with the larger estimate significant at the five percent level of confidence and the smaller estimate significant at the ten percent level. Among those tested at least three times and those tested at least four times, treatment effect estimates range from declines of 10.4 percentage points to declines of 12.3 percentage points (three of the estimates are statistically significant at the one percent level of confidence and one is significant at the five percent level of confidence). For subjects tested at least five times, the treatment effect estimates range from declines of 4.4 to 6.4 percentage points. Only the larger estimate is statistically significant at the ten percent level.

To summarize, we find statistically significant and substantial effects of treatment under HOPE pretrial on the likelihood that pretrial releases fail drug tests. A 23 percent reduction in

failed drug tests provides a solid lower bound estimate. Estimates are larger in magnitude when we contrast the test outcomes of study subjects with similar testing frequency.

C. Jail Days Served Between the Supervised Release Date and Case Disposition Date

Tables 8 and 9 present our analysis of jails days served. For this domain, we test for an effect on jail days served for the overall period between the supervised release open date and case disposition date, for the sub-period between the supervise release open date and the supervised release close date, and for the sub-period between the supervised release close date and the case disposition date. Table 8 presents group averages and the average difference between the treatment and control groups. Table 9 presents the corresponding treatment effect estimates from multivariate models.

Control group members are incarcerated an average of 49.7 days between the supervised release open date and the case disposition dates. The comparable average for the treatment group is 51.8 days, with the difference relative to control not statistically significant. Hence, similar to the original HOPE probation evaluation in Hawken and Kleiman (2009), the program does not impact overall jail days served. There are interesting differences by sub-period however. Between the supervised release open date and close date, treatment group members are incarcerated an average of 20.40 days compared to an average of 1.1 days for the control group (with the difference of 19.34 days statistically significant at the one percent level of confidence). By contrast, between the supervised release close date and the case disposition date (which is often the same date for many of the cases) average jail days for the treatment group is 31.2 compared to an average of 48.6 for the control group (with the difference of -17.45 days statistically significant at the five percent level of confidence). The treatment effect

estimates from the multivariate models presented in Table 9 basically confirm the findings from the comparison of averages presented in Table 8.

Hence, while treatment does not impact total days served during the portion of the pretrial period following the granting of a supervised release motion, treatment reshuffles jail days to earlier in the pretrial period and away from incarceration following a permanent revocation through the disposition of the criminal case.

We should reiterate here that the sanctions structure for HOPE pretrial deviates in an important manner from the sanctions structure used in HOPE probation. In HOPE probation, probation officers can directly file motions to modify probation with the HOPE court, streamlining the modification process and permitting the modifications hearings to be scheduled two business days later. This mean that those probationers arrested on the spot for testing positive can be sanctioned with a two-day jail spell capped off with a modification hearing before the judge. In contrast, while pretrial officers prepare the affidavit for a motion to revoke/modify supervised release cases the prosecuting attorney must file the motion with the court. To allow sufficient time for the preparation and filing of the motion by two different agencies, HOPE pretrial allowed for seven-day sanctions for those arrested on the spot (the practical consequence of scheduling modification hearings seven days following an arrest).

Assuming that a sanctioning structure that scheduled hearings for two days following a violation would have had similar effects on pretrial misconduct as the structure actually implemented, then the intervention in this counterfactual may have reduced average jail days served among the treatment group. Of course, this is a big assumption and we can only evaluate the behavioral effects for the program as implemented. Nonetheless, the two day sanctions have

proven effective in Hawaii for those on probation (Hawken and Kleiman 2009). Should the jurisdiction continue HOPE pretrial, they should consider streamlining the process for filing supervised release modifications and experimenting with shorter sanctions.

D. Pretrial Arrests

Tables 10 and 11 present treatment effect estimates for the effect of HOPE pretrial on arrests occurring between the supervised release open date and the disposition date for the criminal case. We present estimates for all arrests and subsets of arrests. Regarding the subsets, we present estimates for arrests involving a new charge as well as arrests where only community corrections charges or a contempt of court charge is listed for the arrest incident. We also tabulate total arrests during this period by whether there is a felony charge, by whether the most serious charge is a misdemeanor, and other arrests. Finally, we characterize arrest by the nature of the most serious charge (person, property, drugs, other, community corrections).

While average total arrests are similar for the treatment and control groups, members of the treatment group experience fewer average arrests involving a new charge (0.385 for the control group and 0.226 for the treatment group, with the difference statistically significant at the five percent level of confidence). Average felony arrests for treatment group members is 42 percent lower than that of the control group (0.204 felony arrests for the control group compared with 0.118 for the treatment group). Regarding the nature of the offense, treatment group members are 42 percent less likely to be arrested for an offense where the most serious charge is a property offense (with the difference statistically significant at the 10 percent level of confidence) and 70 percent less likely to be arrested for an offense where the most serious charge falls into the “other” category.

The multivariate results in table 11 roughly conform to the analysis of mean differences. Here, however, adding controls leads to some change in coefficient estimates and in some instances improvements in precision as measured by the standard error. The effect on arrests with new criminal charges is comparable and statistically significant in the first specification including the indicator for the later randomization period. Adding all of the control variables shrinks the standard error but also causes a decline in the coefficient estimate and yields an insignificant result. The result for felony arrest and arrests for property offenses and other offenses are similar to the bivariate differences presented in Table 10. Adding controls increases the estimate of the effect of treatment on community corrections arrests, with the treatment group arrested for technical violations at a rate that is 38 percent higher than the control group average.

E. Criminal Case Dispositions

Tables 12 and 13 present estimates of treatment under HOPE pretrial on three criminal case outcomes: whether the case results in any conviction, whether the case results in a felony conviction, and whether the case results in a deferred acceptance of guilt plea. Several recent high quality quasi-experimental studies have found a causal effect of pretrial detention on the likelihood of conviction (Dobbie, Golden, and Yang 2016; Heaton, Mayson, and Stevenson 2017). This may occur through several channels. Defendants may be better able to mount a defense while not incarcerated. Alternatively, doing well on supervised release may lead to leniency by the prosecutor, the presiding judge, or both. Our outcomes are meant to measure whether such factors are in operation. The any-conviction outcome presents an overall gauge of the case outcome. Conviction for a felony is certainly more serious than conviction for a misdemeanor.

Hence, we separately analyze whether treatment impacts the likelihood of a felony conviction. Recall, all of our study subjects are felony defendants. Finally, a DAG plea carries the prospect of not having the conviction on one's record presuming that the individual complies with the community corrections conditions ordered by the presiding judge.

Approximately 76 percent of the control group is ultimately convicted. By contrast 68 percent of the treatment group is ultimately convicted, with the difference relative to the control group statistically significant at the 10 percent level of confidence. Treatment group members are 10.3 percentage points less likely to receive a felony conviction, a rate that is 14 percent lower than that for the control group (with the difference statistically significant at the five percent level). Treatment group members are slightly more likely to receive DAG/no contest plea. The difference relative to the control group, however, is not statistically significant.

The multivariate results in Table 13 basically accord with the results in Table 12. Controlling for the difference in randomization periods causes a decline in the overall conviction difference from -0.081 (estimate presented in Table 12) to -0.068 in the first model specification and -0.063 in the second specification. Combined with the slightly higher standard error this difference is no longer statistically significant. Similarly, the difference in felony conviction rates between the treatment and control groups narrows from -0.103 in Table 12 to -0.082 in the multivariate models presents in Table 13. The treatment effect estimates are still statistically significant at the ten percent level of confidence.

F. Post-Disposition Arrests

Tables 14 through 17 present estimates of the effect of treatment on post-disposition arrest outcomes. Our ability to analyze post-disposition arrests is qualified to some degree by

the timing of case dispositions relative to the time period for which we have criminal history records. In particular, for cases with very late disposition dates as well as the handful of cases where the criminal case has yet to reach a disposition, we do not have sufficient time since case disposition to study post-disposition arrests. We observe at least 180 days of post-disposition arrests for 467 of our study subjects and 365 days of arrest activity for 437 of our study subjects. Hence, here we analyze post-disposition arrests for these sub-samples. Tables 14 and 15 present results for arrests occurring during the 180-day widow following case disposition while tables 16 and 17 present results for arrests occurring during the 365-day window following case disposition. The arrest outcomes that we study parallel the outcome definitions that we use for arrests occurring between the supervised release open date and case disposition date.

Beginning with the comparisons of means for the 180-day window in Table 14, arrest rates are generally lower in nearly all comparisons for the treatment group relative to the control group. However, none of the differences are statistically significant. In the multivariate models presented on Table 15, controlling for being randomized during the latter period leads to slight increases in the treatment effect estimates for all arrests and misdemeanor arrests, and both are now marginally statistically significant (at the 10 percent level of confidence). However, this finding is not particularly robust and adding other controls yields negative yet insignificant treatment effect estimates.

Table 16 presents comparison of average values for arrests occurring over the 365-day post-disposition time window. While arrest rates are lower for treatment group members in most comparisons, only the difference for new drug offenses is statistically significant (with the difference of -0.034 relative to the control group average of 0.052 significant at the 10 percent

level of confidence). The treatment effect estimates from the multivariate models presented in Table 17 closely align with the mean differences in Table 16. Again, while there are many negative estimates, only the treatment effect estimates for new drug offenses are statistically significant.

To summarize the results for this domain, there is suggestive evidence of a negative effect of treatment on post-disposition arrests though very few of the point estimates are statistically significant. We hope to revisit these results in a year when we will have complete information on post-disposition arrests for all of our study subjects.

7. Adjusting Key Inference Statistics for the Multiple Hypotheses Tested

We have presented treatment effect estimates as well as standard errors for single hypotheses tests. We have tested for an impact of HOPE-pretrial on a number of outcomes. One may contend that the large number of outcomes explored may generate a few significant results by chance and that the inferences should be adjusted for the family wise error rate associated with the number of hypotheses tested.

As a final set of results, here we present a comparison of the p-values from single-hypothesis tests to adjusted p-values that take into account the number of inferences being made. We focus here on eleven principal outcomes: all revocations, whether one experiences a permanent revocation, drug test failures, total jail days, arrests for a new crime and separately for a community corrections violations during the supervised release period, whether one is convicted, convicted of a felony, or has a case disposed through a DAG plea, arrests for new crimes 180 days post disposition, and arrests for a community corrections violation 180 days post

disposition. We employ the free step-down resampling method of Westfall and Young (1993) making use of the STATA routine written and made publicly available by Jones, Molitor, and Reif (2018). We focus on the simple difference in means for each outcome between the treatment and control groups since statistically adjusting for covariates has little impact on our results.

Table 18 presents the results from this exercise. The first two columns of figures show the estimated treatment effect for each outcome as well as the robust standard error for each estimate. The third column presents unadjusted p-values from single hypothesis tests. The final column presents the adjusted p-values based on the Westfall Young algorithm with 10,000 replications. The treatment effects for all revocations and permanent revocations remain statistically significant at the one percent level of confidence. While the effect on the drug test failure rate is significant at the one percent level if we base inference on the single hypothesis test, the p-value increases to 0.057 when adjusted for the family wise error rate. Two of the outcomes go from being significant at the five percent level to just missing significance at the 10 percent level of confidence (both arrested during the pre-trial period for a new offense and felony conviction have unadjusted p-values of 0.019 and adjusted p-values of 0.127). The adjusted p-value for the “any conviction” (0.259) greatly exceeds the 10 percent threshold.

8. Conclusion

The findings of this study are several. Applying the HOPE model to pretrial supervised releases reduces drug test failure rates, reduces the likelihood of a permanent revocation of supervised release, reduces arrests during the supervised release period involving new criminal charges, and reduces the likelihood that the case results in any conviction and in a felony

conviction. While there is suggestive evidence of an effect of the intervention on post-disposition arrests, the estimates are imprecisely measured and generally not statistically significant. While the overall number of revocations filed against treatment group members is significantly higher, the effect of the intervention on warrants issued is marginal, likely reflecting revocations that are more akin to HOPE modifications following an on-the-spot arrest for a failed test. There is no overall impact of treatment on pretrial jail days served, though treatment does reshuffle jail days towards earlier in the pretrial supervision period.

Given the recent null findings from the HOPE-DFE and DYT experiments, it is worthwhile to ponder why the results here differ. Several possibilities come to mind. First, HOPE-pretrial was designed and for the most part implemented by Judge Steven Alm, with Judge Alm presiding over supervised release deliberations for most of the experimental period. HOPE pretrial most certainly adhered to the practices in Hawaii that were developed under HOPE probation and found to be effective by Hawken and Kleiman (2009) and Hawken et. al. (2016). Hence, fidelity to the HOPE model is not in question.

Beyond adherence to the model, it may be the case that the findings here are specific to the Hawaiian context or even to Judge Alm himself. Hawaii of course is a cluster of Islands in the middle of the Pacific Ocean, a fact that may fundamentally alter the nature of community corrections. Moreover, criminal justice practices and sentencing in Hawaii may have an effect by altering the relative consequences of failure, a point emphasized in Hawken (2016). Finally, Judge Alm just may be particularly good at working with probation and pretrial officers as well as the individuals on different forms of community corrections supervision. To be sure, substantial effort has been devoted to documenting the HOPE model and to provide technical assistance to

counties and other agencies wishing to implement the model (in other words, to document exactly what is being done in Hawaii and to strategize implementation in other contexts).

Nonetheless, our literature review highlighted several studies that have demonstrated various aspects of the HOPE model to be effective including evaluations of the implementation of the 24/7 Sobriety programs in South Dakota and Montana, the evaluation of the statewide application of swift and certain probation practices in Washington State, and a host of small scale non-experimental evaluations of HOPE-style interventions throughout the country. It may simply be the case the HOPE doesn't work everywhere or that key implementation differences between the DYT and HOPE-DFE programs explains the less positive findings from these analyses.

It is interesting to ponder some of the key deviations of the experiment evaluated here from the standard practice of HOPE probation and what they imply for the interpretation of the findings. We noted the fact that the control group in HOPE pretrial was subjected to random testing as well as the requirement to use the daily call-in line. In the original HOPE probation RCT, only the treatment group was subjected to random testing. Interestingly, while we find that assignment to treatment caused a significant and substantial reduction in the likelihood of a failed drug test (on the order of a 20 to 30 percent decline), the decline observed here was smaller than the proportional decline observed for HOPE probation. In conjunction, these two sets of findings suggest that both random testing as well as the certainty and celerity of sanctions individually impact compliance with the condition that the individual not use drugs. Of course, there may be an interaction effect between random testing and swift and certain sanctions. However, we cannot evaluate interaction effects given the design of our treatment/control contrast.

A further key difference between the implementation of HOPE pretrial and the standard practice in HOPE probation concerns the sanctions possibilities. HOPE pretrial employed the following sanctions: several hours in a court cellblock, one week in jail, two weeks in jail, and permanent revocation of pre-trial supervised release and readmission to jail for the duration of the pretrial period. HOPE probation is similar with the exception that the second least severe sanction is two days in jail and the more severe sanctions may involve more than two weeks in jail or the application of an open prison term. As we have already discussed, the sanction following a failed drug test is set by the practical consideration regarding the time it takes to prepare and submit a modification motion. Probation officers in Honolulu are able to directly submit these motions and thus probation hearings in HOPE court are typically scheduled two days later. By contrast, pretrial officers prepare the motions which are then filed by the prosecutor. This led to one week being the amount of time allotted to schedule a subsequent modification hearing in HOPE court, a factor that likely contributed to average pretrial jail days among the treatment group. It is entirely possible that a version of this demonstration with hearings scheduled two days following a violation may have yielded lower average jail days served. This of course assumes that the impact on behavior would be similar with this less severe sanction to what we observed here. Should Honolulu continue HOPE pretrial, they should consider streamlining the process for filing supervised release modifications and experimenting with shorter sanctions.

This brings us to the issue of the cost of this intervention relative to the benefits associated with less pretrial misconduct. Several of the outcomes that we have analyzed suggest that the intervention was cost neutral in some domains. HOPE treatment group members served

the same number of average jail days between the supervised release open date and the case disposition date. Moreover, there was no significant differences in average warrants issued and served. It is likely the case that more resources were expended on HOPE treatment group members in terms of court time and money for drug testing. However, these additional resources were likely quite small on average. Regarding court time, simple back of the envelope calculations based on our findings regarding revocations and a past process evaluation of HOPE probation suggest that that average HOPE treatment group member required an additional six to seven minutes of court time relative to the control group.[14] Given the admittedly small increment to the overall HOPE case load (no more than 50 HOPE pretrial open cases at any given time for a HOPE court monitoring roughly 2,000 active probation cases at any given time), and the small amount of average court time devoted to each case, marginal court costs generated by this demonstration were likely close to zero.[15]

Regarding drug tests, the average treatment group member was tested 5.2 times while the average control group member was tested 2.8 times. Tests are more expensive if the on-site

[14] Hawken and Kleimn (2009) estimate that the average HOPE probation hearings lasts 7.21 minutes. The average time per person of delivering the HOPE warnings was 3.53 minutes. Assuming that a motion to permanently revoke supervised release takes twice as long as a HOPE modification hearing (due to perhaps the defense contesting the motion), we can use the revocation findings reported here to generate an estimate of court time devoted to the average treatment and control group member. Specifically, HOPE treatment group members acquired 0.853 revocations on average, 0.121 of which were revocations resulting in a permanent revocation of supervised release. For control group members, the proportion experiencing a permanent revocation was 0.316. Assuming 7.21 minutes for motions to modify, 14.42 minutes for a motion to permanently revoke, and 3.51 minutes for the HOPE warning (delivered only to HOPE treatment group members), average court time per case would be 10.53 minutes for treatment group members and 4.56 minutes for control group members (with a difference of roughly six minutes). If we assume that a motion to permanently revoke requires the same amount of time as a motion to modify (7.21 minutes), then the average time per case would be 9.35 for treatment group members and 2.28 for control group members (giving a difference of roughly 7 minutes).

[15] In conversations with Judge Alm who presided over HOPE court for the majority of the time period of the experiment, it was noted that HOPE pretrial cases were rare events on his calendar and that the additional cases were easily accommodated and added to the probation caseload.

result is contested and requires that the sample be submitted for lab test confirmation. Under various assumptions regarding how frequently that occurs, we estimate that the additional drug testing for treatment group members may generate roughly $10 to $17 in additional costs per case for treatment group members relative to control group members.[16]

One domain where there was substantially greater expenditures on treatment group members concerns inpatient and outpatient drug rehabilitation services. In total, $308,165 was spent on treatment services for treatment group members, amounting to $1,621 per treatment study subject. To be sure, some control group members also received treatment services, either at their own expense or services that were covered by the defendant's health insurance plan such as Med-QUEST (Medicaid Hawaii), or more rarely, through private insurance coverage. Hence, $1,621 is likely a substantial over-estimate of the increment in treatment expenditures (regardless of who bears the costs of treatment) above and beyond resources that would normally be expended on pretrial supervised releases.

Putting a dollar value on the benefits is a bit more difficult given the state of research on the economic costs of crime[17] and the difficulty associated with pricing greater compliance with pretrial release conditions and the social costs associated with relatively minor offenses. Nonetheless, we can use our estimates to at least present a range of estimates for the benefits associated with one of the outcomes for which we see a treatment effect: property offenses. Heaton (2010) presents summary estimates of the costs per criminal incident of various offenses

---

[16] Intake Services informed us that the costs per test is $4.25 plus an additional $29 if the sample is sent to a lab. If we assume that all tests incur the $4.25 and that all failed tests are contested, average testing costs per treatment group member is $47.13 while the average cost per control group member if $29.52 (for a difference of $17.61). If we assume that none of the failed tests are contested, the average costs per treatment group member is $22.10 while the average per control group member is $11.90, giving a difference of $10.20.

[17] For a review and discussion of this literature see Dominguez-Rivera and Raphael (2015).

based on a variety of methodological approaches to estimating the costs of crime. After converting to 2016 dollars,[18] the estimates in Heaton (2010) suggest social costs per incident of $15,584 for a burglary, $2,545 for larceny, and $10,804 for motor vehicle theft. Given the distribution of property crime across these categories in Honolulu in 2016 and assuming that a reduction in property offending among felony supervised releases would reflect the composition in crime more generally in the city, these figures suggest that each property crime prevented generates social benefits worth $5,320.[19]

Our estimates indicate a decline in arrests for a new offense of 0.159, 0.076 of which was for a decline in property offenses and the balance for less serious "other offenses." If we multiply 0.076 by $5,320 we arrive at a value for the reduction in crimes cleared by an arrest of $404. Actual savings due to reductions in property crime are likely to be higher given that people who are criminal active are often not apprehended for crimes committed (that is to say, clearance rates especially for property offenses are notoriously low) as well as the fact that many offenses (motor vehicle theft being an important exception) are often not reported to the police.[20] While our sample is too small to generate reliable estimates of treatment effects on detailed property offenses, this range of estimates associated with the measured decline in property offending alone coupled with under-reporting and less than perfect crime clearance suggest that the value

[18] We inflate the figures in Heaton (2010) by 1.19 to convert from 2007 dollars to 2016 dollars (the time period corresponding to our evaluation).

[19] In 2016, data from the Uniform Crime reports shows crimes rates per 100,000 residents of 377.5, 2,258.8, and 436.5 for burglary, larceny, and motor vehicle theft. This suggests that the proportion of property crime due to burglary, larceny, and motor vehicle theft is 12.3, 73.5, and 14.2, respectively. Appling these weights to the cost of crime estimates from Heaton yields the savings per crime figure note above. See https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-4.

[20] In 2016, 13.1 percent of reported burglaries, 20.4 percent of reported larceny thefts, and 13.3 percent of reported motor vehicle thefts were cleared by an arrests (see https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/clearances).

of these benefit may offset a substantial amount the additional treatment expenditures. A full accounting would also include the value of lower offending in the "other offense" category, a category that includes many vehicle code violations, driving under the influence, various weapons offenses, and many misdemeanors and infractions. In addition, one should add the value of a more compliant caseload of pretrial defendants, as well as the value associated with fewer felony convictions and fewer convictions overall.

References

Alm, Steven (2016), "Fair Sanctions, Evidence-Based Principals, and Therapeutic Alliances," *Criminology and Public Policy,* 15(4): 1195-1214.

Becker, Gary S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2): 169-217.

Carson, E. Anne (2018), *Prisoners in 2016*, Bureau of Justice Statistics, U.S. Department of Justice, Washington, D.C. NCJ 251149.

Chalfin, Aaron; Haviland, Amelia and Steven Raphael (2012), "What Do Panel Studies Tell Us About a Deterrent Effect of Capital Punishment? A Critique of the Literature," *Journal of Quantitative Criminology,* 25: 1-39.

Chalfin, Aaron and Justin McCrary (2017), "Criminal Deterrence: A Review of the Literature," *Journal of Economic Literature,* 55(1): 5-48.

Chalfin, Aaron and Justin McCrary (2018), "Are U.S. Cities Underpoliced? Theory and Evidence,"*Review of Economics and Statistics*, 100(1): 167-186.

Clarke, Ronald V. (1995), "Situational Crime Prevention," *Crime and Justice*, 19: 91-150.

Cohen, Thomas H. and Brian A. Reaves (2007), *Pretrial Release of Felony Defendants in State Courts*, Bureau of Justice Statistics Special Report, U.S. Department of Justice Office of Justice Programs, Washington, D.C. NCJ 214994.

Cook, Philip J. (2016), "Behavioral Science Critique of HOPE," *Criminology and Public Policy*, 15(4): 1155-1161.

Cullen, Francis; Pratt, Travis C.; and Jillian J. Turanovic (2016), "It's Hopeless: Beyond Zero-Tolerance Supervision," *Criminology and Public Policy,* 15(4): 1215-1227.

DeVall, Kristen; Lanier, Christina and David Hartmann (2014), *Evaluation of Michigan's Swift & Sure Sanctions Probation Program: October 1, 2011 – September 30, 2013*, University of North Carolina Willmington.

Di Tella, Rafael and Ernesto Schargrodsky, (2004), "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack," *American Economic Review*, *94*, 115–133.

Dobbie, Will; Golden, Jacob and Crystal Yang (2016), "The Effect of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," National Bureau of Economic Research Working Paper #22511.

Dominguez-Rivera, Patricio (2018), "How Potential Offenders and Victims Interact: A Case Study from a Public Transportation Reform," U.C. Berkeley Working Paper.

Dominguez-Rivera, Patricio and Steven Raphael (2015), "The Role of the Costs of Crime Literature in Bridging the Gap Between Social Science Research and Policy Making: Potentials and Limitations," *Criminology and Public Policy*, 14(4): 589-632.

Donohue, John J. and Justin Wolfers (2005). "Uses and Abuses of Empirical Evidence in the Death Penalty Debate," *Stanford Law Review*, 58(3):791-845.

Hamilton, Zachary; Campbell, Christopher M.; van Wormer, Jacqueline; Kigerl, Alex; and Brianne Posey (2016), "Impact of Swift and Certain Sanctions: Evaluation of Washington State's Police for Offenders on Community Supervision," *Criminology and Public Policy*, 15(4): 1009-1072.

Hawken, Angela (2016), "All Implementation is Local," *Criminology and Public Policy,* 15(4): 1229-1239.

Hawken, Angela and Mark Kleiman (2009), "Managing Drug Involved Probationers with Swift and Certain Sanctions: Evaluating Hawaii's HOPE," Final Report to the National Institute of Justice, Grant Number 2007-IJ-CX-0033.

Hawken, Angela; Kulick, Jonathan; Smith, Kelly; Mei, Jie; Zhang Yiwen; Jarman, Sara; Yu, Travis; Carson, Chris and Tifanie Vial (2016), "HOPE II: A Follow-up Evaluation of Hawaii's HOPE Probation," Final Report to the National Institute of Justice Award Number 2010-IJ-CX-0016.

Heaton, Paul (2010), *Hidden in Plain Sight: What Cost-of-Crime Research Can Tell Us About Investing in Policy*, Center on Quality Policing, the RAND Corporation, Santa Monica, CA.

Heaton, Paul; Mayson, Sandra and Megan Stevenson (2017), "The Downstream Consequences of Misdemeanor Pretrial Detention," *Standard Law Review,* 69(3): 711-794.

Helland, Eric and Alexander Tabarrok (2007), "Does Three-Strikes Deter? A Non-Parametric Estimation," *Journal of Human Resources,* 42(2): 309-330.

Hjalmarsson, Randi (2009) "Crime and Expected Punishment: Changes in Perceptions at the Age of Criminal Majority." *American Law and Economics Review*, Volume 11(1): 209-248.

Hoynes, Hilary; Whitmore-Schanzenbach, Diane and Douglas Almond (2016), "Long-Run Impacts of Childhood Access to the Safety Net," *American Economic Review*, 106(4): 903-934.

Institute for Behavior and Health (2015), *The State of the Art of HOPE Probation,* Institute for Behavior and Health, Inc. Rockville, MD.

Johnson, Rucker C. (2011), "Long-Run Impacts of School Desegregation and School Quality on Adult Outcomes," National Bureau of Economic Research Working Paper #16664.

Jones, Damon; Molitor, David and Julian Reif (2018), "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study," National Bureau of Economic Research Working Paper #24229.

Kilmer, Beau; Nicosia, Nancy; Heaton, Paul and Greg Midgette (2013), "Efficacy of Frequent Monitoring with Swift, Certain, and Modest Sanctions for Violations: Insights from South Dakota's 24/7 Sobriety Project," *American Journal of Public Health*, 103(1): e37-e43.

Kleiman, Mark (2009), *When Brute Force Fails: How to Have Less Crime and Punishment,* Princeton University Press: Princeton, NJ.

Klick, Jonathan and Alexander Tabarrok (2005), "Using terror alert levels to estimate the effect of police on crime," *Journal of Law & Economics*, *48*, 267–279.

Lattimore, Pamela K.; Layton Mackenzie, Doris; Zajac, Gary; Dawes, Debbie; Arsenault, Elaine; and Stephen Tueller (2016), "Outcomes Findings from the HOPE Demonstration Field Experiment: Is Swift, Certain, and Fair an Effective Supervision Strategy?" *Criminology and Public Policy* 15(4): 1103-1141.

Lee, David S. and Justin McCrary (2009) "The Deterrent Effect of Prison: Dynamic Theory and Evidence," Industrial Relations Section Working Paper, Princeton University.

Loudenburg, Roland; Drube, Gregg and Gary Leonardson (2011), *South Dakota 24/7 Sobriety Program Evaluation Findings Report,* Mountain Plains Evaluation, LLC.

MacDonald, John and Steven Raphael (2017), "An Analysis of Racial and Ethnic Disparities in Case Dispositions and Sentencing Outcomes for Criminal Cases Presented to and Processed by the Office of the San Francisco District Attorney," UC Berkeley Working Paper.

Midgette, Greg and Beau Kilmer (2015), “The Effect of Montana’s 24/7 Sobriety Program on DUI Re-Arrest: Insights from a Natural Experiment with Limited Administrative Data,” RAND Justice, Infrastructure, and Environment Working Paper, WR-1083-MHP.

Nagin, Daniel S. (2013), “Deterrence: A Review of the Evidence by a Criminologist for Economists,” *Annual Review of Economics*, 5: 83-105.

Nagin, Daniel S. (2016), “Project HOPE: Does it Work?”, *Criminology and Public Policy,* 15(4): 1005-1007.

Nagin, Daniel S. and John V. Pepper (2012), *Deterrence and the Death Penalty*, National Academy of Sciences, Washington, D.C. National Academy of Sciences 2017.

Nagin, Daniel S. and Greg Pogarsky (2004), “Time and Punishment: Delayed Consequences for Criminal Behavior,” *Journal of Quantitative Criminology,* 20(4): 295-317.

National Academies of Sciences, Engineering, and Medicine (2017), *Proactive Policing: Effects on Crime and Communities*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24928.

O’Connell, Daniel J; Brent, John J.; and Christy A. Visher (2016), “Decide Your Time: A Randomized Trial for a Drug Testing and Graduates Sanctions Program for Probationers,” *Criminology and Public Policy* 15(4): 1073-1102.

Petersilia, Joan and Susan Turner (1993), “Intensive Probation and Parole,” *Crime and Justice,* 17: 281-335.

Polinksy, A. Mitchell and Steven Shavell (1984), “The Optimal Use of Fines and Imprisonment,” *Journal of Public Economics*, 24: 89-99.

Raphael, Steven and Jens Ludwig (2003), “Prison Sentence Enhancements: The Case of Project Exile," in *Evaluating Gun Policy: Effects on Crime and Violence*, edited by Jens Ludwig and Philip J. Cook, Brookings Institution Press: Washington, D.C.

Schiraldi, Vincent (2016), “Confessions of a Failed HOPE-er,” *Criminology and Public Policy*, 15(4): 1143-1153.

Shannon, Lisa (2013), *Kentucky SMART Probation Program Year One Report,* Morehead University, Sociology, Social Work, and Criminology.

Snell, Clete (2007), *Fort Bend County Community Supervision and Corrections Special Sanction Court Program Evaluation Report,* Department of Criminal Justice, University of Houston, Downtown; Houston, Texas.

Virginia Criminal Sentencing Commission (2016), *Immediate Sanction Probation Pilot Program Evaluation,* Richmond Criminal Sentencing Commission http://www.vcsc.virginia.gov/Immediate%20Sanction%20Probation%20Pilot%20Program%20Evaluation%20-%20Final%2012-20-2016.pdf

Westfall, Peter H. and S. Stanley Young (1993), *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment,* John Wiley and Sons; New York, NY.

Zeng, Zhen (2018), *Jail Inmates in 2016*, Bureau of Justice Statistics, U.S. Department of Justice, Washington, D.C. NCJ 251210.

Figure 1: Cumulative First-Time Assignments to the HOPE Treatment and Control Groups by Month of Assignment

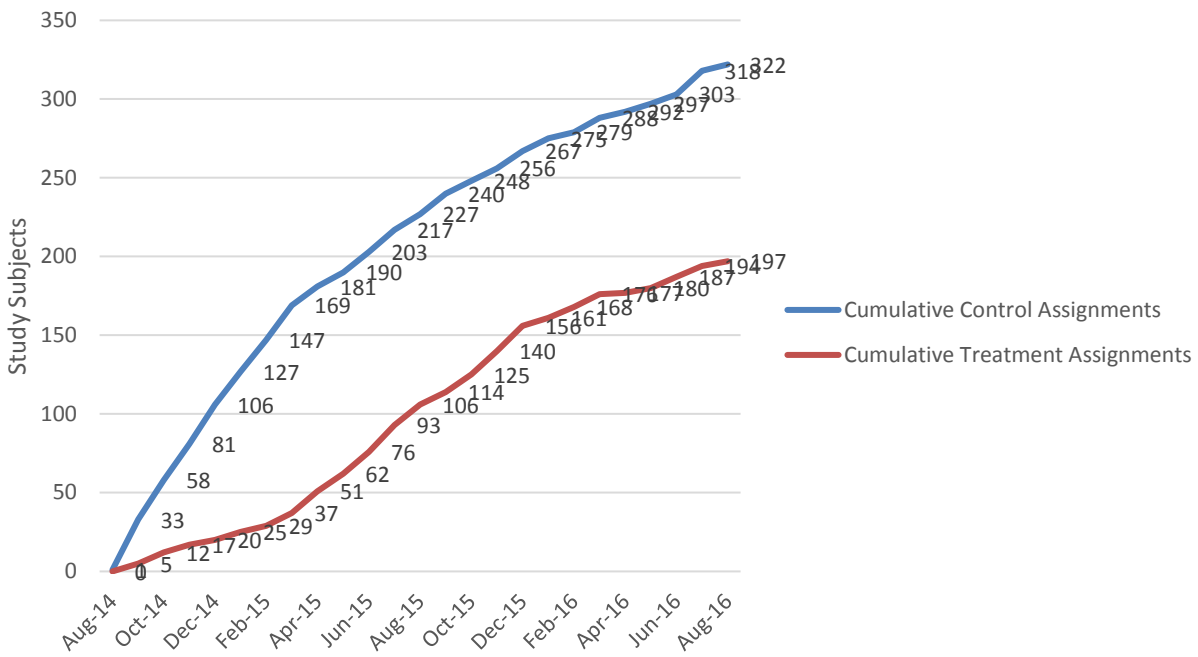


Figure 2: Empirical Distribution of the Time in Days between Supervised Release Close Dates and Revocation Dates

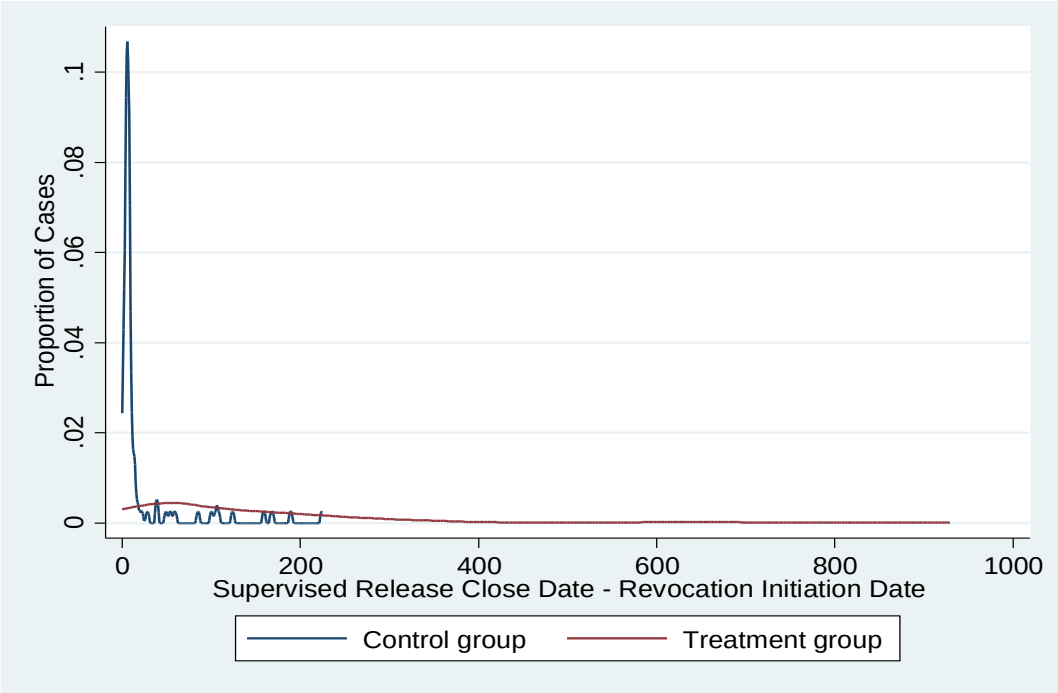


Table 1
Average Demographic Characteristics of HOPE Treatment and Control Group Members and Tests for Demographic Covariate Balance

	Control Group	Treatment Group	Difference, Treatment – Control (Standard Error)
Male	0.79	0.81	0.01 (0.04)
Age	34.24	36.00	1.76 (1.05)[c]
White	0.21	0.24	0.03 (0.04)
Native Hawaiian	0.39	0.36	-0.03 (0.04)
African American	0.04	0.07	0.03 (0.02)
Hispanic	0.03	0.02	-0.02 (0.02)
Samoan	0.09	0.08	-0.01 (0.03)
Filipino	0.07	0.09	0.03 (0.02)
Japanese	0.06	0.07	0.01 (0.02)
Veteran	0.05	0.07	0.02 (0.02)
U.S. Citizen	0.93	0.95	0.02 (0.02)
N	316	190	-

The F-statistic and P-value from a regression of the treatment indicator on the variables listed in this table are 0.91 and 0.53, respectively.

a. The indicated difference is statistically significant at the one percent level of confidence.
b. The indicated difference is statistically significant at the five percent level of confidence.
c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 2
Average Values for Criminal History at Time of Arrest and Current Arrest Charges for HOPE Treatment and Control Group Members and Tests for Criminal History and Current Charge Covariate Balance

	Control Group	Treatment Group	Difference, Treatment – Control (Standard Error)
Number of Prior arrests at time of current arrest			
Felony	1.82	2.01	0.19 (0.27)
Com. Cor. Viol.	5.44	5.41	-0.03 (0.69)
Misdemeanor	7.78	7.28	-0.49 (0.98)
Other	0.55	0.62	0.07 (0.12)
Number of Prior Convictions			
Felony	0.68	0.75	0.07 (0.12)
Misdemeanor	3.73	3.52	-0.22 (0.57)
Age at first arrest	23.52	24.38	0.87 (0.66)
Days Between Arrest and Supervised Release	126.33	110.04	-16.29 (10.94)
Most Serious Arrest Charge, Current Offense			
Felony A: Person	0.03	0.02	-0.01 (0.01)
Felony A: Property	0.00	0.01	0.005 (0.004)
Felony B: Person	0.06	0.02	-0.04 (0.02)[b]
Felony B: Property	0.11	0.13	0.02 (0.03)
Felony B: Drug	0.01	0.05	0.04 (0.01)[a]
Felony B: Other	0.02	0.02	-0.01 (0.01)
Felony C: Person	0.12	0.12	0.00 (0.03)
Felony C: Property	0.29	0.28	-0.01 (0.04)
Felony C: Drug	0.21	0.18	-0.04 (0.04)
Felony C: Other	0.12	0.15	0.03 (0.03)
Arrest Info Missing	0.02	0.02	0.00 (0.01)
N	316	190	-

The F-statistic and P-value from a regression of the treatment indicator on the variables listed in this table are 1.20 and 0.25, respectively.

a. The indicated difference is statistically significant at the one percent level of confidence.
b. The indicated difference is statistically significant at the five percent level of confidence.
c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 3
Average Outcomes by Treatment Group and Differences in Average Outcomes for Number of Revocations by Type, and for Permanent, Case-Closing Revocations

Panel A: All Revocations Regardless of Timing			
	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
Revocations	0.294 (0.027)	0.853 (0.078)	0.558 (0.082)[a]
Abscond/FTA/FTR	0.139 (0.021)	0.432 (0.047)	0.292 (0.051)[a]
Abscond	0.054 (0.013)	0.032 (0.013)	-0.022 (0.017)
FTA	0.000 (0.000)	0.021 (0.010)	0.021 (0.010)[b]
FTR	0.098 (0.018)	0.384 (0.046)	0.286 (0.049)[a]
Failed Drug Test	0.104 (0.018)	0.379 (0.055)	0.274 (0.057)[a]
Drug Treatment	0.104 (0.018)	0.163 (0.030)	0.059 (0.034)[c]
New Arrest	0.028 (0.009)	0.016 (0.009)	-0.013 (0.013)
Other	0.104 (0.017)	0.205 (0.031)	0.101 (0.036)[a]
Warrants issued	0.272 (0.024)	0.347 (0.038)	0.075 (0.046)
Panel B: Permanent Revocations			
	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
SR permanently revoked	0.316 (0.026)	0.121 (0.024)	-0.195 (0.035)[a]
FTA	0.025 (0.009)	0.032 (0.013)	0.006 (0.015)
New Drug Activity	0.041 (0.011)	0.011 (0.007)	-0.031 (0.013)[b]
New Charges	0.022 (0.008)	0.005 (0.005)	-0.017 (0.009)[c]
Other	0.228 (0.024)	0.074 (0.019)	-0.154 (0.030)[a]

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 4
Treatment Effect Estimates on Number of Revocations by Type, and on Permanent, Case-Closing Revocations

Panel A: All Revocations Regardless of Timing			
	Control Group Average	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
Revocations	0.294	0.558 (0.082)[a]	0.555 (0.083)[a]
Abscond/FTA/FTR	0.139	0.305 (0.050)[a]	0.309 (0.049)[a]
Abscond	0.054	0.009 (0.018)	0.014 (0.019)
FTA	0.000	0.022 (0.011)[c]	0.025 (0.012)[b]
FTR	0.098	0.269 (0.049)[a]	0.266 (0.047)[a]
Failed Drug Test	0.104	0.272 (0.057)[a]	0.249 (0.054)[a]
Drug Treatment	0.104	0.069 (0.036)[c]	0.074 (0.036)[b]
New Arrest	0.028	-0.007 (0.011)	-0.003 (0.011)
Other	0.104	0.075 (0.037)[b]	0.093 (0.037)[b]
Panel B: Permanent Revocations			
	Control Group Average	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
SR permanently revoked	0.316	-0.190 (0.038)[a]	-0.170 (0.038)[a]
FTA	0.025	-0.001 (0.016)	0.008 (0.016)
New Drug Activity	0.041	-0.040 (0.016)[b]	-0.043 (0.018)[b]
New Charges	0.022	-0.010 (0.011)	-0.004 (0.009)
Other	0.228	-0.138 (0.031)[a]	-0.131 (0.032)[a]

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 5
Average Outcomes by Treatment Group and Differences in Average Outcomes for All Drug Tests Pooled, All Tests pooled for First Five Tests per Subject, and Separate Estimates per Test Order

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control	N
All tests pooled	0.217 (0.014)	0.167 (0.012)	-0.050 (0.018)[a]	1,870
All tests pooled through five per subject	0.215 (0.016)	0.176 (0.016)	-0.039 (0.023)[c]	1,227
Test #1	0.188 (0.028)	0.151 (0.029)	-0.036 (0.040)	344
Test #2	0.242 (0.034)	0.163 (0.031)	-0.079 (0.046)[b]	302
Test #3	0.223 (0.038)	0.171 (0.034)	-0.052 (0.051)	244
Test #4	0.261 (0.046)	0.204 (0.041)	-0.057 (0.061)	190
Test #5	0.152 (0.044)	0.222 (0.046)	0.070 (0.064)	147

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 6
Estimated Impacts of Hope Treatment on Drug Test Failures: All Tests Pooled, All tests pooled for First Five Tests per Subject, and Separate Estimates per Test Order

	Control Group Mean	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)	N
All tests pooled[#]	0.217	-0.061 (0.034)[c]	-0.067 (0.032)[b]	1,870
All tests pooled through five per subject[#]	0.215	-0.055 (0.033)[c]	-0.055 (0.031)[c]	1,227
Test #1	0.188	-0.065 (0.043)	-0.052 (0.044)	344
Test #2	0.242	-0.102 (0.051)[c]	-0.069 (0.053)	302
Test #3	0.223	-0.055 (0.051)	-0.078 (0.051)	244
Test #4	0.261	-0.054 (0.063)	-0.071 (0.071)	190
Test #5	0.152	0.046 (0.064)	0.027 (0.074)	147

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2. Drug tests failures include positive tests, tests where the subjects was noted as tampering, and tests where the subject was a no-show.
#. Standard errors for these models are clustered by individual to account for multiple tests for individual.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 7
Estimated Impacts of Hope Treatment on Drug Test Failures for Sub Groups with at Least Two, Three, Four, and Five Drug Tests: All Tests per Group Pooled and Individual Test Results

	Control Group Mean	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)	N
Subjects with at least two tests				
Tests pooled[#]	0.205	-0.082 (0.040)[b]	-0.062 (0.038)[c]	604
Test #1	0.167	-0.063 (0.043)	-0.056 (0.042)	302
Test #2	0.242	-0.102 (0.051)[c]	-0.069 (0.054)	302
Subjects with at least three tests				
Tests pooled[#]	0.231	-0.105 (0.040)[a]	-0.104 (0.037)[a]	732
Test #1	0.198	-0.104 (0.048)[b]	-0.098 (0.049)[b]	244
Test #2	0.272	-0.155 (0.057)[a]	-0.135 (0.059)[b]	244
Test #3	0.223	-0.055 (0.051)	-0.077 (0.051)	244
Subjects with at least four tests				
Tests pooled[#]	0.241	-0.109 (0.043)[b]	-0.123 (0.041)[a]	760
Test #1	0.228	-0.160 (0.057)[a]	-0.157 (0.058)[a]	190
Test #2	0.283	-0.183 (0.064)[a]	-0.176 (0.064)[a]	190
Test #3	0.196	-0.039 (0.056)	-0.090 (0.056)	190
Test #4	0.261	-0.054 (0.063)	-0.071 (0.071)	190
Subjects with at least five tests				
Tests pooled[#]	0.191	-0.044 (0.043)	-0.064 (0.038)[c]	735
Test #1	0.182	-0.117 (0.066)[c]	-0.167 (0.066)[a]	147
Test #2	0.212	-0.113 (0.074)[c]	-0.118 (0.078)	147
Test #3	0.182	-0.016 (0.066)	-0.056 (0.069)	147
Test #4	0.227	-0.018 (0.071)	-0.007 (0.080)	147
Test #5	0.152	0.046 (0.064)	0.027 (0.074)	147

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2. Drug tests failures include positive tests, tests where the subjects was noted as tampering, and tests where the subject was a no-show.

#. Standard errors for these models are clustered by individual to account for multiple tests for individual.

a. The indicated coefficient is statistically significant at the one percent level of confidence.

b. The indicated coefficient is statistically significant at the five percent level of confidence.

c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 8
Average Outcomes by Treatment Group and Difference in Average Outcomes for Average Jail Days Served Between the Date of Randomization and the Case Disposition Date for Revocations and for Other Jail Admissions Occurring on or Before the Case Disposition Date

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
Total Jail Days	49.69 (5.24)	51.76 (6.34)	1.88 (8.21)
Between SR Open Date and SR Close Date	1.07 (0.35)	20.40 (2.87)	19.34 (2.89)[a]
Between SR Close Date and Case Disposition Date	48.62 (5.18)	31.17 (5.66)	-17.45 (7.658)[b]

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 9
Treatment Effect Estimates on Jail Incarceration Day: All Jail Days between Randomization and Case Disposition, Days Between Randomization and Supervised-Release Case Closure, and Days Between Supervised Release Case Closure and Case Disposition Date

	Control Group Average	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
Total Jail Days	49.69	4.94 (8.57)	7.94 (8.79)
Pre-SR Closure	1.06	19.85 (3.10)[a]	19.21 (3.05)[a]
Post-SR Closure, Pre-Case Disposition	48.62	-14.90 (8.04)[c]	-11.27 (8.35)

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 10
Average Outcomes by Treatment Group and Difference in Average Outcomes for New Arrests by Type Occurring During the Pre-Trial Period

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
All Pre-Trial Arrests	0.828 (0.086)	0.753 (0.086)	-0.076 (0.122)
New Crime	0.385 (0.056)	0.226 (0.039)	-0.159 (0.068)[b]
Com. Cor. Viol.	0.443 (0.045)	0.527 (0.070)	0.084 (0.083)
Felony Arrests	0.204 (0.031)	0.118 (0.028)	-0.086 (0.042)[b]
Misd. Arrests	0.382 (0.058)	0.366 (0.049)	-0.016 (0.076)
Violation/other	0.243 (0.028)	0.269 (0.043)	0.026 (0.052)
Person offenses	0.065 (0.019)	0.038 (0.016)	-0.027 (0.025)
Property offenses	0.178 (0.032)	0.102 (0.024)	-0.076 (0.039)[c]
Drug offenses	0.039 (0.012)	0.027 (0.012)	-0.011 (0.016)
Other offenses	0.091 (0.019)	0.027 (0.012)	-0.064 (0.022)[c]
Com. Cor. Viol.	0.456 (0.046)	0.559 (0.073)	0.103 (0.086)

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 11
Estimated Impacts of Hope on New Arrests by type Occurring During the Pre-Trial Period

	Control Group Mean	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
All Pre-Trial			
Arrests	0.828	-0.008(0.117)	0.078 (0.114)
New Crime	0.385	-0.125 (0.063)[b]	-0.093 (0.062)
Com. Cor. Viol.	0.443	0.117 (0.083)	0.172 (0.083)[b]
Felony Arrests	0.204	-0.076 (0.045)[c]	-0.076 (0.045)[c]
Misd. Arrests	0.382	0.008 (0.073)	0.067 (0.073)
Violation/other	0.242	0.061 (0.051)	0.087 (0.053)
Person offenses	0.065	-0.016 (0.025)	-0.003 (0.027)
Property offenses	0.178	-0.064 (0.037)[c]	-0.064 (0.039)
Drug offenses	0.039	-0.006 (0.017)	-0.005 (0.015)
Other offenses	0.091	-0.061 (0.023)[a]	-0.047 (0.024)[b]
Pre-Trial Mis.	0.456	0.139 (0.086)	0.197 (0.087)[b]

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 12
Average Outcomes by Treatment Group and Difference in Average Outcomes for the Likelihood that the Current Case Results in any Conviction or Results in a Felony Conviction

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
Conviction	0.764 (0.024)	0.683 (0.034)	-0.081 (0.041)[c]
Felony conviction	0.722 (0.026)	0.618 (0.036)	-0.103 (0.044)[b]
DAG/No Contest Plea	0.152 (0.020)	0.161 (0.027)	0.009 (0.020)

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 13
Estimated Impacts of Hope on the Likelihood that the Current Case Results in any Conviction or Results in a Felony Conviction

	Control Group Average	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
Conviction	0.764	-0.068 (0.043)	-0.063 (0.043)
Felony conviction	0.722	-0.082 (0.046)[c]	-0.082 (0.046)[c]
DAG/No Contest Plea	0.152	0.018 (0.028)	0.023 (0.034)

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 14
Average Outcomes by Treatment Group and Difference in Average Outcomes for New Arrests by Type During the 180-Day Period Following Case Disposition

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
All Arrests	0.351 (0.040)	0.287 (0.045)	-0.065 (0.060)
New Crime	0.162 (0.024)	0.129 (0.032)	-0.034 (0.040)
Com. Cor. Viol.	0.189 (0.028)	0.158 (0.030)	-0.031 (0.041)
Felony arrests	0.064 (0.016)	0.070 (0.023)	0.006 (0.028)
Misd. Arrests	0.179 (0.028)	0.123 (0.028)	-0.056 (0.039)
Violation/other	0.108 (0.020)	0.094 (0.024)	-0.015 (0.031)
Person offenses	0.027 (0.009)	0.023 (0.012)	-0.004 (0.014)
Property offenses	0.071 (0.015)	0.058 (0.020)	-0.012 (0.025)
Drug offenses	0.020 (0.009)	0.012 (0.008)	-0.009 (0.012)
Other offenses	0.041 (0.012)	0.035 (0.016)	-0.005 (0.020)
Com. Cor. Viol.	0.193 (0.028)	0.158 (0.030)	-0.034 (0.041)

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 15
Estimated Impacts of Hope on New Arrests by Type During the 180-Day Period Following Case Disposition

	Control Group Mean	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
All Arrests	0.351	-0.106 (0.063)[c]	-0.085 (0.063)
New Crime	0.162	-0.055 (0.042)	-0.054 (0.041)
Com. Cor. Viol.	0.189	-0.050 (0.045)	-0.031 (0.046)
Felony arrests	0.064	-0.002 (0.028)	-0.010 (0.027)
Misd. Arrests	0.179	-0.083 (0.043)[c]	-0.067 (0.044)
Violation/other	0.108	-0.021 (0.033)	-0.007 (0.035)
Person offenses	0.027	0.007 (0.014)	0.004 (0.015)
Property offenses	0.071	-0.026 (0.026)	-0.016 (0.027)
Drug offenses	0.020	-0.010 (0.012)	-0.015 (0.012)
Other offenses	0.041	-0.021 (0.022)	-0.023 (0.021)
Com. Cor. Viol.	0.193	-0.055 (0.045)	-0.034 (0.046)

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2. The estimates in this table are based on the 467 observations for which the available post-disposition observations period is at least 180 days.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 16
Average Outcomes by Treatment Group and Difference in Average Outcomes for New Arrests by Type During the 365-Day Period Following Case Disposition

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
All Arrests	0.655 (0.061)	0.594 (0.088)	-0.061 (0.107)
New Crime	0.296 (0.039)	0.275 (0.054)	-0.020 (0.067)
Com. Cor. Viol.	0.360 (0.044)	0.319 (0.056)	-0.041 (0.070)
Felony arrests	0.109 (0.022)	0.119 (0.030)	0.010 (0.037)
Misd. Arrests	0.345 (0.043)	0.319 (0.059)	-0.026 (0.073)
Violation/other	0.202 (0.033)	0.156 (0.038)	-0.046 (0.050)
Person offenses	0.052 (0.015)	0.081 (0.027)	0.029 (0.030)
Property offenses	0.097 (0.020)	0.106 (0.029)	0.009 (0.035)
Drug offenses	0.052 (0.015)	0.019 (0.011)	-0.034 (0.018)[c]
Other offenses	0.079 (0.019)	0.069 (0.027)	-0.010 (0.032)
Com. Cor. Viol.	0.375 (0.044)	0.319 (0.056)	-0.056 (0.071)

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Table 17
Estimated Impacts of Hope on New Arrests by Type During the 365-Day Period Following Case Disposition

	Control Group Mean	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
All Arrests	0.655	-0.137 (0.110)	-0.101 (0.107)
New Crime	0.296	-0.059 (0.069)	-0.049 (0.065)
Com. Cor. Viol.	0.360	-0.077 (0.073)	-0.051 (0.078)
Felony arrests	0.109	0.001 (0.038)	-0.006 (0.039)
Misd. Arrests	0.345	-0.080 (0.077)	-0.063 (0.073)
Violation/other	0.202	-0.057 (0.052)	-0.033 (0.055)
Person offenses	0.052	0.039 (0.030)	0.037 (0.029)
Property offenses	0.097	-0.007 (0.036)	-0.010 (0.037)
Drug offenses	0.052	-0.037 (0.020)[c]	-0.039 (0.021)[c]
Other offenses	0.079	-0.034 (0.034)	-0.025 (0.032)
Com. Cor. Viol.	0.365	-0.096 (0.074)	-0.063 (0.078)

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2. The estimates in this table are based on the 437 observations for which the available post-disposition observations period is at least 365 days.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.

Table 18
Summary of Principal Estimates Along with Unadjusted P-Values and P-Values Adjusted for the Family-Wise Error Rate

Outcome	Difference in means, Treatment-Control	Standard Error	Unadjusted P-Value	P-Values Adjusted for Family Wise Error Rate[a]
All revocations	0.558	0.082	0.000	0.000
Permanent revocations	-0.195	0.035	0.000	0.000
Failed drug test	-0.050	0.018	0.007	0.057
Jail days	1.886	8.219	0.819	0.955
Arrested, new crime while on SR	-0.159	0.067	0.019	0.127
Arrested, CC violation on SR	0.084	0.083	0.315	0.826
Any conviction	-0.081	0.042	0.054	0.259
Felony conviction	-0.103	0.044	0.019	0.127
DAG/No contest	0.009	0.034	0.787	0.955
Arrested, new crime 180 days post disposition	-0.034	0.041	0.409	0.883
Arrest, CC violation 180 days post disposition	-0.031	0.041	0.447	0.883

a. We use the Westfall and Young free step-down resampling methodology to compute adjusted p-values. Specifically, we use the wyoung procedure in STATA using 10,000 replications developed by Jones, Molitor and Reif (2018).

Appendix Table A1
Average Outcomes by Treatment Group and Differences in Average Outcomes for Number of Revocations Occurring Within the Thirty-Day Window Preceding the Supervised Release Close Date by Type

	Control Group Average	Treatment Group Average	Difference, Treatment minus Control
Revocations	0.253 (0.024)	0.137 (0.025)	-0.116 (0.035)[a]
Abscond/FTA/FTR	0.123 (0.019)	0.058 (0.019)	-0.066 (0.026)[b]
Abscond	0.054 (0.013)	0.011 (0.007)	-0.043 (0.015)[a]
FTA	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
FTR	0.082 (0.015)	0.047 (0.017)	-0.035 (0.023)
Failed Drug Test	0.085 (0.016)	0.058 (0.017)	-0.027 (0.023)
Drug Treatment	0.089 (0.016)	0.042 (0.015)	-0.047 (0.021)[b]
New Arrest	0.025 (0.009)	0.000 (0.000)	-0.025 (0.009)[a]
Other	0.085 (0.016)	0.053 (0.016)	-0.033 (0.023)

Standard errors are in parentheses. Standard errors on the differences in means are heteroscedastic-robust standard errors.

a. The indicated difference is statistically significant at the one percent level of confidence.

b. The indicated difference is statistically significant at the five percent level of confidence.

c. The indicated difference is statistically significant at the ten percent level of confidence.

Appendix Table A2
Supplemental Treatment Effect Estimates on Number of Revocations Occurring Within the Thirty-Day Window Preceding the Supervised Release Close Date by Type

	Control Group Average	Treatment Effect, Specification (1)	Treatment Effect, Specification (2)
Revocations	0.253	-0.099 (0.038)[a]	-0.078 (0.040)[b]
Abscond/FTA/FTR	0.123	-0.041 (0.026)	-0.040 (0.028)
Abscond	0.054	-0.015 (0.013)	-0.014 (0.015)
FTA	0.000	-	-
FTR	0.082	-0.035 (0.023)	-0.035 (0.025)
Failed Drug Test	0.085	-0.025 (0.025)	-0.022 (0.026)
Drug Treatment	0.089	-0.040 (0.022)[c]	-0.029 (0.023)
New Arrest	0.025	-0.020 (0.008)[b]	-0.015 (0.007)[b]
Other	0.085	-0.042 (0.026)	-0.033 (0.026)

Robust standard errors are in parentheses. Specification (1) includes only the indicator variable for the treatment group and an indicator variable for being randomized after April 1, 2015. Specification (2) adds controls for all of the variables listed in Tables 1 and 2.

a. The indicated coefficient is statistically significant at the one percent level of confidence.
b. The indicated coefficient is statistically significant at the five percent level of confidence.
c. The indicated coefficient is statistically significant at the ten percent level of confidence.