Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI

Interim Report

Respectfully Submitted

D. James Greiner
Ryan Halen
Matthew Stubenberg
Access to Justice Lab
Harvard Law School

Christopher L. Griffin, Jr.
James E. Rogers College of Law
University of Arizona

August 18, 2020

Table of Contents

Executive Summary 4

Introduction 6

I. Implementation and Evaluation of the PSA-DMF System in Dane County 6

A. The PSA-DMF System 6

B. Dane County Basics 8

C. From Arrest to First Appearance in Dane County 9

D. First Appearance: Dane County Practice and Governing Wisconsin Law 10

E. Implementation of the PSA-DMF System in Dane County 15

F. Field Study Description 20

1. Field Study Construction 20

2. Randomization Protocol 21

3. Data And Data Processing 23

a. Data Available 23

i. Data Sources 24

ii. Data Provided 25

iii. Analysis Dataset Description 26

b. Data Still Sought 27

c. Data Unlikely To Become Available 27

II. Findings 30

A. Qualitative Findings 30

B. Quantitative Findings 31

1. Overview of Basic Analysis Techniques 31

2. Analyses Completed 34

a. Balance of Observed Covariates 34

b. Effect of the PSA-DMF System on the Commissioner's First Appearance Decisions 40

c. Effect of the PSA-DMF System on Four Primary Outcome Variables 47

i. Effect of the PSA-DMF System on FTA 55

ii. Effect of the PSA-DMF System on N(V)CA 58

iii. Effect of the PSA-DMF System on Predisposition Incarceration 63
iv. Effect of the PSA-DMF System on Demographic Fairness Measures 68
(1). False Positive Rates 70
(2). False Negative Rates 76
(3). Overall Procedure Error Rates 82
(4). Risk Factor Separated Outcome Based Disparate Impacts 89
3. Anticipated Future Analyses 97
III. Appendices 99
A. Dane Decision Making Framework 99
B. State Public Defender Interview Questions 101
C. Sample PSA-DMF System Printout 102
D. Rule 206 One Year Report 103
E. Wisconsin Bail Statute Criminal Procedure 969.01 104
F. Definitions 105
G. Miscellaneous Data Tables 108
H. PSA Assessor Risk Factor Worksheet 110

Executive Summary

This interim report discusses the Access to Justice ("A2J") Lab's findings to date with respect to its randomized evaluation of the Public Safety Assessment-Decision Making Framework ("PSA-DMF") System in Dane County, Wisconsin.

The PSA is a risk assessment instrument. Scientists funded by Arnold Ventures produced the PSA by reviewing past data on criminal history, demographics, new crimes, failure to appear ("FTA"), and other potential risk factors. The goal of the PSA is to use criminal history information (age is also used, but not race or gender) to produce scores ranking the level of risk that an individual who has been arrested and released pending trial will fail to appear, be re-arrested for a new crime, or be re-arrested for a new violent crime. The PSA scores are applied to the DMF, which brings together the information from the PSA with a community's local policies and values, its laws, and its resources to provide a recommendation regarding pretrial release and, if release is obtained, a supervision level. Judicial officers may use the PSA-DMF System when deciding whether to release an individual before trial, and this decision rests always with the judicial officer. The PSA-DMF System is one tool in the toolbox that the judicial officer can draw upon in the exercise of their professional discretion.

The A2J Lab is a research lab within Harvard Law School. The A2J Lab works with legal systems in jurisdictions around the country to run randomized field experiments. These field experiments are like the randomized clinical trials used for new drugs. The experiments usually divide cases into a treated group that receives a new, to-be-tested system and a control group that receives business as usual. Just as in medical experiments, which compare the health of people receiving a new drug to people receiving an old drug, the purpose of these experiments is to find out whether the new system is better than standard practice. As in those medical experiments, something similar to a coin flip or a dice roll is used to decide who gets the new system and who gets the currently existing system. This kind of randomized experiment is the gold standard in finding out what works.

In the late spring of 2017, Dane County started providing the PSA-DMF System information to judicial officers deciding how much and what kind of bail and supervision to assign to individuals who have been arrested; such decisions affect whether the individual will be released or remain in jail until trial. Working with the A2J Lab and Arnold Ventures, Dane County's randomized field experiment began a month later. The experiment used something like a coin flip to divide cases into two groups. In the treated group, the judicial officer who was deciding how much and what kind of bail and pretrial supervision to assign received a paper printout with the PSA-DMF System information on it. In the control group, the judicial officer did not receive the paper printout. In other words, the control group received standard practice, and the treated group received the new system in which the judicial officer got the PSA-DMF System printout.

Dane County randomized cases from the middle of 2017 until the end of 2019, and information on each arrestee will be collected for a two-year period after the judicial officer's bail/supervision level decision in each case. The A2J Lab agreed to analyze data the County provided to compare the treated and control groups on (among other things) measurements of racial fairness, number of days incarcerated (if any), FTA, new criminal activity, and new violent criminal activity. The Lab also observed hearings, reviewed documents, and interviewed several Dane County officials so as to understand the County's criminal justice operations and its interests. Because of the two-year follow-up period, the County will provide the A2J Lab full information on all arrestees in the study sometime in early 2022. If the data are provided then, the A2J Lab's final report will be available in the summer of 2022.

The A2J Lab also agreed to provide an interim report analyzing decisions from the middle of 2017 to the middle of 2018, based on a one-year follow-up period for those cases. In summary, the A2J Lab's conclusions in the interim report are as follows:

- Qualitatively, Dane County's commitment to improving its pretrial system appears strong. Seeking to reduce the footprint of the criminal justice system in the community, the County made several changes to pretrial operations before and during the study period, such as shortening the time from arrest to first appearance, reducing the frequency of drug testing, increasing the availability of text message reminders for court hearings, and increasing the use of citations in lieu of arrests. Criminal justice officials worked to provide information to the A2J Lab to facilitate the Lab's independent evaluation in a spirit of learning and a desire to improve.
- Quantitatively, the A2J Lab does not yet have enough cases, or enough of a follow-up period on those cases, to make firm conclusions about whether it is better or worse to make the PSA-DMF System printout available to the judicial officer before assignment of bail (if any) and release conditions.
    - The A2J Lab has data on cases for only 12 out of the 30 months that the study ran.
    - For those 12 months of cases, the A2J Lab has data on only half of the two-year follow-up period.
    - The result is that the A2J Lab has one year's worth of follow-up data on 2262 arrests, which constitutes in a rough sense about 20% of the information that will eventually become available. It is too early to draw conclusions about whether the PSA-DMF System is positive, negative, or neutral for Dane County.
- The limited data available thus far, not enough to draw firm conclusions, suggest the following:
    - There is some evidence that providing the PSA-DMF System printout to the judicial officer caused a change in the officer's decisions.
        - Generally, when the printout indicated that an individual presented lower risk, the judicial officer was less likely to require cash bail or, if cash bail was required, the amount was lower than in comparable control group cases. The opposite was generally true in the treated group when the printout indicated that the individual presented higher risk, as compared to the control group. This change was statistically significant but mild, and we cannot yet tell whether the change is policy-relevant.
        - Treated group cases varied less in bail types and amounts than did control group cases. This change was strong and statistically significant.
    - There was no statistically significant difference between treated versus control group cases with respect to:
        - various measures of the racial fairness of the judicial officer's decisions;
        - the number of days (if any) of pretrial incarceration;
        - the frequency with which arrestees failed to appear at court dates; or
        - the frequency with which arrestees were arrested for new crimes, including new violent crimes, during the pretrial period.

The A2J Lab is grateful to Arnold Ventures and to Dane County for the opportunity to work on this project. The Lab looks forward to the final report in mid-2022, assuming provision of the required data.

Introduction

This First Interim Report discusses the Access to Justice ("A2J") Lab's findings to date with respect to its randomized control trial ("RCT") evaluation of the Public Safety Assessment-Decision Making Framework ("PSA-DMF") System in Dane County, Wisconsin. This First Interim Report provides the results of the A2J Lab's analysis of custodial First Appearance hearing events[1] from the first year of randomization, with the analysis dataset covering events occurring up to two years after the RCT kickoff. Assuming the provision of required data, the Lab anticipates producing its Final Report in mid-2022 providing our analysis of the First Appearance release decision events from the full 30 months of randomization, with the analysis dataset covering events occurring within two years of each release decision event.

This Report proceeds in two Parts. Part I addresses the PSA-DMF System, Dane County and its implementation of the System, and the construction and implementation of the RCT. Part II describes the Lab's findings, qualitative and quantitative.

I. Implementation and Evaluation of the PSA-DMF System in Dane County

This Part provides the background needed to understand the findings described in Part II. It proceeds in six subparts. It also includes the bases for our qualitative findings. Subpart I.A. briefly describes the PSA-DMF System. Subpart I.B provides basic information about Dane County. Subpart I.C describes the Dane County system from arrest to First Appearance hearing. Subpart I.D describes Dane County's First Appearance hearing procedures and practices, including applicable Wisconsin law. Subpart I.E describes the implementation of the PSA-DMF system in Dane County. Subpart I.F describes the Lab's field operation.

A. The PSA-DMF System

This Subpart briefly describes the PSA-DMF System for persons unfamiliar with its operation.

---

[1] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). For ease of reference, we use the phrase "First Appearance hearing" in this report to refer to the first judicial hearing of a still-in-custody arrestee before a Dane County Commissioner (see immediately below), *i.e.*, to *custodial* first appearances. Dane County also administers first appearances in criminal proceedings with respect to individuals who were not physically arrested (citation proceedings), and with respect to individuals who bail out according to a statewide bail schedule shortly after being arrested. Our use of the phrase "First Appearance hearing" does not encompass either of these types of events.

The PSA is a predisposition risk assessment instrument ("PRAI") that judges[2] may use when deciding whether to release or detain an arrestee before trial. The PSA takes as inputs data on the arrestee's criminal history, current charge, and age. These inputs (some in combination) are assigned an initial set of integer weights. Those integer weights are further processed to produce two risk scores that can take on values of 1-6, with higher numbers signaling higher risk. The first score bands arrestees on risk of being arrested or cited for new criminal activity ("NCA") if released pending disposition. The second 1-6 scale bands arrestees on risk of failure to appear ("FTA") at the case's court hearings. The PSA also flags arrestees to signal an elevated risk of being arrested for new violent criminal activity ("NVCA") before disposition; the flag operates as a 0-1 variable.[3]

Arnold Ventures ("AV"), formerly the Laura and John Arnold Foundation, funded researchers to produce the PSA in a project that concluded in 2013.[4] AV and the developing researchers sought to construct a PRAI that (i) did not require inputs from an expensive and potentially legally fraught interview with the arrestee, and (ii) produced risk bands informative in any jurisdiction in the United States. "Validation" studies, in which researchers assess whether the PSA's risk bands correspond to differences in released arrestees' misbehavior rates, have been completed in several jurisdictions.[5]

The PSA scores are typically accompanied by the Decision Making Framework ("DMF"), which incorporates the objective information from the PSA with community-specific determinations regarding local policy and values, state statutes, and jurisdictional resources to produce a release recommendation as well as (in locations that choose to use it this way) a supervision level to be imposed if the arrestee is released. The PSA scores rely on objective data, and the scoring system is the same in all jurisdictions. The DMF recommendation system can be different in each jurisdiction.[6] The decision about whether to release or detain an arrestee, and the level of supervision accompanying any release, rests always with the judge. The PSA does not produce a recommendation, and the DMF's recommendation is not binding.

Dozens of jurisdictions have implemented the PSA, including three entire states and several large cities.[7]

---

[2] We use the term "judge" when referring generically to judicial officials who make initial release decisions. Jurisdictions assign various titles to such officials, including "magistrate" or "Commissioner." In Dane County, for example, "Commissioners" conduct First Appearance hearings, and "judges" ordinarily handle case proceedings after First Appearance. When referring to Dane County proceedings in particular, we use the term "Commissioner."

[3] A complete discussion of the PSA's inputs, initial integer weights, and processing of those weights into 1-6 FTA and NCA risk bands is available at https://www.psapretrial.org/about/factors (last visited Feb. 19, 2020).

[4] Support for the assertions in this paragraph appear in https://www.psapretrial.org/about/background (last visited Feb. 19, 2020), which provides a more detailed discussion of the PSA's features and development, as well as links for additional information.

[5] The Access to Justice Lab is currently pursuing validation efforts in four counties.

[6] A copy of the Dane County DMF appears as Appendix A.

[7] See https://www.psapretrial.org/about#jurisdictions-united-states (last visited Feb. 19, 2020).

B. Dane County Basics

This Subpart provides basic information about Dane County and its demographics.

Dane County, Wisconsin is the state's second-most populous jurisdiction (estimated by the U.S. Census Bureau to be 542,364 in 2018[8]), behind only Milwaukee County. It includes the capital city of Madison, which is also home to the University of Wisconsin's flagship campus.

The most recent demographic estimates indicate that the County's population is composed of 84% white, about 5% black, and 5% Asian.[9] Across all races, Dane County's residents are about 5% Hispanic.[10] The age profile is distributed such that just over half of the population is aged between 20 and 54.[11] As expected in the presence of one of the nation's premier public higher education institutions, 96% of the population has at least a high school degree.[12] The median household income is approximately $68,000, and 12% of Dane's residents live below the poverty line.[13] Both figures are slightly higher than the same statistics for the United States as a whole.[14]

The presence of the University of Wisconsin-Madison is an important component of the Dane County PSA/DMF System Study. The County is able to draw on talented graduates of its highly-regarded social science programs. Combined with an atmosphere of intellectual curiosity, the judicial, legislative, and executive branches are amenable to rigorous program evaluations. The local labor market also reflects this commitment to scientific exploration. For example, Epic Systems, Inc., a health services software corporation, is the jurisdiction's largest employer.

Dane County's citizen engagement with criminal justice, especially addressing racially disparate outcomes, is evident. By one account,[15] a debate took shape in 2013, after the publication of the Wisconsin Council on Children and Families' "Race to Equity" report.[16] For an example of a product of this debate, in late 2017, a group petitioned to redirect over $100 million in County

---

[8] U.S. Census Bureau, American FactFinder, https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml?src=bkmk (last visited Feb. 28, 2020).

[9] U.S. Census Bureau, ACS Demographic and Housing Estimates, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk (last visited Feb. 28, 2020).

[10] Id.

[11] Id.

[12] U.S. Census Bureau, American FactFinder, supra note 8.

[13] Id.

[14] See Jessica Semega et al., U.S. Census Bureau, Income and Poverty in the United States: 2018 figs.1, 7 (2019), available at https://www.census.gov/library/publications/2019/demo/p60-266.html.

[15] Steven Elbow, "*None of This Has Changed": Madison's Racial Disparities Have Gotten Worse, Despite Decades of Reports, Task Forces and Funded Programs*, Capital Times, Mar. 4, 2020. https://madison.com/ct/news/local/govt-and-politics/none-of-this-has-changed-madison-s-racial-disparities-have/article_0490a398-46f5-54ea-af5c-66ff1a32dfac.html.

[16] Wis. Council on Children & Families, Race to Equity: A Baseline Report on the State of Racial Disparities in Dane County (2013), available at http://racetoequity.net/wp-content/uploads/2016/11/WCCF-R2E-Report.pdf.

Jail funds toward alternatives to incarceration rather than additional construction of holding facilities.[17] They cited statistics purporting to show that 48% of the incarcerated in Dane were black even though only 5-6% of the County's residents were black. Commentary followed in which various community organizations sought change in County processes and procedures based on concerns of racial equity.[18]

C. From Arrest to First Appearance in Dane County

This Subpart describes the Dane County criminal justice system from arrest to First Appearance hearing.

A completed arrest by law enforcement in Dane County led to a custodial transfer to the Dane County Jail (located in the Public Safety Building) or a citation and summons to appear in court. An arrestee charged with a misdemeanor could post bail and secure release before a First Appearance before a Commissioner. The bail amounts are set by offense, on a statewide basis by the Chief Justice of the Wisconsin Supreme Court, based on the recommendations of a committee of seven circuit court judges from across Wisconsin.[19] An arrestee charged with a felony must appear before a Commissioner.[20] If the arrestee could not bond out, the resulting First Appearance hearings, which are the focus of this report, took place in the courtroom in the Public Safety Building before a presiding judge.

The timing of the First Appearance hearing was set by local Rule 206.[21] The provision covered review of bail set by a judge, the handling of non-custodial initial appearances, and the scheduling of custodial First Appearance hearings. Dane County adopted Rule 206 effective July 1, 2016, ten months before the PSA-DMF Framework was implemented. In relevant part, the Rule set a shorter timeline between booking and the First Appearance Hearing. Specifically, as long as the arrestee had no other holds (*e.g.*, immigration detention, warrants), they were required to be scheduled for First Appearance within 48 hours of the arrest. For example, an arrestee booked before 8:00 AM on a Monday was to have had an initial appearance on Tuesday. The same was true for each pair of weekdays: booking before 8:00 AM on Friday resulted in a mandatory appearance on Monday's calendar.

This rule change, combined with the adoption and launch of the PSA-DMF Framework, generates potential macro-level confounders for understanding the causal effect of the Framework. For example, the reduced time between booking and initial appearance mandated

---

[17] Emilie Burditt, D*erail the Jail Panel Discusses Ways to Fight Mass Incarceration in Dane County*, Badger Herald, Nov. 14, 2017, https://badgerherald.com/news/2017/11/14/derail-the-jail-panel-discusses-ways-to-fight-mass-incarceration-in-dane-county/.
[18] See*, e.g.*, Elbow, supra note 15.
[19] See Wis. Judicial Conference, Uniform Misdemeanor Bail Schedule (2020), available at https://wicourts.gov/publications/fees/docs/bondsched20.pdf.
[20] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).
[21] Wis. R. App. P. L DANE 206

by Rule 206 almost certainly reduced the overall time spent incarcerated among custodial arrestees booked after Framework implementation relative to those booked before.[22] As a result, the local average treatment effect of the Framework on time spent incarcerated was affected by this change (and other components of criminal justice reform adopted alongside the Framework, as discussed below), rather than the Framework's adoption in isolation. Another relevant consideration is the information that was available to the Commissioner at custodial First Appearance hearings and to the Assessors calculating PSA-DMF Framework risk scores. First Appearances occurred in Dane County whether or not the District Attorney ("DA") had filed formal charges in the arrestee's case. These cases, coded as group-file ("GF") matters on the docket, used law enforcement's recommended charges if the DA had not settled on or had enough time to choose the charging statutes. The current charge is a risk factor that partially determines the new violent criminal activity flag. As a result, the difference between law enforcement recommendations and ultimate DA charges could, in the counterfactual, affect whether the violence flag was triggered. The A2J Lab, however, received early evidence from Dane County that Rule 206 had no significant effect on the number of pre-filing initial appearances.[23]

D. First Appearance: Dane County Practice and Governing Wisconsin Law

This Subpart discusses Dane County's procedure at First Appearance, including a brief outline of Wisconsin's somewhat unusual governing law.

Before an in-custody First Appearance hearing would occur, Assessors would create a PSA-DMF System report for all eligible defendants based on the list of arrestees received from the jail each morning. Assessors would use the PSA-DMF System computer program and input the required fields after they had consulted CCAP, NCIC, and publicly available criminal justice databases from other jurisdictions. The program would produce a report in the PSA-DMF System-present condition which the Assessors would print for all relevant parties. The Assessors would complete 15-30 PSAs on average each day the Court was open.[24]

The terms "bail" and "bond" are used interchangeably in this report. Technically, the term "bond" refers to an agreement to follow particular rules or conditions.[25] A bond can sometimes be

---

[22] See Letter from Juan B. Colas, Presiding Judge, to Circuit Court Judges, Amended Rule 206: Earlier Initial Bail Review Hearings (Oct. 7, 2016) (on file with the Access to Justice Lab) (explaining that, when comparing the timing of the first bail hearing or First Appearance after Rule 206 was amended to when the same events would have occurred under the old rule, the median times from booking to hearing were 47.7 and 64.5 hours, respectively, a reduction of 35%).

[23] See Email from Colleen Clark-Bernhardt to James Greiner & Christopher Griffin, Jan. 4, 2017 (on file) ("From January 1, 2016 to June 30, 2016, except for a few days at the beginning of June, we were under old Rule 206. We drew 138 GF case numbers. . . . From July 1, 2016 to December 31, 2016, we had 144 GF case numbers under circumstances covered by new Rule 206.").

[24] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

[25] Wis. Stat Sec. 967.02(1h).

accompanied by a requirement to post money. The posting of money is considered "bail".[26] However, in practice the terms are often used interchangeably as they are in this report.[27]

A Commissioner adjudicating a First Appearance hearing in Dane County faced two primary decisions: whether to impose a "signature" or a "cash" bond, and the level and nature of supervision and conditions to be imposed upon a custodial arrestee who gained predisposition release. With respect to the first decision: Signature bonds were "no-money-down" bonds. The Commissioner specified an amount of money in the bond. The arrestee was not required to deposit money with the court to obtain release vis-à-vis the present arrest,[28] but if the arrestee failed to appear or was re-arrested or violated some other condition of release, the County had the legal right to seek forfeiture of the amount of money named in the bond. In practice, the County rarely sought such forfeiture, so a signature bond operated much like a release on one's own recognizance ("ROR").[29] In a "cash" bond, in contrast, an arrestee gained release only if they deposited the amount specified in the bond with the court. Note that Wisconsin was one of a handful of states with no commercial bail bonds industry.

The mechanism Wisconsin law provided to remand an arrestee without bond was deemed so burdensome as to be unusable.[30] Thus, operationally, to release an arrestee (at least for the purposes of the charges at hand in the particular First Appearance hearing), the Commissioner imposed a signature bond. If the Commissioner sought something other than near-automatic release, the Commissioner imposed a cash bond, determining the amount according to the law described below, but with the knowledge that a high cash bond might well result in the incarceration of the arrestee until case disposition.

Regarding the second major decision, the level of supervision assigned to an arrestee if released, Dane County featured an Office of Pretrial Services ("Pretrial"), known at the beginning of this study as the Alternatives to Incarceration Program.[31] During this study, the management structure of the Pretrial Services Office consisted of Chief Deputy Clerk of Circuit Court Kerry Widish as first-line supervisor and Social Worker Julie Beyler as office Lead Worker, both under the

---

[26] Wis. Stat Secs. 967.02(1d) and 969.001

[27] Email exchange citing Commissioner Hanson, Colleen Clark-Bernhardt & Matthew Stubenberg, Apr. 27, 2020 (on file with the Access to Justice Lab).

[28] An arrestee receiving a signature bond might have been detained for some other reason, such as an immigration hold, or because the present arrest violated a condition of probation or parole or predisposition release in some other case.

[29] ROR is available under Wisconsin law, and has been a topic of some recent discussion in Dane County. Email exchange, Dane County Commissioner Jason Hanson & D. James Greiner, Feb. 21, 2020 (on file with the Access to Justice Lab). ROR has not, however, been a part of Dane County predisposition culture, in part because it does not allow for the imposition release conditions (such as those protecting a victim or a witness). Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). We have no record of an order for ROR release in the data we analyzed for this Report. The closest we observed to ROR were two $50 signature bonds.

[30] We have no record of a case in which the Commissioner or the District Attorney attempted to invoke this process.

[31] See https://courts.countyofdane.com/pretrial-services (last visited Feb. 20, 2020).

direction of Clerk of Circuit Court Carlo Esqueda.[32] Pretrial administered four levels of supervision, numbered 0 through III.[33] Commissioners determined whether to refer an arrestee to Pretrial. Unless the Commissioner specified the supervision level in the First Appearance order,[34] Pretrial deterministically used the supervision level recommended in the arrestee's DMF for events randomized to the PSA-DMF-System-present condition, and made its own judgment for events randomized to the no-PSA-DMF-System-present condition.[35]

In choosing between a signature or a cash bond, in fixing an amount of bail, and in imposing any other conditions, the Commissioner by Wisconsin law considered a list of factors including "the nature, number and gravity of the offenses and the potential penalty the defendant faces," "the defendant's prior record of criminal convictions and delinquency adjudications, if any," "whether the defendant is currently on probation, extended supervision or parole, whether the defendant is already on bail or subject to other release conditions in other pending cases," and "whether the defendant has in the past forfeited bail or violated a condition of release or was a fugitive from justice at the time of arrest."[36] Once the Commissioner decided to fix bail, however, the amount to be imposed was to "be only . . . the amount found necessary to assure the appearance of the defendant."[37] To the extent necessary to serve "the purpose of protecting members of the community from serious bodily harm or preventing intimidation of witnesses," other reasonable conditions could be imposed.[38] This statutory framework (including both the factors to consider and the requirement that bail only be set to assure appearance) appears to be a codification of the Wisconsin Supreme Court's opinion in *Whitty v.*

[32] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

[33] A supervision level zero was created for defendants ordered into Pretrial supervision who fell into a DMF recommendation that did not carry a predisposition supervision level. For levels I through III, Pretrial used an Arnold Ventures-generated table specifying the frequency of telephone contacts, the frequency of in-person check-ins, and the availability of conditions monitoring. Email exchange, Chief Deputy Clerk of Court Kerry Widish & Matthew Stubenberg, Feb. 27, 2020 (on file with the Access to Justice Lab).

[34] Commissioners rarely if ever specify a supervision level in First Appearance orders, although they do specify the tools (such as GPS monitoring or drug testing) that Pretrial may use. Email exchange, Dane County Commissioner Jason Hanson & D. James Greiner, Feb. 21, 2020 (on file with the Access to Justice Lab).

[35] Matthew Stubenberg, Memo, "Meeting with Clerk of Circuit Court and Register in Probate," Memorializing Conversation on Dec. 12, 2019 (on file with the Access to Justice Lab). The description above comes primarily from Clerk of the Circuit Court, Carlo Esqueda, Chief Deputy Clerk of Court Kerry Widish, and Lead Social Worker Julie Byler. All three were present for this conversation.

[36] Wis. Stat. Ann. § 969.01(4). See Appendix E.

[37] Id. See also Wis. Stat. Sec. 967.02(1d), defining "bail" as an amount of money to assure appearance and the person's compliance with conditions of bond. Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

[38] Id.

State.[39] In practice, the Commissioner's primary decision on cash versus non-cash bail focused primarily on the risk of FTA at the next hearing date in the case.[40]

We note here that, according to a Dane County study, Commissioners ordered less restrictive release conditions in a majority of cases. Between 2012 and 2016, 72% of all arrestees received only a signature bond during the pendency of their trials, including 78.7% of all misdemeanors and 93.3% of all traffic cases.[41] Conversely, only 18.9% of all cases involved the imposition of cash bail.[42]

Dane County First Appearance hearings occurred at 1:30pm (2:00pm on Mondays) each non-holiday weekday afternoon in a courtroom attached to the Dane County Jail (the "Jail"). The actual courtroom was secure. Members of the public could observe hearings from an adjacent room, into which sound was piped, by looking through a large panel made of reinforced glass. The arrestee appeared in person.[43]

In contrast to some jurisdictions, Dane County featured both an Assistant District Attorney ("ADA") and a lawyer from the Wisconsin State Public Defender's Office ("SPD") in all First Appearance matters, and both offices ordinarily made some effort to prepare. On the prosecutor side, the ADA assigned to prosecute the case usually compiled a file that included the charges, the sworn law enforcement affidavit, and a short writeup of some criminal history information, including some FTAs and some prior convictions (if any). The criminal history search ordinarily encompassed a review of Wisconsin's Consolidated Court Automation Programs ("CCAP") Case Management system, and sometimes included search results from the FBI's National Crime Information Center ("NCIC"). The ADA usually compiled this file the afternoon or evening before or the morning of the corresponding First Appearance hearing. A different, single ADA reviewed this file quickly, sometimes before the First Appearance hearing, sometimes at the hearing itself, before representing the DA's office in all matters.[44]

---

[39] 149 N.W.2d 557 (Wis. 1967). See, e.g., Rohl v. State, 279 N.W.2d 731, 737 (Wis. Ct. App. 1979) ("The factors to be considered in determining bail were originally set out in Whitty v. State, and are now codified in sec. 969.01(4).") (internal citation omitted).

[40] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI,," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

[41] Judge Nicholas J. McNamara, Frequency of Signature Bonds in Dane County Criminal Cases: 2012-2016 11 tbl.1 (2018), available at https://docs.legis.wisconsin.gov/misc/lc/study/2018/1783/020_september_17_2018_meeting_10_00_a_m_lc_conference_room/sept17handout_esqueda.

[42] Id.

[43] This description stems from our in-person observations of several sessions before and during this Study. Collectively, we observed five sessions, including three "ride-alongs," in which one of us sat next to the Commissioner inside the secured courtroom.

[44] Matthew Stubenberg, Memo, "Meeting with District Attorney," Memorializing Conversation on Dec. 13, 2019 (on file with the Access to Justice Lab). The description above comes primarily from District Attorney Ismael R. Ozanne and ADA Hector Al-Homsi, both present for this conversation. At this meeting, DA Ozanne and ADA Al-Hosmi expressed frustration that the prosecuting ADA's criminal history

On the defense side, an SPD assistant/intern attempted to interview newly arrived arrestees each morning at the Jail. The SPD assistant/intern who attempted to interview arrestees should not be confused with the SPD attorney who represented arrestees at First Appearance hearings. Unless the arrestee was unavailable or refused the request to talk to the SPD, the assistant/intern recorded the basic contents of the interview onto a single-page form,[45] one copy of which was placed in a file folder, and another kept on file at the SPD's office down the street from the Jail. The sheet allowed the assistant/intern to record, among other things, information on an arrestee's ties to Dane County, including the length of residence there; the presence of family there; and employment status. Like the DA's office, the SPD assigned a single attorney to handle all matters at First Appearance, except for the rare cases in which an arrestee had retained a private attorney for the hearing. This SPD lawyer sometimes reviewed an arrestee's file the afternoon or morning before the First Appearance hearing, or perhaps at the hearing itself, with the latter sometimes made necessary by SPD's inability to discern which arrestees would undergo First Appearance on what days.[46] This inability stemmed from sometimes last-minute decisions, discussed above, regarding which arrestees would "go" on a particular date, which in turn stemmed from up-to-the-moment availability of charges and sworn law enforcement affidavits, together with a County-wide desire to have arrestees "go" to First Appearance as soon as possible.[47]

Operationally, after the arrestee was led into the Jail courtroom and the court clerk called the case, the presiding Commissioner quickly re-reviewed[48] the charges and the law enforcement affidavit for the case on a desktop computer, then provided the ADA and the SPD attorneys an opportunity to address the court. ADA statements frequently focused on charges, criminal history, and facts of the incident, as disclosed in the law enforcement affidavit. SPD statements ordinarily focused on whichever of these same aspects were favorable to the arrestee as well as on residence, family, and employment in Dane County, if any. In some cases, the Commissioner might have asked questions of either the ADA or the SPD attorney. In some cases, the Commissioner consulted CCAP on the nearby desktop to ascertain or verify some aspect of the arrestee's criminal history. The arrestee did not ordinarily address the court directly. Ordinarily, if the ADA agreed that a signature bond was appropriate, the Commissioner would so order. Upon reaching a decision, the Commissioner spoke directly to the arrestee to explain the basics of the order to be imposed and to note the next court date; arrestees received a written copy of the Commissioner's order. Each arrestee's First Appearance typically lasted 3-7 minutes. First Appearance hearings in total ordinarily lasted between 90 and 150 minutes,

search was duplicative of, and sometimes less comprehensive than, the search that the Dane County Assessors conducted to obtain the inputs for the PSA-DMF System.

[45] A copy of the SPD interview sheet appears as Appendix B.

[46] Email Exchange, Matthew Stubenberg and Catherine Dorl, Trial Division Director, Wisconsin State Public Defenders, Mar. 4, 2020 (on file with Access to Justice Lab).

[47] The DA complied with the changes to Rule 206 primarily by implementing two technological changes, the increased use of dictation technology for ADAs drafting law enforcement affidavits and a shift to electronic signatures on those affidavits. Neither shift solved the problem of last-minute inclusion of cases on the First Appearance calendar.

[48] As we observed in our ride-alongs, in most cases, the Commissioner had reviewed the charges and the law enforcement affidavit before an arrestee's First Appearance hearing.

with approximately 20-40 arrestees processed, although both the length of the hearing and the number of arrestees processed varied from day to day. Both figures were, for example, typically larger on Tuesdays and on days following holidays.[49]

After First Appearance, an arrestee's charges proceeded according to standard Dane County rules and procedures. Two features of that processing deserve mention here. First, Wisconsin law made available, at the arrestee's request, a periodic review of that arrestee's bail status. The Circuit Court Judge assigned to the case, not the Commissioner, conducted this review, which could take place as frequently as every 72 hours,[50] but in practice occurred only once or, at most, two or three times in each case.[51] Although in cases randomized to the PSA-DMF-System-available condition, the PSA-DMF printouts were available to judges at these reviews, few judges in fact chose to consult them.[52] One judge, the Honorable Nicholas McNamara, a member of the Dane County Criminal Justice Council Committee ("CJC"), did consult those printouts. When he did so, he focused on the FTA and NCA scales, as well as the records of the FTA and NCA events (*i.e.*, the inputs to those scales). He paid little attention to the 0-1 NVCA violence flag because he lacked knowledge of how it was calculated, and little attention to the DMF recommendation.[53]

Second, the County made various efforts to remind arrestees of their post-First-Appearance hearing dates. In particular, beginning in April of 2017, released arrestees could sign up through CCAP to receive text message reminders of their hearing dates regardless of the level of supervision assigned to them.[54] In May of 2018 (approximately a year into this study's randomization period), Dane County Pretrial initiated a mandatory text message reminder system for all arrestees assigned to its monitoring program.[55]

E. Implementation of the PSA-DMF System in Dane County

This Subpart discusses the context and specifics of Dane County's implementation of the PSA-DMF system.

---

[49] The statements in these paragraphs stem from our personal observations of, and ride-alongs during, First Appearance hearings. See note 32, supra.

[50] Wis. Stat. Ann. § 969.01(8). Actually, either party may request a modification of conditions (that is, the state can request an increase in the bail amount).

[51] Email Exchange, Matthew Stubenberg and Judge Nicholas McNamara, Feb. 25, 2020 (on file with Access to Justice Lab). Judge McNamara pointed out that Wisconsin's statutory speedy trial rights can also result in a mandate to convert a cash bond into a signature bond for arrestees who remain incarcerated beyond speedy trial deadlines (90 days after an arrestee request for a speedy trial in the case of felonies, 60 days after first appearance for a misdemeanor). Id.

[52] Matthew Stubenberg, Memo, "Meeting with Judge McNamara," Memorializing Conversation on Dec. 12, 2019 (on file with Access to Justice Lab).

[53] Id.

[54] Email Exchange, Matthew Stubenberg and Kerry Widish, Mar. 4, 2020 (on file with Access to Justice Lab).

[55] Matthew Stubenberg, Memo, "Meeting with Clerk of Court and Register in Probate," Memorializing Conversation on Dec. 12, 2019 (on file with Access to Justice Lab).

Dane County implemented the PSA-DMF System as part of a larger package of reforms to its administration of predisposition criminal justice. One impetus for these reforms was a concern among County stakeholders regarding perceived racial disparities in the administration of the County's criminal justice system, including during the predisposition period.[56] The County produced a comprehensive report in 2009 about the state of the racial disparity in the criminal justice system along with recommendations to correct the problem.[57] The County made an effort, beginning especially in 2014, to adopt evidence-based practices and systems, including a PRAI, and to incorporate an evaluation that would include measures of racial justice.[58]

Prior to this package of reforms, initiated approximately a year before this Study's randomization period, no substantial changes had occurred in County predisposition administration for at least two decades.[59] Examples of changes included in this package of reform include the following:.

- On July 1, 2016, the Court implemented an amendment to its existing Rule 206.[60] This order resulted in a reduction in the time an arrestee who could not bond out according to the Wisconsin bail schedule[61] spent incarcerated awaiting an in-custody First Appearance hearing, from an average of more than 65 hours to an average of approximately 48 hours.[62]
- Throughout the study period, Dane County has made substantial use of citation in lieu of arrest for individuals suspected of committing some misdemeanors and lower-level felonies, potentially reducing the number of matters adjudicated at in-custody First Appearance hearings, and also potentially causing the distribution of such matters to feature a higher fraction of felonies than the A2J Lab has observed in other jurisdictions.[63]

---

[56] See supra Section I.C.
[57] Celia Jackson & Laurie Mlatawou, Dane County Task Force on Racial Disparities in the Criminal Justice System, Sept. 2009. Retrieved from https://danedocs.countyofdane.com/pdf/oeo/final_report.pdf.
[58] See "A2J Lab: 'Behind the Experiment', Dane County, Part II", https://a2jlab.org/a2j-lab-behind-the-experiment-dane-county-part-ii/ (last visited Mar. 1, 2020).
[59] Matthew Stubenberg, Memo, "Meeting with Clerk of Court and Register in Probate," Memorializing Conversation on Dec. 12, 2019 (on file with Access to Justice Lab).
[60] The order is available at https://courts.countyofdane.com/prepare/CompleteRuleList#rule206 (last visited Feb. 27, 2020).
[61] Wisconsin Uniform Misdemeanor Bail Schedule, 2016, Retrieved at https://wicourts.gov/publications/fees/docs/bondsched16.pdf. (last visited Apr. 20, 2020)
[62] A copy of the report appears as Appendix D.
[63] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). During construction of the Lab's field operation, the Lab worked with Chief Deputy Clerk of Court Kerry Widish to determine the number of matters that reached First Appearance Hearing yearly. Data from calendar years 2014 and 2015 suggested that there were approximately 3000-3400 such matters per year. As discussed below, there were approximately 2200 study-eligible First Appearance adjudications in the first year of the Lab's field operation. This fact led Dane County to agree to extend the study's randomization period by six months.

- The State of Wisconsin previously had prohibited for-profit surety bail, i.e., a commercial bail bonds industry.[64]
- On June 30, 2017, about one month into this Study's randomization period, the County switched from requiring drug testing as a release condition for every defendant to requiring drug testing only if ordered by a Judge or Commissioner.[65]
- On June 30, 2017, about one month into this Study's randomization period, the County removed Pretrial case workers' discretion to assign independently GPS monitoring and Remote Breath (alcohol monitoring) as a release condition. After this change, GPS monitoring and Remote Breath could only be required by order of a Judge or Commissioner.[66]
- Around 2017, the County expanded the options available when an arrestee violated a condition of monitored predisposition release. Before 2017, a significant condition violation meant the defendant was removed from the bail monitoring program. Post 2017, after a significant violation of a bail condition, a Judge or Commissioner could make no change, increase supervision level, schedule a bail hearing, or modify bail based on the offense. This change resulted in a higher case count for Pretrial.[67]
- In March, 2018, the director of the State Courts implemented a recommendation from the Wisconsin Circuit Court Access Oversight Committee to remove from CCAP criminal cases that resulted in a dismissal or acquittal after two years. This process is referred to as "sunsetting".[68]
- In May, 2018, about a year into this Study's randomization period, Dane County Pretrial implemented text message reminders for defendants, as discussed above.[69]
- In May, 2018, about a year into this Study's randomization period, Pretrial gained the ability to conduct phone check-ins in addition to in-person check-ins.[70]

These reforms demonstrate that the County viewed the adoption of the PSA-DMF System as one of a package of reforms to its predisposition system implemented with the aspirations of reducing incarceration, reducing the intrusiveness of predisposition supervision, and increasing support for individuals during predisposition release. Although the Lab has not evaluated the effectiveness of these reforms, they provide important context for the findings discussed in Part II. They also suggest that evaluating the effect of the adoption of the PSA-DMF System might

---

[64] See Wis. Stat. Ann. § 969.12(2).
[65] Matthew Stubenberg, Memo, "Meeting with Clerk of Court and Register in Probate," Memorializing Conversation on Dec. 12, 2019 (on file with the Access to Justice Lab).
[66] Id.
[67] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).
[68] Email exchange, Manager of the Division of Policy and Practice Innovation Dane County Board of Supervisors & Matthew Stubenberg, Jul. 19, 2019 (on file with the Access to Justice Lab).
[69] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).
[70] Id.

be challenging with time series, panel data, synthetic controls, or other methods that generally depend on before-after comparisons.

Operationally, implementation of the PSA-DMF System was strong. AV contracted with Luminosity,[71] which provided a team led by Dr. Marie VanNostrand, one of the original lead scientists who developed the PSA and the concept of the DMF. The Luminosity team made several site visits to Dane County to explain the PSA-DMF System to officials there; to provide technical assistance and training for implementation; and to audit the County's operations for a short time after implementation. The Luminosity team also programmed a software tool that the County's Assessors (see below) used to enter the PSA-DMF System inputs, with the tool then producing the PSA-DMF System output. Dr. VanNostrand also facilitated the adoption of some of the reforms listed in the previous paragraph.[72]

AV provided a startup grant to Dane County to fund 1.5 FTEs worth of "Assessors," the name the County gave to the officials charged with producing the PSA risk scores and DMF recommendations and with attending First Appearance hearings to assure the System's availability and proper usage. The County later allocated funding sufficient for .5 FTE, resulting in 2 full-time Assessors. The County has also provided ongoing general purpose revenue funding to maintain the two Assessor positions as well as facility and software needs.[73] The Assessors worked within the Office of Clerk of the Circuit Court. The Assessors consulted CCAP, NCIC, and publicly available criminal justice databases from other jurisdictions (inside and outside of Wisconsin) to produce the PSA-DMF System inputs. For cases randomized to the PSA-DMF-System-present condition, the Assessors printed out at least four copies of the report for the First Appearance hearing, providing one each to the Commissioner, the ADA, and the SPD attorney.[74]

At the custodial First Appearance hearings, for arrestees randomized to the PSA-DMF-System-present condition, the Commissioner ordinarily reviewed the PSA-DMF System printout at the same time they re-reviewed the charges and the law enforcement affidavit. The Commissioner often encouraged the ADA or the SPD attorney to frame their statements by reference to the printout, or to explain why their recommendations were different from those in the printout.[75] ADA and SPD attorneys enjoyed mixed success on this score. The primary challenge for both attorneys was that the printout ordinarily was made available to them only minutes before the start of the First Appearance hearing.[76] Once the deliberation process described above

---

[71] See http://luminosity-solutions.com/ (last visited Mar. 6, 2020).
[72] We personally observed these facts during our own site visits.
[73] Email from Colleen Clark-Bernhardt to Matthew Stubenberg, May. 1, 2020 (on file with Access to Justice Lab).
[74] We personally observed these facts during our own site visits.
[75] Matthew Stubenberg, Memo, "Meeting with Pretrial Assessors," Memorializing Conversation on Dec. 13, 2019 (on file with the Access to Justice Lab). If the ADA's recommendation was not in line with what the PSA/DMF Report recommended, the Commissioner would inquire about why the ADA was deviating from the PSA/DMF recommendation.
[76] Matthew Stubenberg, Memo, "Meeting with District Attorney," Memorializing Conversation on Dec. 13, 2019 (on file with the Access to Justice Lab). As explained above, this difficulty sometimes stemmed from the availability of law enforcement affidavits, and thus charge filings, close to the start time for the First

concluded, the Commissioner ordinarily referenced the PSA-DMF System information when announcing their decision.[77] If an arrestee's matter was about to conclude without some mention of the System information, the Assessor frequently called attention to it.[78] In sum, the PSA-DMF System information played a visibly public role during First Appearance.

Commissioner Jason Hanson, who conducted the overwhelming majority of custodial First Appearance hearings during this Study's randomization period, reported that he began his review of the PSA-DMF System printout by reviewing the FTA score as well as the FTA score inputs, particularly any FTAs within the past two years. Next, he examined whether the arrestee had a pending case when arrested, something that, he stated, the ADA did not always know. Next, Commissioner Hanson examined the NCA score. He paid little attention to the 0-1 NVCA flag because he was skeptical that it differentiated arrestees well; its 0-1 distinction seemed coarse, and NVCA risk had comparatively less to do with risk of FTA at the next court hearing.[79] He also paid less attention to the DMF, and little attention at all to DMF Step 2. He was skeptical that the DMF generally, and Step 2 in particular, was evidence-based. Instead, he suspected that those recommendations were the result of a debate about community standards for which there may have been insufficient time to prepare well. For example, Commissioner Hanson was unsure of the System definition of a violent crime because it encompassed statutes that included words such as "shove" or "strike." According to him, however, the definition did not include offenses such as discharging a firearm into a building. Overall, though, Commissioner Hanson liked the PSA-DMF System, observing: "I think it has been great. I like the tool. It's the best tool I have." Specifically, Commissioner Hanson believed that the tool was objective and that the inputs used to calculate the PSA-DMF System scores included most of the same factors he personally considered important in making custodial First Appearance decisions. Perhaps of greatest note, Commissioner Hanson believed that the System improved his decision making.[80]

Appearance hearing, and the Assessors could not produce the PSA-DMF System printout without knowledge of the charges. But even when the printout was available the day before the First Appearance hearing, the County had no system through which the PSA-DMF System printouts were made available to the ADA or the SPD attorney in advance of the hearing. As a result, both attorneys effectively obtained the printouts upon arrival in the jail's courtroom. Id.

[77] Matthew Stubenberg, Memo, "Meeting with Commissioner Hanson," Memorializing Conversation on Dec. 12, 2019 (on file with the Access to Justice Lab). Commissioner Hanson said he would make note of the PSA/DMF Report on the record. If he forgot the Pretrial Assessor in the courtroom would step in and do so.

[78] Id.

[79] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

[80] Matthew Stubenberg, Memo, "Meeting with Commissioner Hanson," Memorializing Conversation on Dec. 12, 2019 (on file with the Access to Justice Lab). With respect to Commissioner Hanson's belief that the PSA-DMF System improved his decision making, the following is instructive. During the meeting on Dec. 12, 2019, Greiner asked, "This is just a hypothetical; we have no results we can share. But, what would you say if it turned out that the presence of the PSA-DMF System printout had no effect on FTA, NCA, NVCA, or predisposition incarceration time?". Commissioner Hanson replied, "I'd be astonished." He continued by expressing further surprise that such a result would be possible, pointing both to the

Of possible relevance in interpreting the quantitative results presented in Part II is that Commissioner Hanson stated that as his familiarity with the PSA-DMF System increased, he believed that his ability to guess partially what the PSA calculations would have been for PSA-DMF-System-not-present cases increased. Commissioner Hanson recognized that he could ordinarily not be sure about those guesses because the Assessors had access to greater information and greater opportunity to conduct searches than did he.[81] Nevertheless, it could be that with respect to some First Appearance hearing events, Commissioner Hanson's guesses might have been accurate.[82]

F. Field Study Description

This Subpart describes the A2J Lab's field operation. Subsection I.F.1 briefly discusses the A2J Lab's efforts to understand the Dane County criminal justice system prior to study launch. Subsection I.F.2 reviews the study's randomization protocol, reviewing the trade-offs involved in randomizing by arrest event versus by arrestee along with other technical details. Subsection I.F.3 discusses the data available for our analysis, relevant data that are currently unavailable but that may become available later, and relevant data that will likely never become available.

1. Field Study Construction

This subsection discusses the A2J Lab's efforts to construct and maintain this study's field operation. The field operation was a success, largely due to the dedication, diligence, and competence of the Dane County officials involved. These officials maintained a spirit of cooperation and an inquisitive desire to find out what worked with respect to the administration of criminal justice.

The A2J Lab sought to remain diligent before, during, and after study launch. Prior to creating the partnership with Dane County officials that led to the present study, the A2J Lab investigated the County's predisposition criminal justice system. In January of 2015, one of us conducted telephone interviews of Clerk of Circuit Court Esqueda along with Chief Deputy Widish, Deputy Jail Administrator Gordon ("Gordy") Bahler, DA Ozanne, Dorothea ("DeeDee") Watson and Jonathan E. Tradewell from SPD, Pretrial Lead Social Worker Beyler, former Commissioner Todd Meurer, and Colleen Clark-Bernhardt, then Equity and Criminal Justice Council Coordinator, now Manager of the Division of Policy and Practice Innovation for the Dane County Board of Supervisors. All of these individuals or their successors, as well as other individuals from their offices, were involved in dozens of pre-launch meetings in Madison, which also included meetings with Judge McNamara, Commissioner Hanson, and Dane County Circuit Court Judge Juan Colás.[83] As Dane County hired its first two Assessors, Rhonda Frank-Loron

---

additional and systematic criminal history information and the PSA scores as key information in his decision making.

[81] Id.

[82] Id.

[83] See "A2J Lab 'Behind the Experiment': Dane County, Part III," https://a2jlab.org/a2j-lab-behind-the-experiment-dane-county-part-iii/ (last visited Mar. 1, 2020).

and Clark Rodgers, both former federal probation and pretrial services officers, the A2J Lab maintained contact with them, observing and participating in Luminosity's training of them.[84] Similarly, the Lab met repeatedly with Research Analyst Noemi Reyes upon her hiring, working to assure clear communication regarding the nature and integrity of the data required for the study.[85] After study launch, the A2J Lab continued making period site visits and maintained weekly telephone calls with Clark-Bernhardt, Reyes, and eventually Research Associate Isabel Anadon ("Dane County Research Team").

The County implemented the PSA-DMF System in April of 2017. The County and the A2J Lab agreed that the month of April should be a trial period, with the present study kicking off officially in May of 2017. The A2J Lab's most recent site visit occurred December 12-13, 2019, three weeks before randomization closed. The Dane County Research Team transmitted an initial sample data set in August of 2018 which allowed the A2J Lab to understand part of the structure of the data. The transmission of the initial data for this analysis to the A2J Lab started in the summer of 2019, with transmissions continuing through February of 2020.

2. Randomization Protocol

This subsection discusses the study's randomization protocol.

The choice of randomization protocol for this study involved the balancing of several concerns. In discussion with AV and Dane County, the A2J Lab considered the following.

- The assignment protocol for this study could not malfunction; disruption or confusion in the First Appearance hearing could result in delays that might prolong the incarceration of arrestees destined to be released by a Commissioner's decision.
- The assignment protocol had to work nearly instantaneously, or had to be clear at the time of an individual's arrest, as it was sometimes unclear until moments before what arrestee events would "go" at a particular First Appearance hearing.
- The assignment protocol had to be robust to the fact that Dane County officials could not know which arrestees would bail out prior to First Appearance.
- As is true in any RCT, the assignment protocol needed to assure that the unit of the randomization actually received the treatment condition the randomization protocol assigned. This challenge played a substantial role in the A2J Lab's recommendations given that no one involved in the Dane County system had participated in an RCT before, and that there was no way (without undue expense) for the A2J Lab to confirm independently that units received the treatment assigned to them.
- In the A2J Lab's view, it was appropriate to presume that the assignment protocol should follow the science used to construct the PSA-DMF System.

[84] See "A2J Lab: 'Behind the Experiment': Dane County, Part I," https://a2jlab.org/a2j-lab-behind-the-experiment-dane-county-part-i/ (last visited Mar. 1, 2020).
[85] See "A2J Lab: 'Behind the Experiment': Dane County, Part II," https://a2jlab.org/a2j-lab-behind-the-experiment-dane-county-part-ii/ (last visited Mar. 1, 2020).

The certain prospect of repeat-player arrestees presented a challenge to the choice of randomization protocol for this study. Based on both common sense and the A2J Lab's review of a sample of Dane County data from calendar year 2015, it was clear that some arrestees would gain release after being randomized at one First Appearance hearing, then be rearrested and appear at a subsequent First Appearance hearing, with some repeating this cycle several times. The question that arose was whether to randomize each arrest event separately. If yes, then for an individual repeat-player arrestee, some First Appearance hearings might be in the PSA-DMF-System-present condition with other such hearings in the System-not-present condition. If not, then for an individual repeat-player arrestee, all First Appearance hearings would be in the same randomization condition, meaning that the unit of randomization and analysis was the person (*i.e.*, the arrestee).

After extensive discussions with multiple offices within Arnold Ventures and with Dane County officials, the A2J Lab recommended the former scheme described above, with randomization assignment following whether the case number[86] was even (PSA-DMF-System-present) or odd (no-PSA-DMF-System-present). The A2J Lab's recommendation was based on the following considerations. First, from AV reports and from a confidential document reviewed by the Lab, the A2J Lab understood that the dataset compiled to create the PSA analyzed First Appearance hearings (or their equivalents) separately, not arrestees separately. Thus, the A2J Lab's recommended protocol followed the science that gave rise to the PSA-DMF System.

Second, the case number was assigned as soon as charges were filed by a clerk working in Circuit Court Clerk Esqueda's office. Once assigned, the case number was visible to the Commissioner, the Assessor, the ADA, the SPD attorney, the court clerk, and to anyone else present at the First Appearance hearing. Indeed, the court clerk ordinarily announced the case number(s) at the beginning of each arrestee's hearing. The publicly visible and obvious nature of a randomization protocol based on case number effectively deputized all Dane County officials present at the hearing to enforce the randomization protocol. If an even-numbered case was called and there was no PSA-DMF System printout, then efforts could be made to remedy the situation, including by altering the order in which cases were considered (which would not affect when arrestees receiving signature bonds were released) to allow location or generation of the printout. Meanwhile, no Dane County official with whom the A2J Lab spoke could conceive of a situation in which the case number could be manipulated or chosen intentionally.

This protocol also had the advantage that outcomes were well-defined, and well-defined in a manner that followed the research leading to the PSA-DMF System's construction. N(V)CA and FTA could be defined as events occurring during the predisposition period of each First Appearance hearing matter, and could be conceptualized as either 0-1 variables or as counts.

---

[86] Specifically, all case numbers in Dane County ended with a six-digit number, which determined the randomization condition. Note: A small number of First Appearance events had multiple case numbers. The A2J Lab and Dane County officials agreed that in those instances, randomization turned on the odd/even status of the smallest six-digit case number.

It was possible but less conceptually transparent how to do so under a scheme that used arrestees as the unit of randomization.

This protocol included some drawbacks. For some, it was counterintuitive that a particular arrestee could have cases in different randomization conditions simultaneously. Further, repeat-player matters randomized to different treatment conditions challenged an assumption fundamental to causal inference[87] from the RCT's data in that a Commissioner familiar with a repeat-player arrestee's criminal history from previous custodial First Appearance hearings might be able to guess what the PSA-DMF System would have disclosed in a matter randomized to the no-PSA-DMF-System-present condition.

The A2J Lab's recommendation to randomize in Dane County based on First Appearance hearing, not by arrestee, was also informed by the fact that AV funding has allowed the A2J Lab to pursue multiple RCTs in other jurisdictions. The A2J Lab recommended that in other jurisdictions, particularly those in which both the PSA-DMF System and the randomization protocol would be implemented by software, treatment assignment turn on the odd/even status of an arrestee identification number, thus making the arrestee (not the First Appearance hearing) the unit of randomization. And in fact, A2J Lab has fielded three RCTs in another state using this design.

Operationally, in Dane County, the randomization protocol was simple, and we have high confidence that treatment received was faithful to treatment assigned in all but a handful of cases. The software that the Assessors used to calculate the PSA and DMF, and to produce the resulting printouts, blocked the Assessors from printing in odd-numbered cases. And as just mentioned, all Dane County officials at custodial First Appearance hearings participated in assuring that a printout was present in even-numbered cases. Dane County officials uniformly reported that implementation of the study was smooth in this regard.

3. Data And Data Processing

This subsection discusses the data generation, data processing, and resulting analysis dataset for this study. It proceeds in three subsections. Subsection I.F.3.a discusses the data available for this study. Subsection I.F.3.b discusses relevant data that the A2J Lab hopes to obtain at some point. Subsection I.F.4 discusses relevant data that is likely never to become available for this study.

a. Data Available

This subsection discusses the data available for this study. It proceeds in three subsections. Subsection I.F.3.a.i discusses the databases used, subsection I.F.3.a.ii discusses the

---

[87] This is called the "Stable Unit Treatment Value Assumption," or "SUTVA." For a several-paragraph discussion of challenges to SUTVA and the implications for this study, see D. James Greiner & Heidi Liu, An RCT To Assess the Effectiveness of the PSA in Dane County, WI (Sep. 13, 2015, on file with the Access to Justice Lab).

information included in those databases, and subsection I.F.3.a.iii discusses the resulting analysis dataset.

i. Data Sources

This subsection describes the sources of the data comprising the analysis dataset. The data used in the analysis originated from four primary sources:

1) The CCAP database which is a statewide court information case management system ("CMS")
2) the PROTECT database used by the Dane County District Attorney's Office ("DA"),
3) the Spillman database used by the Dane County Sheriff's Office, and
4) the database used by Pretrial Assessors to input information and process the PSA-DMF System.

The A2J Lab did not have direct access to any of these data sources and relied on the Dane County Research Team to write queries to match/combine data from the five sources into the files used for analysis.

Regarding the specifics of each dataset: The DA used a CMS called PROTECT.[88] The PROTECT CMS was custom-developed for statewide use by Wisconsin district attorneys. The Dane County Research Team only had access to the subsystem controlled by the Dane County DA. PROTECT housed NCA data, including on arrests that did not result in prosecutions (*i.e.*, NCA events that did not appear in CCAP). As discussed below, the A2J Lab is exploring an effort to obtain NCA data from the PROTECT subsystems controlled by other counties' district attorneys' offices.

As discussed above, Wisconsin courts, including Dane County, used CCAP. CCAP had basic case information data as well as FTA data. The Dane County Sheriff's Office used a CMS called Spillman, one used by numerous sheriff's offices around the country. The Dane County Sheriff's Office ran the Dane County Jail, and the Spillman CMS held information related to each arrestee's incarceration and release.

The Assessors had their own database which contained information related to the creation of the PSA-DMF System reports. When the PSA-DMF System was implemented in Dane County, a contractor working with the Luminosity implementation team and funded by AV programmed an internal web portal for the Assessors to calculate the scores of the PSA-DMF System. The portal captured all the data entered into it as well as the scores returned; it also housed information on whether an arrestee's matter reached a First Appearance hearing.

[88] See District Attorney IT Program, available at http://dait.state.wi.us/section.asp?linkid=11&locid=13 (last visited Mar. 2, 2020).

ii. Data Provided

This subsection describes in greater detail the data received from the five sources identified in the previous subsection.

To produce data tractable for the type of analysis undertaken here, the A2J Lab had to combine information from multiple sources. For instance, to calculate the amount of time an arrestee spent incarcerated during their pretrial period required joining information on separate booking events held in the Spillman database with case disposition and First Appearance hearing dates found in both levels of the CCAP database[89]. The combined CCAP data provided a valid date range, which was used to filter valid booking instances. Thus, time spent incarcerated for any arrestee outside of the valid pretrial date range would not be counted toward that case's total pretrial incarceration time. This type of calculation, based on joining and matching data from the five separate datasets, was necessary for all the primary outcome variables: FTA, N(V)CA, and total predisposition incarceration. The A2J Lab constructed a relational database from the data provided by the Dane Research Team to aid in the calculation and storage of these data. The structure of this database and the specific information obtained from the main sources of data are described below.

1) PSA/DMF: This table contained primary PSA-DMF-related measurements and case information, including relevant case dates (First Appearance hearing, PSA-DMF System report date), demographic information on arrestees, PSA-DMF System input and output values (including raw scores and recommendations), and treatment group assignment.

2) Charges: This table contained charge-level information for each case, including the charge offense level (*e.g.*, misdemeanor/felony), and the charge disposition date, which were used to calculate the valid end date for each case's pretrial period. A single case can have multiple charges associated with it, and by extension, multiple charge disposition dates. A case's pretrial period was calculated as the last charge disposition date associated with the case.

3) Courts: This table contained the outcome information for each case's First Appearance hearing. Relevant data in this table included reported bail amounts, hearing dates, the judge's release decision, reasons why no release decision occurred, and any release conditions. Cases with valid entries for a missing release decision were filtered out of the analysis dataset as they were not study-eligible.

[89] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). The Dane County Research Team provided statewide CCAP data only for defendants who had a PSA generated in Dane County.

4) NCA: This table contained information on any new criminal activity within the PROTECT or CCAP data. Information included NCA dates and violence flags. These data were used to compute both NCA and NVCA counts.

5) FTA: This table contained information on any failure to appear events, including FTA dates and types.

6) Booking: This table contained information relating to specific booking events and included Dane County Jail entry dates, release dates, and release reasons.

The Dane Research Team linked and combined the PSA data prior to sending it over to the A2J Lab. This linking of multiple datasets proved formidable as there was no unique identifier to link all the datasets. The linking required matching on multiple soft data points such as name and date of birth.[90]

iii. Analysis Dataset Description

This subsection briefly describes the dataset assembled and analyzed for this report. Further details are included in Part II to describe the A2J Lab's findings.

The full analysis dataset was compiled by joining the separate relational tables described in Subsection I.F.3.a.ii into a single, flat dataset where each PSA-DMF System assessment for a single case was considered a single unit. All data were joined on a universal "Internal Assessment ID" variable that generated via the PSA system software and available in all data sources. The final dataset consisted of 2262 rows (*i.e.*, separate cases) and 45 columns (*i.e.*, separate observed variables). Because much of the data in the relational database existed on levels apart from that of a single court case (*e.g.*, charge data, N(V)CA/FTA data, booking data), merging the separate sources required some sort of aggregation function to derive single variable-case observations. For instance, FTA and N(V)CA data were calculated using a count function where each unique case contributed to an overall count of relevant instances. Total predisposition incarceration was the sum of non-overlapping booking duration bounded by First Appearance date and case disposition date. Other variables, such as felony and misdemeanor charges, are binary indicators for whether the charge type was present in the charge table for the relevant case. The broad categories of variables are briefly described below.

- Outcome variables: NCA count, NVCA count, FTA count, total number of days of predisposition incarceration, and time until first failure.
- First Appearance variables: bail amounts, release decision type, and binary indicators for each potential condition that could be placed on release.
- Case variables: disposition date, felony charge present, misdemeanor charge present, and First Appearance hearing date

---

[90] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

- Initial release variables: whether the arrestee exited the Jail after the underlying arrest, release date, and release reason.
- Demographic variables: gender, race, ethnicity, and date of birth
- PSA-DMF System Variables: raw input data used to compute the outputs of the PSA-DMF System, NCA score, FTA score, NVCA flag, and the recommendation.
- Additional calculated variables: FTA and N(V)CA counts were computed as binary indicators; date of birth was translated into an age at First Appearance variable; and separate ordinal release decision was calculated to distinguish among an arrestee ordered released with a signature bond, cash bail less than or equal to $1000, and cash bail greater than $1000.

b. Data Still Sought

This subsection describes relevant data that the A2J Lab hopes to obtain at some point in the future to include in the Final Report.

The primary source of data the A2J Lab continues to seek consists of NCA information arising from incidents that occurred outside of Dane County. The A2J Lab was limited to PROTECT data from the subsystem that the Dane County DA; statewide PROTECT data were not available. The A2J Lab is exploring whether it can acquire N(V)CA data on arrestees randomized in the study from subsystems controlled by other county DAs. The A2J Lab planned to send one of its researchers to Dane County in the spring to pursue this effort, which requires programming on computers located in the Dane County DA's office. This effort has been delayed because of the Covid-19 travel restrictions.

The PROTECT data were the only source of N(V)CA outcomes when an arrestee was arrested, but the DA declined to prosecute. The PROTECT system exists statewide, and each DA's office in Wisconsin had access to the information from every other DA's office. However, each DA's office agreed to access another's data only for case-related activities. Consequently, the Dane County DA granted the Dane County Research Team access to Dane County's PROTECT data but not to the PROTECT data from other counties. There are 72 counties in Wisconsin, and it would have been unduly time-consuming to obtain consent from each DA in each county. However, the A2J Lab has discovered a possible method to search the study participants against the PROTECT CMS to identify which counties have NCAs that were not prosecuted. If successful, this method would allow for the Dane County DA and the A2J Lab to approach the (hopefully) small number of counties with a nontrivial number of hits to request access to the PROTECT data for just the relevant arrestees.

c. Data Unlikely To Become Available

This subsection discusses data relevant to this study that are unlikely to become available in the foreseeable future.

First, data on arrestees' criminal justice involvement outside the state of Wisconsin will likely not become available to the A2J Lab.[91] To the A2J Lab's knowledge, the only source of such data is the FBI's NCIC system, which is rarely available to external researchers.

Second, data on arrestees' N(V)CA arrests that did not result in charges in Wisconsin counties (other than those covered by the PROTECT data acquisition efforts described in the previous subsection) will likely not become available. Obtaining permission from each of the other 71 elected DAs in Wisconsin to search their respective PROTECT subsystems appears at present to constitute a daunting task.

Third, data records expunged or removed from CCAP via Wisconsin legislative sunsetting provisions before we obtain them from the County will likely not become available. The Dane Research Team was required by compliance agreements to transmit only data as allowed under the law.[92] Future extracts of data may be missing data from the current dataset corresponding to events that have been expungement or sunsetted between data pulls. While there may have been some cases that were removed from CCAP via expungement or sunsetting before they were included in the dataset provided the number of cases removed statewide is relatively low so the impact on the dataset is likely low.[93]

Wisconsin allows for the potential expungement of criminal records under certain conditions.[94] The requirements for expunging one's criminal records make expungement restrictive in practice. The requirements vary, but generally, the individual must have been convicted of a particular crime, have been younger than 25 years of age during the commission of the crime, the crime carried not more than a possible six year term of imprisonment, and have requested (and the judge approved) an expungement at the time of sentencing.[95] If a case was expunged before the CCAP data were pulled, it would not appear in our dataset. Once a case is expunged, it is unlikely the A2J Lab would be able to obtain information about that case in subsequent data pulls.

In November 2017, the Wisconsin Circuit Court Access Oversight Committee recommended implementing a policy whereby criminal cases that resulted in a dismissal or acquittal would be

---

[91] Through its work on other RCTs and so-called "validation" studies of the PSA-DMF System, the A2J Lab has obtained criminal history data from other states. One such project concerns a state adjacent to Wisconsin, and thus might include data relevant to the present study. The A2J Lab will seek permission from this jurisdiction to use its data for this study.
[92] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).
[93] Preliminary analysis suggests approximately 1000 cases removed from CCAP statewide through expungement or sunsetting. Email exchange, Noemi Reyes, Research Analyst with the Dane County Criminal Justice Council at the Office of the Dane County Board of Supervisors & Matthew Stubenberg, Mar. 9, 2020 (on file with the Access to Justice Lab).
[94] See Special disposition, Wis. Stat. Ann. § 973.015, https://docs.legis.wisconsin.gov/statutes/statutes/973/015, Last Visited (Mar. 9, 2020)
[95] Id.

removed from CCAP after two years.[96] In March 2018, the Director of the State Courts implemented this recommendation.[97] A criminal misdemeanor or felony case for which all charges were dismissed or resulted in acquittal would no longer be available publicly on CCAP two years after the entry of a final order. This process is automatic and commonly referred to as "sunsetting."

Some cases might have been affected by the sunsetting provision before the pull of CCAP data was completed. The Dane Research Team's access to the CCAP data was the same as that available to the public, except the Dane Research Team received the data in bulk. The CCAP data were pulled over a period of time around August and September, 2019. Cases subject to sunsetting between the start of the study in June 2017 and the pull date were removed by law before they could be captured.

The CCAP data source proved to have issues with data integrity in regards to case disposition and sentencing information. The analysis in this report was calculated only on cases that have a disposition. Approximately 240 cases from the dataset provided were missing case disposition information in CCAP and were therefore not included in this interim analysis. A case lacking a disposition could indicate the case was still active. However, upon deeper investigation some of the cases had indeed been closed with a valid disposition. In these cases with a valid disposition there was a particular issue involving the accuracy or lack of sentencing information. The approximately 240 cases have been excluded from this report as the Dane Research Team and the A2J Lab continue to investigate these cases.[98]

Each of the three data sources identified above likely has some NCA information; the first and second may have some NVCA information. Whether any of these sources has FTA information depends on whether one defines FTA as limited to a failure to appear for a hearing connected to the randomized First Appearance hearing event or more broadly as a failure to appear for any post-randomization case in any jurisdiction. As discussed below, this report uses the former definition, and under that definition, none of the four data sources discussed in this subsection has relevant FTA data.

Finally, we will likely not obtain data from jails, prisons, or incarceration facilities other than the Dane County Jail. It is possible that arrestees in our dataset spent time incarcerated in facilities other than the Dane County Jail, particularly if they have N(V)CA reported in other jurisdictions. Whether such information is relevant depends on study goals. If study goals are limited to the impact of predisposition incarceration on Dane resources, data from outside Dane are

---

[96] The report recommends the Director of State Courts should create a display period on WCCA of two years or less for dismissed felony cases, including deferred prosecutions, and felony cases that resulted in acquittal. See WISCONSIN CIRCUIT COURT ACCESS OVERSIGHT COMMITTEE Report, https://www.wicourts.gov/courts/committees/docs/wccafinalreport2017.pdf (last visited Mar. 9, 2020)
[97] Email exchange, Manager of the Division of Policy and Practice Innovation Dane County Board of Supervisors & Matthew Stubenberg, Jul. 19, 2019 (on file with the Access to Justice Lab).
[98] See Email from Colleen Clark-Bernhardt to Matthew Stubenberg, May. 4, 2020 (on file).

irrelevant. If they include the impact of predisposition incarceration on the arrestee, data from outside Dane County are relevant.

II. Findings

This Part provides the A2J Lab's findings from this study. It proceeds in two subparts. Subpart II.A provides the A2J Lab's qualitative findings, principally related to the strength of the field operation. Subpart II.B provides the A2J Lab's quantitative findings, focusing principally on the primary outcome variables analyzed.

A. Qualitative Findings

This Subpart discusses the A2J Lab's qualitative findings. Because most of the information appears in Part I, we simply restate the most salient points here.

The implementation of the PSA-DMF System was strong. The Luminosity implementation team was extraordinarily knowledgeable and competent. All relevant stakeholders, including SPD attorneys, ADAs, the Clerk of Court, Pretrial, the Commissioners, several judges, County staff, and others received Luminosity's training on the System. Adoption of the PSA-DMF System was the subject of extensive discussion within the County, including the public, a discussion largely mediated through the CJC. Dane County implemented the PSA-DMF System as one of a larger package of reforms to its predisposition process that began before and continued after System adoption. County officials working on the System's implementation were competent, diligent, active, and inquisitive.

The RCT field operation was strong. From the outset, County officials approached the adoption of the PSA-DMF System with an attitude of wanting to know what worked, as opposed to assuming that they already knew what worked. They were active participants in the discussions and the decision making needed for the field operation. They were responsive to the A2J Lab's requests for information. They understood the need for, and took steps to assure, fidelity to the randomization scheme. Although data production was slower than all would have liked,[99] when produced, datasets were high quality.

Lest the A2J Lab be accused of sycophantic reporting, we note that our experiences with the implementation of the PSA-DMF System and with attempts to constitute RCT field operations in other counties were dissimilar in many ways to our experience in Dane County.

---

[99] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). As discussed above, the data production process was iterative resulting in the slower than anticipated data production schedule.

B. Quantitative Findings

In this Subpart, we provide our quantitative findings, proceeding in three sections. Section I.B.1 provides a simplified explanation of the statistical techniques we deployed. Section I.B.2, the bulk of this report, describes our findings on covariate balance as well as the results of our comparisons of the PSA-DMF-System-present versus PSA-DMF-System-not-present groups on First Appearance decisions, FTA, N(V)CA, demographic fairness[100] measures, and predisposition incarceration. Section I.B.3 discusses analyses that we did not include in this report but that we anticipate performing in the future, likely in our final report based on the complete dataset.

1. Overview of Basic Analysis Techniques

This section describes the analysis techniques used in the next section.

Our primary analysis technique is a permutation test, perhaps the simplest causal inference technique available. An intuitive explanation of this technique complete, with illustrations and an example of the testing of two different types of alpaca shampoo, is available online.[101] By way of brief explanation as applied to the present setting: recall that First Appearance decisions were essentially[102] randomly assigned to one of two conditions: PSA-DMF-System-present and PSA-DMF-System-not-present. For each decision, we recorded the value of multiple outcome variables. For purposes of illustration, suppose the outcome of interest is the number of predisposition days the arrestee spent incarcerated in the Dane County Jail. We begin by subtracting the average number of predisposition incarceration days in the PSA-DMF-System-present group from the same average for the PSA-DMF-System-not-present group. Note that, as discussed below, there were approximately 1131 First Appearance hearing events in each of the two groups.

We then ask as follows: The assignment of which 1131 First Appearance hearing events were in the PSA-DMF-System-present group (and which 1131 were in the other group) was random. What if the randomization had assigned a different set of 1131 First Appearance hearing events to the System-present group (and thus, a different set of 1131 to the System-not-present group)? What would the value have then been of the difference between the average number of predisposition incarceration days in the System-present group and the same average for the PSA-DMF-System-not-present one? It is easy enough to compile a hypothetical dataset with a different set of 1131 assigned to the System-present group (and the other, different, 1131

[100] Demographic fairness, which will be discussed in much greater detail in section 2.c.iv, broadly refers to whether the experience of members of different groups within a specific demographic category (i.e. Race or Gender) differs from other groups within the same demographic category based on whether the PSA-DMFsystem is present or not present.

[101] Jared Wilbur, The Permutation Test: A Visual Explanation of Statistical Testing, available at https://www.jwilber.me/permutationtest/ (Mar. 2019) (last visited Mar. 4, 2020).

[102] For the remainder of this report, we assume that the device of assigning First Appearance hearings to treatment condition was random. See Section I.F.2, above for why we believe that this assumption is reasonable.

assigned to the no-System-present group): we tell the computer to ignore temporarily the actual assignment of hearing events to treatment groups and to assign randomly, anew, 1131 to each group.

We would then like to calculate the value of the difference in the average of incarceration days for the newly assigned PSA-DMF-System-present group and the newly assigned PSA-DMF-System-not-present group under this hypothetical randomization. But we have a problem. Some of the First Appearance hearing events that were actually and originally assigned to the System-present group have now, in this new and hypothetical randomization, been moved to the PSA-DMF-System-not-present group. Indeed, mixing observations into different groups this way was the whole point of pretending to do the randomization over again. But for a hearing event that was actually, originally in the System-present group, how do we know how many days of incarceration we would have observed had that event, counterfactually, been in the System-not-present group? After all, we can only observe one state of the world for each First Appearance hearing event, and for that particular event, we observed it in the System-present condition. So how do we fill in what that First Appearance hearing event's number of days of predisposition incarceration would have been had it, counterfactually, been assigned to the System-not-present condition? The short answer is that we cannot definitively do so. We can only do so by making an assumption. For the present, we assume that the number of days of predisposition incarceration for that hearing event was the same in the System-present and the System-not-present conditions. In other words, we assume at first that the presence of the PSA-DMF System has no effect on predisposition incarceration days. This is called a "null hypothesis." This "null hypothesis" allows us to "know," to fill in, the counterfactual number of days of predisposition incarceration for that unit.

We use the same null hypothesis to fill in the missing values for First Appearance hearing events actually assigned to the PSA-DMF-System-not-present condition. In other words, for each hearing event actually assigned to the PSA-DMF-System-not-present condition, we need to know what its number of days of predisposition incarceration would have been had it been assigned to the PSA-DMF-System-present condition. We use the null hypothesis of no treatment effect to fill in the same value that we actually observed for this hearing event, that is, its value under the PSA-DMF-System-present condition.

We make the same assumption for all units under this hypothetical, new randomization. And we calculate a value for the difference in the average number of days incarcerated in the System-present versus the System-not-present groups. We store that number somewhere.

We repeat the process with a new, different hypothetical random allocation of 1131 First Appearance hearing events to each treatment condition. By that, we mean that we repeat this process thousands of times. Each of these hypothetical random allocations is called a "permutation," and the name of this technique is a "permutation test." Each time we permute, we record the value of the difference in the average number of incarceration days, System-present group minus System-not-present group.

Why did we do all of this? We did so because it allows us to end up with tens of thousands of hypothetical values for the difference in average incarceration days based on randomization. And we can compare those tens of thousands of hypothetical values of treatment effects, defined here as the difference in the average number of days of incarceration across our two groups, to the actual, observed value of the difference in average incarceration days. That allows us to see how extreme, if at all, our observed value was.

More precisely, we ask as follows. Recall that all of tens of thousands of hypothetical values were generated under an assumption, called the null hypothesis, that there was no difference in any of the First Appearance hearing events between the number of days incarcerated when the PSA-DMF System was present versus when it was not present. So we ask: assuming that hypothesis were true, how many of the hypothetical values for the average difference between the groups were as big, or bigger, than the one that we actually observed in our real randomization? If only a very small number of those hypothetical values were as big as the value under our actual randomization, then either our actual randomization gave us a pretty strange result (meaning we just had bad luck) or that something else was amiss. If the fraction of those hypothetical values that were as big or bigger than our actual value was really small, say, less than .05 of them were as big or bigger, we might conclude that something else really was amiss. What could that something else be? We infer that the culprit is the null hypothesis, the assumption that there was no difference in the number predisposition incarceration days, for any of the First Appearance hearing events, when the PSA-DMF System was present versus when it was not present. We likely should reject that assumption. We did not know it then, but we think we know it now, meaning that we have statistical evidence for rejecting the null assumption.

By way of defining vocabulary, the fraction of the hypothetical values that were as big or bigger than our actual value is called the "p-value." And the p-value is a measure of "statistical significance." It tells us how likely it is, under the null hypothesis assumption of no treatment effect for any observation, that we would observe what we did observe purely due to chance. Thus, we consider low p-values to be evidence of a treatment effect. The definitions of "permutation," "permutation test," and "null hypothesis," are provided two paragraphs above.

With respect to our permutation testing, there are a few additional details. First, each time we permuted the data, we calculated values under the null hypothesis for all of our outcome variables of interest. As discussed below, there were lots of these, although only some were of primary importance, with the others exploratory or supplemental, helping us understand what we saw with respect to the primary outcomes. Second, we used the permutation technique to calculate 95% confidence intervals for the possible effect sizes using a series of null hypotheses of constant additive treatment effects. Roughly speaking,[103] a 95% confidence interval is a range of numbers that is 95% likely to include the true treatment effect. To repeat, our intervals were produced under the assumptions that the treatment effect for each First Appearance observation was the same (constant) and that it took the form of adding some number as

[103] Key additional assumptions are that (i) the treatment effect is constant and additive, and (ii) that the null hypotheses that produced the interval are true.

opposed to, say, multiplying by some number (additive). These assumptions tend to produce wider intervals than might some other statistical techniques, meaning they tend to suggest less knowledge about where the true effect lies.

Third, in the example above, we defined the causal effect of interest as the difference between the two average number of days of predisposition incarceration, PSA-DMF-System-present group versus PSA-DMF-System-not--present group. But we could have chosen a different definition of causal effect, perhaps the difference in the medians, or some measure of how variable the number of days of incarceration was. And in some cases below, we did choose other measures of causal effects, such as ratios, or medians.

The advantage of the permutation test technique is that it requires fewer assumptions than do other statistical techniques. The disadvantage is that if the variable being analyzed has a non-standard distribution, the permutation test can sometimes fail to detect treatment effects that other statistical techniques might detect. For example, one primary outcome variable with an obviously non-standard distribution is the example used above (*i.e.*, the number of days of predisposition incarceration). The distribution of this outcome variable is described further below, but briefly, it has a lot of observations with 0 days incarcerated, a lot of observations with small numbers, and a lot of observations spread out over values in the tens, twentys, thirties, etc. up through the hundreds. This distribution looks nothing like a bell curve, and nothing like any other standard statistical distribution for which permutation testing is appropriate on its own. We deployed our permutation testing with respect to this model, but we also fit a mathematical statistical model, using something called a “zero-inflation” technique to attempt to account for the large number of zeros. We report both sets of results below.

2. Analyses Completed

This section constitutes the primary results of interest for this report. It provides our quantitative findings and proceeds in three subsections. Subsection II.B.2.a addresses covariate balance (*i.e.*, whether the randomization produced treatment groups that appear to look like one another with respect to background variables). Subsection II.B.2.b addresses whether the presence of the PSA-DMF System affected First Appearance hearing decisions. Subsections II.B.2.c.i through II.B.2.c.iv discuss whether the presence of the PSA-DMF System affected FTA, N(V)CA, measures of demographic fairness, and number of days of predisposition incarceration.

a. Balance of Observed Covariates

This subsection addresses whether, to the extent we can tell, the randomization did what it was supposed to do, which was to create two groups of First Appearance hearing events that looked alike with respect to background variables such as race, gender, age, charges, and other values determined before, or otherwise unaffected by, the presence vel non of the PSA-DMF System printout.

The following figures report distributions for, and the significance of any differences between, PSA/DMF system presence/non-presence for the following variables: race, gender, age, charges, and PSA-DMF System outputs (NCA score, FTA score, NVCA flag). If the randomization scheme was successful, each of these variables should show no significant difference between the PSA-DMF-System-present group and the PSA-DMF-System-not--present group because each of these measurements was determined prior to either providing or not providing the PSA-DMF printout to the judge.

Figure 1: Distribution of Race Overall And by Treatment Condition

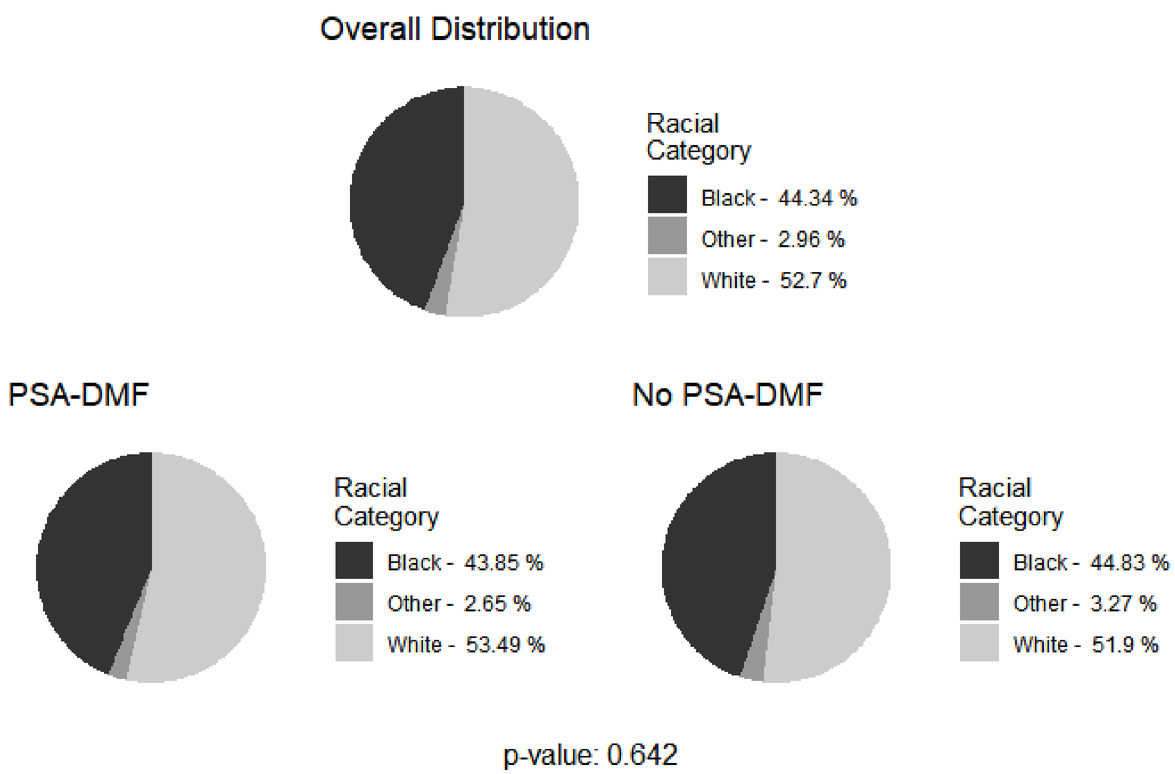

Figure 1: This Figure displays pie charts showing the distribution of race categories, both in the overall data sample as well as by randomization group. Of the three racial categories represented in the data, the large majority of respondents are either Black or White, and thus race based analysis are restricted to a comparison between these two groups. White arrestees represent a slightly larger percentage of the overall population, as well as larger percentages of each randomization group, compared to Black arrestees. Comparing the distribution of race across randomization groups allows for the analysis of potential differences between the groups. The reported p-value of 0.642 indicates we can reject the potential that the randomization groups differ from one another with respect to racial distributions. Relevant Ns for each randomization group are as follows: PSA-DMF: Black- 496, Other- 30, White- 605; No PSA-DMF: Black- 507, Other- 37, White- 587.

Figure 1 reports the distribution of race categories within the data, both for the overall dataset as well as for each of the randomization groups. Because the overall number of individuals falling into the “Other” category is a small proportion of the overall study group, the analysis focuses on comparing the two largest categories: black and white. For the PSA-DMF-system-present group, the percentages of black and white arrestees were 43.85% and 53.49%, respectively; for the PSA-DMF-system-not-present group, the relevant values were 44.83% and 51.90%, resulting in differences between the randomization groups of about 1% and 1.5%. These differences are substantively small and are not statistically significant, meaning technically that we have no evidence to contradict a hypothesis that there is no difference between the two groups. This is what we would expect if the randomization were effective.

Figure 2: Gender Distribution

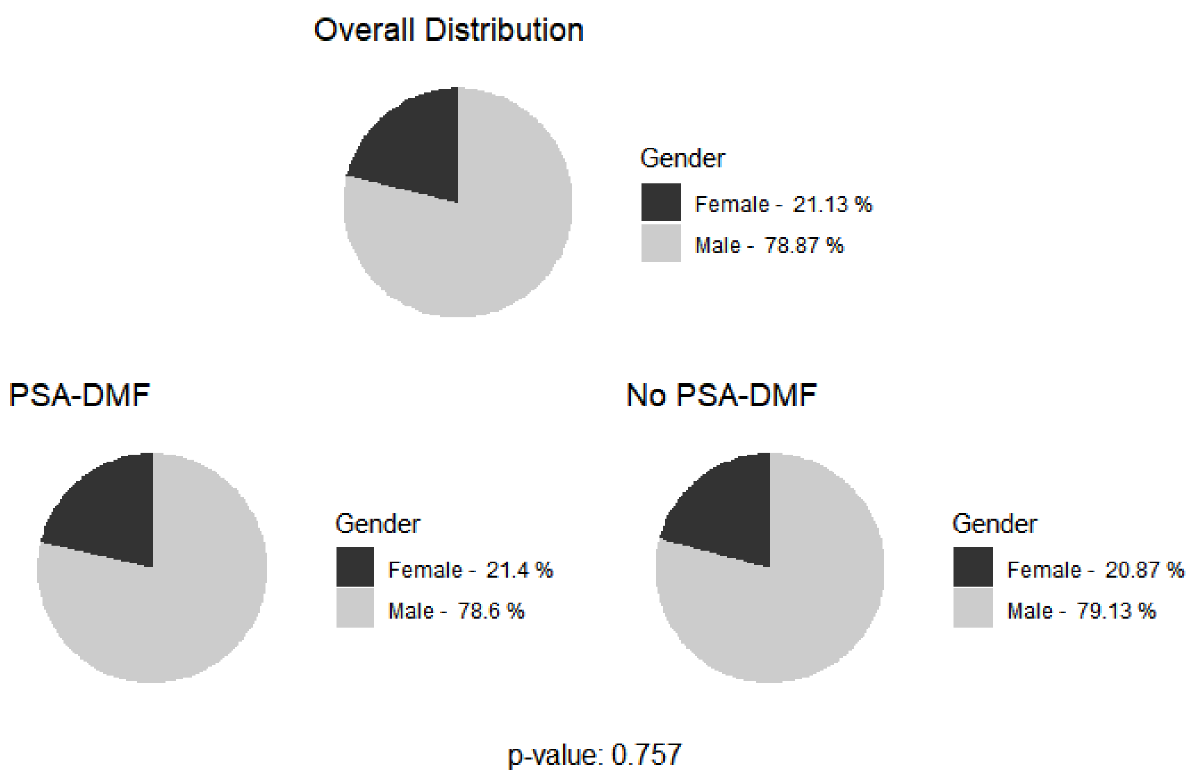

Figure 2: This Figure displays pie charts showing the distribution of gender categories, both in the overall data sample as well as by randomization group. Of the two gender categories represented in the data, the large majority of respondents are male, representing roughly 4 times the number of arrestees as female arrestees. Comparing the distribution of gender across randomization groups allows for the analysis of potential differences between the groups. The reported p-value of 0.757 indicates we can reject the potential that the randomization groups differ from one another with respect to gender distributions. Relevant Ns for each randomization group are as follows: PSA- DMF: Female- 242, Male- 889; No PSA-DMF: Female- 236, Male-895.

Figure 2 reports the distribution of gender in a manner similar to the race distribution report above. Female arrestees represent only about one-fifth of the study group (21.13%), and male arrestees represent about four-fifths (78.87%). This fact will be relevant in the section discussing fairness measurements since direct comparisons of the PSA-DMF System's impact on gender will be impacted by the differences in size of the two groups. However, in terms of the effectiveness of the randomization scheme, the gender distributions of the two treatment groups is again what we would expect if the randomization scheme were effective. The difference between each randomization group for each gender is about 0.6%, small substantively and not statistically significant.

Figure 3: Distribution of Age Overall And by Treatment Group

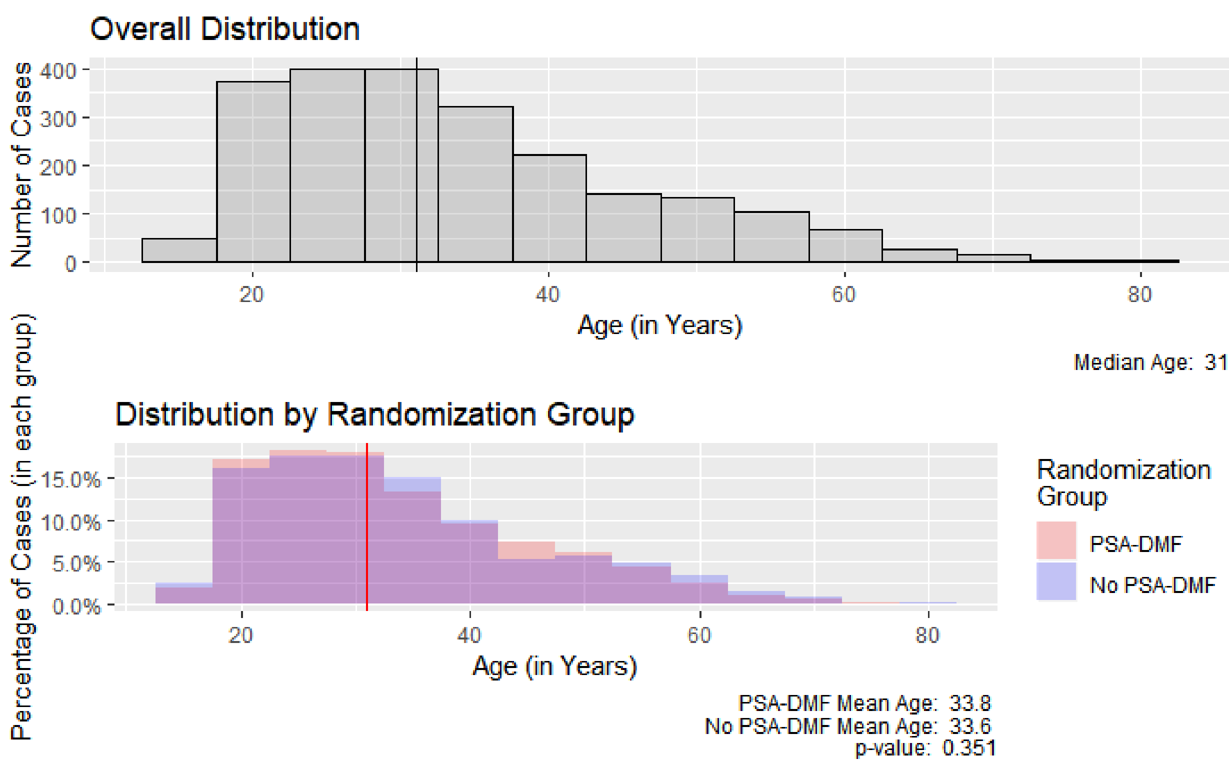

Figure 3: These graphs show histograms of the reported age of arrestees at First Appearance. The top histogram reports the overall distribution of age and indicates a left skew that signifies a large portion of younger arrestees in their 20s, which is confirmed by the median overall age of 31. The bottom graph overlays histograms of each randomization group. The large area of blended color indicates a substantial degree of overlap of the distributions of age in each group. This is confirmed by the similarity in mean ages (33.8 for the PSA-DMF group, 33.6 for the no PSA-DMF group) as well as by the insignificant p-value of 0.351 on the permutation test of randomization group differences in mean age.

Figure 3 reports the overall distribution of age at First Appearance, as well as the distribution of age at First Appearance by randomization group. Overall, the data skew toward the left side of the graph, indicating that a large proportion of arrestees were in their 20s, with a median age of 31. The distribution by randomization group shows the same trend of arrestees skewing younger. The mean age within each randomization group differs by 2.5 months, and the median for each group is the same as the overall median of 31. The difference in mean ages fails to achieve either substantive or statistical significance. Again, this distribution is what we would expect under a properly functioning randomization procedure.

Figure 4: Distribution of Highest Charge

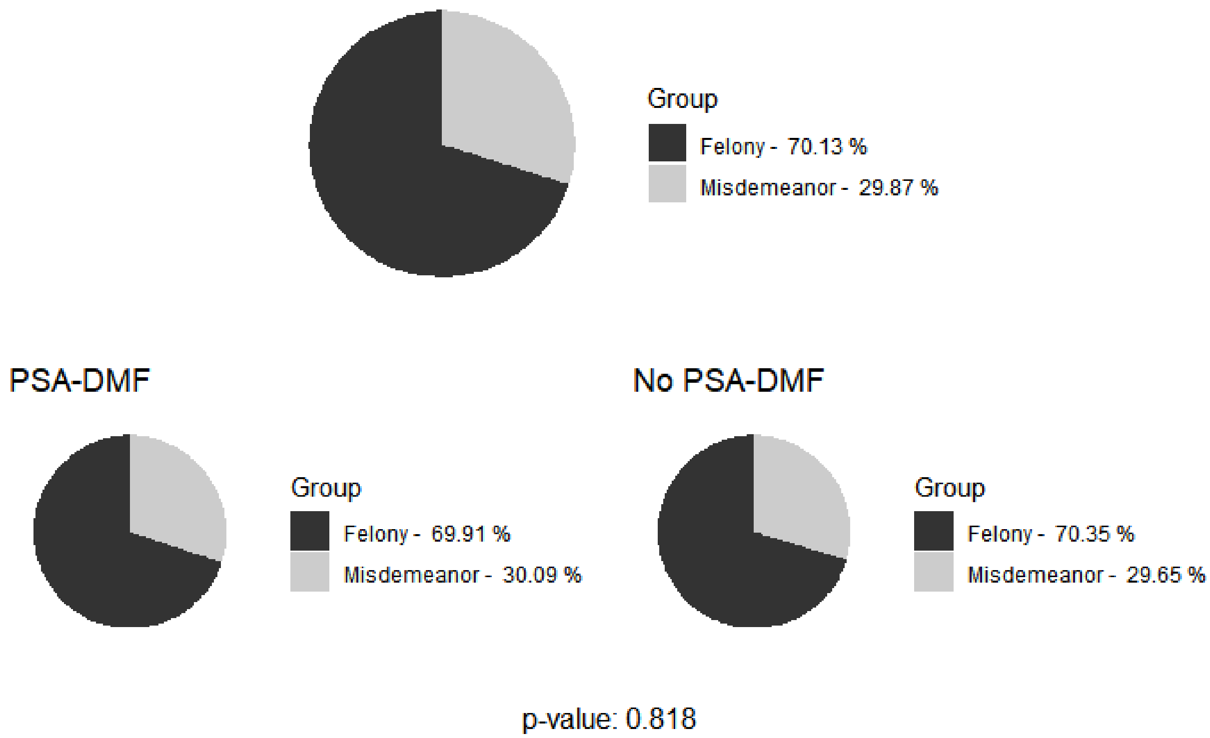

Figure 4: This Figure displays pie charts showing the distribution of highest charge per case categories, both in the overall data sample as well as by randomization group. Of the two charge categories represented in the data, the large majority of first appearance events are marked by a highest charge level of Felony, representing over 2 times the number of first appearance events with a highest charge level of Misdemeanor. Comparing the distribution of charge level across randomization groups allows for the analysis of potential differences between the groups. The reported p-value of 0.818 indicates we can reject the potential that the randomization groups differ from one another with respect to highest charge level distributions.

Figure 4 reports the distribution of highest charge level (misdemeanor or felony), both overall and for each of the randomization groups. The PSA-DMF-system-present group had a felony charge rate of 69.91%, and 30.09% of cases carried a misdemeanor as the highest charge. The corresponding numbers for the System-not-present group were 70.35% and 29.65%. The differences between each group are about half a percentage point. As with the findings for the prior variables, the differences between randomization groups are very small and fail to achieve statistical significance.

Figure 5: Distribution of PSA-DMF System Outputs

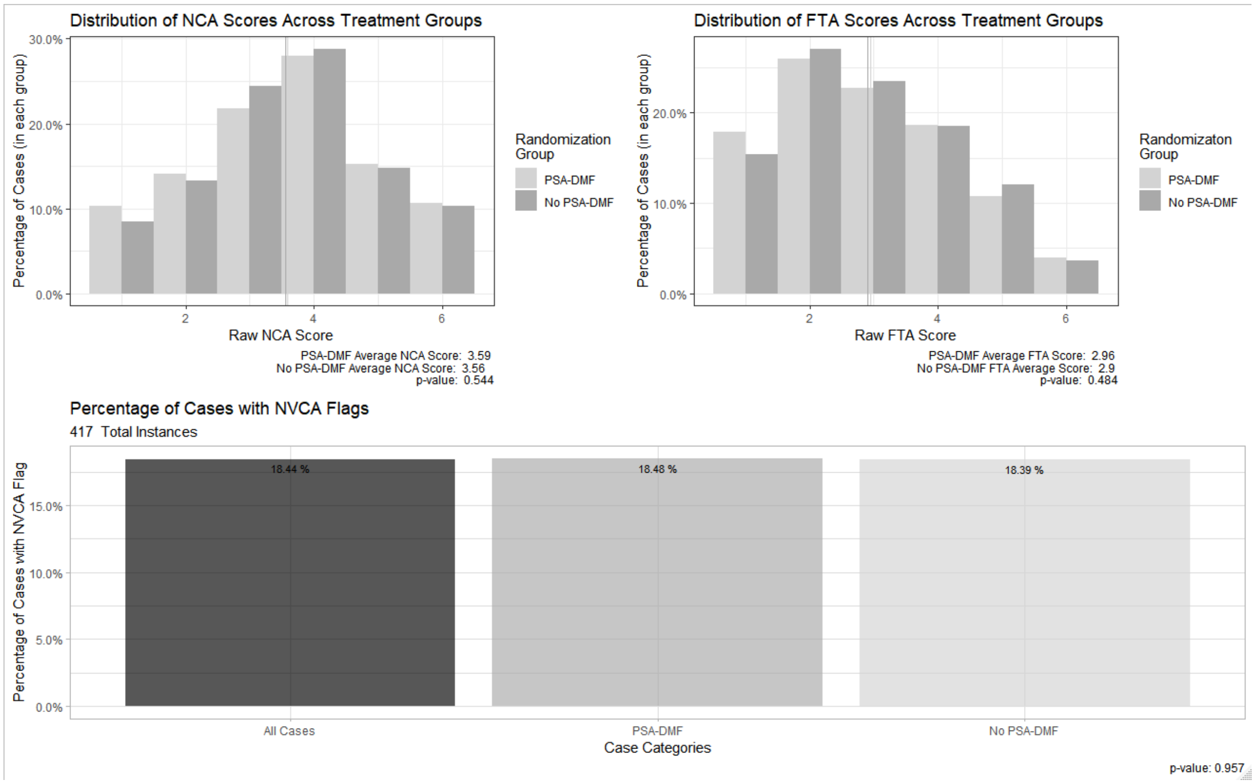

Figure 5: These bar graphs show the distribution of PSA-DMF System Outputs across the randomization groups for the three output types: NCA Score, FTA Score, and NVCA Flag. Because PSA-DMF System Outputs are calculated prior to treatment, comparing the distribution of scores across randomization groups allows for the evaluation of whether the distributions of arrestees is comparable across randomization groups with respect to relevant risk factors captured by the PSA-DMF System. The similarity between average score values for the randomization groups for both NCA and FTA (3.59; 3.56 and 2.96; 2.9, respectively) as well as the similar percentage of NVCA Flags in each group indicate that the population of arrestees in each randomization group are similar with respect to relevant risk factors. This is further demonstrated by the insignificant p-values on permutation tests for randomization group differences of 0.544, 0.484, and 0.957 for NCA Scores, FTA Scores, and NVCA Flags.

Figure 5 reports the distribution of PSA-DMF System outputs (FTA/NCA scores and NVCA flag) for each randomization group. FTA and NCA scores are each reported on a 1-6 scale, so we can detect any differences in the distributions by comparing the mean values for each randomization group. The PSA-DMF-System-present group mean FTA and NCA scores were 2.96 and 3.59, respectively, and the corresponding System-not-present group's means were 2.90 and 3.56. The distribution of the NVCA flag can be compared using the difference in the fraction of cases with a flag in each randomization group, which were 0.184 for the System-present group and 0.185 for the System-not-present. The resulting differences were 0.06, 0.03, and 0.004 for FTA, NCA, and NVCA, respectively. As with the prior variables, these differences are neither substantively nor statistically significant.

Given these findings, we can reasonably conclude that to the extent observable the randomization groups were similar to each other, meaning that the randomization procedure successfully generated balance on observable covariates.

b. Effect of the PSA-DMF System on the Commissioner's First Appearance Decisions

This subsection addresses the effect of the PSA-DMF System on the Commissioner's First Appearance decision.

We analyze the impact of the PSA-DMF System on First Appearance decisions in two ways. The first considers the probability of obtaining a signature bond and the average bail amount as separate outcome variables. The second combines these metrics into an ordinal scale consisting of three groups: arrestees assigned a signature bond, arrestees assigned cash bail of at most $1000, and arrestees assigned cash bail of more than $1000. The $1000 threshold was chosen because it evenly divided the cash bail group. Using these delineators, 1583 arrestees were assigned a signature bond, 403 arrestees were assigned cash bail of at most $1000, and 276 arrestees were assigned cash bail of more than $1000.

Figure 6: First Appearance Outcomes Across Study Groups

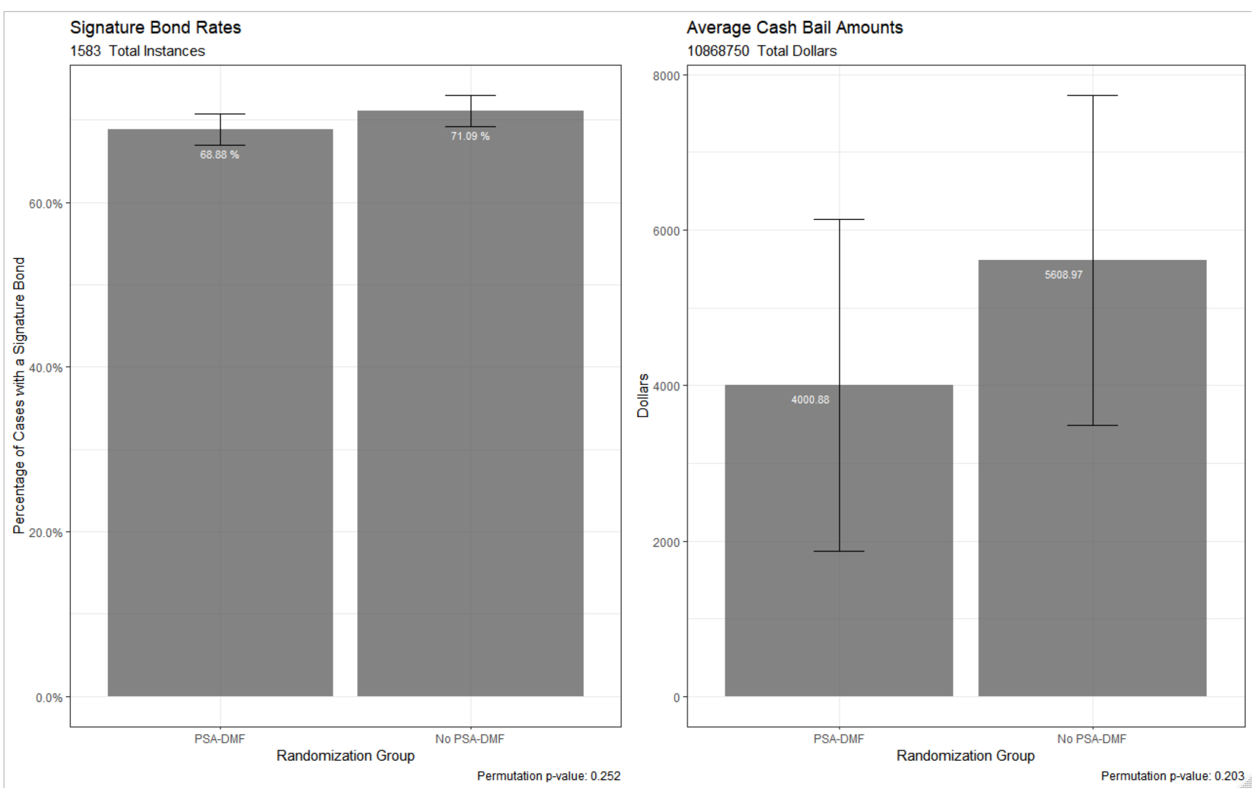

Figure 6: These bar graphs show the average value first appearance outcomes for each randomization group. The left bar graph shows the difference in percentages of cases receiving a surety bond at first appearance for the randomization groups while the right bar graph shows the average cash bail amount determined at first appearance. Signature bonds are coded as a

bond amount of $0, which is what the large majority of first appearance events observe. Despite this, there is substantively large variation in cash bail amounts determined at first appearance. The error lines on the bars indicate the confidence intervals of each randomization group's relevant statistic. The fact that the error bars overlap each indicates that there is no significant difference between randomization groups. The reported insignificant p-values of 0.252 and 0.203 for signature bond rates and average bond amounts, respectively, also confirmed that these metrics do not provide support of a PSA-DMF System treatment effect on overall values of these first appearance outcomes. The use of average bond amounts is sensitive to the presence of outliers in the bond data. This average figure does not mean, for example, that a case selected at random would be expected to have a bond amount at or around $4000, for the PSA-DMF condition, or $5600 for the No PSA-DMF condition. Thus, for example, each randomization group had a median value of $0, indicating that at least 50% of arrestees in each condition were designated for release (with respect to that case) without the requirement of depositing money with the court. The 75th percentile for the PSA-DMF condition was $500, while the same measure was $300 for the no PSA-DMF condition, indicating that at least 75% of cases, in the PSA-DMF group, paid $500 or less to meet release conditions ($300 in the no PSA-DMF group).

Figure 6 reports the mean differences in signature bond rates and cash bail amounts by each randomization group. Because the signature bond variable is a binary indicator for whether a signature bond was assigned, the mean of the variable is the fraction of cases receiving a signature bond. Overall, approximately .7 of cases received a signature bond, .689 of cases in the PSA-DMF-System-present group and .711 of cases in the System-not-present group. The difference between the randomization groups is about .022, a substantively small difference that is not statistically significant.

The average cash bail amount for the overall study population was $4804.93, with a minimum and a median of $0 (recall that signature bonds are coded as $0 in this variable) and a maximum of $1,000,000.[104] The average bail amounts for each randomization group were $4000.88 and $5608.97 for the System-present and System-not-present groups, a difference of roughly $1600. This difference might seem large, but because of the high variation in the bail amounts, it is not statistically significant (p-value of 0.203). Thus, when evaluating signature bond rates and average cash bail amounts separately, the PSA-DMF System does not appear to cause a statistically significant effect.

The high variation on the cash bail amounts is worth a deeper analysis. The System-present and System-not-present groups had cash bail amount standard deviations of 42767 and 59787, respectively, with a p-value derived from an F test on the variances significant at the 0.001 level.

---

[104] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab). The cash bail numbers include defendants who may have been held for additional reasons like other criminal charges or a violation of probation. In this situation a defendant may have chosen not to contest the cash bail amount because they know they will still be held on the other charge. Therefore, the cash bail numbers may be higher than if every defendant was held on just a single case.

The F test conducted an analysis on the ratio of the variances of the two randomization groups, which produced a test statistic of 1.954, indicating that the variation of the System-not present cash bail amount was roughly twice that of the System-present group. Substantively, the presence of the PSA-DMF System caused an approximately 50% decrease in the variation on cash bail amounts. The F-test 95% confidence interval was (1.74 and 2.12), suggesting that the PSA-DMF system decreased variation on cash bail amounts by between 43% and 53%. Practically speaking, this finding provides evidence that the PSA-DMF System caused greater uniformity in bond amounts, meaning that arrestees in the PSA-DMF-System-present group saw more similar cash bail levels compared to arrestees in the System-not-present group.

Despite the lack of significant treatment effect on average cash bail amounts, there are reasons to suspect that a simple, unidirectional, overall effect of the PSA-DMF System on first appearance outcomes would be difficult to detect. For instance, low risk factor scores/recommendations may have a positive effect on signature bonds (or a negative effect on average bond amounts), but high risk factor scores/recommendations may have a negative effect on signature bonds (or a positive effect on average bond amounts), which would cause each effect to diminish the overall treatment effect when pooled with the other. We examine the issue more closely.

The following three figures are heatmaps of the judge's decision (based on the previously described ordinal scale) across levels of the PSA-DMF System recommendation scale. To clarify, the PSA itself does not make recommendations, rather the DMF system translates PSA scores into recommendations based on the input and decisions of actors within the local area the PSA-DMF System has been implemented. Thus the following findings do not speak to the PSA itself, but to the specific PSA-DMF System in Dane county.

Figure 7: First Appearance Outcomes Across NCA Score Categories

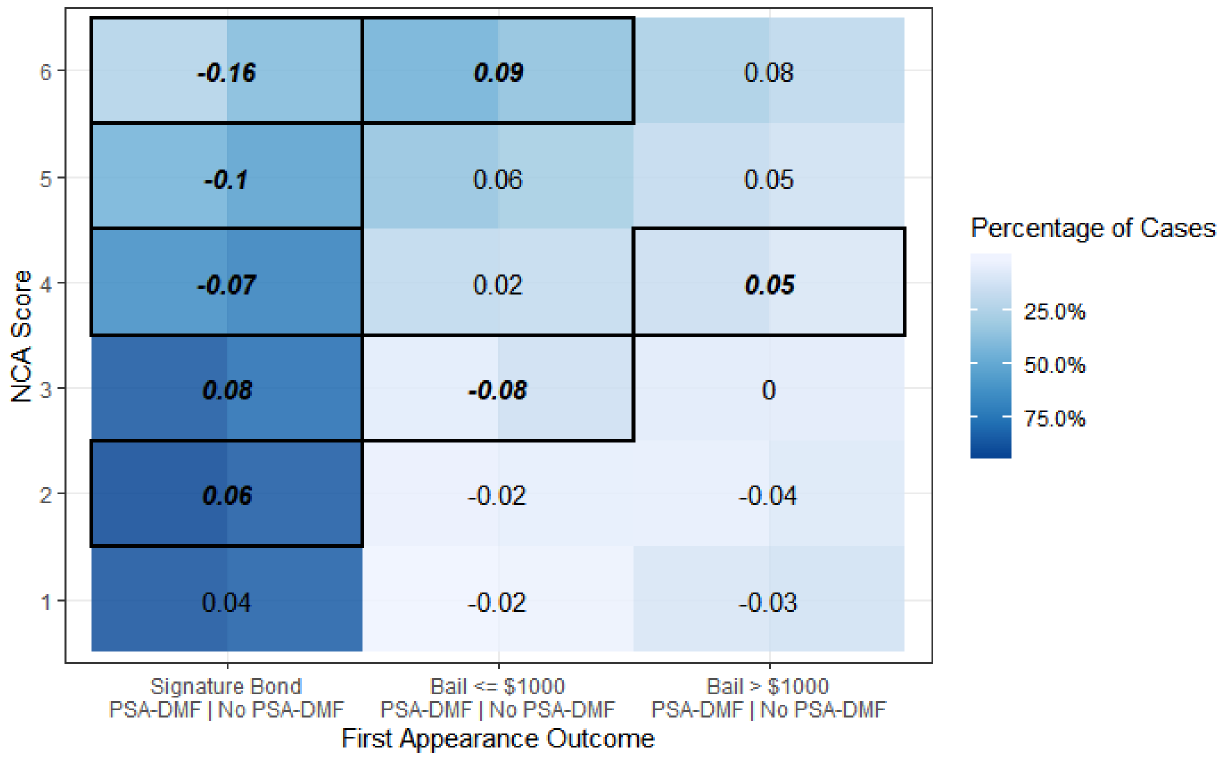

Figure 7: This heatmap shows the distribution of NCA scores across levels of first appearance outcomes as well as between randomization groups. Darker colors represent a higher percentage concentration of cases for a specific NCA score (cell colors are normalized by row), while differences in colors within the cells indicate the magnitude of the difference between randomization groups for the respective NCA score and first appearance ordinal outcome for associated with the cell. The differences in proportions of cases in each cell by randomization group are reported by the text within the cells. Bolded and italicized text indicate differences significant at the p<.010 level, while cells with borders indicate differences significant at the p<0.05 level. Overall this heatmap indicates the existence of statistically significant impacts of NCA Scores reported by the PSA-DMF System on first appearance outcomes, specifically for NCA Scores of 2, 3, 4, 5, and 6 with respect to the probability of a signature bond; scores of 2 and 3 are associated with increases in the proportion of cases receiving a signature bond relative to either level of cash bail, while scores of 4, 5, and 6 are associated with lower proportions of cases receiving signature bonds when the PSA-DMF System is present. NCA scores of 3 are also associated with lower proportions of cases receiving cash bail amounts of $1000 or less when the PSA-DMF System is present while scores of 4 and 6 are associated with higher proportions of cases receiving cash bail amounts of more than $1000 and $1000 or less, respectively, when the PSA-DMF System is present.

Figure 7 shows both the distribution of cases (using the shading key) and the differences between the treatment groups for each NCA score and release order type (using the numbers within each cell). Darker colored boxes represent a higher concentration of cases. In general, the pattern matches what we would expect: lower risk factor score cases are associated with less stringent release conditions at First Appearance, and the opposite for higher risk score cases. Zooming in, each cell is divided into two subcells, with the percentage System-present cases on the left, and the percentage of System-not-present cases on the right. Each cell also displays the difference between the proportion of cases in each cell by treatment group. The figure shows that cases with an NCA score of 4, 5, or 6 were less likely to be assigned a signature bond when in the System-present group, differences that were statistically significant. Cases with an NCA score of 2 were more likely to be given signature bonds. Additionally, cases with an NCA score of 3 were less likely to be given cash bond amounts of at most $1000 when in the PSA/DMF system present group, and such cases with NCA scores of 6 or 4 were more likely to be given cash bond amounts of at most $1000 and at least $1000, respectively, when compared to similar cases in the no PSA/DMF system present group.

Based on Figure 7, the presence of the PSA-DMF System affected the judge's release condition decision making with respect to NCA scores. This effect is most visible in the signature bond figures, with higher scores less likely to receive a signature bond and lower scores more likely. Even in cells without statistically significant effects, differences exist in the same direction.

Figure 8: First Appearance Outcome by Presence of NVCA Flag

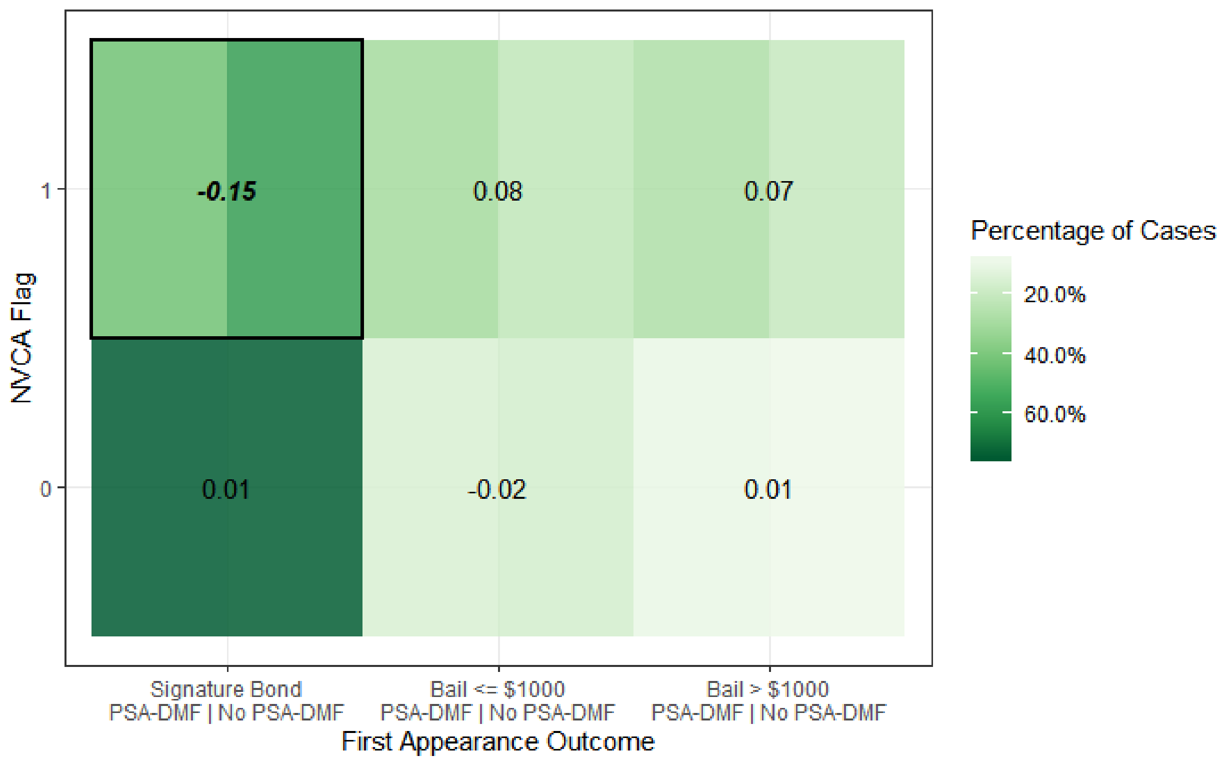

Figure 8: This heatmap shows the distribution of the presence of an NVCA Flag across levels of first appearance outcomes as well as between randomization groups. Darker colors represent a

higher percentage concentration of cases for a specific value on the NVCA Flag indicator (cell colors are normalized by row), while differences in colors within the cells indicate the magnitude of the difference between randomization groups for the respective NVCA Flag indicator and first appearance ordinal outcome for associated with the cell. The differences in proportions of cases in each cell by randomization group are reported by the text within the cells. Bolded and italicized text indicate differences significant at the p<.010 level, while cells with borders indicate differences significant at the p<0.05 level. Overall this heatmap indicates the existence of significant impacts of the presence of an NVCA Flag reported by the PSA-DMF System on first appearance outcomes, specifically when an NVCA Flag is present the proportion of cases receiving a signature bond are lower for the PSA-DMF system present group.

Figure 8 shows both the distribution of cases and the differences between each randomization group by the presence of the NVCA flag. As before, darker boxes represent a higher concentration of cases. In general, the pattern is what we would expect. Cases without the NVCA flag present are more associated with signature bonds. Cases with the flag present (weakly) show the opposite pattern. With respect to difference across treatment conditions, 15% fewer cases in the PSA-DMF-System-present group were given a signature bond when an NVCA flag was present compared to the PSA-DMF-System-not-present group. This finding also supports the conclusion that the PSA-DMF System had observable effects on the judge's decision making.

We note, however, that given Commissioner Hanson's statement to the effect that he paid little attention to the NVCA flag, as detailed in Subpart I.E, it may be that the effects described above are due to the availability of some other aspect of the PSA-DMF System printout which may increase the odds that the person skips bail, perhaps the NCA scores themselves. For example, a defendant who had a NVCA flag likely had a lengthy criminal history. The significance of the extended criminal history presented on the PSA-DMF-System report may have led the judge to grant fewer signature bonds compared to the PSA-DMF-System-not-present group where the criminal history of the defendant was presented in a less clear fashion.[105]

[105] Dane County CJC Research Team, Memo, "Comments/edits on the Interim Report "Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI," Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

Figure 9: First Appearance Outcomes Across FTA Score Categories

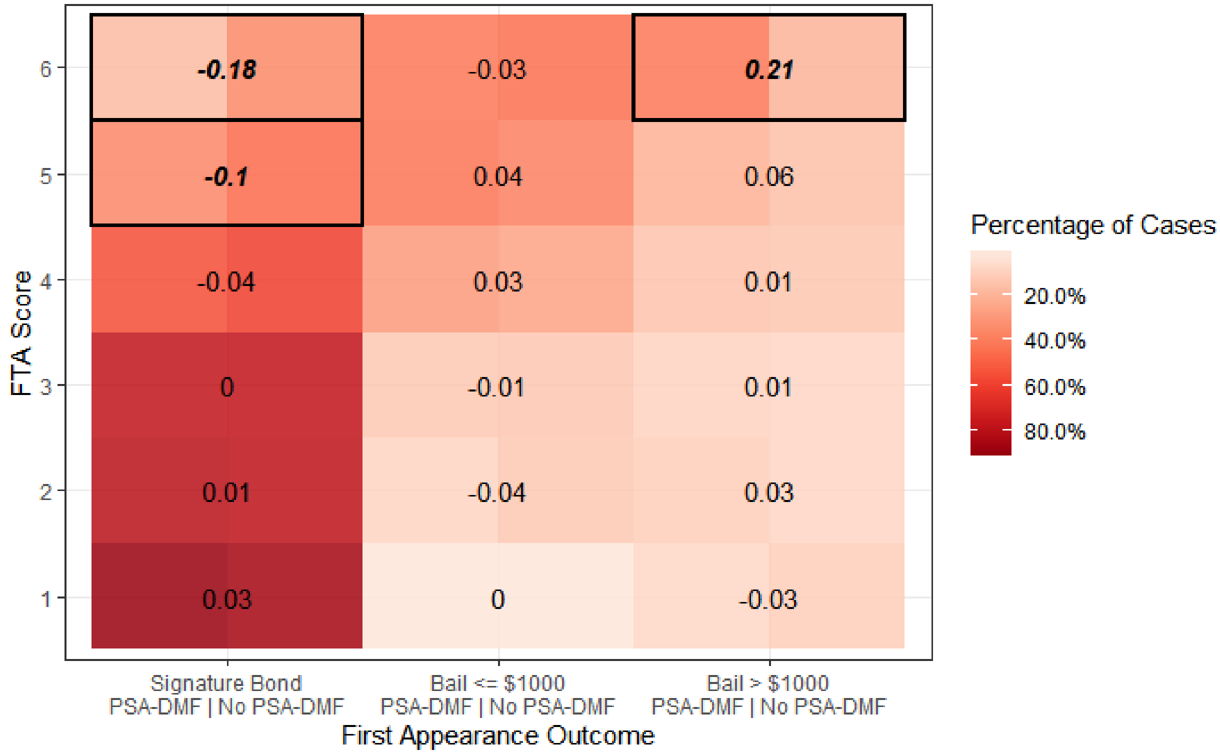

Figure 9: This heatmap shows the distribution of FTA scores across levels of first appearance outcomes as well as between randomization groups. Darker colors represent a higher percentage concentration of cases for a specific FTA score (cell colors are normalized by row), while differences in colors within the cells indicate the magnitude of the difference between randomization groups for the respective FTA score and first appearance ordinal outcome associated with the cell. The differences in proportions of cases in each cell by randomization group are reported by the text within the cells. Bolded and italicized text indicate differences significant at the p<.010 level, while cells with borders indicate differences significant at the p<0.05 level. Overall this heatmap indicates the existence of significant impacts for FTA Scores reported by the PSA-DMF System on first appearance outcomes, specifically for FTA Scores of 5 and 6, which are associated significant decreases in the proportion of cases receiving a signature bond when the PSA-DMF System is present, as well as, in the case of FTA Scores of 6, with increases in the proportion of cases assigned a cash bail amount of $1000 or higher.

Figure 9 shows the same analysis, this time for each FTA score. As before, darker boxes represent a higher concentration of cases. The same pattern emerges: lower FTA scores are associated with lighter release conditions and higher FTA scores with more stringent ones. Significant effects by PSA-DMF System status are observable but only for higher FTA scores. Cases in the PSA-DMF-system-present condition with FTA scores of 5 or 6 were less likely to be assigned signature bonds, and such cases with FTA scores of 6 were more likely to be assigned cash bail greater than $1000. The same conclusion--that the PSA-DMF System had observable effects on the Commissioner’s decision making--emerged from this analysis.

There are two primary reasons the effects above were not detectable in the initial aggregate distributions (*i.e.*, those not differentiated by risk factor scale). First, the impacts of NCA scores for the System-present group were in opposing directions between higher and lower scores, which may have washed out overall. Second, scores of 5 or 6 on both the NCA and FTA scales, as well as cases with NVCA flags present, are much less common than other categories.

c. Effect of the PSA-DMF System on Four Primary Outcome Variables

This subsection describes our findings of the effect of the PSA-DMF System on four primary outcome variables: FTA, N(V)CA, number of days of predisposition incarceration, and measures of demographic fairness. Our overall conclusion is that we find no evidence of a causal effect of the effect of the PSA-DMF System on any primary outcome variable, but that we currently lack statistical power to detect what we consider policy-relevant differences.

A word on the structure of variables such as FTA, NCA, and NVCA: there are various ways to conceptualize these outcomes. One way is as a so-called "binary" variable. In other words, FTA considered as a binary outcome variable has the value 0 if there was no FTA event (defined more fully in the next subsection) and 1 if there was one or more FTAs. This way of understanding FTA focuses on how many arrestees had perfect attendance records, something that might have affected how the criminal justice system treated them. A second way of conceptualizing FTA is as a count; this FTA measure takes on the number of FTAs (as defined below) observed for the proceedings corresponding to a particular First Appearance hearing event. This way of conceptualizing FTA could be considered to focus on the total costs imposed on the criminal justice system and on society by FTA events. For FTA "non-zero count," we consider only the First Appearance hearing events with at least one FTA. This way of conceptualizing FTA focuses on understanding the behavior of bad actors. These explanations all apply equally to NCA and NVCA.

With respect to the number of predisposition incarceration days, we chose to focus on the number of days incarcerated, as opposed to whether the arrestee was ordered incarcerated at the First Appearance hearing or whether the arrestee was actually incarcerated, because an arrestee could be incarcerated, then released, then reincarcerated, then released, and so on during a single predisposition period. The number of days incarcerated predisposition accommodates this reality while providing valuable information about the costs imposed on the arrestee and the County.

The following graph provides an overall summary of the results of our permutation testing for FTA, N(V)CA, and number of days of incarceration. It provides the p-values resulting from permutation tests for the difference in means[106] for the outcomes listed on the y-axis. For a difference to be close to statistically significant, the p-value would need to be near or below .05. None of these p-values are close to .05. We provide the results of other statistical techniques, and further explanation, below. This graph previews and partially demonstrates, however, that

[106] See Section II.B.1 for an explanation of these terms.

we have no statistical evidence of any effect of the presence of the PSA-DMF System on FTA, N(V)CA, or the number of days of predisposition incarceration.

Figure 10: Permutation Test p-Values of Three Primary Outcome Variables

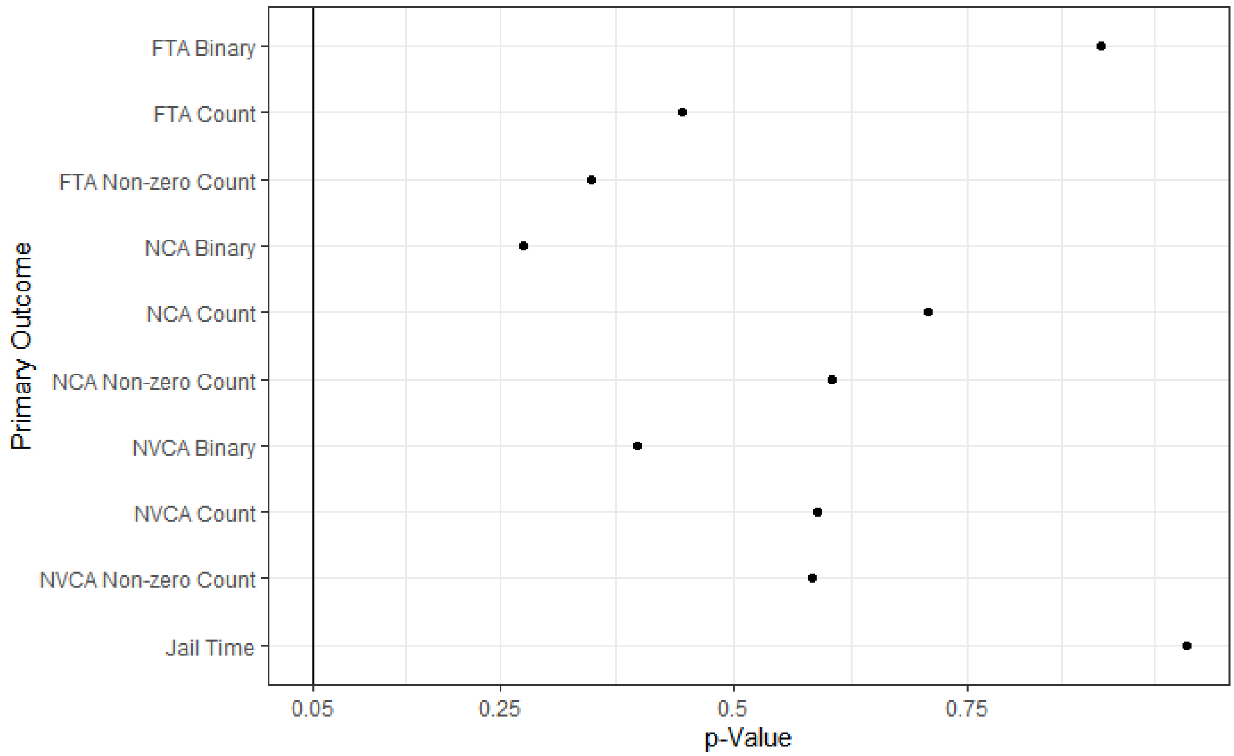

Figure 10: This figure shows p-values resulting from permutation tests of differences in means for FTA, NCA, NVCA, and number of days of predisposition incarceration, with the first three outcomes calculated in different ways. "Binary" refers to whether or not there was any instance (there may have been more than one) of the variable in question. "Count" refers to how many instances of the variable in question occurred. "Non-zero Count" means that we disregarded any First Appearance hearing event that had no instance of the variable in question, focusing exclusively on those hearing events that had at least one instance. "Jail time" refers to the number of days of predisposition incarceration, not any form of post-conviction sentence. None of the p-values are close to or below .05. We have no statistical evidence of a causal effect the PSA-DMF System on these primary outcomes.

Figure 10, above, demonstrates a lack of statistically significant differences from permutation testing for FTA, NCA, and NVCA, each conceptualized in three ways, as well as the number of days incarcerated in the analysis dataset as a whole. This finding continues to hold true when we consider breakdowns by race and gender, with one exception that we do not credit: the NVCA comparison for women.

At present, because we have data from one year, with a lower-than-anticipated 1131 observations in each treatment condition group, we lack the statistical power to detect what we consider policy-relevant outcomes. As demonstrated in Figures 11 and 12, the width of the 95% confidence intervals derived from our permutation tests, assuming constant additive treatment effects, demonstrate this fact.

Figure 11: Count Outcome Treatment Effect Sizes with Confidence Intervals

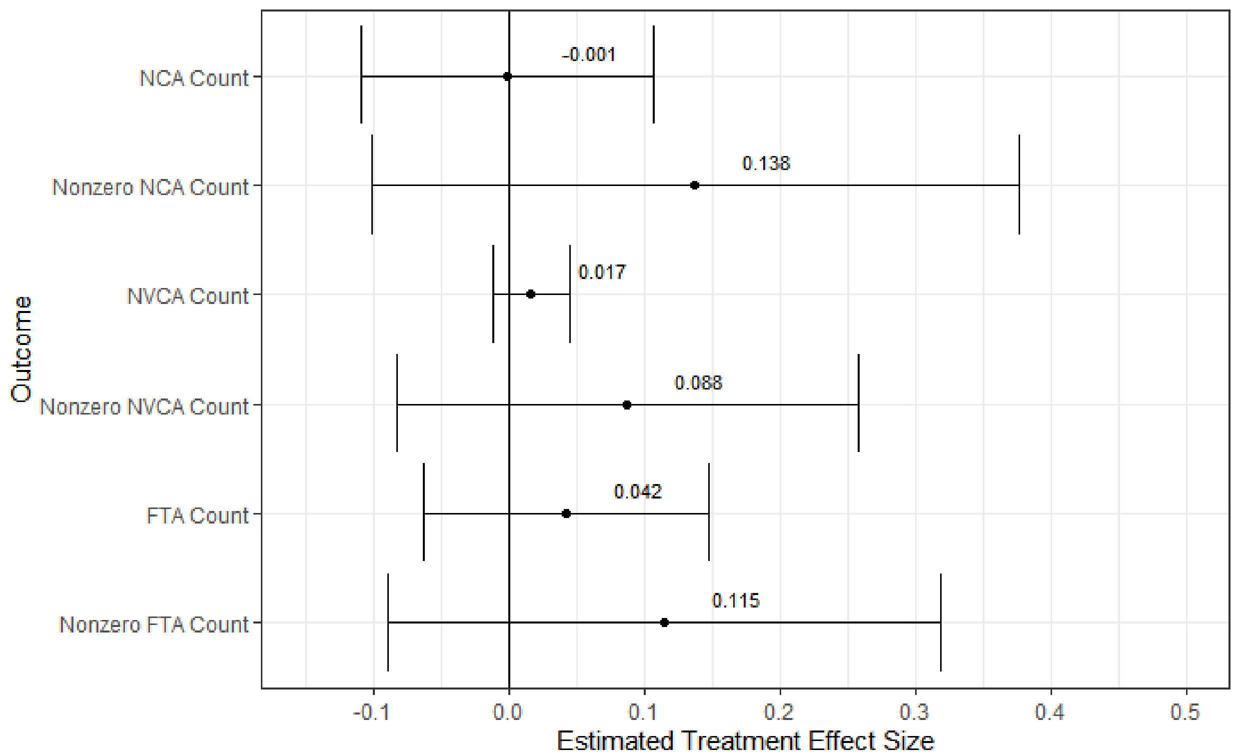

Figure 11: This graph shows the 95% confidence intervals for the primary outcome variables considered as counts. All of the intervals include 0, suggesting that we lack evidence to conclude that the availability of the PSA-DMF System printout caused a statistically significant impact on the outcomes listed on the y axis. That said, several of the intervals are wide. We lack the statistical power to rule out effects of a size that we would consider to be policy-relevant.

Figure 12: Rate Outcome Treatment Effect Sizes with Confidence Intervals

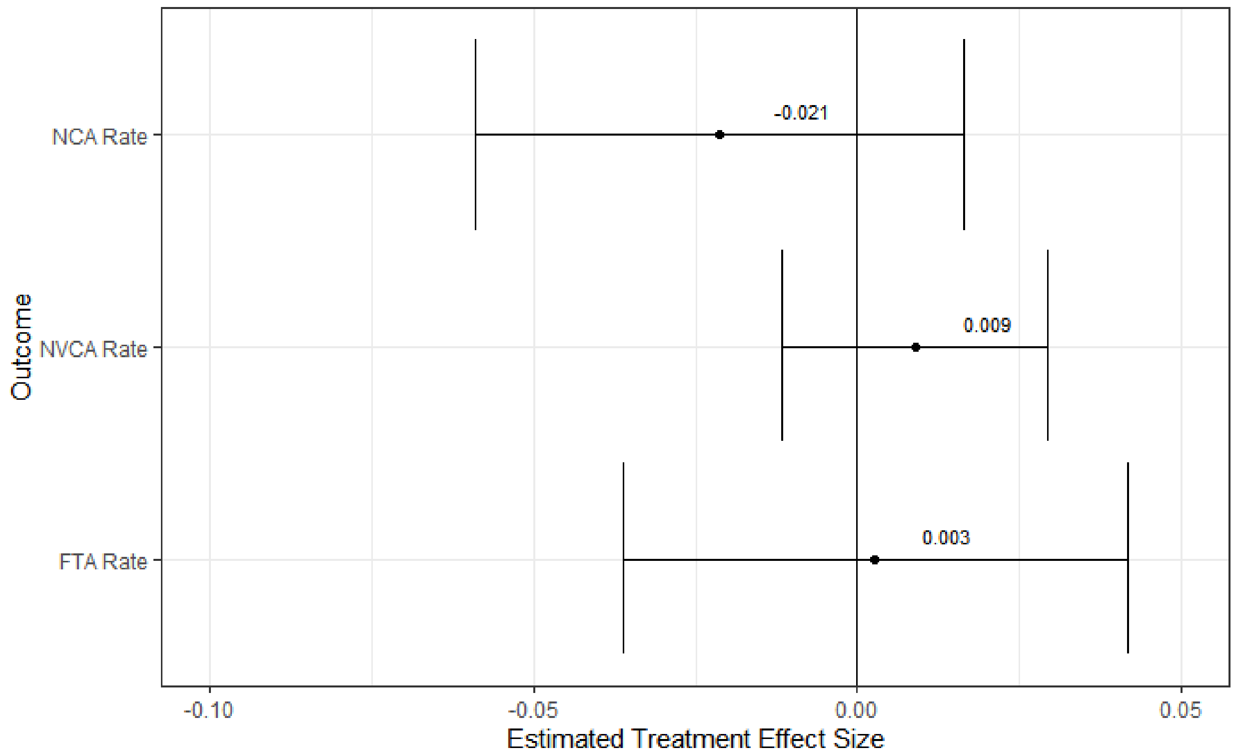

Figure 12: This graph shows the 95% confidence intervals for the primary outcome variables considered as rates or fractions. All of the intervals include 0, suggesting that we lack evidence to conclude that the availability of the PSA-DMF System printout caused a statistically significant impact on the outcomes listed on the y axis. That said, several of the intervals are wide. For example, for the FTA rate, we are unable to detect and rule out differences as large as 3.5 percentage points in either direction. We lack the statistical power to rule out effects of a size that we would consider to be policy-relevant.

The scale of Figures 11 and 12 does not permit inclusion of the permutation-test 95% confidence interval for the total number of predisposition days incarcerated, which is (-10.5, 2.9). This interval is wide. In other words, and taking some liberties with statistical terminology, it is reasonably likely that the effect of the availability of the PSA-DMF System printout is as large as, say, a ten-day reduction in the average number of days of predisposition incarceration, or a two-day increase, but we cannot so detect.

Another way of analyzing the FTA and N(V)CA outcomes is through the lens of time to failure, meaning the length of time from the First Appearance hearing until either the first FTA or N(V)CA event for cases that saw at least one such event. For this analysis, shorter times to failure are bad. Note that an arrestee's opportunity to fail and likelihood of failure may be affected by the length of the predisposition period, something the presence of the PSA-DMF System printout could affect, so this analysis should not be overemphasized.

Figure 13 shows times to failure by treatment conditions. There are no visually obvious patterns on this graph, and the permutation test p-value is high. At least from this permutation test, we have no evidence of a statistically significant treatment effect for time to failure. That said, the usual caveat regarding an inability to detect policy-relevant effect sizes applies, this time coupled with the fact that time to failure does not have a standard distribution, and may be appropriate for modeling.

Figure 13: Time to Failure by Treatment Condition

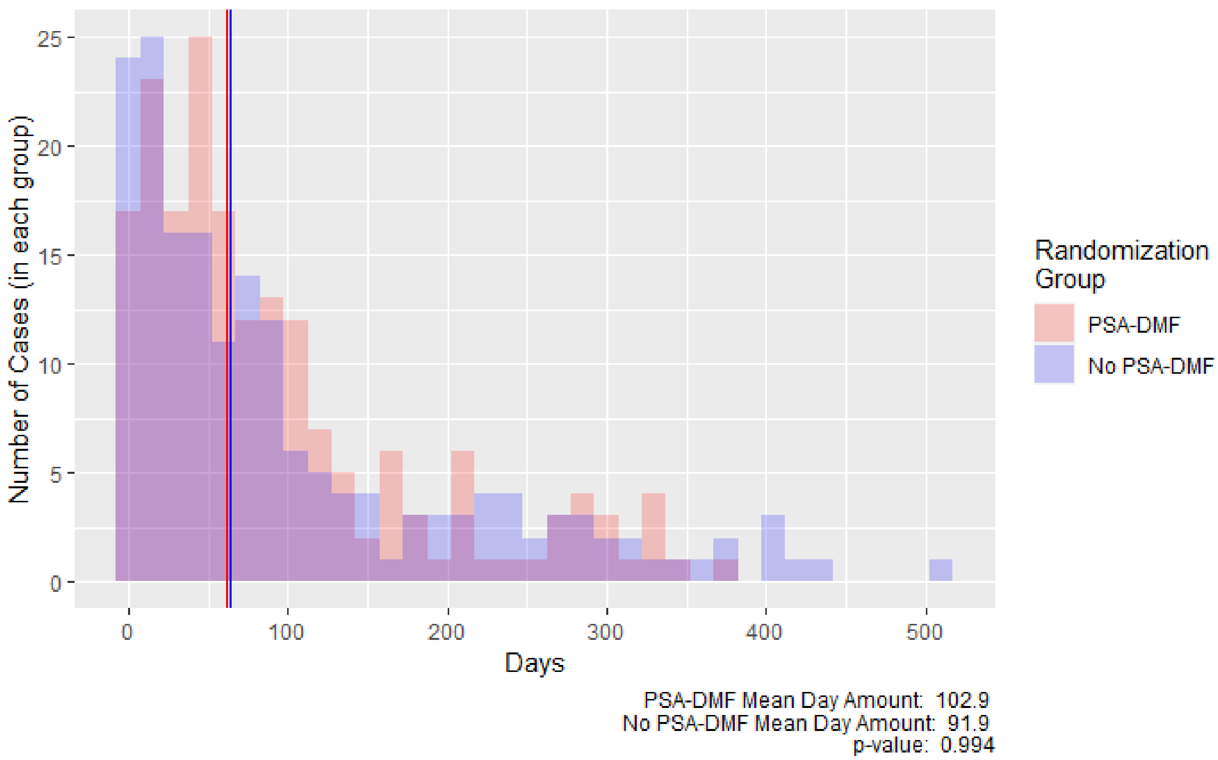

Figure 13: This graph shows overlaid histograms of time lapse from First Appearance hearing to first FTA or N(V)CA by treatment condition. There are no obvious visual patterns, and the permutation test p-value is large, suggesting no evidence of a statistically significant treatment effect with respect to this outcome. That said, we lack sufficient power to detect effect sizes we consider policy-relevant, and this variable's unusual distribution may render it appropriate for modeling. Note that the data used to generate this figure consist of cases in which a failure (either NCA or FTA) actually occurred. Furthermore, as is true of most histograms, none of the bars represent a single x axis value (number of days). Rather, the graph uses bins to represent the number of cases that fall within a fixed interval on the x-axis. Here, each bin has a range of 15 days. For example, the PSA-DMF-present bin centered on the 0 value of the x-axis indicates that roughly 17 cases recorded a failure between 0 and 15 days of the first appearance hearing.

To provide more nuance regarding our primary outcomes, we analyzed them for demographically defined subsets of the arrestee population. Our conclusions are the same with respect to all subsets: we find no evidence of a causal effect of the PSA-DMF System on any primary outcome variable, but we currently lack statistical power to detect what we consider policy-relevant differences.

Figure 14 shows the side-by-side comparisons across our two treatment groups for FTA, NVCA, and NCA, each conceptualized as binary outcome variables for black arrestees. The vertical brackets are standard error bars. Permutation test p-values appear in the caption. There is no evidence of a statistically significant effect for black arrestees.

Figure 14: Binary Outcome Comparisons for Black Arrestees

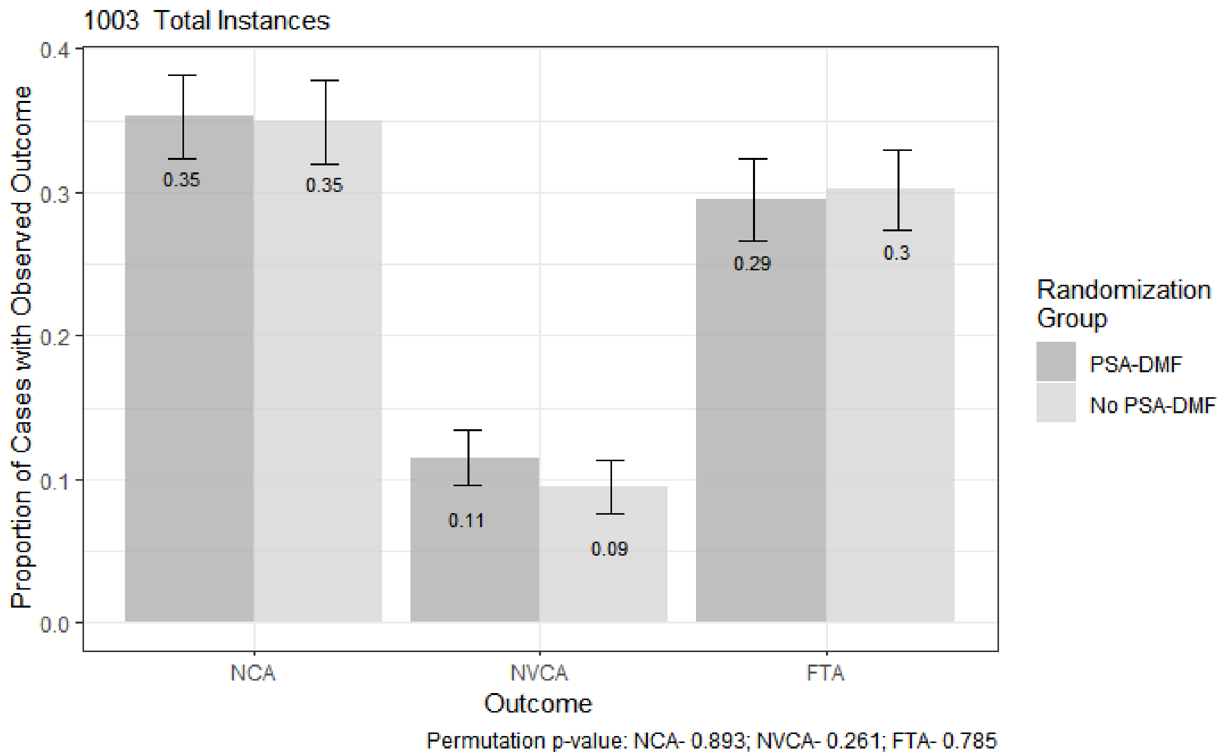

Figure 14: This graph shows side-by-side comparisons across treatment conditions of the proportions of First Appearance hearing events with black arrestees that resulted in at least one instance of the specified outcome. Permutation test p-values appear in the x-axis label. Rates for all three of NCA, NVCA, and NCA are similar, and none of the observed differences is close to statistical significance. We conclude that we have no statistical evidence that the PSA-DMF System had a causal impact on these three outcome variables for black arrestees. That said, the width of the standard error brackets signals substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 15 shows the corresponding graph for white arrestees. Once again, there is no evidence of a statistically significant effect.

Figure 15: Binary Outcome Comparisons for White Arrestees

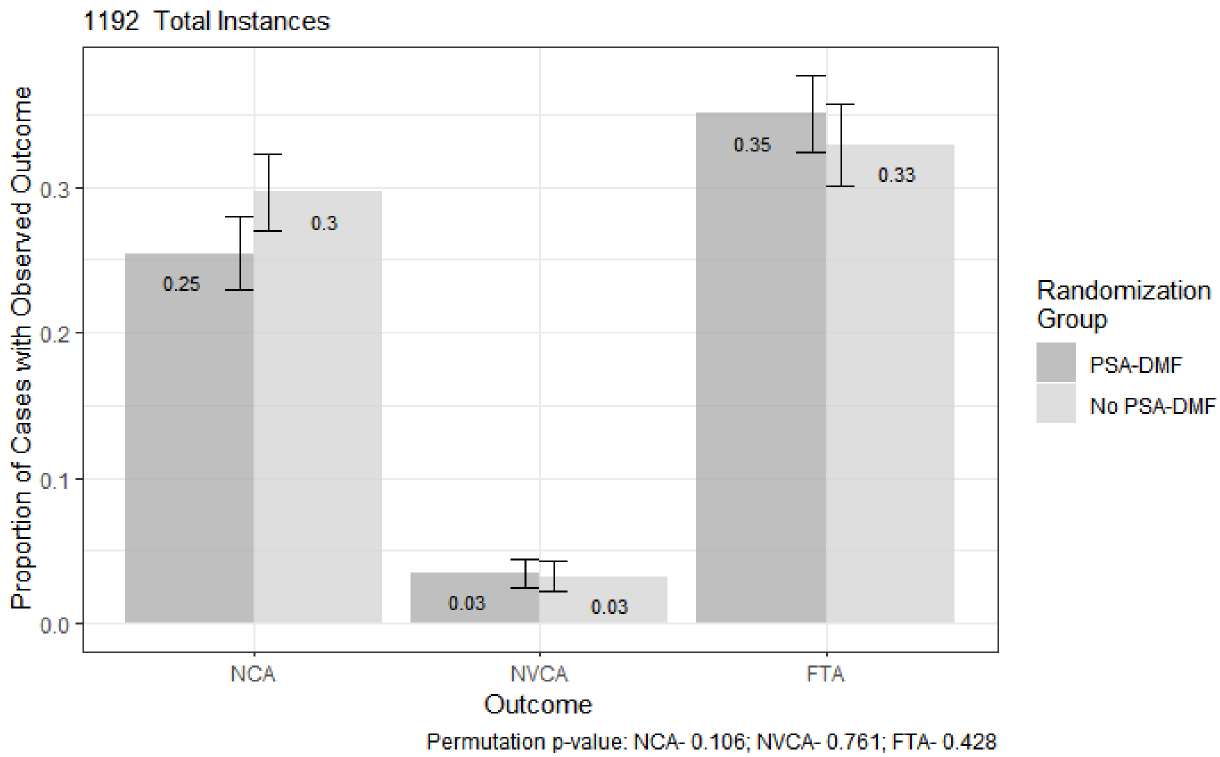

Figure 15: This graph shows side-by-side comparisons across treatment conditions of the proportions of First Appearance hearing events with white arrestees that resulted in at least one instance of the specified outcome. Permutation test p-values appear in the x-axis label. Rates for all three of NCA, NVCA, and NCA are similar, and none of the observed differences is close to statistical significance. We conclude that we have no statistical evidence that the PSA-DMF System has a causal impact on these three outcome variables for white arrestees. That said, the width of the standard error brackets signals substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 16 shows the corresponding graph for male arrestees. Once again, there is no evidence of a statistically significant effect.

Figure 16: Binary Outcome Comparisons for Male Arrestees

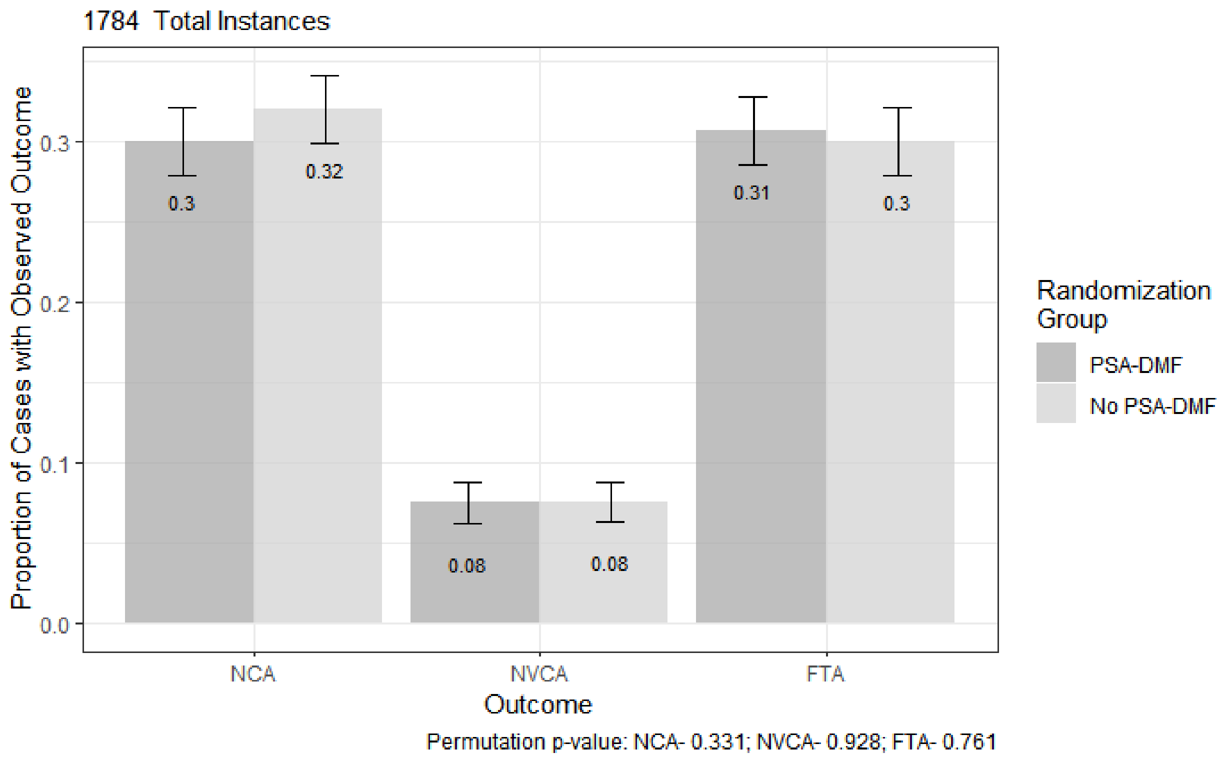

Figure 16: This graph shows side-by-side comparisons across treatment conditions of the proportions of First Appearance hearing events with male arrestees that resulted in at least one instance of the specified outcome. Permutation test p-values appear in the x-axis label. Rates for all three of NCA, NVCA, and NCA are similar, and none of the observed differences is close to statistical significance. We conclude that we have no statistical evidence that the PSA-DMF System has a causal impact on these three outcome variables for male arrestees. That said, the width of the standard error brackets signals substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

The corresponding graph for female arrestees, Figure 17, below, requires slightly more explanation. There were fewer female arrestees than male arrestees, resulting in fewer FTA, NVCA, and NCA events to analyze. Credible analysis was still possible for FTA and NCA, but not so for NVCA. There were only 15 NVCA events across both treatment conditions, out of a total of 478 female arrestees. In our view, counts this low ordinarily render most statistical analysis techniques unreliable, and we do not trust the result of our permutation testing for NVCA among female arrestees. We show it below for completeness and as a preview of what we anticipate that we will produce, reliably, in our final report.

Figure 17: Binary Outcome Comparisons for Female Arrestees

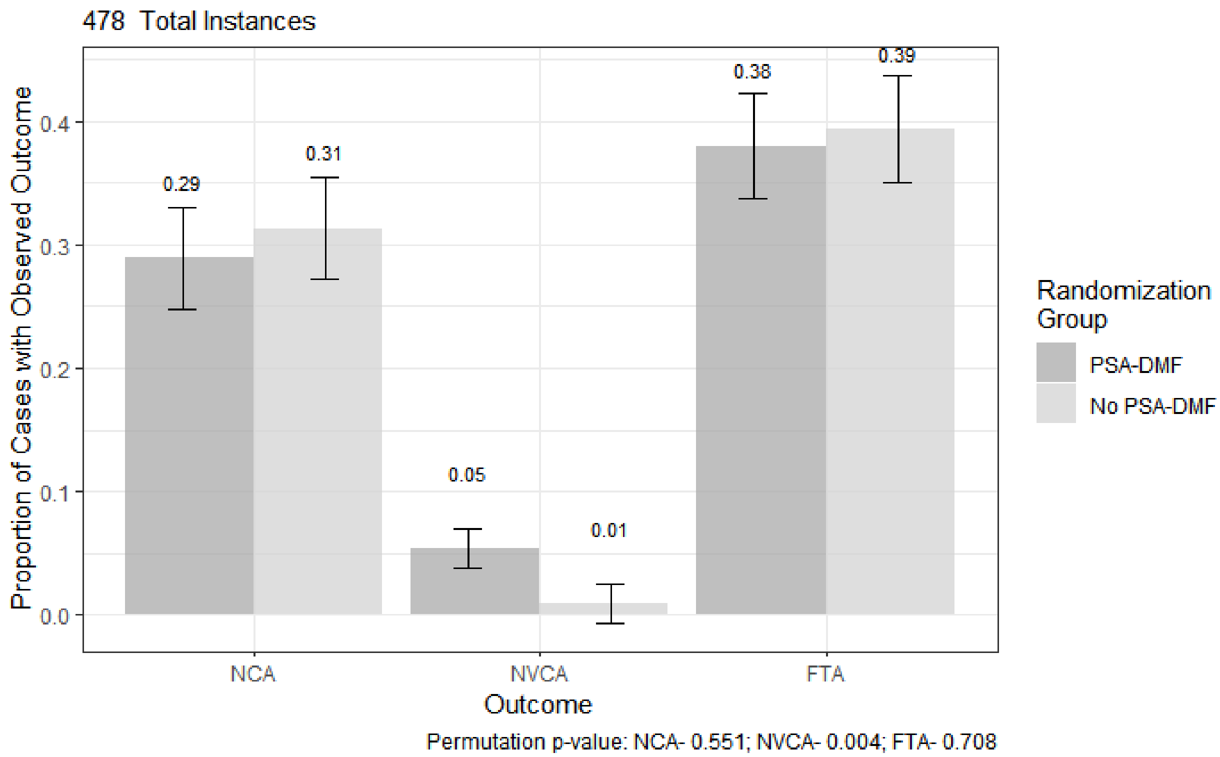

Figure 17: This graph shows side-by-side comparisons across treatment conditions of the proportions of First Appearance hearing events with female arrestees that resulted in at least one instance of the specified outcome. Permutation test p-values appear in the x-axis label. Rates for all three of NCA, NVCA, and NCA are similar. P-values for NCA and FTA are not close to statistical significance. The p-value for NVCA would indicate statistical significance, but the result should not be credited due to the low number of NVCAs (a total of 15 across both treatment conditions) available for inference. We conclude that we have no statistical evidence that the PSA-DMF System has a causal impact on these three outcome variables for female arrestees. That said, the width of the standard error brackets signals substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

We provide additional details regarding each of our primary outcome variables in the following subsections.

i. Effect of the PSA-DMF System on FTA

This subsection addresses the effect of the PSA-DMF System on incidence of FTA.

We begin by sharply defining FTA for the purposes of this report. Under AV definitions, an FTA occurs when an arrestee (i) fails to appear at a hearing, (ii) while on predisposition release, (iii) after receiving clear notice of the appearance date, that (iv) results in a warrant for an arrest. Given the practice among Dane County Commissioners and staff of notifying arrestees of their hearing dates, as well as the text message reminder systems in place,[107] we assume that sufficient notice was provided for all court appearance dates in Dane County. While this assumption might have failed in individual instances, we saw no evidence of systemic failures of notice, much less evidence that notice failures were more or less frequent in the PSA-DMF-System-present group as compared to the System-not-present group. Accordingly, we coded each non-appearance by a released arrestee in our system accompanied by the issuance of a warrant as an FTA.

We note that for the purposes of this report, an FTA can occur only with respect to the legal proceeding corresponding to the particular First Appearance hearing event that constitutes an observation in our data.[108] One implication of this definition is that an FTA can occur only during the time period from First Appearance hearing to case disposition.

With all of this in mind, our fundamental findings with respect to FTA appear in Figures 18 through 19 and in the corresponding confidence intervals, above, with more details below. In short, we find no evidence that the presence or absence of the PSA-DMF System printout has a statistically significant effect on FTA overall, or with respect to subsets defined by race or gender. But, as demonstrated by the FTA 95% confidence intervals produced above, we currently lack power to detect effect sizes that we consider policy-relevant.

First, Figure 18 provides our most basic FTA finding, the comparison of the fraction of First Appearance hearing events that were followed by at least one FTA across our two treatment conditions, *i.e.*, FTA considered as a binary variable.

---

[107] See supra Section I.D.

[108] In other words, for the purposes of this report, we did not consider FTAs from cases or proceedings that might have preexisted the particular First Appearance hearing event randomized in our study, or that might have arisen afterwards. We clarify this point because AV has requested that in our final report we analyze outcome variables with a follow-up period for each observation of two years from First Appearance hearing date, regardless of time to disposition. With a two-year follow-up period, an argument exists that FTA events from cases other than the proceeding corresponding to a particular First Appearance hearing observation should be included in the analysis. This point generalizes to NCA and NVCA.

Figure 18: Binary FTA Rates, Overall And by Treatment Condition

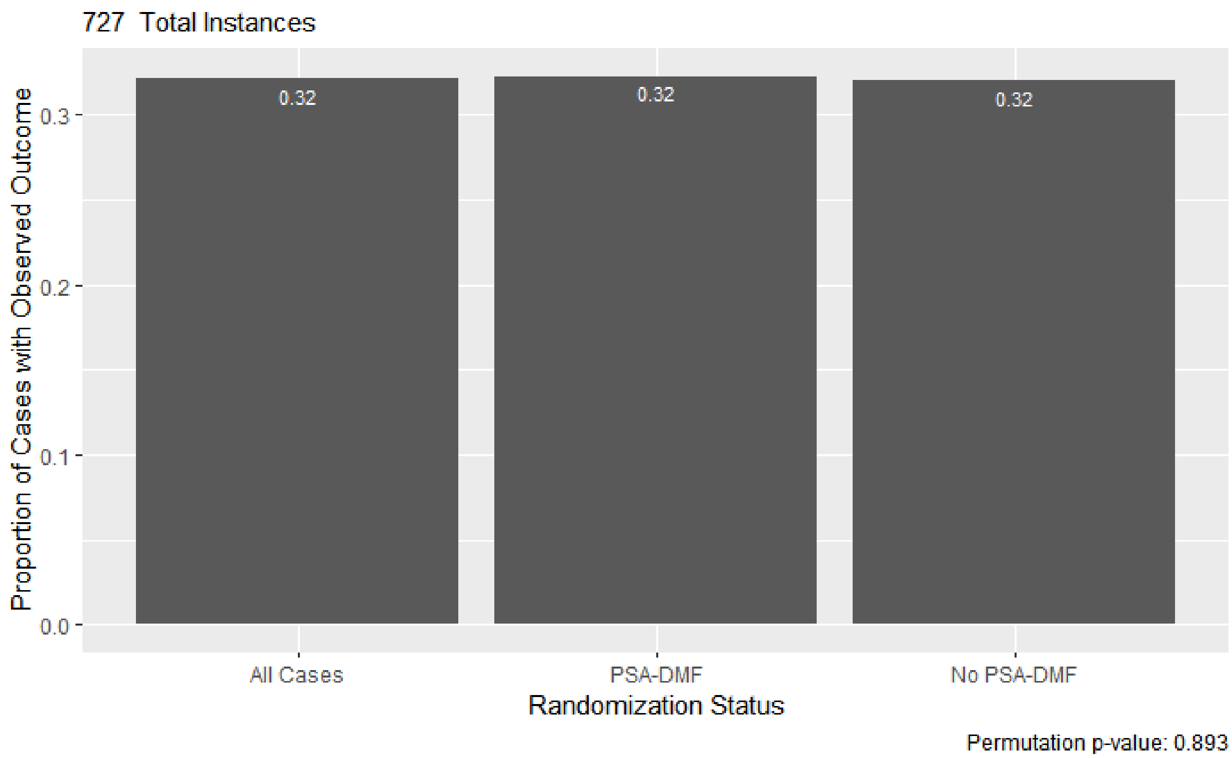

Figure 18: This bar graph shows binary FTA, overall and by treatment condition. The rates for the PSA-DMF-System-present and not-present groups are nearly identical, and the p-value is high (not near the traditional .05 threshold for statistical significance). We have no evidence of a statistically significant treatment effect with respect to FTA considered as a binary variable. That said, the 95% intervals for the FTA treatment effect provided above reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 19 provides the same side-by-side comparison, this time with FTA considered as a count variable, *i.e.* how many FTAs occurred in each proceeding corresponding to the First Appearance Hearing event. For both Figures 18 and 19, the counts or rates across treatment conditions are similar, with p-values not near .05, indicating that we have no evidence of a statistically significant treatment effect for FTA.

Figure 19: Count FTA by Treatment Condition

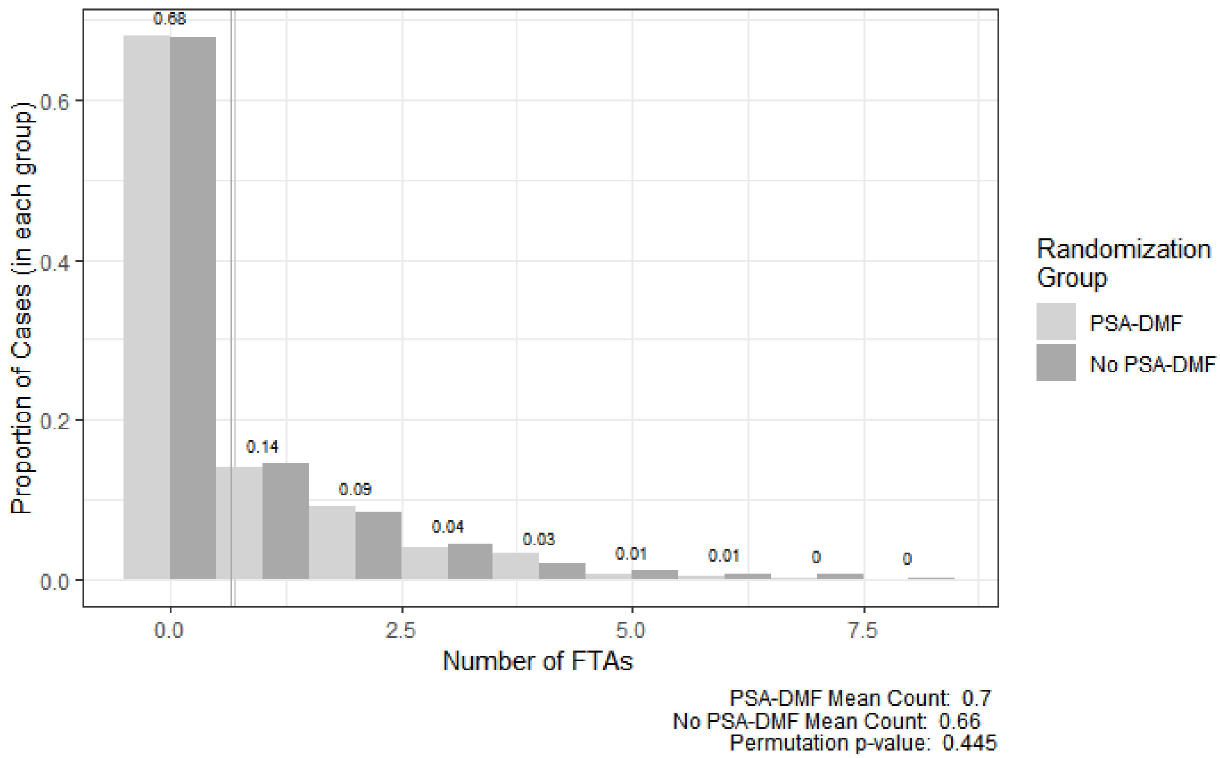

Figure 19: This bar graph shows count FTA by treatment condition. The counts for the PSA-DMF-System-present and not-present groups are similar, and the p-value is not near the traditional .05 threshold for statistical significance. We have no evidence of a statistically significant treatment effect with respect to FTA considered as a count variable. That said, the 95% intervals for the FTA treatment effect provided above reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

ii. Effect of the PSA-DMF System on N(V)CA

This subsection addresses the effect of the PSA-DMF System on incidence of NCA and NVCA.

We begin by sharply defining NCA and NVCA for purposes of this report. Under AV definitions, an NCA occurs when an arrestee (i) is arrested or issued a criminal citation, (ii) while on predisposition release, (iii) for an offense that could result in any amount of jail or prison time. An NVCA occurs when an NCA occurs, but when the offense is classified as violent under the laws of the jurisdiction in which the new offense allegedly occurred.

The Dane County PSA Working Group and many Dane County officials believed that prediction of whether any future criminal activity will be violent as opposed to nonviolent was impossible. Some Dane County Officials also questioned the factors upon which the violence flag is based, contending that they classified charges based on an over- and under-inclusive definition of a violent offense.[109]

For purposes of this report, an N(V)CA can occur only during the time period from First Appearance hearing to case disposition. As we explain in Section II.B.3, below, we will examine incidence of arrest or citation occurring up to two years after each First Appearance hearing in our final report.

We begin with NVCA. Our fundamental findings appear in Figures 20 through 21 and in the corresponding confidence intervals, above, with more details below. In short, we find no evidence that the presence or absence of the PSA-DMF System printout has a statistically significant effect on NVCA overall, or with respect to subsets defined by race or gender. But, as demonstrated by the NVCA 95% confidence intervals produced above, we currently lack power to detect effect sizes that we consider policy-relevant.

Figure 20 provides our most basic NVCA finding, the comparison of the fraction of First Appearance hearing events that were followed by at least one NVCA across our two treatment conditions, *i.e.*, NVCA considered as a binary variable.

---

[109] Dane County CJC Research Team, Memo, “Comments/edits on the Interim Report “Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI,” Memorializing comments and suggested edits to the Interim Report on Apr. 13, 2020 (on file with Access to Justice Lab).

Figure 20: Binary NVCA Rates, Overall And by Treatment Condition

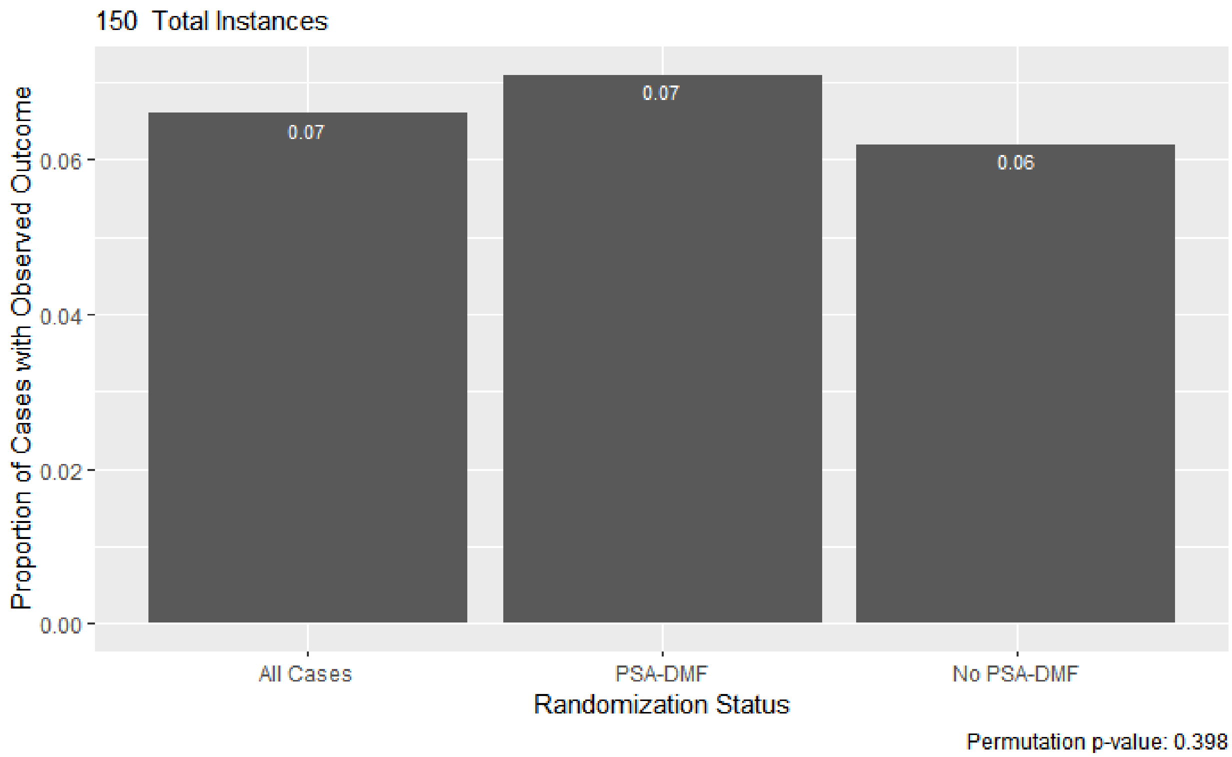

Figure 20: This bar graph shows binary NVCA, overall and by treatment condition. The rates for the PSA-DMF-System-present and not-present groups are similar, and the p-value is high (not near the traditional .05 threshold for statistical significance). We have no evidence of a statistically significant treatment effect with respect to NVCA considered as a binary variable. That said, the 95% intervals for the NVCA treatment effect reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 21 provides the same side-by-side comparison, this time with NVCA considered as a count variable, *i.e.*, how many NVCAs occurred during the pretrial period for the corresponding First Appearance hearing event. For both Figures 20 and 21, the counts or rates across treatment conditions are similar, with p-values not near .05, indicating that we have no evidence of a statistically significant treatment effect for NVCA. That said, the 95% intervals for the NVCA treatment effect provided above reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 21: Count NVCA by Treatment Condition

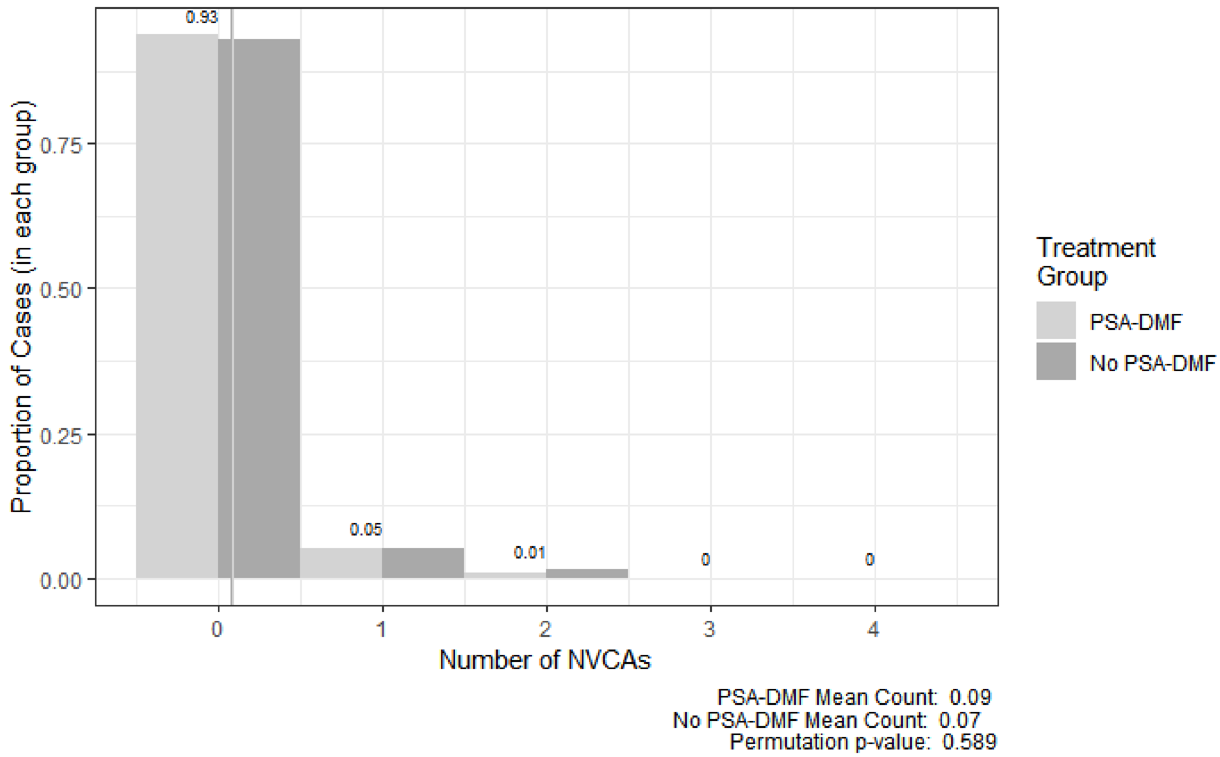

Figure 21: This bar graph shows count NVCA by treatment condition. The counts for the PSA-DMF-System-present and not-present groups are similar, and the p-value is not near the traditional .05 threshold for statistical significance. We have no evidence of a statistically significant treatment effect with respect to NVCA considered as a count variable. That said, the 95% intervals for the NVCA treatment effect reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Our findings with respect to NCA are similar. Once again, our fundamental findings appear in Figures 22 through 23 and in the corresponding confidence intervals, above, with more details below. In short, we find no evidence that the presence or absence of the PSA-DMF System printout has a statistically significant effect on NCA overall, or with respect to subsets defined by race or gender. But, as demonstrated by the NCA 95% confidence intervals produced above, we currently lack power to detect effect sizes that we consider policy-relevant.

Figure 22 provides our most basic NCA finding, the comparison of the fraction of First Appearance hearing events that were followed by at least one NCA across our two treatment conditions, *i.e.*, NVCA considered as a binary variable.

Figure 22: Binary NCA Rates, Overall And by Treatment Condition

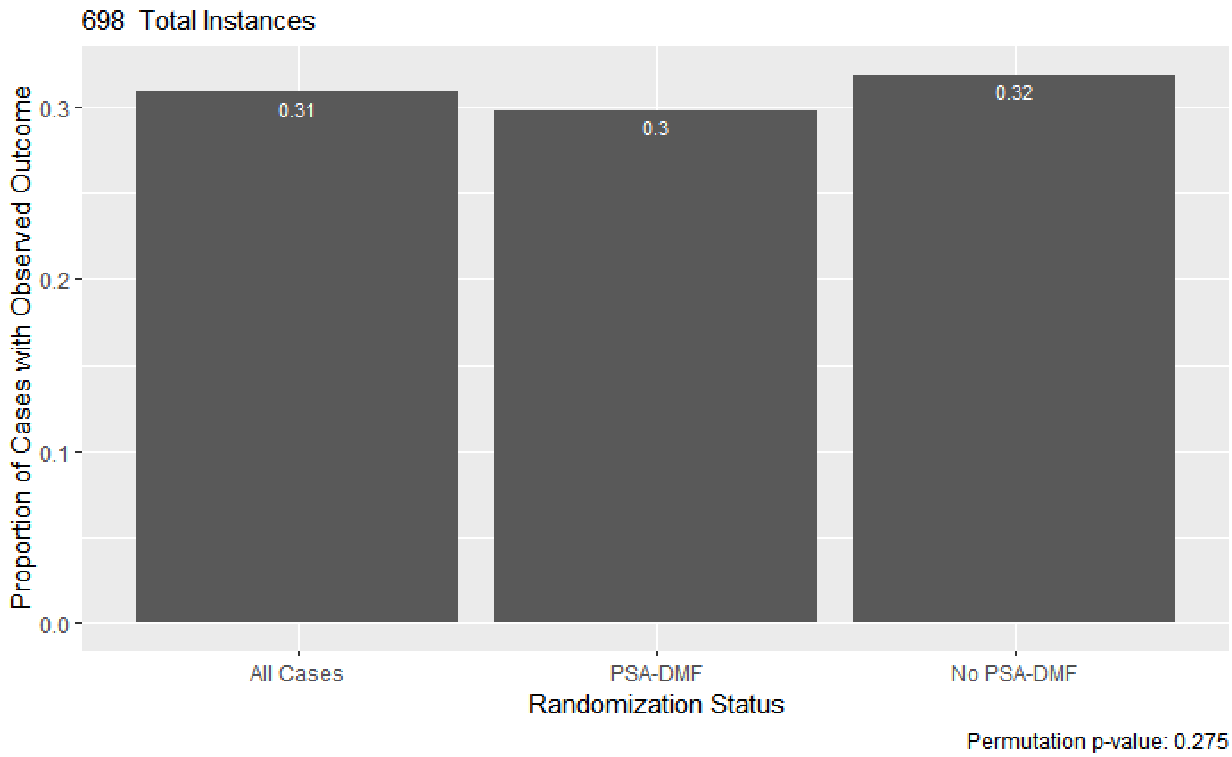

Figure 22: This bar graph shows binary NCA, overall and by treatment condition. The rates for the PSA-DMF-System-present and not-present groups are similar, and the p-value is high (not near the traditional .05 threshold for statistical significance). We have no evidence of a statistically significant treatment effect with respect to NCA considered as a binary variable. That said, the 95% intervals for the NCA treatment effect reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 23 provides the same side-by-side comparison, this time with NCA considered as a count variable, *i.e.*, how many NCAs occurred during the pretrial period for the corresponding First Appearance hearing event. For both Figures 22 and 23, the counts or rates across treatment conditions are similar, with p-values not near .05, indicating that we have no evidence of a statistically significant treatment effect for NCA. That said, the 95% intervals for the NCA treatment effect provided above reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

Figure 23: Count NCA by Treatment Condition

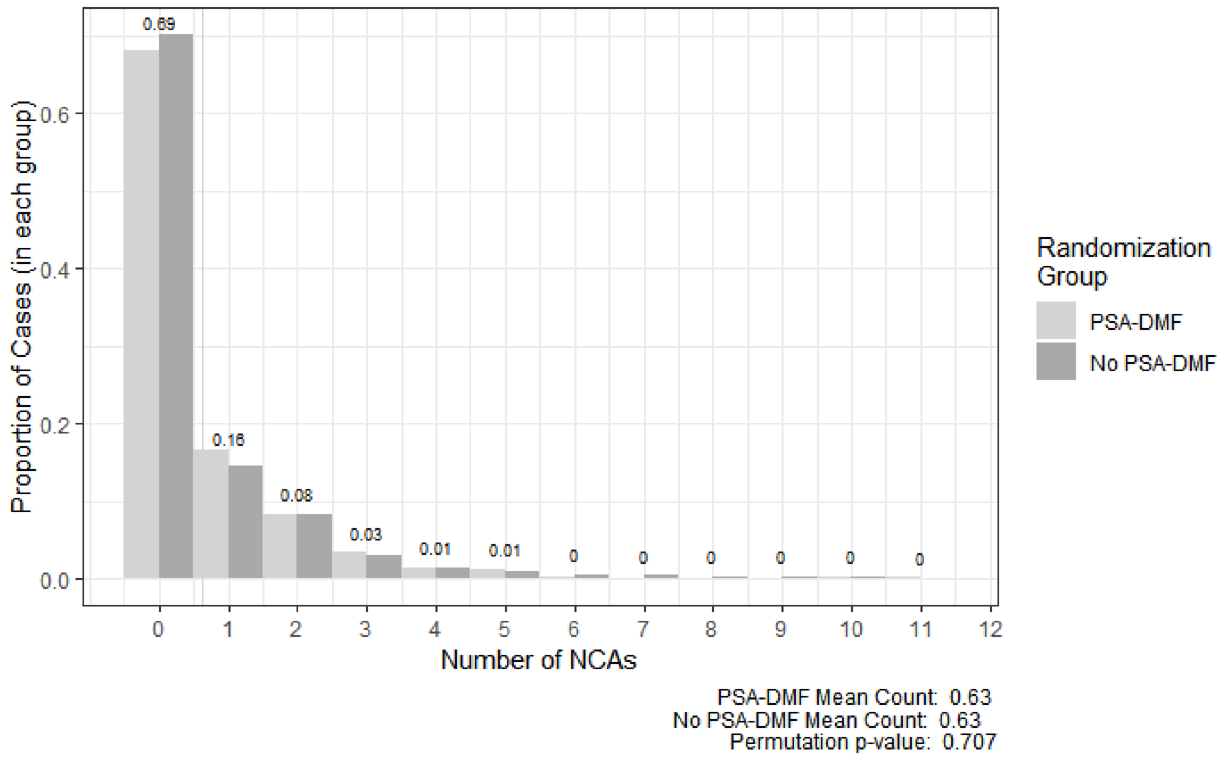

Figure 23: This bar graph shows count NCA by treatment condition. The counts for the PSA-DMF-System-present and not-present groups are similar, and the p-value is not near the traditional .05 threshold for statistical significance. We have no evidence of a statistically significant treatment effect with respect to NCA considered as a count variable. That said, the 95% intervals for the NCA treatment effect reflect substantial uncertainty. We lack statistical power to detect causal effects of a size we consider policy-relevant.

iii. Effect of the PSA-DMF System on Predisposition Incarceration

This subsection addresses the effect of the PSA-DMF System on the number of days incarcerated during the predisposition time period associated with each randomized First Appearance hearing event. It reports the results of both permutation testing and the zero-inflated model.

Predisposition incarceration differs from the other main outcomes because it measures a length of time, in days, as opposed to an incident rate of a specific post-release outcome. Arrestees may have increased incarceration lengths because of an inability to post bail or because of an arrest on new criminal charges while released, among other reasons. The length of predisposition incarceration also is a function of the pretrial period itself; criminal cases that reach disposition sooner naturally have shorter potential predisposition incarceration lengths. These attributes make analyzing the impact of the PSA-DMF system on predisposition incarceration somewhat more complicated than the other main outcome variables.

Figure 24: Overall Distribution of Predisposition Incarceration Length

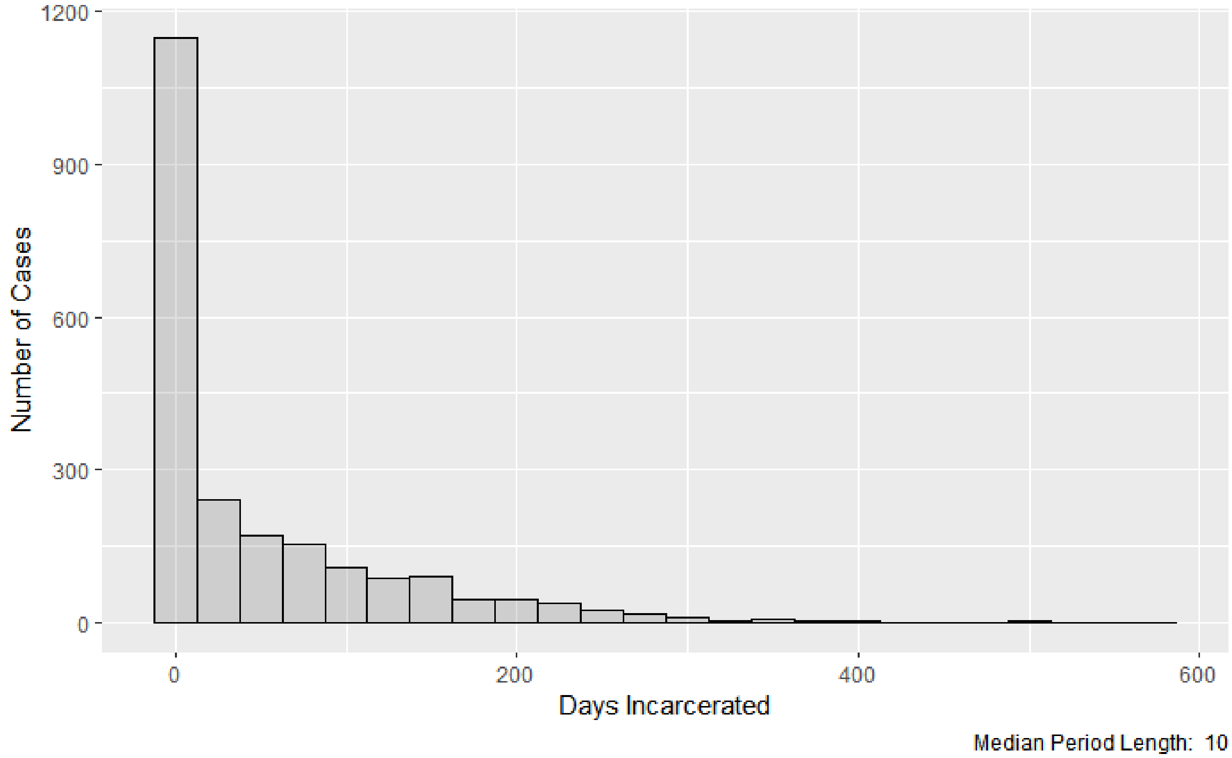

Figure 24: This histogram displays the distribution of predisposition incarceration length. Each bar represents a period of 25 days, indicating that the majority of arrestees spend 25 or less total days of their pretrial period incarcerated. However, the distribution of predisposition trial length has significantly long tales, with a maximum value in the high 500s. The median predisposition incarceration period is 10 days, indicating that roughly half of the arrestee populations spends less than one and half weeks incarcerated prior to case disposition.

Figure 25: Distribution of Predisposition Incarceration Length by Randomization Group

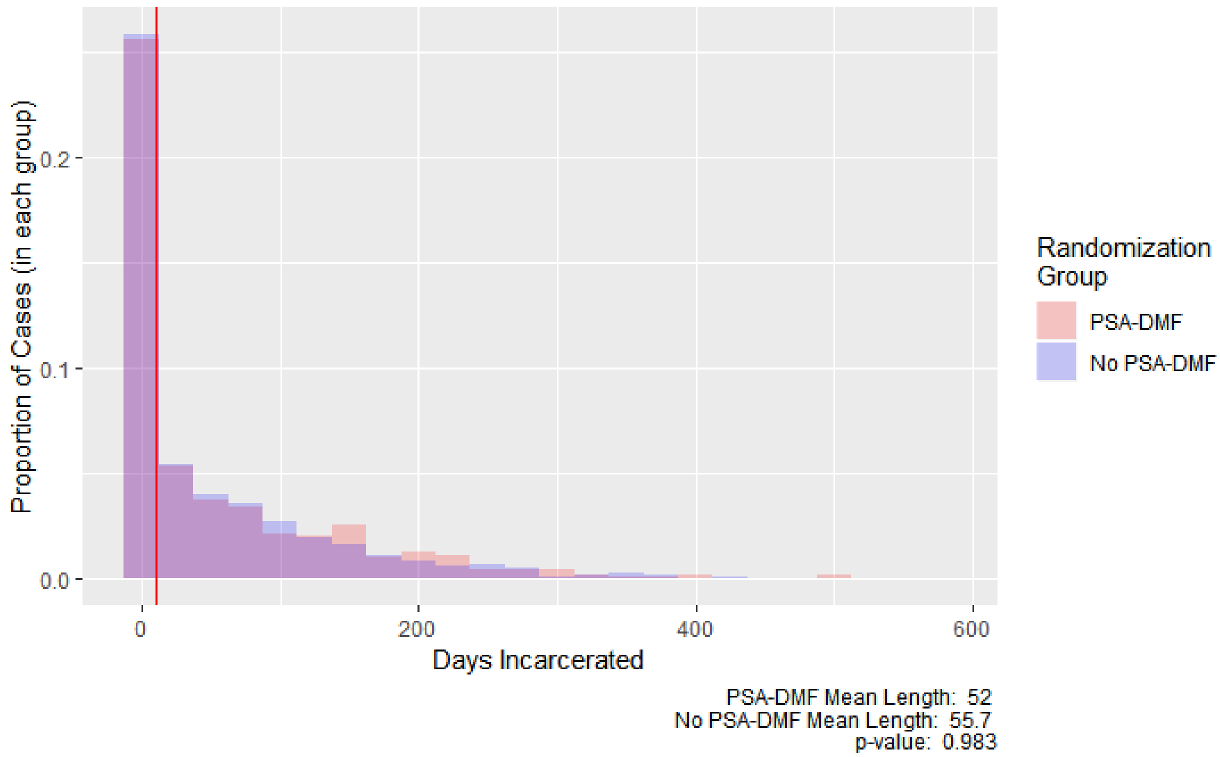

Figure 25: This figures shows overlapping histograms of the distribution of predisposition incarceration length for each randomization group. The blended color area represents overlap between the randomization groups, while areas of primary color represent places where the two distributions do not intersect. The large proportion of blended color relative to primary color indicates that the two groups have very similar distributions of predisposition incarceration length. This is additionally confirmed by the similar mean predisposition incarceration length values (52 and 55.7 for PSA-DMF System present cases and no PSA-DMF System present cases, respectively), as well as the highly insignificant p-value of 0.983 on the permuted difference of means statistic.

Figures 24 and 25 report the distribution of predisposition incarceration, overall and by treatment condition, respectively. Predisposition incarceration is calculated as the time spent in the Dane County Jail between two dates: the First Appearance hearing date and the case disposition date. The overall distribution has a minimum value of 0 days incarcerated and a maximum value of 568 days. Despite this dispersion, the vast majority, about 59% of arrestees, spent less than 30 predisposition days incarcerated. Of those cases, 188, about 8%, spent 0 days incarcerated predisposition. The median duration was 10 days. The overall mean was 53.8 days, indicating that large values affected the mean.

We initially estimate the treatment effect of the PSA-DMF System on the length of pretrial incarceration by calculating the difference in means for the two treatment conditions. The PSA-DMF-System-present group had a mean predisposition incarceration length of 52 days, and the nPSA-DMF-System-not-present group had a mean of 55.7 days. The resulting estimated treatment effect size was -3.7 days, or a decrease of about 6.8% from the overall mean. The effect size estimate had a 95% confidence interval between -10.5 days and 2.9 days. Because the interval includes 0, we have no statistical evidence to contradict an initial hypothesis that the true difference of means, *i.e.*, the treatment effect size, is 0. The large p-value of 0.983 captures the same idea. The data do not provide support for the conclusion that the PSA-DMF System significantly caused a change in total predisposition incarceration lengths.

The above calculations, as with all the calculations involving the study groups, include only a subset of the total criminal defendant population. Defendants who were not arrested, who bonded out prior to a bail hearing, or whose cases did not receive a PSA for reasons like staleness are not included. This data should not be interpreted to reflect the entire criminal defendant population.

Despite the initial lack of evidence for a treatment effect from permutation testing, there are reasons to believe that the distribution predisposition incarceration time would benefit from a different estimation methodology. Recall that a substantial fraction of arrestees spent 0 days incarcerated, but that some arrestees had lengthy spells in the Dane County Jail. For variables distributed in this way, analyses based on means and medians can sometimes suppress real differences. Moreover, we might believe that cases resulting in 0 days of predisposition incarceration were sufficiently different from those with lengthy Jail times. This phenomenon is known as "zero inflation." Zero inflation occurs when a subset of zeros observed in the data are a result of the same process as the rest of the data, but some zeros are the result of a separate process (meaning, here, a different kind of case). We used a zero-inflation model to re-estimate the treatment effect of the PSA-DMF System on pretrial incarceration length. The zero-inflation model works by estimating effect sizes in two steps. The first estimates the probability that a zero in the data set is a result of the underlying data process that generates the rest of the data or a separate process. The second estimates the effect sizes of variables on the main count dependent variable utilizing the corrective estimates from the first model.

The variables used to estimate the first stage of the zero-inflation model were pretrial length and felony charge. The rationale is that cases with shorter lengths are less likely to have long predisposition incarceration.[110] Moreover, cases with at least one felony charge are more complex, and as a result, less likely to be quickly resolved. The variables used to estimate the second stage of the zero-inflation model included the treatment condition, the DMF recommendation, demographic information (*e.g.*, race, gender, age), and the total count of different release conditions. The main independent variable of interest here is the treatment condition.

[110] We recognize that case length may be itself affected by treatment condition, but deemed that it would have predictive power for the present analysis. In our final report, we will fit additional zero-inflation models that rely entirely on pretreatment variables.

Figure 25: Zero-Inflation Model Predictions of Predisposition Incarceration Length

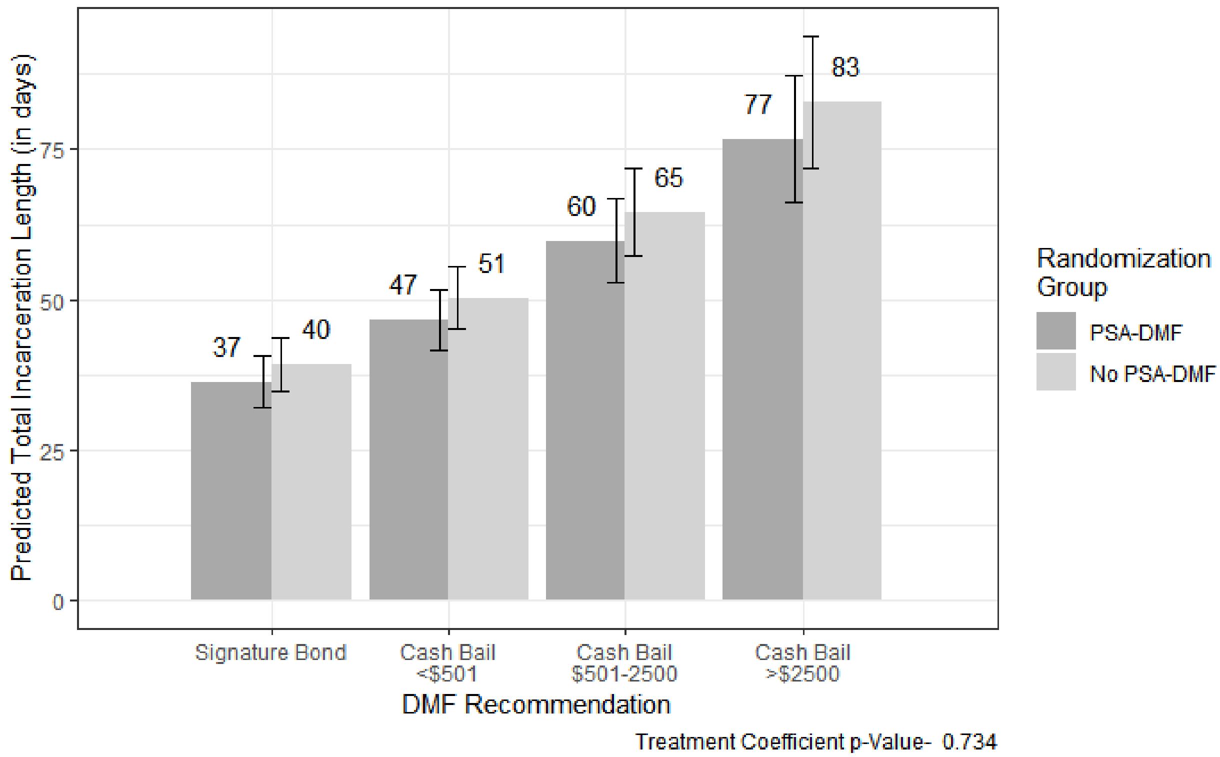

Figure 25: This bar graph shows predicted predisposition incarceration lengths for a zero-inflation model across all permutations of two substantively important model covariates: Randomization group and DMF Recommendation. The predictions are in line with expectations of the performance of each covariate: the coefficient estimate of the PSA-DMF System treatment effect serves to decrease predicted predisposition incarceration lengths while higher risk factor levels of the DMF Recommendation serve to increase predicted incarceration lengths. Despite the estimated coefficient direction of the treatment effect, the estimate itself does not achieve significance, with a p-value on the estimate of 0.734.

Figure 25 reports predicted average predisposition incarceration lengths for four categories of the DMF recommendation variable by randomization group. The estimated treatment effect from the model is -0.078, which indicates that cases assigned to the PSA-DMF-System-present group were associated with lower average predisposition incarceration lengths after correcting for inflated zero counts. By exponentiating this effect size, we recover what is known as a "rate ratio," or the multiplicative effect on the predicted predisposition incarceration length between the PSA-DMF-System-present group and the System-not-present group. The rate ratio of 0.925means that the PSA-DMF System is associated with a 7.5% decrease in mean predisposition incarceration length. However, this effect size exists on a 95% confidence interval from 0.817 to 1.031, meaning that the true value very likely exists between an 18.3% decrease and a 3.1% increase in predisposition incarceration length. Therefore, we cannot reject the possibility that the true effect size, corrected for the presence of inflated zeros, is 0, or even slightly positive. This is further confirmed by the p-value obtained for this effect size of 0.734.

In summary, neither the initial difference of means permutation tests nor the zero-inflation model provide evidence that the PSA-DMF System caused a difference in the total number of days spent incarcerated during the predisposition period. Both methods generated negative treatment effect size estimates, but neither produced the requisite statistical confidence to rule out t true effect sizes of 0. As noted above, we lack statistical power to detect causal effects of a size we consider policy-relevant.

iv. Effect of the PSA-DMF System on Demographic Fairness Measures

This subsection addresses the effect of the PSA-DMF System with respect to measures of demographic fairness. After an introduction defining terms used throughout, this subsection proceeds in four subsections, corresponding to the four different types of demographic fairness analysis we conducted. Overall, with the exception of a few (of many) permutation tests, mostly related to gender, none of the comparisons we conducted resulted in statistically significant differences in the demographic fairness metrics as compared across the PSA-DMF System-present and not-present groups. For reasons we explain below, we do not credit the few comparisons producing statistically significant differences as substantively significant, not at least until we have more data.

We begin with basic background. The potential for bias, however defined, to emerge from the use of criminal justice risk assessment tools has been of interest to criminal justice reformers. A well-publicized report by *ProPublica* in 2016 alleged that one popular risk assessment instrument was biased.[111] Since then, the existence, extent, and measurement of bias in risk assessment tools has garnered more attention.

The A2J Lab's objective in this report is not to determine whether a particular tool is, in an absolute sense, fair or biased, or whether a specific metric or methodological approach should be used over another. The literature contains multiple approaches for assessing bias that present analytical tradeoffs.[112] The goal of this subsection is to identify key commonalities among the different approaches and to develop an approach tailored to the nature of the data used in this study.

Many analyses of bias in algorithmic assessment tools begin with accuracy measures of classification devices used in machine learning and computer science. Often referred to as "classification" or "confusion matrices," these tables relate predictions to observations in a cross-tabular binary format. These matrices provide the two most common starting points in

---

[111] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias," available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing ProPublica (last visited Mar. 7, 2020)

[112] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, & Aaron Roth. Fairness in Criminal Justice Risk Assessments: The State of the Art, Soc. Meth. & Res. (2018).

bias assessment: false positive and false negative rates. We borrow the nomenclature from Berk et al. to produce the following confusion matrix.

	Failure Observed (Negative Outcome Observed)	Success Observed (No Negative Outcome Observed)	Conditional Use Error
Predicted Failures (Negative Outcome Predicted)	TP True Positives	FP False Positives	$FP/(TP+FP)$ Failure Prediction Error
Predicted Successes (No Negative Outcome Predicted)	FN False Negative	TN True Negative	$FN/(TN+FN)$ Success Prediction Error
Conditional Procedure Error	$FN/(TP+FN)$ False Negative Rate	$FP/(TN+FP)$ False Positive Rate	$\frac{(FP+FN)}{(TP+FP+TN+FN)}$ Overall Procedure Error

Despite the widespread use of false positive and false negative rates, these metrics exist in tension; as one decreases the other, by the nature of conditional probability must increase. This fact has led to the prevalence of other metrics based on joint use of false positive and false negative rates, like the "overall procedure error" metric detailed in Berk et al.'s work and defined in the table above.

Another common theme in the risk assessment bias literature is the concept of disparate impact.[113] Disparate impacts exist when a facially neutral instrument generates a disproportionately adverse impact for a particular group or groups.

Much of the risk assessment disparate impact literature focuses on differential accuracy and performance. Specifically, the concern is about properly attributing disparate impact to risk assessment tools rather than some other external factor(s). As one scholar observed regarding in the ProPublica study, "the differences in false positive and false negative rates cited as evidence of racial bias . . . are a direct consequence of applying an instrument that is free from predictive bias to a population in which recidivism prevalence differs across groups."[114] This phenomenon, the misattribution of measured disparate impact due to other sources of differential group metrics, is a common theme in the literature.

In this subsection, we propose a method for leveraging the data generated by the RCT study design. This subsection proceeds by discussing four groups of fairness metrics: false positive rates, false negative rates, overall procedure error, and mean outcome disparate impact

---

[113] Alexandra Chouldechova, Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, 5 Big data 153 (2017).
[114] Id. at 154.

measures. We construct each measure in ways that depart from prior approaches in the literature because of the properties of the RCT-generated data. Instead of making absolute assessments of bias, we measure the causal effect of the availability of the PSA-DMF System printout on the demographic fairness measures in the literature.

(1). False Positive Rates

This subsection discusses the false positive demographic fairness measure. We begin with an explanation of this metric before providing results of comparisons focusing on race and gender.

A false positive incident occurs when a decision maker predicts that an outcome of interest will occur, but it does not. Under this measure, purportedly, a predictive tool is unbiased if the relative occurrence of false positives is roughly equivalent across groups of interest. The relative occurrence of false positives is measured against the total number of actual negative cases, both those predicted to be negative (true negatives) and those predicted to be positive (false positives). The resulting false positive rate is the number of false positive observations divided by the total number of observed negative cases (True Negatives + False Positives). The false positive rate and the false positive group equivalency definition are stated in equations 1.1 and 1.2, below.

False Positive Rate

$$FP_R = \frac{False\ Positives}{True\ Negatives + False\ Positives} \#1.1$$

False Positive Group Equivalency

$$FP_{R|j=1} = FP_{R|j=2} \#1.2$$

Several factors complicate the calculation and analysis of the false positive rate for the PSA-DMF System in Dane County. First, the PSA-DMF System does not provide binary, or any, predictions about specific arrestees. Second, the complexity of the criminal justice system means that, there are multiple potential sources of bias that might also affect the false positive rate other than the risk assessment tool. The first concern complicates the actual calculation of the false positive rate, and the second complicates our ability to draw direct conclusions from a simple evaluation of the group equivalency definition in equation 1.2.

Because the PSA-DMF System does not provide predictions, we develop several false positive rates based on hypothetical binary predictions generated from the raw PSA-DMF scores. Take for instance the NCA score, which is calculated on a 1-6 scale. Under the PSA-DMF System, a higher NCA score is understood to mean a higher risk of an arrestee having being arrested again for new charges if released. Thus, we can construct a hypothetical false positive rate by making an assumption that all individuals with an NCA score of at least 4 are predicted to have an NCA incident. We then tally the number of arrestees in the dataset with a score of less than 4 who did not have an NCA incident (true negatives) and those with a score of 4 or more who also did not have an NCA incident (false positives). Using these sums, we can calculate a false positive rate for the hypothetical threshold of 4. This routine is repeated for the potential threshold values of 2, 3, 5, and 6.[115] This process produces five separate hypothetical false positive rates and permits inference about potential disparate impacts across the NCA scale.

Although we can calculate the difference between group-specific false positive rates across all hypothetical score values for relevant outcomes, we cannot attribute any resulting differences to the PSA-DMF System alone. The reason, as noted above, is that the false positive rate will also capture non-System sources of potential disparate impact. Fortunately, the underlying, randomized data generating process mitigates this concern. Instead of drawing comparisons directly between groups, we measure the difference in false positive rates for the same group across treatment groups; any difference between them would be directly attributable to the PSA-DMF System alone. As a result, if we observe a statistically significant PSA-DMF System treatment effect for one group but not another, we would have some evidence of disparate impact. We can still make direct group comparisons by utilizing odds ratios of the false positive rate for two groups because the false positive rate for a group is equivalent to the probability that a random arrestee drawn from that group will be subject to false positive. Equation 1.3 formally shows this measure below.

False Positive Odds Ratio

$$FP_{OR} = \frac{\frac{FP_{R|j=1}}{\left(1 - FP_{R|j=1}\right)}}{\frac{FP_{R|j=2}}{\left(1 - FP_{R|j=2}\right)}} \#1.3$$

Using data from the RCT to analyze false positive rate treatment effects means that any conclusions about the fairness of the PSA-DMF System were relative to the fairness of a criminal justice system without the PSA-DMF System. Equations 1.4 and 1.5 below express the test statistics used for both within-group treatment effects and between-group effects. An increase in the within-group false positive rate for any race or gender category would mean that the PSA-DMF System predicted that those individuals would experience a relevant outcome but did not in reality. Conversely, a decrease in the within-group false positive rate would mean that an individual in that group would be less likely to experience a false positive with the PSA-DMF System in place relative to a First Appearance hearing without the PSA/DMF system.

Translating hypothetical false positive rates into meaningful treatment effects is difficult because a judge is likely to implement more complex decision rules than the hypothetical thresholds discussed above. That said, higher false positive rates are most likely to be associated with a higher incidence of over-incarcerating arrestees who would otherwise not commit the relevant outcome. Significant differences in either direction away from 1 for the within-group false

---

[115] We do not use a hypothetical threshold of 1 for the NCA score because all arrestees must be scored at least this value.

positive odds ratio indicate less identical experiences by group. In other words, the null hypothesis is an odds ratio of 1.

False Positive Rate Within-group Treatment Effect

$$TE_{FP|J=j} = FP_{R|Z=1} - FP_{R|Z=0} \#1.4$$

False Positive Odds Ratio Between-Group Treatment Effect

$$TE_{FP|j\in\{1,2\}} = \frac{\frac{FP_{R|j=1,Z=1}}{\left(1 - FP_{R|j=1,Z=1}\right)}}{\frac{FP_{R|j=2,Z=1}}{\left(1 - FP_{R|j=2,Z=1}\right)}} - \frac{\frac{FP_{R|j=1,Z=0}}{\left(1 - FP_{R|j=1,Z=0}\right)}}{\frac{FP_{R|j=2,Z=0}}{\left(1 - FP_{R|j=2,Z=0}\right)}} \#1.5$$

Figure 26: False Positive Rate NCA Graphs

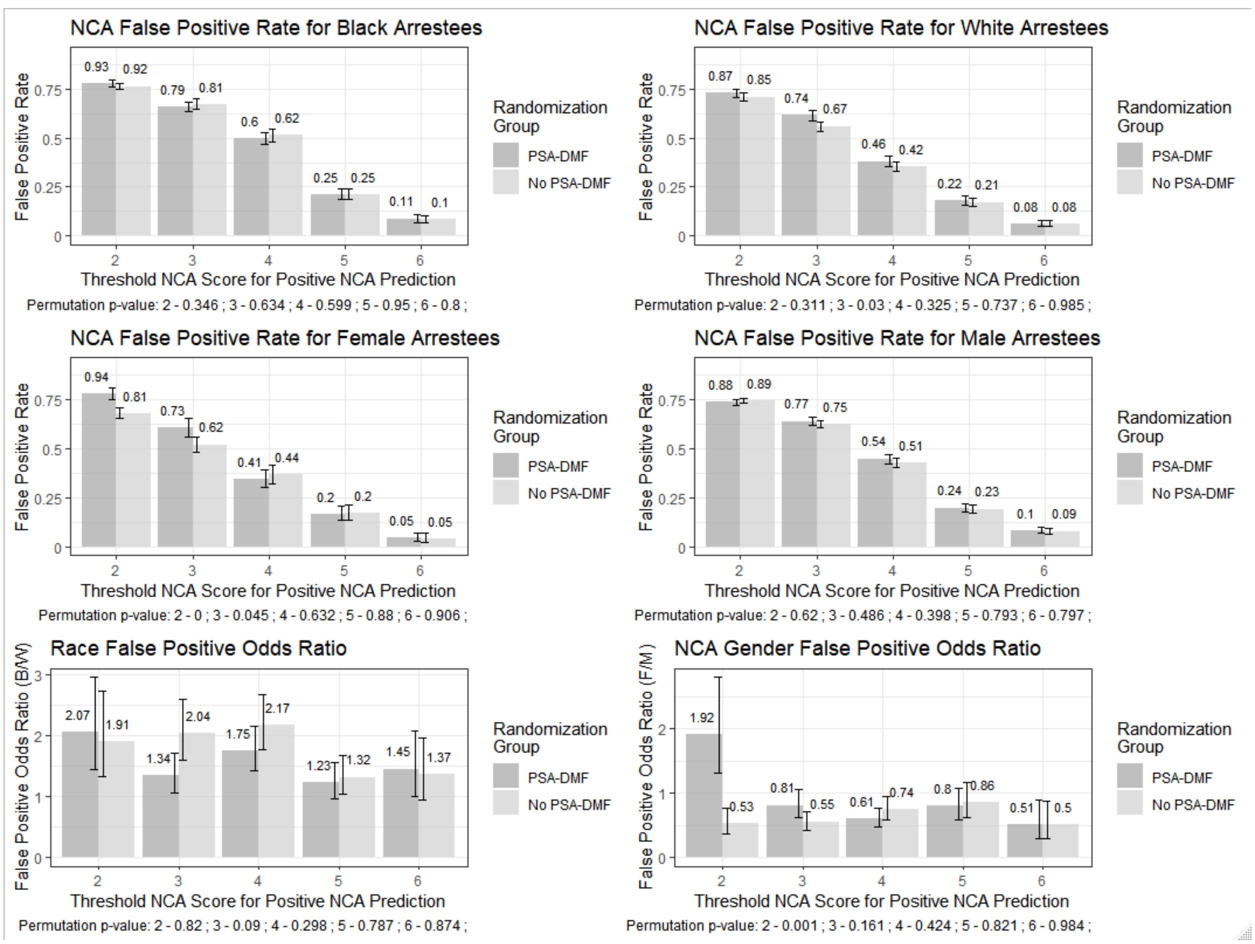

Figure 26: This series of bar graphs report the calculations for false positive rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on NCA outcomes. The top two graphs are intra group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other.

The middle two graphs repeat the intra group false positive rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false positive rates for both the racial group pairing and the gender group pairing. The top four graphs focus on calculating the treatment effect of the PSA-DMF System via a difference in false positive rates for within a specific group, while the bottom two graphs calculate potential treatment effects via false positive odds ratios between two groups. In this way, the bottom graphs can evaluate potential treatment effects on the basis of whether they move the false positive odds ratios towards an between-group parity value of 1, or away from it. The graphs report a single race based significant treatment effect (white arrestees at a hypothetical threshold of 3), and three gender based significant treatment effects (female arrestees at a hypothetical threshold of 2 and 3 and gender based false positive odds ratios at a hypothetical threshold of 2. Given the small number of significant findings and their concentration at lower levels of the NCA scale, these findings do not show a policy-relevant treatment effect.

Figure 26 reports the false positive rates for each hypothetical NCA score level alongside observed NCA outcomes. We calculated hypothetical false positive rates for black and white arrestees, for female and male arrestees, and for PSA-DMF-System-present and not-present cases at each possible threshold on the NCA scale (2-6). We repeated this process for the false positive odds ratio calculation for each race-gender pairing, creating 30 permutation tests. We observed only one statistically significant difference with respect to race among these 30 tests: with an NCA score threshold value of 3, white arrestees had significantly higher false positive rates when the PSA-DMF System was provided. Female arrestees with hypothetical thresholds set at 2 and higher or 3 and higher on the NCA scale had significantly higher false positive rates when the PSA-DMF System was present, and the false positive odds ratio measure for gender at a threshold value of 2 supports the same conclusion.

A few observations based on Figure 26: First, is that because these results were based on hypothetical constructions of potential decision rules, we expected that treatment effects would have been detected at multiple thresholds on the any outcome scale. If so, there would have been evidence robust to threshold values chosen. Second, higher threshold values should be more heavily weighted than lower threshold values because higher risk factor values were associated with more statistically significant treatment effects.[116] These considerations mean that the four significant treatment effects detected in this subsection are likely unrepresentative of the true treatment effects; they are concentrated at lower NCA risk values.

[116] See supra Subsection II.B.2.b.

Figure 27: False Positive Rate NVCA Graphs

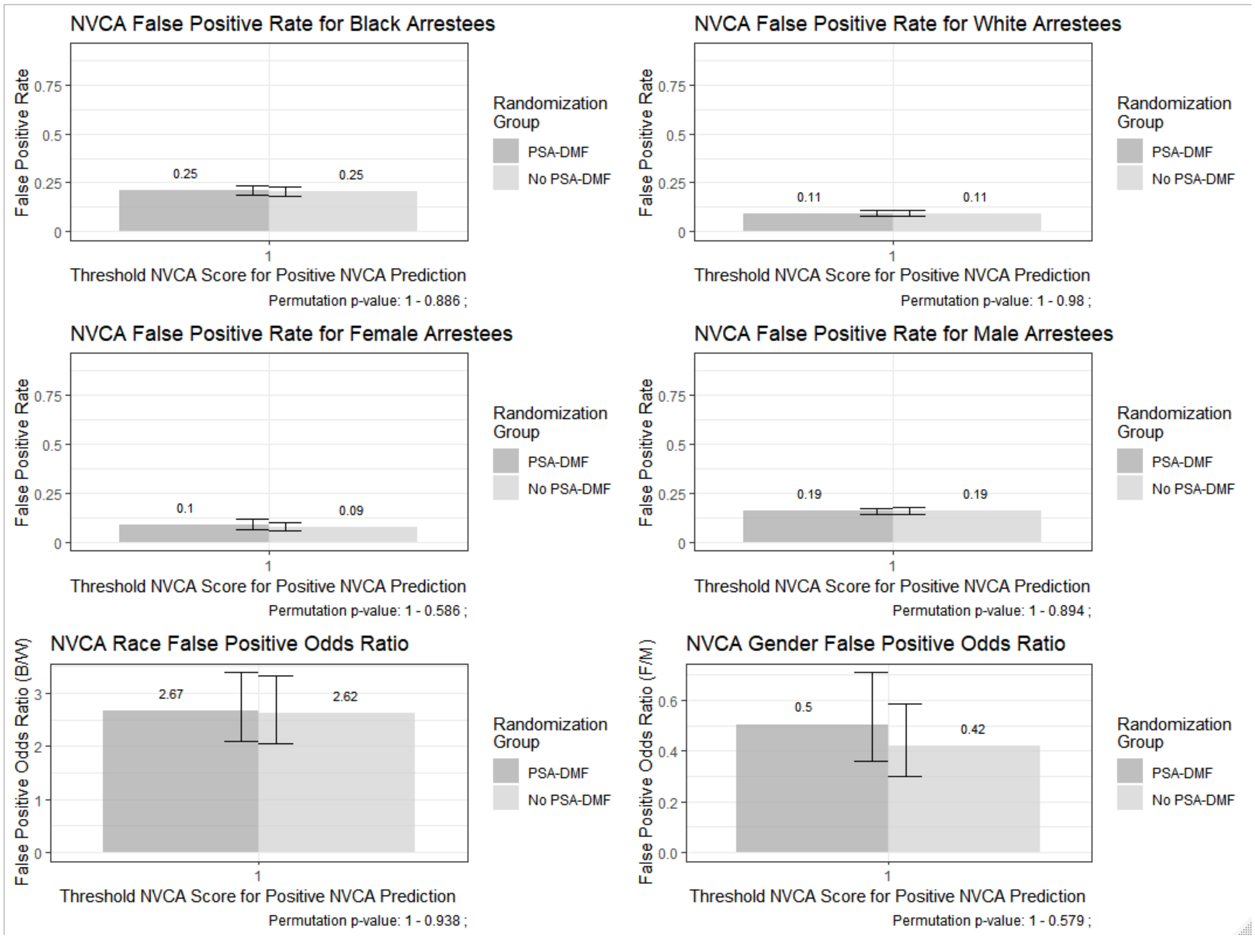

Figure 27: This series of bar graphs report the calculations for false positive rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on NVCA outcomes. The top two graphs are intra group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the intra group false positive rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false positive rates for both the racial group pairing and the gender group pairing. There are no significant treatment effects to report.

Figure 27 reports the false positive rates and odds ratios for NVCA. None of the differences are statistically significant.[117]

[117] To clarify one aspect of interpretation, we note that the PSA-DMF System does not make specific predictions about any arrestee. The presence of the NVCA Flag purports to indicate an increased risk of observing an NVCA outcome if the arrestee secures release. The false positive rate reported remains hypothetical in the sense that it reflects a judge's prediction of NVCA even if that is not how judges used or interpreted the NVCA flag.

Figure 28: False Positive Rate FTA Graphs

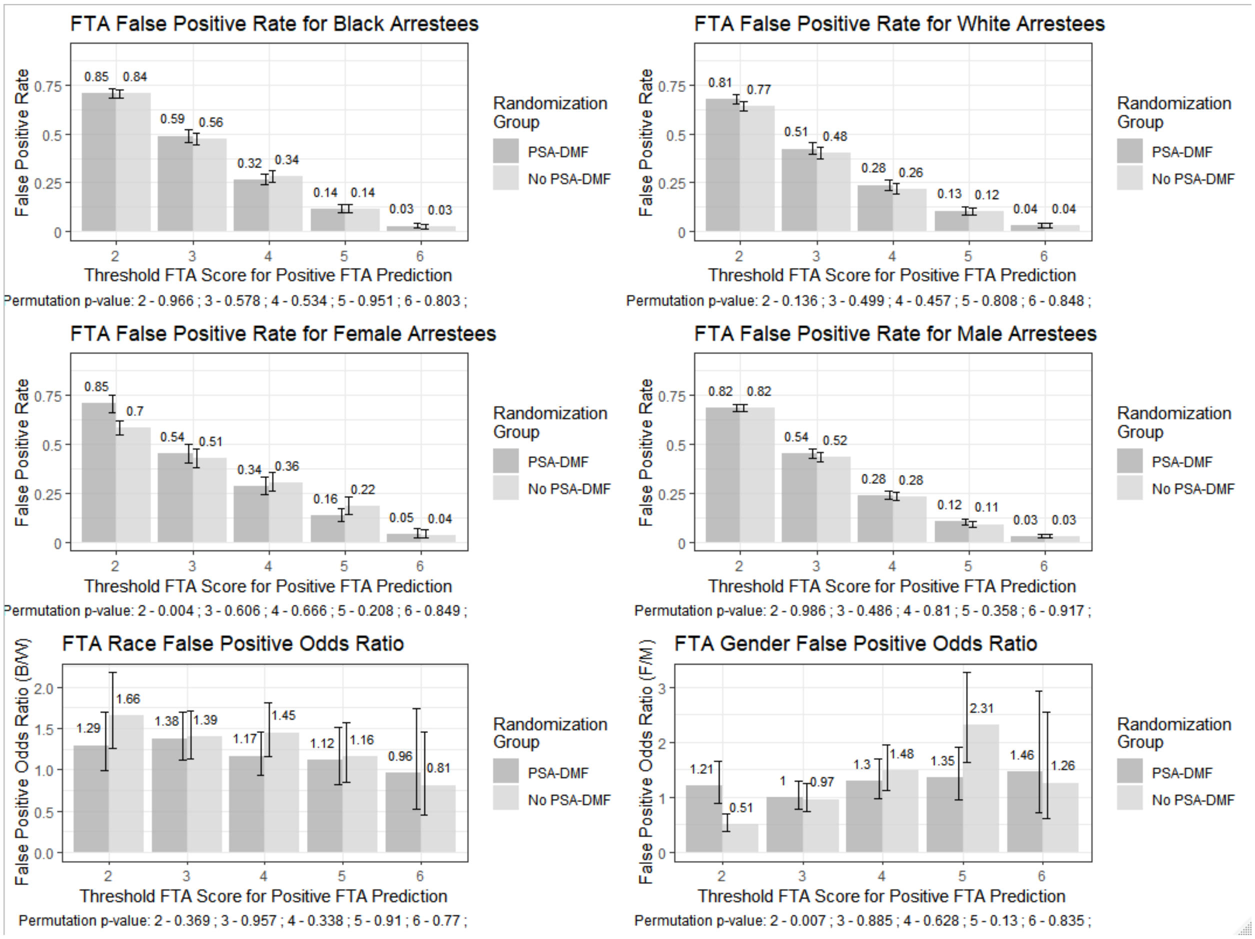

Figure 28: This series of bar graphs report the calculations for false positive rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on FTA outcomes. The top two graphs are intra group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the intra group false positive rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false positive rates for both the racial group pairing and the gender group pairing. No significant race based effects were detected, while a hypothetical threshold of 2 produced significant treatment effects for female arrestees and gender based false positive odds ratios. These findings were not robust to changes in hypothetical threshold values and therefore do not provide evidence for true intra or between-group false positive treatment effects.

Figure 28 reports the group-based false positive rates and odds ratios for FTA. The interpretation and analysis is the same as in Figure 26 for NCA false positives. And the findings for FTA are qualitatively similar to those for NCA. No significant treatment effects existed for FTA false positive rates by race. Differences by gender existed when a hypothetical FTA risk score threshold of 2 or was used. This effect likely explains the result for the false positive odds ratios by gender at the same hypothetical threshold of 2 or higher on the FTA scale. It is unlikely, however, that these findings are representative of the true treatment effect because, as with the NCA false positive analysis, treatment effects are both concentrated at the lower end of the scale (where we did not detect significant PSA-DMF System effects on from FTA scores on First Appearance outcomes) and are not robust to changes in the hypothetical threshold value.

Figures 26-28 did not provide evidence for a PSA/DMF System effect in either direction on group-based fairness using false positive rates.

(2). False Negative Rates

This subsection discusses the false negative demographic fairness measure. We begin with an explanation of this metric before providing results of comparisons focusing on race and gender.

A false negative incident occurs when a decision maker predicts that some outcome of interest will not occur, but it does in reality. For those who focus on this kind of analysis, a tool is unbiased if the relative incidence of false negatives is roughly equivalent across relevant groups. The relative incidence of false negatives is measured against the total number of observed positives cases, both those predicted to be positive (true positives) and those predicted to be negative (false negatives). This result is a false negative rate: the number of false negative observations divided by the total number of observed positive cases (true positives + false negatives). The false negative rate and the group equivalency definition are stated in equations 2.1 and 2.2, below.

False Negative Rate

$$FN_R = \frac{False\ Negatives}{True\ Positives + False\ Negatives} \#2.1$$

False Negative Equivalency

$$FN_{R|j=1} = FN_{R|j=2} \#2.2$$

False negatives differ from false positives in a number of important ways. Primarily, where false positive rates are essentially a measure of the proportion of total negative observation cases (cases where the outcome of interest was not observed) that were incorrectly predicted to observe positive instances of the outcome, false negative rates invert this. False negative rates are the proportion of all observed positive outcome cases (cases where the outcome of interest was observed in the data) that were incorrectly predicted to have negative instances of the outcome observed (where the outcome was not seen in the data). Thus, while false positive and false negative rates are both measurements of the incidence of incorrect predictions, they differ in structure in not only the numerator (false positive counts vs. false negative counts), but also in terms of their denominators (total observed negative counts vs. total observed positive counts).

A last measurement construction difference between false negative rates and false positive rates concerns the placement and direction of the hypothetical threshold values used to construct the measurements. Because false positive rates turn on the rate of false prediction for all observed negative outcomes, the hypothetical thresholds are constructed by assigning positive predictions to all values at or higher than a specific threshold value on the relevant outcome scale. False negative rates, however, work in the opposite manner. Because false negative rates measure the incidence of false prediction for all observed positive outcomes, the hypothetical thresholds are constructed by assigning a negative prediction to all cases at or below a specific threshold value on the relevant outcome scale. Thus, where false positive rates had no measurement for the threshold value of 1 since all cases we either at or higher than a value of 1 (or 0 in the case of the NVCA Flag), false negative rates have no measurement for the threshold value of 6 since all cases are at or below the value of six on the relevant outcome scale (or 1 in the case of the NVCA Flag).

The practical interpretation of the false negative rate with respect to potential arrestee experience and overall System outcomes also differs from that of the false positive rate. Whereas higher false positive rates signaled potential over-incarceration of individuals who would not commit an outcome of interest if released, higher false negative rates could signal higher observed N(V)CA or FTA because the mistake undercounted those who would experience these events if released).

The calculation and analysis of false negative rates proceeds in a similar manner as for false positive rates. We once again construct within-group and between-group treatment effects for false negatives. The formal definitions for false negative odds ratios, false negative rate within-group treatment effects, and false negative odds ratio between-group treatment effects are provided below in equations 2.3, 2.4, and 2.5, respectively.

False Negative Odds Ratio

$$FN_{OR} = \frac{\frac{FN_{R|j=1}}{\left(1 - FN_{R|j=1}\right)}}{\frac{FN_{R|j=2}}{\left(1 - FN_{R|j=2}\right)}} \#2.3$$

False Negative Rate Within-group Treatment Effect

$$TE_{FN|J=j} = FN_{R|Z=1} - FN_{R|Z=0} \#2.4$$

False Negative Odds Ratio Between-Group Treatment Effect

$$TE_{FN|j\in\{1,2\}} = \frac{\frac{FN_{R|j=1,Z=1}}{\left(1-FN_{R|j=1,Z=1}\right)}}{\frac{FN_{R|j=2,Z=1}}{\left(1-FN_{R|j=2,Z=1}\right)}} - \frac{\frac{FN_{R|j=1,Z=0}}{\left(1-FN_{R|j=1,Z=0}\right)}}{\frac{FN_{R|j=2,Z=0}}{\left(1-FN_{R|j=2,Z=0}\right)}} \#2.5$$

Interpretation of any group differences in these measurements is identical to that for false positive calculations: An increase in the within-group false negative rate for any in the System-present condition relative to System-not-present condition meant that the PSA-DMF System's provision led to "predictions" that they would not experience a relevant outcome but did in reality. A decrease conversely meant that an individual in a group had a lower risk of a false negative under the PSA-DMF System than without it. We interpret odds ratios greater than 1 as positive treatment effects and odds ratios less than 1 as negative. As with false positives, white arrestees and male arrestees are the baseline groups.

Figure 29: False Negative Rate NCA Graphs

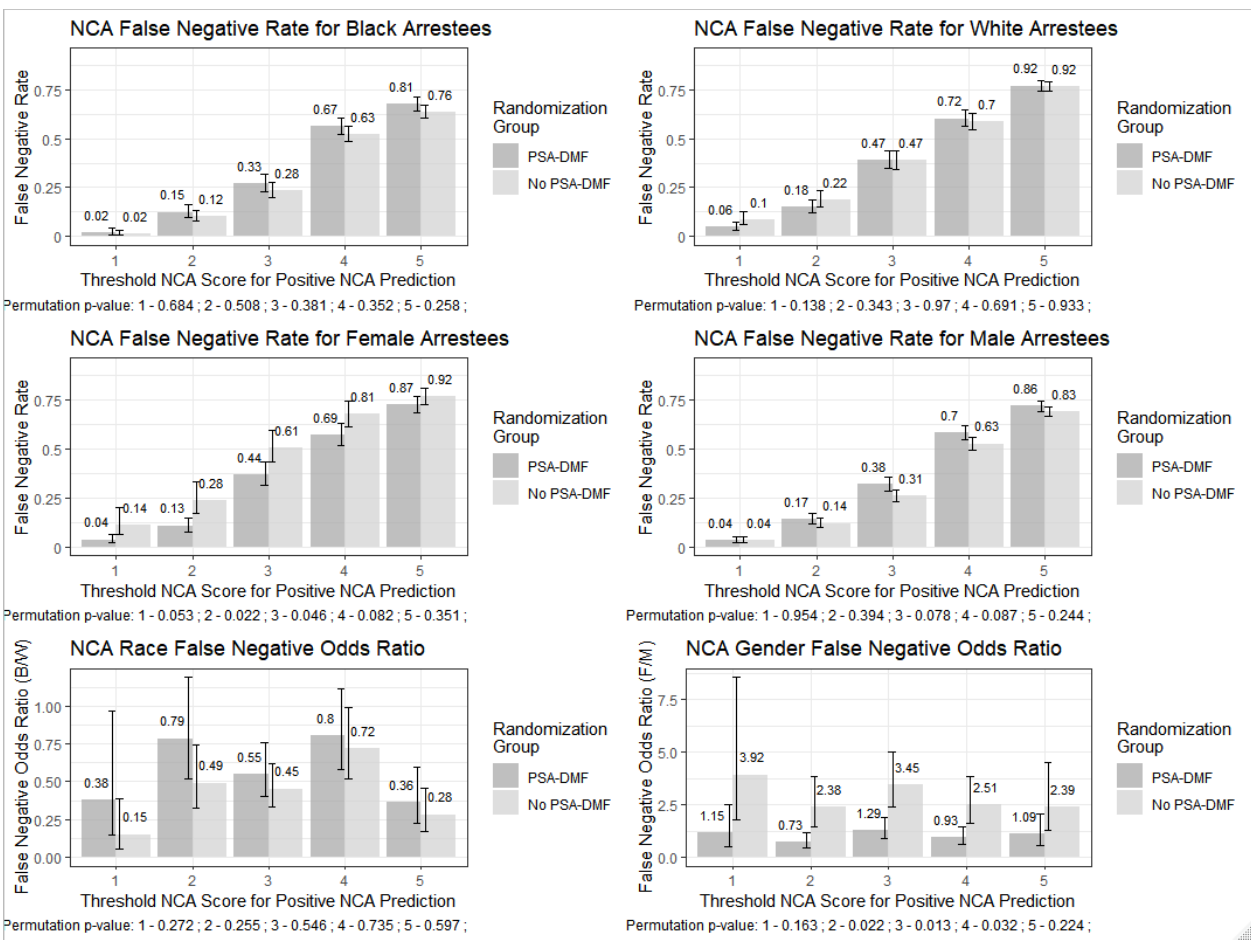

Figure 29: This series of bar graphs report the calculations for false negative rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact

of the PSA-DMF System on NCA outcomes. The top two graphs are intra group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the intra group false negative rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false negative rates for both the racial group pairing and the gender group pairing. The top four graphs focus on calculating the treatment effect of the PSA-DMF System via a difference in false negative rates for within a specific group, while the bottom two graphs calculate potential treatment effects via false negative odds ratios between two groups. There is no significant race based treatment effects to report. Significant gender based treatment exist at hypothetical thresholds of 2 and 3 for female arrestees and values of 2, 3, and 4 for gender based false negative odds ratios. The observed significant treatment effects push the gender based false negative rates closer to the parity value of 1, and thus do not constitute a disparate impact.

Figure 29 reports the relationship between the PSA-DMF System and group-specific false negative rates and odds ratios for NCA. There are no statistically significant treatment effects for race comparison. There are several statistically significant treatment effects among the gender-based false negative rates. At thresholds of 3 or lower, female arrestees were less likely to experience NCA false negatives with the PSA-DMF System printout provided then without it. The treatment effect sizes are -0.17, -0.15, and -0.10 with associated 95% confidence intervals of -0.33 to -0.01, -0.28 to -0.02, and -0.19 to -0.01, for threshold values of 3 or lower, 2 or lower, and 1, respectively. The same statistically significant treatment effects do not exist among examining male arrestees, which potentially indicates a disparate impact. Yet, when examining the gender-based between-group false negative odds ratios, male and female experiences with false negatives tends closer to equity. Gender-based between-group false negative odds ratios show statistically significant treatment effects for hypothetical NCA Score values of 4, 3, and 2 or lower. For each of these threshold values, the PSA-DMF-System-not-present group had false negative rates significantly higher than 1 (indicated by the confidence interval bars on the graph), and the statistically significant negative treatment effect of the PSA-DMF System pushes these odds ratios closer to 1. These effect sizes for the threshold values of 4, 3, and 2, or lower are -1.46, -2.16, and -1.65, and exist on the related 95% confidence intervals of -2.84 to -0.08, -3.79 to -0.53, and -3.03 to -0.27, respectively, all of which include the group parity value of 1. Thus, Figure 29 does not provide evidence that the PSA-DMF System disparately impacts arrestees by gender.

Figure 30: False Negative Rate NVCA Graphs

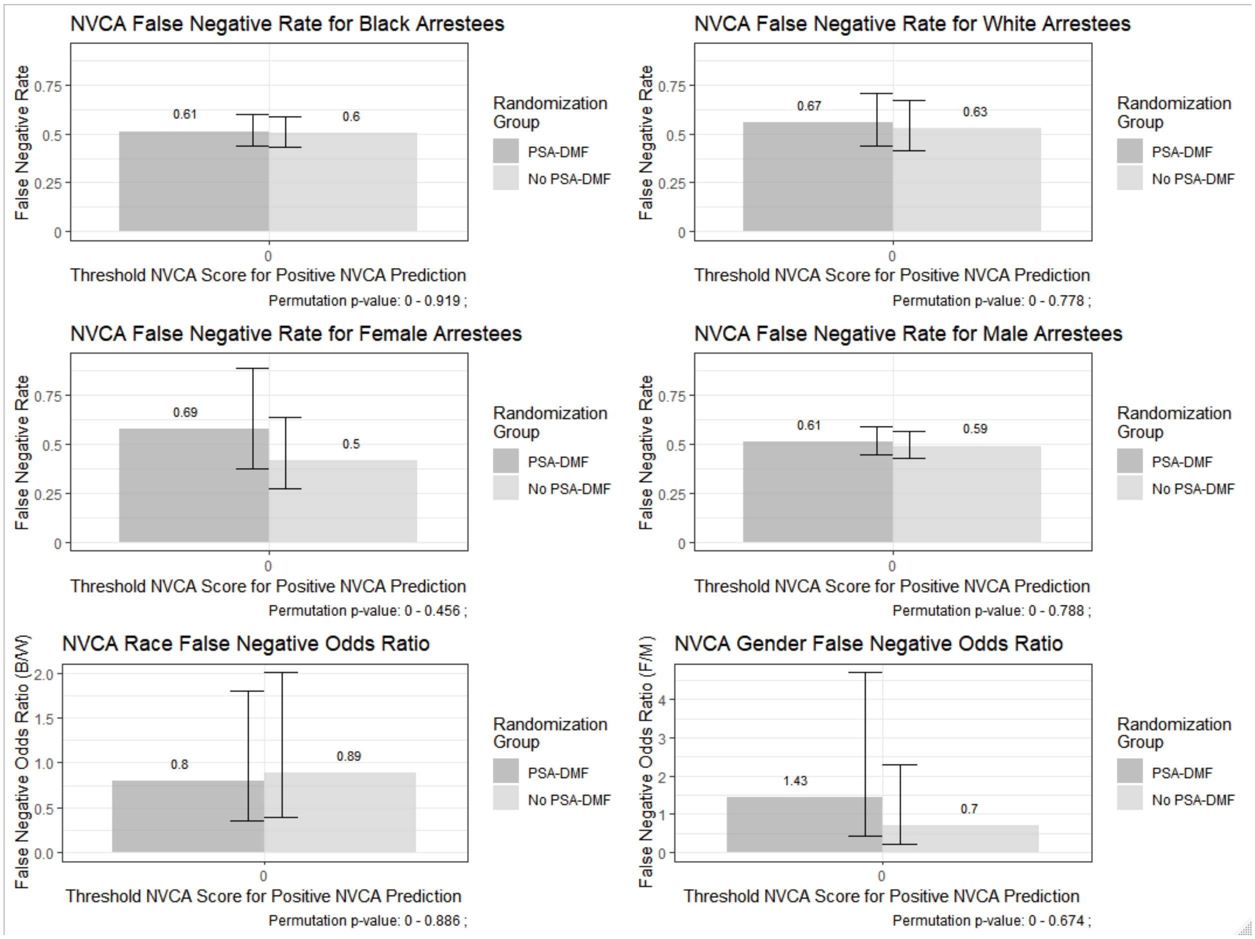

Figure 30: This series of bar graphs report the calculations for false negative rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on NVCA outcomes. The top two graphs are within-group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the within-group false negative rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false negative rates for both the racial group pairing and the gender group pairing. There are no significant treatment effects to report.

Figure 30 reports group-based false negative rates and odds ratios for NVCA. The analysis and findings are qualitatively similar to the prior subsection on false positive rates, with the exception of only one hypothetical threshold. As with the false positive calculations, no statistically significant group-based treatment effects emerge. As a result, there is no evidence to support any PSA-DMF System effect on the fairness of criminal justice outcomes when measured using group-based false negative rates.

Figure 31: False Negative Rate FTA Graphs

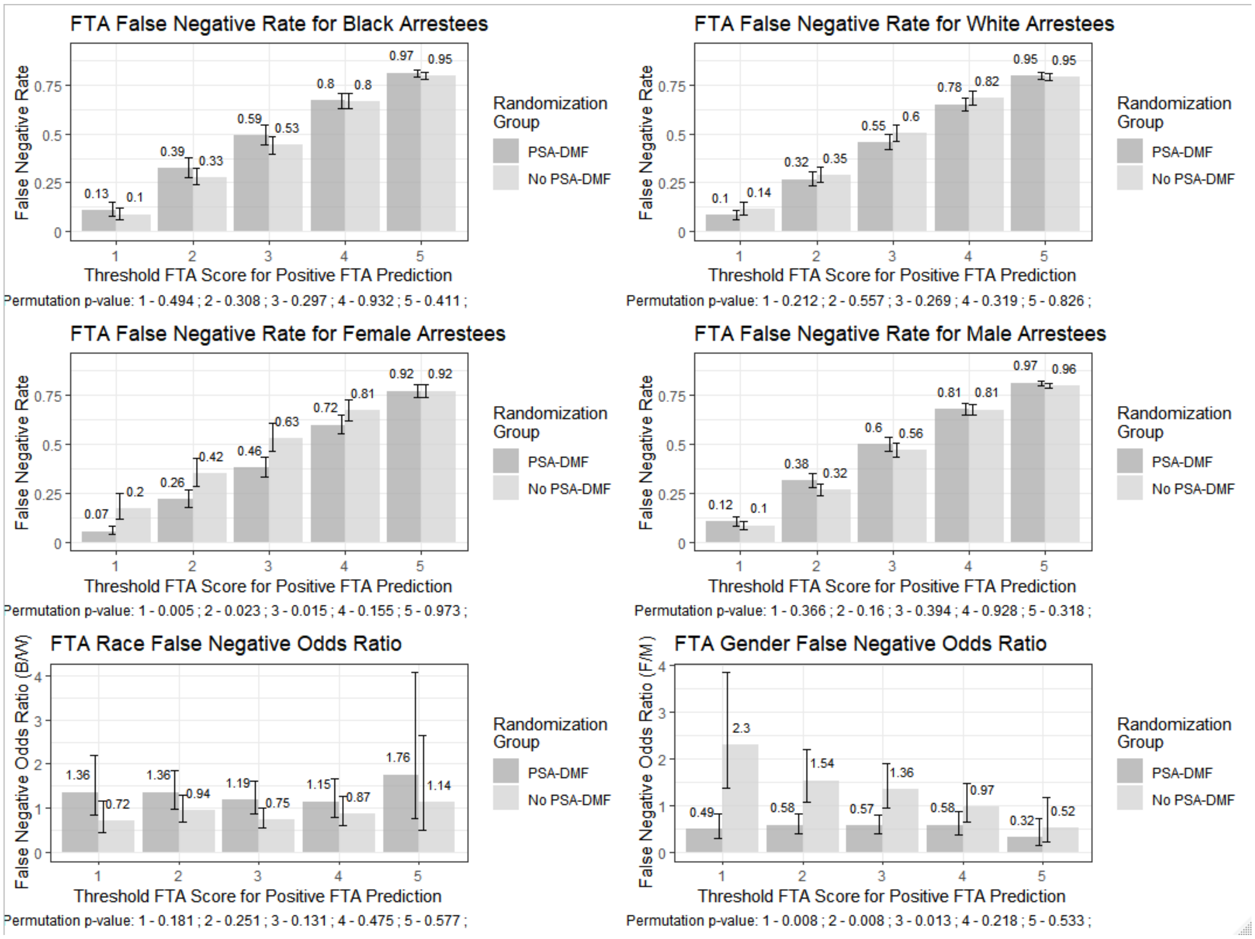

Figure 31: This series of bar graphs report the calculations for false negative rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on FTA outcomes. The top two graphs are within-group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the within-group false negative rate metric for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on false negative rates for both the racial group pairing and the gender group pairing. There are no significant race based treatment effects to report. Significant gender based treatment exist at hypothetical thresholds of 2 and 3 for female arrestees and values of 1, 2, and 3 for gender based false negative odds ratios. The observed significant treatment effects push the gender based false negative rates closer to then away from the parity value of 1, thereby inverting potential gender based disparate impacts.

Figure 31 shows the group-based false negative calculations for FTA. The findings here were roughly analogous to the earlier findings on NCA false negative calculations, with no race-related statistically significant treatment effects or for male arrestees. Statistically significant treatment effects were detected for female arrestees and gender-based between-group false negative odds ratios. Specifically, significant negative treatment effects at theoretical FTA score threshold values of 3 or lower, 2 or lower, and 1 emerged for female arrestees, with estimated treatment effect sizes of -0.17, -0.16, and -0.13 and relevant 95% confidence intervals of -0.31 to -0.03, -0.30 to -0.02, and -0.23 to -0.03, respectively. For gender-based false negative odds ratios, significant, negative treatment effects existed at hypothetical FTA score thresholds of 3 or lower, 2 or lower, and 1. Unlike the NCA calculations above, however, none of the FTA gender-based between-group false negative odds ratios significantly differed from 1.

Figures 29-31 and the related PSA-DMF System treatment effect calculations do not provide any evidence of race-related impacts on fairness as measured by group-based comparisons of false negative rates. Some evidence, however, supports gender-based impacts on fairness as measured by group-based comparisons of false negative rates. For NCA false negatives, the PSA-DMF System seemed to equalize experiences relative to a world without the System. With respect to FTA false negatives, there was evidence that the System inverted the disparate gender-related impacts of false negative rates relative to a world without the System.

(3). Overall Procedure Error Rates

This subsection discusses the overall procedure error demographic fairness measure. We begin with an explanation of this metric before providing results of comparisons focusing on race and gender.

False positive and false negative rates are inversely related. The number of true positives and false negatives must equal the total number of observed positives, and the number of true negatives and false positives must equal the total number of observed negatives. Thus, any attempt to decrease the number of false positives (which will increase the number of true negatives) will also increase the number of false negatives (and decrease the number of true positives). With this tension, we can map the two measures onto different criminal justice values. Attempts to lower false positive rates are associated with seeking less incarceration; higher false positive rates likely increase the over-incarceration of arrestees who would not have committed new N(V)CAs or FTAs if released. Similarly, attempts to lower the false negative rate are associated with higher carceral rates; higher false negative rates could lead to higher incidence of N(V)CA or FTA because higher-risk arrestees would be more likely to be mistakenly released. This value mapping is complicated by the fact that lowering false negative rates might also correspond to valuing less incarceration rates. Lowering false negatives might decrease the probability that a released arrestee will misbehave after release and thus spend less time incarcerated Likewise, aiming to reduce false positive rates can also promote public safety if it protects arrestees from crimes committed in correctional facilities. The tradeoff between false positive and false negative rates has generated alternative measurements that do not value one rate over the other, or at least value them equally.

Berk et al. proposed one such measurement that combines false positive and false negative rates, known as “overall procedure error.”[118] This measurement calculates the total instances of

[118] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, & Aaron Roth. Fairness in Criminal Justice Risk Assessments: The State of the Art, Soc. Meth. & Res. (2018).

incorrect prediction (both false positives and false negatives) and divides that sum by the total number of observed cases. The overall procedure error provides equal weighting to both false positives and false negatives by pooling them into a common failed prediction measure. These failed predictions are then expressed as a share of the overall cases, producing what is essentially a failed prediction rate. Under this approach, a predictive tool is fair if the relative frequency of incorrect predictions is roughly equivalent across groups. The formal definition of the overall procedure error rate and the overall procedure error rate group equivalency definition are expressed in equations 3.1 and 3.2, respectively, below.

Overall Procedure Error Rate

$$OPE_R = \frac{False\ Positives + False\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives} \#3.1$$

Overall Procedure Error Rate Group Equivalency

$$OPE_{R|j=1} = OPE_{R|j=2} \#3.2$$

The calculation of the overall procedure error rate follows the same pattern as false positive and false negative rates: we choose hypothetical threshold values for a relevant outcome such that cases at or above those values are considered positive predictions for the outcome of interest while values below that threshold are considered negative predictions for the same outcome. Unlike false positive and false negative rates, however, overall procedure error rates can be calculated for all values on a scale. Consider a hypothetical threshold value of 1 on the NCA scale. Any case with a NCA score of at least 1 will represent a positive prediction for NCA, but all cases have an NCA score of at least 1. Therefore, true negatives cannot be observed because there are no negative predictions; the false positive rate given in equation 3.1 reduces to the number of false positives divided by the number of false positives, which is 1. But the overall procedure rate, because it includes all incorrect predictions, can be calculated. In this example, we will observe false positives (predicted positive cases with observed negative outcomes) but no true or false negatives (because there are no negative predictions). The overall procedure error rate reduces to the number of false positives divided by the number of all positive predictions (true positives and false positives). Because there are only positive predictions in this example, the result provides the proportion of incorrect predictions. Thus, the overall procedure error rate allows calculations at each possible threshold value for a relevant outcome scale.

There are benefits and drawbacks to using overall procedure error rates rather than false positive and false negative rates. One drawback is that overall procedure error rates, without more information, preclude analysis of false positives and false negatives because they pool both errors into a singular metric. On the other hand, interpreting overall procedure rates is relatively easier than interpreting either false positive or false negative rates. Increases in the overall procedure error rate increase the likelihood that a predictive tool will generate some error. If the goal of risk assessment algorithms is to help judges correctly distinguish high-risk and low-risk arrestees, then any positive overall procedure error indicates the tool may not be working as well as might be hoped.

As with false positives and false negatives, we use comparative within-group and between-group treatment effects. Within-group treatment effects use the overall procedure error rates, and between-group treatment effects use overall procedure error odds ratios. The formal definitions for overall procedure error odds ratios, overall procedure error rate within-group treatment effects, and overall procedure error odds ratio between-group treatment effects are given below in equations 3.3, 3.4, and 3.5, respectively.

Overall Procedure Error Odds Ratio

$$OPE_{OR} = \frac{\frac{OPE_{R|j=1}}{\left(1 - OPE_{R|j=1}\right)}}{\frac{OPE_{R|j=2}}{\left(1 - OPE_{R|j=2}\right)}} \#3.3$$

Overall Procedure Error Rate Within-group Treatment Effect

$$TE_{OPE|j=j} = OPE_{R|Z=1} - OPE_{R|Z=0} \#3.4$$

Overall Procedure Error Odds Ratio Between-Group Treatment Effect

$$TE_{OPE|j\in\{1,2\}} = \frac{\frac{OPE_{R|j=1,Z=1}}{\left(1 - OPE_{R|j=1,Z=1}\right)}}{\frac{OPE_{R|j=2,Z=1}}{\left(1 - OPE_{R|j=2,Z=1}\right)}} - \frac{\frac{OPE_{R|j=1,Z=0}}{\left(1 - OPE_{R|j=1,Z=0}\right)}}{\frac{OPE_{R|j=2,Z=0}}{\left(1 - OPE_{R|j=2,Z=0}\right)}} \#3.5$$

A positive difference in within-group overall procedure error rates between the PSA-DMF System-present and not-present conditions mean that the PSA-DMF System printout caused individuals to experience a greater risk of misclassification. And statistically significant differences in either direction away from 1 indicate significant differences between race or gender groups. Values greater than 1 indicate a possible disparate impact against black and female arrestees, and values less than 1 indicate a possible disparate impact against white and male arrestees.

Figure 32: Overall Procedure Error NCA Graphs

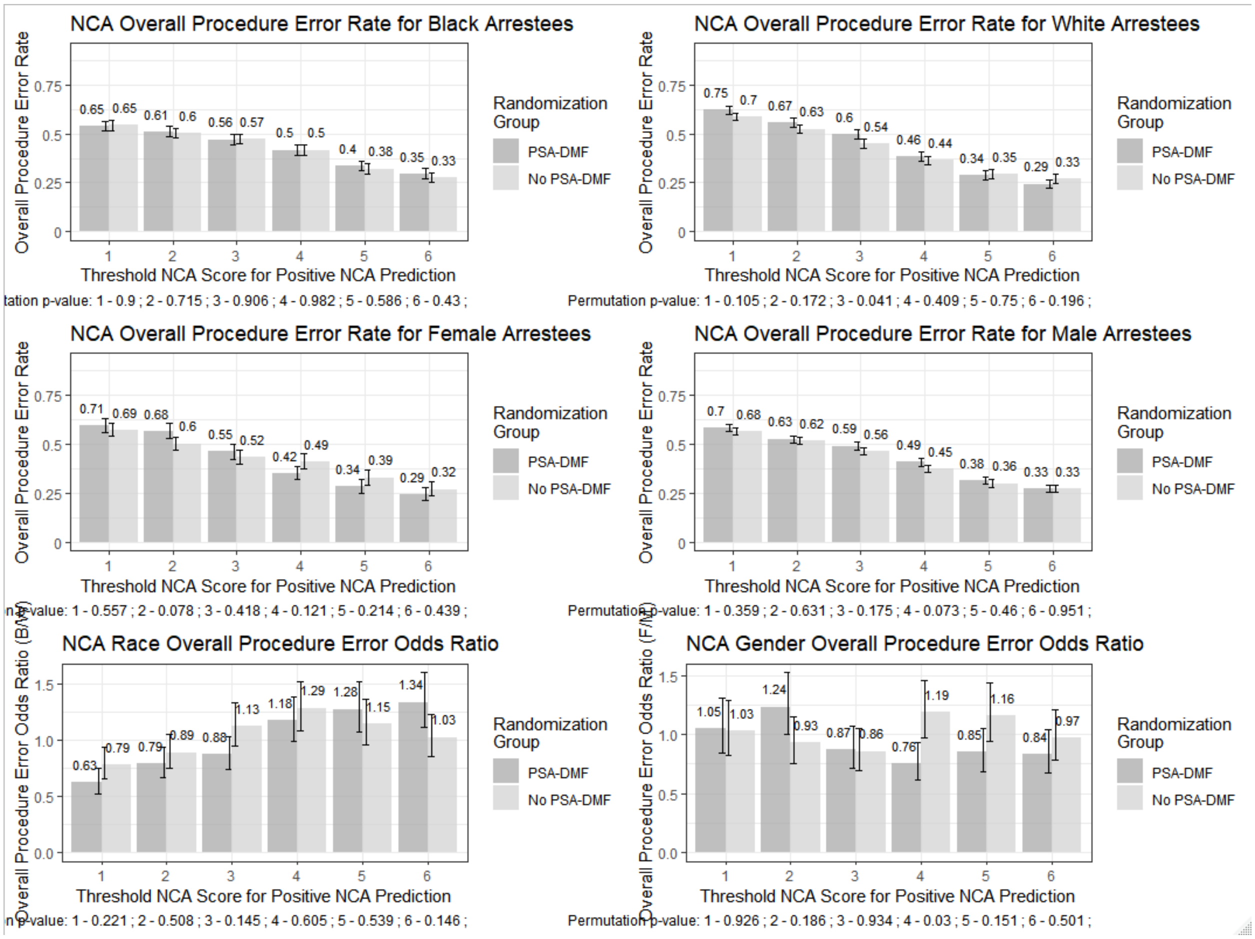

Figure 32: This series of bar graphs report the calculations for overall procedure error rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on NCA outcomes. The top two graphs are within-group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the within-group overall procedure error rate metrics for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on overall procedure error odds ratios for both the racial group pairing and the gender group pairing. The top four graphs focus on calculating the treatment effect of the PSA-DMF System via a difference in overall procedure error rates for within a specific group, while the bottom two graphs calculate potential treatment effects via overall procedure error odds ratios between two groups. There is a single significant raced based finding for a hypothetical threshold of 3 for white arrestees and a single significant gender based finding for a hypothetical threshold value of 4 for an overall procedure error odds ratios. The lack of over confirmatory significant treatment effects indicates that these do not provide support for potential disparate impact treatment effects.

Figure 32 reports the overall procedure error calculations for NCA. Significant differences in the effect sizes based on the permutation tests indicate a significant treatment effect from providing the PSA-DMF System printout at First Appearance hearings. The overall procedure error calculations for NCA are comprised of 36 separate permutation tests across the two randomization groups: PSA-DMF System-present and not-present. Of those 36 tests, only one, gender-based overall procedure error between-group odds ratios, yields a statistically significant treatment effect when the hypothetical threshold value for a positive NCA prediction is set to 4 or higher. The effect size is a -0.43 difference in the odds ratio in the System-present condition, which exists on a 95% confidence interval of -0.82 to -0.04. As with NCA false positives above, this treatment effect does not emerge at other levels of the hypothetical threshold value, which suggests that the effect is not robust to changes in that value. As a result, this finding is unlikely to be reflective of the true population treatment effect.

Figure 33: Overall Procedure Error NVCA Graphs

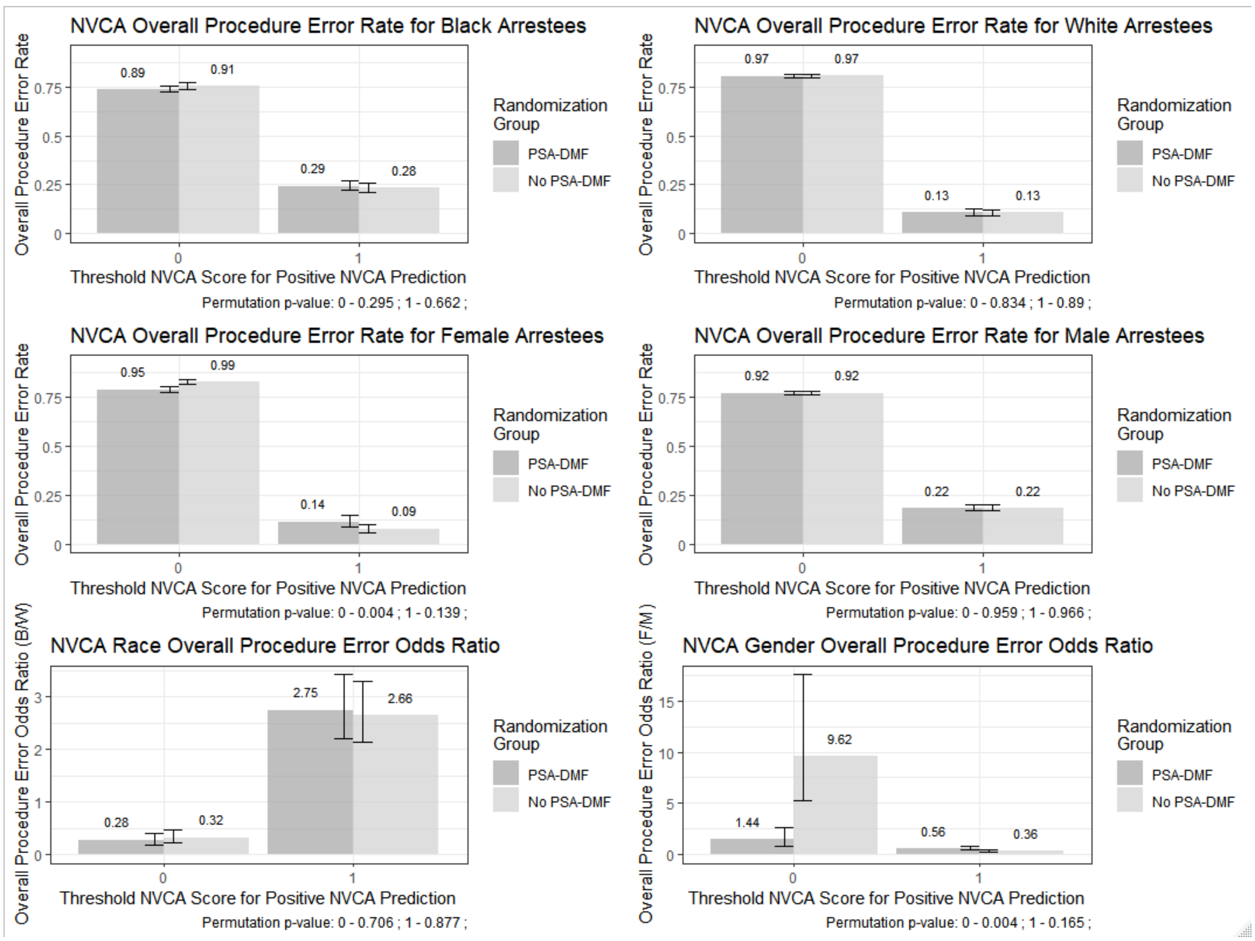

Figure 33: This series of bar graphs report the calculations for overall procedure error rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on NVCA outcomes. The top two graphs are within-group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the within-group overall procedure error rate metrics for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on overall procedure error odds ratios for

both the racial group pairing and the gender group pairing. There are two significant treatment effects detected, both gender based and both at a hypothetical threshold of 0 for the presence of an NVCA Flag. These treatment effects move the overall procedure rate closer to parity, however, given the paucity of observed NVCA instances in the data and the fact that a hypothetical threshold of 0 would give all arrestees a positive prediction, the significant treatment effects here make little practical sense.

Figure 33 reports the overall procedure error calculations for the NVCA flag. In this analysis, the two calculations essentially show the difference in hypothetical value thresholds when categorizing all arrestees as positive predictions vs. just arrestees with the violence flag present. As a result, significant values generated only for setting the hypothetical value threshold to capture all arrestees should be considered with skepticism because the relatively few instances of observed NVCA outcomes (see Figure 20) means a hypothetical rule based on categorizing all arrestees as positive predictions for NVCA is a poor fit for the data. That said, with respect to race-related NVCA overall procedure error calculations, there are no significant treatment effects to report. For gender-related NVCA overall procedure error calculations, two significant treatment effects, one for a hypothetical threshold value of all female arrestees, and another for the same threshold for gender based overall procedure error between-group odds ratios. In both situations, the treatment effect is negative, meaning providing the PSA-DMF System printout at First Appearance hearings significantly lowers the NVCA overall procedure error rate for female arrestees and reduces between-group disparity for NVCA overall procedure error. As previously stated, however, the fact that these significant treatment effects only exist for unrealistic hypothetical threshold values means they are unlikely to be reflective of the true population treatment effect. We do not consider them to be policy-relevant.

Figure 34: Overall Procedure Error FTA Graphs

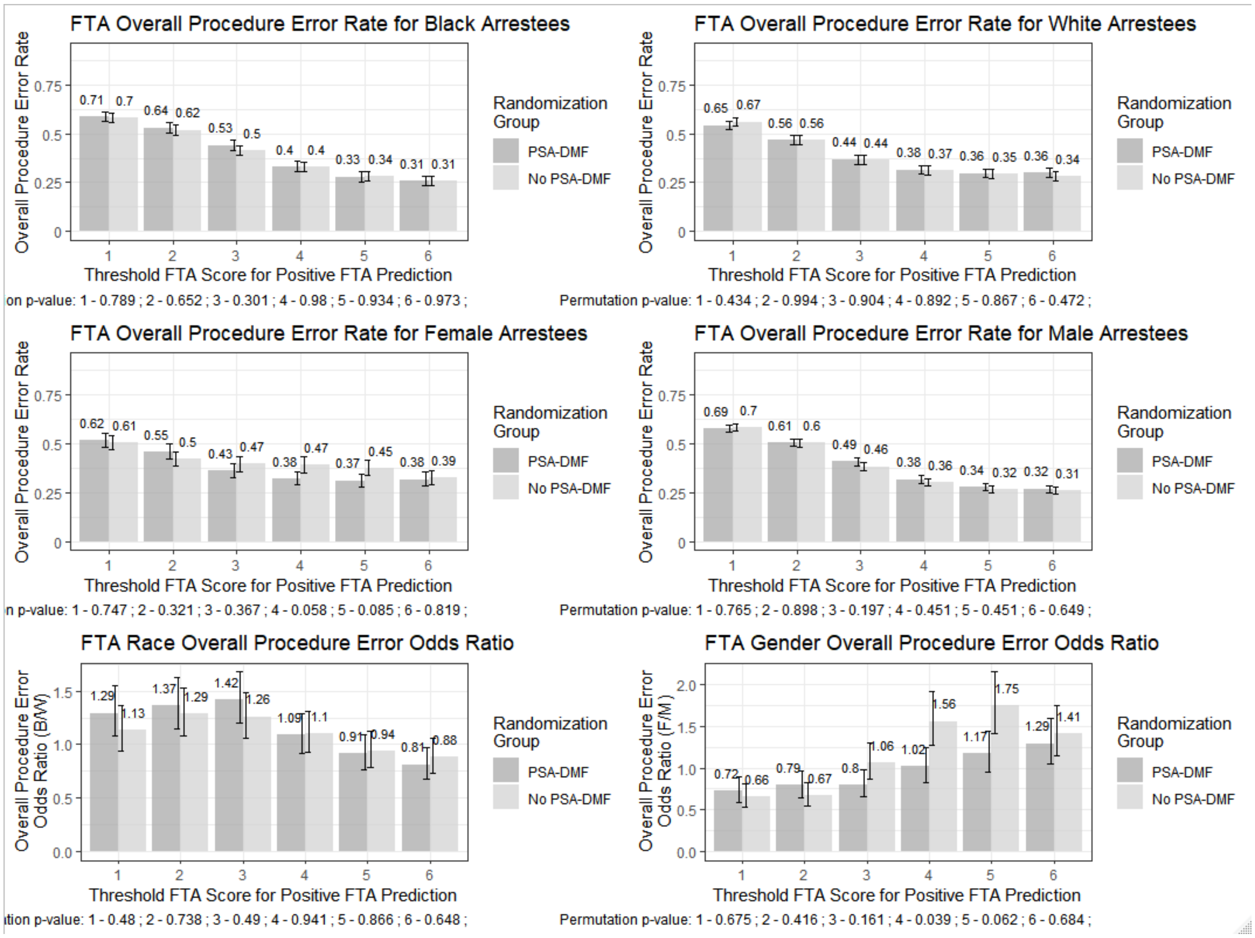

Figure 34: This series of bar graphs report the calculations for overall procedure error rates constructed with reference to measuring the fairness impacts on the accuracy and potential disparate impact of the PSA-DMF System on FTA outcomes. The top two graphs are within-group racial metrics meant to ascertain potential treatment effects for one group that are not observed on the other. The middle two graphs repeat the within-group overall procedure error rate metrics for a gender variable pair. The bottom two graphs provide direct between-group assessments of the effect of the PSA-DMF System on overall procedure error odds ratios for both the racial group pairing and the gender group pairing. There is a single observed significant treatment effect for gender based overall procedure error odds ratios for a hypothetical FTA threshold of 4. Given the lack of other confirmatory treatment effects, this does not provide support for disparate impacts of the PSA-DMF System on overall procedure error rate based metrics.

Figure 34 reports overall procedure error calculations for FTA. Of the 36 permutation tests conducted, only one achieved significance: gender-related overall procedure error between-group odds ratios at a hypothetical FTA score of 4 or higher. The reported effect size was -0.54 with a confidence interval of -1.06 to -0.02, which brings the odds ratio closer to the parity score of 1. This treatment effect does not appear at other levels of the hypothetical threshold value, indicating that it is not robust to changes in that value. Thus, this finding is unlikely to be reflective of the true population treatment effect.

The prior three figures consist of 84 separate permutation tests across three different outcomes in an attempt to identify potential group-based fairness impacts as measured by changes in the overall procedure error. Although a few significant treatment effects were observed, none was either robust to changes in the hypothetical threshold value or based on reasonable hypothetical prediction rules. The findings do not support a conclusion of significant group-based fairness impacts, as measured by overall procedure error, under the PSA-DMF System.

(4). Risk Factor Separated Outcome Based Disparate Impacts

This subsection discusses the disparate impact demographic fairness measure. We begin with an explanation of this metric before providing results of comparisons focusing on race and gender.

The prior three group-based fairness metrics are based on diagnostic measurements of the predictions generated by the PSA-DMF System: false positive rates, false negative rates, and overall procedure errors. They use different hypothetical threshold values associated with predictions rules, but some of these hypothetical values are more sensible than others. These three metrics frame fairness in terms of differential classification when the predictive tool is available versus not available. Our final alternative uses observed outcomes rather than classification accuracy to measure fairness.

This version of a fairness measure focuses on whether individuals who had the same risk scores but come from different groups had similar incident rates for the outcomes of interest. As in the prior subsections, we use comparative, group-based fairness calculations to compare the presence of the PSA-DMF System to its absence. We examined five outcomes across risk score values for potential disparate impact: the incident rate of signature bonds at first appearance, average bond amount at first appearance, observed NCA incidence, observed NVCA incidence, and observed FTA incidence.

The formal statement of the mean outcome value and mean outcome group equivalency definition are stated in equations 4.1 and 4.2 below.

Mean Outcome Value

$$\underline{Y_k} = \frac{\sum_{i=1}^{n} \quad k_i}{n} \#4.1$$

Mean Outcome Group Equivalency

$$\underline{Y}_{k|j=1} = \underline{Y}_{k|j=2} \#4.2$$

Incorporating the relevant risk scores assures that similar individuals across different groups are compared. Between-group comparisons are made through either mean outcome group ratios, for non-probabilistic outcomes, or outcome group odds ratios for probabilistic outcomes. We then have the score-separated mean outcome group equivalency definition, score-separated mean outcome group ratio, and score-separated outcome group odds ratio in equations 4.3, 4.4, and 4.6, below.

Score-Separated Mean Outcome Group Equivalency

$$\underline{Y}_{k|S=s,j=1} = \underline{Y}_{k|S=s,j=2} \#4.3$$

Score-Separated Mean Outcome Group Ratio

$$R_{Y_k} = \frac{\underline{Y}_{k|S=s,j=1}}{\underline{Y}_{k|S=s,j=2}} \#4.4$$

Score-Separated Outcome Group Odds Ratio

$$OR_{Y_k} = \frac{\frac{Pr\ Pr\ (S = s, j = 1)}{(1 - Pr\ Pr\ (S = s, j = 1)\ )}}{\frac{Pr\ Pr\ (S = s, j = 2)}{(1 - Pr\ Pr\ (S = s, j = 2)\ )}} \#4.5$$

Finally, as above, we use a comparative framework that generates an attributable group-based fairness impact with and without the PSA-DMF system. The results are score-separated mean outcome within-group treatment effect, the score-separated mean outcome between-group ratio treatment effect, and the score-separated mean outcome between-group odds ratio treatment effects reported in equations 4.6, 4.7, and 4.8, below.

Score-Separated Mean Outcome Within-group Treatment Effect

$$TE_{Y_k} = \underline{Y}_{k|S=s,J=j,Z=1} - \underline{Y}_{k|S=s,J=j,Z=0} \#4.6$$

Score-Separated Mean Outcome Between-Group Ratio Treatment Effect

$$TER_{Y_k} = \frac{\underline{Y}_{k|S=s,j=1,Z=1}}{\underline{Y}_{k|S=s,j=2,Z=1}} - \frac{\underline{Y}_{k|S=s,j=1,Z=0}}{\underline{Y}_{k|S=s,j=2,Z=0}} \#4.7$$

Score-Separated Mean Outcome Between-Group Odds Ratio Treatment Effect

$$TEOR_{Y_k} = \frac{\frac{Pr\ Pr\ (S = s, j = 1, Z = 1)}{(1 - Pr\ Pr\ (S = s, j = 1, Z = 1)\ )}}{\frac{Pr\ Pr\ (S = s, j = 2, Z = 1)}{(1 - Pr\ Pr\ (S = s, j = 2, Z = 1)\ )}} - \frac{\frac{Pr\ Pr\ (S = s, j = 1, Z = 0)}{(1 - Pr\ Pr\ (S = s, j = 1, Z = 0)\ )}}{\frac{Pr\ Pr\ (S = s, j = 2, Z = 0)}{(1 - Pr\ Pr\ (S = s, j = 2, Z = 0)\ )}} \#4.8$$

Interpreting these metrics is more straightforward than the diagnostic-based metrics discussed in the prior three sections. Significant treatment effects denote attributable positive or negative impacts on relevant score-separated group outcome means. Practically speaking, for similarly situated (in terms of risk score) individuals across demographic groups, significant treatment effects indicated potential disparate impacts of the tool, in the case of moving comparison groups away from parity, or potential increases in the fairness of the bail-setting process, in the case of moving comparison groups towards parity. In the case of between-group comparisons, significant treatment effects mean that arrestees in the group with some risk score experience different levels of the outcome when the PSA-DMF System printout is provided relative to when it is not. This metric allows us to gauge potential disparate impacts by measuring each group-risk score pairing separately and determining the existence and parity consequences of any differential impacts. We can then utilize the between-group treatment effect measures to make direct comparisons for various risk score levels. Between-group ratios are used for the First Appearance outcomes of probability of surety bond and average bail amount. Odds ratios are not used because average bail amounts are not probabilities and thus not transformable into odds ratios. Between-group odds ratios are used for the main outcome variables of NCA, NVCA, and FTA.

Figure 35: DMF Recommendation-Separated Signature Bond Rate Graphs

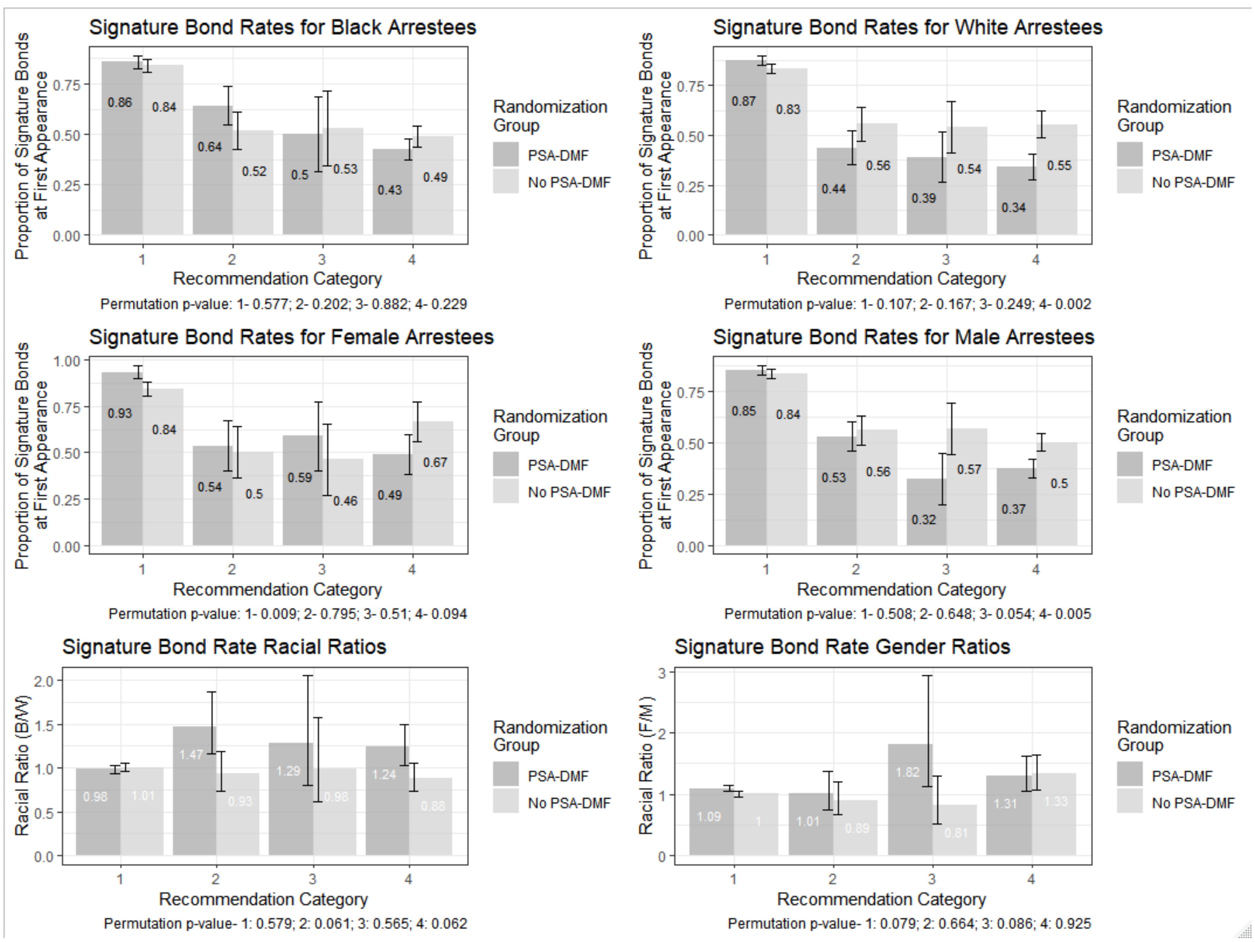

Figure 35: These bar plots report DMF Recommendation level separated mean outcome differences for groups for the proportion of signature bond incident rates. The top two graphs focus on identifying potential disparate impacts between race categories by identifying significant treatment effects that exist for one group but not the other. This process is repeated

for gender categories in the middle series of graphs. The bottom series of graphs attempt direct between-group outcome comparisons to determine if any significant treatment effects move the signature bond incident rate closer to a parity value of one, or away. There exists a single race based significant treatment effect for white arrestees with the highest DMF risk factor recommendation. Two gender based significant treatment effects were detected; one for female arrested at the lowest DMF risk factor recommendation, and one for male arrestees at the highest DMF risk factor recommendation. These treatment effects do not extend to the between-group signature bond gender ratios, indicating a lack of strong support for potential disparate impacts.

Figure 35 reports the rate of signature bond orders for each group across the four DMF recommendations: signature Bond, cash bail $500 or less, cash bail between $501 and $2500, and cash bail greater than $2500. For race-related comparisons, there is one significant treatment effect, which is for white arrestees recommended for the highest DMF level (cash bail greater than $2500). Provision of the PSA-DMF System printout decreases the proportion of signature bonds for this group from 0.55, to 0.34, an effect size of -0.21, which has a 95% confidence interval between -0.34 to -0.08. Although this decrease of 21 percentage points is fairly substantial, and a potential disparate impact because black arrestees do not have a similar result, the relevant risk score, race based signature bond ratio is not significant, providing at most mixed evidence of a group based fairness impact.

For gender-related comparisons, female arrestees with DMF recommendations at the lowest level (signature bond) were significantly more likely to receive signature bonds, with an effect size of 0.09, on a 95% confidence interval of 0.019 to 0.16. Male arrestees, on the other hand, who had the highest DMF recommendation, were less likely to receive signature bonds (effect size -0.13).   These two findings are consistent with what we would expect when the PSA-DMF System printout is provided: less severe recommendations were more likely to have signature bonds ordered, and more severe recommendations were less likely to have signature bonds ordered. The lack of comparable effects across gender groups suggests a potential disparate impact. However, neither gender-related signature bond ratio for the relevant DMF recommendation levels is significant, again providing at most mixed evidence for a group-based fairness impact from the PSA-DMF System.

Figure 36: DMF Recommendation-Separated Average Bond Amount Graphs

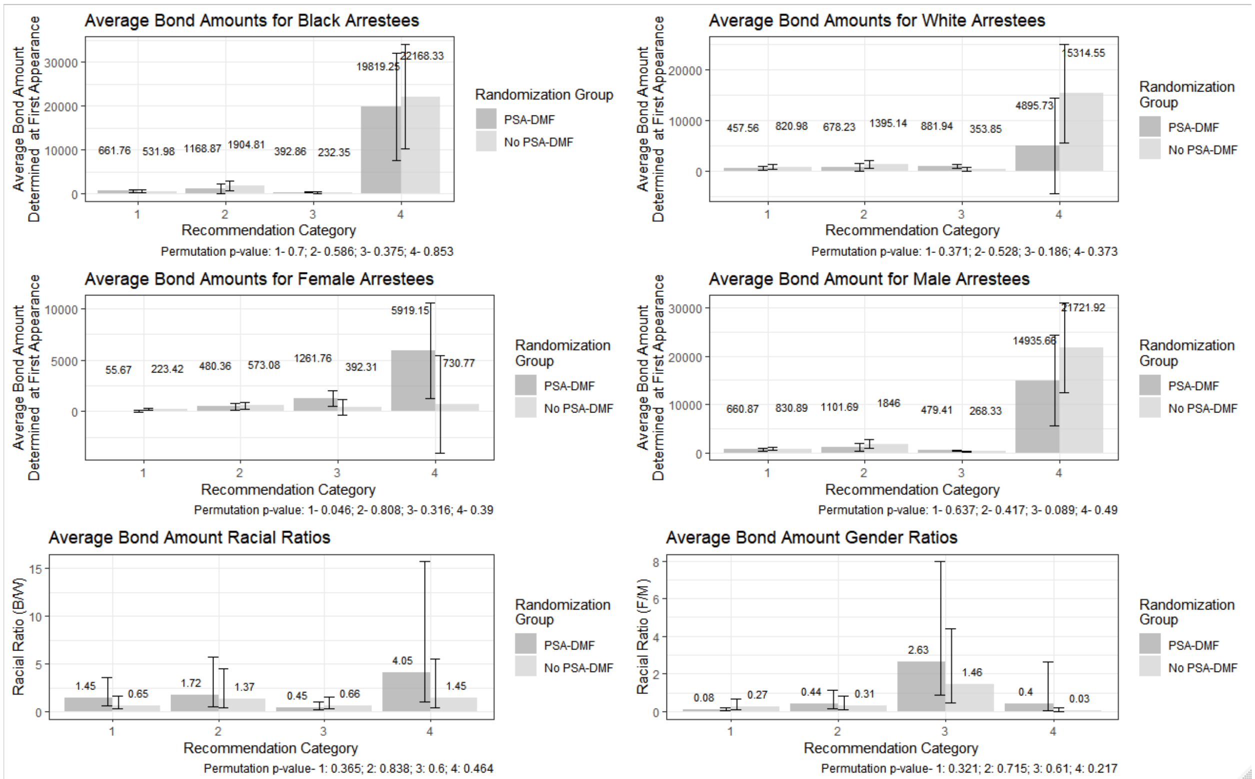

Figure 36: These bar plots report DMF Recommendation level separated mean outcome differences for groups for average bond amount determined at a first appearance hearing. The top two graphs focus on identifying potential disparate impacts between race categories by identifying significant treatment effects that exist for one group but not the other. This process is repeated for gender categories in the middle series of graphs. The bottom series of graphs attempt direct between-group outcome comparisons to determine if any significant treatment effects move the between-group average bond ratio closer to a parity value of one, or away. No significant treatment effects were detected.

Figure 36 reports the results of the average bail amount calculations for each DMF recommendation and for each group. The interpretations of the within-group results are straightforward: significant treatment effects indicated an increase or decrease in the average bail amount for a fixed combination of DMF recommendation groups. Significant treatment effects for the between-group results indicated an increase or decrease in the ratio of bail amounts. As with the diagnostic measures in the previous sections, movements toward 1 indicated an increase in parity, and movements away from 1 indicated a decrease. We detected no significant treatment effects for either within or between-group DMF-separated average bail amounts.

Figure 37: NCA Score-Separated NCA Incidence Rate Graphs

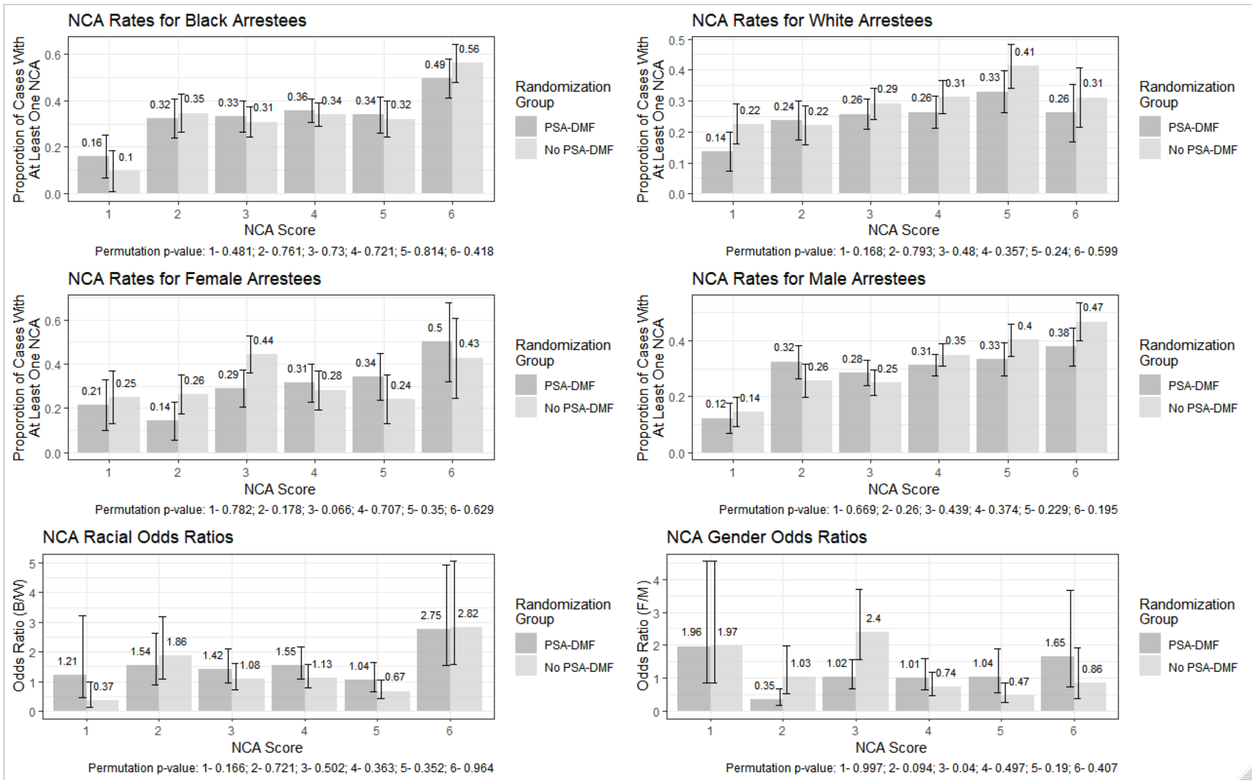

Figure 37: These bar plots report PSA-DMF NCA Score level separated mean outcome differences for groups NCA incident rates. The top two graphs focus on identifying potential disparate impacts between race categories by identifying significant treatment effects that exist for one group but not the other. This process is repeated for gender categories in the middle series of graphs. The bottom series of graphs attempt direct between-group outcome comparisons to determine if any significant treatment effects move the between-group NCA incident rate closer to a parity value of one, or away. There are no significant race based treatment effects and a single gender based treatment effect for gender based NCA odds ratios at an NCA score of 3. Given the lack of other significant treatment effects, this finding does not provide credible evidence for outcome-based disparate impacts.

Figure 37 reports the results of the NCA incident rate analysis across each level of the PSA-DMF System NCA score scale. The interpretations follow the same metric as the prior graphs. For the NCA calculations, we detected one significant treatment effect: arrestees with NCA scores of 3 had a significant decrease in the relevant gender-related odds ratio of 2.4 in the PSA-DMF-System-not-present group compared to 1.02 in the System-present group. This is an effect size of -1.38 with a 95% confidence interval of -2.68 to -0.08. Although this is a practically significant movement towards group parity, we detected no other significant treatment effects, suggesting at most limited evidence for an outcome-based group fairness impact.

Figure 38: NVCA Flag-Separated NVCA Incidence Graphs

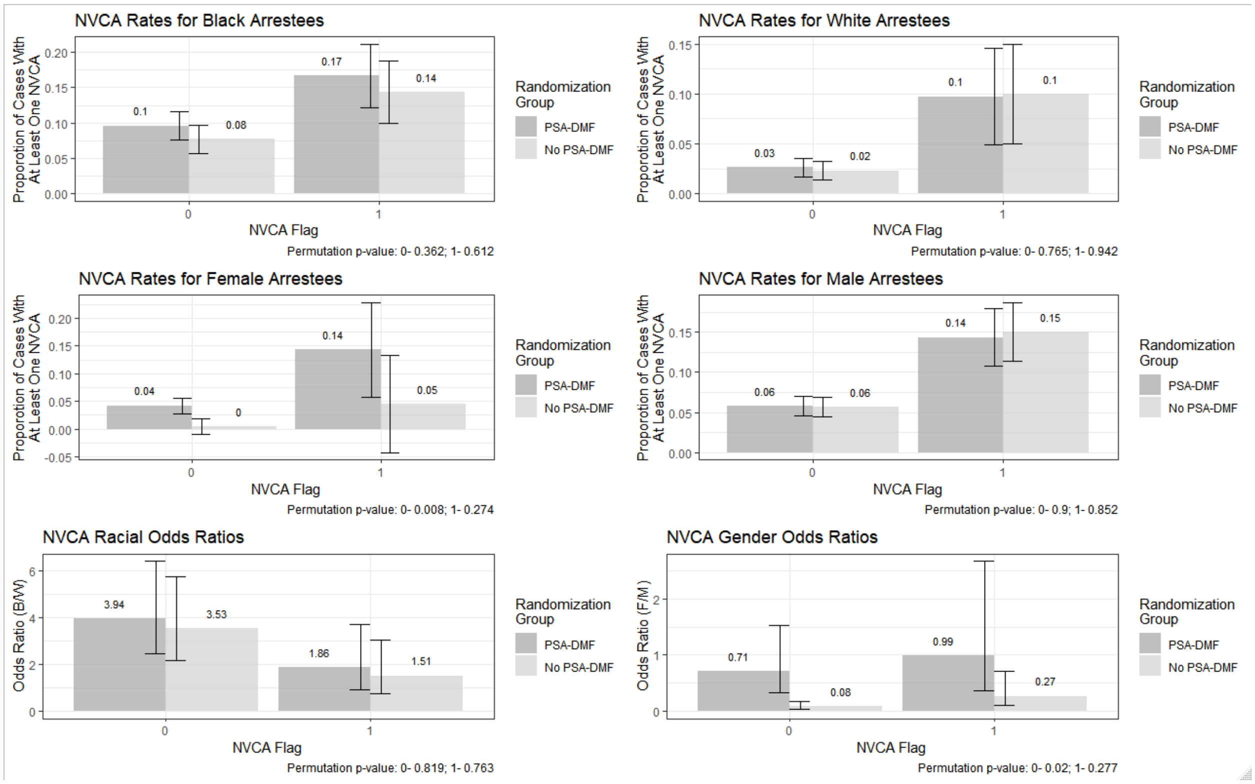

Figure 38: These bar plots report PSA-DMF NVCA Flag level separated mean outcome differences for groups NVCA incident rates. The top two graphs focus on identifying potential disparate impacts between race categories by identifying significant treatment effects that exist for one group but not the other. This process is repeated for gender categories in the middle series of graphs. The bottom series of graphs attempt direct between-group outcome comparisons to determine if any significant treatment effects move the between-group NVCA incident rate closer to a parity value of one, or away. There are no significant race based treatment effects and two significant gender based treatment effects for gender based NVCA odds ratios and NVCA incident rates for female arrestees at an NVCA Flag value of 0. There are only 10 individual arrestees that match the NVCA Flag score of 0, NVCA incident, and female arrestee cell, which is sufficiently rare that this treatment effect is not considered evidence of a true disparate impact.

Figure 38 reports the results of the NVCA incident rate calculations for each level of the PSA-DMF NVCA Flag. For race-related NVCA comparisons, we detected no significant treatment effects. There were, however, two significant effects for gender-related NVCA outcomes. The first is a significant positive treatment effect on the incidence of NVCA for female arrestees with no NVCA flag present. The effect size is roughly 0.037 with a 95% confidence interval of 0.008 to 0.066. The second is a significant treatment effect for the gender-related odds ratio for NVCA incidence, which had an effect size of 0.63on a 95% confidence interval of 0.089 to 1.17. These two results might provide some indication of an impact on fairness, but we are suspicious of them given the small incidence rate of NVCA among women. Specifically, the number of female arrestees with no NVCA flag who were arrested for an NVCA was negligible. Of the 2262 cases in the dataset only 10 cases fit these criteria.

Figure 39: FTA Score-Separated FTA Incidence Graphs

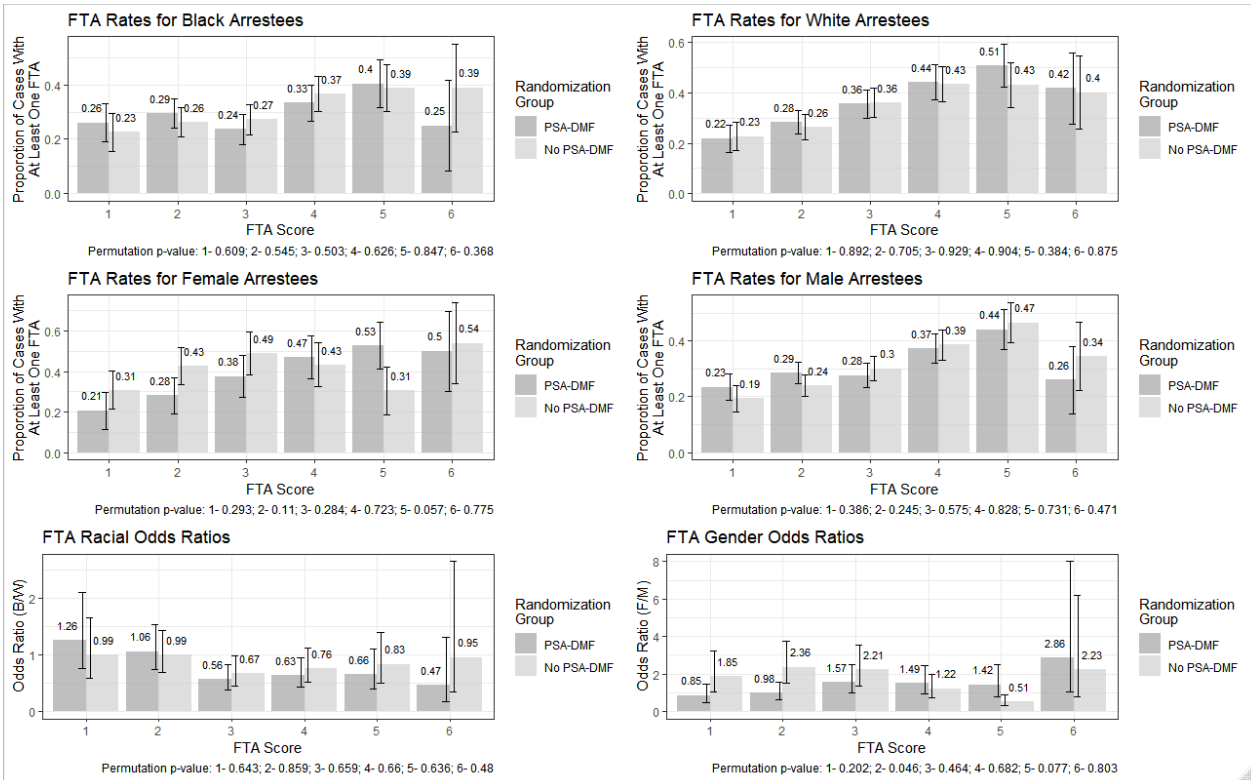

Figure 39: These bar plots report PSA-DMF FTA Score level separated mean outcome differences for groups FTA incident rates. The top two graphs focus on identifying potential disparate impacts between race categories by identifying significant treatment effects that exist for one group but not the other. This process is repeated for gender categories in the middle series of graphs. The bottom series of graphs attempt direct between-group outcome comparisons to determine if any significant treatment effects move the between-group FTA incident rate closer to a parity value of one, or away. There are no significant race based treatment effects and a single gender based treatment effect for gender based FTA odds ratios at an FTA score of 2. Given the lack of other significant treatment effects, this finding does not provide credible evidence for outcome-based disparate impacts.

Figure 39 reports the results of the FTA incidence rate calculations for each level of the PSA-DMF FTA score scale. For race-related FTA, we detected no significant treatment effects. With regard to gender-related FTA comparisons, we detected one statistically significant treatment effect for gender based FTA odds ratios at the score of 2. The effect size is -1.38 with a 95% confidence interval of -2.7 to -0.03. As before, with no other significant treatment effects detected, we do not consider this result to be substantial evidence of an outcome-based group fairness impact.

Figures 37-39 provided at best mixed, and more likely no, evidence for group-based fairness impacts. Although a few metrics--and all of those related to gender--yielded statistically significant treatment effects, they were either non-existent at other risk score levels or applicable to such rare circumstances as to not provide serious robust support for the existence of such fairness impacts.

We note that we have not formally applied multiple testing penalties to any of our results because of the provisional nature of this report. As has been clear from our discussion, however, given the large number of tests conducted, we have not credit as important occasional indications of statistical significance. Given the large number of tests conducted, one would expect to detect a few significant differences even if there are no underlying disparate treatment effects. We anticipate formalizing our treatment of multiple testing in our final report.

3. Anticipated Future Analyses

We anticipate conducting additional analyses in our final report, which we will produce in due course after we receive the final data submission from Dane County. In addition to having 30 (instead of 12) months of data with a two-year (instead of a 12-month) follow-up period, we hope to have the additional data identified in subsection I.F.3.b. Thus, our analyses should be more complete.

We anticipate performing additional statistical analysis with respect to our four primary outcome variables of FTA, N(V)CA, number of days of predisposition incarceration, and demographic fairness measures. First, we anticipate exploring additional modeling. As discussed in Section II.B.1, our workhorse statistical technique for this report was a permutation test, which has the advantage of fewer assumptions but the disadvantage of sometimes producing wide confidence intervals, particularly when working with non-standard distributions (*e.g.*, the number of days of predisposition incarceration). Fitting statistical models sometimes provides more precise inference (including narrower confidence intervals), but at the risk of additional assumptions. Given the imprecise inferences available from the data thus far, and the unexpectedly low case volume, additional modeling may be worth the risk.

We anticipate examining First Appearance decision making with respect to at least the FTA, NVCA, and NCA outcomes using a more sophisticated statistical framework from the causal inference literature. We are working with two quantitative methodologists to adapt the technique of principal stratification[119] to this setting. We also anticipate further work examining the assumptions that underlie our analysis techniques. In particular, we have begun preliminary work with the same two quantitative methodologists to explore how to address potential violations of the Stable Unit Treatment Value Assumption ("SUTVA"). Both of these efforts require advancing the state of the art in causal inference.

[119] See Constantine E. Frangakis & Donald B. Rubin, Principal Stratification in Causal Inference, 58 Biometrics 21 (2002).

We anticipate developing and formalizing demographic fairness metrics other than the four frameworks we included in the report. The literature has numerous metrics, and including them all may cause undue length and complication. We will keep abreast of this literature and respond accordingly.

We anticipate formalizing our treatment of multiple testing penalties. As noted in our discussion of demographic fairness, when one performs a large number of tests, one should anticipate that some will show statistical significance even if the underlying data-generating mechanisms are free of any differences. Statistical formalization of multiple testing can partially address this concern.

We anticipate analyzing outcomes other than the four primary variables examined in this report. For example, we currently have the data to evaluate whether the presence of the PSA-DMF System printout affects case length and conviction probability. We may be able to obtain data on sentencing and post-conviction incarceration.

As noted above, at AV's request, we will analyze FTA, N(V)CA, and number of days of predisposition incarceration using not the predisposition period as our timeframe but rather two years after randomization. The shift in timeframe will require some redefinition of outcomes and may well require additional methodological work.

Finally, we will perform any reasonable additional analyses requested by AV or Dane County. We look forward to discussing with all relevant stakeholders how the Access to Justice Lab can be most helpful in the continuing conversation regarding the PSA-DMF System and predisposition risk assessment instruments more generally.

III. Appendices

A. Dane Decision Making Framework

Public Safety Assessment

Decision Making Framework – Dane County, WI [Effective 10-27-2016]

Decision Making Framework (DMF) Instructions

Step 1: Complete the PSA to generate the FTA scale, NCA scale and NVCA flag.

Step 2: Determine if (1) the defendant was extradited from another state for the current charge; (2) any current charge is escape (946.42), murder/homicide (940.01 – 940.10 excluding 940.04), felony sexual assault 1st and 2nd degree [940.225 (1) and (2)], armed robbery [943.32 (2) if display/recover a firearm as determined by complaint], including attempting, soliciting, or conspiring to commit any of these charges, or an FTA for any of these charges; or (3) the current charge is violent and the PSA resulted in an NVCA flag.

- If yes = Cash Bail, Pretrial Supervision Level III and Maximum Conditions.
- If no, continue to step 3.

Step 3: Apply the FTA and NCA scales derived in Step 1 above to the DMF Matrix to determine the preliminary recommendation release type and corresponding conditions level.

Step 4: Determine if any current charge is domestic abuse (940.19 & 968.075/973.055), robbery [943.32 (if no display/recover a firearm as determined by complaint)], stalking (940.32), violates a temporary restraining order or injunction [813.12(8)], kidnapping (940.31), arson (943.02 and 943.03), or the current offense involved the use of a firearm. Attempting, soliciting, or conspiring to commit any of these offenses or an FTA for any of these offenses are included.

- If no, the preliminary recommendation type and corresponding supervision level if applicable identified in Step 3 is the final recommendation.
- If yes, increase the preliminary recommendation release type and corresponding supervision level if applicable as shown below:
  - Signature Bond with No Conditions or with Conditions if Appropriate = Signature Bond with Pretrial Supervision Level (PSL) I (conditions if appropriate)
  - Signature Bond with PSL I = Modest Cash Bail with PSL II (conditions if appropriate)
  - Modest Cash Bail with PSL II = Moderate Cash Bail with PSL III and GPS
  - Moderate Cash Bail with PSL III and GPS = Cash Bail, PSL III and Maximum Conditions

Pretrial Supervision Information

All pretrial supervision includes criminal history checks and court reminder notifications before each court date. Other services will be provided as shown in the table below.

Risk Level (by Color)	Pretrial Supervision Level	Phone Contact[1]	Face-to-Face Contact[1]	Conditions Monitoring
Dark Green	None	None	None	No
Light Green	None	None	None	No
Yellow	Level I	None	Monthly	Yes
Light Orange	Level II	None	Biweekly	Yes
Dark Orange/Red	Level III	Biweekly	Biweekly	Yes

[1]Biweekly = every other week

PROPERTY OF THE LAURA AND JOHN ARNOLD FOUNDATION
FOR USE ONLY BY AUTHORIZED PSA SITES

DMF Matrix

	NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
FTA 1	Signature Bond with No Conditions	Signature Bond with No Conditions				
FTA 2	Signature Bond with No Conditions	Signature Bond with No Conditions	Signature Bond with Conditions if Appropriate	Signature Bond with PSL I - Conditions if Appropriate	Modest Cash Bail with PSL II - Conditions if Appropriate	
FTA 3		Signature Bond with Conditions if Appropriate	Signature Bond with Conditions if Appropriate	Signature Bond with PSL I - Conditions if Appropriate	Modest Cash Bail with PSL II - Conditions if Appropriate	Cash Bail, PSL III and Maximum Conditions
FTA 4		Signature Bond with Conditions if Appropriate	Signature Bond with Conditions if Appropriate	Signature Bond with PSL I - Conditions if Appropriate	Modest Cash Bail with PSL II - Conditions if Appropriate	Cash Bail, PSL III and Maximum Conditions
FTA 5		Modest Cash Bail with PSL I - Conditions if Appropriate	Modest Cash Bail with PSL I - Conditions if Appropriate	Modest Cash Bail with PSL II - Conditions if Appropriate	Moderate Cash Bail with PSL III and GPS	Cash Bail, PSL III and Maximum Conditions
FTA 6				Cash Bail, PSL III and Maximum Conditions	Cash Bail, PSL III and Maximum Conditions	Cash Bail, PSL III and Maximum Conditions

B. State Public Defender Interview Questions

Location: ____________ Date: _____________ ☐ **Eligible** ☐ **Private Attorney** ____________

☐ **Not Eligible** ☐ **Does Not Want SPD**

☐ **Eligible Pending Verification**

NAME: __________________________ DOB: _________ Sex: ☐ M ☐ F Arrest Date: _____________

Case Type: ☐ Homicide ☐ Class A/B/C Felony ☐ Felony ☐ Misd ☐ Traffic ☐ Extradition

BOOKING CHARGES:

Arrested on Warrant? ☐ Y ☐ N How? ☐ Turned Self In ☐ Arrested at Home ☐ Police Contact/New Charges

Other Pending Charges: __________________ Attorney: ______________ Appoint for Pending? ☐ Y ☐ N

Prior Adult Convictions: ___________________ Missed Court? Y ☐ N ☐ Why: _____________________

RESIDENTIAL HISTORY AND CONTACT

Address: ______________________________ Apt:____ City: ☐ Madison ☐ Other: _________ Zip: _______

☐ Own ☐ Rent ☐ Name on Lease ☐ Sec. Deposit Amt: $_________ Live with alleged victim? ☐ Y ☐ N

Reside: ☐ Dane ☐ WI _____ yrs./all life Came to WI from: ___________ ☐ To Reside ☐ To Visit ☐ Other

Phone: ______________________________ E-mail: ____________________________

Message Phone: _______________________ Who: _______________________________

OTHER TIES

☐ Spouse ☐ Parents ☐ Siblings ☐ Extended Family ☐ Friends

☐ Children Ages: ________________________ Live with Client? Y ☐ N ☐

☐ School Name: ________________________ ☐ Full ☐ Part Time Grade/Year: __________

☐ Co-Signer Name: ________________________ Phone: _____________________

EMPLOYMENT

Employer: _______________________________________ Since: ______________________

☐ Unemployed What are you currently doing? ____

Prior Employer ___________________________________ When: ______________________

Currently Receiving: ☐ SSI ☐ SSDI ☐ SS Retirement ☐ VA Disability ☐ W2 ☐ Unemployment **Veteran**: ☐ Y ☐ N

HOLDS

☐ Probation Hold

Agent: __________________

☐ Serving Sentence Until

Date: ___________________

☐ Other Holds

____

____

INITIAL APPEARANCE INFORMATION ☐ BBH Date: ____________

IA Date: _____________ ☐ Prelim Time Limits Waived

Bond: ☐ Signature ☐ Cash $_______ ☐ Co-Signer: _______________

Conditions:

No Contact: ☐ Victim ☐ BMP

☐ Victim's Address ☐ No Weapons

☐ Codefendant ☐ No Alcohol/Drugs

☐ Geographical ______________ ☐ Not Operate MV

☐ Other ______________________

☐ Other Conditions:____

Trial Court Br: _______________

C. Sample PSA-DMF System Printout

DANE COUNTY CLERK OF COURTS
Public Safety Assessment – Report

215 S Hamilton St #1000
Madison, WI 53703
Phone: (608) 266-4311

Name: [redacted] **Spillman Name Number:** [redacted]
DOB: [redacted] **Gender:** Male
Arrest Date: 03/25/2017 **PSA Completion Date:** 03/27/2017

New Violent Criminal Activity Flag

No

New Criminal Activity Scale

1	2	3	4	5	6

Failure to Appear Scale

1	2	3	4	5	6

Charge(s):

961.41(1)(D)(1) MFC DELIVER HEROIN <3 GMS F 3

Risk Factors:	Responses:
1. Age at Current Arrest	23 or Older
2. Current Violent Offense	No
a. Current Violent Offense & 20 Years Old or Younger	No
3. Pending Charge at the Time of the Offense	No
4. Prior Misdemeanor Conviction	Yes
5. Prior Felony Conviction	Yes
a. Prior Conviction	Yes
6. Prior Violent Conviction	2
7. Prior Failure to Appear Pretrial in Past 2 Years	0
8. Prior Failure to Appear Pretrial Older than 2 Years	Yes
9. Prior Sentence to Incarceration	Yes

Recommendations:

Release Recommendation - Signature bond
Conditions - Report to and comply with pretrial supervision

Notes -

2/25/2020
1:21:29 PM 1

D. Rule 206 One Year Report

Honorable Juan B. Colás
Presiding Judge, Dane County Circuit Court
215 South Hamilton Street, Room 7103
Madison, WI 53703-3291

Telephone: (608) 266-4460 **Facsimile: (608) 266-4079**

Ashley Sanders, Clerk Renee Treasure, Judicial Assistant

October 20, 2017

To: Circuit Court Judges

From: Juan B. Colás, Presiding Judge

Re: One-Year Report: Amended Rule 206, Earlier Initial Bail Review Hearings

Effective July 1, 2016 Local Rule 206 was amended to provide automatic, prompt bail review hearings or initial appearances for persons booked into the jail. The former rule required that a hearing be requested by counsel. This memo summarizes the effects of the change over one full year.

Commissioner Hanson tracked all cases to which the rule applied from July 1, 2016 through June 30, 2017. He compared when the first bail hearing or the initial appearance occurred to when it would have occurred under the old rule. He made the assumption most favorable to the old rule: that a request for a hearing would have been made, even though such requests were actually rare. In other words, he used the earliest time the hearing *could have* occurred as the time the hearing *would have* occurred.

The rule applied to 1,873 hearings. After the amendment the median time from booking to hearing was 48.42 hours, down from the 65 hours the median would have been under the old rule. The new rule resulted in a 26% reduction in time to hearing and the equivalent of 400 jail days saved, (conservatively assuming that everyone who got cash bail remained in jail).

82% of those who had a hearing were released immediately, the rest had cash bail. 86% of hearings were initial appearances and 14% were bail review only. On average there were 22 bail-review only hearings each month, lower than in the first quarter after the amendment (26) and slightly lower than under the previous rule (23). Some feared that the amendment might cause more people to have a CCAP record of a bail hearing whom a prosecutor with more time might have released without charges or a bail hearing record under the previous rule. The stability of the ratio of initial appearances to bail-only hearing suggests this has not happened.

E. Wisconsin Bail Statute Criminal Procedure 969.01

CHAPTER 969

BAIL AND OTHER CONDITIONS OF RELEASE

969.001 Definitions.
969.01 Eligibility for release.
969.02 Release of defendants charged with misdemeanors.
969.03 Release of defendants charged with felonies.
969.035 Pretrial detention; denial of release from custody.
969.04 Surety may satisfy default.
969.05 Endorsement of bail upon warrants.
969.065 Judicial conference; bail alternatives.
969.07 Taking of bail by law enforcement officer.
969.08 Grant, reduction, increase or revocation of conditions of release.
969.09 Conditions of bond.
969.10 Notice of change of address.
969.11 Release upon arrest in another county.
969.12 Sureties.
969.13 Forfeiture.
969.14 Surrender of principal by surety.

Cross–reference: See definitions in s. 967.02.

969.001 Definitions. In this chapter:

(1) "Bail" means monetary conditions of release.

(2) "Serious bodily harm" means bodily injury which causes or contributes to the death of a human being or which creates a substantial risk of death or which causes serious permanent disfigurement, or which causes a permanent or protracted loss or impairment of the function of any bodily member or organ or other serious bodily injury.

History: 1981 c. 183; 1987 a. 399.

969.01 Eligibility for release. **(1)** BEFORE CONVICTION. Before conviction, except as provided in ss. 969.035 and 971.14 (1r), a defendant arrested for a criminal offense is eligible for release under reasonable conditions designed to assure his or her appearance in court, protect members of the community from serious bodily harm, or prevent the intimidation of witnesses. Bail may be imposed at or after the initial appearance only upon a finding by the court that there is a reasonable basis to believe that bail is necessary to assure appearance in court. In determining whether any conditions of release are appropriate, the judge shall first consider the likelihood of the defendant appearing for trial if released on his or her own recognizance.

(2) AFTER CONVICTION. (a) Release pursuant to s. 969.02 or 969.03 may be allowed in the discretion of the trial court after conviction and prior to sentencing or the granting of probation. This paragraph does not apply to a conviction for a 3rd or subsequent violation that is counted as a suspension, revocation, or conviction under s. 343.307, or under s. 940.09 (1) or 940.25 in the person's lifetime, or a combination thereof.

(b) In misdemeanors, release may be allowed upon appeal in the discretion of the trial court.

(c) In felonies, release may be allowed upon appeal in the discretion of the trial court.

(d) The supreme court or a justice thereof or the court of appeals or a judge thereof may allow release after conviction.

(e) Any court or judge or any justice authorized to grant release after conviction for a misdemeanor or felony may, in addition to the powers granted in s. 969.08, revoke the order releasing a defendant.

(3) BAIL FOR WITNESS. If it appears by affidavit that the testimony of a person is material in any felony criminal proceeding and that it may become impracticable to secure the person's presence by subpoena, the judge may require such person to give bail for the person's appearance as a witness. If the witness is not in court, a warrant for the person's arrest may be issued and upon return thereof the court may require the person to give bail as provided in s. 969.03 for the person's appearance as a witness. If the person fails to give bail, the person may be committed to the custody of the sheriff for a period not to exceed 15 days within which time the person's deposition shall be taken as provided in s. 967.04.

(4) CONSIDERATIONS IN SETTING CONDITIONS OF RELEASE. If bail is imposed, it shall be only in the amount found necessary to assure the appearance of the defendant. Conditions of release, other than monetary conditions, may be imposed for the purpose of protecting members of the community from serious bodily harm or preventing intimidation of witnesses. Proper considerations in determining whether to release the defendant without bail, fixing a reasonable amount of bail or imposing other reasonable conditions of release are: the ability of the arrested person to give bail, the nature, number and gravity of the offenses and the potential penalty the defendant faces, whether the alleged acts were violent in nature, the defendant's prior record of criminal convictions and delinquency adjudications, if any, the character, health, residence and reputation of the defendant, the character and strength of the evidence which has been presented to the judge, whether the defendant is currently on probation, extended supervision or parole, whether the defendant is already on bail or subject to other release conditions in other pending cases, whether the defendant has been bound over for trial after a preliminary examination, whether the defendant has in the past forfeited bail or violated a condition of release or was a fugitive from justice at the time of arrest, and the policy against unnecessary detention of the defendant's pending trial.

History: 1977 c. 187; 1979 c. 112; 1981 c. 183; 1993 a. 486; 1995 a. 77; 1997 a. 232, 283; 2009 a. 100, 214.

The trial court exceeded its authority in granting bail to a revoked probationer pending review of a probation revocation. State ex rel. Shock v. DHSS, 77 Wis. 2d 362, 253 N.W.2d 55 (1977).

Habeas corpus is available to persons released on personal recognizance bonds. State ex rel. Wohlfahrt v. Bodette, 95 Wis. 2d 130, 289 N.W.2d 366 (Ct. App. 1980).

The court may impose a monetary condition of release under sub. (2) (b). State v. Barnes, 127 Wis. 2d 34, 377 N.W.2d 624 (Ct. App. 1985).

A warrant under sub. (3) must be supported by probable cause to believe that the testimony of the person is material and that it may become impractical to secure the person's presence by subpoena. State v. Brady, 130 Wis. 2d 443, 388 N.W.2d 151 (1986).

Indigency under this section relates to current economic status and does not involve consideration of whether the defendant is shirking unless the shirking relates to another statutory factor. Cash bail is not prohibited against an indigent convicted misdemeanant who takes an appeal. However, where there is no risk that the indigent misdemeanant will not appear, cash bail is inappropriate. State v. Taylor, 205 Wis. 2d 664, 556 N.W.2d 779 (Ct. App. 1996), 96–0857.

The conditions that a court is authorized to impose under this section and s. 969.03 govern the release of a defendant from custody and do not apply if the defendant cannot post bond and is not released. A court may impose pretrial, no–contact provisions on incarcerated defendants under s. 940.47 if the terms of that statute are met. State v. Orlik, 226 Wis. 2d 527, 595 N.W.2d 468 (Ct. App. 1999), 98–2826.

A circuit court that followed a blanket policy that mandated participation in a pretrial program as a condition of release for all persons based solely on the nature of the offense, without making an individualized determination that that condition is appropriate, erroneously exercised its discretion in setting conditions of bail. State v. Wilcenski, 2013 WI App 21, 346 Wis. 2d 145, 827 N.W.2d 642, 12–0142.

Under sub. (1), judges and court commissioners have the power, prior to the filing of a complaint, to release on bail persons arrested for the commission of a felony. 65 Atty. Gen. 102.

The public defender may represent indigent material witnesses subject to sub. (3) bail provisions so long as there is no conflict of interest with another client, but may not represent indigents in civil forfeiture actions unless that action is reasonably related to one for which an indigent is entitled to counsel. 72 Atty. Gen. 61.

Pretrial release; Wisconsin bail reform. 1971 WLR 594.

The presumption of release in bail decisions. Adelman and Schulenburg. Wis. Law. July 1989.

Nationwide Trend: Rethinking the Money Bail System. Okocha. Wis. Law. June 2017.

F. Definitions

New Criminal Activity: A New Criminal Activity (NCA) occurred when an arrestee experiences a new booking/arrest/citation event that results in at least one charge that carries the potential of incarceration as a penalty. An NCA is considered valid for a specific case if it occurred during the pretrial period for that case, i.e. between First Appearance and Disposition. This measurement is constructed as both a total count of all NCA activities that occurred between the IA date and the disposition date for a case as well as a binary measure that takes on a value of 1 if any NCA was observed during the pretrial period, and 0 otherwise. It should be noted that NCA was used as both a risk score output of the PSA-DMF System and as an outcome measure for evaluating the PSA-DMF System's effectiveness.

New Violent Criminal Activity: A New Violent Criminal Activity (NVCA) occurs when an arrestee experiences a new booking/arrest/citation event that results in at least one charge that carries the potential of incarceration as a penalty and is considered to be a violent charge according to state statute. Any observed NVCA is, by definition, also a valid NCA observation. An NVCA is considered valid for a specific case if it occurred during the pretrial period for that case, i.e. between First Appearance and Disposition. This measurement is constructed as both a total count of all NVCA activities that occurred between the IA date and the disposition date for a case as well as a binary measure that takes on a value of 1 if any NVCA was observed during the pretrial period, and 0 otherwise. It should be noted that NVCA was used as both a risk score output of the PSA-DMF System and as an outcome measure for evaluating the PSA-DMF System's effectiveness.

Failure to Appear - Outcome: A Failure to Appear (FTA) outcome event represents any unexcused absence by a defendant from a prior scheduled court appearance which resulted in a bench warrant being issued. An FTA is considered valid for a specific case if it occurred during the pretrial period for that case, i.e. between First Appearance and Disposition. This measurement is constructed as both a total count of all FTA activities that occurred between the IA date and the disposition date for a case as well as a binary measure that takes on a value of 1 if any FTA was observed during the pretrial period, and 0 otherwise. It should be noted that FTA was used in several ways during the PSA-DMF System process. First, FTA was a factor used to generate one of an arrestee's PSA-based risk scores. Second, the PSA generated that FTA risk score (on a scale between one and six). Finally, FTA was an outcome measurement for evaluating the effectiveness of the PSA-DMF System.

Failure to Appear - PSA input: An FTA input referred to events in an arrestee's criminal history that were used to generate PSA-based risk scores. FTA inputs include: 1) the arrestee missing any pre-disposition court appearance (e.g., misdemeanor, felony, or criminal traffic hearing) or a bench warrant issued for such non-appearance within a two-year period before the current arrest; and 2) the same events occurring more than two years before the current arrest.[120]

[120] A copy of the worksheet that the Assessors use appears in Appendix H.

Predisposition Incarceration: Predisposition Incarceration is a metric that represents the total amount of time, in days, that an arrestee spent incarcerated during the pretrial period of a specific case, i.e. from a case's First Appearance until it's disposition. There are two primary sources of incarceration time: 1) the initial booking event: from the IA hearing until first release, for example due to posting a bond; and 2) subsequent booking events due to being held for an N(V)CA.

Time Until First Failure: Time Until First Failure is a metric that represents the total time, in days, between a specific case's First Appearance hearing and the date of either the first NCA or FTA. In the event that a case had no valid NCA of FTA observations, this metric is accorded a Null or missing value. Any analyses including this metric represent only cases that had at least one failure, i.e. at least one valid N(V)CA or FTA.

Bail Amounts: A bail amount is the amount of dollars necessary for an arrestee to secure release from jail after the First Appearance hearing in a specific case. This is functionally the Bond amount set by the Commissioner. For Signature Bonds, the bail amount may have been any amount (including $0) that must have been paid if the arrestee failed to appear for the next scheduled hearing.

Release Decision Type: Release Decision refers to the monetary conditions imposed by the Commissioner on an arrestee for a specific case at that case's First Appearance hearing. The Dane County Research Team has traditionally relied on a metric of release decisions that focus on 4 types of categories: Signature Bonds, Cash Bond less than $500, Cash Bond between $501 and $2500, and Cash Bond greater than $2500. The A2J Lab also utilized a similar construct that collapsed the Cash Bond categories into only two categories: bonds exceeding $1000 and bonds below $1000. The $1000 threshold was chosen for ease of representation in graphs and statistical tests.

Demographic Fairness Measures: Demographic Fairness refers to metrics that compare some relevant outcome of the pretrial process or PSA assessment across categories of a relevant demographic variable between the two randomization groups, to determine whether the availability or non-availability of the PSA-DMF system moves these comparative outcomes closer to or farther away from parity between the two demographic groups. For instance, one demographic fairness measure would be Race based False Positive Rates, where the False Positive Rate (the number of incorrect high risk predictions for a hypothetical threshold on the PSA) for both Black and White arrestees are compared under the PSA-DMF system present group and the no PSA-DMF system present group. A ratio of the two False Positive Rates of 1 would represent perfect parity between the two groups, so if the PSA-DMF system present group has a significantly different ratio than the no PSA-DMF system present group that is closer to 1, than this represents some evidence that the PSA-DMF system increases overall demographic fairness.

Incarceration Prior to Disposition Measures: This metric is functionally equivalent to the Predisposition Incarceration metric; both are used in the relevant academic literature, so both terms are used interchangeably here as well.

Completed Arrest: A completed arrest was an apprehension, either custodial or non-custodial, by law enforcement. A custodial arrest was an apprehension that resulted in a person's transfer to the Dane County Jail (located in the Public Safety Building), and a non-custodial arrest was one that only involved receipt of a citation and summons to be present in court for the First Appearance.

G. Miscellaneous Data Tables

Table 1: Data Sources

Sources	Description	Data	Start Date	Last Observation Date
CCAP	CCAP is a statewide case management system that houses various information for court cases in Wisconsin	IA Outcomes, FTA, N(V)CA	06/01/2017	05/31/2019
PROTECT	PROTECT is a case management system used by the Dane County District Attorney	N(V)CA	06/01/2017	05/31/2019
Spillman	Spillman is a database that contains jail booking information and is in use by the Dane County Sheriff's Office	Pretrial Incarceration Length	05/19/2017	06/06/2019
Pretrial Assessor Database	The Pretrial Assessor's database is used to input PSA related information, calculate PSA scores, and output PSA scores and PSA-DMF system recommendations; this system was developed and implemented via funding from Arnold Ventures	PSA Inputs, PSA Outputs, Demographic Information	06/01/2017	05/31/2018

Table 2: NCA Heatmap Raw Row (within Score) Proportions

NCA Score	Signature Bond		Cash Bond <= $1000		Cash Bond > $1000	
1	0.885	0.845	0.010	0.026	0.104	0.129
2	0.920	0.862	0.033	0.050	0.047	0.088
3	0.870	0.793	0.062	0.142	0.069	0.065
4	0.647	0.719	0.199	0.180	0.153	0.101
5	0.455	0.558	0.359	0.302	0.186	0.140
6	0.250	0.413	0.466	0.380	0.284	0.207
	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF

Table 3: NVCA Heatmap Raw Row (within Score) Proportions

	Signature Bond		Cash Bond <= $1000		Cash Bond > $1000	
No NVCA Flag	0.748	0.742	0.151	0.168	0.101	0.090
NVCA Flag Present	0.426	0.572	0.301	0.221	0.273	0.207
	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF

Table 4: FTA Heatmap Raw Row (within Score) Proportions

FTA Score	Signature Bond		Cash Bond <= $1000		Cash Bond > $1000	
1	0.891	0.856	0.029	0.030	0.080	0.114
2	0.801	0.786	0.088	0.129	0.111	0.085
3	0.777	0.782	0.128	0.136	0.094	0.082
4	0.560	0.602	0.292	0.261	0.148	0.137
5	0.360	0.459	0.426	0.385	0.213	0.156
6	0.171	0.356	0.415	0.444	0.415	0.200
	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF	PSA-DMF	No PSA-DMF

H. PSA Assessor Risk Factor Worksheet

The worksheet below is no longer considered confidential; Arnold Ventures has publicized the factors associated with calculating the PSA risk scores. More information on the factors and how the score is calculated can be found at https://advancingpretrial.org/psa/factors/.

PUBLIC SAFETY ASSESSMENT – RISK FACTOR WORKSHEET

DEFENDANT NAME	GENDER	RACE		CASE IDENTIFIER
DOB (AGE)	**ARREST DATE**			**PSA COMPLETION DATE**
RISK FACTOR	**CIRCLE CORRECT RESPONSE**			**DOCUMENT DATA SOURCE, DATE, LOCALITY, CASE NUMBER, CHARGE/FTA**
1. Age at Current Arrest Defendant age at time of current arrest. If no arrest or arrest date unknown, use PSA completion date	20 or younger	21 or 22	23+	
2. Current Violent Offense Consult the violent offense list in the appendix of the PSA Risk Factor and Outcome Definitions document	No	Yes		
3. Pending Charge At The Time Of The Offense Defendant was arrested, released, and allegedly committed a new offense while on release pending trial	No	Yes		
4. Prior Misdemeanor Conviction Conviction for criminal offense defined as a misdemeanor by statute (excludes civil traffic or ordinance violations)	No	Yes		
5. Prior Felony Conviction Conviction for criminal offense defined as a felony by statute	No	Yes		
6. Prior Violent Conviction Conviction for violent offense per violent offense list in PSA Risk Factor and Outcome Definitions document appendix	0	1 or 2	3+	
7. Prior FTA Pretrial In Past 2 Years Missed pre-disposition court appearance (misdemeanor, felony or criminal traffic) & warrant issued in past 2 years	0	1	2+	
8. Prior FTA Pretrial Older Than 2 Years Missed pre-disposition court appearance (misdemeanor, felony or criminal traffic) & warrant issued more than 2 years prior to arrest	No	Yes		
9. Prior Sentence To Incarceration Prison or jail sentence (misdemeanor, felony or criminal traffic) of 14 days or more (not suspended) imposed at sentencing or re-sentencing	No	Yes		