Overriding (In)justice: Pretrial Risk Assessment Administration on the Frontlines

Sarah A. Riley
sriley23@stanford.edu
Stanford University
Stanford, CA, USA

ABSTRACT

A small but growing number of empirical studies have attempted to measure the impacts of algorithmic pretrial risk assessments on discrete policy goals such as decarceration, racial equity, and public safety. A separate but related body of work explores frontline worker resistance and discretion related to sociotechnical systems in criminal legal contexts. I build on work that aims to bridge the gaps between these literatures by offering an ethnographic account of pretrial risk assessment administration across the United States. I draw on semi-structured interviews with 74 pretrial actors and site observations across 8 jurisdictions. I highlight the process of risk assessment administration and the frontline workers who perform that labor. Like judges, pretrial officers have the autonomy to override risk assessment recommendations, unlike judges however, their decisions are made outside the courtroom and far removed from public scrutiny. This paper makes three contributions. First, it provides a detailed account of the personal, professional, and organizational dynamics that lead pretrial officers to override risk assessment recommendations. Second, it presents a taxonomy of override behavior among pretrial officers in an effort to promote more effective policy decisions. Lastly, it provides further empirical evidence that pretrial risk assessments are unlikely to guarantee racial or economic equity or decarceration in the long term.

CCS CONCEPTS

• **Applied computing → Law, social and behavioral sciences**.

KEYWORDS

ethnography, risk assessment, criminal justice, pretrial

ACM Reference Format:
Sarah A. Riley. 2024. Overriding (In)justice: Pretrial Risk Assessment Administration on the Frontlines. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3630106.3658920

1 INTRODUCTION

A small but growing number of empirical studies aim to measure the impacts of algorithmic pretrial risk assessments on discrete policy goals such as decarceration, racial equity, and public safety. A separate but related body of work explores frontline worker resistance and discretion related to sociotechnical systems in criminal legal contexts. I build on work that bridges gaps between these literatures [e.g., 2, 23] by offering an ethnographic account of pretrial risk assessment administration across the United States. I draw on semi-structured interviews with 74 pretrial actors and site observations across 8 jurisdictions, and in doing so, I shed light on an understudied phase of the criminal legal process and underappreciated group of legal actors. Whereas most research focuses on risk assessment output or judicial decision-making, I highlight the process of risk assessment administration and the frontline workers who perform that labor. Like judges, pretrial officers have the autonomy to override risk assessment recommendations, unlike judges however, their decisions are made outside the courtroom and far removed from public scrutiny. This paper makes three contributions. First, it provides a detailed account of the personal, professional, and organizational dynamics that lead pretrial officers to override risk assessment recommendations. Second, it presents a taxonomy of override behaviors among pretrial officers in an effort to promote more effective policy decisions. Lastly, it provides further evidence that pretrial risk assessments are unlikely to guarantee racial and economic equity, decarceration, and public safety in the long term, due in part to implementation variation.

1.1 Measuring the Impacts of Pretrial Risk Assessments

The evidence as to whether or not pretrial risk assessments achieve their intended policy goals is mixed, in part because the body of empirical research documenting their impacts is still relatively small [12]. In an environment where algorithms supplant judicial decisions entirely, it seems possible that pretrial risk assessments could lead to lower jail populations, reductions in crime, and greater racial equity [18]. However, as many point out, judges—not algorithms—make the final pretrial decisions, and understanding the effects of pretrial risk assessments in terms of jail populations, racial disparities, and criminal activity requires empirical study [e.g., 7, 12, 26, 28]. And the results of multiple quasi-experimental studies suggest that pretrial risk assessments do not fulfill their potential. In fact, the evidence suggests that their impact on decarceration is often negligible [12, 16, 26, 28]; if it is positive, it diminishes, sometimes within months of implementation [25, 26]. Pretrial risk assessments may also drive racial discrimination in the criminal legal system [3, 12], perhaps because judicial overrides more often favor white defendants and punish Black and Latinx defendants [1, 12]. This aligns with growing evidence that risk assessments primarily benefit white and more affluent defendants

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658920

and can have the opposite effect for Black and indigent defendants [21, 24].

1.2 Frontline Worker Discretion and Resistance

As "entangled, relational, emergent, and nested" [15] systems, pretrial risk assessments are both shaped by and actively shape social reality. This makes discrete policy outcomes such as decarceration and racial equity difficult to predict, as system outcomes are highly context dependent. Little qualitative research on pretrial risk assessment administration exists, however studies on judicial decision-making in pretrial and sentencing contexts, as well as frontline decision-making in other municipal contexts, provide clues as to how social context can influence outcomes in unpredictable ways.

Prior studies have documented a variety of ways that frontline workers resist, manipulate, or override algorithmic systems. For example, law enforcement agents may ignore predictive policing "hot spot" recommendations or interfere with patrol car antennae to prevent managers from hearing their conversations in the field [5, 11]. There are also examples of judges in pretrial [28] and sentencing [9, 26] contexts overriding risk assessment recommendations, which can systematically disadvantage Black or older defendants [28]. At the same time, in the child welfare context, caseworker overrides may actually reduce racial disparities [10] and improve accuracy [13]. These contradictions create further confusion about the role that human discretion should play in the algorithmic system administration process. There are many reasons that frontline workers intervene. For example, they may oppose managerial surveillance, or resist deskilling [5]. Alternatively, workers may distrust algorithmic output [9], or lack institutional guidance on navigating disagreements between the algorithmic output and frontline worker recommendations [4]. Lastly, interventions may be explained by organizational structures such as local norms or information dissemination practices that influence frontline worker behavior [23].

1.3 Pretrial Workers and Pretrial Agencies

Pretrial risk assessments cannot administer themselves; they require an extensive infrastructure: databases to collect and store risk factors, people to organize and conduct pretrial interviews, and networks to disseminate risk assessment scores to judges, defense attorneys, and prosecutors. Pretrial agencies house and maintain that infrastructure, and they are an example of what Celeste Watkins-Hayes calls catch-all bureaucracies [30]. As catch-all bureaucracies, pretrial agencies are distinct from other public agencies such as the Department of Motor Vehicles (DMV) because they primarily serve people who face a multitude of interconnected issues stemming from social and economic disadvantage [30]. For example, people in the pretrial system are much more likely to suffer from mental health disorders or drug addiction than those who never get arrested [29]. As a result, while pretrial officers must execute certain mandates, they are also challenged to address a wide array of issues far beyond the boundaries of those mandates [30].

While prosecutors, defense attorneys, and judges each have well-defined roles at discrete stages of the criminal legal process, the role of pretrial officers is more ambiguous. I spoke to pretrial officers who described a complex, and sometimes conflicting, set of responsibilities. Pretrial officers are both expected to advocate for defendants' wellbeing and convey their misdeeds to the court. At the same time, pretrial agency funding often comes from state agencies. This renders pretrial officers triadic advocacy workers [17], who must manage and maintain relationships with defendants while navigating the constraints of multiple organizational bureaucracies, including a court system, state criminal justice services agency, and pretrial agency.

Pretrial officers might also be called "satellites of social control" [8] within the criminal legal system because although pretrial agencies are institutionally distinct from jails and courtrooms, they exert influence on defendants throughout every stage of the legal process. Prior to arraignment, pretrial officers exercise power by conducting defendant interviews—usually in jails—and subsequently determining employment status and history of drug abuse. At arraignment, pretrial officers wield control via pretrial reports, which judges consult as they make pretrial decisions. And post arraignment, pretrial officers assume supervision duties, which requires that they report defendants' behavior to the court, particularly if it violates bond conditions.

2 DATASET AND METHODS

I collected data in two separate phases. In the first phase, I aimed for breadth. I conducted semi-structured interviews via Zoom with 26 pretrial actors across 11 states, all of whom were either pretrial supervisors or pretrial data analysts. I targeted winners of the MacArthur Foundations Safety and Justice Challenge, a grant program with a goal of "reducing jail incarceration and increasing equity for all" (SJC). I outreached jurisdictions who explicitly mentioned pretrial risk assessments among their decarceration strategies. Of the eight jurisdictions I emailed, seven agreed to participate. In addition to the 16 pretrial workers from those jurisdictions, I also spoke to 10 additional pretrial actors who I identified via snowball sampling.

This phase of my data collection revealed several key aspects of the pretrial risk administration landscape across the United States, which in turn informed my approach in the subsequent phase. First, a great degree of inter-agency collaboration is required to administer a pretrial risk assessment. For example, two common risk factors include charge type and criminal history. Thus, at a minimum, pretrial agencies must collaborate to some degree with the state-level agency that stewards crime data and the local law enforcement agency that arrested the defendant. Second, there is considerable variation across jurisdictions with respect to pretrial processes broadly, and risk assessment administration specifically. This is true even among jurisdictions that implement the same risk assessment. For example, I found that the manner in which risk assessment recommendations are presented to judges and the gamut of supportive interventions available to defendants upon pretrial release vary greatly. Lastly, it follows that pretrial risk assessment administration is a complex, multi-stage process involving multiple legal actors. Although greater attention is paid to judicial decision-making, I show that pretrial officers wield considerable influence over pretrial outcomes.

In the second phase of data collection, I aimed for depth. To mitigate some of the variation I observed in first phase, I chose to focus on the state of Virginia. Pretrial programs across the entire state administer the same risk assessment, draw on the same data infrastructures, and operate under the guidance of a single state agency. Conducting research in an environment where these attributes were held constant enabled me to focus on other sources of variation. The Virginia Pretrial Risk Assessment Instrument (VPRAI) is administered by all 35 pretrial agencies across the state, which serve 115 of Virginia's 133 cities and counties. All pretrial agencies in the state are managed by the Department of Criminal Justice Services (DCJS), which sets policies and standards and administers general appropriation funds and small grants to pretrial localities. DCJS is also responsible for maintaining the Pretrial and Community Corrections Case Management System (PTCC), which contains all the administrative data required for a portion of the VPRAI. I outreached all 35 pretrial agencies and ultimately spoke with pretrial workers from 22 of them. In total, I conducted semi-structured interviews with 48 pretrial workers across Virginia. While the majority were pretrial officers or supervisors, I also spoke to 16 public defenders, 1 DCJS employee, 1 judge, and 1 bail fund activist. In addition to semi-structured interviews, I also conducted site visits to eight jurisdictions, during which I shadowed pretrial officers, attended bail hearings, and visited two jails.

3 RESULTS

3.1 Decomposing the Pretrial Process

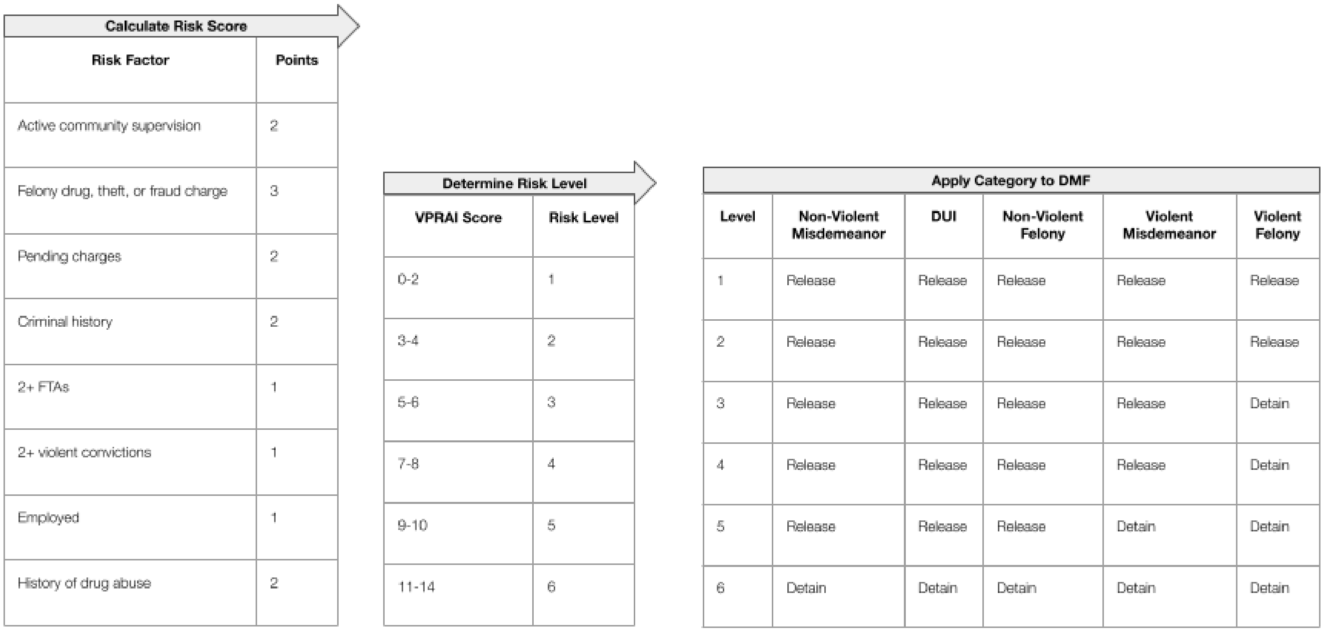
Although research tends to conceptualize the pretrial process as a point-in-time decision involving a single judicial actor, during the course of my fieldwork, it became clear that it is in fact a complex, multi-actor process. First, calculating a risk score requires the cooperation of multiple actors across multiple agencies, and determining risk factor values is a subjective process. Next, once a score is calculated and a pretrial release/detention recommendation is generated, the recommendation is communicated to judges, defense attorneys, and prosecutors through another highly subjective process. Here I describe those two stages of the pretrial process. In doing so, I detail the ways that pretrial workers shape risk assessment inputs and outputs, formally and informally, and with or without a digital trace, for moral, political and professional reasons. I refer to all of these behaviors as overrides, which I take from Virginia's VPRAI instruction manual. In Virginia, pretrial officers have the discretion to override the system's recommendation if they believe there are sufficient "aggravating/mitigating considerations" (see Figure 2). Performing this type of override requires that pretrial officers provide a written justification of their decision to the court. This process is both formal and traceable. During the course of my fieldwork, however, many pretrial workers acknowledged intervening in the risk assessment administration process in a manner they understood to shift the inputs or outputs to the algorithmic system. Pretrial officers have found creative ways to resist [20], tamper with [14], and misuse [19] risk assessments when they create discomfort or violate pretrial officers' sense of duty or morality. In decomposing the pretrial risk assessment administration process, I detail the myriad ways this occurs. And by taxonomizing overrides, I contribute to our collective understanding of pretrial risk assessment administration on the ground, promoting more honest policy discussions about the role of pretrial risk assessments in criminal legal reform.

3.2 Stage 1: Calculating the Risk Score

Stage 1 generally occurs in three steps: first, pretrial officers use a set of attributes, each assigned a point value, to generate a scaled risk score; next, this risk score is converted to a risk level or category; and finally, the risk category is applied to a decision-making framework (DMF) that translates the risk category into a recommended course of action (i.e., release or detain). The Praxis, Virginia's DMF, uses the risk level and current charge type (i.e., nonviolent felony, driving under the influence, nonviolent felony, violent misdemeanor, violent felony) to generate a detention/release recommendation (see Figure 1).

While this process is roughly consistent across jurisdictions, some risk assessments require a defendant interview in Stage 1, which introduces distinct forms of variation. Interviews require that defendants, pretrial officers, and correctional staff cooperate, and when that does not happen, defendants do not receive a risk score. In practice, this means that the proportion of defendants at arraignment with risk scores tends to be higher in jurisdictions that implement pretrial risk assessments without an interview component. In Virginia, pretrial interviews happen inside jail, facilitated by correctional officers who usher defendants and pretrial officers to a secure location. Sometimes, however, interviews do not happen due to the actions of pretrial or correctional officers. This has significant downstream consequences because judges may interpret the absence of a VPRAI score as evidence of noncompliance. For example, a public defender told me, "if they don't do one and there's a note that the person was not cooperative, that is used against the client. That has nothing to do with their ability to appear for court just because they didn't wanna talk to the pretrial officer." Pretrial officers have the discretion to determine whether a defendant is too intoxicated or hostile to interview, information that can later be used by prosecutors or judges to justify pretrial detention. Sometimes defendants decline interviews simply because the purpose or benefits of participating in the interview are unclear to a defendant, either due to miscommunications with pretrial officers or carelessness on the part of corrections officers. For instance, pretrial officers in one Virginian jurisdiction told me that sometimes correctional officers fail to convey the option to interview, thereby eliminating defendants' freedom of choice. Pretrial officers suspected that correctional staff falsely reported that defendants declined their interviews just to avoid the trouble of arranging them. Other times, correctional officers describe the interview process to defendants in a manner that discourages participation—for example, by asking, "do you want to talk to [pretrial agency]," without describing the process or purpose of the interview. Pretrial staff believed this promoted feelings of frustration, confusion, or suspicion in defendants, thus reducing participation rates. I spoke to one pretrial manager who attempted to resolve this issue with his sheriff, but he ultimately determined that little could be done to change correctional staff behavior, due to practical constraints and cultural norms. The considerable distance between the sheriff's office and the jail complex prevented supervisors from physically overseeing jail operations day-to-day, which allowed correctional

Figure 1: generating a VPRAI recommendation



Calculate Risk Score

Risk Factor	Points
Active community supervision	2
Felony drug, theft, or fraud charge	3
Pending charges	2
Criminal history	2
2+ FTAs	1
2+ violent convictions	1
Employed	1
History of drug abuse	2

Determine Risk Level

VPRAI Score	Risk Level
0-2	1
3-4	2
5-6	3
7-8	4
9-10	5
11-14	6

Apply Category to DMF

Level	Non-Violent Misdemeanor	DUI	Non-Violent Felony	Violent Misdemeanor	Violent Felony
1	Release	Release	Release	Release	Release
2	Release	Release	Release	Release	Release
3	Release	Release	Release	Release	Detain
4	Release	Release	Release	Release	Detain
5	Release	Release	Release	Detain	Detain
6	Detain	Detain	Detain	Detain	Detain

staff to minimize their workloads by reporting that all defendants declined their interviews. Without anyone to monitor the conversations between correctional staff and defendants, the pretrial agency had no grounds to intervene. All local government agencies depend on external agencies to varying degrees, perhaps for funding or operational support, but pretrial officers cannot perform any aspect of their role without the cooperation of correctional staff. This creates a power imbalance, where pretrial officers must perpetually seek buy-in and avoid conflict.

While this may suggest that pretrial risk assessments without an interview component yield more standardization, the trade offs are not so straightforward. Take the Public Safety Assessment (PSA). Among other risk factors, it considers "prior violent conviction," and among other outcomes, it predicts "new violent criminal arrest." But those categories are broad and do not map neatly onto local criminal codes. Ultimately, the task of labeling specific crimes as either violent or nonviolent falls to an individual, or to a state or local entity. I spoke to the pretrial manager of a rural county who described how this ambiguity undermines the goal of standardization. Different localities across her state compute PSA scores differently according to their local infrastructure, which has given rise to multiple interpretations of the same risk factors. Whereas some counties receive PSA scores from the state, others receive risk factor values and compute PSA scores on their own. Still other counties receive raw data and complete the entire process on their own, with little guidance or oversight. The pretrial manager shared her frustration with the lack of clarity:

> We need to make sure that all the jurisdictions are entering data the same way. Right now, the [state criminal justice agency] has told all the jurisdictions to just use the [software] fields in a way that makes the most sense to them. That's nuts. Like we have to have really defined ways of how to enter data into each field because otherwise we are going to be comparing apples to oranges.

Even states that attempt to standardize the interpretation of risk factors at the local level only get so far, particularly if risk assessment administration requires an interview. For example, Virginia's VPRAI manual—while thorough—still affords pretrial workers significant latitude. I spoke to one pretrial officer about the employment status risk factor:

> I would say there's an argument on how that one is scored … What does employment mean? And their threshold is that they have to be working for 20 hours a week. And in the training, it's just that simple. So to be considered like you met employment criteria you're employed for 20 hours a week, or you are primary child caregiver or caregiver for like an adult or parent or something like that … Like, no—I'm no Uncle Sam, man. I don't care, honestly, if that person is paying their taxes the way that we value. Like, it comes out of my check. I don't care if they have a 1099 or any of that. Are they working? Do they have a job for 20 hours a week that occupies their time? Because in my mind, I think the gist of that is it's more about their structured use of time than it is—whether or not that person's paying their income tax … So that's—it's a convoluted mess.

I conducted this initial interview via Zoom, and during a subsequent site visit, his office inadvertently exposed the pitfalls of inter- and intra-agency communication. He repeated his interpretation of the state's employment criteria, but I later spoke to another pretrial officer in his agency who explained that defendants are required to work at least 30 hours per work to be classified as employed. "Odd jobs," he said, do not count. Even within the same pretrial agency, staff interpret state policies differently. These findings foreshadowed a similar inconsistency I observed with respect to the risk factor history of drug abuse. To define history of drug abuse, Virginia's VPRAI manual provides a set of examples:

> Indications of history of drug abuse could include (a) previously used illegal substance(s) repeatedly, distinguishing from short-term experimental use; (b) admits to previously abusing illegal or prescription drugs; (c) the criminal history contains drug related convictions; and (d) the defendant received drug treatment in the past.

But without bounding "the past," pretrial staff are left to infer meaning, ultimately imposing their own definitions. One pretrial officer focuses on explicit evidence of drug use: attending drug counseling, receiving drug convictions, or testing positive for drugs while on supervision are all interpreted as having a history of drug abuse. However, a pretrial officer in a different jurisdiction relies solely on her conversations with defendants to glean their patterns of drug use. She pointed out that certain questions require follow-up, telling me that "a while" to a pretrial officer can mean 2-3 years, but to an addict can mean 3 days. Getting clear or truthful answers can be difficult, "a lot of it has to do with the demeanor of the person giving the questions."

Sometimes, implementation variation at this stage of the pretrial process arises from purposeful manipulation on the part of pretrial officers. Speaking about employment status and history of drug abuse, one pretrial officer told me, "if they are on the cusp of being not recommended and recommended, I will go back and play with it and change it to see where it falls, to see if it actually makes an impact and then, I can better fine tune my decision. Usually, it doesn't have an impact overall but there have been a few cases where it does and I will give it to them if I think it is legitimate and then sometimes, I don't. It just depends."

While some amount of implementation variation in phase 1 is unavoidable, state and local policies can encourage or suppress it. In particular, highly centralized governance structures and more precisely defined protocols are likely to result in greater uniformity. In Kentucky, PSA scores are calculated in state-run call centers. Consolidating this process within a single office makes it easier to conduct training sessions, disseminate information, and exercise oversight. In contrast to Kentucky, Virginia cedes far more control to localities to conduct pretrial programming. Localities administer risk instruments independently, with little inter-jurisdictional visibility, but are nonetheless expected to operate with uniformity. That said, imposing standardization as Kentucky does in phase 1 many only defer implementation variation to later phases of the pretrial process.

3.3 Stage 2: Making a Pretrial Recommendation

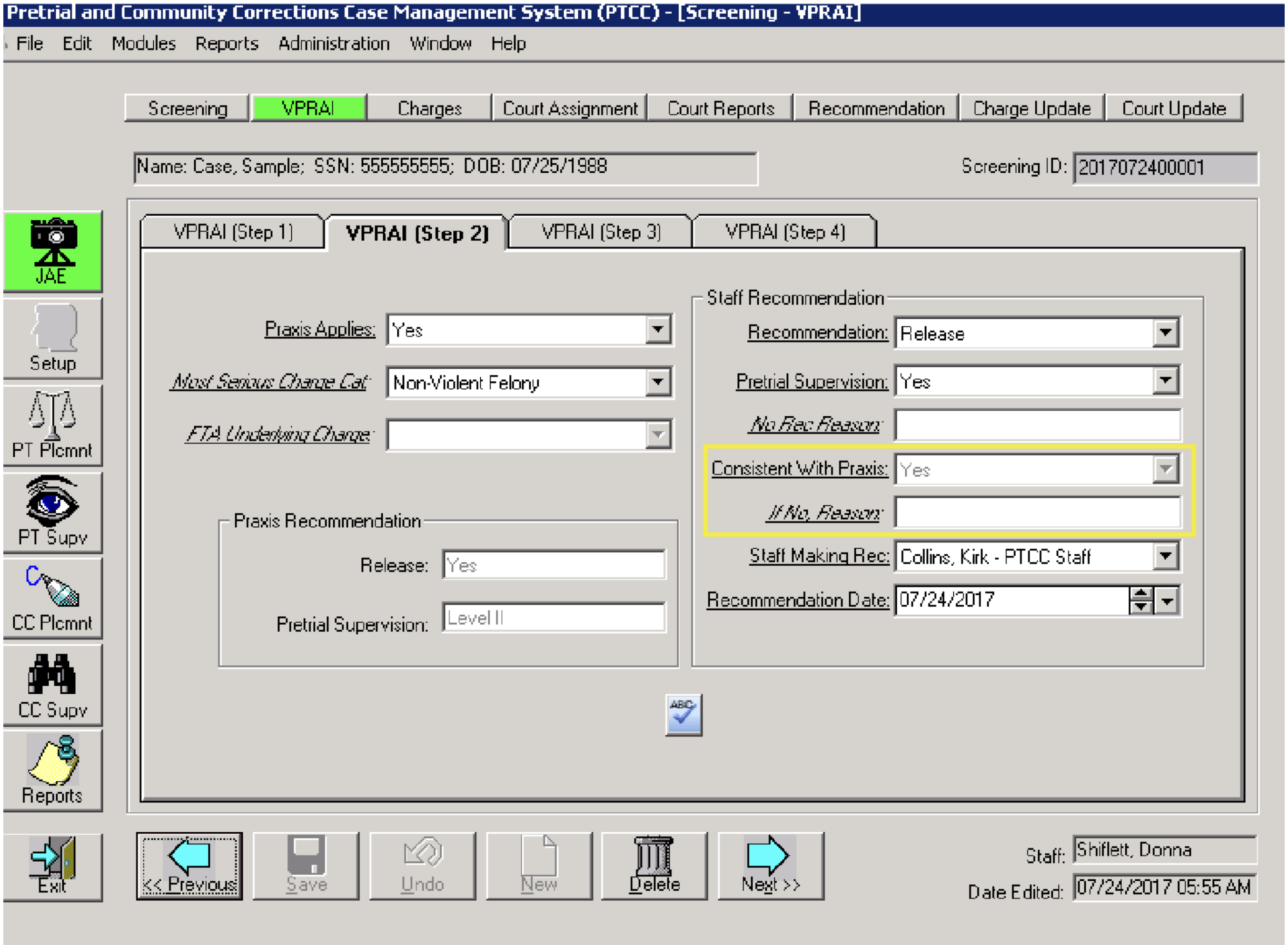
In the next phase of the pretrial process, pretrial officers prepare a report, which generally contains a defendant's criminal history, risk factor values, and the risk assessment recommendation. This report is circulated at arraignment, although in Virginia, judges and prosecutors have access to the report in advance. In certain jurisdictions, pretrial officers have the authority to override the risk assessment recommendation. In Virginia, this occurs by indicating a staff recommendation in the state's Pretrial and Community Corrections Case Management System (PTCC) (See Figure 2). When that happens, staff are asked to justify their override in a short paragraph (see Figure 3). Pretrial officers can also communicate "mitigating/aggravating circumstances" to court actors without performing a formal override. State guidelines require that pretrial officers not override more than 15 percent of the risk assessment's recommendations, so written explanations to the court are one way to circumvent that rule. Not all officers make use of override privileges. One pretrial officer told me, "[w]hen the tool is administered as it is, you should have the best results. I am a firm believer that assessments should not be overridden because if you [administer] the tool with fidelity, the tool will give you the best result." In contrast, a pretrial officer in a neighboring jurisdiction made such liberal use of the policy that the state questioned his behavior directly. He told me that while he was never sanctioned, he understood that his compliance rate was "reviewed from time to time."

I heard different justifications for performing overrides. Some pretrial officers fundamentally disagreed with the risk assessment's definition of public safety, and they overrode recommendations that violated their notions of public safety. For example, several pretrial officers believed that regardless of the pretrial risk assessment's recommendation that drug addicts were safer in jail than at home. One pretrial officer told me, "I think your professional judgment has to be put into that because [the] risk score is not always gonna capture what's going on with the defendant. If I know that you're a daily heroin user, you are homeless, you have no job, no ties, no friends, no family, what am I supposed to do with you? Because if you're—if I let you out, you're gonna go use. You're just gonna end up right back here or dead." While risk assessments generally define threats to public safety in terms of rearrest or failure to appear, her definition included self-harm, leading her to override release recommendations when she believed that defendants were a risk to themselves. Another officer justified overriding decisions about defendants who did not have any criminal history out of concern over too much uncertainty:

> For an example, someone comes in and it is an arson charge. This is their first charge. They don't have a criminal history. It's arson or something like that and the tool is saying, you know, release. We would say release however we're going to say there's a caveat. We're recommending release with a community release plan. We wanna partner with the Community Services Board and basically have this person on a contract to say that they're gonna follow this release plan. Because this is an out-of-the-norm type situation where we don't have any background, we don't have any information, like, we don't have any history.

Other pretrial officers justified overrides in cases where they might otherwise experience public backlash. This happened most frequently when the pretrial risk assessment recommended release for individuals charged with particular crimes, including rape, murder,

Figure 2: recording a formal override in PTCC



Name: Case, Sample; SSN: 555555555; DOB: 07/25/1988

Screening ID: 2017072400001

Praxis Applies: Yes

Most Serious Charge Cat: Non-Violent Felony

FTA Underlying Charge:

Praxis Recommendation

Release: Yes

Pretrial Supervision: Level II

Staff Recommendation

Recommendation: Release

Pretrial Supervision: Yes

No Rec Reason:

Consistent With Praxis: Yes

If No, Reason:

Staff Making Rec: Collins, Kirk - PTCC Staff

Recommendation Date: 07/24/2017

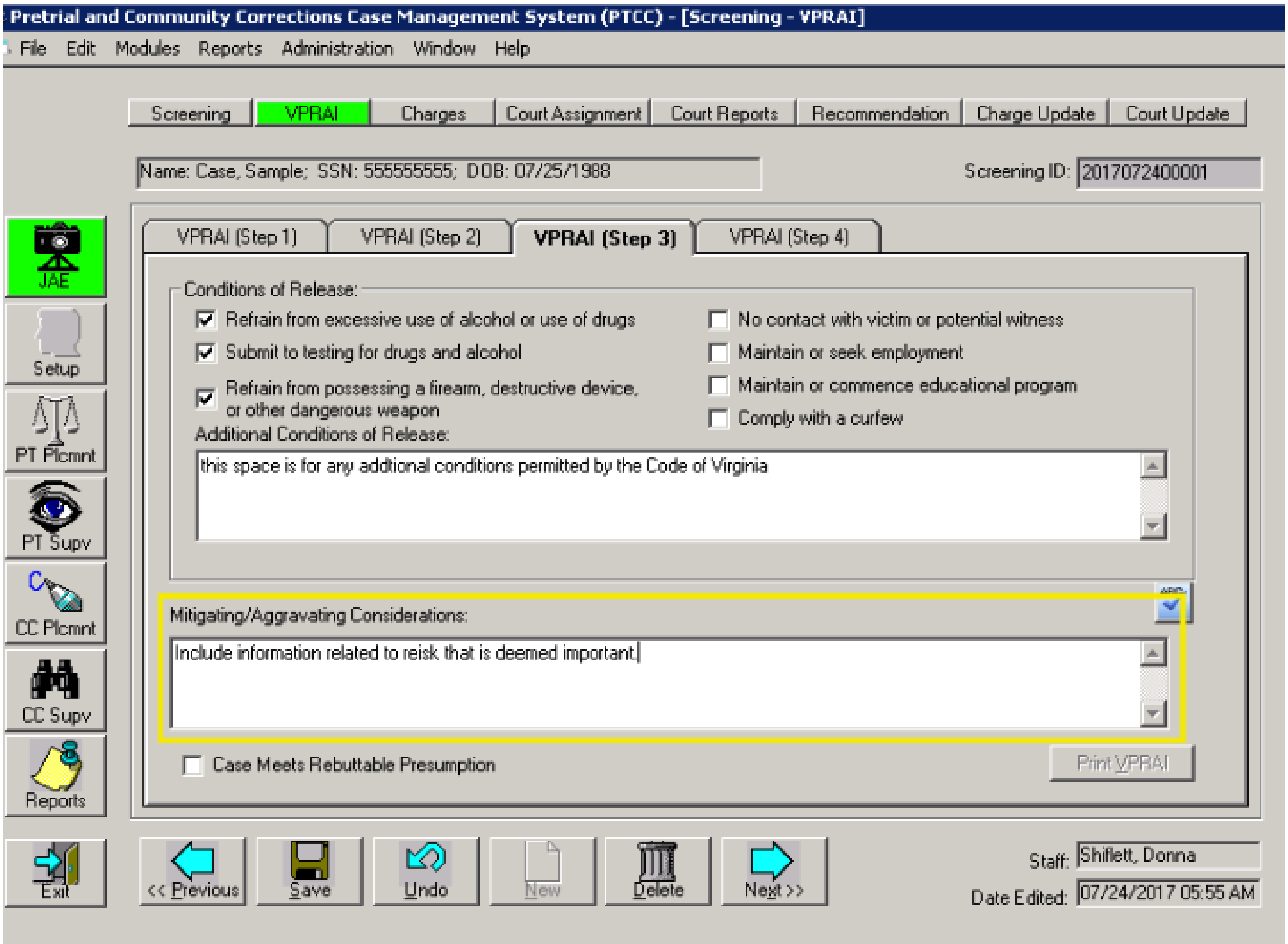
Staff: Shiflett, Donna

Date Edited: 07/24/2017 05:55 AM

or child sexual abuse. One pretrial officer told me, "I also have to keep in mind the optics of it, if someone is brought in—let's take child sex offenders, for instance—if they are charged with one of those types of crimes, usually, a lot of time, those individuals don't have a record but I don't recommend bond for those individuals based just on the optics of the charge." Some also feared liability in the event that defendants committed crimes upon release: "[i]f somebody is in here on rape charges and that is a violent felony offense, I, from a pretrial standpoint, will just say no bond because I don't want it to come back on me, well, pretrial said release this guy and then, he goes out and he rapes somebody else. And that is more something that I want the judge to override my decision; whereas, I am overriding a computer, let the judge override me so, that way, it can't come back on pretrial." Fears of liability were not uncommon. One officer described her pretrial agency's creative policy of including a disclaimer on each pretrial report, explaining that VPRAI recommendations are auto-generated, and do not reflect the personal or professional opinions of anyone in the agency. This policy was adopted after many "uncomfortable" interactions between pretrial officers and prosecutors, in which pretrial officers felt personally blamed for risk assessment recommendation. This most often happened in cases involving domestic violence or child sex offenses.

I found that where formal override policies did not exist, pretrial staff simply executed overrides in more creative ways. For example, one participant told me, "[T]he woman who runs our pretrial services, she's worked really hard at establishing a good rapport with our district court judges. She goes to their meetings and brings them cake, and she also has found a way to include a narrative written by her pretrial supervisors that gets included with all of the documentation that district court judges review in a case. And those narratives are not based in science or best practices." Elsewhere, in Kentucky, the State Supreme Court directs pretrial officers to prepare investigative reports that assess defendants' "financial resources, past conduct, history relating to drug or alcohol abuse, criminal history, and record concerning appearance at court proceedings." These reports give pretrial officers significant latitude to persuade judges to override risk assessment recommendations,

Figure 3: recording mitgating/aggravating considerations in PTCC



Name: Case, Sample; SSN: 555555555; DOB: 07/25/1988

Screening ID: 2017072400001

Conditions of Release:

☑ Refrain from excessive use of alcohol or use of drugs

☑ Submit to testing for drugs and alcohol

☑ Refrain from possessing a firearm, destructive device, or other dangerous weapon

☐ No contact with victim or potential witness

☐ Maintain or seek employment

☐ Maintain or commence educational program

☐ Comply with a curfew

Additional Conditions of Release:

this space is for any addtional conditions permitted by the Code of Virginia

Mitigating/Aggravating Considerations:

Include information related to reisk that is deemed important.

☐ Case Meets Rebuttable Presumption

Staff: Shiflett, Donna

Date Edited: 07/24/2017 05:55 AM

by highlighting certain defendant characteristics that make either release or detention seem more appropriate. So despite Kentucky's highly centralized operation, discretion nonetheless pervades the pretrial process.

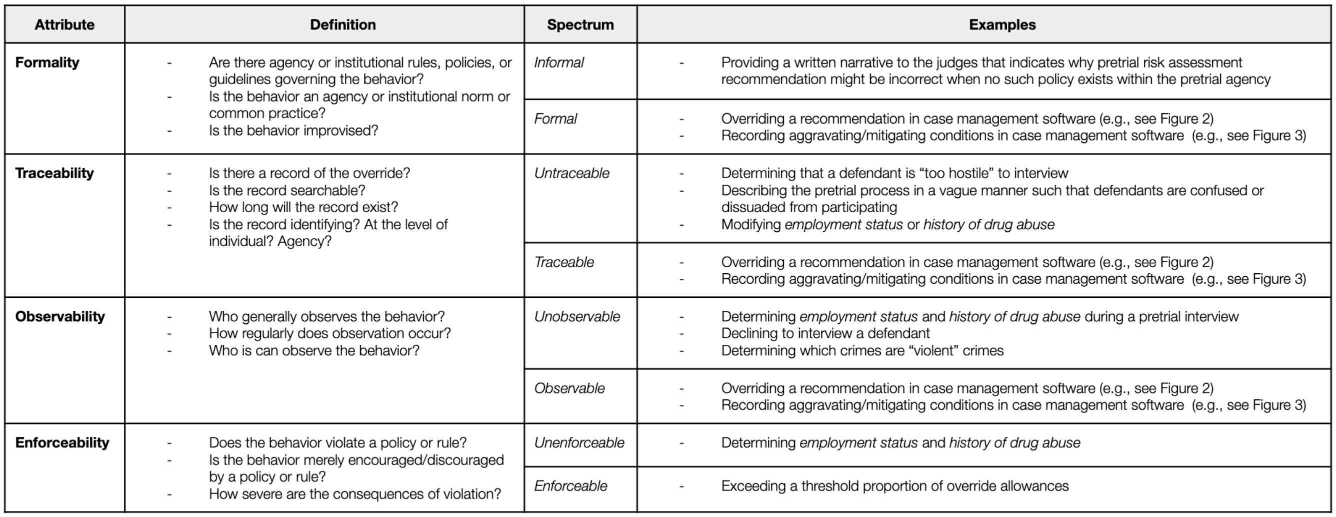
4 DISCUSSION

It is clear from my fieldwork that discretion pervades the pretrial process, even in the presence of algorithmic guardrails. Override behaviors threaten to undermine standardization, introduce opportunities for discrimination, and conceal critical moments of pretrial decision-making from the public by shifting them from courtrooms to jail cells. One way that states and localities attempt to guide and supervise discretion among pretrial officers is to specify override policies. However, even when formal override policies exist, pretrial officers use other methods to modify risk factor values and risk assessment recommendations.

Below I describe four dimensions of override behaviors: formality, traceability, observability, and enforceability. Formality captures the extent to which the behavior is governed by convention, rules, or policies. Traceability refers to the existence of a record of the behavior and whether or not the record is associated with a particular pretrial officer. Observability refers to visibility. Who witnesses the behavior? Do judges? Can the public? Enforceability indicates whether or not the behavior violates an explicit rule. Of course, this list is not exhaustive. I focus on these attributes in particular because they seemed to guide decision-making practices among pretrial officers.

The taxonomy does not say anything about either the motivations of pretrial officers or whether or not certain behaviors nudge defendants closer to either pretrial detention or release. The pretrial officers I spoke to were motivated to intervene in the risk assessment administration process by a variety of factors, such as fears of public backlash or concern for defendants' personal safety. Similar motivations may lead two pretrial officers to make very different decisions. For example, I described how one pretrial officer tended to recommend detention for people with substance use disorders, but a different pretrial officer might instead recommend release and drug abuse counseling.

Figure 4: Attributes of Override Behaviors



Attribute	Definition	Spectrum	Examples
Formality	- Are there agency or institutional rules, policies, or guidelines governing the behavior? - Is the behavior an agency or institutional norm or common practice? - Is the behavior improvised?	*Informal*	- Providing a written narrative to the judges that indicates why pretrial risk assessment recommendation might be incorrect when no such policy exists within the pretrial agency
		Formal	- Overriding a recommendation in case management software (e.g., see Figure 2) - Recording aggravating/mitigating conditions in case management software (e.g., see Figure 3)
Traceability	- Is there a record of the override? - Is the record searchable? - How long will the record exist? - Is the record identifying? At the level of individual? Agency?	*Untraceable*	- Determining that a defendant is "too hostile" to interview - Describing the pretrial process in a vague manner such that defendants are confused or dissuaded from participating - Modifying *employment status* or *history of drug abuse*
		Traceable	- Overriding a recommendation in case management software (e.g., see Figure 2) - Recording aggravating/mitigating conditions in case management software (e.g., see Figure 3)
Observability	- Who generally observes the behavior? - How regularly does observation occur? - Who is can observe the behavior?	*Unobservable*	- Determining *employment status* and *history of drug abuse* during a pretrial interview - Declining to interview a defendant - Determining which crimes are "violent" crimes
		Observable	- Overriding a recommendation in case management software (e.g., see Figure 2) - Recording aggravating/mitigating conditions in case management software (e.g., see Figure 3)
Enforceability	- Does the behavior violate a policy or rule? - Is the behavior merely encouraged/discouraged by a policy or rule? - How severe are the consequences of violation?	*Unenforceable*	- Determining *employment status* and *history of drug abuse*
		Enforceable	- Exceeding a threshold proportion of override allowances

In response to these forms of frontline worker discretion jurisdictions *could* choose to exercise more oversight over the process of pretrial risk assessment administration. Kentucky's pretrial system is highly centralized and offers one example of how to enforce greater standardization in this area. The state administers the PSA, which does not require a pretrial interview, thus blocking one discretionary pathway entirely. Additionally, state-run call centers are responsible for calculating risk scores, making oversight easier and minimizing the variation associated with having many local agencies, each with different rules and operations. Of course, this alternative also introduces trade offs. Suppressing discretion requires more intense worker surveillance and minimizes worker autonomy. The empowerment of street-level bureaucrats has been recognized as an important aspect of improving organizational outcomes [22] and client responsiveness [6]. Beyond that, the clinical judgment of pretrial officers can be useful when a defendant's case is unusual, instances where a case reflects a set of covariates that is relatively rare in the training data. And, of course, eliminating pretrial officer discretion does not solve outcome disparities that result from risk factors beyond their control, or from judicial discretion.

5 CONCLUSION

In this paper, I provided an ethnographic account of risk assessment administration, highlighting a group of people and stage of the pretrial process that has remained understudied despite their outsized influence over system outcomes. I describe the personal, professional, and organizational dynamics that lead pretrial officers to override risk assessment recommendations. I also name four attributes of override behaviors: formality, traceability, observability, and enforceability. Lastly, I offer additional empirical evidence that pretrial risk assessments are unlikely to guarantee racial or economic equity or decarceration in the long term.

In the shorter term, understanding the conditions that make certain override behaviors more or less likely is useful for creating more effective state policies and pretrial agency rules. If pretrial risk assessments are intended to promote standardization and decarceration, agencies will have to think carefully about the local contexts in which they are deployed and how they promote or undermine those goals.

In the longer term, I provide further evidence that jurisdictions should look to systemic reforms rather than rely on so-called evidence-based, "system-conserving" [27]interventions. Policies like presumptive pretrial release, "value-added" pretrial programs that provide wrap-around services and do not communicate with courts, and redirection programs that divert people to mental health and addiction services *pre*-arrest are all places to start.

ACKNOWLEDGMENTS

Thank you to the many pretrial officers and experts I spoke to and shadowed for their support throughout this project. Thank you Karen Levy, Solon Barocas, my anonymous reviewers, and members of Stanford University's Communication Works in Progress (CWIP) group for their invaluable feedback and advice on this project. This work was supported by the Horowitz Foundation for Social Policy, The John D. and Catherine T. MacArthur Foundation, the Microsoft Ada Lovelace Fellowship, and UCLA's Center for Critical Internet Inquiry.

REFERENCES

[1] Alex Albright. 2019. If you give a judge a risk score: evidence from Kentucky bail decisions. *Law, Economics, and Business Fellows' Discussion Paper Series* 85 (2019).

[2] Ali Alkhatib and Michael Bernstein. 2019. Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[3] Utsav Bahl, Chad M Topaz, Lea Obermüller, Sophie Goldstein, and Mira Sneirson. 2023. Algorithms in Judges' Hands: Incarceration and Inequity in Broward County, Florida. (2023).

[4] Emily Bosk and Megan Feely. 2020. The goldilocks problem: Tensions between actuarially based and clinical judgment in child welfare decision making. *Social Service Review* 94, 4 (2020), 659–692.

[5] Sarah Brayne and Angèle Christin. 2021. Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems* 68, 3 (2021), 608–624.

[6] Evelyn Z Brodkin. 2012. Reflections on street-level bureaucracy: past, present, and future.

[7] Timothy P Cadigan and Christopher T Lowenkamp. 2011. Implementing risk assessment in the federal pretrial services system. *Fed. Probation* 75 (2011), 30.

[8] Ursula Castellano. 2009. Beyond the courtroom workgroup: Caseworkers as the new satellite of social control. *Law & policy* 31, 4 (2009), 429–462.

[9] Steven L Chanenson and Jordan M Hyatt. 2016. The use of risk assessment at sentencing: Implications for research and policy. *Hyatt, JM & Chanenson, SL (2016). The Use of Risk Assessment at Sentencing: Implications for Research and Policy. Bureau of Justice Assistance, Washington, DC* (2016).

[10] Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. How child welfare workers reduce racial disparities in algorithmic decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–22.

[11] Angèle Christin. 2017. Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4, 2 (2017), 2053951717718855.

[12] Jennifer E Copp, William Casey, Thomas G Blomberg, and George Pesta. 2022. Pretrial risk assessment instruments in practice: The role of judicial discretion in pretrial reform. *Criminology & Public Policy* 21, 2 (2022), 329–358.

[13] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova. 2020. A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[14] Tarleton Gillespie. 2006. Designed to 'effectively frustrate': Copyright, technology and the agency of users. *New Media & Society* 8, 4 (2006), 651–669.

[15] Vern L Glaser, Neil Pollock, and Luciana D'Adderio. 2021. The biography of an algorithm: Performing algorithmic technologies in organizations. *Organization Theory* 2, 2 (2021), 26317877211004609.

[16] Kosuke Imai, Zhichao Jiang, D James Greiner, Ryan Halen, and Sooahn Shin. 2023. Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment. *Journal of the Royal Statistical Society Series A: Statistics in Society* 186, 2 (2023), 167–189.

[17] Summer Rachel Jackson and Katherine Cissel Kellogg. 2023. Triadic advocacy work. *Organization Science* 34, 1 (2023), 456–483.

[18] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. Human decisions and machine predictions. *The quarterly journal of economics* 133, 1 (2018), 237–293.

[19] Guillaume Latzko-Toth, Johan Söderberg, Florence Millerand, and Steve Jones. 2019. Misuser Innovations The Role of "Misuses" and "Misusers" in Digital Communication Technologies. *digitalSTS: A Field Guide for Science & Technology Studies* (2019), 393.

[20] Karen EC Levy. 2015. The contexts of control: Information, power, and truck-driving work. *The Information Society* 31, 2 (2015), 160–174.

[21] Douglas B Marlowe, Timothy Ho, Shannon M Carey, and Carly D Chadick. 2020. Employing standardized risk assessment in pretrial release decisions: Association with criminal justice outcomes and racial equity. *Law and Human Behavior* 44, 5 (2020), 361.

[22] John Petter, Patricia Byrnes, Do-Lim Choi, Frank Fegan, and Randy Miller. 2002. Dimensions and patterns in employee empowerment: Assessing what matters to street-level bureaucrats. *Journal of public administration research and theory* 12, 3 (2002), 377–400.

[23] Dasha Pruss. 2023. Ghosting the Machine: Judicial Resistance to a Recidivism Risk Assessment Instrument. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 312–323.

[24] Jennifer Skeem, Nicholas Scurich, and John Monahan. 2020. Impact of risk assessment on judges' fairness in sentencing relatively poor defendants. *Law and human behavior* 44, 1 (2020), 51.

[25] CarlyWill Sloan, George Naufal, and Heather Caspers. 2023. The effect of risk assessment scores on judicial behavior and defendant outcomes. *Journal of Human Resources* (2023).

[26] Megan Stevenson. 2018. Assessing risk assessment in action. *Minn. L. Rev.* 103 (2018), 303.

[27] Megan T Stevenson. 2023. Cause, Effect, and the Structure of the Social World. *Available at SSRN* (2023).

[28] Megan T Stevenson and Jennifer L Doleac. 2022. Algorithmic risk assessment in the hands of humans. *Available at SSRN 3489440* (2022).

[29] Linda A Teplin, Karen M Abram, and Gary M McClelland. 1996. Prevalence of psychiatric disorders among incarcerated women: I. Pretrial jail detainees. *Archives of general psychiatry* 53, 6 (1996), 505–512.

[30] Celeste Watkins-Hayes. 2019. *The new welfare bureaucrats: Entanglements of race, class, and policy reform.* University of Chicago Press.