
Adverse Impacts of Mental Health Needs Assessment on Jail Outcomes

Evidence from Transition Age Youth and Adults

Scott Cunningham
Jonathan A. Seward
Karen Clay
Vivian S. Vigliotti

ABSTRACT

We investigate the causal impact of jail mental health needs scores, assigned by clinicians, on youth and adult outcomes. Utilizing an instrumental variable (IV) approach leveraging quasi-random assignment of intake clinicians, we find that worse scores increase incarceration duration and suicide attempts in both groups and recidivism in adults. For the marginal person, worse scores may lead to consequences like prolonged jail stays, which counteract potential benefits. We discuss possible mechanisms, including the existence of a mental health docket for misdemeanors, and effects of scores on bond posting, mental health care, type of housing within the jail, and length of stay.

Scott Cunningham is the Ben H. Williams Professor of Economics at Baylor University (scunning@gmail.com). Jonathan A. Seward is a Postdoctoral Research Associate in the Economics Department at Baylor University and holds a joint appointment as Adjunct Staff in the Economics, Sociology, and Statistics

(continued on next page)

[Submitted December 2023; accepted August 2023]; doi:10.3368/jhr.1222-12713R2

JEL Classification: I12, K14, I14, and I18

ISSN 0022-166X E-ISSN 1548-8004 © 2024 by the Board of Regents of the University of Wisconsin System

 Supplementary materials are available online at: <https://jhr.uwpress.org>.

 This open access article is distributed under the terms of the CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>) and is freely available online at: <https://jhr.uwpress.org>.

Special Issue: Causes and Consequences of Child Mental Health

Janet Currie, Guest Editor

Sponsored by the NOMIS Foundation and the Center for Health and Wellbeing at Princeton University.

Scott Cunningham <https://orcid.org/0000-0002-0602-2089>

Jonathan A. Seward <https://orcid.org/0000-0001-7352-8170>

Karen Clay <https://orcid.org/0000-0003-3978-6656>

Vivian S. Vigliotti <https://orcid.org/0000-0001-5436-2566>

Special Issue: Causes and Consequences of Child Mental Health
THE JOURNAL OF HUMAN RESOURCES • 59 • Supplement

I. Introduction

Transition age for youth—ages 17–25—is a period of emergence or worsening of mental health problems and of increasing interaction with the criminal justice system.¹ As a result, a sizeable share of transition age youth in jail have serious mental health problems either alone or comorbid with alcohol or substance abuse (Bronson and Berzofsky 2017). Several outcomes are of significant policy concern for transition age youth while in jail, including length of incarceration (that is, length of stay), suicide attempts, and upon release, recidivism. Length of stay is of concern because longer stays can impact mental health and be criminogenic. Suicide is the leading single cause of death in U.S. jails, accounting for 31 percent of all deaths. Even when suicide attempts do not lead to death, they are an important marker of distress (Carson and Cowhig 2020). Recidivism is of concern because it makes it less likely that youth will successfully transition to stable employment, in addition to crime's other costs. While there is sizeable literature on length of stay in jail or prison and recidivism, primarily using judge leniency designs, there is limited causal literature on mental health screening and how it might relate to these policy-relevant outcomes. Further, the literature has not focused on youth, despite the importance of this subpopulation.²

This work provides new causal evidence on the effect of mental health needs scores assigned by jail clinicians on the outcomes of transition age youth and adults in jail.³ Our analysis draws on administrative data from a large urban county jail that includes the universe of inmates booked in the county between 2016 and 2019. In the jail, the first point of contact with mental health resources is the intake clinicians, who assess each inmate within 36 hours of their arrival. After a 15–30-minute interview, intake clinicians write a summary of the inmate's functioning and various risks, including a mental health

Department at RAND Corporation. Karen Clay is a Professor of Economics and Public Policy at Carnegie Mellon University's Heinz College of Information Systems and Public Policy and an NBER Fellow. Vivian S. Vigliotti is a Health Outcomes Researcher II, Team Lead at Yale New Haven Health System. Authorship has been randomized. The authors thank jail officials for the data and detailed information on operations in the jail, Daniel Nagin for early comments and suggestions, as well as Alison Cueller, Gordon Dahl, and other participants at the Causes and Consequences of Child Mental Health conference hosted by the Center for Health and Wellbeing at Princeton University. The authors gratefully acknowledge financial support from Baylor University and Carnegie Mellon University. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The raw data supporting the findings of our study are not publicly accessible. This is due to the data use agreement in place with the organization that granted access to the data. This agreement explicitly restricts the sharing of raw data. The authors are able to advise interested researchers on the process to acquire the data themselves (scunning@gmail.com).

-
1. Throughout the paper, we may refer to transition age youth, who are 17–25 in our study, as youth.
 2. Younger individuals are also important, but it can be particularly difficult to get data on this population. An important paper in this literature is Cuellar, McReynolds, and Wasserman (2006).
 3. It is worth emphasizing that this paper does not examine the effects of preexisting mental health on outcomes or preexisting mental health comorbidities with substance use disorder on outcomes. We have very limited data on substance use in this setting. This paper also does not examine the effect of being assigned to mental health care in jail. This paper does focus on the effect of quasi-random mental health needs scores, as it examines the causal effect of being labeled (being given a mental health needs score) as having moderate or severe mental health needs.

needs score. We collapse the four-point mental health needs score into two bins, none/mild (0) and moderate/severe (1).

Being labeled with a moderate or severe mental health needs score, conditional on underlying mental health, may affect outcomes through several mechanisms. It can shift misdemeanor cases to the mental health docket, which is associated with significant delays. Having a moderate to severe mental health needs score may affect judge issuance of bonds and the posting of bond by friends and family, leading to an even longer length of stay. Such scoring could increase suicide attempts through its impact on length of stay, despair, stigma, medication, or type of housing within the jail. Such delays could affect recidivism as research has linked length of stay to recidivism (Gupta, Hansman, and Frenchman 2016; Heaton, Mayson, and Stevenson 2017; Dobbie, Goldin, and Yang 2018; Leslie and Pope 2018; Loeffler and Nagin 2022). Assignment to the mental health docket may also influence recidivism through reductions in expected punishment.

Estimating the causal impact of receiving a worse (moderate to severe) mental health needs score on inmate outcomes is complicated by two factors. First, an inmate's score is endogenously assigned to people with moderate to severe mental health outcomes in general and thus naive comparisons to inmates with even a slightly lower score will be contaminated with both supply and demand effects. Second, the set of interventions attached to moderate to severe mental health needs scores are difficult to disentangle from one another, if they are even observed.

Our instrumental variables (IV) estimation strategy takes advantage of the quasi-random assignment of intake clinicians at booking to correct for selection bias associated with the mental health needs score. We measure clinician leniency using a leave-one-out residualized measure, which is the average tendency of the randomized intake clinician to score inmates at booking as moderate/severe based on other inmates seen at intake in the same month. This IV strategy is similar to Aizer and Doyle (2015) and Dobbie, Goldin, and Yang (2018), as well as other papers in criminal justice using a leniency design.⁴ Hence, the question we are addressing is whether the marginal person, who is assigned a higher mental health needs score solely because they are assigned to a clinician who gives higher scores, benefits from receiving this higher score. If the marginal person needs mental health treatment and a higher score enables them to get it, then a higher score could potentially be beneficial. However, if the mental health needs score has unforeseen consequences, such as extending jail time without facilitating access to effective treatment, it could potentially undermine or even negate any medical benefits derived from the classification.

We have two main findings. First, IV estimates suggest that quasi-random assignment of moderate to severe mental health needs scores for youth increases length of stay in jail and suicide attempts but did not have a statistically significant effect on recidivism.⁵ Being classified as having moderate/severe mental health needs increases length of stay

4. This IV approach leverages quasi-random and systematic "tendencies" towards leniency among evaluators, like judges, and was first suggested by Imbens and Angrist (1994) in their seminal work on the local average treatment effect.

5. In this study, recidivism is defined as an individual reentering into the jail administrative data set post booking. Reentry means they were arrested, and we are using this rearrest measure as a proxy for recidivism. To address censoring related to recidivism, we examine three subsamples in which we truncate the sample to match the recidivism period. That is, when examining recidivism within one year, the sample is limited to individuals with a release date at least one year before the end of the sample period. Similarly, the sample is restricted for recidivism within 18 months and two years.

for youth jailed for misdemeanors (7.7 days) and felonies (54.1 days). It also increases suicide attempt rates for youth jailed for misdemeanors both overall (4.0 percentage points) and per day (0.7 percentage points per day). In all cases, there are large increases relative to the complier means.

Second, having a worse (moderate to severe) mental health needs score leads to increases in length of stay in jail, suicide attempts, and recidivism for adults. Being classified as having moderate/severe mental health needs increases length of stay for adults jailed for felonies (22.6 days). It also increases suicide attempt rates for adults jailed for misdemeanors and felonies overall (1.8 and 1.8 percentage points, respectively) and per day (0.4 and 0.1 percentage points per day, respectively). Both the length of stay and the suicide outcomes are smaller in magnitude than comparable outcomes for youth, suggesting that the adverse impact on the adult population is somewhat smaller. Being classified as having moderate/severe mental health needs increases recidivism by adults jailed for misdemeanors (25 percentage points within 1 year). As with youth, in all cases there are large increases relative to the complier means.

This paper contributes to three literatures. The first is the literature on mental health of youth and adult inmates in jail. Drawing on the 2011–2012 National Inmate Survey, Bronson and Berzofsky (2017) report detailed information on mental health status of individuals in jail by demographic characteristics, including age. They find that youth in jail have similar levels of serious psychological distress to other age groups in jail.⁶ This study builds on the literature by providing new causal evidence on how mental health needs scores relate to youth and adult outcomes in jail, including length of stay, suicide attempts, and recidivism.⁷

The second is the literature on length of stay and recidivism. On the policy side, the Stepping Up Initiative is focused on reducing length of stay and recidivism of people with mental illness in jail (NACO 2018). This paper is closest to the subset of the literature that examines pre-trial incarceration on recidivism using quasi-random case assignment across the judges setting bail (Gupta, Hansman, and Frenchman 2016; Heaton, Mayson, and Stevenson 2017; Dobbie, Goldin, and Yang 2018; Leslie and Pope 2018; Loeffler and Nagin 2022). The literature finds adverse effects of longer pre-trial incarceration on recidivism. We extend the literature by examining the causal effect of mental health needs scores on length of stay and recidivism.

The third is the literature on suicide in prisons and jails. In a recent meta-analysis of 77 descriptive studies, Zhong et al. (2021) described the clinical, institutional, and criminological factors associated with suicide in prison. They note that psychiatric diagnosis, suicidal ideation, and single cell occupancy are potentially modifiable. Our paper contributes to this literature by examining the causal effect of mental health needs scores on suicide attempts.

In the following, Section II includes background information on mental illness in youth and details on the operations of our large urban county jail. Section III discusses the data and provides descriptive analysis of the mental health needs score used to

6. See also Kaba et al. (2015), who examine a large data set of individuals entering New York City jails for the first time.

7. A sizeable related literature in economics examines the relationship between measures of mental health or access to mental health care and future incarceration. See Chatterji and Cuellar (2006); Fletcher and Wolfe (2009); Anderson, Cesur, and Tekin (2015); Heller et al. (2017); Bondurant, Lindo, and Swensen (2018); Jácóme (2020); and Deza, Maclean, and Solomon (2022).

construct our binary mental health needs score. Section IV presents the research design. Section V discusses the results of the IV analysis of the mental health needs score on length of stay, suicide attempts, recidivism, and a range of robustness checks. Section VI highlights the policy implications of our results and the need for further work on outcomes in jail.

II. Background

This section begins with a discussion of mental illness and transition age youth. We then provide some background information on the large urban county jail and how inmates are processed. The final subsection discusses national data on mental illness and comorbidities.

A. Mental Illness of Transition Age Youth

Transition age youth, individuals aged 17–25 years old, sit at the confluence of two sets of risks—both a period of emergence or worsening of mental health issues and a period of increasing interaction with the criminal justice system. This period represents a delicate moment in the lives of these individuals, as the worsening symptoms may lead to an arrest prior to reduction in symptoms.

Regarding the first set of risks, a recent global meta-analysis found one-third of individuals have their first mental disorder before age 14, with most being neurodevelopmental. The modal age of onset for substance use, personality, mood, and schizophrenia-spectrum and other psychotic disorders is around 20 years old (Solmi et al. 2022). Almost two-thirds of individuals developed their first mental disorder before the age of 25.

The second risk is that this period corresponds to the peak years of criminal activity. Peak age-crime distributions are below 25 years old for all crimes except gambling in the FBI UCR database (Ulmer and Steffensmeier 2014). Further, the age-crime distributions have been trending towards younger ages. The peak age-crime distribution and trends appear to be driven by a range of biological, sociological, economic, and demographic factors.

As a result of these two sets of risks, significant numbers of youth with mental illness end up in jail. The prevalence estimates for individuals with mental illnesses within criminal justice settings are dramatically higher than the general community (Prins 2014). Consequently, jails have become the de facto mental hospital of last resort.

B. Large Urban County Jail

The large urban county jail is a 3,050-bed capacity facility that services the county's detention needs, has been in operation for several decades, and is spread over two sites. It averages about 2,150 inmates per day across these sites.

Nearly everyone who is arrested and charged for a misdemeanor or felony in the county will arrive at the jail for “booking” and bail setting.⁸ Booking is the process where

8. It is important to note that a felony in this context does not equate to violence. Whereas there is some overlap between the two classifications, they are each addressing different dimensions of criminality. Felonies include

information about the individual is entered into police records upon arrival. Once an individual has been arrested and booked, the arresting officer will file a probable cause affidavit with the court. The affidavit is then reviewed by a judge, who also investigates the defendant's criminal history and sets the bail. There is no set bail schedule, and the amount could be more or less depending on the judge who initially reviews the case.

Once the judge has set the bail amount, a defendant may be released from jail by personal bonds, cash bonds, lawyer-assisted bonds, or surety or bail bonds. A personal bond is a sworn agreement by the defendant that they will return to court as ordered and will comply with the conditions placed on their release. No money is required at the time of release, but the defendant is required to pay an administrative fee of the bond amount to the jail within a certain number of days of release. The county's pretrial services or an attorney may submit a request for release on personal bond to a judge, but defendants cannot request a personal bond themselves. Factors influencing the availability of a personal bond include residence in the community, employment, and criminal history. Only a judge can approve release on a personal bond. In cases where a personal bond is not available, bond must be posted for the individual to be released.

Inmate mental health is evaluated within 36 hours in compliance both with the jail's own historic administrative policies and with statewide regulations. The formal process of booking starts with a topical mental health screening by the officer receiving the individual into residency and then followed by an interview with a randomly assigned trained clinician. Clinicians employed at the jail all have a minimum of a masters of social work.⁹ Most of the interviews are done by clinicians who have substantial experience with these intake interviews. In a robustness check, we dropped clinicians with fewer than 500 interviews in our sample. In doing so, we lost 5 percent of the sample, indicating that 95 percent of the interviews are done by clinicians with substantial experience.

Clinicians are alphabetized daily by last name and then assigned one-to-one to a newly booked inmate as they arrive at the facility, creating a quasi-randomization of intake clinician to inmates at booking. The purpose of the intake clinician is to screen inmates and collect information, not to treat mental health problems. These clinicians use a structured survey questionnaire developed by the facility to assess the mental health needs of each inmate. The questionnaire also includes the Columbia Suicide Severity Rating Scale (Posner et al. 2008) to provide the facility with information about suicide risk. A short description of the survey is contained in the [Online Appendix](#). Completion of the interview can take anywhere from 15 to 30 minutes, depending on the inmate and their responses.

Based on the survey and their own subjective judgment, clinicians produce a mental health needs score. Subjective judgement can come both in the scoring of individual items and in the aggregation of these items into a final score. Clinicians classify the inmate into one of four groups: 0 ("high daily functioning"), 1 ("mild functioning problems"), 2 ("moderate functioning problems"), or 3 ("severe functioning problems"). For the empirical analysis, we collapse the mental health needs score into a dummy variable equal to zero if none/mild and one if moderate/severe. Inmates with moderate/severe scores and

both violent and nonviolent crimes, and in fact, many felonies are nonviolent, such as certain types of drug possession, burglary, theft, etc.

9. The data on educational attainment were not provided to us; however, the director of the clinicians stated that the minimum degree is a masters of social work.

charged with misdemeanors are routed to the county's mental health docket.¹⁰ As a result, this is a natural comparison and is salient to officials at the jail.

C. National Data on Mental Illness and Comorbidities

Mental illnesses include many different conditions that vary in degree of severity and may also be comorbid with alcohol or substance abuse. Two broad categories describe such conditions: any mental illness and serious mental illness. Any mental illness includes all recognized mental illnesses, while serious mental illness is a smaller and more severe subset of any mental illness. More specifically, serious mental illness is defined as a mental, behavioral, or emotional disorders resulting in serious functional impairment, which substantially interferes with or limits one or more major life activities.

In 2020, young adults (18–25) had higher incidence of mental illness with and without substance use disorder than adults (ages 26–49) in the past year (SAMHSA 2021). The incidence of any mental illness without substance use disorder was 18.4 percent for young adults and 16.1 percent for adults. The incidence for having both mental illness and substance use disorder was 12.2 percent for young adults and 9.2 percent for adults. The differences also hold for serious mental illness without substance use disorder, where the incidence was 5.1 percent for young adults and 3.7 percent for adults, and for serious mental illness with substance use disorder, where the incidence was 4.6 percent for young adults and 3.2 percent for adults. Young adults are more likely than adults to have received inpatient care in the past year, 1.4 percent versus 1 percent. Finally, young adults and adults are equally likely to have taken prescription medication for mental illness in the past year, 14.5 percent versus 14.4 percent.

In 2020, young adults were more likely than adults to have had suicidal thoughts, a suicide plan, and attempted suicide in the past year (SAMHSA 2021). Among young adults, 11.3 percent had serious thoughts of suicide, 4.0 percent made a suicide plan, and 1.9 percent attempted suicide. Among adults, 5.3 percent had serious thoughts of suicide, 1.3 percent made a suicide plan, and 0.4 percent attempted suicide.

Deaths from suicide, and from alcohol and drug intoxication, in the overall population have been trending up over time and are lower for youth than adults (CDC 2021). The mortality rate from suicide was 10.7 per 100,000 in 2001 and 13.5 per 100,000 in 2020. Alcohol-induced deaths rose from 7.1 per 100,000 in 2000 to 10.4 per 100,000 in 2019 and further to 13.1 per 100,000 in 2020 (Spencer et al. 2022a). The mortality rate from drug intoxication rose from 6.1 per 100,000 in 2001 to 21.6 per 100,000 in 2019 to 28.3 per 100,000 in 2020 (Spencer et al. 2022b). It is difficult to make age-based comparisons directly with the previous statistics because mortality statistics are often reported for different age ranges. That said, the rates of death are lower for those 15–24 years old than those 25–39 years old. In 2020, the suicide rate was 14.2 per 100,000 for ages 15–24 and 18.4 for ages 25–34. For both alcohol-induced deaths and drug intoxication, the rates were also lower for youth than adults.

10. Many counties have sought to reduce the number of offenders with mental illness in jail and prison by diverting defendants with mental illness from traditional dockets into specialty dockets that, in exchange for dismissing charges, help connect defendants to mental health care. These dockets, known as mental health dockets, are utilized in approximately 500 counties across the nation. See the website for the U.S. Substance Abuse and Mental Health Services Administration (<https://www.samhsa.gov/>) for an annually updated list. These mental health dockets exhibit considerable heterogeneity across U.S. counties.

In jail, young adults (18–24) also have lower rates than adults (25–34) of death from suicide and death from alcohol and drug intoxication per 100,000 inmates (Carson 2021). For young adults, average mortality rates over 2000–2016 were 28 per 100,000 for suicide and five per 100,000 for alcohol and drug intoxication. For adults, the rates were 41 per 100,000 for suicide and ten per 100,000 for alcohol and drug intoxication. Suicide rates in jail for all ages were relatively stable over time—they were 48 per 100,000 in 2000 and 46 per 100,000 in 2016. In contrast, deaths from alcohol and drug intoxication for all ages rose from six per 100,000 in 2000 to 15 per 100,000 in 2016. It is worth noting that the patterns of death for the two causes were quite different. The median time in jail for deaths from alcohol and drug intoxication was one day, while it was nine days for deaths from suicide.

III. Data and Descriptive Analysis

A. Data

We use de-identified administrative data from a large urban county correctional complex (“jail”), which encompasses the universe of inmates booked in the county between 2016 and 2019. These data were collected during routine mental and physical health assessments on inmates and include information on each inmate’s offense type (felony, misdemeanor), demographics, results from the clinician interview (mental health needs score), and outcomes, including length of stay, suicide attempts, and recidivism.¹¹ A unique inmate ID and unique booking ID are jointly used to identify a unique inmate booking event, as well as to link inmates over time. In our analyses, we focus on inmates who received a mental health assessment.¹²

Our main sample spans January 2016 to November 2019. It includes 79,571 inmate–booking observations; 20,430 are for youth and the remaining 59,141 are for adults. For some of the analysis, we limit the sample period. Richer data on mental health of the inmates including self-reported data on psychiatric medications and hospitalization became available in July 2018. This sample includes 31,608 observations; 7,597 observations are for youth and 24,011 are for adults.¹³ While the data we have are quite rich in some dimensions, it is worth noting the data we do not have. For inmate privacy reasons, we were unable to obtain data on medical, psychiatric, and psychological treatment in jail; housing assignments within the jail; specifics of charges beyond misdemeanor levels (A, B, C) and felony levels (1, 2, 3); and outcomes beyond jail, such as whether they were convicted. These are fundamental limitations on our analysis.

B. Youth and Adult Characteristics

Table 1 contains summary statistics for the full sample for 2016–2019 and shows that youth and adults in jail generally have similar demographic characteristics. Most

11. Institutional review board (IRB) approval was granted from Baylor University in April 2019.

12. There is a small subset of people who filter in and out of the jail before they are assessed. Quick release is typically due to transfers or rapid bailouts for misdemeanors.

13. We omitted veterans from the data set as they go into a separate docket.

S290 **Table 1**
Summary Statistics 2016–2019

	Transition Age		Adults (over 25)	
	Misdemeanor <i>N</i> = 11,147	Felony <i>N</i> = 9,227	Misdemeanor <i>N</i> = 30,643	Felony <i>N</i> = 27,515
Panel A: Inmate Characteristics				
Age at booking	21.62 (2.47)	21.51 (2.52)	39.08 (10.46)	38.62 (9.93)
Male	0.70 (0.46)	0.81 (0.39)	0.73 (0.44)	0.78 (0.41)
White	0.73 (0.44)	0.67 (0.47)	0.73 (0.44)	0.71 (0.46)
Black	0.26 (0.44)	0.32 (0.47)	0.25 (0.43)	0.29 (0.45)
Asian	0.0111 (0.1049)	0.0086 (0.0921)	0.0118 (0.1079)	0.0079 (0.0885)
Race other	0.0007 (0.0268)	0.0008 (0.0275)	0.0007 (0.0255)	0.0006 (0.0241)
Hispanic	0.35 (0.48)	0.36 (0.48)	0.24 (0.43)	0.28 (0.45)
Number of priors	3.04 (5.56)	4.44 (6.11)	6.07 (11.20)	4.53 (6.85)
Mental health needs rating				
None/mild	9,289 (83%)	7,817 (85%)	22,680 (74%)	21,979 (80%)
Moderate/severe	1,858 (17%)	1,410 (15%)	7,963 (26%)	5,536 (20%)
Panel B: Outcomes				
<i>LOS</i>	8.14 (19.93)	58.56 (99.26)	11.11 (25.66)	58.84 (91.05)
Suicide attempt (<i>SA</i>)	0.0057 (0.0756)	0.0061 (0.0777)	0.0064 (0.0799)	0.0069 (0.0826)
<i>SA</i> / <i>(LOS + 1)</i>	0.0010 (0.0194)	0.0002 (0.0055)	0.0012 (0.0194)	0.0003 (0.0071)
Recid. within 1 year	0.37 (0.48)	0.39 (0.49)	0.43 (0.49)	0.35 (0.48)
Recid. within 18 months	0.42 (0.49)	0.44 (0.50)	0.46 (0.50)	0.40 (0.49)
Recid. within 2 years	0.44 (0.50)	0.47 (0.50)	0.48 (0.50)	0.42 (0.49)

Notes: Mean (SD); *N* (%).

inmates are male, white, and many have prior offenses within 365 days. Although the majority of inmates are white, Black inmates are significantly overrepresented relative to the general population. The county as a whole is 9 percent Black and 34 percent Hispanic. Across the four columns, inmates are 25–32 percent Black and 24–36 percent Hispanic. Compared to adults in jail, youth are more likely to be Hispanic. The Hispanic population is growing rapidly and is on average quite young. Youth who are charged with a misdemeanor are less likely to have been charged with a prior offense.

The distributions of crimes are also similar for youth and adults; 55 percent of youth and 53 percent of adults are charged with misdemeanors, and the remainder are charged with felonies. The charges are concentrated in more serious misdemeanors and lower-level felonies.¹⁴ [Online Appendix Figure A.1](#) shows the aggregate distribution of case types for all ages.¹⁵ The county does not separately report misdemeanor thefts and felony thefts.

Furthermore, [Online Appendix Tables A.2 and A.3](#) present descriptive statistics by the initial assessment of mental health needs (none/mild vs. moderate/severe) for misdemeanors and felonies, respectively. Interestingly, within both age groups (youth and adults), individuals assessed with moderate/severe mental health needs tend to be slightly older and have a higher number of prior convictions. These trends are consistent for both misdemeanors and felonies, indicating that more severe mental health needs and more frequent contact with the criminal justice system are interconnected.

C. Evaluation of Mental Health

Initial evaluations provide three measures of mental health: (i) the clinician evaluation of mental health needs; (ii) Columbia Suicide Severity Rating Scale, which measures suicide risk; and (iii) self-reported prior psychiatric hospitalization and use of psychiatric medication.¹⁶ [Online Appendix Table A.1](#) shows that youth score lower than adults for mental health needs and self-reported hospitalization and medication but were similar on suicide risk. For mental health needs, 9–11 percent of youth score moderate or severe, as compared to 14–20 percent of adults. For psychiatric hospitalization, 13–14 percent of youth and 17–20 percent of adults report prior hospitalization. For psychiatric medication, 23–24 percent of youth and 32 percent of adults report taking medication. For suicide risk, 13–16 percent of youth score moderate or severe, as compared to 13–16 percent of adults.

In Tables 2 and 3, we include descriptive regressions for mental health needs ratings for both the sample we use in our main IV analysis (Table 2) and for the last five quarters of our data, where additional suicide measurements became available (Table 3). For individual i , booking b , and month–year time t , consider the following model:

$$(1) \quad \text{Score}_{ibt} = \beta_0 + \beta_1 \mathbf{X}_{ibt} + \tau_t + \varepsilon_{ibt}$$

where the inmate's mental health needs score, Score_{ibt} is a 0/1 dummy variable, indicating that the clinician's classification was none/mild (0) or moderate/severe (1).

14. In the case of multiple charges, the analysis is for the highest charge.

15. Note that the individual-level data do not include the specific case type, only the level of the offense.

16. In our IV analyses, only the clinician evaluation results are available for the full time period (2016–2019). The suicidality measures as shown in [Online Appendix Table A.1](#) are only available from 3Q2018 to 4Q2019.

Table 2
Mental Health Classification Descriptive Regressions (IV Sample)

Variables	Transition Age		Adults	
	Misdemeanor (1)	Felony (2)	Misdemeanor (3)	Felony (4)
Age at booking	0.0006 (0.0012)	0.0043*** (0.0015)	0.0025*** (0.0003)	0.0027*** (0.0003)
Male	−0.0557*** (0.0076)	−0.0783*** (0.0103)	−0.0660*** (0.0067)	−0.0669*** (0.0064)
Black	0.0120 (0.0109)	−0.0186** (0.0083)	0.0087 (0.0071)	0.0151** (0.0063)
Asian	0.0357 (0.0352)	0.0355 (0.0446)	−0.0354* (0.0203)	0.0223 (0.0279)
Race other	0.0804 (0.1381)	0.3560** (0.1652)	−0.0145 (0.0762)	−0.1989*** (0.0285)
Hispanic	−0.0433*** (0.0062)	−0.0549*** (0.0078)	−0.0754*** (0.0055)	−0.0645*** (0.0060)
Number of prior charges	0.0124*** (0.0009)	0.0007 (0.0005)	0.0108*** (0.0004)	0.0064*** (0.0004)
Month–year fixed effects	Yes	Yes	Yes	Yes
Observations	11,147	9,227	30,643	27,515
Outcome mean	0.1667	0.1528	0.2599	0.2012

Notes: Each column represents a linear probability model with the binary outcome of interest being the mental health classification of moderate/severe as opposed to none/mild. Clustered (month–year) standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

$Score_{ibt}$ is related to an inmate’s booking and defendant-level characteristics, $\mathbf{X}_{ibt}$, as well as month–year fixed effects τ_t , and ε_{ibt} , an error term. The baseline inmate characteristics are age at booking, sex, race, ethnicity, and the number of prior offenses.

Table 2 indicates that the mental health needs score is generally higher for older inmates, female inmates, non-Hispanic inmates, and inmates with prior offenses. Table 3 shows that the mental health needs score captures important features of inmate mental health. For example, suicide risk ratings have large positive and statistically significant coefficients. Prior psychiatric hospitalization also has a positive and statistically significant coefficient in most specifications.

D. Description of Outcomes

Table 1 indicates that, during their time at the jail, youth and adults have similar length of stay and recidivism rates. For misdemeanors, the average length of stay is eight days for

Table 3*Mental Health Classification Descriptive Regressions (3Q2018–4Q2019)*

	Transition Age		Adults	
	Misdemeanor	Felony	Misdemeanor	Felony
Age at booking	0.0012 (0.0020)	0.0044*** (0.0014)	0.0021*** (0.0003)	0.0025*** (0.0004)
Male	0.0040 (0.0073)	−0.0127 (0.0096)	−0.0182** (0.0064)	−0.0158** (0.0061)
Black	0.0332** (0.0117)	−0.0100 (0.0099)	0.0124 (0.0077)	0.0328*** (0.0080)
Asian	0.0375 (0.0379)	0.0075 (0.0470)	0.0423 (0.0269)	0.0144 (0.0285)
Race other	−0.0510* (0.0285)	0.1213 (0.3345)	−0.1117** (0.0425)	−0.1742** (0.0614)
Hispanic	−0.0095 (0.0108)	−0.0230** (0.0090)	−0.0351*** (0.0062)	−0.0275*** (0.0084)
Number of prior charges	0.0095*** (0.0011)	0.0008 (0.0006)	0.0080*** (0.0004)	0.0039*** (0.0005)
Suicide risk: Mild	0.0673*** (0.0120)	0.0452*** (0.0131)	0.1363*** (0.0109)	0.1206*** (0.0139)
Suicide risk: Moderate/severe	0.3382*** (0.0341)	0.4090*** (0.0282)	0.3847*** (0.0232)	0.4048*** (0.0258)
Prior psych. hospitalization	0.0600** (0.0246)	0.0172 (0.0240)	0.1278*** (0.0150)	0.0838*** (0.0128)
Prior psych. medications	0.0204 (0.0152)	0.0087 (0.0140)	−0.0296*** (0.0066)	−0.0033 (0.0084)
Self-report drug/alcohol use	0.0267 (0.0339)	0.0138 (0.0131)	−0.0433*** (0.0130)	−0.0077 (0.0125)
Month–year fixed effects	Yes	Yes	Yes	Yes
Observations	3,969	3,577	11,806	11,275
Outcome mean	0.1114	0.0865	0.2027	0.1396

Notes: Each column represents a linear probability model with the binary outcome of interest being the mental health classification of moderate/severe as opposed to none/mild. Clustered (month–year) standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

youth and 11 days for adults. One-year recidivism rates are 37 percent for youth and 43 percent for adults. For felonies, average length of stay is 59 days for both youth and adults. One-year recidivism rates are 39 percent for youth and 35 percent for adults.¹⁷

Furthermore, [Online Appendix Tables A.2 and A.3](#) present descriptive statistics by the initial assessment of mental health needs (none/mild vs. moderate/severe) for misdemeanors and felonies, respectively. Individuals initially assessed as moderate/severe typically have longer lengths of stay, more suicide attempts, and higher recidivism levels compared to those assessed as none/mild. These trends are consistent for both misdemeanors and felonies, suggesting that the severity of initial assessment (moderate/severe) is associated with more negative outcomes overall.

Table 1 reports that 0.57–0.61 percent of youth and 0.64–0.69 percent of adults attempted suicide while in jail. Our large urban county jail defines a suicide attempt as self-harm, expressed ideation, or other actions, regardless of how effective they might be.¹⁸ Note, since inmates stay 8–11 days for misdemeanors and 59 days for felonies, these attempts are happening over a fairly short time frame.

From January 2018 through November 2019, we have detailed information on what jail officials considered the most serious suicide attempts. The information from this subsample is summarized in [Online Appendix Table A.8](#). During this time, 85 attempts occurred, of which 30 required medical intervention to address physical harm, psychiatric needs, or both at the jail, and 19 required hospitalizations. Similar shares of youth and adults appear in the data set, and they needed medical intervention at the same rate. The most common methods by which inmates attempted suicide were strangulation, banging their head, hanging, and cutting. At the end of the period covered by these data, the recording of serious suicide attempts shifted to electronic medical records. These data were not available to our research team.

[Online Appendix Figure A.2](#) shows that the most frequent causes of death in this corrections facility are cardiac arrest, followed by suicide. Over our sample period, there were nine cardiac deaths, two suicides, and one death each from substance toxicity, hematoma, and hypertension.

IV. Research Design

A. Overview

To better understand the research design, let us consider a model in which an inmate i at booking b remains in jail longer, attempts suicide, or recidivates based on interventions given in response to their mental health needs score, $Score_{ibt}$.

17. The data presented in Table 1 are intended to preserve the maximum number of observations available within the four-year sample period. To account for the censoring issues associated with recidivism, the recidivism variables are restricted accordingly. For instance, when examining recidivism within one year, the sample is limited to individuals with a release date at least one year before the end of the sample period. Similarly, the sample is restricted for recidivism within 18 months and two years. These limitations are implemented to address the potential impact of censoring on the analyses.

18. For instance, if an inmate were to wrap a towel around their neck, an officer would record this as an attempt regardless of intent. In other words, determining intent is not as important to the jail, given that the downside risk of being incorrect is large.

$$(2) \quad Y_{ibt} = \beta_0 + \beta_1 \text{Score}_{ibt} + \beta_2 \mathbf{X}_{ibt} + \tau_t + \varepsilon_{ibt}$$

where Y_{ibt} is the outcome of interest for individual i in booking b in month-year t : (1) length of stay, (2) suicide attempt, or (3) recidivism within a year of release. The matrix $\mathbf{X}_{ibt}$ contains baseline inmate and booking characteristics, τ_t is a vector of month-year dummies, and ε_{ibt} is an error term.

Ordinary least squares (OLS) estimates of Equation 2 cannot identify the causal effect of classification scores on outcomes if scores are systematically assigned to people already at an elevated risk level. Identifying the causal effect of the score requires exogenous variation. We use the inmate's quasi-randomly assigned clinician's residualized average mental health needs score as an instrument for an inmate's own classification score. Under the standard IV assumptions, we identify the local average treatment effect associated with having moderate to severe mental health needs.

B. Instrumental Variable Calculation

We construct our instrument of clinician tendency to assign moderate to severe mental health needs classification scores. Our instrument is the residualized, leave-one-out mean intake clinician tendency measure conditional on month-of-year fixed effects. This method is similar to the judge leniency designs used in other criminal justice studies (Aizer and Doyle 2015; Arnold, Dobbie, and Yang 2018).

To isolate the effect of clinician scoring on the mental health needs score of inmate i at booking b in time t , we first estimate a linear model that includes a vector of month-of-year fixed effects. These fixed effects account for any countywide trends that might affect mental health needs scores:

$$(3) \quad \text{Score}_{ibt} = \gamma_1 + \gamma_2 \tau_t + \varepsilon_{ibt}$$

where ε is an idiosyncratic error. We then residualize the inmate's mental health needs score using the fitted coefficients from Equation 3:

$$(4) \quad \text{Score}_{ibt}^* = \text{Score}_{ibt} - \hat{\gamma}_1 - \hat{\gamma}_2 \tau_t$$

This residualized score isolates the part of the inmate's mental health needs score that is due to the clinician's scoring, independent of the month-of-year fixed effects.

Next, we construct a measure of each clinician's propensity to assign a high residualized score. This measure is the average residualized score assigned by the clinician, excluding the score for the inmate in question. We construct a measure of residualized clinician c propensity to assign a high score Z_{btc} as:

$$(5) \quad Z_{btc} = \left(\frac{1}{n_{tc} - n_{itc}} \right) \left[\sum_{k=0}^{n_{tc}} (\text{Score}_{ikt}^*) - \sum_{b=0}^{n_{itc}} (\text{Score}_{ibt}^*) \right]$$

where n_{tc} is the number of cases seen by clinician c in month-year t , and n_{itc} is the number of bookings of inmate i seen by clinician c in month-year t .¹⁹ Score_{ikt}^* is the residualized mental health needs score that clinicians give at time t , and Score_{ibt}^* is

19. For more than 99 percent of the observations, n_{it} is 1.

the residualized score by inmate i . In other words, we remove the residualized mental health needs score assignment of all of an inmate's bookings seen by clinician c in each month.²⁰

C. Clinician Leniency

The screening process itself is structured. Clinicians meet one-on-one with individuals within 36 hours after booking. During this 15–30-minute encounter, a specific set of questions are asked, and the clinician records the inmate's self-reported answers.²¹ Despite this structured process, variation can still occur. Clinicians may vary in the recording of specific answers and in how they view the overall set of answers. They are not given specific guidelines for aggregating responses, such as summing the responses and comparing the number to a threshold. This variation can be influenced by a range of factors, including experience, workloads, and personal biases, such as implicit biases towards certain demographic groups or mental health conditions.

Figure 1 presents four distributions of normalized intake clinician residualized leniency measure for high scores stratified by each of our four subpopulations: youth misdemeanor, adult misdemeanor, youth felony, and adult felony. Moving from left to right on each of the subfigures shows the association between the share of the sample assigned a moderate to severe mental health needs score and intake clinicians whose average tendency is to assign worse classification scores. The spread contained in each of the histograms represents the spread of the intake clinician leave-one-out-mean for each of the subpopulations. If clinicians are randomly assigned to inmates and there is no scope for clinician subjectivity in assessing inmate functioning, there would be no variation in our instrument, as all clinicians would agree. Moreover, if clinician assignment is indeed random, then the positive slope in each of the subfigures represents the causal effect of the residualized leave-one-out-mean on the inmate's own classification score.

D. First Stage

We explore this positive association between the residualized leave-one-out-mean clinician scores and an inmate's own score by estimating the following linear probability model:

$$(6) \quad \text{Score}_{ibt} = \alpha + \pi \mathbf{Z}_{btc} + \mathbf{X}_{ibt} + \tau_t + \varepsilon_{ibt}$$

where Score_{ibt} is the binary treatment variable ("classification score"), indicating whether an inmate received a mental health needs score of "Moderate" or "Severe," and $\mathbf{Z}_{btc}$ is a vector of the residualized leave-one-out-mean clinician score, $\mathbf{X}_{ibt}$ is an array of pre-booking inmate characteristics, including race, sex, age at booking, whether they had a prior offense in the last year, the number of offenses per booking, τ_t are month-of-year fixed effects, and ε_{ibt} is the inmate specific error term. Standard errors are two-way clustered by clinician and inmate. Since intake clinicians are quasi-randomly assigned, the first stage has a causal interpretation. Thus, Equation 6 is a linear estimate of the average treatment effect of being assigned a clinician with a higher average mental health needs score on one's own mental health needs score.

20. Following Dobbie, Goldin, and Yang (2018), the instrument is calculated across the full sample.

21. The questions are described in more detail in [Online Appendix A.1](#).

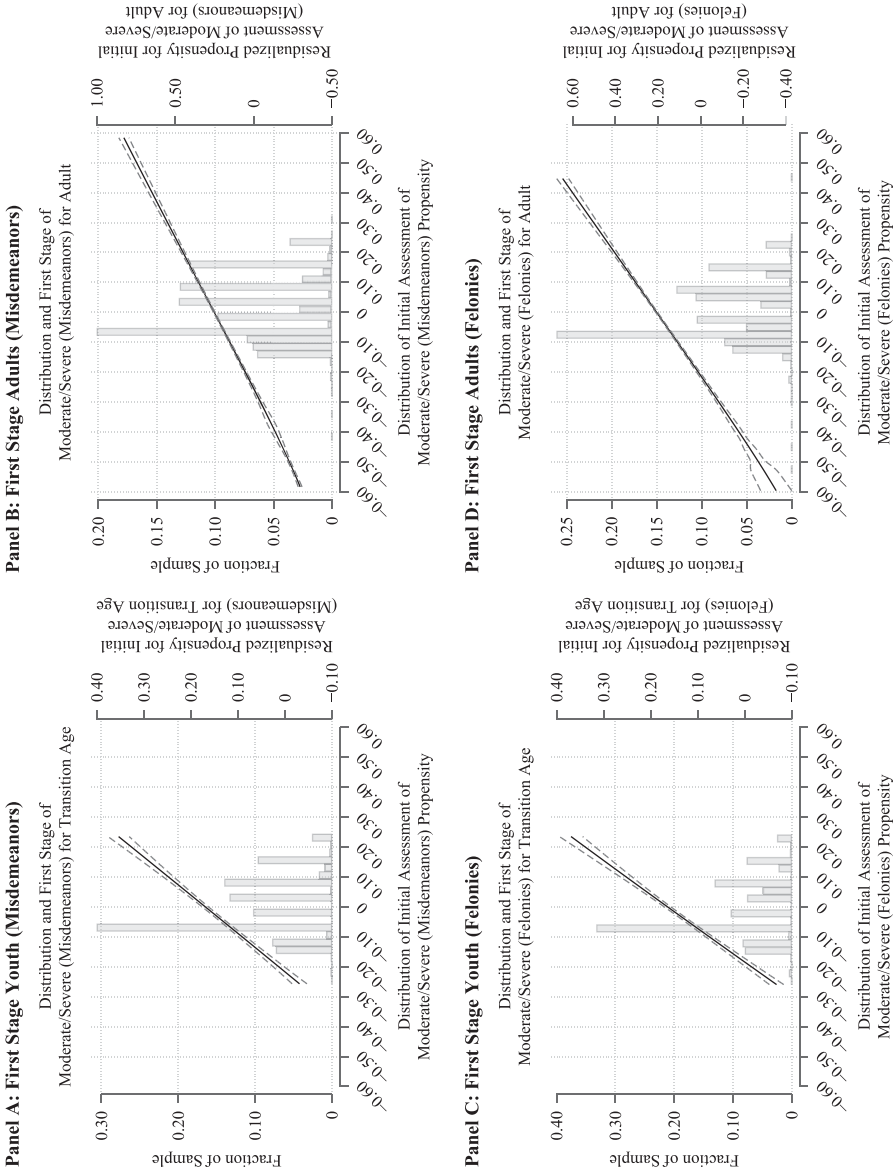


Figure 1
Distribution of Clinician Leniency and First Stage

Table 4
First-Stage Regressions for Initial Assessment of Moderate/Severe (Misdemeanors)

	Transition Age		Adults	
	(1)	(2)	(3)	(4)
Z: Clinician's leave-out mean mental health score	0.783*** (0.033)	0.758*** (0.028)	0.981*** (0.030)	0.861*** (0.030)
Kleibergen–Paap <i>F</i>	575	715	1,048	808
Time fixed effects	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes
Observations	11,147	11,147	30,642	30,642

Notes: We report the first-stage results of a linear probability model stratified by age group. The binary outcome of interest is the initial assessment of an inmate's mental health needs being either none/low or moderate/severe. The propensity to assign the most severe score is estimated using data from other cases assigned to the clinician following the procedure described in the text. Columns 1 and 2 limit the sample to transition age youth, whereas Columns 3 and 4 limit the sample to adults. Columns 1 and 3 show the results by controlling only for month–year fixed effects, whereas Columns 2 and 4 also include the inmate baseline controls as shown in Table 1. Each column gives the corresponding clinician and inmate robust two-way clustered standard errors in parentheses. Robust (Kleibergen–Paap) first-stage *F*-statistic is reported. Note this is equivalent to the effective *F*-statistic of Montiel Olea and Pflueger (2013) in this case of a single instrument. **p* < 0.10, ***p* < 0.05, ****p* < 0.01.

The first-stage relationship is very strong for youth and adults across both misdemeanor and felony offenses. Table 4 shows the first-stage results for misdemeanor populations. Our instrument has a standard deviation of 0.101 in the misdemeanor sample, and we multiply that by the first-stage coefficient to better interpret the magnitudes associated with the leniency instrument.

Shifting clinician leniency one standard deviation would cause a 7.7–7.9 percentage point increase in youth with misdemeanors being classified as moderate to severe mental health needs, and an 8.7–9.9 percentage point increase in adults with misdemeanors. The Kleibergen–Paap first-stage *F*-statistics for misdemeanors are 575–715 for youth and 981–861 for adults, which are significantly higher than commonly employed benchmarks.²² Table 5 shows the same statistics for the felony cases. A one standard deviation (0.098) increase in clinician leniency caused an 8.4–8.6 percentage point increase in receiving a high score among youth with felonies and a 9.2–9.6 percentage point increase among adults with felonies. The effective *F*-statistics for these two subpopulations are also very high and range from 370 to 455.

E. Instrument Validity

Imbens and Angrist (1994) delineates three additional assumptions for the IV approach to identify the local average treatment effect (LATE). The first assumption posits that the

22. The Kleibergen–Paap first-stage *F*-statistic is equivalent to the Olea and Pflueger (2013) effective *F*-statistic given our model is just identified.

Table 5
First-Stage Regressions for Initial Assessment of Moderate/Severe (Felonies)

	Transition Age		Adults	
	(1)	(2)	(3)	(4)
Z: Clinician's leave-out mean mental health score	0.881*** (0.043)	0.860*** (0.045)	0.977*** (0.046)	0.940*** (0.048)
Kleibergen–Paap <i>F</i>	428	370	455	389
Time fixed effects	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes
Observations	9,226	9,226	27,515	27,515

Notes: We report the first stage results of a linear probability model stratified by age group. The binary outcome of interest is the initial assessment of an inmate's mental health needs being either none/low or moderate/severe. The propensity to assign the most severe score is estimated using data from other cases assigned to the clinician following the procedure described in the text. Columns 1 and 2 limit the sample to transition age youth, whereas Columns 3 and 4 limit the sample to adults. Columns 1 and 3 show the results by controlling only for month-year fixed effects, whereas Columns 2 and 4 also include the inmate baseline controls as shown in Table 1. Each column gives the corresponding clinician and inmate robust two-way clustered standard errors in parentheses. Robust (Kleibergen–Paap) first-stage *F*-statistic is reported. Note this is equivalent to the effective *F*-statistic of Olea and Pflueger (2013) in this case of a single instrument. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

instrument is as good as randomly assigned. This random assignment, or independence, allows both the first stage and the reduced form to possess causal interpretations. We offer indirect evidence supporting this assumption through randomization tests and balance tables.

Tables 6 and 7 provide support that the assignment of clinicians is random for youth and adults. In Table 6, we evaluate the correlation between an inmate's classification score and baseline inmate covariates, specifically among youth and misdemeanors. As expected, individuals with a more severe mental health needs score exhibit differences along observable dimensions compared to those with less severe scores. In the second column, we regress the residualized leave-one-out-mean intake clinician score onto the same inmate characteristics. The correlation weakens, with many coefficients becoming insignificant or zero, and the joint significance of all covariates in the model yields an *F*-test value of 2. In the third and fourth columns, we conduct randomization tests for adults with misdemeanors. The *F*-test value for the excludability of covariates drops from 33 to 4, and any discernible differences are relatively small compared to a covariate measurement, such as age in years. Table 7 further provides evidence supporting randomization for both youth and adults with felonies.

In Tables 8 and 9, we follow Aizer and Doyle (2015) and use an alternative test for randomization by showing that covariate means are balanced across the bottom, middle, and top terciles of the instrument's distribution. The mean values of covariates are roughly equivalent across the instrument distribution. In certain instances, such as age for misdemeanors and felonies, and race for felonies, the values exhibit statistical differences, albeit small. However, the balance is significantly less for felony priors.

Table 6
Test of Randomization for Moderate/Severe (Misdemeanors)

	Transition Age		Adults	
	Moderate/ Severe (Misdemeanors) (1)	Z: Moderate/ Severe (Misdemeanors) (2)	Moderate/ Severe (Misdemeanors) (3)	Z: Moderate/ Severe (Misdemeanors) (4)
Asian	0.036 (0.041)	0.004 (0.017)	-0.035 (0.037)	-0.001 (0.007)
Black	0.012 (0.014)	-0.002 (0.002)	0.009 (0.010)	-0.003** (0.001)
Race other	0.080 (0.110)	-0.043** (0.019)	-0.015 (0.097)	-0.020 (0.018)
Hispanic	-0.043*** (0.011)	-0.004 (0.004)	-0.075*** (0.011)	-0.004 (0.003)
Male	-0.056*** (0.014)	-0.012** (0.006)	-0.066*** (0.011)	-0.011** (0.005)
Age at booking	0.001 (0.001)	0.000 (0.000)	0.003*** (0.000)	0.001*** (0.000)
Number of priors	0.012*** (0.002)	0.000 (0.000)	0.011*** (0.001)	0.001*** (0.000)
Time fixed effects	Yes	Yes	Yes	Yes
F-test	11	2	33	4
Observations	11,147	11,147	30,642	30,642

Notes: These linear probability models control for the baseline characteristics used in the instrumental variables analyses. The binary dependent variable in Columns 1 and 3 is being assigned a moderate to severe mental illness score at initial assessment. The dependent variable in Columns 2 and 4 is the propensity to assign a high or low score to inmates. Time fixed effects include month–year fixed effects. Clinician and inmate two-way clustered standard errors shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

While the principle of independence lends a causal interpretation to the reduced form equation, the second assumption, exclusion restriction, confines the sole causal pathway to the inmate’s mental health needs score. This is a challenging assumption to uphold, as there are no straightforward tests, such as balance tables, which can help validate it.

One such violation pertains to the “cohort” of other inmates used to generate the clinician’s leave-one-out mean. Since these inmates will all receive systematic scores, any interaction between the inmate and that cohort could introduce bias through peer effect channels once the inmate is placed in a specific type of housing within the jail. However, it is important to note that housing assignments are based on the inmate’s exogenous criteria, not the clinician’s decision, so this problem of mixing is more likely

Table 7
Test of Randomization for Moderate/Severe (Felonies)

	Transition Age		Adults	
	Moderate/ Severe (Misdemeanors) (1)	Z: Moderate/ Severe (Misdemeanors) (2)	Moderate/ Severe (Misdemeanors) (3)	Z: Moderate/ Severe (Misdemeanors) (4)
Asian	0.036 (0.048)	−0.001 (0.013)	0.022 (0.035)	−0.005 (0.005)
Black	−0.019 (0.013)	−0.004 (0.003)	0.015 (0.010)	0.001 (0.002)
Race other	0.356** (0.158)	0.023 (0.026)	−0.199*** (0.037)	0.010 (0.022)
Hispanic	−0.055*** (0.014)	−0.003 (0.003)	−0.064*** (0.010)	−0.005 (0.004)
Male	−0.078*** (0.017)	−0.015** (0.007)	−0.067*** (0.009)	−0.010* (0.006)
Age at booking	0.004*** (0.002)	0.001*** (0.000)	0.003*** (0.000)	0.001*** (0.000)
Number of priors	0.001 (0.001)	−0.001*** (0.000)	0.006*** (0.001)	0.000 (0.000)
Time fixed effects	Yes	Yes	Yes	Yes
F-test	4	3	25	4
Observations	9,226	9,226	27,515	27,515

Notes: These linear probability models control for the baseline characteristics used in the instrumental variables analyses. The binary dependent variable in Columns 1 and 3 is being assigned a moderate to severe mental illness score at initial assessment. The dependent variable in Columns 2 and 4 is the propensity to assign a high or low score to inmates. Time fixed effects include month–year fixed effects. Clinician and inmate two-way clustered standard errors shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to occur with other inmates in the same housing unit than it is with other inmates from the “clinical scoring cohort.” In other words, the clinician does not influence these other downstream factors, nor does the inmate’s “scoring cohort” necessarily follow them.²³

23. Some researchers have asked if the assignment of the score itself could induce deterioration in the inmate causing suicidality and that therefore this is a possible violation of exclusion. We fully acknowledge that the assignment of the score could have complex effects that include actions based on updating one’s beliefs about one’s own mental health. We consider such an effect to not be a violation of exclusion, but rather a plausible channel by which the score impacts suicidality. Since the randomization of the instrument breaks it from all confounders between the score and suicidality, exclusion violations come solely from the instrument operating

Table 8
Balance of Instrument and Inmate Characteristics for Moderate/Severe (Misdemeanors)

	Bottom Tercile	Middle Tercile	Top Tercile	Middle vs. Bottom <i>p</i> -Value	Top vs. Bottom <i>p</i> -Value
Panel A: Transition Age					
Z: Clinician's leave-out mean mental health score	-0.097	-0.015	0.119	(0.000)	(0.000)
Inmate characteristics					
Asian	0.013	0.009	0.012	(0.257)	(0.772)
Black	0.253	0.271	0.257	(0.261)	(0.837)
Race other	0.001	0.001	0.000	(0.689)	(0.143)
Hispanic	0.355	0.349	0.353	(0.651)	(0.955)
Male	0.720	0.697	0.690	(0.438)	(0.285)
Age at booking	21.569	21.656	21.647	(0.182)	(0.193)
Number of priors	2.754	3.245	3.179	(0.279)	(0.115)
Panel B: Adults					
Z: Clinician's leave-out mean mental health score	-0.097	-0.009	0.126	(0.000)	(0.000)
Inmate characteristics					
Asian	0.011	0.012	0.012	(0.151)	(0.417)
Black	0.251	0.262	0.247	(0.105)	(0.464)
Race other	0.001	0.001	0.000	(0.412)	(0.081)
Hispanic	0.246	0.248	0.240	(0.967)	(0.636)
Male	0.743	0.723	0.726	(0.379)	(0.401)
Age at booking	38.392	39.082	39.718	(0.053)	(0.005)
Number of priors	4.920	6.198	7.018	(0.012)	(0.007)

Notes: Data are from a large county correctional complex. Time fixed effects include month-year fixed effects. Clinician and inmate two-way clustered standard errors shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The final condition needed to interpret our estimates as a LATE parameter associated with the effect of higher mental health needs scores on outcomes is monotonicity. Monotonicity is a shape restriction without which the IV estimates do not have causal interpretations. But it also has a behavioral interpretation. In this context, monotonicity means that if there are two clinicians at intake, and one tends to give more moderate to severe mental health needs scores than the other, then that clinician would weakly always do so if we could rotate the inmates to be seen by the other

on the outcome in ways other than through the score. Consequently, exclusion violations in this context would involve the "clinical scoring cohort."

Table 9*Balance of Instrument and Inmate Characteristics for Moderate/Severe (Felonies)*

	Bottom Tercile	Middle Tercile	Top Tercile	Middle vs. Bottom <i>p</i> -Value	Top vs. Bottom <i>p</i> -Value
Panel A: Transition Age					
Z: Clinician's leave-out mean mental health score	-0.102	-0.032	0.108	(0.001)	(0.000)
Inmate characteristics					
Asian	0.008	0.009	0.009	(0.991)	(0.907)
Black	0.334	0.322	0.317	(0.306)	(0.056)
Race other	0.001	0.001	0.001	(0.895)	(0.453)
Hispanic	0.354	0.370	0.358	(0.275)	(0.873)
Male	0.833	0.804	0.783	(0.148)	(0.119)
Age at booking	21.424	21.512	21.615	(0.348)	(0.000)
Number of priors	4.764	4.530	3.990	(0.108)	(0.003)
Panel B: Adults					
Z: Clinician's leave-out mean mental health score	-0.101	-0.028	0.111	(0.001)	(0.000)
Inmate characteristics					
Asian	0.009	0.007	0.007	(0.288)	(0.266)
Black	0.279	0.285	0.294	(0.479)	(0.151)
Race other	0.001	0.000	0.001	(0.235)	(0.985)
Hispanic	0.284	0.281	0.268	(0.539)	(0.422)
Male	0.797	0.775	0.774	(0.064)	(0.332)
Age at booking	38.221	38.510	39.125	(0.226)	(0.002)
Number of priors	4.379	4.622	4.575	(0.865)	(0.572)

Notes: Data are from a large county correctional complex. Time fixed effects include month-year fixed effects. Clinician and inmate two-way clustered standard errors shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

clinician. This is not something that can be directly tested because it involves counterfactual inmate scoring by different therapists than the ones they had seen. A commonly employed indirect test is to evaluate whether the first-stage estimates are nonnegative for all subsamples. We present that evidence in Table 10 using the full sample of inmates to calculate our measure of clinician leniency. All first-stage coefficients are nonnegative and highly significant for all subsamples of youth misdemeanors and adult misdemeanors. Differences in magnitudes represent stronger roles for clinician discretion, and Black inmates are the ones with the largest magnitudes among misdemeanors. Table 11 shows that the first-stage relationship is also non-negative for all felony subsamples.

Table 10
Average Monotonicity Tests (Misdemeanors)

	Male (1)	Female (2)	Black (3)	White (4)	Hispanic (5)
Panel A: Transition Age					
Z: Clinician's leave-out mean mental health score	0.738*** (0.045)	0.796*** (0.083)	0.824*** (0.062)	0.729*** (0.040)	0.632*** (0.045)
Observations	7,841	3,306	2,897	8,118	3,930
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Panel B: Adults					
Z: Clinician's leave-out mean mental health score	0.846*** (0.035)	0.901*** (0.057)	0.941*** (0.049)	0.841*** (0.045)	0.734*** (0.059)
Observations	22,380	8,262	7,762	22,498	7,502
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: This table presents a test for satisfying average monotonicity as proposed in Frandsen, Lefgren, and Leslie (2020), where they show that average monotonicity can suffice in lieu of strict monotonicity if the average treatment propensities move in the same direction as their potential treatment decisions. In the context of our paper, we can relax strict monotonicity for any given clinician if the individual monotonically complies with enough other judges. Thus, the coefficients should all be significant and the same direction to support average monotonicity.

F. Instrumental Variable Modeling

We present two IV models—two-stage least squares (2SLS) and instrumental variables LASSO (IVLASSO). In a just identified 2SLS model, the residualized leave-one-out-mean is similar to a propensity score that absorbs information from all clinician fixed effects into a single scalar. If any of these clinician fixed effects are weak predictors of mental health needs scores, then the bias of 2SLS will be centered on the bias of OLS.²⁴ IVLASSO, on the other hand, is robust to many instruments, as well as many controls (Chernozhukov, Hansen, and Spindler 2015). When using IVLASSO, our instrumental variables are the full vector of intake clinician fixed effects.

One of the criticisms of the LATE as a policy parameter is that it describes the causal effects of an unknown and invisible subpopulation, the compliers. As we do not observe a person's potential treatment status under different instrument values, we cannot

24. The advantage of using the residualized leave-one-out mean is that similar to the propensity score, it reduces dimensionality problems associated with the instrument. Hull (2017) notes that it is typically simpler to use the just-identified model with the residualized leave-one-out mean as the instrument for the treatment than to invert a multidimensional matrix in 2SLS. Alternative models that are robust to many instruments, like jack-knived instrumental variables estimates (JIVE), are sometimes used instead, but JIVE is biased if there are many covariates (Kolesár et al. 2015).

Table 11
Average Monotonicity Tests (Felonies)

	Male (1)	Female (2)	Black (3)	White (4)	Hispanic (5)
Panel A: Transition Age					
Z: Clinician's leave-out mean mental health score	0.797*** (0.040)	1.078*** (0.129)	0.884*** (0.107)	0.848*** (0.045)	0.672*** (0.054)
Observations	7,449	1,776	2,995	6,145	3,328
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Panel B: Adults					
Z: Clinician's leave-out mean mental health score	0.931*** (0.039)	0.977*** (0.106)	0.913*** (0.088)	0.956*** (0.047)	0.752*** (0.052)
Observations	21,508	6,005	7,873	19,409	7,642
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: This table presents a test for satisfying average monotonicity as proposed in Frandsen, Lefgren, and Leslie (2020), where they show that average monotonicity can suffice in lieu of strict monotonicity if the average treatment propensities move in the same direction as their potential treatment decisions. In the context of our paper, we can relax strict monotonicity for any given clinician if the individual monotonically complies with enough other judges. Thus, the coefficients should all be significant and the same direction to support average monotonicity.

directly identify who is and is not a complier. However, we can learn information about their mean characteristics by regressing $X_i D_i$ on D_i using 2SLS (Angrist, Hull, and Walters 2023). [Online Appendix Tables A.4 and A.5](#) lists those estimates. In general, the compliers do not look vastly different than the overall population.

V. Results

In this section, we present evidence examining the effects of the mental health needs score on inmate outcomes. Tables 12 and 13 present analyses of the effect of a moderate to severe mental health needs score on the three outcomes of interest for the misdemeanor and felony subpopulations, respectively. Columns 1, 2, and 3 report the OLS, 2SLS, and IVLASSO results, respectively, for youth. Similarly, Columns 4, 5, and 6 report the OLS, 2SLS, and IVLASSO results, respectively, for adults. Month-of-year fixed effects and baseline controls are included in each of the regressions as previously discussed, and we present confidence intervals based on Anderson–Rubin tests for weak instruments for the 2SLS models. The complier means for each of the outcomes of interest are included in [Online Appendix Tables A.6 and A.7](#).

Table 12
Effects of Initial Assessment of Moderate/Severe (Misdemeanors)

	Transition Age			Adults		
	OLS (1)	2SLS (2)	IVLASSO (3)	OLS (4)	2SLS (5)	IVLASSO (6)
<i>LOS</i>	5.649*** (0.473)	6.082** (2.928) [0.641, 11.524]	7.650*** (2.256)	6.510*** (0.323)	-0.128 (2.068) [-3.978, 3.722]	1.339 (1.900)
Suicide attempt (<i>SA</i>)	0.019*** (0.006)	0.046*** (0.011) [0.025, 0.067]	0.040*** (0.013)	0.012*** (0.002)	0.021*** (0.006) [0.011, 0.032]	0.018** (0.007)
<i>SA/(LOS+1)</i>	0.003 (0.002)	0.007*** (0.002) [0.003, 0.012]	0.007*** (0.002)	0.002*** (0.000)	0.004*** (0.001) [0.001, 0.007]	0.004*** (0.001)
Recid. within 1 year	0.053*** (0.008)	0.049 (0.056) [-0.055, 0.154]	0.054 (0.066)	0.162*** (0.02)	0.245*** (0.065) [0.124, 0.366]	0.249*** (0.071)

(continued)

Table 12 (continued)

	Transition Age			Adults		
	OLS (1)	2SLS (2)	IVLASSO (3)	OLS (4)	2SLS (5)	IVLASSO (6)
Recid. within 18 months	0.064*** (0.011)	−0.004 (0.069) [−0.131, 0.123]	−0.019 (0.061)	0.149*** (0.02)	0.216*** (0.063) [0.099, 0.332]	0.195*** (0.075)
Recid. within 2 years	0.074*** −0.01	0.031 −0.072 [−0.102, 0.164]	−0.039 −0.044	0.140*** −0.017	0.245*** −0.069 [0.132, 0.385]	0.157 −0.099
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the ordinary least squares (OLS), two-stage least squares (2SLS), and the instrumental variables LASSO (IVLASSO) estimates of the impact of being assigned a moderate/severe mental health needs rating. The dependent variable is listed in each row. The recidivism within one year sample is truncated to one year prior to the last date our data set. Similarly, the recidivism within 18 months and two years samples are truncated to 18 months and two years prior to the last date. The 2SLS and IVLASSO specifications instrument for assignment of a high mental health needs score using a clinician leniency measure that is estimated using data from other cases assigned to a clinician as described in the text. We include month-year fixed effects and baseline controls for all specifications. The clinician and inmate robust two-way clustered standard errors are shown in parentheses. For the 2SLS estimates, confidence intervals based on the inversion of the Anderson–Rubin test are shown in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13
Effects of Initial Assessment of Moderate/Severe (Felonies)

	Transition Age			Adults		
	OLS (1)	2SLS (2)	IVLASSO (3)	OLS (4)	2SLS (5)	IVLASSO (6)
<i>LOS</i>	23.165*** (2.548)	47.372*** (14.415) [20.597, 74.147]	54.055* (27.763)	23.112*** (1.178)	24.658** (11.104) [3.988, 45.328]	22.628* (12.974)
Suicide attempt (<i>SA</i>)	0.017*** (0.002)	0.038** (0.017) [0.007, 0.070]	0.015 (0.011)	0.017*** (0.002)	0.030*** (0.008) [0.016, 0.046]	0.018* (0.009)
<i>SA</i> /(<i>LOS</i> +1)	0.001*** (0.000)	0.001 (0.001) [-0.001, 0.003]	0 (0.001)	0.001*** (0.000)	0.002*** (0.001) [0.001, 0.003]	0.001*** (0.000)
Recid. within 1 year	0.054*** (0.008)	0.055 (0.053) [-0.043, 0.152]	0.028 (0.062)	0.039*** (0.006)	0.06 (0.047) [-0.019, 0.157]	0.051 (0.050)

(continued)

Table 13 (continued)

	Transition Age			Adults		
	OLS (1)	2SLS (2)	IVLASSO (3)	OLS (4)	2SLS (5)	IVLASSO (6)
Recid. within 18 months	0.041** (0.019)	−0.038 (0.053) [−0.147, 0.061]	0.024 (0.052)	0.029*** (0.006)	−0.008 (0.043) [−0.079, 0.079]	0 (0.055)
Recid. within 2 years	0.054*** (0.015)	0.036 (0.081) [−0.113, 0.184]	0.025 (0.062)	0.017** (0.007)	−0.054 (0.051) [−0.139, 0.050]	−0.062 (0.057)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the ordinary least squares (OLS), two-stage least squares (2SLS), and the instrumental variables LASSO (IVLASSO) estimates of the impact of being assigned a moderate/severe mental health needs rating. The dependent variable is listed in each row. The recidivism within one year sample is truncated to one year prior to the last date our data set. Similarly, the recidivism within 18 months and two years samples are truncated to 18 months and two years prior to the last date. The 2SLS and IVLASSO specifications instrument for assignment of a high mental health needs score using a clinician leniency measure that is estimated using data from other cases assigned to a clinician as described in the text. We include month-year fixed effects and baseline controls for all specifications. The clinician and inmate robust two-way clustered standard errors are shown in parentheses. For the 2SLS estimates, confidence intervals based on the inversion of the Anderson–Rubin test are shown in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We present the main results on length of stay, suicide attempts, and recidivism and then discuss five robustness checks. The robustness checks include: (i) limiting the sample to clinicians with at least 500 evaluations, (ii) restricting the sample to inmates scored as having mild or moderate mental illness (excluding inmates scored as having no mental illness or as having severe mental illness), (iii) examining a subsample where we can control for self-reported previous psychiatric hospitalization and psychiatric medication, (iv) dropping bookings in which individuals ever reported using drugs or alcohol in the last 12 months, and (v) investigating the results separately by race and ethnicity.

A. Length of Stay

In Table 12, youth charged with misdemeanors as well as youth and adults charged with felonies who have worse a mental health needs score have statistically significantly longer stays in jail. In the IV specifications for youth charged with misdemeanors, the increase is 6.1–7.7 days. In the IV specifications for adults charged with misdemeanors, there is no statistically significant effect. In the IV specifications for youth charged with felonies, the increase is 47.4–54.1 days. In the IV specifications for adults charged with felonies, the increase is 22.6–24.7 days. For youth charged with misdemeanors, and youth and adults charged with felonies, the Anderson–Rubin confidence intervals exclude zero. These groups all experience large increases relative to the complier mean lengths of stay.

A moderate to severe mental health needs score may influence length of stay through several channels. Judges may set bond based on the mental health needs score. If family and friend networks learn new information about the inmate's mental health status, this may affect their decisions about whether to post bail. New information could be revealed, for example, through a recommendation that the case be sent to the mental health docket. The mental health docket, for which individuals charged with misdemeanors and classified as moderate or severe mental health needs are eligible, often operates with delays.²⁵

B. Suicide Attempts

In Table 12, youth and adults charged with misdemeanors and adults charged with felonies who have moderate to severe mental health needs scores are statistically significantly more likely to attempt suicide while in jail. In the IV specifications for youth charged with misdemeanors, the increase is 4.0–4.6 percentage points. In the IV specifications for adults charged with misdemeanors, the increase is 1.8–2.1 percentage points. In the IV specifications for youth charged with felonies, the 2SLS and IVLASSO differ. The coefficient is positive in both specifications but is only statistically significant for 2SLS. In the IV specifications for adults charged with felonies, the increase is 1.8–3.0 percentage points. In all the specifications, the Anderson–Rubin confidence intervals exclude zero.

25. In addition to delays in the mental health docket, individuals with severe mental health needs may require mental health competency hearings. These hearings often involve extensive delays. Furthermore, if the individual is in need of inpatient services to restore competency, there are lengthy wait times to be admitted to a state hospital. A 2017 survey found that average wait times vary by state from 30 days up to a year (Fuller et al. 2017).

Given we find that in many cases a moderate to severe mental health needs score causes compliers to spend, on average, weeks to months longer time in jail, it's possible that the increase in suicide attempt is simply a mechanical function of time. That is, if an inmate has a constant suicide attempt hazard per day in jail, then suicide attempts will increase for no other reason than increasing their length of stay. Thus, to check whether our suicide attempt result is simply a mechanical artifact of length of stay, we reestimate our OLS and IV models for all four subpopulations using suicide attempts per day in jail, measured as $SA/(LOS + 1)$.²⁶

In Table 12, youth and adults charged with misdemeanors and adults charged with felonies who have moderate to severe mental health needs scores have statistically significantly higher suicide attempts per day. In the IV specifications for youth charged with misdemeanors, the increase is 0.7 percentage points per day. In the IV specifications for adults charged with misdemeanors, the increase is 0.4 percentage points per day. In the IV specifications for youth charged with felonies, neither specification is statistically significant. In the IV specifications for adults charged with felonies, the increase is 0.1–0.2 percentage points per day. For youth and adults charged with misdemeanors and adults charged with felonies, the Anderson–Rubin confidence intervals exclude zero.

These analyses suggest that worse classification scores cause suicide attempts to rise along two margins. First, by increasing their time in jail, it increases inmate suicide attempts mechanically because the longer a person is in jail, the more likely they are to attempt suicide. However, we also find that worse classification scores increase suicide attempts per day in jail, suggesting that not all of the increased risk is mechanical. Every day they are there, inmates with worse classification scores face higher risks than compliers over the same days.

The previous section documented one mechanism, length of stay, through which a moderate to severe mental health classification score might cause increases in suicide attempts. There are several other possible mechanisms through which a moderate to severe mental health classification score might cause increases in suicide attempts or suicide attempts per day. Labeling of more severe mental health needs could exacerbate mental health deterioration due to its impact on despair, stigma, medication, or placement in a different type of housing within the jail, thereby leading to suicide attempts. The heightened supervision that accompanies more severe classifications could also result in the same behavior being more frequently observed and consequently recorded as a suicide attempt.

One question is how to interpret suicide attempts. An attempt could be an expression of distress or could be a strategic action to achieve some other objective. This raises the question of what the objective might be and why this would differ across mental health scores. The food is the same throughout the jail. The most likely issue would be housing, since almost all inmates have cellmates. In instances of conflict between cellmates, an inmate might attempt suicide to instigate a change in cellmates or housing type. However, this prompts the question of why the labeling itself would differentially influence this behavior. For this to be the case, it would necessitate that the matches in the population with worse mental health needs scores are inherently more problematic, that jail personnel are more responsive to prisoners with worse mental health needs scores, or a combination of both. While some attempts may have been strategic, the

26. We add a one to length of stay because some individuals bond out on the day of arrival and do not spend a full day.

records of the more severe attempts in [Online Appendix Table A.8](#) and the two deaths by suicide suggest that some attempts probably reflected distress.

C. Recidivism

Table 12 shows that adults charged with misdemeanors who have moderate to severe mental health needs scores are statistically significantly more likely to recidivate in the first year after release. In the IV specifications for adults charged with misdemeanors, the increase is 24.5–24.9 percentage points. For adults charged with misdemeanors, the Anderson–Rubin confidence intervals exclude zero. These are large increases relative to the complier mean recidivism. We also see increased recidivism in the first 18 months after release and the first two years after release. In the IV specifications for youth and adults charged with felonies, there is not a statistically significant effect of classification on recidivism. This may reflect incapacitation.

One mechanism through which a moderate to severe mental health classification score may cause increases in recidivism is length of stay. Previous studies have established a causal relationship between length of stay and recidivism (Gupta, Hansman, and Frenchman 2016; Heaton, Mayson, and Stevenson 2017; Dobbie, Goldin, and Yang 2018; Leslie and Pope 2018; Loeffler and Nagin 2022).

Another relevant mechanism for inmates charged with misdemeanors is assignment to the mental health docket. For inmates charged with misdemeanors, the mental health needs score determines whether their case is routed to a traditional or a mental health docket. The primary intervention at the mental health docket is the dismissal of charges. In theory, dismissal is contingent upon the completion of the court's recommended mental health care regimen, but monitoring is minimal. Assignment to the mental health docket may influence recidivism through reductions in expected punishment. Other mechanisms may affect recidivism as well.

D. Robustness Checks

In this section, we present five robustness checks. One concern is that our results might be driven by clinicians who evaluated a small number of inmates. [Online Appendix Tables A.9 and A.10](#) present the results when we restrict the sample to clinicians who evaluated at least 500 inmates, which retains 95 percent of the sample. The results are very similar to our main results in Tables 12 and 13.

Another concern is that the sample includes inmates who are scored as having no mental illness and severe mental illness, when the compliers are likely to be individuals scored as having mild or moderate mental illness. To address this, we restrict the sample to individuals scored as having mild or moderate mental illness ([Online Appendix Tables A.11 and A.12](#)).²⁷ Compared to Table 12, the coefficients for youth and adults in [Online Appendix Table A.11](#) are generally similar but are less likely to be statistically significant. This reflects the fact that this sample is much smaller, only 38 percent of the original sample. Notably, however, coefficients on $SA/(LOS + 1)$ for youth in the 2SLS

27. The dummy variable in the robustness check is equal to zero for mild and one for moderate. The dummy variable in the main specification is equal to zero for none and mild and one for moderate and severe.

and IVLASSO specifications are positive and statistically significant. The coefficients for youth and adults in [Online Appendix Table A.12](#) are also generally similar to Table 13 but are less likely to be statistically significant.

A third concern is that omitted variables regarding mental health history might be influencing the estimates. For a subsample of the period, we have additional self-reported information about psychiatric medication and psychiatric hospitalization. In [Online Appendix Tables A.8–A.16](#) we focus on this subsample and include psychiatric medication and psychiatric hospitalization as controls. The balance tests in [Online Appendix Tables A.8 and A.9](#) find that most inmate characteristics are not statistically significantly different across the terciles of the leave-out mean mental health score (our instrument). If we focus on psychiatric hospitalization, psychiatric medication, and drug use in the last 12 months, only three out of 24 differences are statistically significantly different, and all the differences are at the 10 percent level. This suggests that the instrument is not correlated with these observable characteristics and supports the validity of our instrument. Additionally, the point estimates for misdemeanors in [Online Appendix Table A.15](#) are larger than estimates in Table 12. The effects on the coefficients for felonies are more mixed. Some estimates for felonies in [Online Appendix Table A.16](#) are no longer statistically significant, although the point estimates on the coefficients often remain large. Others, notably the IV Lasso coefficient on suicide attempts for youth, become statistically significant.

A fourth concern is that omitted variables regarding substance use might be influencing the estimates. While we do not have good measures of substance use disorder, for the subsample with information on psychiatric hospitalization and psychiatric medication, we also have additional self-reported information on whether the individual used drugs or alcohol within the last 12 months.²⁸ The results in [Online Appendix Tables A.17 and A.18](#) when we drop individuals who report alcohol or drug consumption are similar to the results in [Online Appendix Tables A.15 and A.16](#) when we include psychiatric medication and psychiatric hospitalization as controls.

A final question is whether the effects of worse classification scores differ across race or ethnicity. In [Online Appendix Tables A.19–A.24](#) we estimate the results separately for inmates who are white, Black, and Hispanic. The results for length of stay, suicide, suicide per day, and recidivism are generally similar across the three groups.

VI. Conclusion

Drawing on administrative data from a large urban county jail, we provide new causal evidence on the effect of mental health needs scores by jail clinicians on the outcomes of transition age youth and adults in jail. Using an IV estimation approach, we found that worse mental health needs scores increased incarceration duration and suicide attempts in youth and adults and recidivism in adults. Thus, the marginal person,

28. This information is from the jail's modified version of the TCU Drug Screen, which is described in [Online Appendix A](#). We construct a binary measure of the TCU score, which is one if the individual ever indicates alcohol or drug consumption in the interview with the clinician and zero otherwise. The share of inmates reporting any drug or alcohol consumption is small—6 percent for misdemeanors and 13 percent for felonies. This suggests that underreporting is an issue.

who was assigned a higher mental health score solely because they were assigned to a clinician who gives higher scores, had worse outcomes. We discussed possible mechanisms, including bonds, the existence of a mental health docket for misdemeanors, despair, stigma, medication, and placement in a different type of housing within the jail. Length of stay itself is a possible mechanism for suicide attempts, suicide attempts per day, and recidivism.

Our findings regarding the adverse effects of receiving a moderate to severe mental health needs score are highly policy-relevant but should be interpreted in light of certain limitations of our data. A crucial caveat is that our data originate from a single large urban jail, so we cannot disentangle the extent to which our findings are attributable to the jail itself or to state policy. Furthermore, our data do not include comprehensive information on substance use disorder, and due to the data use agreement, we are unable to link to disposition or other post-incarceration outcomes.

Based on our findings, we offer three recommendations. First, considering the substantial variation in the assignment of scores across clinicians, it would be prudent to invest in comprehensive training to ensure more accurate and consistent classification of individuals' mental health needs. Second, the development and implementation of superior residential programs for individuals with mental illness could potentially facilitate earlier release from jail. Depending on the program, this may reduce suicide attempts and recidivism for the marginal person. Third, a more in-depth examination of the experiences of inmates within the jail is necessary to gain a better understanding of the mechanisms through which moderate to severe mental health scores at booking translate into adverse outcomes for both youth and adults.

References

- Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130(2): 759–803.
- Anderson, D. Mark, Resul Cesur, and Erdal Tekin. 2015. "Youth Depression and Future Criminal Behavior." *Economic Inquiry* 53(1):294–317.
- Angrist, Joshua, Peter Hull, and Christopher Walters. 2023. "Methods for Measuring School Effectiveness." In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, forthcoming. New York: Elsevier.
- Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133(4):1885–932.
- Bondurant, Samuel R., Jason M. Lindo, and Isaac D. Swensen. 2018. "Substance Abuse Treatment Centers and Local Crime." *Journal of Urban Economics* 104:124–33.
- Bronson, Jennifer, and Marcus Berzofsky. 2017. "Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates 2011–12." NCJ 250612. Washington, DC: Bureau of Justice Statistics.
- Carson, E. Ann. "Mortality in Local Jails 2000–2018." Tables 2021, NCJ 251921. Washington, DC: Bureau of Justice Statistics.
- Carson, E. Ann, and Mary P. Cowhig. 2020. "Mortality in Local Jails 2000–2016." Tables 2020, NCJ 251921. Washington, DC: Bureau of Justice Statistics.
- CDC. 2021. "Data & Statistics Fatal Injury Report for 2021." Atlanta, GA: Centers for Disease Control and Prevention.

- Chatterji, Pinka, and Alison Evans Cuellar. 2006. "How Do Youth with Mental Disorders Fare in the Juvenile Justice System?" NBER Working Paper 12437. Cambridge, MA: NBER.
- Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. "Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments." *American Economic Review: AEA Papers and Proceedings* 105(5):486–90.
- Cuellar, Alison Evans, Larkin S. McReynolds, and Gail A. Wasserman. 2006. "A Cure for Crime: Can Mental Health Treatment Diversion Reduce Crime among Youth?" *Journal of Policy Analysis and Management: Journal of the Association for Public Policy Analysis and Management* 25(1):197–214.
- Deza, Monica, Johanna Catherine Maclean, and Keisha Solomon. 2022. "Local Access to Mental Healthcare and Crime." *Journal of Urban Economics* 129:103410.
- Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108(2):201–40.
- Fletcher, Jason, and Barbara Wolfe. 2009. "Long-Term Consequences of Childhood ADHD on Criminal Activities." *Journal of Mental Health Policy and Economics* 12(3):119–38.
- Fuller, D.A., E. Sinclair, H.R. Lamb, J.D. Cayce, and J. Snook. 2017. "Emptying the 'New Asylums': A Beds Capacity Model to Reduce Mental Illness behind Bars." Arlington, VA: Treatment Advocacy Center.
- Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." *Journal of Legal Studies* 45(2):471–505.
- Heaton, Paul, Sandra Mayson, and Megan Stevenson. 2017. "The Downstream Consequences of Misdemeanor Pretrial Detention." *Stanford Law Review* 69:711–94.
- Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *Quarterly Journal of Economics* 132(1):1–54.
- Hull, Peter. 2017. "Examiner Designs and First-Stage F Statistics: A Caution." Unpublished.
- Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–75.
- Jácóme, Elisa. 2020. "Mental Health and Criminal Involvement: Evidence from Losing Medicaid Eligibility." Unpublished.
- Kaba, Fatos, Angela Solimo, Jasmine Graves, Sarah Glowa-Kollisch, Allison Vise, Ross MacDonald, Anthony Waters, Zachary Rosner, Nathaniel Dickey, Sonia Angell, and Homer Venters. 2015. "Disparities in Mental Health Referral and Diagnosis in the New York City Jail Mental Health Service." *American Journal of Public Health* 105(9):1911–16.
- Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and Inference with Many Invalid Instruments." *Journal of Business & Economic Statistics* 33(4):474–84.
- Leslie, Emily, and Nolan G. Pope. 2018. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *Journal of Law and Economics* 60(3):529–57.
- Loeffler, Charles E., and Daniel S. Nagin. 2022. "The Impact of Incarceration on Recidivism." *Annual Review of Criminology* 5:133–52.
- NACO. 2018. "Four Key Measures #2: Shortening the Length of Stay in Jail for People with Mental Illnesses." Washington, DC: NACO.
- Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31(3):358–69.
- Posner, K., D. Brent, C. Lucas, M. Gould, B. Stanley, G. Brown, P. Fisher, J. Zelazny, A. Burke, M. Oquendo, and J. Mann. 2008. "Columbia-Suicide Severity Rating Scale (C-SSRS)." Menands, NY: Research Foundation for Mental Hygiene.

- Prins, Seth J. 2014. "Prevalence of Mental Illnesses in US State Prisons: A Systematic Review." *Psychiatric Services* 65(7):862–72.
- SAMHSA. 2021. "Key Substance Use and Mental Health Indicators in the United States: Results from the 2020 National Survey on Drug Use and Health." HHS Publication no. PEP21-07-01-003, NSDUH Series H-56. Technical Report. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.
- Solmi, Marco, Joaquim Radua, Miriam Olivola, Enrico Croce, Livia Soardo, Gonzalo Salazar de Pablo, Jae Il Shin, James B Kirkbride, Peter Jones, Jae Han Kim, Jong Yeob Kim, André F. Carvalho, Mary V. Seeman, Christoph U. Correll, and Paolo Fusar-Poli. 2022. "Age at Onset of Mental Disorders Worldwide: Large-Scale Meta-Analysis of 192 Epidemiological Studies." *Molecular Psychiatry* 27:281–95.
- Spencer, Merianne Rose, Sally C. Curtin, and Matthew F. Garnett. 2022a. "Alcohol-Induced Death Rates in the United States, 2019–2020." NCHS Data Brief 448. Hyattsville, MD: National Center for Health Statistics.
- Spencer, Merianne Rose, Arialdi M. Miniño, and Margaret Warner. 2022b. "Drug Overdose Deaths in the United States, 2001–2021." NCHS Data Brief 457. Hyattsville, MD: National Center for Health Statistics.
- Ulmer, Jeffrey Todd, and Darrell J. Steffensmeier. 2014. "The Age and Crime Relationship: Social Variation, Social Explanations." In *The Nurture versus Biosocial Debate in Criminology: On the Origins of Criminal Behavior and Criminality*, ed. Kevin M. Beaver, J.C. Barnes, and Brian B. Boutwell, 377–96. Thousand Oaks, CA: SAGE Publications.
- Zhong, Shaoling, Morwenna Senior, Rongqin Yu, Amanda Perry, Keith Hawton, Jenny Shaw, and Seena Fazel. 2021. "Risk Factors for Suicide in Prisons: A Systematic Review and Metaanalysis." *Lancet Public Health* 6(3):e164–e174.