Social Workers and Suicidality in Jail: Evidence from Travis County's Mental Health Court

Vivian S. Vigliotti (r), Jonathan Seward (r), Scott Cunningham (r)

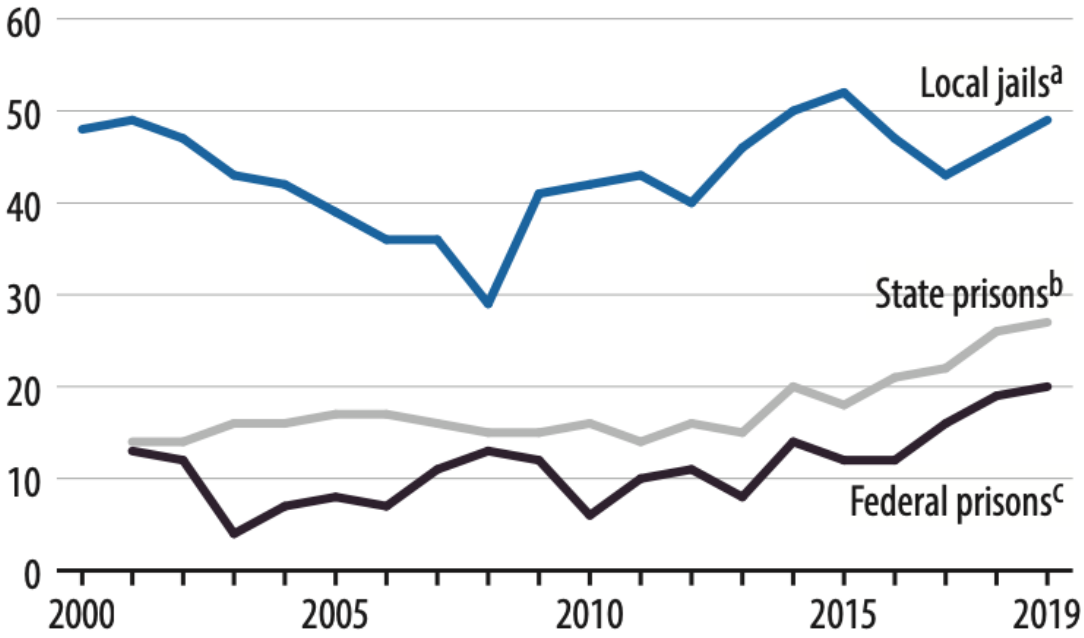
November 2021

Baylor University

Rising Suicides in Corrections

FIGURE 2

Rate of suicides per 100,000 inmates in local jails and 100,000 prisoners in state and federal prisons, 2000–2019



Suicide in jails

- Suicide is the leading cause of death in jails nationally and the second leading cause of death in Texas accounting for a third of all deaths
- Pre-trial detainees have a suicide attempt rate 8x higher than the general population
- Suicides represented an average of 6% of deaths in state prisons from 2001 to 2016
- Suicide counts and rates in corrections are even higher than military and veterans

Lawyers-with-social-workers

- Travis county, Texas (Austin) has a mental health court (MHC) where mentally ill offenders are diverted away from traditional courts into a friendly court seeking dismissals
- Indigent defense is provided by either a public defender with a team of social workers or a private moonlighting defense attorney (from the "wheel") depending on symptom severity
- Treatment is a high score on mental health assessment at booking representing low functioning and very high mental health symptoms
- We call this sometimes a "high score" and other times "lawyers-with-social-workers"

Summary of findings I

- We instrument for lawyers-with-social-workers (i.e., high score) using the randomized evaluators' average propensity to score high among (i.e., the leniency design)
- Lawyers-with-social-workers have no effect on repeat offending
- Lawyers-with-social-workers **improve** mental health scores by as much as 25%
- Lawyers-with-social-workers **reduce** suicide attempts by 9-22%, and self-reported suicidal ideation by 1-2%

These results appear driven by those without a prior treatment, suggesting early onset of symptoms and/or having not been identified for some other reason (e.g., lower attachment to healthcare, lower income)

Summary of findings II: MTE results

Marginal treatment effect (MTE) results

- MTE for suicide attempts cannot be calculated due to strict monotonicity violation
- No heterogeneity found for suicidal ideation (i.e., ATE $\approx$ LATE)

Summary of findings III: MTE results

MTE results

- Impact of social workers on subsequent functioning gets larger as we move closer to those with worse functioning
- But at some point, it plateaus and returns fall
- But returns are always improving even for those with the worst functioning

Introduction to Mental Health Courts

Jails and prisons are the mental health hospitals of last resort

- Inmates are 64% or up to 12 times more likely to have a mental illness than the general community (Prins, 2014)
  - In most states, there is at least one jail or prison that houses more mentally ill individuals than the largest psychiatric hospital in the area (Torrey et al. 2014)
  - $\sim$20 percent of inmates in our data require treatment for their mental illness
- On any given day, 7 percent of inmates with mental illness are experiencing severe symptoms such as psychosis, delusions or suicidal thoughts (Corrections Officers Receive Specialized Mental Health Training, 2020)
  - One study found a 77% prevalence rate of mental illness among inmates who attempted suicide (Goss et al. 2002)

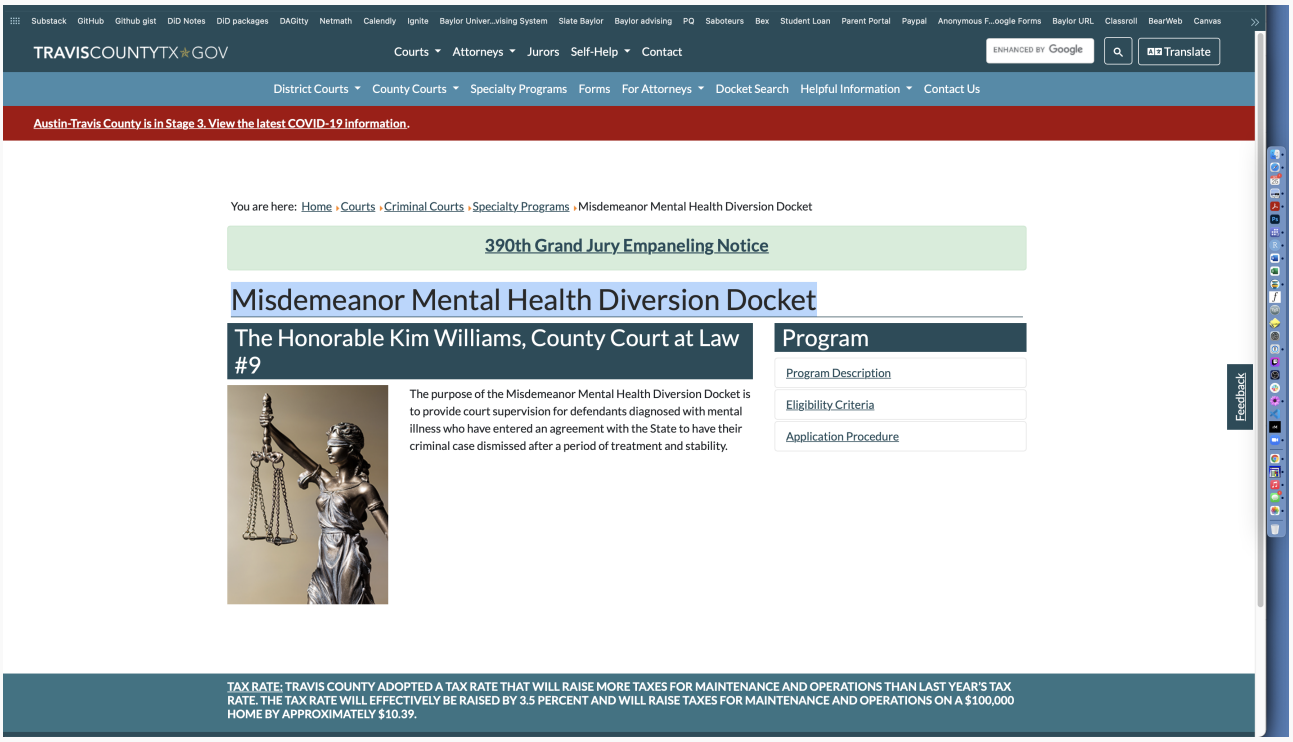
Changing philosophies and therapeutic jurisprudence

- Mental health court (MHC) movement emerged out of inequities in the experiences among people with mental illnesses, growth in therapeutic jurisprudence and the drug court movement
- MHCs are specialty courts endogenously adopted by counties to care for the growing mentally ill population caught in criminal justice institutions

Emergence of mental health courts

- With **typical courts**, a defendant is booked, screened for mental illness; if convicted, goes to jail, likely receives medication to treat mental illness depending on making bail
- MHCs are **diversion** interventions (e.g., drug court, battery court) that engage defendants with mental illnesses in lieu of incarceration
- In counties with an MHC, the inmate will be redirected to a specialty court if supervisors believe mental illness contributed to offense, the defendant meets criteria and there is capacity

Misdemeanor Mental Health Diversion Docket



Data and research design

Travis county correctional complex

Data Description

Travis County Texas Correctional Complex (2016-2019)

- Universe of bookings (n= 5,222 using only the mental health court inmates)
- Inmate ID (unique) & booking ID (unique case/event) allows us to measure recidivism (we use only misdemeanors)
- Administrative data including offense type (felony, misdemeanor), demographics, mental health, charges, suicide attempt, suicide ideation, etc.

Raw data shows signs of selection bias (see summary statistics below)

Table 1: Descriptive Statistics by Public Defender Assignment

	Private Defender	Public Defender
Outcomes		
Suicide attempt in next booking	0.051	0.030
Suicide ideation in next booking	0.006	0.004
Next booking mental health score improves	0.431	0.547
Inmate Characteristics		
White	0.731	0.704
Asian	0.009	0.011
Black	0.259	0.284
Race other	0.001	0.001
Hispanic	0.218	0.177
Male	0.630	0.702
Age at booking	35.653	37.204
Prior offense w/in 365 days	0.379	0.449
Number of offenses per booking	1.597	1.654
First time in jail	0.019	0.014
Prior treatment	0.140	0.087
Prior medications	0.129	0.089
Prior hospitalization	0.103	0.080
Homeless	0.055	0.042
Jobless	0.073	0.052
Observations	4,294	928

What is a leniency design?

- Leniency design dates back to the original LATE paper by Imbens and Angrist (1994)
- Assume that in order to be assigned to counsel, an inmate must first be seen by a clinician who after interviewing them scores the severity of their symptoms (high/low)
- If symptoms were a blood test, then it wouldn't matter who saw the inmate – they'd all give the same score in counterfactual even if randomized
- But symptoms are based on observation, interpretation and professional judgment, and they're seeing them in high volume daily for only 15 minutes usually
- Leniency uses the clinicians average tendency to give inmates high scores as an instrument for what score they did give an inmate

IV Assumptions: (1) Independence

Independence – director of inmate mental health has explained they use a random number generator to assign therapists to inmates which we check with balance

IV Assumptions: (2) SUTVA

SUTVA – an inmate's score is a function of their personally assigned clinician and not someone else's (no spillover from instrument assignment)

IV Assumptions: (3) Exclusion

Exclusion – randomized therapist during assessment can only impact suicidality via assignment to lawyers-with-social-workers

We hypothesized that if the scores are used to deliver services to inmates while in jail, then this could violate exclusion

We interviewed the director of inmate mental health who employees and directs the clinicians about this several times. He assured us that the rating does not have any specific impact on choices made while they are in jail – it does not affect housing, psych criteria or follow-up criteria. Psych criteria is shaped by making the minimum score, but does not vary once you meet that score.

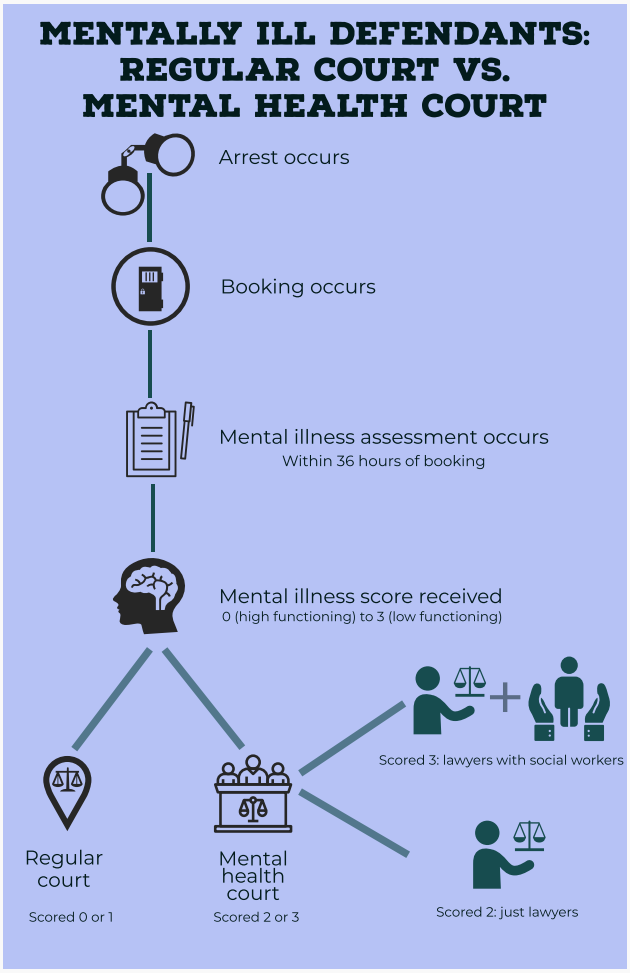
IV Assumptions: (4) Non-zero first stage

Non-zero first stage – relationship between instrument and high score

IV Assumptions: (5) Monotonicity

Monotonicity – if clinician A strictly gives more high scores than clinician B, then any time clinician B would have given a high score, clinician A would have as well (they do not change places in strictness)

From arrest to social workers



Randomized assignment to therapists

- Each inmate is randomly assigned a therapist who interviews them for 15 minutes within 36 hours of booking
- Therapists assign a score (0-3) measuring the severity of mental and behavioral health symptoms as it relates to their **functioning** (important)
  - Inmates with no (0) or mild (1) functioning related symptoms skip MHC and are assigned to typical courts
  - Inmates with moderate (2) symptoms are assigned to private indigent defense attorneys (paid for by the county) in MHC
  - Inmates with severe (3) symptoms are assigned to MHC public defenders office
- After therapists score the inmate, he is assigned to a court and the therapist never sees them again (i.e., no therapy) which ensures exclusion holds in the data

Calculating the residualized leave-one-out mean

1. Regress observed *PubDef* onto a vector of time controls (day of year time fixed effects)
2. Calculate the residual, $\widetilde{D}_{dkt}$, from this regression
3. Use the residualized public propensity to recommend public defense rate to calculate the therapist recommendation instrument $\widetilde{Z}_{cl}$ as a leave-one-out mean rate of Public defense recommendation associated with each randomly assigned therapist l and inmate c

$$\begin{aligned} \widetilde{Z}_{cl} &= \left(\frac{1}{n_l - n_c}\right)\left(\sum_{k=0}^{n_l} \widetilde{D}_{dkt} - \sum_{k\in\{c\}} \widetilde{D}_{dkt}\right) \\ &= \frac{1}{n_l - 1}\sum_{k\neq c}^{n_l-1} \widetilde{D}_{dkt} \end{aligned} \tag{1}$$

2SLS estimating equations

$$PubDef_{dct} = \beta \widetilde{Z}_{cl} + \psi X_{dct} + \tau_t + \varpi_{dct} \quad (2)$$

$$Y_{dct} = \delta \widehat{PubDef}_{dct} + \gamma X_{dct} + \tau_t + \varepsilon_{dct} \quad (3)$$

where Y is the outcome of interest (e.g., repeat offending, suicide), *PubDef* is an indicator equalling 1 if the inmate was assigned to the public defender and 0 if the private indigent defense attorney; X are pre-court controls; τ_t are time fixed effects; $\widetilde{Z}$ is the residualized "leave-one-out-mean" average assignment to mental health court and errors are at the end of each equation.

Criticism of 2SLS: Over identification and bias

- Even though the 2SLS model is just identified with residualized leave-one-out-mean, our instrument is actually multi-dimensional in the number of clinicians and with weak instruments, this creates finite sample bias for 2SLS due to *many instruments*
- To help pin this down, consider that the propensity score theorem which states the propensity score is a scalar based on dimension reduction in X (Rosenbaum and Rubin 1983)
- This isn't really a just identified model so we should explore alternative models that are more appropriate for our data and design: we consider three

Alternative to 2SLS: JIVE

- One popular alternative to the 2SLS model in these applications has been the jackknife IV estimator (JIVE), but JIVE can be extremely biased with numerous *covariates*.
- Kolesar (2013) notes that in a finite sample, JIVE will be noisy and this estimation error will be correlated with the outcome since it depends on the treatment status of a particular inmate.
- This will cause JIVE to be biased when the number of covariates is large as is the case in our context – we have 14 covariates and 84 time fixed effects.
- We face therefore a tradeoff between a set of time fixed effects that ensure conditional randomization and the biases created by large numbers of covariates for our JIVE estimator.

Alternative to 2SLS: UJIVE

- To resolve this, we accompany our 2SLS and JIVE estimates with models that are more robust for large number of covariates as well as large number of instruments.
- Our first alternative is to estimate LATEs using the UJIVE estimator (Kolesar 2013)
- UJIVE estimates are robust to a large set of *covariates* and *instruments* by excluding inmate i from estimation, thus guaranteeing that the prediction error be uncorrelated with the outcome (Kolesar 2013)
- This means that UJIVE estimates are consistent for a convex combination of LATEs even when we have a large number of covariates.
- We implement this using Kolesar's matlab code.

Alternatives: LASSO

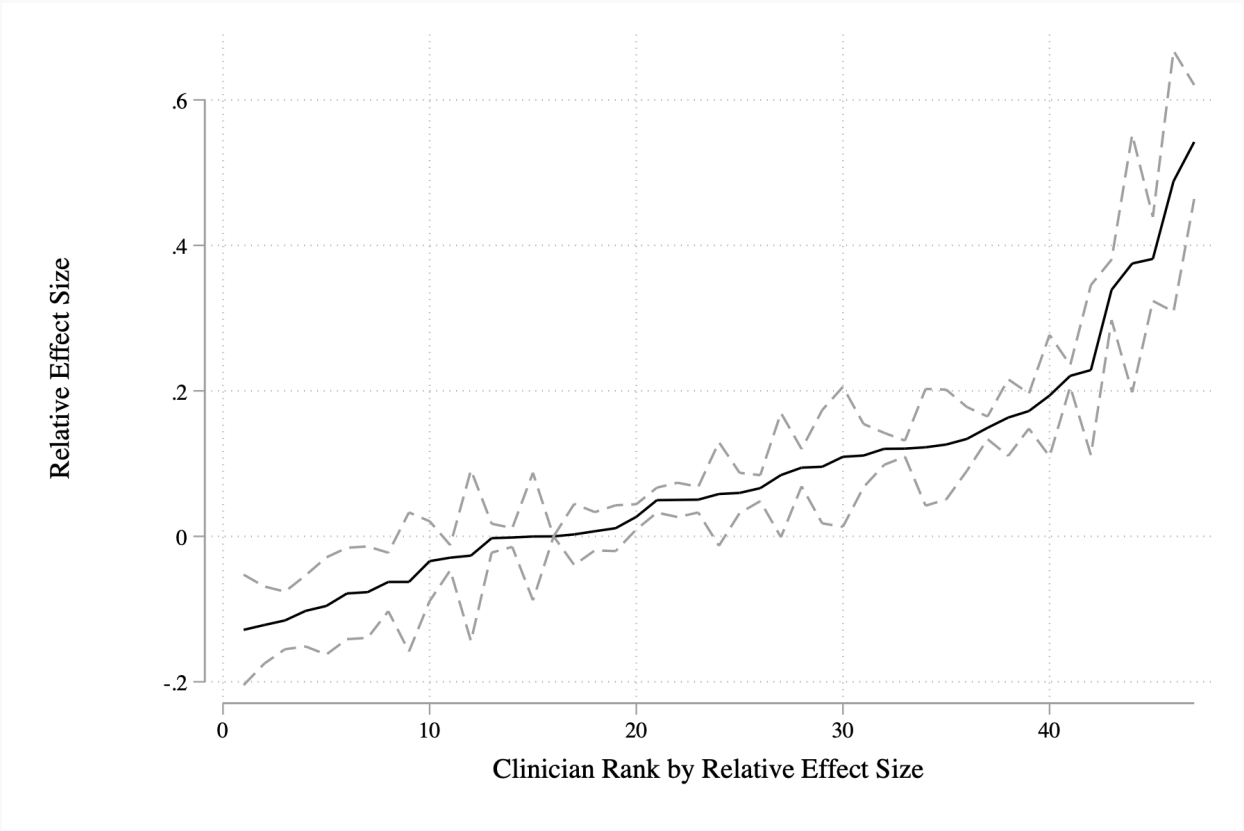
- We also present two machine learning selection IV models:
  - the post-double-selection model and
  - the post-lasso-orthogonalization method described by Chernozhukov (2015) which we loosely term our LASSO and post-LASSO models, respectively.
- These machine models models are designed to minimize the problems of including a large number of *instruments* (columns 3-4) as well as a large number of *controls* (columns 5-6).
- We use the Stata command `lassopack` for its implementation

Description of Data and Identification Tests

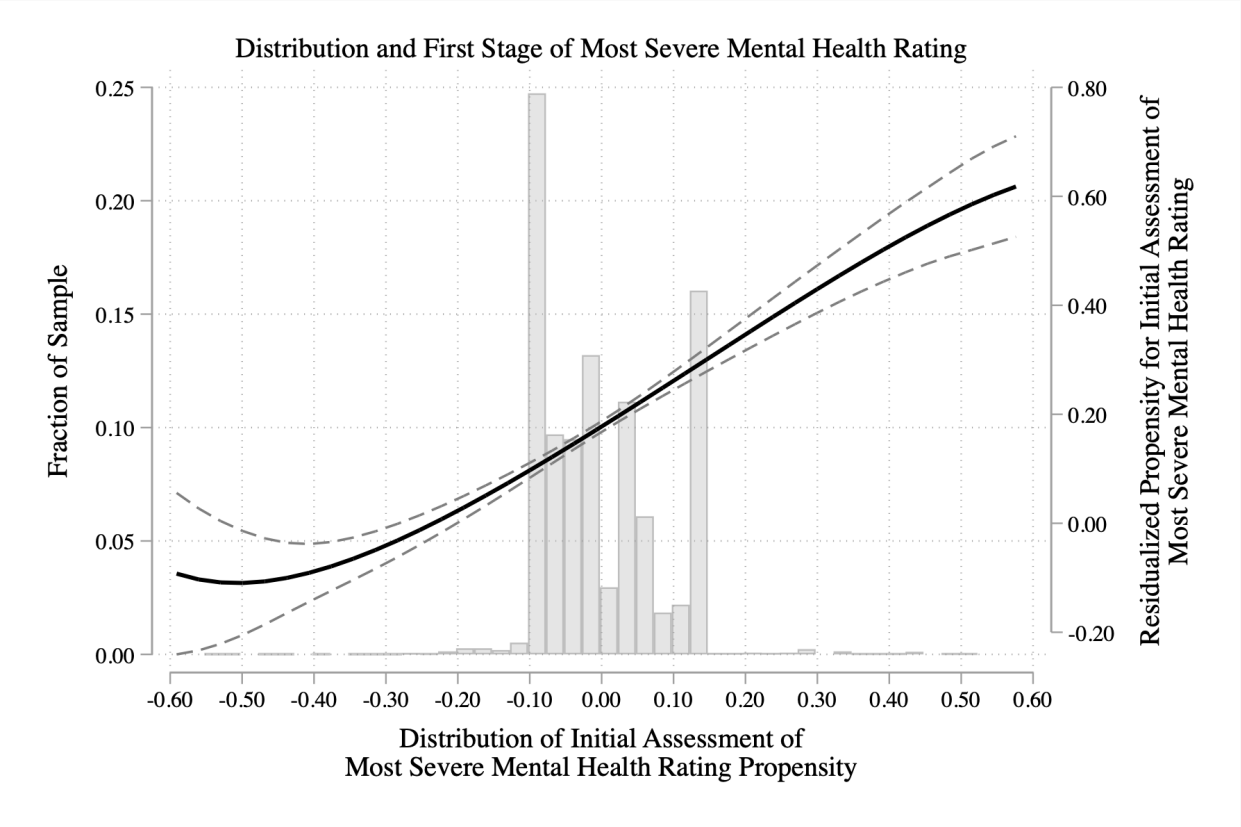
Systematic "biases" in first-time screeners"

- Assessment is brief and one-time event
- Imagine rating mental illness symptoms was like a blood test – then there'd be no variation in assessment
- As therapists see on average same types, without bias there would be no variation in recommendations
- But we do not find this – therapists appear to disagree, caused likely by discretion and "tendencies"

Visualizing therapist fixed effects



Visualizing residualized leave-one-out mean



Strength of first stage

Table 2: First Stage Regressions for Initial Assessment of Most Severe Mental Health Rating

	(1)	(2)
Z: Clinician's Leave-Out Mean Mental Health Score	0.635***	0.619***
	(0.152)	(0.150)
Kleibergen-Paap F	17.3653	17.1609
Time Fixed Effects	Yes	Yes
Baseline Controls	No	Yes
Observations	5,215	5,215

We report the first stage results of a linear probability model with outcome of interest being the initial assessment of an inmate's mental health being most severe as opposed to moderately 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. Column (1) shows the results by controlling only for day-of-week-month fixed effects, whereas Column (2) also includes 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 reported (which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013) in this case of a single instrument). * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 3: Instrument v. Inmate Characteristics for Public Defender

	Bottom Tercile	Middle Tercile	Top Tercile	Middle v. Bottom P-Value	Top v. Bottom P-Value
Z: Clinician's Leave-Out Mean Mental Health Score	-0.088	-0.020	0.107	(0.000)	(0.000)
Inmate Characteristics					
Asian	0.010	0.009	0.009	(0.717)	(0.679)
Black	0.279	0.256	0.253	(0.069)	(0.003)
Race other	0.001	0.001	0.002	(0.365)	(0.768)
Hispanic	0.202	0.227	0.202	(0.248)	(0.795)
Male	0.643	0.639	0.649	(0.976)	(0.892)
Age at booking	36.445	35.793	35.523	(0.372)	(0.133)
Prior offense w/in 365 days	0.380	0.372	0.421	(0.706)	(0.082)
Number of offenses per booking	1.606	1.581	1.637	(0.820)	(0.467)
First time in jail	0.025	0.018	0.011	(0.434)	(0.174)
Prior treatment	0.176	0.118	0.098	(0.420)	(0.363)
Prior medications	0.163	0.112	0.092	(0.451)	(0.380)
Prior hospitalization	0.136	0.089	0.073	(0.413)	(0.339)
Homeless	0.062	0.050	0.045	(0.658)	(0.688)
Jobless	0.090	0.074	0.045	(0.745)	(0.293)

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

Strict monotonicity test

- Frandsen, Lefgren and Leslie (2020) test joint null of exlcusion and monotonicity
- If you can argue one hold, then rejections mean the other doesn't hold
- We think exclusion plausibly holds given the research design, so rejecting the null speaks to strict monotonicity violations
- This test requires that the ATE among individuals who violate monotonicity be identical to the ATE among some subset of individuals who satisfy it.
- Their proposed test is based on two observations: that the average outcomes, conditional on judge assignment, should fit a continuous function of judge propensities, and secondly, the slope of that continuous function should be bounded in magnitude by the width of the outcome variable's support.

Monotonicity tests

Table 4: Frandsen, Lefgren, Leslie (2020) Test of Joint Null of Exclusion and Monotonicity

Outcome	FLL P-Value
Suicide attempt in next booking	0.000
Suicide ideation in next booking	1.000
Next booking mental health score improves	0.267

This table presents results from the test proposed in Frandsen, Lefgren, and Leslie (2020) for the joint null hypothesis that the monotonicity and exclusion restrictions hold. A failure to reject the null implies that we cannot reject the hypothesis that the monotonicity and exclusion restrictions jointly hold. This test was implemented in Stata via the package testjfe (Frandsen, 2020).

Average monotonicity tests

- Traditional monotonicity checks first stage's qualitative sign within subsample
- This checks whether the instrument is operating "in the same direction" weakly on average for a variety of demographic sub populations
- We divide samples by gender, race and age and first stage sign are the same, with some fluctuation in point estimates for female and Black samples

Table 5: Average Monotonicity for Initial Assessment of Most Severe Mental Health Rating

	Male	Female	Black	White	Hispanic	Age < 25	Age > 45
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Z: Clinician's Leave-Out Mean Mental Health Score	0.562***	0.766***	0.899***	0.547***	0.556***	0.598**	0.568***
	(0.164)	(0.152)	(0.179)	(0.147)	(0.200)	(0.254)	(0.170)
Observations	3,355	1,860	1,371	3,790	1,097	1,031	1,219
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the first stage results by subsamples as listed in the column headers, which serves as informal evidence of average monotonicity if the estimate significant across all subsamples. * p<0.10, ** p<0.05, *** p<0.01

Results

Table 6: Effects of Initial Assessment of Most Severe Mental Health Rating on Health Outcomes

	OLS		2SLS		JIVE	
	(1)	(2)	(3)	(4)	(5)	(6)
Suicide attempt in next booking	-0.020***	-0.016***	-0.158**	-0.122**	-0.221***	-0.174***
	(0.006)	(0.006)	(0.064)	(0.060)	(0.055)	(0.061)
			[-0.325, -0.053]	[-0.290, -0.035]		
Suicide ideation in next booking	-0.002	-0.002	-0.019**	-0.014*	-0.034*	-0.029
	(0.003)	(0.003)	(0.008)	(0.008)	(0.018)	(0.021)
			[-0.037, -0.003]	[-0.033, -0.000]		
Next booking mental health score	0.115***	0.136***	0.964***	0.981***	-4.848	-1.969*
	(0.036)	(0.037)	(0.274)	(0.249)	(6.378)	(1.173)
			[0.518, 1.619]	[0.577, 1.575]		
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	No	Yes	No	Yes	No	Yes

This table reports the ordinary least squares, two-stage least squares, and jacknived instrumental variables (Angrist et al 1999) estimates of impact of a clinician's initial assessment of a most severe mental health rating on inmates' subsequent mental health. The outcome variable interest are given in each row along with the corresponding estimates of the impacts of an initial assessment of a most severe mental health rat Two-stage least squares specifications instrument for severe mental health rating using a clinician leniency measure that is estimated using c from other cases assigned to a clinician as described in the text. We include day-of-week-month fixed effects for all specifications and base controls for Columns (2), (4), and (6). The clinician and inmate robust two-way clustered standard errors are shown in parentheses. For the 2S estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 7: IVLASSO Results for Initial Assessment of Most Severe Mental Health Rating and Suicidality Outcomes

	UJIVE		LASSO		Post-LASSO	
	(1)	(2)	(3)	(4)	(5)	(6)
Suicide attempt in next booking	-0.146***	-0.102***	-0.099***	-0.090***	-0.109***	-0.072*
	(0.034)	(0.029)	(0.033)	(0.032)	(0.039)	(0.043)
Suicide ideation in next booking	-0.018*	-0.012	-0.020***	-0.021***	-0.020***	-0.021***
	(0.010)	(0.010)	(0.004)	(0.005)	(0.004)	(0.004)
Next booking mental health score improves	1.012***	1.116***	0.469	0.526*	0.509*	0.577**
	(0.259)	(0.164)	(0.289)	(0.306)	(0.277)	(0.254)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes	No	Yes

This table reports Kolesar's (2013) Unbiased Jacknife Instrumental Variables Estimator (UJIVE) and the Instrumental Variables LASSO (IVLASSO) estimates of the impact of a clinician's initial assessment of a most severe mental health rating on inmates' subsequent mental health and criminality. The outcome variables of interest are given in each row along with the corresponding estimates of the impacts of an initial assessment of a most severe mental health rating. We include day-of-week-month fixed effects and baseline controls for all specifications; however, the IVLASSO procedure penalizes the controls as well as the instruments and can penalize them to zero as discussed in the text. The IVLASSO procedure is run using both methods: lasso-orthogonalization and post-lasso-orthogonalization as shown in Columns (5) and (6). The clinician and inmate robust two-way clustered standard errors are shown in parentheses for IVLASSO and Kolesar's (2013) robust standard errors are shown in parentheses for UJIVE. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Mechanisms

1. Lawyer selection – examine by focusing on recidivism
2. Social workers – cannot be directly evaluated since it is collinear with public defender assignment

But we argue if (1) is not there, then (2) is the most likely explanation

If (2), then it suggests it has a mental healthcare explanation, so we look at effects by whether they had a prior diagnosis

Table 8: Effects of Initial Assessment of Most Severe Mental Health Rating on Recidivism Outcomes

	OLS results		2SLS results			
	(1)	(2)	(3)	(4)	(5) No Prior Offense	(6) Prior Offense
Recid after current booking	0.053**	0.024	0.123	-0.016	0.056	-0.257
	(0.020)	(0.019)	(0.182)	(0.145)	(0.114)	(0.276)
			[-0.283, 0.458]	[-0.311, 0.279]	[-0.174, 0.265]	[-1.494, 0.506]
Recid within 1 year	0.029	0.023	0.006	-0.064	-0.010	-0.198
	(0.022)	(0.023)	(0.156)	(0.143)	(0.118)	(0.273)
			[-0.400, 0.293]	[-0.437, 0.199]	[-0.245, 0.180]	[-1.529, 0.507]
Count of future recidivism	0.135	0.043	0.700	0.278	0.243	0.132
	(0.082)	(0.083)	(0.589)	(0.469)	(0.255)	(1.086)
			[-0.501, 1.901]	[-0.677, 1.324]	[-0.273, 0.710]	[-2.876, 5.006]
LOS	13.442***	12.344***	16.056	17.292	10.626	19.776
	(1.905)	(1.648)	(15.244)	(14.792)	(10.486)	(26.409)
			[-9.105, 58.977]	[-4.222, 58.885]	[-6.545, 31.837]	[-23.087, 224.003]
Days to recidivism	-20.884*	-18.330	29.171	-17.727	-3.769	12.697
	(12.060)	(11.705)	(90.727)	(77.893)	(135.021)	(126.284)
			[-118.927, 229.537]	[-144.483, 153.767]	[-242.925, 235.386]	[-330.825, 332.529]
Next offense felony	-0.011	-0.020*	0.034	-0.020	0.039	-0.098
	(0.010)	(0.010)	(0.083)	(0.075)	(0.058)	(0.146)
			[-0.151, 0.202]	[-0.171, 0.132]	[-0.079, 0.146]	[-0.866, 0.251]
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	No	Yes	No	Yes	Yes	Yes

This table reports the ordinary least squares and two-stage least squares estimates of the impact of a clinician's initial assessment of a most severe mental health rating on inmates' subsequent mental health. The outcome variables of interest are given in each row along with the corresponding estimates of the impacts of an initial assessment of a most severe mental health rating. Two-stage least squares specifications instrument for severe mental health rating using a clinician leniency measure that is estimated using data from other cases assigned to a clinician as described in the text. We include day-of-week-month fixed effects for all specifications and baseline controls for Columns (2) and (4)-(6). The clinician and inmate robust two-way clustered standard errors are shown in parentheses. For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 9: No Prior Treatment Results for Initial Assessment of Most Severe Mental Health Rating and Suicidality Outcomes

	(1) OLS	(2) 2SLS	(3) JIVE	(4) UJIVE	(5) LASSO	(6) Post-LASSO
Suicide attempt in next booking	-0.016***	-0.122**	-0.175***	-0.102***	-0.049	-0.059
	(0.006)	(0.060)	(0.061)	(0.029)	(0.036)	(0.043)
Suicide ideation in next booking	-0.002	-0.014*	-0.029	-0.012	-0.016***	-0.016***
	(0.003)	(0.008)	(0.021)	(0.010)	(0.004)	(0.003)
Next booking mental health score improves	0.136***	0.981***	-2.100	1.116***	0.511	0.580*
	(0.037)	(0.249)	(1.307)	(0.164)	(0.366)	(0.298)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes

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

Marginal Treatment Effects

MTE and average treatment effects

- Why do we use marginal treatment effects (MTE)?
  - If there is heterogeneity in unobservables
  - To get a distribution of treatment effects
- What assumptions do we need?
  - Same ones as LATE including monotonicity
  - Additive separability of treatment effects (i.e., unobserved plus observed heterogeneity)
  - Only required if not full support, and we don't have full support

What are MTE?

- Derivative of symptom proxies with respect to propensity to assign the highest mental illness score
- Average effects of lawyer-with-social-workers on margin (i.e., between the low and high score)
- Heckman and Vytacil (2005, 2006, 2007, etc.) show that all policy relevant parameters (e.g., ATE, ATT, ATU) can be reconstructed using weighted averages over the MTE
- Since we don't have full support, we rescaled the weights so that they integrate to one over the trimmed sample so that we can calculate different weighted averages of the MTEs

Steps

1. Estimate the propensity score using logit (here issues of common support arise as we want it across all cells of Z)
2. Model suicidality/symptoms as a function of covariates and propensity score with second degree polynomials
3. Calculate the derivative of $\widehat{suicidality/score}$ with respect to the propensity score and plot it as a curve

Calculating aggregates

- We use Andresen's **mtefe** command in Stata
- ATE is the equally weighted average over the entire MTE curve, ATT as a weighted average over the left group of "low resistance, high propensity score", and ATU as the weighted average of the right group of ("high resistance, low propensity")
- Downward slope means selection on unobserved returns to lawyers-with-social-workers (i.e., ATT¿ATE¿ATU)
- Upward slope means the reverse (i.e., ATU¿ATE¿ATT)

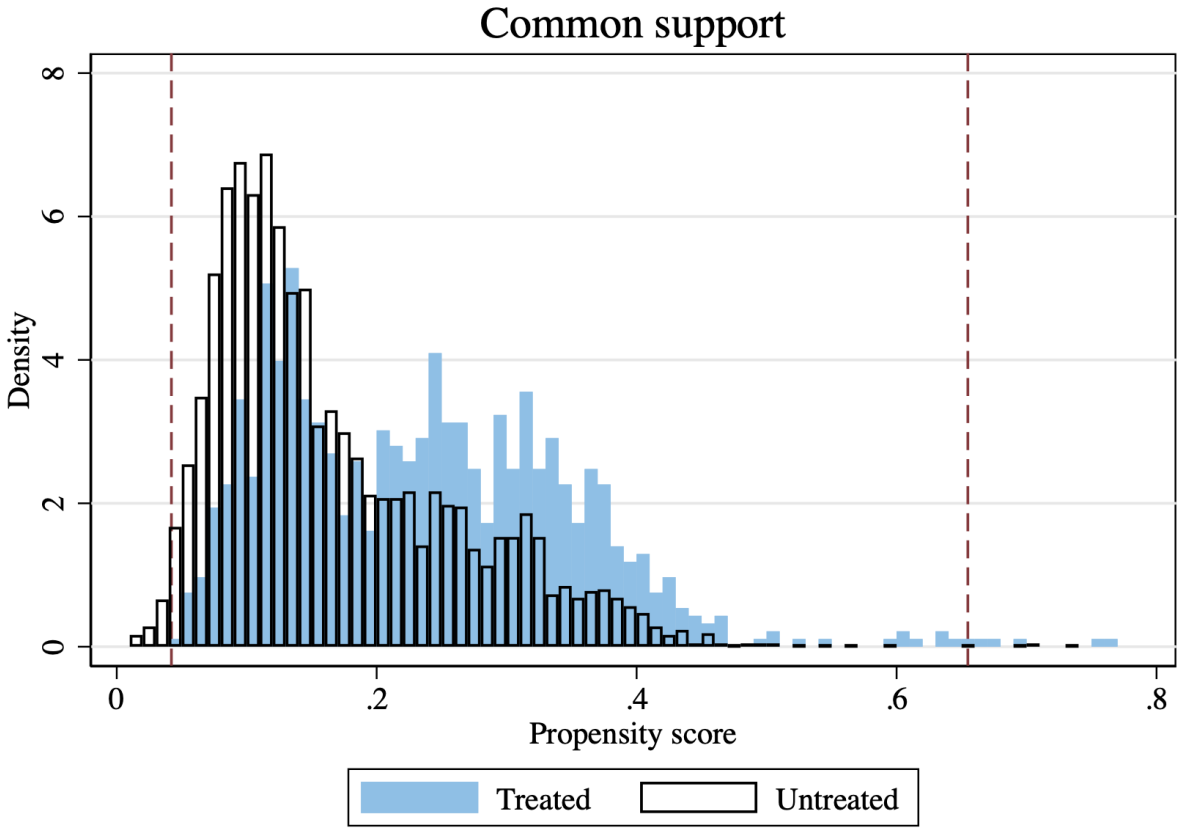
Preview of findings

- Suicide attempt only satisfies average monotoncity, so it's MTEs don't mean anything. We present it just in case someone in the audience still wants to see it.
- Suicidal ideation are small (around -1%) which is similar to our LATEs
- Future symptoms show signs of improving with lawyers-with-social-workers but mostly downward sloping suggesting that the returns are declining with the probability of being assigned a high score
- It could mean the severely mentally ill (high propensity of being assigned a high symptoms score) are more treatment resistant with respect to lawyers-with-social-workers

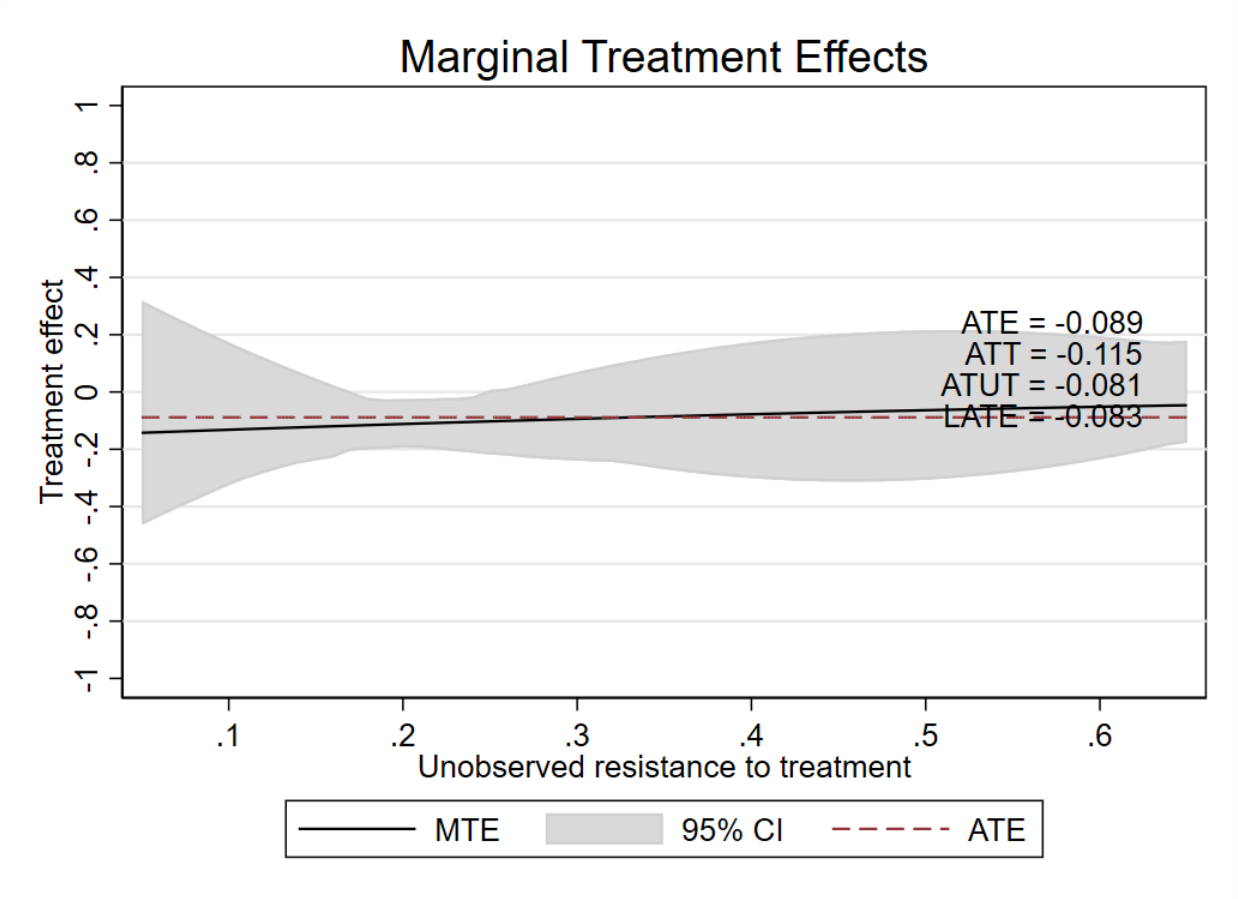
Some sources of confusion I had

- So it may help if you replace the phrase "high propensity score / low resistance to treatment" with "severe mental illness" (high scores)
- If I have schizophrenia with extreme displays of psychosis, then I am **less** resistance to treatment because the treatment is a high score, and I will almost certainly get a high score (high propensity score)
- If I come in with depression but am functional, my score is lower and so I both have a lower propensity score *and* therefore have a *higher* resistance to treatment
- Note the subtle language: "propensity score" and "resistance to treatment" are reversed – people with less resistance have higher propensity scores

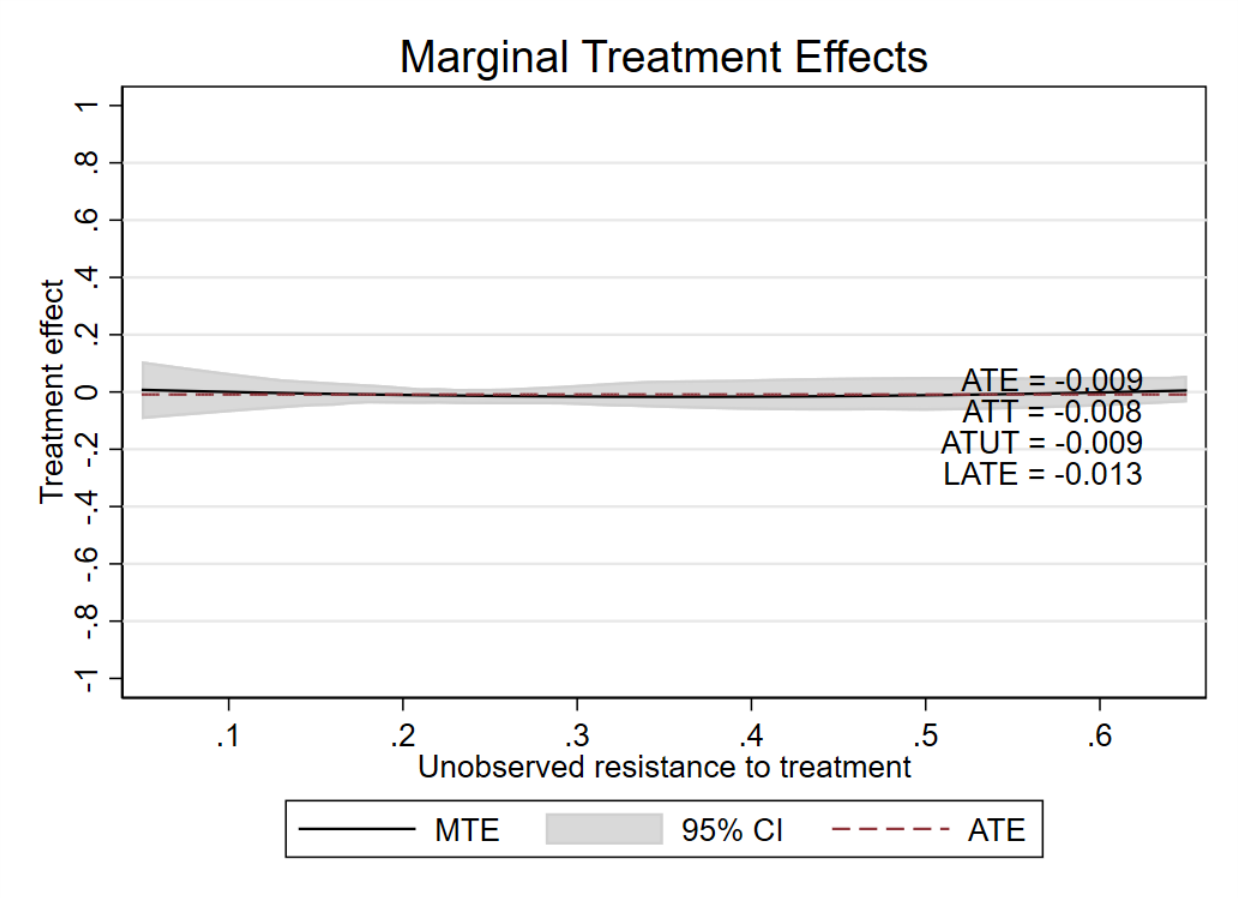
Common support



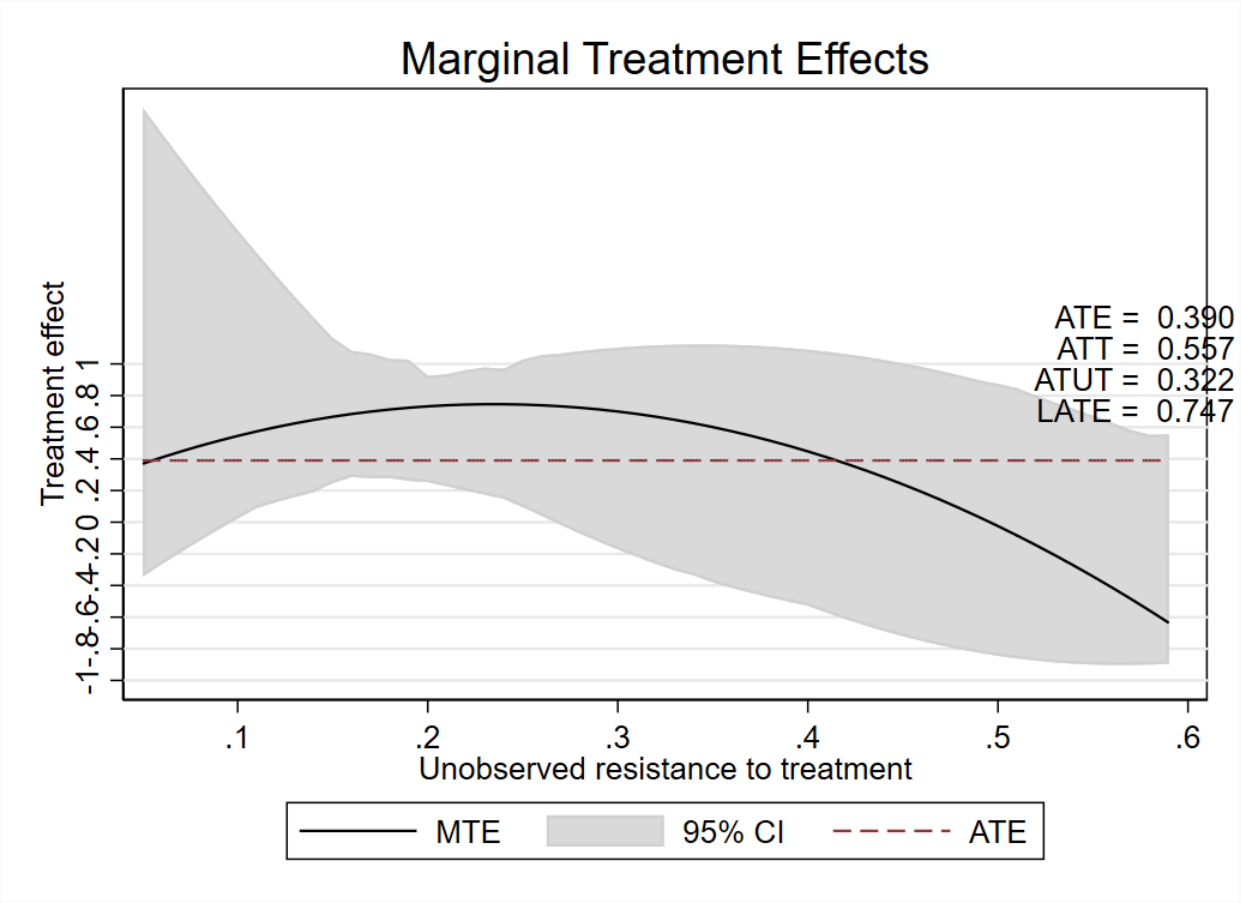
MTE and aggregate parameters for suicide attempt



MTE and aggregate parameters for self reported suicidal ideation



MTE and aggregate parameters for next visit symptom score



Discussion

Diminishing marginal returns to mental health

- Assuming modest statistical value of life, high rates of mental illness in jails, and high rates of suicide in jails, mental health court, when accompanied with social workers, could be productive inputs
- But MTE analysis suggests there is some evidence that the returns do taper off as we move closer to the "least well off" (i.e., the severely mentall ill)

Design elements

- Severely mentally ill are a deeply treatment resistant group, often noncompliant with treatment entirely, cycling in and out of homelessness, jails, emergency rooms and likely much higher incidence of mortality than we may know (or that I currently know)
- Some evidence suggests that the social costs of severe mental illness are extremely high (Biasi, et al. 2021) and may even far exceed that of more moderate forms of mental illness (Jaffe 2017)
- Mental health courts and social workers seem to help, but the effects taper off as we move closer to the severely mentally ill for some. but not all things
- Expect to see more of this as police reform movers into broader criminal justice reform focusing on all levels of the mental illness criminal justice pipeline