For Online Publication

Appendix for "How do mental health treatment and treatment delays impact long term mortality?"

Sydney Costantini

A1 Alternative Samples

A1.1 Patients with no outside insurance

In Section III.E.ii, I discuss a potential issue with the interpretability of my results: what if patients also receive mental health care outside of the VA? To address this issue, I create a subsample of patients who have no outside or non-VA insurance, making it highly likely that they receive care only at the VA. This subsample comprises 69.7% of patients in the primary sample. In Table A1, I show that across characteristics, the no insurance subsample is highly analagous to the main sample, with two exceptions. First, the mean age is slightly younger compared to that of the main sample. Second, two-year mortality in this subsample is 5.3%, as opposed to 7.5% in the main sample. Both of these facts likely reflect the fact that most patients with outside insurance have Medicare, which is only available for patients above 65 or for those who qualify due to disability status.

Mortality results for this sample are presented in Table A4.

A1.2 Patient without scheduled follow up visits

In the sample selection described in Section II.A, I exclude patients with no scheduled follow up visit, as W_{-i} depends on type of appointment scheduled. Since this group is comprised of a substantial 314,554 patients, in Table A1, I also compare characteristics of patients with no scheduled follow up appointments with those of the main sample. As expected, I find that these patients have much lower prior attachment to VA mental health services and are more likely to have Medicare, implying

that more of these patients are using care outside of the VA. Additionally, patients with no follow up appointment are older and are much more likely to be diagnosed with SUD, suggesting that these types of patients may be particularly unlikely to return for mental health services.

Finally, these patients have slightly higher mean wait time than that of the main sample. This fact implies that those who arrive during more congested times may be less likely to have a follow up appointment scheduled at all. As this presents a potential non-random attrition problem that could bias results, I add these patients back to the main sample in Appendix Section A3.1.

A2 Additional balance checks

As the VA flags patients considered to be high risk for suicide, a natural additional balance test is to see whether these high risk patients are just as likely to be assigned low and high values of W_{-i}. I restrict to patients flagged as high risk before the ED visit, as being assigned this flag at the ED visit itself could potentially be endogeous to my experiment. Overall, only 0.3% of patients in my sample are assigned the high risk flag before the ED visit. In Figure A2 Panel A, I show a binscatter of percent of patients assigned the high risk flag versus W_{-i}, residualizing by the baseline controls. I find a small and statisitically insignificant coefficient, providing more evidence that mean wait time is approximately randomly assigned.

Secondly, another key metric indicative of patient severity is whether a patient has a prior hospitalization. Let I_i indicate that a patient was hospitalized in the year prior to the index ED visit (for any cause). In Figure A2 Panel B, I show a binscatter of I_i versus W_{-i}, residualizing by the baseline controls. I find that the coefficient on W_{-i} is both reverse-signed and insignificant, providing further evidence that W_{-i} is uncorrelated with patient severity.

A3 Additional robustness checks

A3.1 Patients without scheduled follow up appointments

To eliminate concern that dropping these patients has implications for my findings, I add patients with no scheduled follow up appointments back to the sample, increasing the sample size to 935,843. This combined sample has a somewhat higher mean two-year mortality of 9.6%.

As the primary construction of W_{-i} and controls rely on assigned appointment type, I make a few modifications to the design in order to run regressions using this larger sample. First, since patients with no scheduled follow up do not have an assigned appointment type, I use two alternative approaches to approximate congestion levels for these patients. I start with a simple average approach: I group all appointment types together during two-week-hospital periods to construct W_{-i} for these patients. Alternatively, I let $W_{-i} = \hat{W}_{-i}$, as described in Section III.C.i. of the main text.[1] In this latter approach, $\hat{W}_{-i}$ is essentially a weighted average mean wait time, weighted by the probability of the patient being assigned a particular appointment type. As appointment type is also a baseline control in my main analysis, in both specifications, I control for either appointment type directly (for the baseline sample) or the predicted probability of being assigned each appointment type (for patients with no appointment type).

In Table A3, I show regression results using both of the aforementioned methods. I show results both with the baseline controls and all additional controls used in Table 1; results are both robust and very similar to those of the baseline analysis sample.

A3.2 Miscellaneous robustness checks

In Figure A3, I show that the baseline results are stable when both 1) detailed patient controls are added, and 2) sensible modifications to the sample or treatment variable are made. Under "Additional controls," I first show the main result with the baseline controls– two-week periods, year-month × hospital, appointment type, and age buckets. Then, I incrementally add groups of additional controls (the next row adds demographics controls, the following row adds demographic and prior utilization controls, etc.).[2] Demographic controls include sex, race, and ethnicity. Prior utilization controls includes count of inpatient and outpatient mental health visits in the previous year, count of mental health appointment cancellations, and count of no shows. Diagnosis controls are 3-digit ICD-9 diagnosis codes associated with the ED visit. Priority status refers to a patient's eligibility for VA healthcare and responsibility for paying copays, which depends on both disability status and income. The second to last row in this section adds leave-out mean characteristics of patients who form a given patient's mean wait time. Leave-out characteristics include mean age, mean prior visits, and

[1] W_{-i} for observations also in the baseline sample are held fixed in these analyses.

[2] These controls overlap with the controls added in Table 1.

mean predicted mortality. Finally, the last row controls for first half versus second half of the month indicators. The motivation behind these final controls is that different types of patients could arrive to the hospital during the beginning or end of the month, perhaps driven by holidays, paychecks, or benefit schedules.[3]

Under "Sample changes", I consider some possible reasonable modifications to the sample and construction of W_{-i}.[4] First, the primary sample considers any appointment within a 90-day window post ED visit to be a follow up appointment. In contrast, in the first row, I only consider follow up appointments within a 60-day window; patients with a first appointment scheduled between days 61 and 90 are dropped. In the second row of this section, related to the concern that holidays may create non-idiosyncratic variation in crowding, I drop all ED visits in December.

Next, I present results using a slightly modified version of W_{-i}. One potential concern is that wait times of patients who arrive during the same time window as a given patient may be influenced by the patient's own wait time (for example, if a high risk patient is pushed to an earlier appointment slot, this may cause more delays for other patients who arrive at a similar time). To mitigate this concern, in this alternative version of W_{-i}, I leave out wait times of not only the patient themselves, but of all patients who arrive on the same day. This robustness check also accounts for the potential of correlated same-day patient shocks, such as multiple veterans ingesting the same contaminated drug.

In Appendix Table A5, I perform an additional series of robustness checks. To capture potential correlation in treatment assignment within a given hospital, in column 2, I use hospital level rather than hospital-month level clustering. In column 3, I add day of week fixed effects. In column 4, to capture potential seasonal variation in types of patients arriving, I interact month of the year with ICD-9 3-digit diagnosis group.

Additionally, one potential concern is that those with prior mental health utilization are differentially affected by congestion. Those with prior mental health utilization likely have an advantage in obtaining timely follow up appointments, as they may have pre-existing relationships with providers. However, W_{-i} does not use a patient's wait time directly; instead, it relies on the wait times of patients arriving at the same time as the index patient. Consequently, W_{-i} should exogenously capture clinic congestion, rather than capturing anything about a patient's prior utilization. However, as an

[3]Note that ability to pay is not a major concern here, since the vast majority of mentally ill patients receive free mental health services at the VA.

[4]When the main sample selection is modified, the construction of W_{-i} is held constant, unless otherwise noted.

additional test to ensure that W_{-i} only captures congestion, rather than variation in patients' prior utilization which may affect wait times, in column 5, I perform a robustness check where I construct W_{-i} only using patients with no prior mental health utilization.

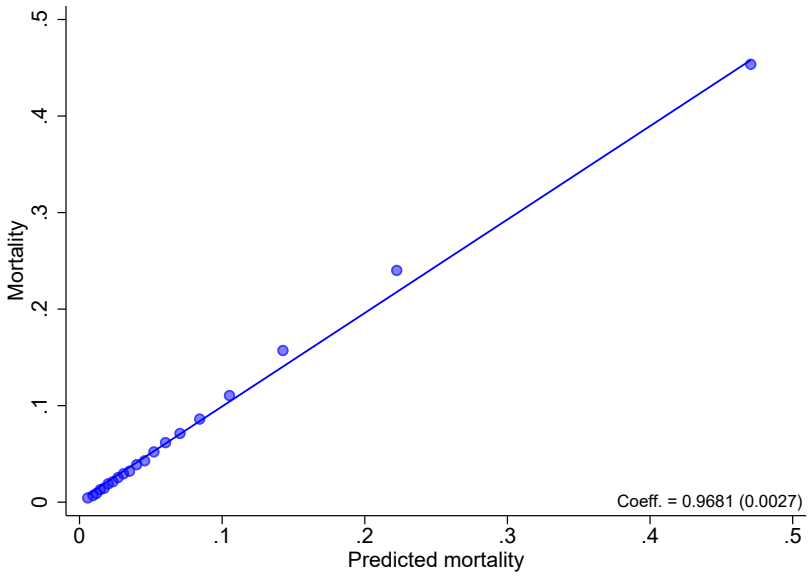
Finally, although all patients in my sample have a mental health diagnosis code connected with their visit, some patients who visit the ED may be experiencing both a mental and physical health emergency. Specifically, 39.9% of the sample also has a *physical* health diagnosis connected to their ED visit, the most common of which are hypertension, sleep disturbances, and respiratory distress. Some of these may be directly connected to mental health – for instance, those with anxiety may experience breathing issues or difficulty sleeping. In column 6, I add controls for physical health diagnoses assigned at the ED visit, including an indicator for no physical health diagnosis.

Table A1: Patient characteristics across samples

	Main sample	Patients with no other health insurance	Patients with no scheduled follow up visit
Male	0.908	0.905	0.927
Black	0.232	0.254	0.205
Hispanic	0.057	0.059	0.048
Age	51.4	47.9	57.8
Prior outpatient visit	0.573	0.547	0.188
Prior inpatient visit	0.096	0.103	0.027
Substance use disorder	0.285	0.321	0.446
Mood disorders	0.299	0.300	0.142
Anxiety	0.120	0.115	0.140
Psychosis	0.092	0.083	0.070
PTSD	0.103	0.101	0.048
Medicare	0.168	0.000	0.247
Other non-VA insurance	0.135	0.000	0.131
Mean wait time	17.73	17.55	18.30
Mortality	0.075	0.053	0.137
N	621,289	432,863	314,554

Note: This table shows characteristic means for the baseline sample compared to both the subsample of patients with no non-VA insurance and the dropped sample of those who have no scheduled follow up mental health visit. Prior oupatient and inpatient visit are indicators for attending a mental health appointment in the previous year. Mean wait time is equivalent to W_{-i} for the baseline sample. To form mean wait times for patients without a follow up mental health visit, I group all appointment types together, as I cannot observe which type of mental health clinic they would have been scheduled for.

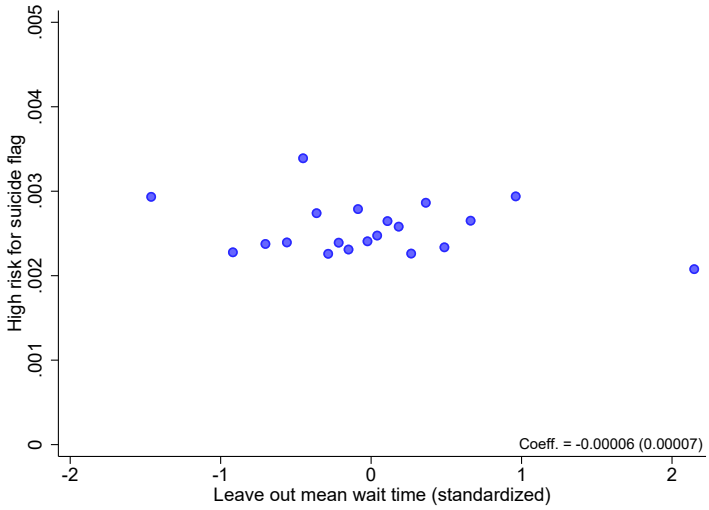
Figure A1: Predictive power



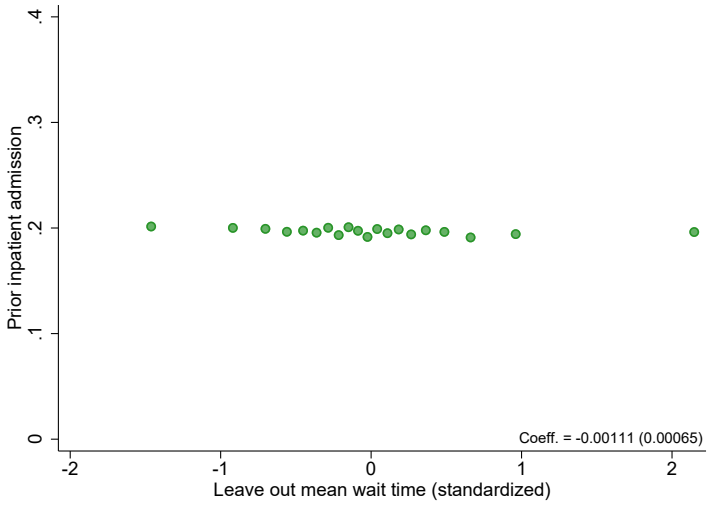
Note: This figure plots observed mortality, Y_i, versus predicted mortality, $\hat{Y}_i$, in twenty equal sized bins. The coefficient shows the regression fit of observed mortality on predicted mortality.

Figure A2: Alternative balance tests

A: Balance with respect to high risk for suicide flag



B: Balance with respect to prior inpatient admission



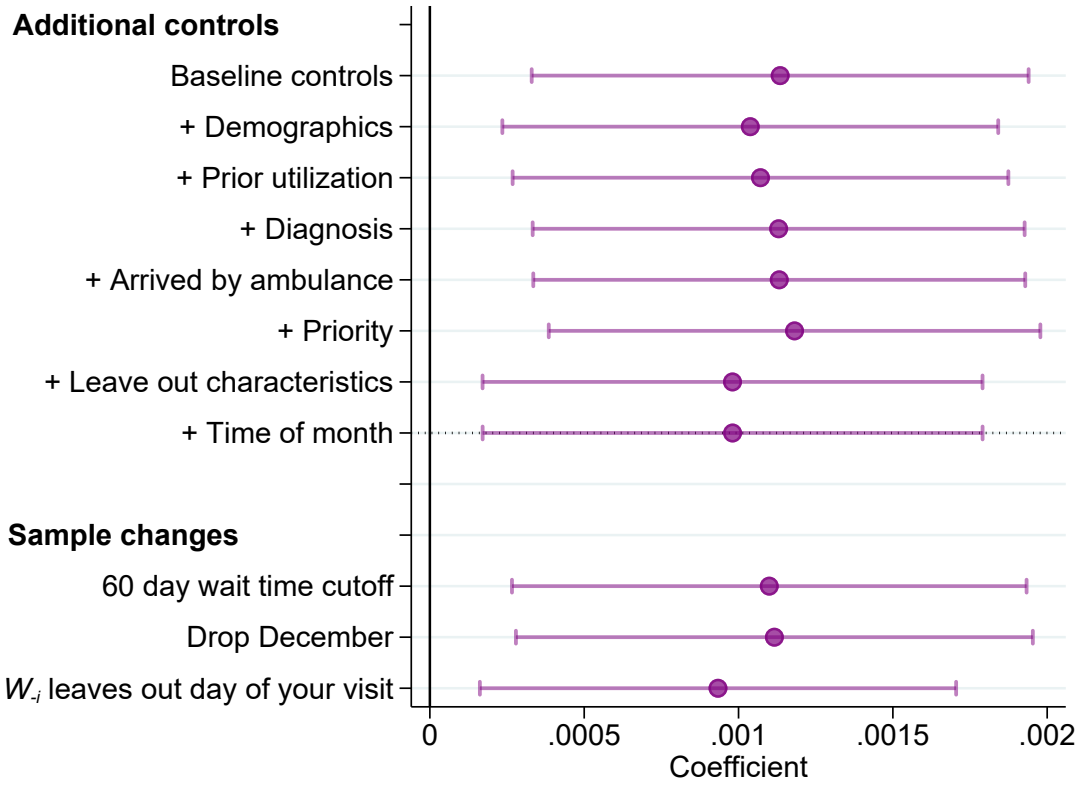
Note: Panel A shows a binscatter of the percent of patients assigned the high risk for suicide flag versus W_{-i}, residualized by the baseline controls. Panel B shows a binscatter of prior inpatient admission, I_i, versus W_{-i}, residualized by the baseline controls.

Table A2: Selection into appointment type

	General	Psychiatry	PTSD	Substance use	Other
Mean wait time$_k$	-0.0047	-0.0035	-0.0027	-0.0018	-0.0017
	(0.0011)	(0.0007)	(0.0011)	(0.0005)	(0.0007)
Mean wait time$_k \times \hat{Y}$	-0.0043	-0.0053	-0.0013	0.0002	-0.0013
	(0.0011)	(0.0007)	(0.0011)	(0.0006)	(0.0008)
Outcome mean	0.612	0.108	0.069	0.059	0.152
Observations	621,289	621,289	621,289	621,289	621,289

Note: This table shows regression results corresponding to Eq. 5 in the main text. The left hand side variable is an indicator for the patient being assigned a given appointment type (general, psychiatry, PTSD, substance use, or other specialty appointment). The first row shows β_1, the coefficient on residualized wait time for appointment type k, and the third row shows β_3, the coefficient on the interaction between residualized wait time and predicted mortality. Wait times are residualized by hospital $\times$ month, two week period, and age buckets. $\hat{Y}_i$ is standardized for ease of interpretation. General includes therapy visits with psychologists and social workers. All standard errors are clustered by hospital-month.

Figure A3: Robustness to additional controls and sample changes



Note: This figure shows the effect of not attending a mental health appointment on 2-year mortality when various changes to the controls, sample, or wait time measure are made. Dots are point estimates; bars show 95% confidence intervals. Under "Additional controls," I first show the main estimate using the baseline controls, two-week period, year-month × hospital, appointment type, and age buckets. I then incrementally add additional groups of controls. Demographics includes sex, ethnicity, and race. Prior utilization includes prior mental health outpatient visits, no shows, and cancellations, as well as prior inpatient mental health visits. Diagnosis refers to 3-digit ICD-9 diagnosis codes from the ED visit. Priority status, a designation between 1 and 8 assigned by the VA based on income and disability status, indicates eligibility for VA healthcare and copay responsibilities. Leave out characteristics include mean age, prior visits, and predicted mortality among patients in one's hospital-time-appointment group (the same group that forms a patient's W_{-i}). Time of month are indicators for first or second half of the month. Under "Sample changes," I make various modifications to the sample selection or construction of W_{-i}. The first row amends the definition of follow up appointment from 90 days to 60 days, the second drops ED visits in December, and the third row excludes all visits on the same day as the patient themselves from the construction of W_{-i}.

Table A3: Main results, including patients with no scheduled follow up

	2-year mortality			
	(1) Average wait time	(2) Average wait time	(3) Predicted wait time	(4) Predicted wait time
Mean wait time	0.00070	0.00095	0.00118	0.00124
	(0.00034)	(0.00035)	(0.00035)	(0.00035)
Outcome mean	0.0956	0.0956	0.0956	0.0956
Observations	935,843	935,843	935,843	935,843
Baseline controls	Yes	Yes	Yes	Yes
Additional patient controls	No	Yes	No	Yes
Leave out controls	No	Yes	No	Yes

Note: This table shows baseline mortality results, adding patients back to the sample who had no scheduled follow up appointment. "Average wait time" refers to forming W_{-i} by grouping all appointment types together (for those without a scheduled follow up). Predicted wait time refers to setting $W_{-i} = \hat{W}_{-i}$ for these patients. All standard errors are clustered by hospital-month.

Table A4: Effect of mental health wait times on mortality for no insurance subsample

	(1)	(2)	(3)	(4)
	No visit	Mortality	Mortality	Mortality
Mean wait time	0.0218	0.00075	0.00082	0.00078
	(0.0010)	(0.00044)	(0.00043)	(0.00044)
Outcome mean	0.3624	0.0528	0.0528	0.0528
Observations	432,863	432,863	432,863	432,863
Baseline controls	Yes	Yes	Yes	Yes
Additional patient controls	No	No	Yes	Yes
Leave out controls	No	No	No	Yes

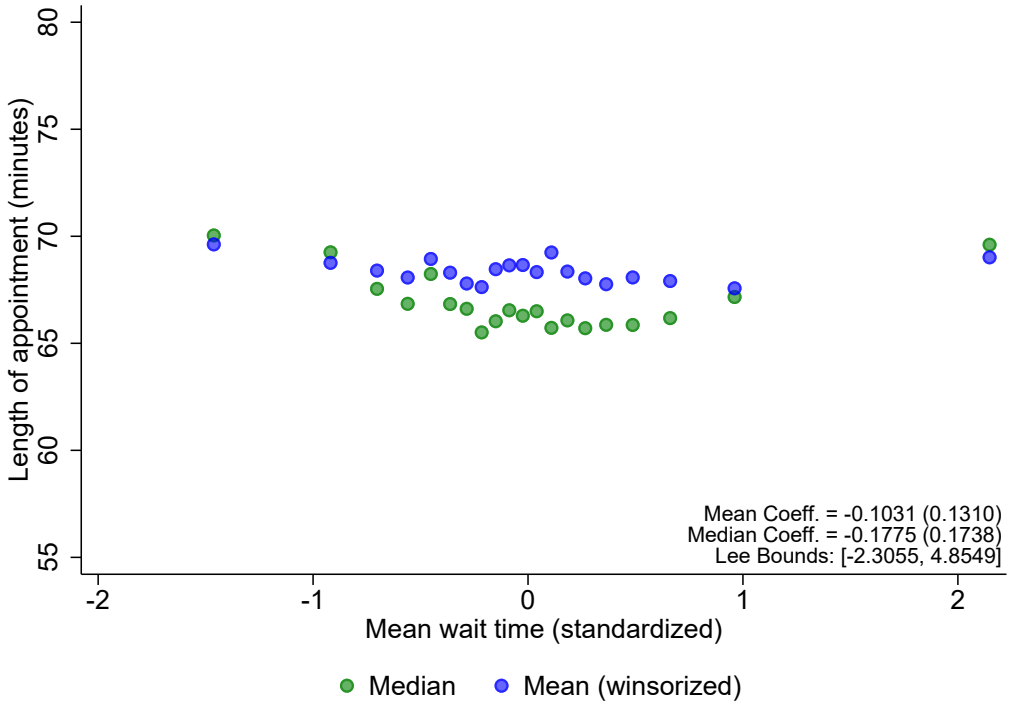
Note: This figure shows the main results for the subsample who have no other non-VA insurance. Column 1 shows the effect of mean wait time on the probability of not attending a follow up mental health visit. Column 2 shows the main mortality estimate, only using baseline controls. Columns 2 shows the estimate when I add additional patient controls, including gender, race, diagnosis, and prior utilization. Finally, column 4 shows the estimate when I control for characteristics of other patients who form a given patient's value of W_{-i}, including (leave-out) mean age, mean prior visits, and mean predicted mortality. Baseline controls include month-year $\times$ hospital, appointment type, and age buckets. Standard errors are clustered by hospital-month.

Table A5: Additional robustness checks

	(1) Baseline	(2) Hospital-level clustering	(3) Day of week controls
Mean wait time	0.00113	0.00113	0.00114
	(0.00041)	(0.00044)	(0.00041)
Outcome mean	0.0746	0.0746	0.0746
Observations	621,289	621,289	621,289
	(4) Month × disease type controls	(5) Mean wait time excludes those with prior utilization	(6) Physical health diagnosis controls
Mean wait time	0.00121	0.00096	0.00098
	(0.00041)	(0.00040)	(0.00041)
Outcome mean	0.0746	0.0751	0.0746
Observations	621,289	575,734	621,289

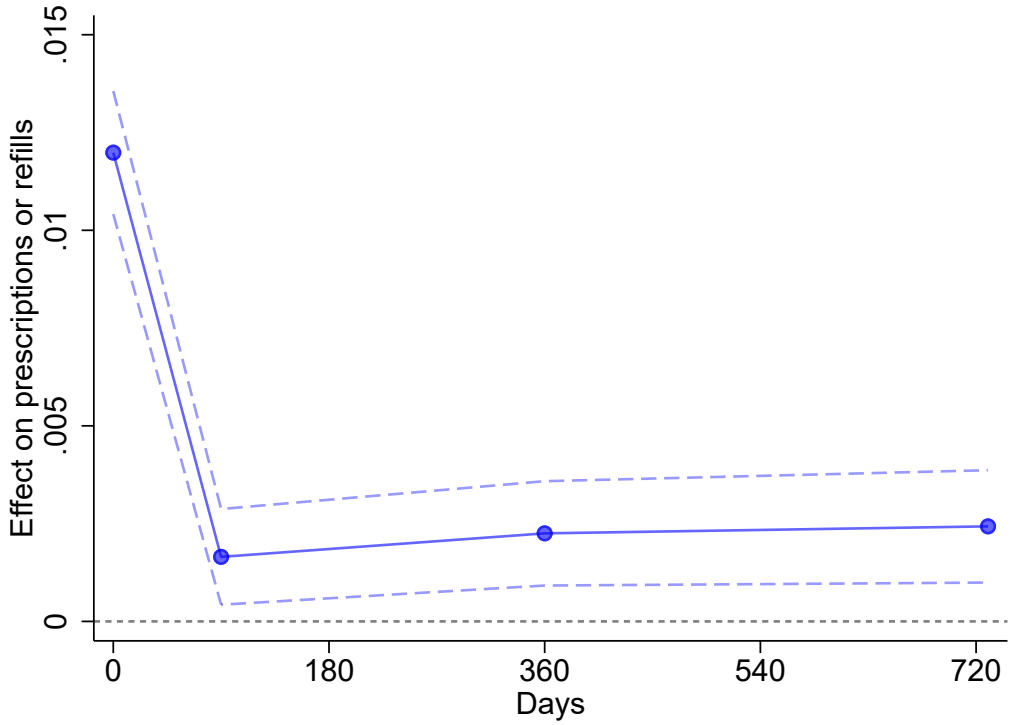
Note: This table shows additional robustness checks for the mortality estimate. The fourth column interacts month of the year with ICD-9 3-digit diagnosis group. The fifth column constructs W_{-i} only using patients with no prior mental health utilization. The sixth column controls for physical health ICD-9 codes assigned at the ED visit.

Figure A4: Length of follow up appointment vs. mean wait time



Note: This figure shows a binned scatterplot of the length of one's follow up appointment, in minutes, versus W_{-i}, residualized by the baseline controls. The mean reported coefficient winsorizes observations at the 5th and 95th percentiles, while median reported coefficient calculates the median wait time in hospital-two-week-appointment type bins. Lee Bounds reports worst case scenario bounds using median wait times.

Figure A5: Effect of mean wait time on prescriptions over time



Note: This figure shows the effect of mean wait time on mental health prescriptions or refills. The first dot shows the effect of mean wait on having any drug prescribed at the ED visit. The second dot shows the effect on having any drug prescribed between days 0 and 90 (including the day of the ED visit). Finally, the third and fourth dots show the effect on prescriptions or refills between days 90-360 and 360-730, respectively.

Table A6: Cause of death

	High wait time patients	Full sample	Risk ratio
All cause mortality	0.0756	0.0745	1.015
	(0.0004)		
Physical disease	0.0660	0.0648	1.018
	(0.0004)		
Accident, suicide, or overdose	0.0096	0.0096	0.993
	(0.0002)		
Heart disease	0.0169	0.0163	1.034
	(0.0002)		
Cancer	0.0131	0.0130	1.006
	(0.0002)		
Respiratory diseases	0.0071	0.0068	1.045
	(0.0001)		
Liver disease	0.0049	0.0047	1.025
	(0.0001)		
Other cardiovascular	0.0038	0.0036	1.034
	(0.0001)		
Infections	0.0030	0.0029	1.039
	(0.0001)		
Other physical disease	0.0203	0.0203	1.001
	(0.0002)		

Note: This table shows the probability of dying from a particular cause for both the full sample and for those with a mean wait time one standard deviation above the mean. The last row shows the relative risk of dying from that cause for those with high wait times versus the full sample.

Table A7: Effect of mean wait time on intermediate outcomes for substance use subsample

	A: Mental health outcomes		
	(1) No follow up visit	(2) Returned to mental health care	(3) Mental health inpatient treatment
Mean wait time	0.0239	-0.0104	-0.0088
	(0.0016)	(0.0009)	(0.0015)
Outcome mean	0.3385	0.9286	0.4151
Observations	175,721	175,721	175,721
	B: Physical health outcomes		
	(4) Primary care visit	(5) Inpatient visit - non-mental health	(6) New physical comorbidity
Mean wait time	0.0027	0.0044	0.0032
	(0.0011)	(0.0013)	(0.0014)
Outcome mean	0.8317	0.2323	0.6529
Observations	175,721	175,721	175,721

Note: This table shows intermediate outcomes for the SUD subsample. Panel A shows results for engagement with mental health care. Returned to mental health care refers to attending any oupatient appointment. Mental health inpatient treatment refers to any mental health inpatient visit, including residential stays (excluding admissions within one day of the initial ED visit). Panel B presents results for non-mental health outcomes. The first column looks at whether a patient attended a primary care appointment, the second column looks at non-mental health inpatient admissions, and the third column looks at whether the patient had any new Elixhauser comorbidity (only including physical conditions) during the two-year follow up period.