Pathways Into Opioid Addiction: Evidence From Practice Variation in Emergency Departments

Sarah Eichmeyer and Jonathan Zhang

Online Appendix

Appendices

A. Comparison with Barnett, Olenski, and Jena (2017) and Barnett et al. (2019)

We would like to begin by thanking Michael L. Barnett, Walid Gellad, Anupam B. Jena, and their coauthors for their suggestions, comments, and clarifications. In this appendix, we describe the differences between our paper and Barnett et al. (2017, 2019), as well as study how and where our findings depart from theirs.

The two papers listed above are the two most closely related to ours. In Barnett, Olenski, and Jena (2017), the authors study long-term use (180 days supply in 12 months) following an ED visit for opioid-naïve Medicare beneficiaries who see a high or low intensity prescriber. With a 20% random sample of Medicare claims from 2008-2011, physicians are classified as high (low) intensity if their overall prescription rate over those four years falls in the top (bottom) quartile within their hospital. The authors find that being treated by a high intensity prescriber is associated with a 0.35pp (30%) increase in the probability of long-term use. They also study a set of secondary outcomes including hospitalizations, ED visits, falls or fractures, constipation, respiratory failure, and opioid poisoning in the following year. They find higher rates of falls or fractures and opioid poisoning associated with high intensity prescribing.

Barnett et al. (2019) use 2012 VHA data to replicate their previous study (identical sample selection and research design) and find a 0.13pp (11%) increase in the probability of long-term opioid use among veterans. They study the same secondary outcomes and fail to

find any statistically significant difference.

Differences Between the Papers

The key differences between these two papers and ours can be grouped into two categories: i) patient outcomes, and ii) econometric specification and sample construction. In terms of patient outcomes, both Barnett, Olenski, and Jena (2017) and Barnett et al. (2019) focus primarily on long-term prescription opioid use (180 day supply in the first year after the ED visit) as their main outcome, along with opioid-related hospitalizations such as falls, fractures, and poisonings as secondary outcomes. Our paper studies additional long-term outcomes including opioid use disorder, proxies for opioid-seeking behavior, overdose mortality, and proxies for illicit opioid use. In addition, we supplement the observed VHA opioid prescriptions in Barnett et al. (2019) with Medicare and Medicaid claims and VHA reimbursed community care.

Econometrically, Barnett, Olenski, and Jena (2017) and Barnett et al. (2019) classify emergency physicians as high and low "intensity" prescribers, similar in spirit to our "leniency" instrument. They do this by first calculating each physician's raw opioid-prescribing rate as the number of emergency visits resulting in a prescription, divided by the total number of emergency visits. They construct one aggregate rate (lumping all years together in the 2017 paper) per physician. They then classify physicians as high (low) intensity prescribers if they fall in the top (bottom) quartile within their hospital.

Our paper utilizes a residualization approach, as described in subsection 3.2, leveraging detailed information about time of day, day of week, age, diagnosis, and pain score, thus eliminating some selection of patient arrival to ED or physician work schedules. Further, we leave out patient-physician pairs' own residual, eliminating the mechanical bias that stems from a patient's own case entering into the instrument. When the number of cases observed for each physician is small, this bias is large and approaches the OLS bias. Our leniency measure is also year-varying, allowing for physicians learning about the risks and benefits of prescription opioids during this time period.

The papers differ in terms of our sample selection as well. Barnett, Olenski, and Jena (2017) focus on all non-admitted emergency department conditions (diagnoses) of opioid-naïve

patients between the years 2008 and 2011. Barnett et al. (2019) focus on VHA emergency department and urgent care clinic visits in 2012. We are not as restrictive regarding prior opioid use, excluding only the top 15th percentile (3,150 mg of morphine in the prior year). However, we are more restrictive regarding conditions, excluding diagnoses that are rarely prescribed (anything less than a 10% prescription rate). Our study years also do not align; we focus on 2006-2016. This affects the interpretation of the estimates. Their estimates are for "new" opioid users following their first opioid prescription, whereas our estimates are for one (additional) prescription for veterans who come to the ED for particular conditions.

Reconciling the Differences in Long-Term Use Estimates

In this section, we investigate how the differences in the studied samples, and in methods in measuring prescriber intensity affect the estimate on long-term prescription opioid use (the only shared outcome studied in both their papers and ours). We begin by replicating Barnett et al. (2019), then we make incremental changes to the sample construction, eventually ending up at the baseline sample studied in this paper. We do this all while keeping the high/low intensity classification based on a physician's opioid prescription rate within a facility, as in their papers. Then, we move to our residualization approach as described in subsection 3.2, also incrementally including more controls, finally arriving at the estimate reported in this paper. With each incremental step, we report the mean long-term prescription opioid use associated with high and low intensity physicians, the ratio between the two (odds ratio), and the Wald estimate (an analog to the 2SLS estimate but with a binary high vs. low "instrument" to aid in comparison and interpretation with Barnett, Olenski, and Jena (2017) and Barnett et al. (2019)).

Table G.15 reports the result of this exercise. The first three columns of row 1 are taken directly from Barnett et al. (2019); the Wald estimate[29] (column 4) of 0.903. column (2) is our best attempt at replicating their main finding. The odds ratio and Wald estimate are very similar; however, the base long-term use means are greater, presumably due to minor

[29]This wald estimate is called "number needed to harm" in Barnett et al. (2017). It is not reported in Barnett et al. (2019), but scaling their high vs. low long-term differences by their prescription rate, yields 0.903.

differences in data definitions. Next, we make incremental changes to the sample restrictions and data definitions to arrive at the baseline sample in this paper. High and low intensity physicians are classified by top and bottom quartile opioid prescribing rate, within a facility, after the corresponding sample restriction change. Some examples of such changes include: changing the definition of long-term opioid use to days supply of opioids filled[30] (row 3), excluding urgent care clinics (row 4), including admitted patients and some prior users (rows 7 and 8), excluding diagnosis conditions that are rarely prescribed (row 9), adding opioid prescriptions from Medicare and Medicaid (row 10), and including all years from 2006-2016 (row 11). Since these changes alter the relevant sample of veterans, they have varying effects on the Wald estimate. For example, including CMS opioid prescriptions increases the Wald estimate, implying that patients who see a more lenient ED physician, are also more likely to fill new opioid prescriptions through Medicare or Medicaid. With the within-facility intensity classification of Barnett, Olenski, and Jena (2017) and Barnett et al. (2019) on our baseline sample, we have a Wald estimate of 2.75 (column 4 of row 11), more than double the main effect reported in this paper. If we allow physician prescribing intensity to vary across years (i.e., top vs. bottom quartile within a facility-year; row 12), then the Wald estimate drops to 1.75, still 50% larger than our estimate of 1.17 with our residualization approach.

In the next four rows of Table G.15 (rows 13-16), with our baseline sample, we now classify physicians as high/low-intensity with our residualization approach, incrementally residualizing for additional covariates. The first level of residualization is at the hospital-year-month level. That is, we construct our physician leniency as described in subsection 3.2, but with only hospital-month fixed effects to control for hospital specific seasonality. We then select the top and bottom quartiles of prescribers per hospital based on their mean residuals. Finally, we compute the difference in (residualized) long-term use divided by (residualized) prescription rate—the Wald estimate—in column (4). By residualizing for hospital specific seasonality, the Wald estimator drops in magnitude substantially. This implies that much of the variation between physicians, even within a facility, is endogenous. The next three rows controls for

[30]If a patient has two on-going opioid prescriptions with overlapping days, Barnett et al. (2019) do not count the overlapping days towards the 180 days supply needed to be classified as a long-term user, whereas we would count it overlapping days, because those opioid pills are available to be abused. Therefore, their measure of long-term use is days of opioids consumed, while ours is days of opioids available.

"shift-level" variation in physician work schedule and patient arrival, diagnosis condition, and patient covariates including age, Elixhauser comorbidity index and pain score, finally arriving at a Wald estimate of 1.25. Recall that our baseline 2SLS estimate (with the continuous leniency instrument) was 1.17. This exercise implies that residualization in both the leniency construction and the second stage can yield different estimates.

Ranking Physicians by Prescribing Leniency Using Barnett et al. (2017, 2019) vs. Our Method

The comparison in the previous section teaches us that sample selection and physician leniency construction lead to differences in estimates of an ED prescription's effects on long-term use. Our long-term use probabilities are larger because they include some prior users and focus on diagnoses that are typically prescribed opioids. Moreover, even by keeping the sample fixed, the two empirical approaches used to construct prescribing leniency arrive at different estimates. The classification of lenient physicians hinges on patient diagnosis, age, risk, and time of arrival at the ED. Figure G.5 demonstrates this by graphing the reshuffling of prescribing ranking after controlling for said covariates for the Tampa VA Medical Center (the largest ED in 2012). Each physician (provided they have treated 30 cases) is sorted by his/her ranking after our residualization method on the x-axis. The y-axis represents their corresponding ranking using the Barnett et al. intensity measure. If both methods yield identical rankings, the physicians align perfectly on the dashed diagonal line. Next, we classify physicians as low and high intensity prescribers based on the top and bottom quartiles using either method. The blue squares correspond to physicians who are classified in the top or bottom quartile by both methods, and the red triangles correspond to physicians about whom the two methods disagree. The physicians at the tails of the distribution tend to be classified as top or bottom prescribers by both methods; however, there is substantial disagreement outside of the tails. There are 46 physicians whom both methods agree are either high or low intensity prescribers, and 34 who are classified by one method but not the other.

B. Identifying VHA Emergency Departments and Linking Opioid Prescriptions

In this section we describe in detail how we identify VHA emergency visits, linking opioid prescriptions to its originating emergency department (what counts as prescribed), and identifying primary care PACT visits.

Emergency Departments

Emergency departments in the VHA were standardized beginning in 2006 with VHA Directive 2006-051 "Standards for Nomenclature and Operations in VHA Facility Emergency Departments". Therefore, we start looking for ED visits in 2006.

Emergency department visits are identified off VA stop codes. We do not consider urgent care centers are emergency departments. After March 2007, we use visits with primary stopcode of 130. Prior to March 2006, we use i) primary-secondary stopcode combination 102-101 OR ii) primary stopcode of 102 with an emergency department CPT procedure code. In addition, we require the visit to originate in a station number (`DivisionSID` that is listed as an emergency department (excluding facilities that have joint emergency and urgent care) in the 2007 Survey of Emergency Departments and Urgent Care Clinics in the VHA. Lastly, we also require emergency departments to have at least 5000 annual visits and non-negligible visit share between 12-4am, following VHA Directive 2006-051, which required emergency departments to operate 24 hours a day, seven days a week.

Opioid Prescriptions

Opioid prescriptions need to be linked back to its origin (i.e., was it from an emergency department or primary care clinic?). We employ the following algorithm in coding an emergency department as prescribed an opioid:

1. We restrict attention to opioid prescriptions that are written (`IssueDate` within a day of the emergency encounter.

2. If there is a perfect provider-prescriber ID match, we code the emergency encounter as $Prescribed = 1$.

3. If the prescription was written on the same day, or on the next day (provided the emergency visit happened after 8pm) and the facility ID (`DivisionSID` match, we code the emergency encounter as $Prescribed = 1$.

4. All other emergency cases are coded as $Prescribed = 0$.

We do not require a perfect provider match because the a patient may see more than one clinician in the ED, and the (head) attending physician may not be the prescriber name on the prescription. Out of the cases we code as *Prescribed*, 88% of them have a prescriber and provider ID match, and the other 12% that match on facility ID and date/time, we code the prescription as *Prescribed* by the attending physician for the purpose of constructing leniency. Here we are assuming that the attending physician influences the decision to prescribe and has oversight what other providers (e.g., nurse practitioner) are doing. Note that if a patient is admitted and prescribed an opioid following their hospitalization, the patient will be considered prescribed provided the prescription was written within a day of the emergency visit, and the prescription will be assigned to the emergency physician.

C. Construction of Secondary Outcomes

Our secondary outcomes comprise additional measures of our main outcomes, as well as measures of illicit drug use. We provide details on variable construction for each outcome below.

Secondary measures of opioid use

To also capture opioid use not covered by our prescription data, we use positive opioid drug screens (urine or blood) within three years of the ED visit as a secondary measure of opioid use. For ease of interpretability, this variable is constructed unconditional on screening—patients who do not receive a screen receive a value of zero. It is thus subject to the limitation that ED assignment may lead to differential screening rates.

To capture the strength of prescription opioids used, we measure the total milligrams of morphine equivalent (MME) of all prescription opioids filled in the three years after the ED visit (excluding the ED opioid prescription).

Secondary measure of opioid seeking behavior

As secondary proxies for opioid seeking behavior, we consider each individual proxy that enters our primary measure, as well as a patient's self-reported pain score (on a 0-10 scale) averaged across all outpatient encounters. This score can be exaggerated by the patient to obtain opioid prescriptions. All measures are based on the first 12 months post-ED visit.

Secondary measures of opioid overdose events

We employ three secondary measures of overdose events.

As a measure predictive of opioid overdose risk, we use an indicator for accidental falls, which is a proxy for impulsivity or sedation (Oliva et al., 2017).

Our measure of non-fatal opioid overdose events is based on ICD-9 diagnosis codes. They are identified via codes 965.x, E850.0-E850.2, E935.0-E935.2, and E980.0.

To investigate the type of opioid involved in a veteran's overdose death, we turn to ICD-9 codes in cause-of-death files. We distinguish heroin, synthetic (excluding methadone; e.g., fentanyl), and natural and semi-synthetic opioids (e.g., morphine, codeine, oxycodone). The

corresponding codes are: heroin (T40.1), synthetic non-methadone opioids (T40.4), and natural and semi-synthetic opioids (T40.2 only).

All measures are based on data from the three years post-ED visit.

Illicit drug use

In addition to heroin and synthetic opioid overdose deaths, we investigate illicit drug use via self-reported survey responses, as well as two proxies obtained from medical records. Our measure based on self reported survey responses is described in the main manuscript, in subsection 4.4.

Our first proxy of illicit drug use based on medical records is an indicator for a physician's intent to screen for heroin/fentanyl. Unfortunately, it is difficult to distinguish heroin and fentanyl from prescription opioids in drug screens—often the same test is used for both—making it impossible to ascertain a physician's intent to screen specifically for heroin/fentanyl from the drug screen alone.[31] Therefore, we code any test that mentions heroin, fentanyl or 6-MAM (the specific metabolite unique to heroin and fentanyl) in the order form, regardless of test result, as an intent to screen for heroin/fentanyl; for veterans with no such tests ordered, the outcome is coded as zero.

Our second proxy for illicit drug use based on medical records is a hepatitis C (HCV) diagnosis. HCV is an infection that is commonly transmitted by sharing needles, and the opioid epidemic has contributed to the rise in HCV infections (Powell et al., 2019; Zibbell et al., 2018). The CDC (2016) identifies injection drug use as the main risk in over half of new HCV cases, and it is estimated that 32% of injection drug users are diagnosed with HCV within one year of injection and 53% within five years.[32]

For both proxies, we concentrate on the three years post-ED visit. It is important to note that we only observe the results of patients who are tested or diagnosed; patients who do not take the test or are not diagnosed are coded as zero.

Alcohol abuse

[31] The specific metabolite unique to heroin and fentanyl called 6-monoacetylmorphine (6-MAM) is detectable in urine for only up to eight hours after heroin use (Moeller et al., 2008) and physicians often do not know the distinction between 6-MAM and standard morphine screens (Starrels et al., 2012).

[32] Hagan et al. (2008); see Degenhardt et al. (2017) for overview.

Besides opioid use and use of illicit drugs, we also investigate alcohol abuse—the most common form of substance use disorder observed among veterans (Seal et al., 2011).

Our measure of alcohol abuse is based on AUDIT-C, an alcohol screening questionnaire widely used among primary and specialty care physicians in the VA. For 89% of veterans in our baseline sample we observe at least one AUDIT-C response in the three years following their ED encounter. Following the AUDIT-C manual, a respondent is coded as alcohol abuse positive if they score a 4 or higher on the questionnaire's 0-12 scale (Babor et al., 2001) within the first three years following their ED encounter. Veterans who did not complete an AUDIT-C questionnaire are coded as zero.

D. Robustness

The 2SLS and IV results presented in Section 4 are robust to key alternative specifications probing into potential violations of the identifying assumptions listed in Section 3.4. Table G.8 summarizes these findings. The underlying analyses are described in detail below.

Addressing threats to conditional independence

Previously, we showed balance of patient observables with respect to physician prescribing leniency; however, there might be selection along unobservable margins. Such selection could occur in two ways: "mechanically", via our use of physicians' choice of diagnosis code (which we use both to construct our sample and as a control in our instrument construction); or via non-random assignment of patients to physicians, even conditional on our detailed set of hospital and date/time fixed effects. We address these concerns below and find our results robust to both.

To address the concern related to endogenous diagnosing, we run a specification that no longer excludes any diagnosis from the analysis sample, and that replaces the original diagnosis control in both leniency construction and the 2SLS model with the most recent outpatient diagnosis code observed *prior* to the ED visit. Results are displayed in column (2) of Table G.8. The magnitude of our coefficients stays virtually unchanged, but we lose significance on OUD and opioid overdose mortality outcomes.

To address concerns related to non-random assignment, we leverage across-shift variation in the composition of physicians working on a particular shift (on average 2-3 physicians per shift). The idea behind this approach is that conditional on our detailed set of hospital-date/time fixed effects, the composition of physicians on shift is as good as random. Given the small number of physicians per shift, and substantial variation in leniency across physicians, this approach yields considerable team-level leniency differences across shifts. We define team-level leniency as the average leniency across physicians on shift, leaving out all cases of that shift when constructing physician leniency (see Appendix H for details). Reduced form results in placebo sample of patients coming to the ED with rarely prescribed health conditions lend credence to this strategy: using the team leniency measure, we find no association between leniency and outcomes in the placebo sample (column 2 of Table G.9).

Results from our re-estimated 2SLS model, replacing the physician prescribing instrument with our team-based alternative, are displayed in column (3) of Table G.8. Apart from the coefficient on OUD, which becomes noisier, our estimates remain largely unchanged. We conclude that patient-physician selection is unlikely to be driving our findings.

Addressing threats to the exclusion restriction

In Section 4.2, we performed placebo checks indicating that the reduced form results observed for opioid-related outcomes operate through an opioid prescription channel, as opposed to other channels. However, we also document a failure of our placebo check for preventable hospitalizations and all-cause mortality, highlighting a potential threat to interpreting our IV findings as identifying the causal effects of receiving prescription opioids: physicians may differ in many dimensions beyond just prescribing leniency. If these dimensions are correlated with leniency and affect our outcomes of interest, then the exclusion restriction is violated. While this concern is slightly alleviated due to the short-term nature of emergency physician and patient relationships, emergency physicians may still make non-opioid related decisions that can impact patient outcomes. In this section, we probe the robustness of our findings to adjusting for chief observable margins of care that may be correlated with leniency, namely: i) decision to admit patients to the hospital, ii) intensity of procedures performed, iii) quality of care provided, and iv) amount of opioids prescribed (i.e. intensive margin). To preview our results, we find that our 2SLS estimates are robust—magnitudes stay essentially unchanged, as does precision.

To address the first two margins, we model them as endogenous decisions as in Mueller-Smith (2015) and Bhuller et al. (2020). That is, we first construct instruments for admission and procedure propensities analogous to prescribing leniency in Equation 1. Hospital admission is a binary variable, while intensity of procedures is proxied for with work-Relative Value Units (w-RVU), which is the part of the CMS fee schedule that converts procedure codes to a payment amount.[33] We then include predicted admission and predicted total w-RVU as controls in the baseline 2SLS regression. Column (4) of Table G.8 reports the

[33]The CMS fee schedule converts procedure codes to payments based on time, technical skill, and effort required. One caveat is that the VHA does not pay physicians on a fee-for-service basis, hence there is an under-reporting of procedures. To the extent that all physicians consistently under-report, this would only be a level-change without biasing our intensity of procedure estimates.

2SLS estimates on the opioid prescription dummy. The estimated coefficients are virtually unchanged, suggesting that an ED physician's admission and intensity of procedure decisions do not affect the patient's long-term outcomes, but rather, opioid prescriptions do.

To address the third margin, we construct a measure of a physician's quality of care and include it as a control in our 2SLS model. Our measure of quality is a physician's average impact on patient immediate (one-month) mortality after visiting the ED—a proxy previously used to assess hospital quality in settings with quasi-random assignment of patients to hospitals (Hull, 2020). Immediate mortality is unlikely to be caused by the physician's prescribing decision, and thus provides a useful measure of physician quality in other dimensions of care. We estimate this physician quality proxy analogous to our prescribing leniency instrument and the two admission and procedure propensities above. This estimated physician quality proxy is included as a control in the baseline 2SLS regressions in column (5). The coefficients on the opioid prescription dummy are nearly identical and our main findings are robust.

Finally, we find that our results are robust to a potential violation of the exclusion restriction relating to the intensive margin decision of the *amount* of opioids to prescribe. We have modeled opioid prescriptions as a binary decision; however, physicians are also deciding on prescription length and dosage. Panel D of Figure G.6 plots the relationship between total MME and extensive margin leniency and finds a small, positive, non-monotonic relationship.[34] Nevertheless, we adopt the standard approach in accounting for the intensive margin in the judges design (Bhuller et al., 2020). We include an endogenous MME prescribed in Equation 4, construct an intensive margin propensity, and run a 2SLS regression with two endogenous variables and two IVs. Then we evaluate the average treatment effect conditional on being prescribed the average ED morphine equivalent dosage. We report this estimate in column (6) of Table G.8, which represents the average treatment effect of being prescribed an average ED prescription, controlling for both the intensive and extensive margin decisions in opioid prescribing. The estimates remain nearly identical, implying that intensive margin prescribing differences are not biasing our results.

[34]The average physician in the top decile of extensive margin prescribing leniency prescribes ca. 6 mg of morphine more than the bottom decile, conditional on being prescribed. The mean MME conditional on being prescribed is 153mg.

Addressing threats to monotonicity

The monotonicity assumption requires lenient physicians to be consistently lenient. We describe standard monotonicity checks following Dobbie et al. (2018) and Bhuller et al. (2020) in Section 3.4, display results in Table G.12, and find that our instrument passes both tests. As an additional robustness check, we allow physicians to have differential prescribing leniency measures across different major diagnosis categories (MDC) and construct a physician-year-MDC-specific instrument. Column (7) of Table G.8 reports 2SLS estimates with these mutually exclusive instruments. Our main estimates retain their sign and approximate magnitude; again, OUD diagnoses and overdose mortality coefficients become noisier.

E. Share of Opioid Overdose Deaths Due to ED Physicians

In this appendix we outline the back-of-the-envelope calculation attributing the universe of VHA veteran opioid overdose deaths between 2006 and 2016 to exposure to prescription opioids through a leniently prescribing ED physician. In the 11 year period, approximately 9,200 veterans died of an opioid overdose. Of these 9,200 veterans, a total of 3,077 visited a VA ED. The prescription rate in the ED is 12%, meaning 369 veterans were prescribed an opioid. Next, we make two assumptions. First, we assume that the local average treatment effect is equal to the average treatment effect. Second, we assume that the average treatment effect for our baseline sample is the same as the universe of ED samples. These two assumptions imply that we can multiply the number of veterans who were prescribed an opioid by the relative effect size of dying from an opioid overdose (0.075/0.167 from Table 5). This means that 165.8 of the 9,200 veterans (or 1.8%) experienced the event because of exposure to prescription opioids through a leniently prescribing ED physician.

F. Who Are Lenient Opioid-Prescribing Physicians and How Do They Vary Along Other Dimensions?

In this appendix we summarize the characteristics of lenient physicians based on their observables, then correlate prescribing leniency with other physician dimensions along four margins: i) decision to admit a patient to an inpatient hospital, ii) decision to perform invasive procedures, iii) likelihood of causing a patient death within one month (proxy for physician quality), and iv) intensive margin decision regarding amount of opioids to prescribe, conditional on prescribing an opioid (based on total milligrams of morphine equivalent).

Table G.16 presents characteristics of lenient and strict physicians-years. Recall that our leniency measure is defined at the year level. Physicians are classified as lenient (strict) if they are in the top (bottom) quartile of our leniency measure each year. Lenient physicians are much more likely to be male and slightly older in age and on average work more. Lenient physicians on average work nine extra days per year compared to strict physicians. They also see more patients per day, however, it is unclear whether this is due to working longer shifts or working quicker. This could be in line with findings that physicians prescribe more opioids when they are busier and more fatigued (i.e., later in the workday or when appointments are running behind schedule as seen in Neprash and Barnett, 2019).

Next, we investigate how physician opioid-prescribing leniency correlates with other physician dimensions. In particular, we study dimensions that may violate our exclusion restriction. We study the graphical first stage of our baseline physician prescribing leniency on four different dimensions: admission, intensity of procedures, physician quality, and the intensive margin opioid-prescribing decision (conditional on prescribing an opioid). All four proxies are discussed in greater detail in subsection 4.5. Figure G.6 displays these first stage correlations over the histogram of opioid-prescribing leniency values. Across all four dimensions, there is a positive relationship with opioid-prescribing leniency in the main mass of the histogram. The relationships are generally small and often non-monotonic at the tails. For instance, the average physician in the top decile of prescribing leniency performs on average 1.628 w-RVU compared to 1.557 in the bottom decile, a 4.6% increase in payment if paid for by CMS. In terms of one-month mortality, for physicians in the top decile of

opioid-prescribing leniency, 0.634% of their ED patients die within a month, compared to 0.614% in the bottom decile. These effects are modest and we've shown in subsection 4.5 that they do not have significant effects on our findings.

G. Calculating and Characterizing Compliers

In this section we describe the method we use to calculate the share of compliers (and always-takers and never-takers), and its characteristics. The method follows Dahl et al. (2014) and Dobbie et al. (2018).

First, compliers are defined as patients who would not have been prescribed an opioid if they had been seen by the most strict physician, but would have been prescribed an opioid if they had been seen by the most lenient physician:

$$\pi_{complier} = P(D_{\bar{z}i} > D_{\underline{z}i}) = E(D_{\bar{z}i} - D_{\underline{z}i}) = P(D_i|Z_i = \bar{z}) - P(D_i|Z_i = \underline{z})$$

where D_i represents the prescription decision for veteran i, Z_i represents the leniency of veteran i's physician, and $\bar{z}$ and $\underline{z}$ represent the most and least lenient physicians.

Similarly, always-takers are patients who would be prescribed an opioid by every physician:

$$\pi_{always-taker} = P(D_{\bar{z}i} = D_{\underline{z}i} = 1) = P(D_{\underline{z}i} = 1)$$

where the last step follows from the monotonicity assumption. Last, the share of never-takers (patients who would never be prescribed an opioid by any physician) is found by:

$$\pi_{never-taker} = P(D_{\bar{z}i} = 0)$$

By defining the most lenient physicians ($\bar{z}$) as physicians with a leniency instrument in the top percentile and the most strict physicians ($\underline{z}$) as physicians with a leniency instrument in the bottom percentile, we can calculate the share of compliers, always-takers, and never-takers from moments in the first stage. For instance, we fit a local linear regression of $Prescribed_i$ on physician leniency, take the share of veterans who are prescribed by the top percentile of leniency, and subtract the share of veterans who are prescribed by the bottom percentile of leniency.

We can also characterize our compliers by observable characteristics. For example, we can calculate the share of veterans who are prior users, conditional on being a complier. In particular, we can compute $P(X_i = x|complier)$:

$$
\begin{aligned}
P(X_i = x|complier) &= P(X_i = x|D_{\bar{z}i} > D_{\underline{z}i}) \\
&= \frac{P(X_i = x \cap D_{\bar{z}i} > D_{\underline{z}i})}{P(D_i|Z_i = \bar{z} - P(D_i|D_i = \underline{z}} \\
&= \frac{P(D_{\bar{z}i} > D_{\underline{z}i}|X_i = x)P(X_i = x)}{\pi_{complier}} \\
&= \frac{\pi_{c|x}P(X_i = x)}{\pi_c}
\end{aligned}
$$

This moment is calculated by computing the share of compliers for the subsample $X_i = x$ (i.e., checking the moments of the first stage for that subsample) and scaling it by the unconditional share of that subsample, divided by the overall share of compliers. This is the second column in Table G.13.

H. Team (Across-Shift) Leniency

As mentioned in subsection 3.4, the identification of our *within*-"shift" quasi-random assignment strategy breaks down if there is selection in patient-physician assignment along unobserved margins. Examples of such violations include such situations as senior physicians delegating difficult, frequent ED visitors who refuse to leave to newer physicians or physicians taking cases of severe conditions on which they are experts. In such cases, assignment to a physician, say A vs. B, is non-random at a given point in time, t. However, if only physicians A and B are working at that ED at time t, we can use the average leniency of that "team" by utilizing the fact that at some other time t', physicians A and B have been replaced with physicians C and D. Specifically, as an alternate robustness strategy, we no longer rely on random assignment to patients conditional on showing up at the ED (*within*-"shift"), but leverage variation in the timing of their visit and the available personnel working at that ED (*across*-"shift").

For individual i arriving at the emergency department at time t, we define $s = [t-1h, t+1h]$, a two hour "shift" window, and the leniency of their potential physician:

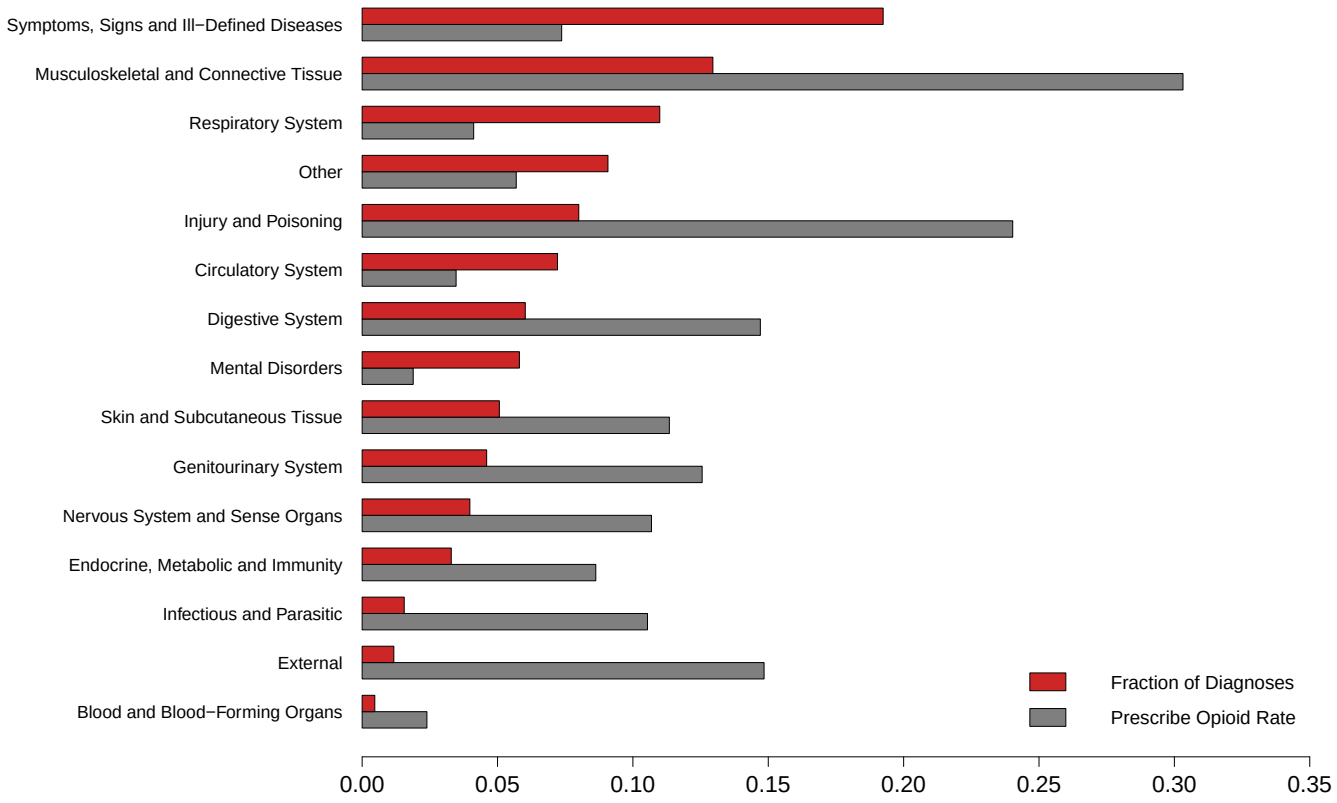
$$Leniency_s^{team} = \frac{1}{\sum_{j \in \mathbb{S}} N_{jy}} \left(\sum_{j \in \mathbb{S}} N_{jy} \times Leniency_{-s,jy}^{phys} \right) \tag{6}$$

where $Leniency_{-s,jy}^{phys}$ is as defined in Equation 2 except leaving out all patient cases occurring in shift s, $\mathbb{S}$ is the team of physicians j who are working at any point during shift s, and N_{jy} is the total number of cases seen in year y by physician j. This team-based leniency instrumental variable is a weighted average of the potential physicians a patient could have seen at the time they arrive in the ED. The weighting is based on the number of cases the physician sees that year to account for variance in our measure of individual physician leniency.

Figure G.7 graphs the histogram of the team leniency along with its first stage in comparison with the baseline physician leniency. As expected, the range of possible values shrinks, however, the first stage slope remains unchanged.

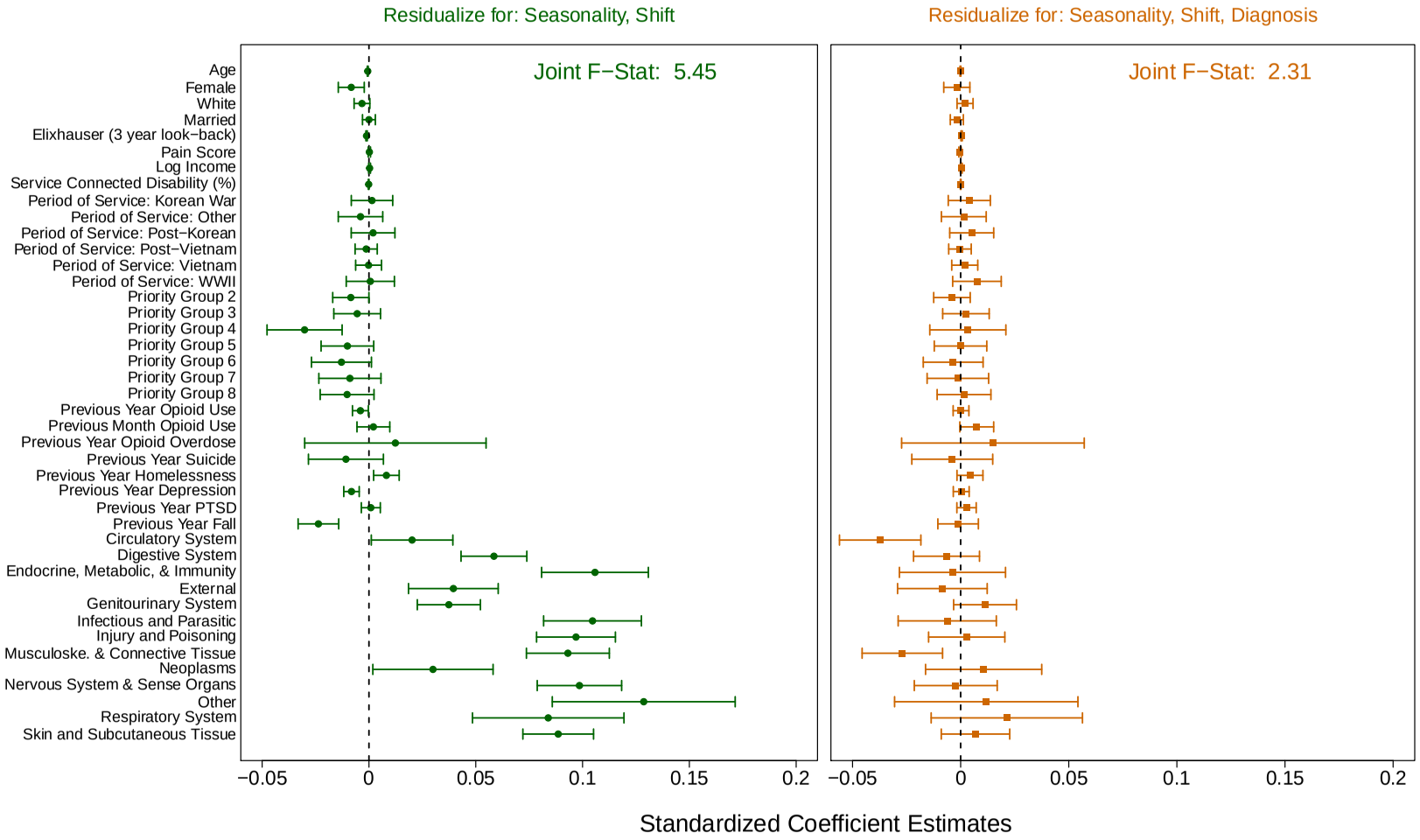
I. Additional Figures and Tables

Figure G.1: Frequent Diagnoses Occurring in Emergency Departments



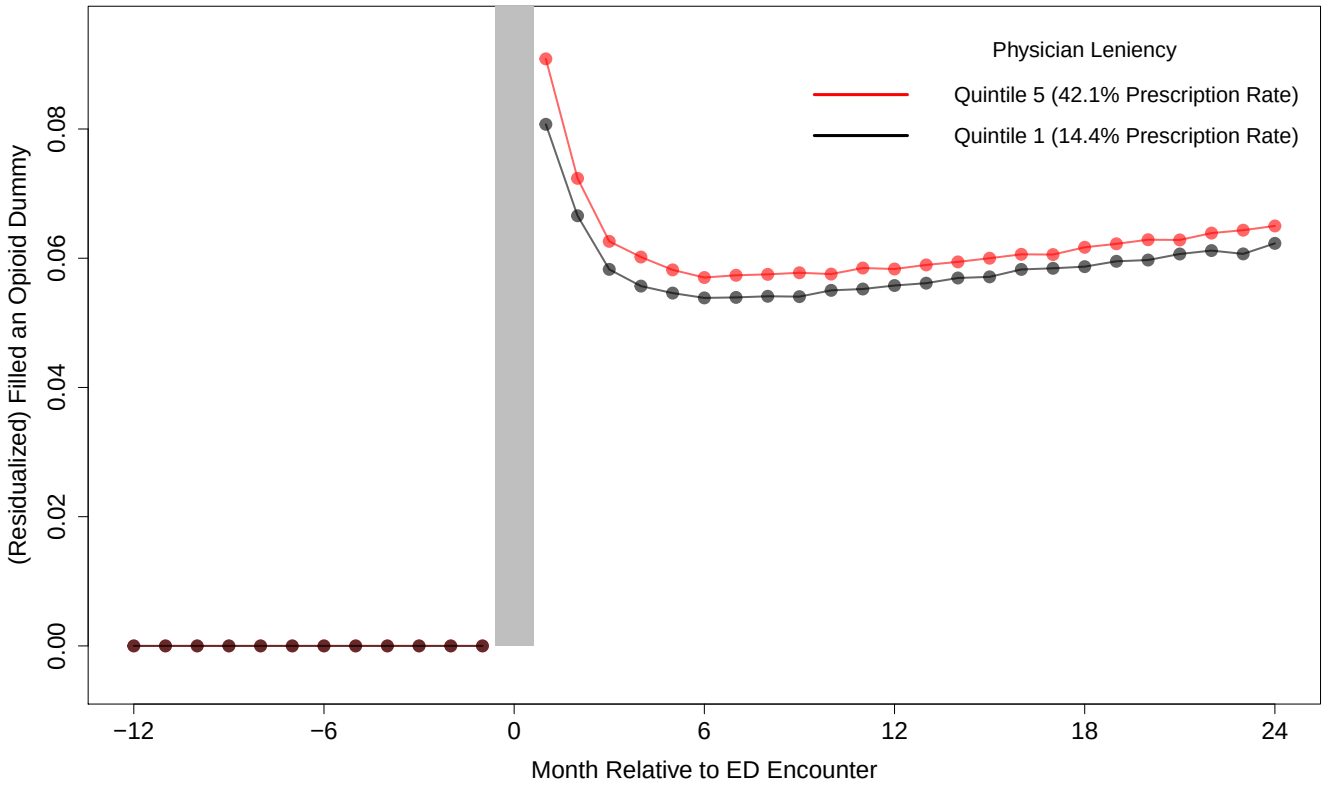
Notes: The 15 most common major diagnosis categories (ICD-9 major chapters) for *all* ED visits and the un-adjusted rate they are prescribed opioids

Figure G.2: Balance Test at Varying Levels of Residualization Controls



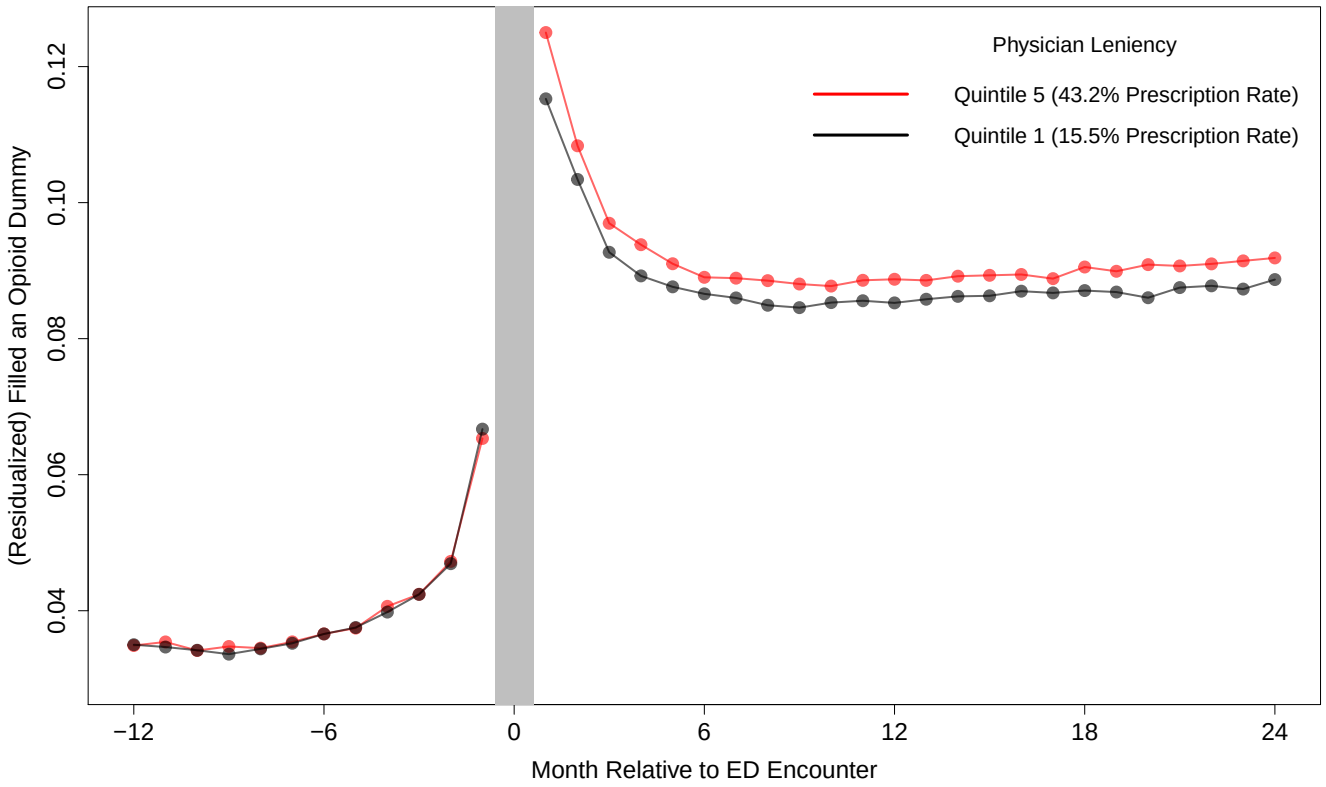
Notes: This figure displays the standard balance test (as seen previously) for the physician leniency measure constructed with different controls in the residualization of Equation 1, effectively, testing for different quasi-random assignment assumptions. In the left panel, the physician leniency instrument is constructed with only controls for seasonality and shift. The right panel, the diagnosis condition is included as an additional control in the residualization.

Figure G.3: Reduced Form: Subsequent Opioid use for Opioid-Naïve Patients



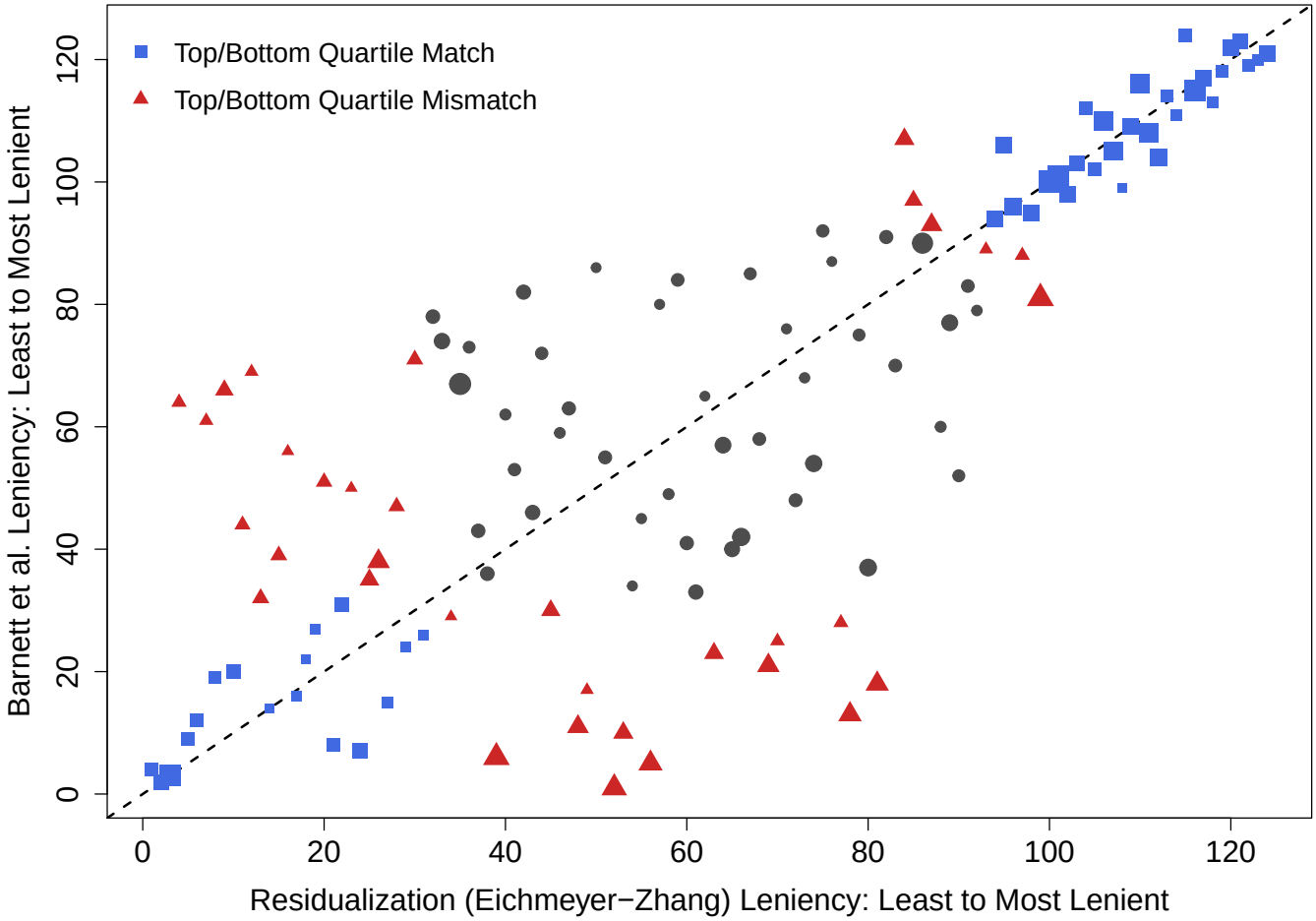
Notes: The reduced form event-study figure corresponding to Figure 3, but for opioid-naïve patients.

Figure G.4: Reduced Form: Subsequent Opioid use for Patients Without an ED Visit in the Prior Year



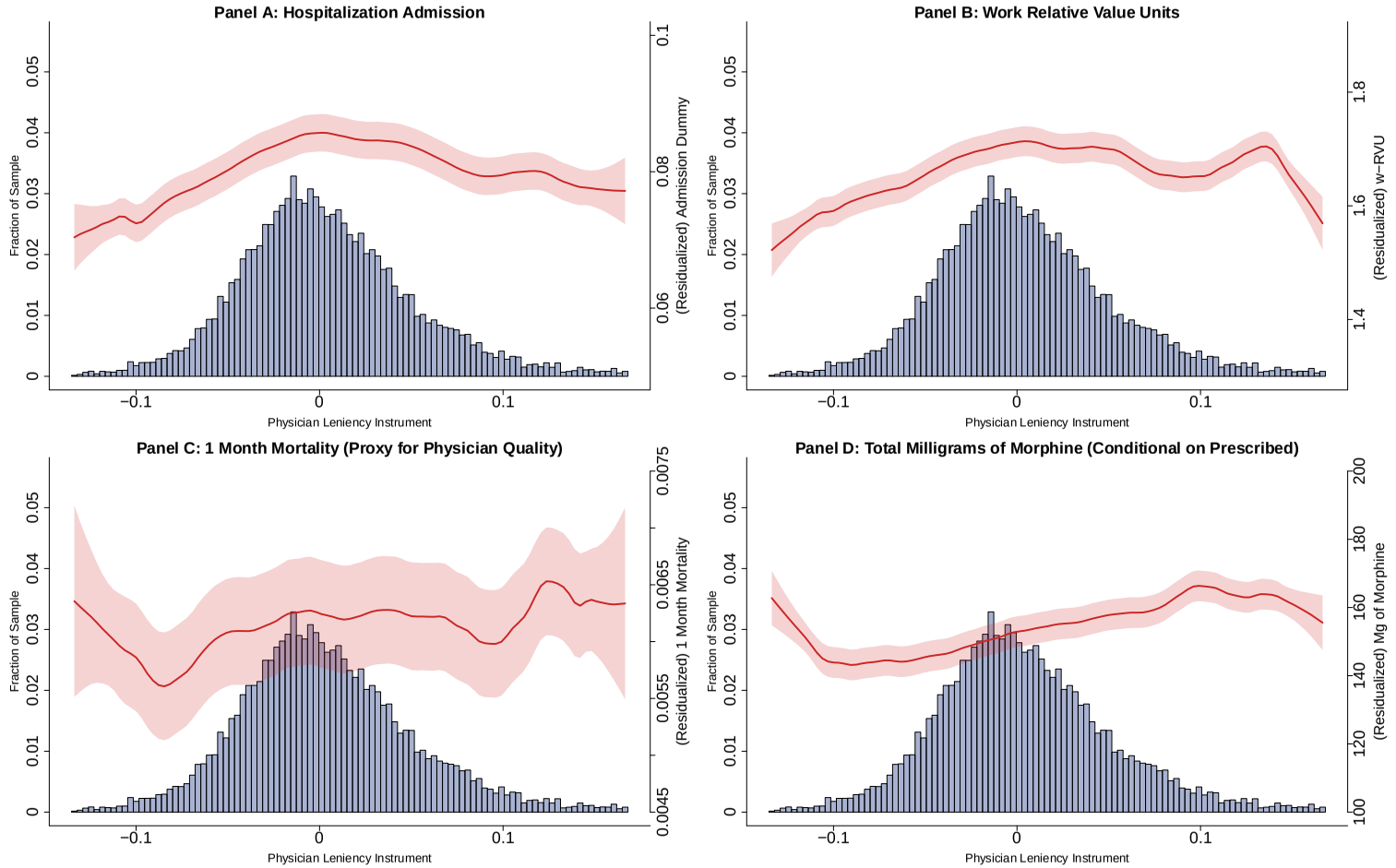
Notes: The reduced form event-study figure corresponding to Figure 3, but for patients who did not visit an ED in the prior year (for any condition), but did utilize VHA outpatient care. Presumably this group are not ED shopping for opioids.

Figure G.5: Ranking Physician Prescribing Leniency in Tampa Veteran Affairs Medical Center



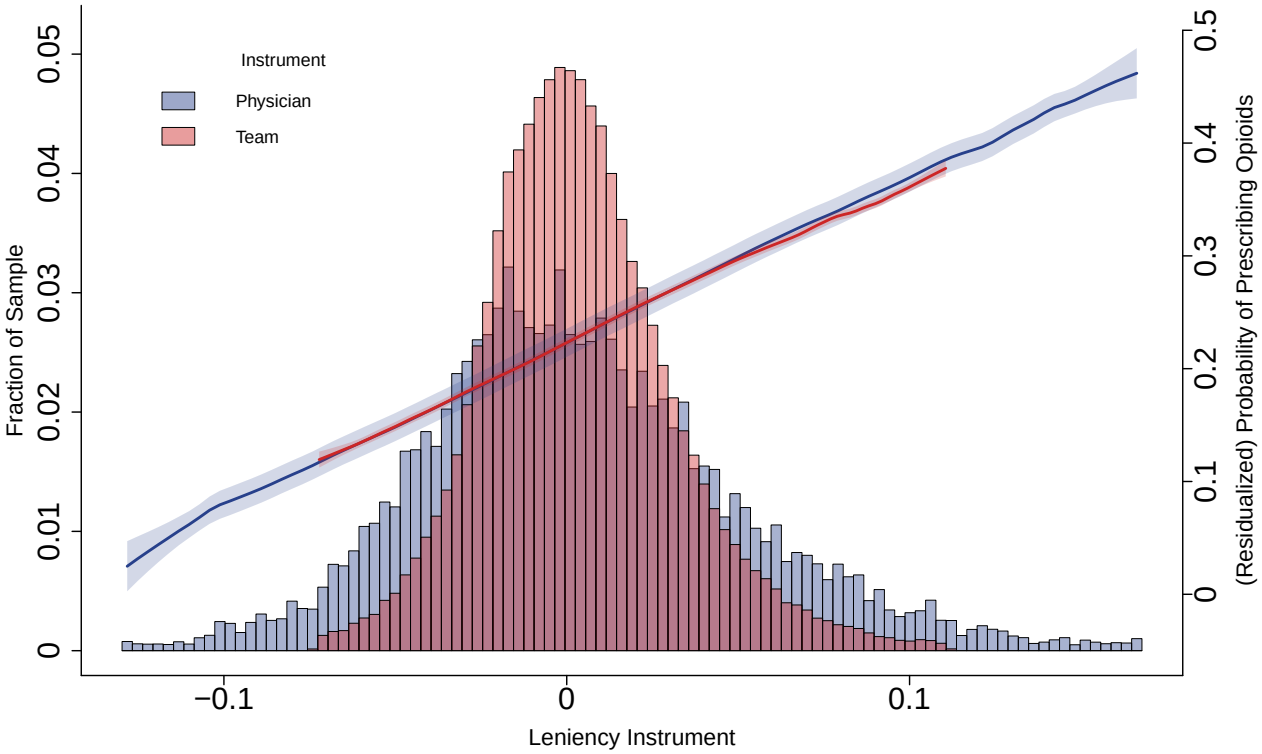
Notes: This graph shows the re-shuffling of physician ranking based on Barnett et al. method and our residualization method for physicians with at least 30 cases in Tampa Veteran Affairs Medical Center, the largest in the country by ED volume in 2012. Each point is a physician and the size of the point is proportional to the log number of cases seen. The blue boxes correspond to physicians that would be classified in the top or bottom quartile by both methods, and the red triangles correspond to physicians that the two methods disagree on.

Figure G.6: First Stage of Baseline Physician Opioid-Prescribing Leniency Instrument on Other Dimensions of Physician Characteristics



Notes: Each panel overlays a first stage local linear regression of a particular (residualized) physician dimension on a histogram of our baseline physician opioid-prescribing leniency. Panel A corresponds to the decision of admitting a patient, panel B is the total work relative value units as a proxy for intensity of procedures performed, panel C is physician quality, proxied by the one month mortality rate, and panel D is the intensive margin of total volume of milligrams of morphine equivalent, conditional on being prescribed an opioid. 95% confidence bands are also displayed in the shaded red region.

Figure G.7: Distribution and First Stage of Team Instrument



Notes: This figure plots the histogram of the alternate team leniency instrument (overlaid on top of the baseline physician leniency instrument) along the x-axis and the left y-axis. A local-linear regression of the fitted probability of prescribed opioids on the instrument after residualizing is overlayed and displayed on the right y-axis. 95% confidence bands are also shown.

Table G.1: NSAID prescriptions for patients who are not prescribed opioids

	Mean
P(Prescribed NSAID | No Opioid)	0.21
Days Supply	15.8
Quantity of Pills	40
Ibuprofen	0.42
Naproxen	0.27
Ketorolac Tromethamine	0.10
Etodolac	0.06
Indomethacin	0.05

Notes: This table reports basic summary statistics for prescriptions of nonsteroidal anti-inflammatory drug for patients who are not prescribed opioids.

Table G.2: Ordinary Least Squares Regression on Main Opioid-Related Outcomes

	Dependent variable: (×100)			
	Long-Term Use	Opioid-Seeking Behavior	Opioid Use Disorder	Opioid Overdose Mortality
	(1)	(2)	(3)	(4)
Prescribed in ED	2.629***	4.329***	0.013	0.044***
	(0.061)	(0.083)	(0.039)	(0.008)
Mean Dep. Var. (×100)	5.8	14.8	3.2	0.17
Residualization FEs?	Yes	Yes	Yes	Yes
Baseline Controls?	Yes	Yes	Yes	Yes
N=	1,879,150	1,879,150	1,775,800	1,846,133

Notes: This table reports the estimated coefficients of an ordinary least squares regression of our main opioid-related outcomes on *Prescribed*. Long-term use is defined as 180 days of opioid supply in the first year following the ED visit (excluding the first 7 days), opioid-seeking behavior in the first year is a composite proxy as described in the text. Opioid use disorder and opioid overdose mortality are defined as within three years. See text for residualization fixed effects and baseline controls. Mortality is calculated within three years of the ED visit. The samples are constrained such that the patients are alive for the entire period the outcome is measured except for mortality outcomes. Regression coefficients, standard errors, and mean dependent variables are scaled as indicated. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.3: 2SLS Estimates on Opioid-Related Outcomes Without Baseline Controls

	Dependent Variable: (×100)			
	Long-Term Use	Opioid-Seeking Behavior	Opioid Use Disorder	Opioid Overdose Mortality
	(1)	(2)	(3)	(4)
Prescribed in ED	1.32***	2.55***	0.323*	0.074**
	(0.209)	(0.332)	(0.161)	(0.034)
Mean Dep. Var. (×100)	5.80	14.79	3.27	0.167
Residualization FE?	Yes	Yes	Yes	Yes
Baseline Controls?	No	No	No	No
N=	1,879,150	1,879,150	1,775,800	1,846,133

Notes: This table reports the 2SLS effect of an opioid prescription on our main outcomes in Table 5 without baseline controls. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, and 3-digit diagnosis codes; the even numbered columns include baseline controls as described in the text. The samples are constrained such that the patients are alive for the entire period the outcome is measured except for mortality. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01.

Table G.4: 2SLS Results for Secondary Outcomes

	Dependent Variable:						
	Mg of Morphine	Positive Drug Screen	Overlapping Prescriptions	Pharmacy Shopping	Back Pain & Headaches	Opioid Overdose	Falls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prescribed in ED	467.7***	0.020***	1.90***	0.30***	0.55***	0.007	0.504***
	(59.5)	(0.253)	(0.20)	(0.07)	(0.27)	(0.012)	(0.257)
Mean Dep. Var	2,353	0.082	9.9	0.6	6.2	0.6	8.4
Residualization FE?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N=	1,775,800	1,775,800	1,840,595	1,840,595	1,532,610	1,775,800	1,775,800

Notes: This table reports of 2SLS effect of an opioid prescription on our secondary outcomes: total milligrams of morphine equivalents (column 1); indicator for an urine or blood drug screen that is positive for opioids in the first three years following the ED visit (column 2); first year overlapping prescriptions (column 3); first year pharmacing shopping (column 4); five or more encounters for back pain and headaches in the first year–excluding patients whose index ED encounter was for backpains or headaches (column 5). Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, and 3-digit diagnosis codes; see text for baseline controls. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.5: 2SLS Results for Substance Use and Abuse Outcomes, Questionnaires, and Proxies

Dependent variable (×100)	Coef (1)	Mean Dep. Var. (2)	N (3)
Any drug *abuse* (illicit or prescription)†	0.156* (0.088)	0.776	1,775,800
Opiate *use*†	0.066 (0.054)	0.274	1,775,800
Cocaine/Crack *use*†	0.143** (0.057)	0.317	1,775,800
Sedatives *use*† (e.g., benzodiazepines)	0.061 (0.042)	0.172	1,775,800
Other stimulant *use*† (e.g., amphetamines)	0.047 (0.040)	0.155	1,775,800
Marijuana *use*†	0.004 (0.068)	0.471	1,775,800
Positive Alcohol screen	0.214 (0.355)	20.5	1,775,800
Intended Heroin/Fentanyl Drug Screen	0.015 (0.072)	0.070	1,775,800
Hepatitis C Diagnosis	0.259 (0.209)	5.90	1,775,800
Pain Score	0.08*** (0.02)	2.70	1,682,968

†: in the past 30 days

Notes: This table reports the 2SLS effect of an opioid prescription on outcomes and proxies for substance abuse and opioid-seeking behavior. All outcomes are binary and veterans who are not screened are coded as zero. Drug abuse outcomes are from the Brief Addiction Monitor questions 6 and 7. The questions ask "In the past 30 days, how many days did you use..." and all non-zero answers are coded as positive. Alcohol screen is based on the AUDIT-C which identifies "hazardous drinkers or active alcohol use disorders"; scores of 4 or greater are coded as positive screens. All regressions include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins fixed effects and standard baseline controls. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Regression coefficients, standard errors, and mean dependent variables are scaled as indicated. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.6: Opioid Overdose Mortality by Type of Opioid

	Dependent Variable: (×100)			
	Heroin	Natural & semi-synthetic opioid only	Synthetic non-methadone	Heroin or synthetic non-methadone
	(1)	(2)	(3)	(4)
Prescribed in ED	0.022	0.034*	0.020	0.030
	(0.019)	(0.020)	(0.017)	(0.023)
Mean Dep. Var. (×100)	0.055	0.050	0.038	0.084
Residualization FE?	Yes	Yes	Yes	Yes
Baseline Controls?	Yes	Yes	Yes	Yes
N=	1,846,133	1,846,133	1,846,133	1,846,133

Notes: This table reports the 2SLS effect of an opioid prescription on opioid overdose mortality by type of opioid. ICD-10 mortality codes: heroin (T40.1; column 1); natural and semi-synthetic opioids only (T40.2 and no other opioid type indicated; column 2); synthetic non-methadone opioids (T40.4; column 3); and heroin or synthetic non-methadone opioids (T40.1 or T40.4, column 4). Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, and 3-digit diagnosis codes; see text for baseline controls. Regression coefficients, standard errors, and mean dependent variables are scaled as indicated. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.7: Reduced Form Regressions on Rarely-Prescribed Samples for Non-Opioid Related Outcomes

	Sample Based on Diagnoses' Prescription Rates		
	Baseline: Sample	Prescription Rate $\in [0.03, 0.1)$	Prescription Rate < 0.03
Dependent Variable ($\times$100):	(1)	(2)	(3)
Homelessness	0.047	−0.091	0.064
	(0.060)	(0.077)	(0.095)
Mean Dep. Var. ($\times$100)	11.9	12.1	16.6
Suicide	−0.006	0.006	0.015
	(0.022)	(0.023)	(0.040)
Mean Dep. Var. ($\times$100)	1.6	1.4	2.4
All-Cause Mortality	0.098*	0.187***	0.137
	(0.050)	(0.065)	(0.085)
Mean Dep. Var. ($\times$100)	9.3	12.7	16.1
Preventable Hospitalizations	0.040	0.165**	0.273**
	(0.059)	(0.076)	(0.112)
Mean Dep. Var. ($\times$100)	8.7	10.7	15.1
Residualization FEs?	Yes	Yes	Yes
Baseline Controls?	Yes	Yes	Yes
Observations	1,958,209	1,897,297	1,449,315

Notes: This table reports the estimated coefficients of a reduced form regression of secondary non-opioid outcomes on physician prescribing leniency. Column 1 estimates the regression on our baseline sample of conditions that are prescribed at least 10% of the time, and columns 2 and 3 are for conditions that are prescribed 3-10% of the time and conditions that are <3% of the time. All coefficients are scaled by the difference in leniency between the 90th and 10th lenient physicians (11.6pp) for interpretability. See text for residualization fixed effects and baseline controls. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.8: Alternate Specifications and Robustness Checks

	2SLS Estimates of *Prescribed* on Main Outcomes						
	Main Baseline	All Diagnoses	Team Leniency	Admit & Procedures	Physician Quality	Intensive Margin	MDC IV
Dep. Var. (×100):	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Long-Term Use	1.172***	1.193***	1.313***	1.165***	1.104***	1.092***	1.921***
	(0.202)	(0.293)	(0.305)	(0.202)	(0.203)	(0.201)	(0.258)
Opioid-Seeking Behavior	2.456***	2.615***	2.035***	2.405***	2.368***	2.410***	2.977***
	(0.314)	(0.503)	(0.463)	(0.312)	(0.317)	(0.314)	(0.383)
Opioid Use Disorder	0.335*	0.418	0.375	0.308*	0.344**	0.321**	0.232
	(0.160)	(0.267)	(0.237)	(0.160)	(0.161)	(0.176)	(0.182)
Opioid Overdose Mortality	0.075**	0.050	0.104*	0.076**	0.077**	0.074***	0.066
	(0.034)	(0.050)	(0.055)	(0.034)	(0.034)	(0.035)	(0.042)
N=	1,775,800	2,547,150	1,775,800	1,775,800	1,739,337	1,775,800	982,679

Notes: This table reports 2SLS regression coefficients of *Prescribed* on the main outcomes for the main baseline empirical model in column 1 and various alternate specifications in columns 2-7. Column 2 takes first visits for the entire sample of diagnosis codes and controls for most recent outpatient diagnosis *prior* to the ED visit (instead of diagnosis recorded *during* the ED visit), both in the leniency construction step and in the 2SLS regression. Column 3 uses team leniency as the instrumental variable. Column 4 includes predicted propensity to admit (hospitalize) and intensity of procedure (measured with w-RVUs) as controls to the baseline via an indirect least squares. Column 5 constructs a proxy for physician quality analogous to propensity to prescribe, but replacing opioid prescription indicator with 1 month mortality indicator. Column 6 includes an intensive margin (total milligrams of morphine) endogenous variable and instrument, and evaluates the average treatment effect at the mean ED opioid prescription (152mg of morphine). Column 7 uses physician-diagnosis-year leniency instruments (physician-diagnosis-year bins with fewer than 20 cases are dropped). See subsection 4.5 for more details. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.9: Alternate Specifications For Placebo Sample

	2SLS Estimates of *Prescribed* on Main Outcomes			
	Main Baseline	Team Leniency	Admit & Procedures	Physician Quality
Dep. Var. (×100):	(1)	(2)	(3)	(4)
Homelessness	0.064	0.217	0.051	0.031
	(0.095)	(0.150)	(0.093)	(0.096)
Suicide	0.015	−0.034	0.008	0.007
	(0.040)	(0.065)	(0.039)	(0.040)
All-Cause Mortality	0.137	0.050	0.115	0.046
	(0.085)	(0.122)	(0.081)	(0.081)
Preventable Hospitalizations	0.273**	0.033	0.242**	0.192*
	(0.112)	(0.144)	(0.106)	(0.111)
Residualization FEs?	Yes	Yes	Yes	Yes
Baseline Controls?	Yes	Yes	Yes	Yes
Observations	1,449,315	1,449,315	1,449,315	1,449,315

Notes: This table illustrates our various robustness strategies to address violation of identifying assumptions for secondary non-opioid outcomes. It reports the estimated coefficients of a reduced form regression of secondary non-opioid outcomes on physician prescribing leniency for diagnosis conditions that are prescribed <3% of the time. Column 1 reports the baseline estimate from column 3 of Table G.7, column 2 reports the estimates using a team leniency instrument scaled by the difference in leniency between the 90th and 10th lenient physicians (11.6pp), columns 3 and 4 control for "non-focal" propensities in admission and procedures, and one-month immediate mortality. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.10: Main Outcomes with Physician Leniency IV Constructed at Varying Levels of Residualization

	IV Residualization Level:		
	Seasonality & Shift	+ Diagnosis	+ Additional Ctrl. (baseline)
Dependent variable:	(1)	(2)	(3)
Long-Term Use	1.113***	1.171***	1.172***
	(0.215)	(0.201)	(0.202)
Opioid Use Disorder	−0.133	0.300*	0.335*
	(0.186)	(0.175)	(0.160)
Opioid Overdose Death	0.068*	0.074**	0.075**
	(0.037)	(0.034)	(0.034)

Notes: This table reports the 2SLS causal effect of an opioid prescription on the main outcomes with three different physician leniency instruments. The three instruments are constructed with varying levels of controls in the residualization in Equation 1: hospital-year-month and hospital-day of week-time of day (Column 1), hospital-year-month and hospital-day of week-time of day, and diagnosis (Column 2), and the above including Elixhauser Comorbidity Index, pain score, five-year age bins, and number of prior ED visits (i.e., the baseline IV; Column 3). All three regressions include the standard controls described in the text. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.11: 2SLS Regression of Main Outcomes on Intensive Margin Quintile Thresholds

	Dependent variable: (×100)			
	Long-Term Use	Opioid-Seeking Behavior	Opioid Use Disorder	Opioid Overdose Mortality
	(1)	(2)	(3)	(4)
MME Quintile 1	0.902*	0.798	1.006**	0.130
	(0.507)	(0.771)	(0.423)	(0.098)
MME Quintile 2	0.783**	2.364***	0.262	0.146**
	(0.350)	(0.581)	(0.287)	(0.069)
MME Quintile 3	1.233***	2.471***	−0.129	0.004
	(0.390)	(0.611)	(0.302)	(0.065)
MME Quintile 4	0.958**	2.075***	0.246	0.033
	(0.455)	(0.652)	(0.294)	(0.069)
MME Quintile 5	1.736***	3.172***	0.512**	0.096
	(0.350)	(0.533)	(0.257)	(0.062)
Test for joint equality (p value):	0.196	0.073	0.166	0.457

Notes: This table reports the estimated coefficients of a 2SLS regression of our main opioid outcomes on being prescribed an opioid that is at least above a certain MME threshold, where each indicator is instrumented by its analogous instrument. The MME thresholds are 60, 100, 150, and 200 milligrams of morphine equivalents corresponding to five quintiles, conditional on being prescribed an opioid; the excluded category is not being prescribed any opioids. Five instrumental variables are constructed, one for each endogenous intensive margin variable. We also report a Wald test on the joint equality of all five coefficients. Residualization fixed effects and baseline controls are also included. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.12: Testing the Monotonicity Assumption

	Dependent variable: Prescribed Opioid	
	Baseline Leniency	Reverse-Sample Leniency
Sub-sample margin:	(1)	(2)
Male	1.696***	1.046***
	(0.012)	(0.014)
Female	1.758***	1.917***
	(0.027)	(0.035)
Black	1.836***	1.951***
	(0.02)	(0.028)
White	1.649***	1.378***
	(0.013)	(0.016)
Opioid-Naïve	1.745***	1.524***
	(0.015)	(0.024)
Prior Users	1.584***	1.628***
	(0.016)	(0.024)
No Depression or PTSD	1.688***	1.608***
	(0.014)	(0.022)
Depression or PTSD	1.74***	1.839***
	(0.015)	(0.019)
Priority Groups 1-4	1.738***	1.851***
	(0.014)	(0.019)
Priority Groups 5-8	1.695***	1.733***
	(0.014)	(0.018)
Injury and Poisoning	1.714***	1.905***
	(0.028)	(0.039)
Musculoskeletal & Connective Tissue	2.267***	2.845***
	(0.023)	(0.043)
Digestive System	1.683***	1.807***
	(0.035)	(0.039)
Circulatory System	0.711***	0.746***
	(0.088)	(0.095)

Notes: Column 1 displays the first stage coefficient of prescribed opioid on the baseline physician leniency instrument for the corresponding sub-sample. Column 2 constructs a new physician leniency instrument using *all* emergency visits, excluding the corresponding sub-sample ("reverse-sample"), and displays the coefficient of the first stage regression back on that sub-sample. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table G.13: Characterization of Compliers

	$P(X = x)$	$P(X = x\|\text{complier})$	$\frac{P(X=x\|\text{complier})}{P(X=x)}$
White	0.690	0.657	0.952
Black	0.239	0.238	0.999
Age < 40	0.201	0.197	0.977
Age $\in [40, 60)$	0.363	0.379	1.046
Age ≥ 60	0.436	0.377	0.865
Opioid-Naïve	0.752	0.766	1.018
Prior Opioid User	0.248	0.222	0.896
Depression or PTSD	0.294	0.296	1.005
No Depression, No PTSD	0.706	0.696	0.986
Musculoskeletal and Connective Tissue	0.311	0.364	1.170
Injury and Poisoning	0.202	0.195	0.966
Digestive System	0.070	0.057	0.819
Other Major Diagnosis Categories	0.417	0.321	0.770
Above Avg Risk for Opioid Overdose Death	0.376	0.400	1.064
Below Avg Risk for Opioid Overdose Death	0.376	0.338	0.898

Notes: This table reports for each demographic subgroup: its unconditional share, its conditional probability given they are a complier, and the relative likelihood.

Table G.14: Average Life Years Lost Associated with Lenient Prescribers

	Dependent variable:	
	All-Cause Life Years	Opioid Overdose Specific Life Years
	(1)	(2)
Physician Leniency	0.0249*	0.0078
	(0.0131)	(0.0073)
Residualization FEs?	Yes	Yes
Baseline Controls?	Yes	Yes
Observations	1,954,608	1,954,608

Notes: This table reports the estimated coefficients of a reduced form regression of average life years lost on physician leniency. Life years lost is calculated by subtracting each veteran's life expectancy (onditional on being alive at each age in five-year increments; taken from VA, 2017) by their age at death. Life years lost is imputed as zero for veterans still alive in our sample and veterans who do not die from an opioid overdose for the regression in column 2. Estimated coefficients are scaled by the difference in prescribing rates between 90th and 10th percentile leniency for interpretability. Standard errors are clustered at the physician-level. *p<0.1; **p<0.05; ***p<0.01

Table G.15: Long-Term Use Estimate: Incrementally Moving From Barnett et al. (2019) to Eichmeyer and Zhang (2021)

Outcome: Long-Term Prescription Opioid Use	High Intensity (1)	Low Intensity (2)	High/Low Ratio (3)	Wald Estimate (4)
1. **Barnett et al. (2019)**	1.39	1.26	1.10	0.903
2. **Replicating Barnett et al. (2019)**	1.96	1.79	1.10	0.987
Incremental changes to sample restriction and data definition:				
3. +Extend long-term use defn. to opioid avail.	2.59	2.33	1.11	1.46
4. +Exclude urgent care clinics	2.53	2.30	1.10	1.39
5. +No prior enrollment/encounter restriction	2.70	2.38	1.14	2.01
6. +No post-ED cancer restriction	2.74	2.44	1.12	1.84
7. +Include admitted patients	2.97	2.68	1.11	2.00
8. +Include prior users	5.36	5.17	1.04	1.32
9. +Exclude rarely prescribed conditions	6.73	6.37	1.06	1.36
10. +Add CMS prescriptions	7.63	7.17	1.06	1.97
11. +Include all years (2006-2016)	6.05	5.36	1.13	2.75
12. **Year-varying physician intensity**	6.01	5.60	1.07	1.75
Incremental controls in leniency residualization:				
13. +Hospital-Year-Month (seasonality)	5.87	5.67	1.04	1.08
14. +Hospital-DayOfWeek-TimeOfDay (shift)	5.90	5.71	1.03	1.10
15. +Diagnosis	5.90	5.70	1.04	1.20
16. +Age, Elixhauser, pain score	5.90	5.69	1.04	1.25

Notes: This table begins with the estimate on long-term opioid use obtained in Barnett et al. (2019), and incrementally alters the sample and empirical approach (i.e., residualization in leniency construction) to arrive at the main estimates in this paper. Column 1 reports the mean long-term use associated with physicians in the top quartile of intensity (defined based on that specific sample restriction and/or leniency construction). Column 2 reports the same mean long-term use for physicians in the bottom quartile. The ratio of the two (odds ratio) is reported in column 3. The fourth column is a Wald estimate—mirroring the 2SLS estimate—for veterans treated by the top and bottom quartile physicians; BOJ 2017 calls this "number needed to harm". Row 1 reports the estimates found in Barnett et al. (2019) and row 2 is our best attempt at replication. Rows 3-11 *incrementally* alter the sample selection and data definitions, moving from Barnett et al. (2019) to our baseline sample in this paper. High/low intensity is defined with respect to the particular sample restriction, across all years. Row 12 classifies physicians-year as high/low within a facility-year for our baseline sample (row 11). Rows 13-16 employ our residualization approach described in Equation 1, *incrementally* including additional controls. In these four rows, both the outcome variable (long-term use) and endogenous variable (prescribed) are residualized with the baseline controls described in the text.

Table G.16: Average Characteristics of Physicians in the Top and Bottom Quartile of Leniency

	Lenient	Strict
Male	0.717	0.612
Age	47.4	46.1
Cases per year	929	789
Days worked per year	114	105
Patients per day	8.25	7.68

Notes: This table displays the simple mean of each variable for physician-years classified as lenient or strict. Lenient and strict are based on the top and bottom quartile of our leniency instrument measure each year. Only physician-years that treat at least 200 patients per year are included.