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

Sarah Eichmeyer and Jonathan Zhang*

August 2021

Abstract

This paper uses variation in practice behavior across physicians to uncover the role of medical care in causing opioid addiction. Using electronic health records of nearly two million US veterans with emergency department (ED) visits, we leverage quasi-random variation in patient assignment to physicians who vary in their propensity to prescribe opioids. We find that assignment to a provider in the top (vs. the bottom) decile of opioid prescribing significantly increases opioid use and misuse rates in the subsequent three years. Instrumental variable results show that being treated by an emergency physician who prescribes an opioid leads to a 1.2 percentage point increase in the probability of long-term prescription opioid use, a 0.34 percentage point increase in development of an opioid use disorder within three years, and a 0.075 percentage point increase in opioid overdose mortality. We find suggestive evidence of transition into heroin and fentanyl use after exposure to prescription opioids being the primary driver of increased overdose mortality.

Keywords: opioids, prescription drugs, physician variation, emergency department, patient outcomes, illicit drugs
JEL Classification Codes: I12, I18, H12, K42

*Eichmeyer: Economics Department, Ludwig Maximilian University of Munich. sarah.eichmeyer@econ.lmu.de; Zhang: McMaster University, and Office of Mental Health and Suicide Prevention, Veterans Health Administration, jonathanzhang@mcmaster.ca.

The views expressed herein are those of the authors and do not necessarily reflect those of the United States Department of Veterans Affairs or the Veterans Health Administration. A version was previous circulated as "Can a Single Opioid Prescription Make a Difference? Evidence from Physician Prescribing Variation in Emergency Departments". This paper has benefited tremendously from the guidance of David Deming, Seema Jayachandran, and two anonymous referees. We are extremely grateful to our advisors Liran Einav, Matthew Gentzkow, and David C. Chan, and our VHA supervisor Jodie Trafton for their guidance and support on this project. We would like to thank Abby Alpert, Michael L. Barnett, Luca Braghieri, Tim Bresnahan, Kate Bundorf, Janet Currie, Rebecca Diamond, Mark Duggan, David Green, Atul Gupta, Caroline Hoxby, Anupam Jena, Petra Persson, Ciaran Phibbs, Joseph S. Shapiro, Molly Schnell, Lucas Stich, Todd Wagner, Heidi Williams, and seminar participants at LMU Munich, McMaster, Ryerson, Simon Fraser, University of Alberta, Stanford, VA, and Wharton, and the VHA Ci2i, HERC, and PERC teams for their comments and suggestions. This project is supported by VHA Enterprise Opiate Safety Team's Comprehensive Addiction and Recovery Act funds and the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship via funds given to the Stanford Institute for Economic Policy Research.

1. Introduction

Medical prescription of opioids involves a clear trade-off between the benefits of immediate pain relief and the costs of longer-term dependence, including addiction and other adverse health impacts. The right balance is unknown and designing optimal opioid prescribing policies is difficult largely because the few randomized controlled trials on the impact of prescription opioids are relatively small-scale and do not follow patients long enough to observe a broad range of health outcomes (Busse et al., 2018).

In this paper, we use medical and pharmaceutical records from two million US veterans to quantify the cost side of opioid prescribing, focusing on its consequences for developing opioid dependence and addiction. Our empirical strategy exploits quasi-random assignment of patients to emergency department (ED) physicians who differ in their propensity to prescribe opioids. When patients arrive at the ED, they are assigned to a physician based on availability, with almost no discretion on either side. Thus patients who arrive at the ED at similar times and with similar symptoms are effectively being randomly assigned to physicians who vary in their willingness to prescribe opioids. We measure physician leniency to prescribe opioids using a leave-out, residualized measure based on all other patients the physician has seen in the ED in the same calendar year. The leniency measure strongly predicts the ED prescription outcome, but is uncorrelated with patient and ED visit characteristics.

We start by evaluating the reduced form effects of ED provider opioid prescribing leniency on downstream prescription opioid use and adverse opioid-related outcomes. We find that practice variation as captured by physician opioid prescribing leniency has large and significant consequences. Being treated by a provider in the top decile of the leniency distribution, compared to being treated by someone in the bottom decile, increases the probability to become a long-term opioid user by 0.24pp (or 4%). It increases the probability to exhibit opioid-seeking behavior, diagnoses of opioid use disorder, and opioid overdose mortality within three years of the ED visit significantly, by 3.4%, 2% and 8.8%, respectively.

Physician care takes on multiple dimensions besides the decision to prescribe opioids (e.g. intensity of medical procedures administered, decision to admit patient to the hospital,

quality of care provided). Employing a placebo exercise, we find evidence that the increase in a patient's downstream opioid-related outcomes after having seen a high prescribing provider is due to exposure to prescription opioids from this provider, as opposed to other differences across providers in care that correlate with prescribing leniency: we find no effects of physician opioid prescribing leniency on patient downstream opioid-related outcomes for a "placebo" sample of patients who visit the ED for conditions that are rarely prescribed an opioid.

With the caveat that other unobserved dimensions of physician care may impact patient outcomes, we employ physician leniency as an instrumental variable for receiving an opioid prescription in the ED to quantify the effect of such a prescription directly.[1] We find that exposure to an opioid prescription increases the probability of long-term use by 1.2 percentage points, on a base of 5.8%. The effect persists for over 24 months and never declines back to pre-ED opioid use levels. It is accompanied by a 0.34pp increase in the probability to develop opioid use disorder, as well as an 2.5pp increase in the probability to exhibit opioid seeking behavior, and thus consistent with increased misuse and dependence rather than appropriate medical care. Most disturbingly, we estimate that exposure to prescription opioids in the ED increases opioid overdose mortality within three years of the ED visit by 0.075pp.

We find suggestive evidence, albeit not statistically significant, that opioid exposure via a prescription from the ED can trigger veterans to begin using illicit injection drugs (such as heroin and fentanyl), as measured by self-reports, proxies from medical records, and overdose mortality from illicit opioids.

Taken together, our results suggest that gains in pain management from opioid treatment relative to non-opioid treatment would have to be very large to outweigh the increased risk of addiction and overdose death that arises from prescription opioid exposure among the patients in our sample. However, we find no evidence for such gains in pain management: our IV estimates show that prescription opioid exposure leads to worse pain according to veterans' self-reported pain score (a "fifth vital sign" in the VA) in the year post-ED visit,

[1] Leniency instruments have been used in other contexts. See, for example, (Doyle, 2007, entry into foster care) ; (Kling, 2006; Bhuller et al., 2020, incarceration); (Dobbie et al., 2017, bankruptcy cases); (Dobbie et al., 2018, pretrial detention); (Duggan, 2005, psychiatrists and antipsychotic drugs); (Doyle et al., 2015, ambulance companies); and (Farre-Mensa et al., 2020, patent examiners), among others.

despite increased opioid use, which should be providing pain relief.

Our findings are robust to analyses probing the key assumptions of random assignment, exclusion, and monotonicity. These checks include a shift- as opposed to physician-level leniency design, augmenting the baseline model to either control for physician behavior in other dimensions or include an instrument for other care decisions, inclusion of controls for physician quality, and monotonicity checks adopted from the judge stringency literature.

This paper relates to a large and growing literature on the health and economic consequences of opioid use.[2] The most closely related paper is Barnett, Olenski, and Jena (2017)—henceforth BOJ—who find that Medicare-insured patients treated by "high-intensity" prescribers are 30% more likely to become long-term users of legal prescription opioids compared to patients treated by "low-intensity" prescribers in the ED. In concurrent work, Barnett et al. (2019) replicate their 2017 study in the VA and find similar but attenuated differences in long-term use. Our paper differs from these two papers in two main ways. First, we study a broader range of relevant outcomes beyond long-term use, including detailed measures of opioid-seeking behavior, illicit drug use, and overdose mortality. Second, we control for the patient arrival and provider assignment process in EDs in both the construction of the instrument and in the main econometric model.[3]

Other closely related papers studying the extent and consequences of supply-side variation in opioid availability to patients are Laird and Nielsen (2016) and Finkelstein et al. (2021). Both focus on patients who move and exploit variation in opioid prescribing rates across origin and destination primary care providers and regions, respectively.[4] In ongoing work (Eichmeyer and Zhang, 2020), we study practice variation across primary care physicians in VHA facilities, and document large effects on patient opioid use and opioid dependence. These papers highlight the importance of broad supply-side factors for developing opioid dependence, but leave the exact mechanism unclear.

[2]See Maclean et al. (2020) for a thorough review of the economics literature.

[3]Appendix A provides a detailed discussion of how this paper differs from BOJ and where our findings depart from theirs.

[4]Both papers find large positive impacts of a high-prescribing environment (either via one's primary care physician or one's region) on opioid use and misuse.

Because of the institutional features of the ED, our research design closely approximates an RCT that dispenses a single prescription. In the ED, patients have no discretion over choosing providers, and physicians' discretion over choosing patients is typically limited, alleviating major selection issues present in other healthcare settings. Furthermore, physicians exhibit wide variation in practice behavior in prescribing opioids, even within the same hospital, while following the same guidelines, thereby providing ample variation in the supply of prescription opioids. Finally, patient-physician interactions in the ED are typically well-documented, short and one-off, constraining physician decision-making to a more limited, better observed choice set than present in settings such as specialty or primary care.[5] In sum, exploiting practice variation in ED settings shuts down other (but not all) potential channels besides opioid prescribing that are present in other settings, determine prescription opioid exposure, and impact patient outcomes. This approach allows us to move closer to identifying the causal impact of medically prescribed opioids on patient outcomes.

It is important to note that this paper studies the impact of an opioid prescription through a prescribing decision requiring clinical judgment—*within practice norms*—rather than through specific VHA policies, differences in adherence to clinical practice guidelines, or substandard care. Moreover, our period of study spans eleven years (2006-2016), and the national narrative regarding the opioid epidemic, clinical guideline recommendations, and VHA practice has changed dramatically over this time (Sandbrink et al., 2020). Opioid prescribing and care delivered at the VA, and studied here, were within clinical guidelines during this period.

The remainder of this paper is structured as follows. The next section describes the data source and outlines our baseline sample. The empirical strategy and its accompanying identifying assumptions are laid out in Section 3. Section 4 presents the results. Section 5 draws implications for opioid prescribing policies and performs welfare calculations. The last section concludes.

[5]Focusing on ED settings provides more than just identification advantages; EDs are a critical part of the health care system in the United States. In 2016, 19.4% of Americans and 12.7% of veterans, visited an ED (CDC, 2018; Huang et al., 2018), often for conditions involving pain; 14.3% of patients seen at EDs in 2011 were prescribed opioids (CDC, 2018), making them non-negligible sources of prescription opioids.

2. Data and Definitions

2.1 Data Source

We analyze four data sources from the Veterans Health Administration (VHA): i) VHA electronic health records, ii) Medicare and Medicaid claims iii) community care provided at non-VHA facilities, and iv) date and cause of death records.

The VHA provides health benefits to roughly 9 million veterans. Generally, veterans who served on active military duty for at least 24 consecutive months are eligible, as well as veterans who served for any length of time prior to 1980. Once eligible, veterans are enrolled in priority groups based on honorable decorations, disability and exposure to adverse war events, and income thresholds. These priority groups only determine copayment rates as there is no annual premium. Eligible veterans can be treated at any of the 152 VA medical centers and approximately 1,400 community-based clinics across the country.

Our primary data source is the VHA Corporate Data Warehouse (Veterans Health Administration, 2019), which includes electronic health records (EHR), enrollment tables, and other health records. The EHR data encompasses standard inpatient, outpatient, and pharmacy data, including patient and facility identifiers, diagnosis and procedure codes, and time of visit. Unlike claims data, EHR data also include referrals, patient questionnaires, clinical notes, lab results, etc. Because of the vast programs covered by the VHA, we also have data on non-medical clinics, including residential treatment programs for substance use disorders. ED records begin in 2006.

Since VHA health records only cover care that occurs at VHA facilities, we supplement it with data from other sources to obtain a more complete picture of medical care received by each veteran. This is important because 80% of VHA benefit enrollees have additional coverage through public or private insurance: 51% have Medicare, 7% have Medicaid, and 28% have private coverage (Huang et al., 2018). We retrieve Medicare claims from 2011-2016 and Medicaid claims records from 2011-2014 for all veterans. We observe medical claims for both Medicare (Part A and B) and Medicaid, along with prescription claims for patients

enrolled in any Medicare Part D plan, and Medicaid (VHA/CMS). In addition, we include all community medical and pharmaceutical claims that the VHA reimburses when it cannot provide the services themselves (e.g. nursing homes, inpatient hospice care, ED visits).

Our last data sources provide us with date and cause of death for all veterans from the CDC National Death Index (NDI) Plus files (Veterans Affairs, 2018). This includes detailed information about cause of death, allowing us to differentiate between opioid drug overdose deaths from deaths by heart disease, among others, along with death by specific type of opioid (e.g., natural and semi-synthetic versus heroin and synthetic opioids). We observe all deaths regardless of place of death/treatment or cause of death.

2.2 Sample Construction

Our research design focuses on adult veterans who visit a VHA ED. We observe approximately 20 million such visits until the end of 2019.

To improve power, we drop encounters with broad diagnosis categories that are prescribed an opioid less than 10% of the time. These encounters are common, and since opioid prescriptions are rare for these cases, our physician leniency instrument could suffer from a weak instrument problem were we to include them. Example diagnoses dropped include heart attacks and mental health episodes. Excluding these conditions does not introduce selection bias only if physician opioid prescribing tendency is orthogonal to physician diagnosing behavior. While this assumption may be violated if we were to use a very detailed level of diagnosis information upon which to base our exclusion criterion, it is plausibly satisfied when using broad diagnosis categories.[6] Therefore, we truncate ICD-9 codes at the three-digit level to account for potential endogenous diagnosing. This selection criterion cuts our sample to approximately 7.1 million visits. In subsection 4.5 we present our main outcomes without excluding such conditions.

Furthermore, we restrict the sample to patients who are not already very heavy prescription

[6]The idea being that, for a patient who presents with an ankle fracture, different physicians may choose different detailed diagnosis codes within the 3-digit ICD-9 824 ("Fracture of ankle"), but are unlikely to disagree at the 3-digit ICD-9 code level.

opioid users prior to the ED visit. The restriction is made because we are interested in understanding what causes *new* cases of long-term opioid use and dependence, and because an additional opioid prescription is unlikely to have an effect on the heaviest prior users. Specifically, visits by patients for whom we observe opioid prescriptions totaling to over 3,150 milligrams of morphine equivalent (MME)—a standardized unit of measurement to represent the potency of various opioids—in the prior year are excluded (this is equivalent to 5 weeks of 90 daily MME, approximately the 84th percentile of prior year opioid usage). After removing visits with non-prescribed diagnoses and high prior opioid users, we are left with approximately 5.9 million visits.

It is a consensus in the medical community that opioids are generally appropriate for end-of-life patients, since pain levels may be high, while the development of opioid dependence is of secondary concern given limited life expectancy. Therefore, as our third sample selection criterion, we exclude ED visits of patients who have terminal cancer or are on end-of-life hospice care at the time of visit.

In order to estimate a precise measure of physician-level opioid prescription tendency, we further restrict our sample to the 5.3 million encounters involving physicians who treat over 200 ED cases per year. Finally, we retain the first ED visit per veteran that satisfies the above criteria. Our baseline sample consists of 1,958,209 emergency visits (and veterans) treated by 5,313 physicians from 2006-2016.

2.3 Variable Definitions

Our explanatory variable in the IV analysis, $Prescribed_i$, is an indicator for whether patient i is prescribed *any* prescription opioid at their ED encounter (see Appendix B for details on how prescriptions are located and assigned). While the patient could decide not to fill the prescription, this is rare in practice, where we observe that approximately 97% of all VHA opioid prescriptions are filled.

Below we detail our main outcomes: a) opioid use, b) opioid seeking behavior, and c)

opioid use disorder and overdose mortality.[7] The first two measures are measured at the first year, relative to the date of the ED visit, since they are realized more immediately, whereas outcomes in group c) are measured within three years, since they are more downstream. Secondary outcomes, including measures of illicit drug use, are detailed in Appendix C.

Opioid use

Our main measure of long-term prescription opioid use follows the standard definition from the medical literature. It is an indicator for at least 180 days supply of opioids filled in the first 12 months after the ED encounter, excluding the first seven days and the initial ED prescription.[8] Since ED prescriptions are non-refillable, this means a long-term user needs to fill 180 days supply of opioids from other physicians. Secondary outcomes that also capture opioid use not covered in prescription data (in the form of positive opioid drug screens), as well as total strength of prescriptions used post-ED visit, are detailed in Appendix C.

Opioid-seeking behavior

To distinguish between medically appropriate long-term use versus inappropriate misuse, we construct three proxies for opioid-seeking behavior: i) a new prescription is filled when more than 25% of a previous prescription remain ("overlapping prescriptions"), ii) prescriptions are filled at three or more pharmacies over a 90-day period ("pharmacy shopping"), and iii) five or more encounter days with back problems, headaches/migraines in one year ("repeated back pain and headaches"). The first two proxies are commonly used in the literature and are strong predictors of overdose (e.g., Yang et al., 2015; Finkelstein et al., 2021). The last proxy is based on conversations with VA physicians, who identified this proxy as possibly indicative of opioid seeking behavior. We summarize the three proxies into a single indicator that equals one if any of the three behaviors is observed for a patient. This composite measure serves as our primary measure of opioid seeking behavior.

[7] Outcomes relying on pharmaceutical records or diagnosis codes are constructed using all data sources i)-iii) listed in subsection 2.1; remaining outcomes are constructed based on VHA records only (with the exception of cause of death, which also relies on CDC records).

[8] See Barnett et al. (2017, 2019); Jena et al. (2016); Dunn et al. (2010) for medical papers using this definition.

Opioid use disorder and overdose mortality

Our measures of severe adverse consequences of opioid use are diagnoses for opioid use disorder (OUD), and mortality from opioid overdose.

OUD is a chronic, life-long disorder broadly defined as a "problematic pattern of opioid use leading to clinically significant impairment or distress", according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V). We identify it via its ICD diagnoses codes.[9] The measure is likely strongly under-estimating true prevalence, since many cases may go undiagnosed by clinicians. For example, only 49% of all veterans for whom we observe an opioid overdose death had a recorded diagnosis for OUD in the three years leading up to their death.

Disjoint from OUDs, opioid overdose events occur when individuals take high doses of opioids that lead to slowing or stopping of breathing and potentially unconsciousness, coma, and death. The vast majority are non-fatal with timely treatment. To avoid issues of under-counting of non-fatal overdose events (because many get treated outside of the hospital and are not captured in our data—e.g., naloxone rescue administered outside the VHA), we use opioid overdose death as our main overdose measure. Different from OUD and non-fatal opioid overdose events, we can observe virtually all opioid overdose deaths in our data, regardless of place of death or type of overdose.[10] We also observe the *type* of opioid that caused the death, and use it as a secondary outcome (see Appendix C).

2.4 Summary Statistics

Table 1 describes relevant statistics of the opioid prescription, ED encounter, veteran, and their medical morbidities for our baseline sample. Approximately one in four (26.1%) veterans in our baseline sample receive an opioid prescription at their ED encounter.[11] The average

[9] The ICD-9 codes for OUD are: 304.0x, 305.5, and 304.7x.

[10] Opioid overdose deaths are constructed using ICD-10 mortality codes in which the entity axis is one of X40-44, X60-64, X85, or Y10-Y14, and at least one of the record axis conditions: T40.0, T40.1, T40.2, T40.3, T40.4, T40.6. We use *diagnoses* for opioid overdoses as a secondary measure (see Appendix C).

[11] Across all ED visits, including those diagnoses that are rarely prescribed opioids and thus excluded from our sample, 12% are prescribed opioids, comparable to the 14.3% national average in 2011 (NCHS, 2016).

ED prescription is for 8.2 days with 21 MME per day.[12] The ED is a common point of contact for patients with pain. Accordingly, musculoskeletal and connective tissue, and injury and poisoning are the two most common major diagnosis categories in our sample, combining to 40% of all ED visits. Frequent diagnoses in these two categories include back or knee pain, fractures, sprains/strains, and dislocations. Figure 1 shows the top 15 most common diagnosis categories and the share that are prescribed opioids in the ED for the baseline sample; Figure G.1 summarizes the major diagnosis categories across all ED visits.

The average veteran is a middle-aged, white male (average age of 55; 90% are male, and 69% are white), and thus belongs to the demographic group that has been hit the hardest by the opioid epidemic (Case and Deaton, 2015; Scholl et al., 2019). Many veterans suffer from mental illness, homelessness, and drug use, with 24% and 13% diagnosed with depression and PTSD, and 6.2% experiencing homelessness at some point in the prior year; 27% used an opioid in the previous year and 10% in the previous month. Rates of depression and PTSD among these veterans are much higher than in the US population as a whole,[13] and prior-year opioid use is double that of the average ED patient in a privately insured Optum Data Mart sample. While one should be cautious about generalizing results from the veteran sample to the overall population, studies have shown that mental health history and prior opioid use are strong predictors of opioid abuse (Boscarino et al., 2010; Edlund et al., 2014; Seal et al., 2012)–suggesting that veterans are more representative of the population that is at-risk of dependence after a brief exposure to opioids.

3. Empirical Strategy

Our empirical strategy closely follows the literature that relies on quasi-random assignment of agents to cases, often referred to as the "judges design." Papers in this literature typically

[12]Patients who are not being prescribed opioids are often prescribed nonsteroidal anti-inflammatory drugs (NSAID). In Table G.1 displays summary statistics of the average NSAID for patients who are not prescribed an opioid.

[13]Based on the NSDUH 2017 and the 2005 National Comorbidity Survey, 7% and 3% of adult Americans, respectively, were diagnosed with depression and PTSD in any year.

exploit variation in the sentencing leniency of judges who work in the same court. Similarly, we explore prescribing variation across physicians who work in the same emergency department. In its reduced form, under the assumption of quasi-random assignment, this approach allows researchers to identify the causal effect of being assigned to different types of physicians. Under additional assumptions, an instrumental variable approach identifies the causal effect of a given medical decision. We employ both approaches, and lay out their details in the next subsections.

3.1 Institutional Details on Patient-Physician Assignment

The ED is a setting where patients have very little choice regarding which provider they see, alleviating much of the selection issues in traditional healthcare settings. Leading up to ED visits, patients cannot look up reviews and schedule an appointment with their preferred physician. Instead, when patients arrive in EDs—sometimes due to an unexpected, sudden health shock—they are typically assigned to an emergency physician based on a triage system: First, a triage nurse assigns the patient an severity index based on health factors including life-threatening status, number of resources required, and vital signs. This index determines the patient's position in queue. Next, when physicians are available, they treat the first patient in the queue and the next after that. Physicians may specialize in certain diagnosis conditions. To summarize, conditional on showing up to the *same ED* at the *same time* with the *same diagnosis*, physician assignment to patients is as good as random.

3.2 Leniency Construction

Our physician leniency measure is a physician's year-varying, residualized leave-out average opioid prescription rate. It is obtained in two steps, following the approach of Doyle et al. (2015) and Dobbie et al. (2018). First, obtain the residuals from the following regression equation, estimated including *all* ED encounters in our sample period, including the ED

visits we dropped in subsection 2.2:

$$Prescribed_{ikt} = \alpha_0 + \alpha_{hym} + \alpha_{hdt} + \alpha_{diagnosis} + \alpha_{agebins} + \gamma W_{ik} + \epsilon_{ikt}, \tag{1}$$

where $Prescribed_{ikt}$ is a dummy that equals one if patient i was prescribed an opioid at their kth ED encounter, which took place on date t. Fixed effects include hospital-year-month fixed effects, α_{hym}, to control for time and seasonal variation in opioid prescribing such as hospital-specific policies (e.g., initiatives to limit prescribing) or hospital-specific seasonality in ED visits. We also control for "shift-level" variations that include both physician scheduling and patient arrival with hospital-day of week-time of day fixed effects, α_{hdt}.[14] Diagnosis fixed effects, $\alpha_{diagnosis}$, are included to account for physician specialization with respect to diagnosis conditions and to increase precision.[15] As mentioned, these three set of controls are what is required for our quasi-random assignment assumption. To improve statistical precision in our leniency measure, we include controls for five-year age bins, $\alpha_{agebins}$, and W_{ik}: Elixhauser comorbidity index, pain score in the ED, and number of prior visits. Under the assumption that we have captured the observables under which quasi-random assignment occurs in the ED, the unexplained variation—the physician's contribution—resides in the error term, ϵ_{ikt}.

In step two, the leniency measure for patient i seen by physician j in year y, is computed as the average residual across all other patients seen by the physician that year:

$$Leniency^{phys}_{i,jy} = \frac{1}{N_{-i,jy}} \sum_{k} \sum_{i' \in \{\mathbb{J} \setminus i\}} \hat{e}_{i'k} \tag{2}$$

where $\hat{e}_{i'k} = \hat{Prescribed}_{i'k} - Prescribed_{i'k}$ is the residual from Equation 1, $\mathbb{J}$ is the set of all ED encounters treated by physician j in year y, and $N_{-i,jy} = |\{\mathbb{J} \setminus i\}|$, the number of cases that physician has seen that year, excluding patient i. This leave-out mean eliminates

[14] Day of week takes on seven values: Sunday, Monday, etc. and time of day are six mutually exclusive four-hour bins: 8am-12pm, 12pm-4pm, etc.

[15] We truncate each ICD-9 diagnosis code to its three-digit value; all ICD-10 codes are crosswalked to ICD-9 (NBER, 2014b,a). We provide a robustness check that replaces the ED diagnosis control with a control for most recent diagnosis prior to the ED visit, in column (2) of Table G.8.

the mechanical bias that stems from patient i's own case entering into the instrument. The measure is interpreted as the average (leave-out) prescription rate of patient i's physician, relative to other physicians in that hospital-year-month, hospital-day of week-time of day, controlling for patient age and diagnosis.

Appendix F summarizes which physicians are more likely to be lenient prescribers according to our measure, and correlates prescribing leniency with other dimensions along which physicians can vary. In subsection 4.5, we examine how these physician characteristics affect our findings and interpretations. As a preview, we attempt to control for and address these physician differences and find our results are robust.

We document that VA ED physicians exhibit wide, systematic variation in their propensity to prescribe opioids. Figure 2 graphs the histogram of the leniency measure along the x-axis and the left y-axis. A local-linear regression of the fitted probability of prescribed opioids on leniency after residualizing is overlaid and displayed on the right y-axis. Table 2 presents the same "first stage" in a regression table: being assigned to a 10pp more lenient physician is associated with a 17pp increase in the likelihood of being prescribed an opioid in the ED. The F-statistic is 25 when all controls and fixed effects are included. The coefficient is greater than 1 because *all* emergency visits are used to construct the leniency instrument, while the first stage is calculated using the baseline sample only, which excludes the rarely prescribed diagnoses.

3.3 Empirical Specification

3.3.1 Reduced Form Approach

To estimate the reduced form effects of being treated by a lenient physician, we estimate the following equation:

$$Y_i = \mu_0 + \mu_1 Leniency_{i,jy}^{phys} + \gamma X_i + \eta_i, \tag{3}$$

where Y_i is patient i's outcome of interest post-ED visit, and where individual-level baseline controls, X_i, include the Elixhauser comorbidity index (three-year look-back excluding the

ED visit), measures of prior opioid use (dummy for prior month and log 1+ total MME for prior year), pain score, prior ED visit count, gender, race, and the set of fixed effects from Equation 1.

3.3.2 Instrumental Variable Approach

To study the effects of being prescribed an opioid in the ED on a outcome Y_i, we estimate the following 2SLS equations using our baseline sample:

$$Y_i = \beta_0 + \beta_1 Prescribed_i + \theta X_i + \epsilon_i \tag{4}$$

$$Prescribed_i = \delta_0 + \delta_1 Leniency^{phys}_{i,jy} + \delta_2 X_i + \nu_i, \tag{5}$$

where X_i is the same as in the reduced form approach. $Prescribed_i$ variable suffers from potential endogeneity concerns. For example, injury severity may be unobserved and correlated with opioid prescription, which in turn also affects long-term prescription opioid use. Hence, we instrument $Prescribed_i$ with the assigned physician j's underlying propensity to prescribe opioids, $Leniency^{phys}_{i,jy}$. We cluster robust standard errors at the physician-level to account for the assignment process of patients to physicians.

3.4 Identifying Assumptions

The reduced form approach delivers an unbiased estimate of the causal effect of being treated by a more leniently prescribing physician, if assignment of patients to ED physicians is random, conditional on seasonality, shift, and diagnosis ("conditional independence"). One can test for whether lenient physicians are routinely assigned patients with particular characteristics. Figure 4 shows evidence of balance on observables. The left panel is the output of a regression of *Prescribed* on observables including patient demographics, priority groups, income, previous medical diagnoses, and ED diagnoses. Unsurprisingly, many observables predict opioid receipt status. The right panel displays the coefficients from a regression of physician leniency instrument on the same set of observables. With so many coefficients, we

fail to reject that the instrument violates conditional independence (the joint F-stat drops by two orders of magnitude, from 280 to 2.4). In fact, some of the variables that are significant in the second column go in the opposite direction of the decision to prescribe.

The residualization in Equation 1 controls for more controls than required to achieve quasi-random assignment; they are included for statistical precision in measuring physician leniency. Figure G.2 displays the outcome of the balance test residualizing for only seasonality and shift (left panel) and also diagnosis (right panel) in Equation 1. This lends support to our assumption that assignment in the ED is nearly random at the seasonality, shift, and diagnosis level.[16] As an alternate design that no longer relies on within-shift random assignment, we leverage variation in the team of physicians working in the ED when a patient arrives in subsection 4.5.

Our instrumental variable approach, which aims to recover the causal effect of being prescribed *an opioid*, relies on three additional assumptions: relevance, exclusion, and monotonicity. We report a strong first stage (i.e. relevance) at the end of subsection 3.2. The exclusion restriction requires that the instrument must influence the outcome of interest only through its effect on initial prescription opioid use. This is perhaps our strongest assumption and is at its core, untestable. However, several features of the ED setting suggest that such violation may likely only have a small impact and may be less concerning than in other healthcare settings. First, unlike in primary care settings, where the patient and primary care provider have many repeat encounters, the scope of what the emergency physician can do to impact medium-term outcomes is limited and well-observed by the researcher. Second, any violation of the exclusion restriction needs to directly affect the specific outcome of interest. The channel by which ED physicians can influence opioid-related outcomes is likely through opioid prescribing. Nevertheless, we take this assumption seriously and perform a placebo check in subsection 4.2 as well as various robustness checks in subsection 4.5.

Finally, the monotonicity assumption is necessary for interpreting the coefficient estimates

[16]The level of residualization does have some effect on our baseline 2SLS causal estimates, as shown in Table G.10. The main estimates are smaller in magnitude when the leniency construction only residualizes for seasonality and shift; however, when diagnosis is included in the residualizing, the 2SLS estimates are very similar to the baseline, while standard errors are slightly larger.

obtained from the IV approach as Local Average Treatment Effects (LATEs) if there are heterogeneous treatment effects. It requires that any patient who is (not) prescribed opioids by a strict (lenient) physician, would also (not) be prescribed opioids by a less strict (lenient) physician. The literature leveraging the judges design typically perform two informal tests for its implications. The first one provides that the first stage should be weakly positive for all sub-samples (Dobbie et al., 2018). The second implication asserts that the instrument constructed by leaving out a particular sub-sample has predictive power over that same left-out sub-sample (Bhuller et al., 2020).[17] Table G.12 presents both of these tests in the two columns for various sub-samples of interest. Both columns exhibit significant and positive first stage coefficients. The coefficient magnitudes differ across subgroups because rates of opioid prescription differ. Finally, we check whether our main results hold using differential, mutually exclusive leniency measures (i.e., by major diagnosis category) in subsection 4.5.

4. Results

4.1 Reduced Form Results

In this section, we examine the causal effects of provider prescribing leniency. We first provide event study results on month-by-month prescription opioid use. Then we turn to presenting effects on our main outcomes of long-term opioid use, opioid-seeking behavior, and severe adverse consequences of opioid misuse (OUD and overdose mortality).

Our event study results (presented in Panel A of Figure 3) reveal a clear pattern: veterans who are treated by lenient providers start using prescription opioids at persistently higher rates than their counterparts treated by stricter providers. The figure plots the reduced form "event-study" analysis of new monthly opioid prescriptions filled for patients who see lenient and strict physicians. Physicians are classified as lenient (strict) if they fall into the top (bottom) quintile of the physician leniency measure. We plot average monthly (relative

[17] Frandsen et al. (2019) provide formal motivation for these two tests by proposing a weaker "average monotonicity" which prescribes that as long as each individual patient complies with monotonicity for a sufficient amount of judges, the 2SLS estimand can still be interpreted as a well-defined weighted average.

to ED encounter date) residualized prescription opioid use indicators for the two groups of patients. The ED prescription and any prescriptions filled in the seven days after the visit are excluded; only new filled prescriptions count toward each 30-day period data point. On average, veterans are getting sicker, experiencing more pain, and hence prescription opioid use rises in the months leading up to their ED encounter. Throughout the entire pre-period, the lenient and strict groups are on parallel, precisely overlapping paths.[18] In contrast, following the ED encounter, we observe a clear increase in prescription opioid use of patients treated in the ED by lenient physicians relative to their counterparts treated by strict physicians. The difference in opioid use is remarkably stable over time, and persists for at least 24 months. It is on the order of approximately 0.25 percentage points (average difference between the two groups in months 2 to 24) and can be attributed to a 27.4pp (42.3% - 14.9%) difference in opioid prescription rates between the lenient and strict prescribers.

Turning to our main outcomes of long-term opioid use, misuse, and its adverse consequences, we find statistically and economically significant positive effects of assignment to a leniently prescribing physician on every outcome (Table 3). Estimated coefficients are scaled by the difference in leniency going from the 90th to 10th percentile in physician leniency—equal to 11.6pp—for interpretability. Assignment to a physician in the top prescribing decile (relative to one in the bottom decile) is associated with a 0.24pp increase in the probability of long-term prescription opioid use—a 4.1% increase. The effect on opioid seeking behavior is 0.50pp—a 3.4% increase. Effects on development of opioid use disorder, and on opioid overdose mortality, are 2.1% and 8.8%, respectively. These findings highlight the substantial role ED physicians can play in putting patients on a path of long-term opioid use and dependence.

The fact that opioid-related outcomes in the form of downstream opioid use and misuse respond so strongly is suggestive of increased opioid use due to the initial ED prescription being the underlying mechanism behind the effects. However, physicians could differ in

[18]Thus, we do not find evidence for differential "opioid shopping" ahead of the ED visit for patients assigned to lenient prescribers. Nevertheless, in Figure G.3 and Figure G.4, the same reduced form figure is displayed for veterans who are opioid-naïve and veterans who never visited an ED in the prior year, but did have some other VHA outpatient encounter.

other dimensions of care—some observable and others not—which could be correlated with prescription leniency. The next section provides an attempt to distinguish between and identify the mechanisms behind the observed reduced form effects. This mediation analysis provides a crucial step on the path towards a well-identified IV analysis.

4.2 Placebo Check

In this section, we investigate whether the reduced form effects observed in Section 4.1 are due to differences in opioid prescription rates across providers, or due to other provider differences correlated with prescribing leniency. We start by studying reduced form effects among patients with diagnosis conditions that are never prescribed opioids, as a "placebo/falsification check". By way of example, consider a patient who arrives at the ED with a heart attack—a condition for which patients are rarely prescribed opioids. For such patients, we should expect to see no impact of leniency only if lenient and strict opioid prescribing physicians do not systematically differ in other dimensions of care relevant to patient outcomes. Conversely, if we do find a reduced form effect for these patients, then lenient physicians must systematically differ from strict physicians in other dimensions of care, beyond opioid prescribing.

To that end, we restrict attention to ED visits for health conditions prescribed no more than 10% of the time (recall that our baseline sample only includes conditions with a >10% prescription rate). We further split this sample into two equally sized bins, a bin with a 0-3% prescription rate, and one with a 3-10% prescription rate. The latter sample of visits provides the basis for our strongest “placebo” test. We estimate a reduced form regression of each main outcome on physician prescribing leniency for each of the sub-samples, following Equation 3. The results of this exercise are displayed in Table 4. They show that in contrast to results for our main sample, the association between physician leniency and a given opioid-related outcome is statistically indistinguishable from zero and much smaller in magnitude for the samples of patients who visit the ED with health conditions that are occasionally (column (1)) or rarely (column 2) prescribed opioids.

To further probe robustness, we perform the placebo exercise for a set of non-opioid

outcomes—in the form of homelessness, suicide, all-cause mortality, and preventable hospitalizations in Table G.7. Note that these outcomes are likely to be more sensitive to other dimensions of physician care than our opioid-related outcomes, because the causal link to prescription opioid exposure is less tight, while other margins of care may matter relatively more. Indeed, we do find a sizeable and statistically significant effect of being assigned to a high prescribing physician on preventable hospitalizations and all-cause mortality for our placebo conditions. For patients with ED diagnoses that are rarely prescribed an opioid (column 3), assignment to a physician in the top decile of the opioid prescribing range increases the likelihood of a preventable hospitalization within three years by 0.27pp, or 1.8%.[19]

The failure of our placebo check for some non-opioid outcomes suggests that there may be dimensions of care correlated with opioid prescribing tendency that vary across ED physicians and may impact non-opioid patient outcomes. While the set of potentially relevant dimensions of care is very large and may include dimensions unobservable to the researcher (such as, for example, instructions on medication adherence and use, physician-nurse scheduling that lead to complementarities), we can make progress by accounting for the chief observable dimensions likely to correlate highly with patient health outcomes in ED settings: quality of care as measured by immediate patient mortality, intensity of procedures, and tendency to admit patients for inpatient hospitalizations. In moving towards a well-identified IV design that estimates the causal impact of an opioid prescription on patient outcomes, we address these potential violations to the exclusion restriction by estimating physician "propensities" along these “non-focal” (relative to opioid prescribing) dimensions and including them as controls (details are provided in Section 4.5). This approach follows recent advances from the judge stringency literature (Mueller-Smith, 2015; Bhuller et al., 2020), which deals with similar concerns.

We recognize that given the multidimensionality of physician behavior, potential for exclusion restriction violations remains even after controlling for chief dimensions of behavior.

[19]In contrast, the associated coefficient estimate for our baseline sample is much smaller and statistically indistinguishable from zero. The difference could be explained by care differences across lenient and strict providers relevant to preventing future hospitalizations being minimal for health conditions that are often prescribed opioids.

For this reason, we also present reduced form results, and we keep a tight focus on opioid-related outcomes, for which such concerns are less pronounced relative to general health outcomes.

4.3 Instrumental Variable Results

In this section, we examine the causal effects of receiving an opioid prescription in the ED. Mirroring our presentation of the reduced form results, we first provide "event study" results on month-by-month prescription opioid use. Then we turn to presenting effects on our main outcomes of long-term opioid use, opioid-seeking behavior, and severe adverse consequences of opioid misuse (OUD and overdose mortality).

Our event study results for month-by-month opioid use are presented in Panel B of Figure 3; they represent the 2SLS analog of Panel A and are estimated by running separate regressions of Equation 4 for each month relative to a patient's ED encounter, with opioid use dummy as the outcome, instrumenting for $Prescribed_i$ with $Leniency_i$. The causal effect of an initial ED opioid on opioid use in subsequent months is highest for the first month at 5pp, and settles in at around 1.5% at around the six-month period, over an average opioid use rate of 9.5%.

Long-term prescription opioid use

We find that an ED opioid prescription increases long-term prescription opioid use by 1.17 pp, a 20% increase on the overall average long-term use rate (column 1 of Table 5). This number is robust to inclusion and exclusion of controls for baseline characteristics, as shown in Table G.3. By means of comparison, naïve OLS estimates are presented in Table G.2. Their magnitude is more than double that of the IV. The upward bias in the OLS is expected if, for example, patients with more unobservably severe conditions are more likely to be prescribed opioids and severe conditions require longer-term opioid treatment. Opioid use measured with urine and blood drug screens are presented in Table G.4; they are also large in magnitude and highly statistically significant.

Turning to the amount of prescription opioids filled by the patient, we find that an opioid prescription in the ED increases three-year morphine equivalent filled by 468 mg, excluding the initial ED prescription (Table G.4). This is equivalent to over 30 tablets of Oxycontin 10 mg—approximately a 20% increase in total observed mg of morphine filled in the three-year period after an average emergency visit. In subsection 4.5 we find that the differences in outcomes stemming from prescribing differences along the intensive margin (i.e., drug dosage) are minimal relative to the extensive margin.

Our estimate of 1.2pp on long-term use is about half the size of what is implied by the methods in Barnett et al. (2017, 2019). In Appendix A, we provide a detailed comparison of the methods and how they affect estimates of long-term use; the key difference is controlling for level of quasi-random assignment in EDs. Our estimate is also much smaller than estimates obtained from studies evaluating broader supply-side variation, such as opioid prescription rates in one's area of residence, or of one's primary care provider (Laird and Nielsen, 2016; Finkelstein et al., 2021; Eichmeyer and Zhang, 2020). This difference is due to our isolation of the impact of a one-off prescription opioid more precisely, relative to more "bundled" treatments evaluated in the other studies.

The effects on subsequent opioid use originating from prescribing variation in the ED are large and go against current medical guidelines. Both the CDC (Dowell et al., 2016) and the CMS (2019) have established recommendations and guidelines to limit opioid prescriptions to three days or fewer for acute pain and to encourage or to require physicians to consider non-opioid therapy first.

Opioid-seeking behavior

We provide IV results for our index of opioid-seeking behavior as a measure of potential prescription opioid abuse in column (2) of Table 5. We find a large positive effect of 2.46 pp (or 16.6%). The large positive effect is unanimously observed among all three individual proxies of opioid seeking behavior that enter the index (Table G.4). A prescription in the ED increases the one-year post-ED likelihood of overlapping prescriptions by 1.9pp on a base of

9.9% and pharmacy shopping by 0.3pp on a base of 0.6%. Veterans are 0.55pp more likely to be seen by a clinician for back pain, headaches, and migraines at least fives times in the first year following the ED visit, on a base of 6.2%. For this last outcome, we exclude patients whose ED visit is for back pain, headaches or migraines, in order to avoid the concern that increases in this proxy could be due to the condition being treated poorly in the ED.

The consistent, large positive impact on opioid seeking behavior provides evidence that an ED prescription increases not only long-term opioid use, but also medically unnecessary prescriptions and potential abuse. The increase in overlapping prescriptions suggests that the long-term use results cannot be explained by appropriate medical care for the emergency condition. Furthermore, the observed increase in "pharmacy shopping" for opioids suggests increased patient demand for opioids following exposure to opioids through the ED visit. Finally, the findings on increased visits for back pain and headaches, as well as increased average self-reported pain scores we document in Table G.5 are troubling. They further suggest that the benefits of prescription opioids may be out-weighted by the risks: Being prescribed opioids should decrease experienced pain, rather than increase it.

Adverse opioid outcomes: OUD and opioid overdose mortality

Our 2SLS findings on the most adverse consequences of opioid use, namely opioid use disorder and opioid overdose mortality, are presented in columns (3) and (4) of Table 5. We find large and troubling effects.

In the three years following the ED prescription, patients who are prescribed an opioid because of the leniency of their physician are 0.335pp more likely to develop an OUD, on a base of 3.27%, a 10% increase in development of an OUD within three years. An ED prescription increases the three-year probability of opioid overdose mortality by 0.075pp—a 45% increase over the base of 0.167%.[20]

In Table G.6, we distinguish between the type of opioid identified in post-mortem toxicology

[20]Our estimate is much lower than correlational studies like Bohnert et al. (2011) which find that conditional on dying from an opioid overdose, two-thirds of patients received opioid prescriptions; patients who were prescribed very high doses had over 660% higher risk-adjusted mortality rates.

tests, specifically, heroin or synthetic opioids (which mainly capture illicit opioids), and natural/semi-synthetic opioids. Although statistical power suffers with such mortality splits, some overall patterns arise: a substantial fraction (30-40%) of the total overdose mortality effect is accounted for by increased overdose deaths from heroin and synthetic opioids. This finding is consistent with prescription opioid exposure increasing the risk of transition into illicit drug use (which may carry higher overdose mortality risk). We investigate this hypothesis further in the next section.[21]

For comparison, OLS estimates are presented in columns (3) and (4) of Table G.2. Across both outcomes, the OLS estimates are biased downward relative to the IV estimates. The difference between OLS and IV is consistent with physicians recognizing patients who already exhibit clear signs of opioid use disorder or overdose risk, and prescribing at lower rates to such patients. This endogeneity in prescribing decisions would lead to a downward bias in the OLS relative to the IV. It is also reflected in Figure 4, where patients with prior opioid overdoses are less likely to be prescribed an opioid.

4.4 Investigating Mechanisms: Transitions Into Illicit Drug Use

Our estimates on heroin and synthetic opioid mortality suggest a causal link between prescription opioid exposure and transition into illicit drug use, which may in turn significantly increase overdose mortality risk. This pattern is consistent with striking correlational facts about opioid abuse: the vast majority (79.8%) of heroin users report having used prescription opioids non-medically prior to heroin initiation (Muhuri et al., 2013), and the majority of veterans who die from an opioid overdose do not fill opioid prescriptions in the prior year (Lin et al., 2019). However, due to lack of credible variation in individual exposure to medically prescribed opioids, causal evidence is missing from the literature. To make progress, we apply our IV approach to two additional outcomes: veterans' surveyed illicit drug use, and proxies for illicit drug use from veterans' medical records.

[21] Results for remaining secondary outcomes (opioid overdose diagnoses and accidental falls) are reported in Table G.4.

Our measure of self-reported illicit drug use is based on responses to the Brief Addiction Monitor (BAM)—a survey fielded to veterans in substance use disorder care (Gaddy et al., 2018). Responses were submitted by ca. 31,000 veterans in our baseline sample in the three years after the ED encounter. Since its sampling skews heavily towards veterans with drug addictions by design, we code veterans who do not respond to the survey as not using illicit drugs. Each respondent is asked whether they used various illegal or street drugs, and whether they abused prescription medications in the past 30 days. The survey distinguishes opiates, cocaine/crack, sedatives (such as benzodiazepines, valium), other stimulants (such as amphetamine, crystal meth), and marijuana. Results are reported in Table G.5. We find that self reported illicit/illegal or prescription drug abuse increases by 0.16pp, or about 20%, as a function of receipt of an ED opioid prescription. There is evidence of increased cocaine/crack use, suggesting that prescription opioids may be a gateway to other illicit drugs or that veterans are using mixed/laced illicit substances.[22]

Detailed medical records allow us to construct additional proxies for illicit drug use that do not rely on self reports. We focus on proxies for heroin and fentanyl use—hard drugs of the opiate class that account for the majority of overdose deaths (O'Donnell et al., 2017). Our proxies are i) physician intent to screen for heroin/fentanyl and ii) positive hepatitis C diagnosis (as a proxy for heroin injection).[23] 2SLS results are reported in Table G.5. The 2SLS coefficients for both proxies are positive and sizeable, but statistically insignificant.

Taken together, we find suggestive, but consistent evidence of increased use of illicit drugs due to previous exposure to medically prescribed opioids.

4.5 Robustness

The 2SLS results presented in the previous sections are robust to several alternative specifications probing the key identifying assumptions of independence, exclusion, and monotonicity.

[22]The increase in cocaine use could be due to increased heroin use: the two are sometimes combined (called "speedball") to enhance the effects of each drug (Ives and Ghelani, 2006; Leri et al., 2003).

[23]Both proxies are only noisy measures of heroin and fentanyl use, and likely severely undercount actual use. See Appendix C for details on variable construction.

We describe each robustness check in detail in Appendix D, and provide a brief overview here. To probe the independence assumption, we perform two checks: First, we take into account potential endogeneity in diagnosing behavior by no longer excluding any diagnosis from our analysis sample, and by replacing the original diagnosis control with the most recent diagnosis code observed *prior* to the ED visit (results are presented in Table G.8, column 2). Second, we replace the physician leniency instrument with a shift-level instrument that captures the average leniency across all physicians on shift (column 3). To provide robustness checks to potential exclusion restriction violations, we adjust 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 (columns 4-6). To probe monotonicity, we we construct physician-diagnosis specific leniency instruments (column 7), and we check whether our instrument has a positive first stage across key demographic- and health-related sub-samples (Table G.12).

Across all robustness checks and main outcomes, we find magnitudes of our estimates broadly unchanged. Precision of estimates remains unchanged for the outcomes of long-term use and opioid seeking behavior, while estimates for the outcomes of opioid use disorder and opioid overdose mortality become noisier in some specifications.

5. Implications and Welfare Calculation

In this section, we draw on the results from Section 4 to derive implications for opioid prescribing regulation, characterize the fraction and characteristics of compliers, as well as perform stylized welfare calculations that help gauge welfare losses due to lenient prescribing practices.

5.1 Intensive vs. Extensive Margin Prescribing Regulation

Our robustness check accounting for intensive margin decisions about the amount of opioids prescribed revealed that estimates of the extensive margin effect of any opioid prescribed remain unchanged— implying that intensive margin prescribing differences are not biasing our results. This finding is consistent with two scenarios: a) within observed practice variation, the strength of an initial prescription has no impact on patient outcomes; b) correlation between extensive and intensive margin prescribing tendencies are so small that controlling for the intensive margin does not impact our results. Since scenario a) carries big policy implications for optimal prescribing regulation, we investigate it further by constructing quintile dummies for the strength of prescription received. We then estimate each physician's propensity to prescribe an opioid above a given strength quintile, and then run a 2SLS regression with the five strength quintiles on the right hand side, using the physician's propensity to prescribe above each quintile as instruments. Results are reported in Table G.11. We find treatment effects that are similar across prescription strength bins, in line with the first scenario. This finding has important implications for optimal regulation of opioid prescribing: It suggests that for a sample of relatively opioid-naïve patients, regulations relating to the *intensive* margin of opioid prescriptions may not be as successful in preventing new cases of opioid use disorder as regulation of the *extensive* margin.

5.2 Complier Analysis

Our 2SLS model identifies a LATE for those patients for whom the assigned ED physician's leniency determined whether they received an opioid prescription. Patients with certain pre-existing conditions and demographics are at higher risk of developing opioid dependence (Ives et al., 2006; Zedler et al., 2014); therefore, it is important to characterize the compliers (Abadie, 2003; Dahl et al., 2014, see Appendix G for more details).

Approximately 39.4% of our baseline sample—veterans who show up in the ED with conditions that are prescribed at least sometimes and are not heavy prior opioid users—are

compliers. These patients would be sent home without an opioid prescription if they saw the most strict physician, and would be sent home with a prescription if they saw the most lenient physician. The share of compliers is large and it underscores the wide variation and lack of consensus among physicians in prescribing opioids.[24] In comparison, 7.6% are always-takers and 53% are never-takers.[25]

Table G.13 reports for each demographic subgroup, its unconditional share, its conditional probability given patients are compliers, and the relative likelihood. Compliers are 4.6% more likely to be middle-aged (ages 40-60) and 17% more likely to have musculoskeletal or connective tissue conditions compared to the baseline sample. Another interesting avenue to study is whether compliers are more or less likely to be at risk for severe opioid outcomes. We predict ex-ante risk of opioid overdose death using veteran characteristics and medical history.[26] Veteran compliers are 6.4% more (10.2% less) likely to be above (below) average risk for opioid overdose death prior to their ED visit. In sum, compliers tend to be middle-aged veterans who show up with musculoskeletal or connective tissue conditions and are at higher risk for opioid overdose mortality. The higher risk (complier) patients tend to be the ones for whom physicians differ in their treatment approach to pain management (i.e., opioid vs. non-opioid treatment), explaining why the 2SLS magnitudes for overdose and overdose mortality are larger than their OLS counterparts.

5.3 Welfare Calculations

In this section, we apply our reduced form approach to construct an estimate of the welfare loss generated by leniently prescribing physicians. We approximate welfare loss in terms of patient life years lost. Each veteran's life years lost can be calculated by their life expectancy

[24] In other contexts, Dobbie et al. (2017) find 13% of consumer bankruptcy cases are compliers, 14% of Norwegian criminals facing incarceration are compliers in Bhuller et al. (2020). For Norwegians applying for disability insurance, 25% are compliers (Dahl et al., 2014).

[25] For the universe of veteran emergency visits (including repeat visits, heavy prior users, all diagnosis conditions, etc.), the fractions of compliers, always-takers, and never-takers are 24%, 1.3%, and 75% respectively.

[26] Risk is predicted using LASSO for veteran observables, medical morbidities prior to ED (suicide, mental health, falls, etc.), prior opioid use, and prior opioid-seeking behavior.

conditional on their current age (imputed from VA, 2017) minus their actual age at death. Veterans still alive at the end of the data period are coded as having zero life years lost. This outcome is then regressed on physician leniency, using the baseline reduced form model from Equation 3; the results are displayed in Table G.14. At the individual patient-level, being treated by a 90th percentile lenient physician leads to a 0.025 (or 9 days) increase in life years lost relative to a 10th percentile lenient physician. Given an annual total of 17,741 patients in our baseline sample who are seen by physicians prescribing at the 90th percentile or above, this estimate aggregates to 442 life years lost per year, 139 of which directly attributed to opioid overdose deaths.[27] As a point of comparison, the annual life years lost among veterans from seasonal influenza is 9,330 (Young-Xu et al., 2017)—approximately 21× that of leniently prescribing ED physicians.[28]

These sizeable, troubling costs of lenient prescribing stand in contrast to the small, or lack thereof, benefits of opioid treatment as a means of pain management. Krebs et al. (2018) find that opioid treatment is no more effective than acetaminophen or non-steroidal anti-inflammatory drugs at treating chronic back pain or knee or hip osteoarthritis pain. The same is true for acute lower back pain (Friedman et al., 2015), kidney stone pain (Teichman, 2004; Holdgate and Pollock, 2004), and minor fractures (Dodwell et al., 2010). Opioid dose reduction is also not associated with increased pain severity among veterans (Frank et al., 2020).

6. Conclusion

Our results highlight that long-term opioid use, misuse and dependence can arise as a consequence of small variations in medical care received in a single medical encounter, at

[27] Note that this estimate is a slight under-estimate, because we assume, for simplicity, 90th percentile prescribing rates for all patients at or above the 90th percentile.

[28] Additionally, we can use our 2SLS results on overdose mortality to construct a stylized estimate of the number of overdose deaths attributable to an initial prescription from the ED. Between 2006-2016, we estimate that approximately 1.8% of all veteran opioid overdose deaths can be attributed to exposure to prescription opioids through a leniently prescribing ED physician (see Appendix E for this back-of-the-envelope calculation).

the emergency department. We find evidence that exposure to opioids via a prescription from the ED physician is the margin of medical care most likely triggering these long-term effects. Our instrumental variable approach suggests that the causal effects of short-term exposure to prescription opioids are substantial: An opioid prescription originating from a ED visit increases the probability of long-term prescription opioid use, opioid use disorder, and overdose mortality by 1.2 pp, 0.34pp, and 0.075pp, respectively. At the same time, we do not find evidence for improved health due to use of prescription opioids. Finally, we find evidence that the effects of differences in the amount of opioids prescribed (i.e., dosage; intensive margin) are dwarfed by the effects from whether the patient gets *any* opioid prescription (i.e., extensive margin).

Taken together, the findings suggest that for individuals like the ones in our study—patients who are not already heavy prescription opioid users, and may have mental health co-morbidities—the short-term benefit of pain relief from prescription opioids comes at a high cost. Beyond physician decision-making, our findings have important implications for opioid prescribing regulation: They suggest that regulation restricting strength and length of prescriptions, a type of regulation widely used across states in the US, is likely to be less impactful in preventing new cases of opioid dependence compared to regulation restricting the extensive margin.

Note, however, that these findings apply to the patient population for which clinicians differ in their choices to prescribe opioids-a population we estimate to be 39% of all patients in our sample. Our research design cannot provide insight into the impact of opioid prescribing for cases where clinicians universally chose to prescribe or not prescribe.

References

Abadie, Alberto (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113, 231–263.

Babor, Thomas F., John C. Higgins-Biddle, John B. Saunders, and Maristela G. Monteiro (2001), "The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care." *World Health Organization*, Second Edition.

Barnett, Michael L., Andrew Olenski, and Anupam B. Jena (2017), "Opioid-Prescribing Patterns of Emergency Physicians and Risk of Long-Term Use." *New England Journal of Medicine*, 376, 663–673.

Barnett, Michael L., Xinhua Zhao, Michael J. Fine, Carolyn T. Thorpe, Florentina E. Sileanu, John P. Cashy, Maria K. Mor, Thomas R. Radomski, Leslie R. M. Hausmann, Chester B. Good, and Walid F. Gellad (2019), "Emergency Physician Opioid Prescribing and Risk of Long-term Use in the Veterans Health Administration: an Observational Analysis." *Journal of General Internal Medicine*, 34, 1522–1529.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad (2020), "Incarceration, Recidivism, and Employment." *Journal of Political Economy*, 128, 1269–1324.

Bohnert, Amy S. B., Marcia Valenstein, MD Matthew J. Bair, Dara Ganoczy, John F. McCarthy, Mark A. Ilgen, and Frederic C. Blow (2011), "Association Between Opioid Prescribing Patterns and Opioid Overdose-Related Deaths." *Journal of American Medical Association*, 305, 1315–1521.

Boscarino, Joseph A., Margaret Rukstalis, Stuart N. Hoffman, and John J. Han Porat M. Erlich Glenn S. Gerhard Walter F. Stewart (2010), "Risk factors for drug dependence among out-patients on opioid therapy in a large US health-care system." *Addiction*, 105, 1776–1782.

Busse, Jason W., Li Wang, and Mostafa Kamaleldin et al. (2018), "Opioids for chronic noncancer pain: A systematic review and meta-analysis."

Case, Anne and Angus Deaton (2015), "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century." *PNAS*, 112, 15078–15083.

CDC (2016), "Surveillance for Viral Hepatitis – United States, 2016. Figure 4.6a."

CDC (2018), "FastStats - Emergency Department Visits."

CMS (2019), "A Prescriber's Guide to the New Medicare Part D Opioid Overutilization Policies for 2019 ."

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad (2014), "Family Welfare Cultures." *The Quarterly Journal of Economics*, 129, 1711–1752.

Degenhardt, Louisa, Amy Peacock, Samantha Colledge, Janni Leung, Jason Grebely, Peter Vickerman, Jack Stone, Evan B Cunningham, Adam Trickey, Kostyantyn Dumchev, Michael Lynskey, Paul Griffiths, Richard P Mattick, Matthew Hickman, and Sarah Larney (2017), "Global prevalence of injecting drug use and sociodemographic characteristics and prevalence of HIV, HBV, and HCV in people who inject drugs: a multistage systematic review." *The Lancet Global Health*, 5, 1192–1207.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang (2018), "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *Amercian Economic Review*, 108, 201–240.

Dobbie, Will, Paul Goldsmith-Pinkham, and Crystal S. Yang (2017), "Consumer Bankruptcy and Financial Health." *Review of Economics and Statistics*, 99, 853–869.

Dodwell, Emily R., Julius Gene Latorre, Emilio Parisini, Elisabeth Zwettler, Divay Chandra, Kishore Mulpuri, and Brian Snyder (2010), "NSAID exposure and risk of nonunion: a meta-analysis of case-control and cohort studies." *Calcified Tissue International*, 87, 193–202.

Dowell, Deborah, Tamara M. Haegerich, and Roger Chou (2016), "CDC Guideline for Prescribing Opioids for Chronic Pain — United States."

Doyle, Joseph J., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner (2015), "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy*, 123, 170–214.

Doyle, Jospeph J. (2007), "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97, 1583–1610.

Duggan, Mark (2005), "Do new prescription drugs pay for themselves? The case of second-generation antipsychotics." *Journal of Health Economics*, 24, 1–31.

Dunn, Kate M., Kathleen W. Saunders, Carolyn M. Rutter, Caleb J. Banta-Green, Joseph O. Merrill, Mark D. Sullivan and; Constance M. Weisner, Michael J. Silverberg, Cynthia I. Campbell, Bruce M. Psaty, and Michael Von Korff (2010), "Opioid prescriptions for chronic pain and overdose: a cohort study." *Annals of Internal Medicine*, 152, 85–92.

Edlund, Mark, Bradley C. Martin, Joan E. Russo, Andrea Devries, Jennifer Brennan Braden, and Mark D. Sullivan (2014), "The Role of Opioid Prescription in Incident Opioid Abuse and Dependence Among Individuals with Chronic Non-Cancer Pain: The Role of Opioid Prescription." *The Clinical Journal of Pain*, 30, 557–564.

Eichmeyer, Sarah and Jonathan Zhang (2020), "Prescription Opioid Use, Physical and Mental Health: Evidence from Primary Care." Working paper.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist (2020), "What Is a Patent Worth? Evidence from the U.S. Patent 'Lottery'." *Journal of Finance*, 75, 639–682.

Finkelstein, Amy, Matthew Gentzkow, Dean Li, and Heidi Williams (2021), "What Drives Prescription Opioid Abuse? Evidence from Migration." Working paper.

Frandsen, Brigham R., Lars J. Lefgren, and Emily C. Leslie (2019), "Judging Judge Fixed Effects." Working paper.

Frank, Joseph W., Evan Carey, Charlotte Nolan, Anne Hale, Sean Nugent, and Erin E. Krebs (2020), "Association Between Opioid Dose Reduction Against Patients' Wishes and Change in Pain Severity." *Journal of General Internal Medicine*, 35, 910–917.

Friedman, Benjamin W., Andrew A. Dym, Michelle Davitt, Lynne Holden, Clemencia Solorzano, David Esses, Polly E. Bijur, and E. John Gallagher (2015), "Naproxen With Cyclobenzaprine, Oxycodone/Acetaminophen, or Placebo for Treating Acute Low Back Pain A Randomized Clinical Trial." *The Journal of American Medical Association*, 315, 1572–1580.

Gaddy, Melinda A, Hilary G Casner, and Jenny Rosinski (2018), "Factor structure and measurement invariance of the Brief Addiction Monitor." *J Subst Abuse Treat*, July.

Hagan, Holly, Enrique R. Pouget, Don C. Des Jarlais, and Corina Lelutiu-Weinberger (2008), "Meta-Regression of Hepatitis C Virus Infection in Relation to Time Since Onset of Illicit Drug Injection: The Influence of Time and Place." *American Journal of Epidemiology*, 168, 1099–1109.

Holdgate, Anna and Tamara Pollock (2004), "Systematic review of the relative efficacy of non-steroidal anti-inflammatory drugs and opioids in the treatment of acute renal colic." *The British Medical Journal*, 328, 1401–8.

Huang, Grace, Sharon Kim, Benjamin Muz, and Joseph Gasper (2018), "2017 Survey of Veteran Enrollees' Health and Use of Health Care." Technical report.

Hull, Peter (2020), "Estimating Hospital Quality with Quasi-Experimental Data." Working paper.

Ives, Richard and Preeti Ghelani (2006), "Polydrug use (the use of drugs in combination): A brief review." *Drugs: Education, Prevention and Policy*, 13, 225–232.

Ives, Timothy J, Paul R Chelminski, Catherine A Hammett-Stabler, Robert M Malone, J Stephen Perhac, Nicholas M Potisek, Betsy Bryant Shilliday, Darren A DeWalt, and Michael P Pignone (2006), "Predictors of opioid misuse in patients with chronic pain: a prospective cohort study." *BMC Health Services Research*, 6.

Jena, Anupam B., Dana Goldman, and Pinar Karaca-Mandic (2016), "Hospital Prescribing of Opioids to Medicare Beneficiaries." *The Journal of American Medical Association Internal Medicine*, 176, 990–997.

Kling, Jeffrey R. (2006), "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96, 863–876.

Krebs, Erin E., Amy Gravely, Sean Nugent, Agnes C. Jensen, Beth DeRonne, Elizabeth S. Goldsmith, MD4 Kurt Kroenke, Matthew J. Bair, and Siamak Noorbaloochi (2018), "Effect of Opioid vs Nonopioid Medications on Pain-Related Function in Patients With Chronic

Back Pain or Hip or Knee Osteoarthritis Pain: The SPACE Randomized Clinical Trial." *The Journal of American Medical Association*, 319, 872–882.

Laird, Jessica and Torben Nielsen (2016), "The Effects of Physician Prescribing Behaviors on Prescription Drug Use and Labor Supply: Evidence from Movers in Denmark." Working paper.

Leri, Francesco, Julie Bruneau, and Jane Stewart (2003), "Understanding polydrug use: review of heroin and cocaine co-use." *Addiction*, 98, 7–22.

Lin, Lewei (Allison), Talya Peltzman, John F. McCarthy, Elizabeth M. Oliva, Jodie A. Trafton, and Amy S.B. Bohnert (2019), "Changing Trends in Opioid Overdose Deaths and Prescription Opioid Receipt Among Veterans." *American Journal of Preventive Medicine*, 57, 106–110.

Maclean, Johanna Catherine, Justine Mallatt, Christopher J. Ruhm, and Kosali Simon (2020), "Economic Studies on the Opioid Crisis: A Review." Nber working paper 28067.

Moeller, Karen E., Kelly C. Lee, and Julie C. Kissack (2008), "Urine Drug Screening: Practical Guide for Clinicians." *Mayo Clinic Proceedings*, 83, 66–76.

Mueller-Smith, Michael (2015), "The criminal and labor market impacts of incarceration." Working paper.

Muhuri, Pradip K., Joseph C. Gfroerer, and M. Christine Davies (2013), "Associations of nonmedical pain reliever use and initiation of heroin use in the United States." *CBHSQ Data Review*.

NBER (2014a), "ICD-9-CM to and from ICD-10-CM and ICD-10-PCS Crosswalk or General Equivalence Mappings." https://www.nber.org/research/data/icd-9-cm-and-icd-10-cm-and-icd-10-pcs-crosswalk-or-general-equivalence-mappings. Date Accessed: 2018-12-16.

NBER (2014b), "International Classification of Diseases (ICD) to Clinical Classifications Crosswalk." https://www.nber.org/research/data/international-classification-diseases-icd-clinical-classifications-crosswalk. Date Accessed: 2018-12-16.

NCHS (2016), "National Hospital Ambulatory Medical Care Survey: 2016 emergency department summary tables."

Neprash, Hannah T. and Michael L. Barnett (2019), "Association of Primary Care Clinic Appointment Time With Opioid Prescribing." *Journal of American Medical Association Network Open*, 2.

Oliva, Elizabeth M., Thomas Bowe, Sara Tavakoli, Eleanor T. Lewis, Ilse Weichers, Michael Harvey, Amanuel Medhanie, Susana Martins, Meenah Paik, Patricia Henderson Tigran Avoundjian, and Jodie A. Trafton (2017), "Development and Applications of the Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to

Improve Opioid Safety and Prevent Overdose and Suicide." *Psychological Services*, 14, 34–49.

O'Donnell, Julie K., R. Matthew Gladden, and Puja Seth (2017), "Trends in deaths involving heroin and synthetic opioids excluding methadone, and law enforcement drug product reports, by census region—United States, 2006–2015." *CDC Morbidity and Mortality Weekly Report 2017*, 66.

Powell, David, Abby Alpert, and Rosalie L. Pacula (2019), "A Transitioning Epidemic: How The Opioid Crisis Is Driving The Rise In Hepatitis C." *Health Affairs*, 38, 287–294.

Sandbrink, Friedhelm, Elizabeth M. Oliva, Tara L. McMullen, Amy R. Aylor, Michael A. Harvey, Melissa L. Christopher, Francesca Cunningham, Taeko Minegishi, Thomas Emmendorfer, and Jenie M. Perry (2020), "Opioid Prescribing and Opioid Risk Mitigation Strategies in the Veterans Health Administration." *Journal of General Internal Medicine*, 35, 927–934.

Scholl, Lawrence, Puja Seth, Mbabazi Kariisa, Nana Wilson, and Grant Baldwin (2019), "Drug and Opioid-Involved Overdose Deaths — United States, 2013–2017." Working paper.

Seal, Karen H, Greg Cohen, Angela Waldrop, Beth E. Cohen, Shira Maguen, and Li Ren (2011), "Substance use disorders in Iraq and Afghanistan veterans in VA healthcare, 2001–2010: Implications for screening, diagnosis and treatment." *Drug and Alcohol Dependence*, 116.

Seal, Karen H., Ying Shi, Gregory Cohen, Beth E. Cohen, Shira Maguen, Erin E. Krebs, and Thomas C. Neylan (2012), "Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan." *The Journal of American Medical Association*, 307, 940–947.

Starrels, Joanna L., Aaron D. Fox, Hillary V. Kunins, and Chinazo O. Cunningham (2012), "They Don't Know What They Don't Know: Internal Medicine Residents' Knowledge and Confidence in Urine Drug Test Interpretation for Patients with Chronic Pain." *Journal of General Internal Medicine*, 27, 1521–1527.

Teichman, Joel M.H. (2004), "Acute renal colic from ureteral calculus." *The New England Journal of Medicine*, 350, 684–693.

VA (2017), "Mortality Rates and Life Expectancy of Veterans From 1980 to 2014, and by Education and Income." Technical report.

Veterans Affairs (2018), "VA Mortality Data Repository." https://www.mirecc.va.gov/suicideprevention/Data/data_index.asp.

Veterans Health Administration (2019), "Corporate Data Warehouse." https://www.data.va.gov/dataset/Corporate-Data-Warehouse-CDW-/ftpi-epf7.

Veterans Health Administration and Centers for Medicare Medicaid Services (2016), "VA/CMS Data." https://www.virec.research.va.gov/VACMS/About.asp#About.

Yang, Zhuo, Barth Wilsey, Michele Bohm, Meghan Weyrich, Kakoli Roy, Dominique Ritley, Christopher Jones, and Joy Melnikowx (2015), "Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-Term Opioid Users in Medicaid." *The Journal of Pain*, 16, 445–453.

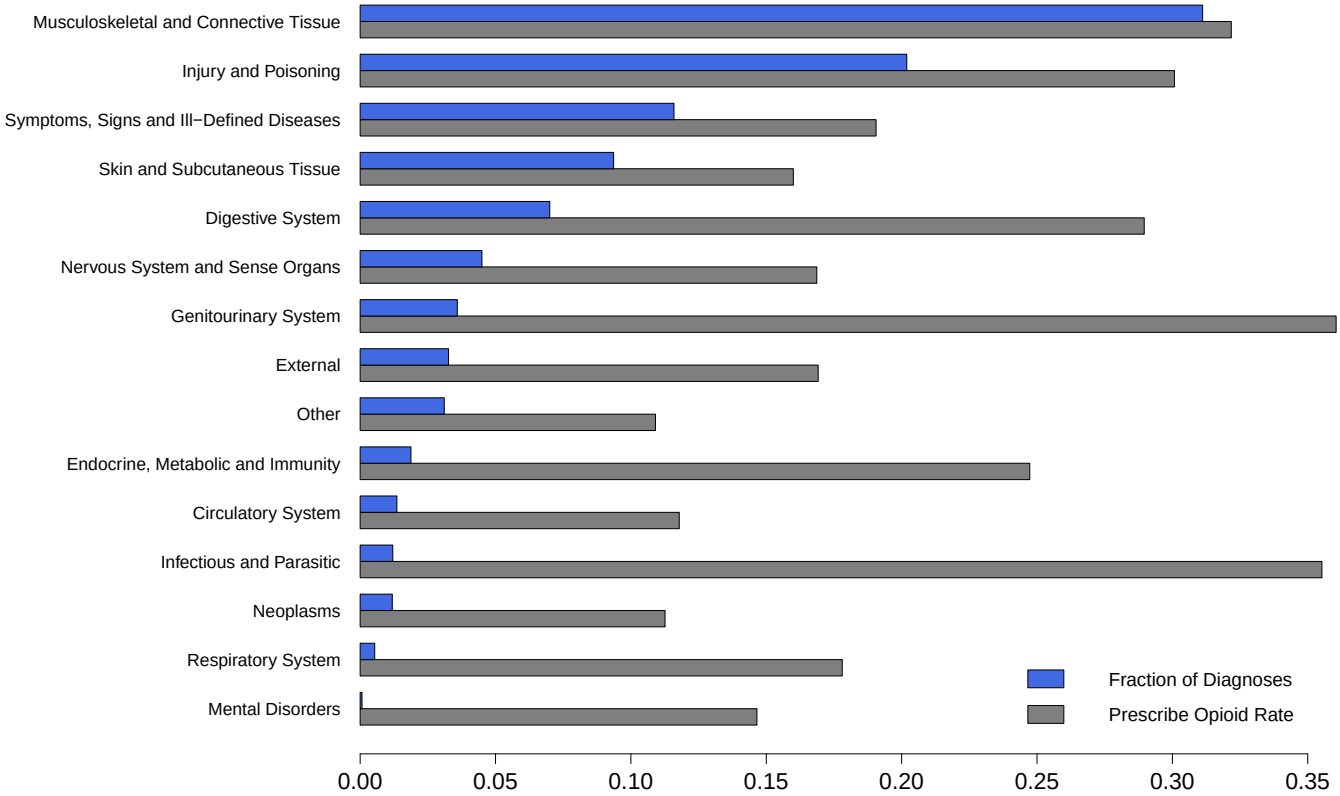
Young-Xu, Yinong, Robertus van Aalst, Ellyn Russo, Jason K. H. Lee, and Ayman Chit (2017), "The Annual Burden of Seasonal Influenza in the US Veterans Affairs Population." *PLoS One*, 12.

Zedler, Barbara, Lin Xie, Li Wang, Andrew Joyce, Catherine Vick, Furaha Kariburyo, Pradeep Rajan, and Onur Baserand Lenn Murrelle (2014), "Risk Factors for Serious Prescription Opioid-Related Toxicity or Overdose among Veterans Health Administration Patients." *Pain Medicine*, 15, 1911–1929.

Zibbell, Jon E., Alice K. Asher, Rajiv C. Patel, Ben Kupronis, Kashif Iqbal, John W. Ward MD, and Deborah Holtzman (2018), "Increases in Acute Hepatitis C Virus Infection Related to a Growing Opioid Epidemic and Associated Injection Drug Use, United States, 2004 to 2014." *American Journal of Public Health*, 108, 175–181.

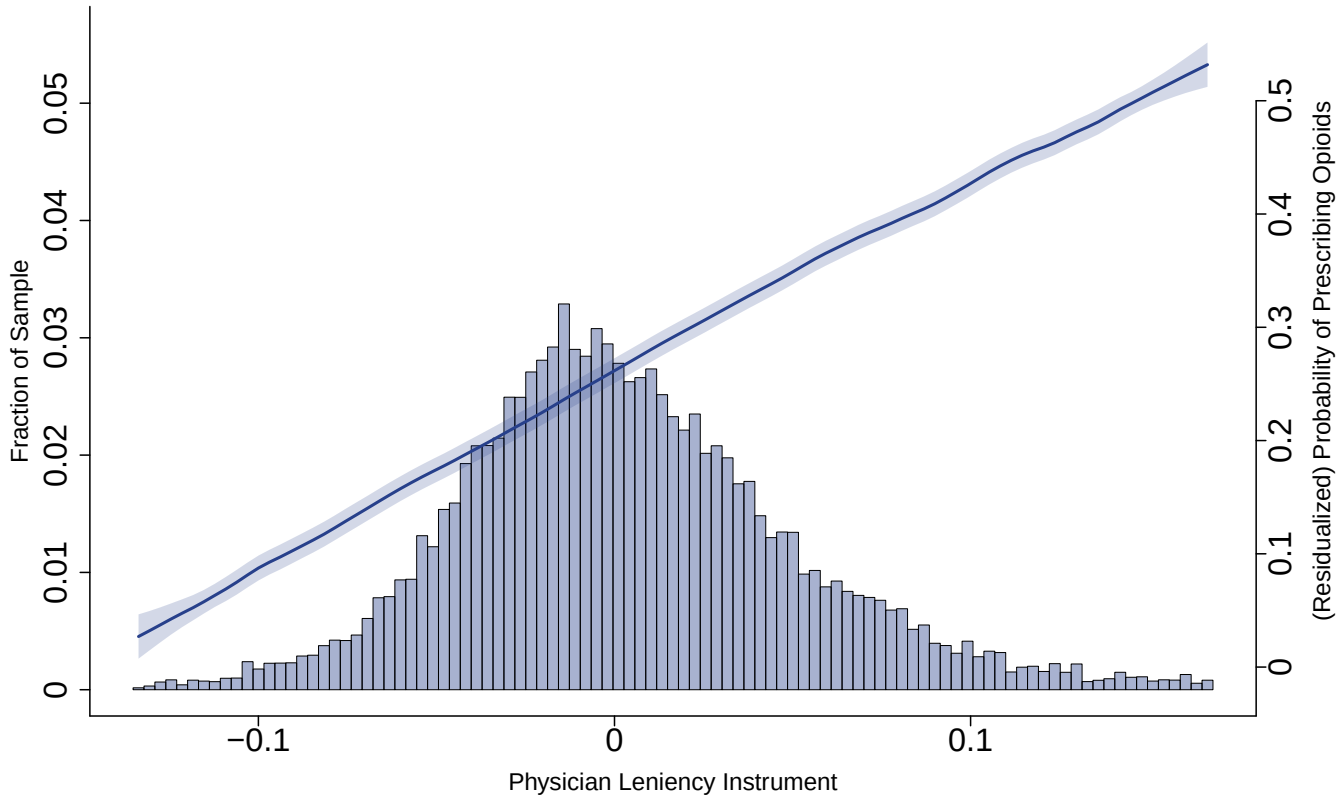
Figures

Figure 1: Frequent Diagnoses Occurring in Emergency Departments



Notes: This figure displays two statistics for the 15 most common major diagnosis categories (ICD-9 major chapters) observed in our baseline ED visit sample: fraction of all ED visits that pertain to a given diagnosis category (blue bars), and un-adjusted opioid prescription rate (gray bars). See subsection 2.2 for details on baseline sample construction.

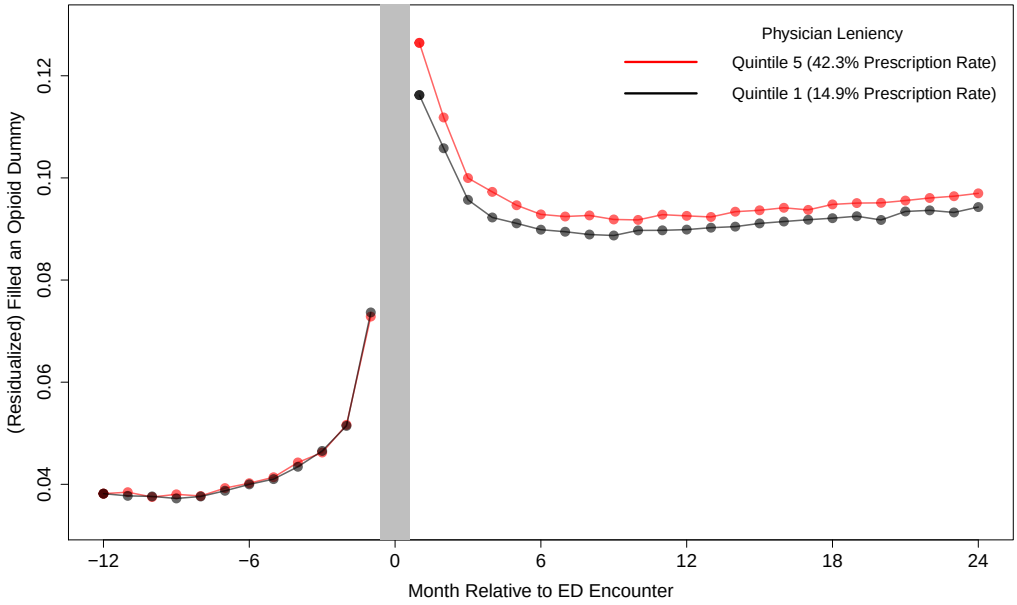
Figure 2: Distribution and First Stage of Instrument



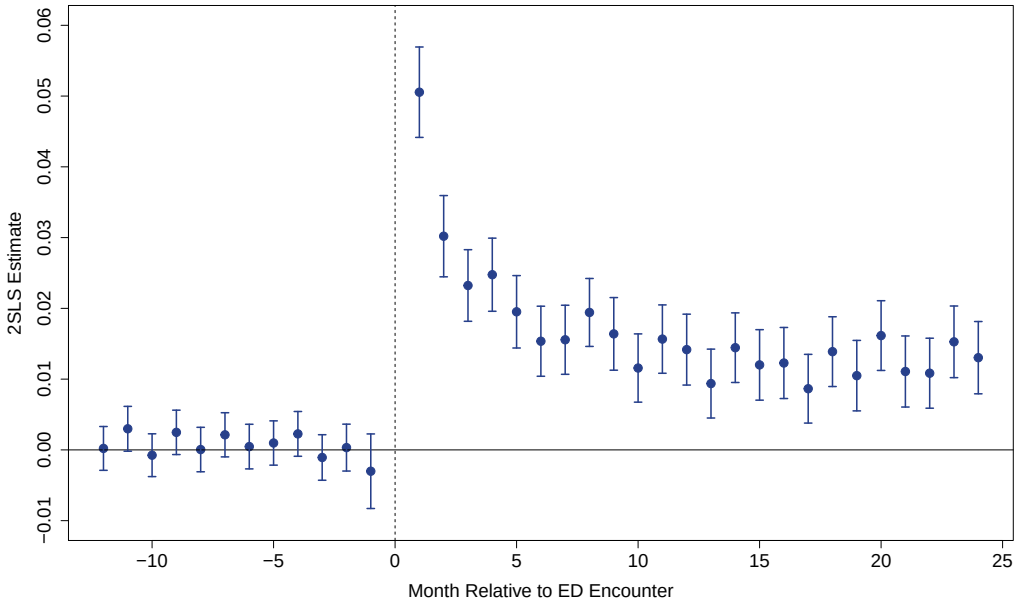
Notes: This figure plots the histogram of physician leniency along the x-axis and the left y-axis for our baseline sample. A local-linear regression of the fitted probability of prescribed opioids on leniency after residualizing (see text for baseline fixed effects and controls in residualization) is overlaid and displayed on the right y-axis. 95% confidence bands are also shown.

Figure 3: Reduced Form and 2SLS Event Study for Prescription Opioid Use

Panel A. Reduced Form

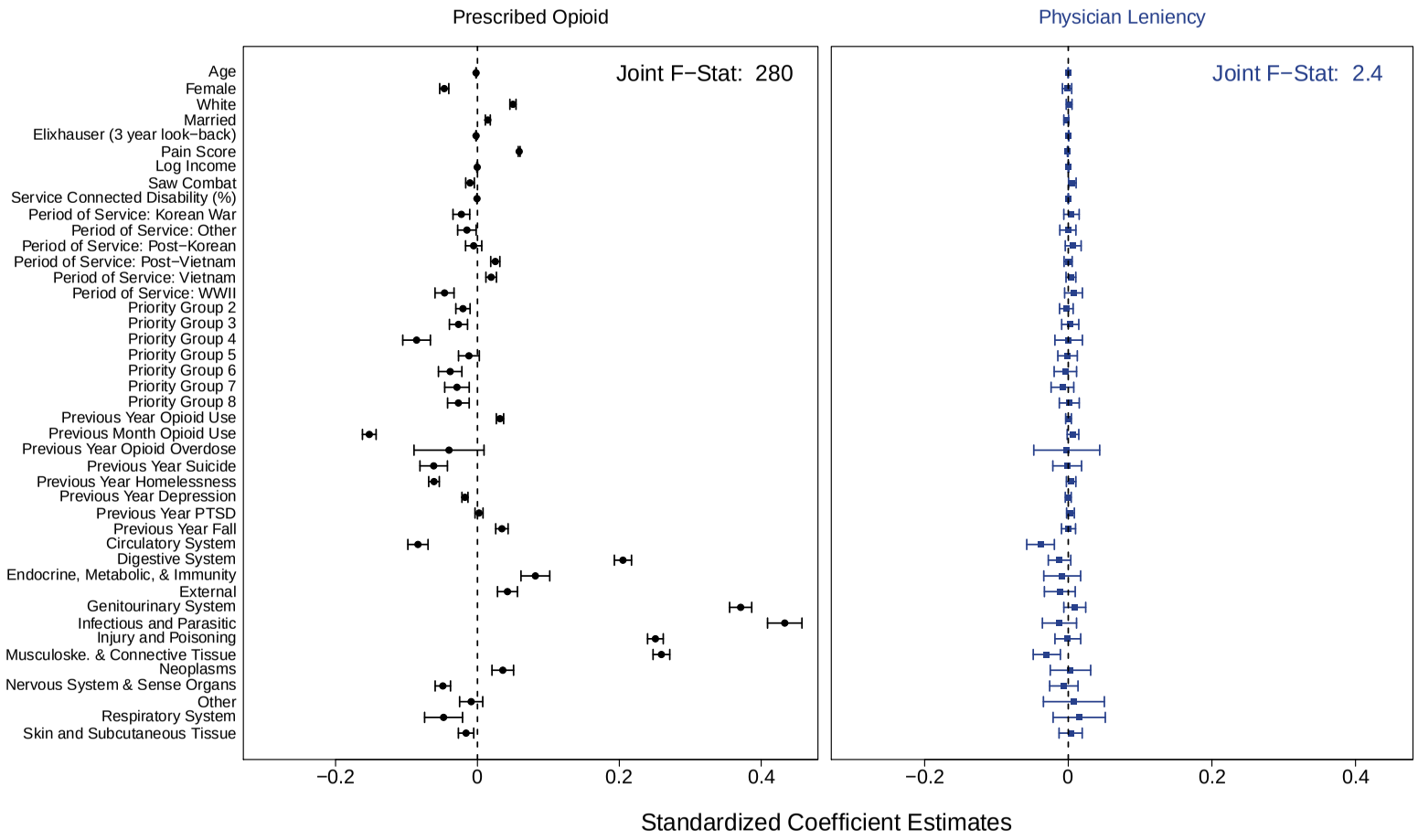


Panel B. 2-Stage Least Squares



Notes: Panels A and B are based on observations at the patient-month level and include all patients in the baseline sample. Panel A plots monthly residualized opioid use rates (see text for baseline fixed effects and controls in residualization), for patients who see a strict physician (blue) and for patients who see a lenient physician (red). Panel B is the 2SLS analog: each point is an estimate from a separate regression of an opioid use dummy that month on $Prescribed_i$ (a dummy for opioid prescription receipt at ED visit), instrumented with physician leniency. In both panels, only new filled opioid prescriptions contribute to each patient-month observation; the initial ED prescription, as well as prescriptions filled in the week after the ED encounter are excluded. Month is defined as 30-day periods relative to ED encounter. 95% confidence intervals constructed using robust standard errors clustered at the physician level are also displayed.

Figure 4: Balance Test for Quasi-Random Assignment



Notes: This figure plots a test for quasi-random assignment of patients to physicians in EDs for our baseline sample. The first column regresses prescribed opioid indicator on all observables, jointly (patient demographics, ED visit variables, and previous medical history), and the second column is the same regression but with the physician leniency instrument as the dependent variable. Both dependent variables are standardized. Construction of instrument is described in the text. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, and 3-digit diagnosis codes. The joint F-statistics are reported. The number of observations is $1,672,553$ for both regressions. Robust standard errors are clustered at the physician level.

Tables

Table 1: Summary Statistics

	Mean	Q1	Median	Q3
Panel A: Prescription Characteristics				
Prescribed	0.261			
Prescribed | ED visit in 2011	0.285			
Prescribed | ED visit in 2016	0.229			
Days Supply | Prescribed	8.2	4	5	10
Daily Milligrams of Morphine | Prescribed	21.0	15	20	25
Opioid: Hydrocodone | Prescribed	0.576			
Opioid: Tramadol | Prescribed	0.165			
Opioid: Oxycodone | Prescribed	0.127			
Panel B: Emergency Department Characteristics				
Patient is admitted	0.083			
ED Diagnosis: Musculoskeletal and Connective Tissue	0.311			
ED Diagnosis: Injury and Poisoning	0.202			
ED Diagnosis: Symptoms, Signs and Ill-Defined Diseases	0.116			
ED Diagnosis: Skin and Subcutaneous Tissue	0.094			
ED Diagnosis: Digestive System	0.07			
Panel C: Veteran Characteristics				
Age	55.1	44	57	66
Income	21,009	4,068	15,000	31,128
Female	0.104			
White	0.69			
Black	0.24			
Married	0.43			
Panel D: Patient Medical History				
Prior Year Opioid Use Indicator	0.27			
Prior Month Opioid Use Indicator	0.10			
Prior Year Total MME	214.3	0	0	60
Prior Year Depression Diagnosis Indicator	0.241			
Prior Year PTSD Diagnosis Indicator	0.134			
Prior Year Homeless Indicator	0.062			
Elixhauser Comorbidity Index (3-year look-back)	1.2	-1	0	2
Observations: 1,958,209				

Notes: This table reports summary statistics for the baseline sample of emergency department visits between 2006-2016 described in the text. Panel A and B summarize characteristics related to the ED opioid prescription and ED visit. Panel C and D summarize patient veteran and patient medical history.

Table 2: First Stage: Effect of Physician Leniency on ED Opioid Prescription

	Dependent Variable: Prescribed in ED		
	(1)	(2)	(3)
Physician Leniency	1.691***	1.702***	1.710***
	(0.012)	(0.012)	(0.012)
Hospital, Seasonality, Shift FE?	Yes	Yes	Yes
Diagnosis and Elixhauser?	No	Yes	Yes
Patient Observables?	No	No	Yes
F-Stat	12	21	25
Observations:	1,958,209	1,958,209	1,958,209

Notes: Estimates of the first stage for the baseline sample described in the text. Hospital, seasonality, shift fixed effects include Hospital-Year-Month and Hospital-Day of week-Hour of day fixed effects. Elixhauser comorbidity is constructed with a 3-year look-back period, excluding the ED encounter. Patient observables include female dummy, black race/ethnicity dummy, prior month opioid use, age bins, and log prior year total milligrams of morphine equivalent. Column 3 corresponds to the baseline controls. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table 3: Reduced Form Results: Opioid Use, Misuse, Addiction, and Overdose Mortality

	Dependent variable: (×100)			
	Long-Term Use	Opioid-Seeking Behavior	Opioid Use Disorder	Opioid Overdose Mortality
	(1)	(2)	(3)	(4)
Physician Leniency	0.237***	0.496***	0.067*	0.015**
	(0.041)	(0.063)	(0.038)	(0.007)
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 a reduced form regression of our main opioid-related outcomes on physician prescribing leniency among patients in our baseline sample. 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. All coefficients are scaled by 100 and multiplied again 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. 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. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table 4: Placebo Exercise: Reduced From Results for Rarely-Prescribed Samples

	Sample Based on Diagnoses' Prescription Rates	
	Prescription Rate $\in [0.03, 0.1)$	Prescription Rate < 0.03
Dependent Variable ($\times$100):	(1)	(2)
Long-Term Use	0.063*	0.065
	(0.035)	(0.088)
Mean Dep. Var. ($\times$100)	4.5	4.0
Opioid-Seeking Behavior	0.039	0.042
	(0.050)	(0.063)
Mean Dep. Var. ($\times$100)	7.3	7.5
Opioid Use Disorder	0.020	0.035
	(0.034)	(0.049)
Mean Dep. Var. ($\times$100)	2.9	4.6
Opioid Overdose Mortality	0.003	−0.004
	(0.007)	(0.010)
Mean Dep. Var. ($\times$100)	0.14	0.21
Residualization FEs?	Yes	Yes
Baseline Controls?	Yes	Yes
Observations	1,897,297	1,449,315

Notes: This table reports the estimated coefficients of a reduced form regression of our main outcomes on physician prescribing leniency for two separate samples based on the patient's index ED (3-digit ICD-9) diagnosis code. Column 1 estimates the regression on conditions that are prescribed 3-10% of the time and column 2 on conditions that are less than 3% of the time. All coefficients are scaled by the difference in baseline leniency between the 90th and 10th lenient physicians (11.6pp) for interpretability. 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. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table 5: 2SLS Results: Opioid Use, Misuse, Addiction, and Overdose Mortality

	Dependent Variable: (×100)			
	Long-Term Use	Opioid-Seeking Behavior	Opioid Use Disorder	Opioid Overdose Mortality
	(1)	(2)	(3)	(4)
Prescribed in ED	1.17***	2.46***	0.335*	0.075**
	(0.202)	(0.314)	(0.160)	(0.034)
Mean Dep. Var. (×100)	5.80	14.79	3.27	0.167
Residualization FE?	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 2SLS effect of an opioid prescription on long-term use (180 days supply in the first 12 months), opioid-seeking behavior in the first year, opioid use disorder within 3 years, and opioid overdose mortality. The endogenous variable of whether the patient is prescribed an opioid in the ED is instrumented with physician leniency in prescribing opioids. Long-term use excludes the initial ED prescription and any prescriptions filled in the first 7 days after the emergency visit. 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.

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:

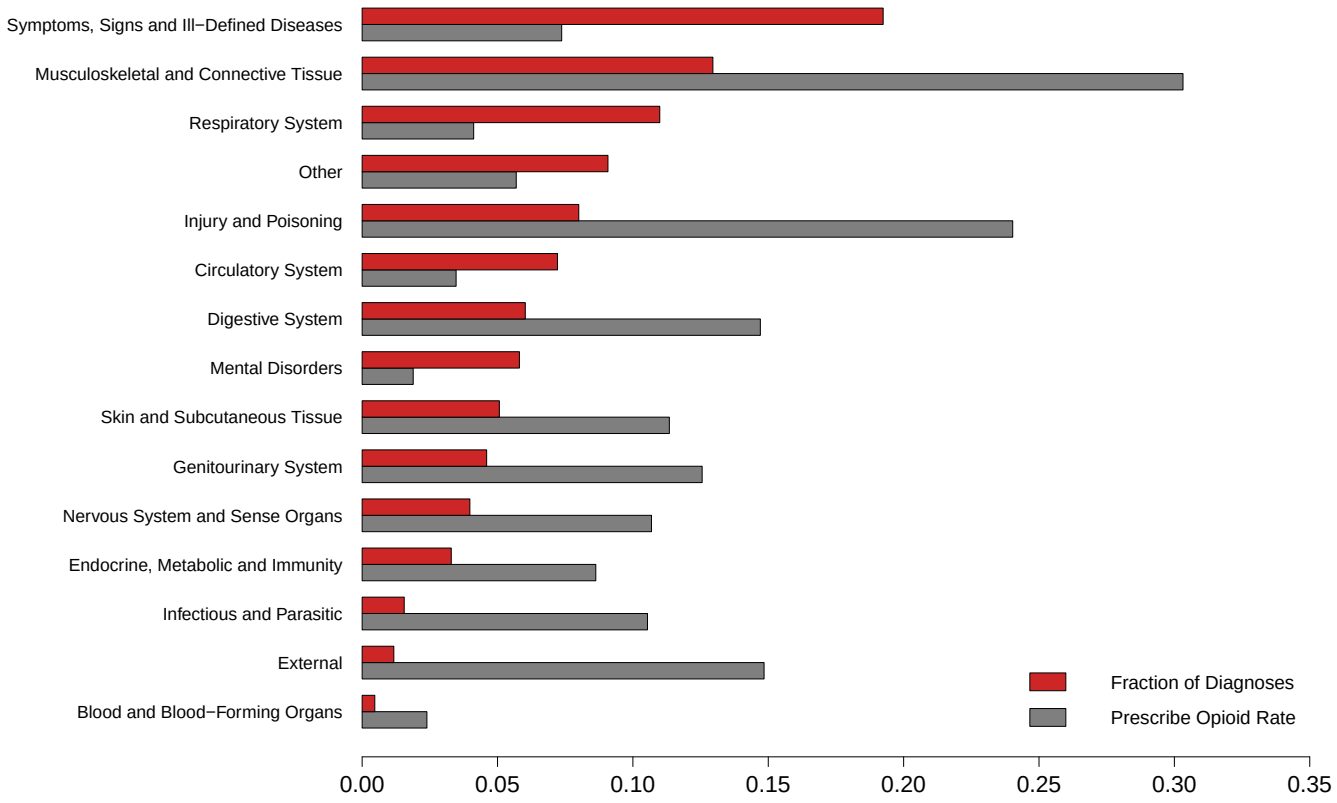
$$Leniency_s^{team} = \frac{1}{\sum_{j \in \mathbb{S}} N_{jy}} \left(\sum_{j \in \mathbb{S}} N_{jy} \times Leniency_{-s,jy}^{phys} \right) \tag{6}$$

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.

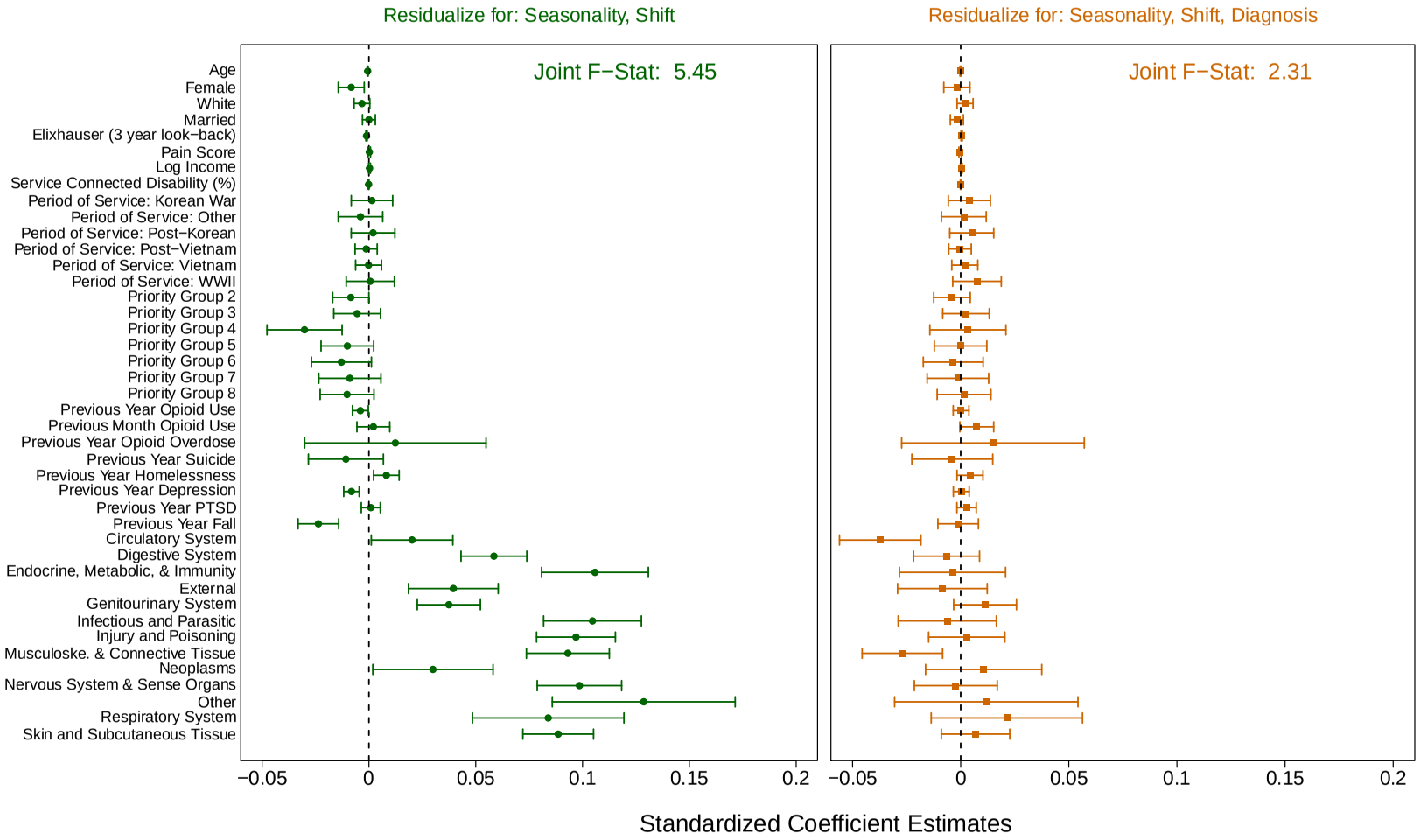
I. Additional Figures and Tables

Figure G.1: Frequent Diagnoses Occurring in Emergency Departments



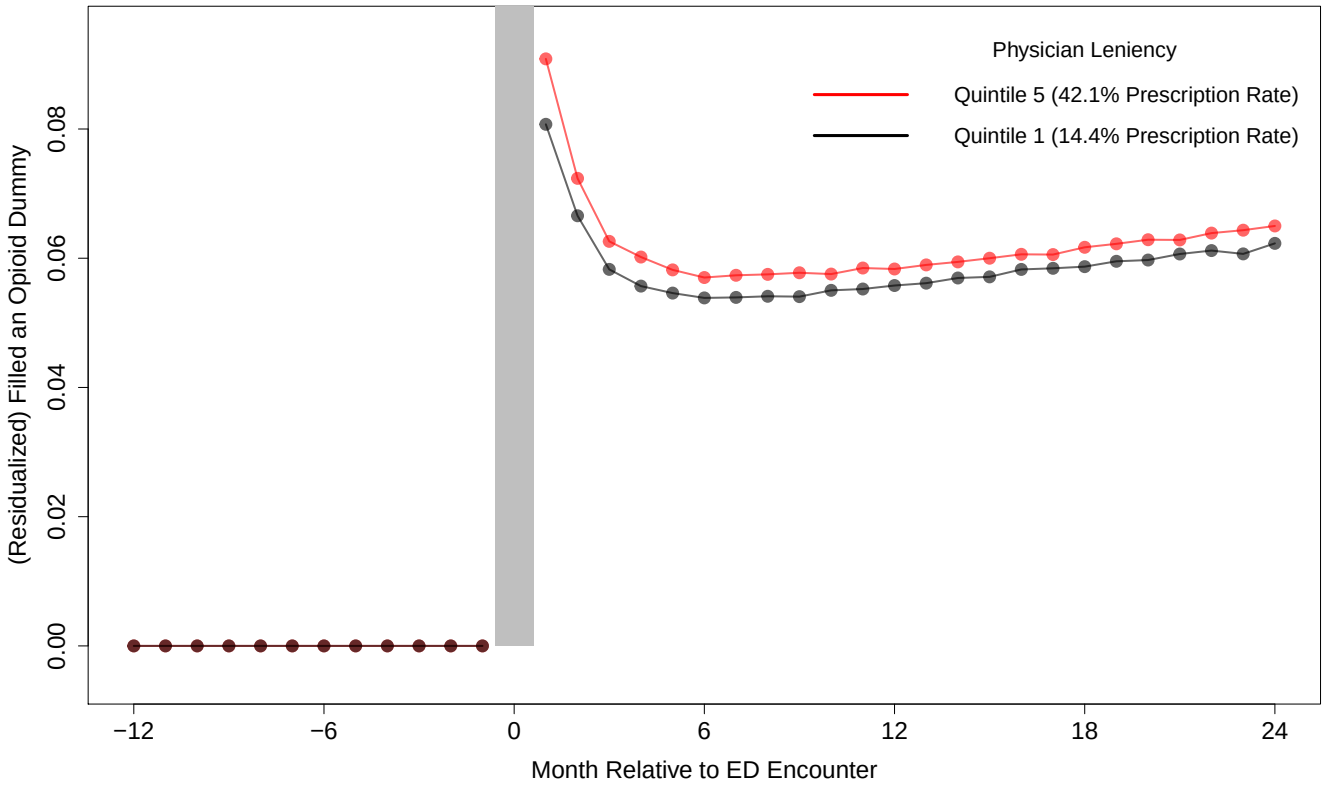
Notes: The 15 most common major diagnosis categories (ICD-9 major chapters) for *all* ED visits and the un-adjusted rate they are prescribed opioids

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



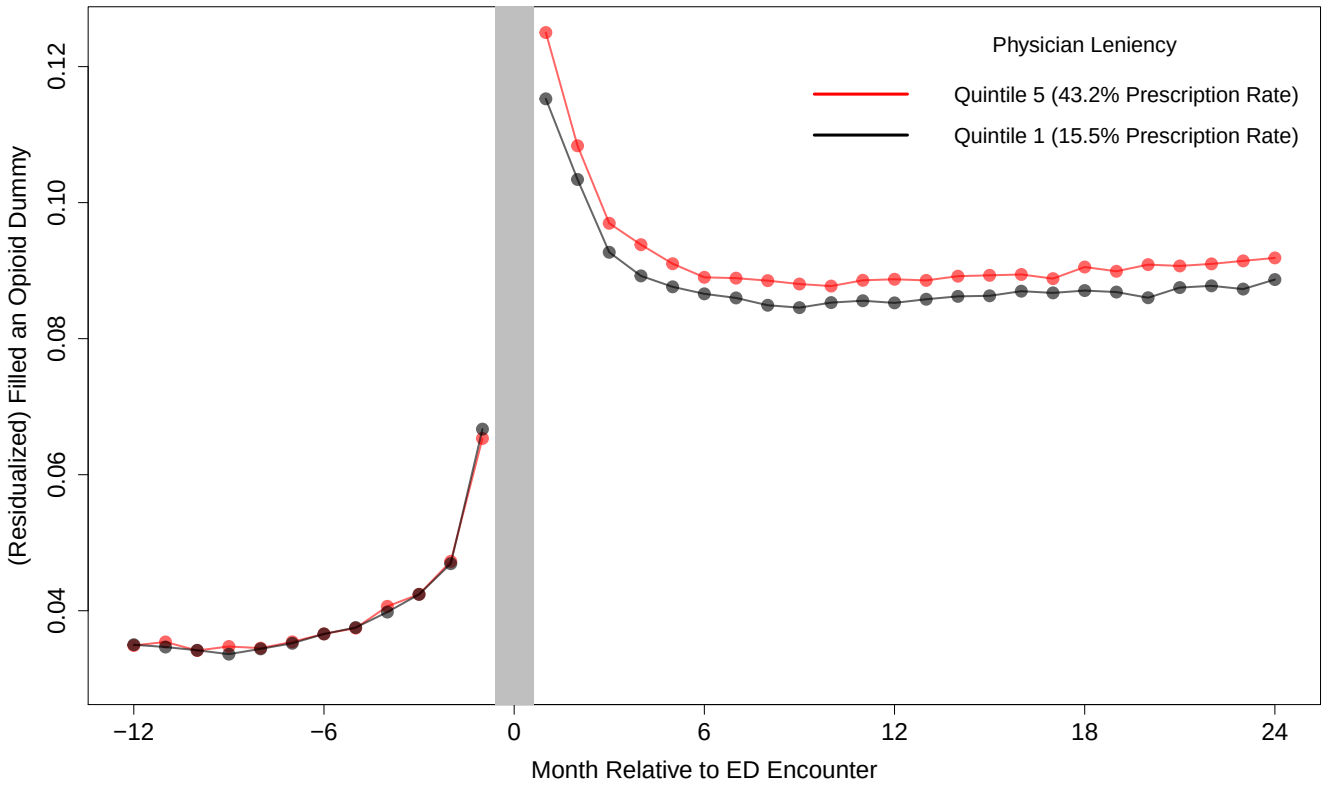
Notes: This figure displays the standard balance test (as seen previously) for the physician leniency measure constructed with different controls in the residualization of Equation 1, effectively, testing for different quasi-random assignment assumptions. In the left panel, the physician leniency instrument 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



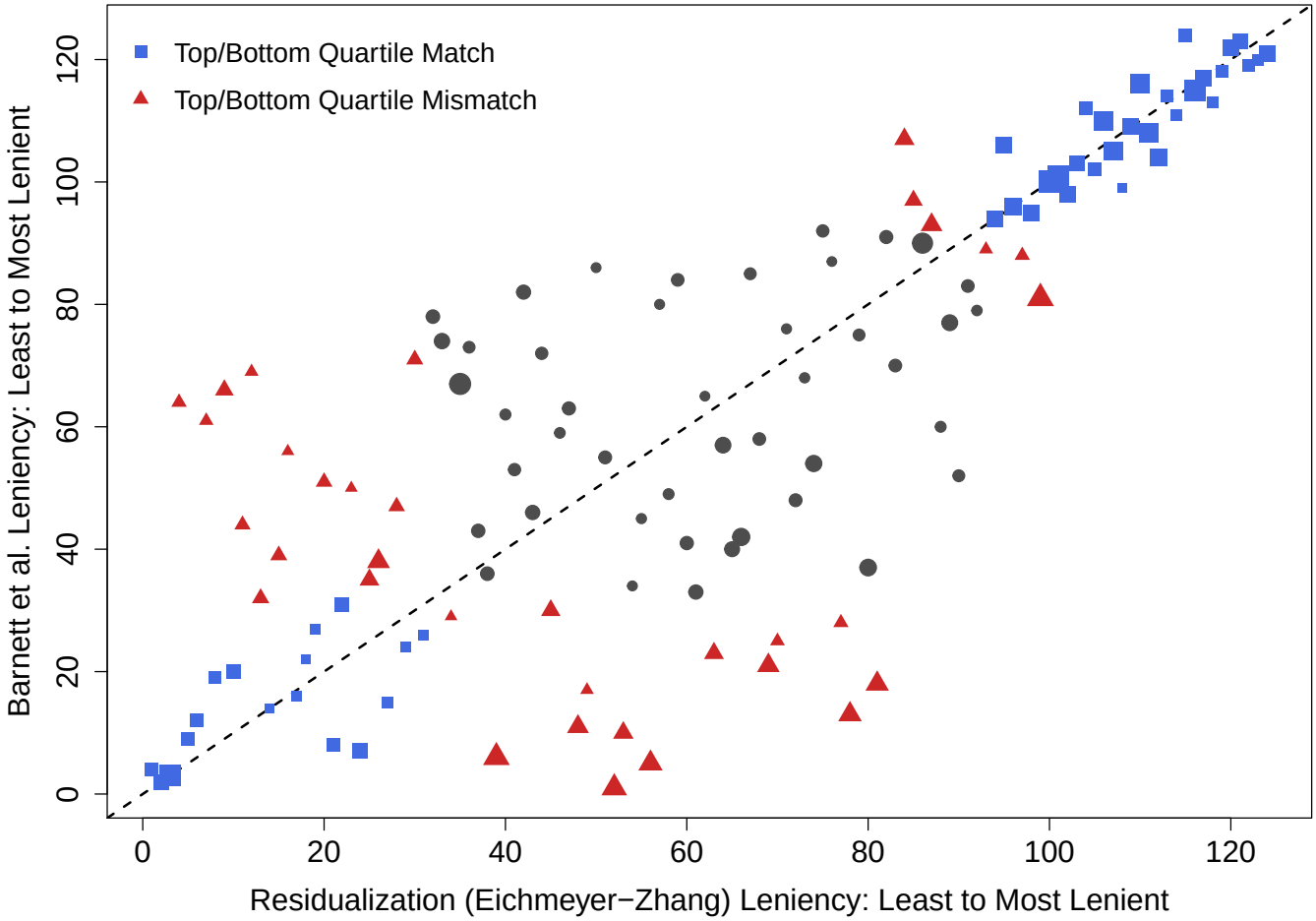
Notes: The reduced form event-study figure corresponding to Figure 3, but for opioid-naïve patients.

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



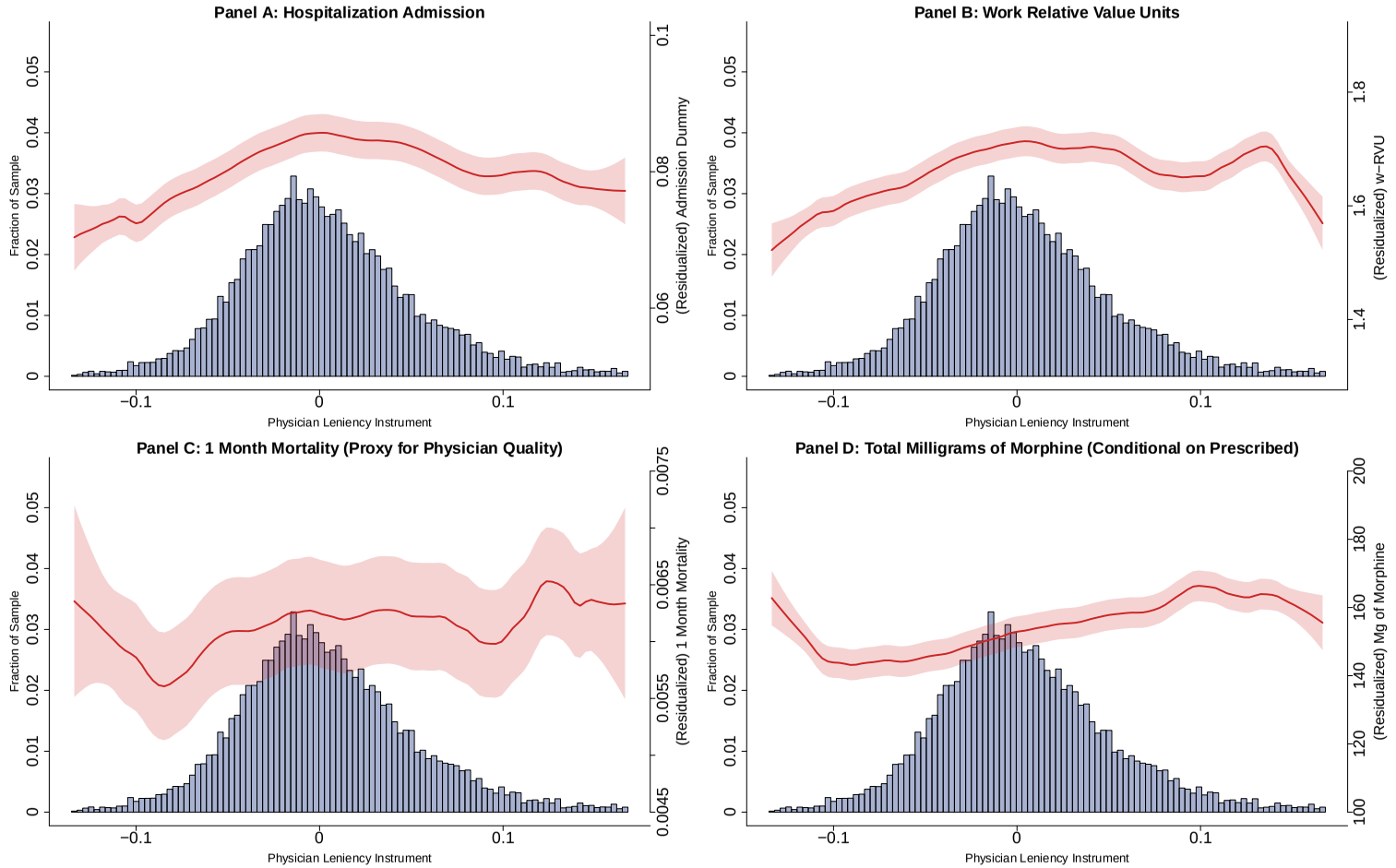
Notes: The reduced form event-study figure corresponding to Figure 3, but for patients who did not visit an ED in the prior year (for any condition), but did utilize VHA outpatient care. Presumably this group are not ED shopping for opioids.

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



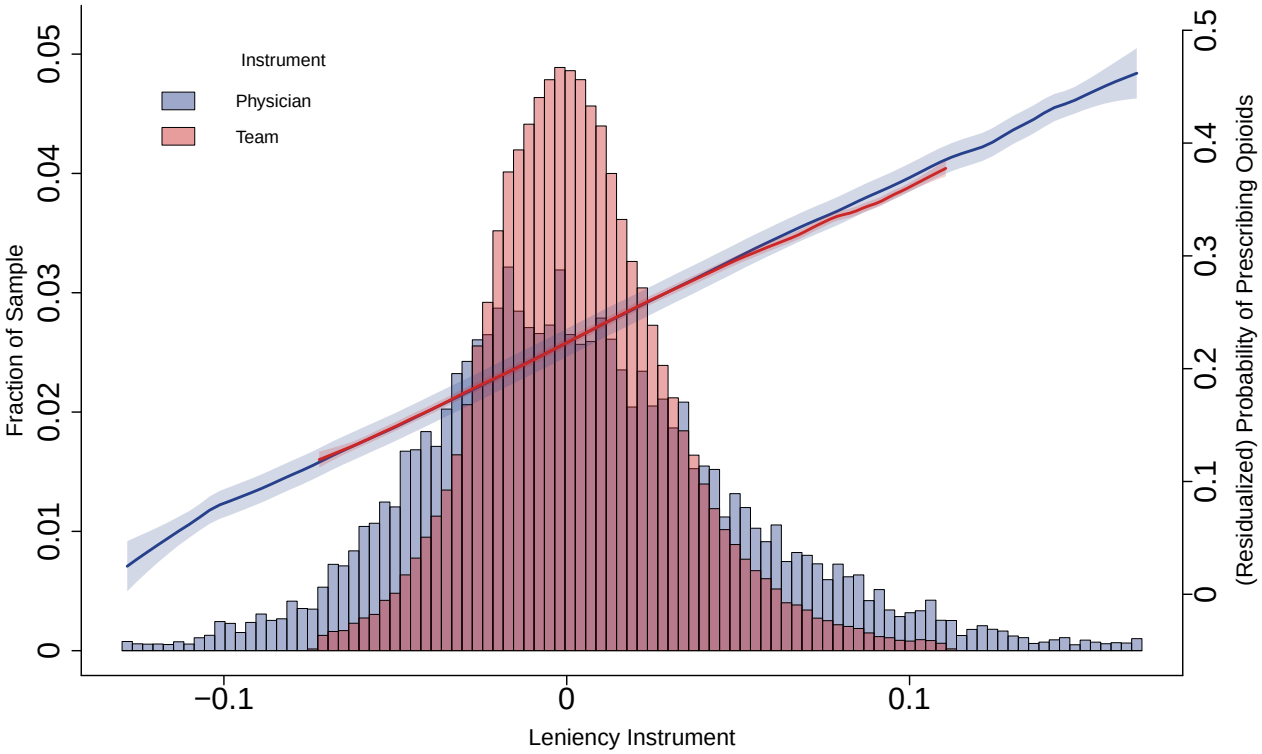
Notes: This graph shows the re-shuffling of physician ranking based on Barnett et al. method and our residualization method for physicians with at least 30 cases in Tampa Veteran Affairs Medical Center, the largest in the country by ED volume in 2012. Each point is a physician and the size of the point 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.