Opioid Prescribing Variation in Emergency Departments: Consequences for Patient Health

Sarah Eichmeyer and Jonathan Zhang*

January 2021

Abstract

Clinical guidelines leave many opioid prescribing decisions to physician judgment. We study the extent to which physician decisions can induce long-term opioid dependence and impact downstream illicit drug use, health and mortality of a patient, by leveraging quasi-random assignment of patients to physicians in emergency departments, in combination with physician variation in their propensity to prescribe opioids. Analyzing the universe of electronic health records for a vulnerable population—veterans—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. We provide evidence that the mechanism is likely operating through the opioid prescription itself. Instrumental variable estimates imply that being treated by an emergency physician who prescribes an opioid (compared to being treated by a physician who does not) leads to a 1.2 percentage point (pp) increase in the probability of long-term prescription opioid use, a 0.34pp increase in development of an opioid use disorder, and a 0.075pp increase in opioid overdose mortality. We also find suggestive evidence of subsequent use of and death from illicit opioids like heroin and fentanyl.

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: Princeton School of Public and International Affairs, Princeton University. jxzhang@princeton.edu.

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 distributed 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, Todd Wagner, Heidi Williams, participants at various Stanford and VHA workshops, 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

Opioids prescribed in medical settings are subject to a clear trade-off between the benefit of immediate pain relieve on the one side, and the cost of longer term dependence, addiction, and its adverse health and mortality consequences on the other side.[1] Policy makers can design optimal opioid-prescribing regulation[2] and physicians can make informed prescribing decisions only when they have accurate information about these cost and benefit parameters—information that is typically gathered via pharmacological randomized controlled trials (RCT).

Yet, those parameters are not well understood, in large part because of high attrition rates from prescription opioid RCTs, a paucity of trials that follow patients beyond a few weeks, a narrow set of patient outcomes, and ethical considerations that prohibit the recruitment of study participants with emotional co-morbidities prevalent in many patients treated with opioids in clinical practice (Busse et al., 2018). As a consequence, the medical community is urgently calling for further research to inform practitioner decision-making in the domain of prescription opioids, in order to provide safe and effective care to patients (Lowenstein et al., 2018; Ross et al., 2011).

In this paper, we quantify the cost side of opioid prescribing. We exploit variation in patient exposure to prescription opioids stemming from within-facility across-physician variation in prescribing leniency in emergency department (ED) settings, and study subsequent opioid dependence, illicit drug use, health and mortality outcomes.

We investigate our research question with the universe of veterans receiving health care from the Veterans Health Administration (VHA), the largest integrated health care system in the United States. The VHA has some unique features that make it an attractive setting for our research design. First, due to its integrative nature, we observe a long history of

[1] Opioid analgesics are substances that bind to mu-opioid receptors. These receptors are primarily located in the brain regions that regulate pain and pleasure, and in the brain stem, which regulates respiration. The former explains the widespread use of opioids for pain relief (and its addicting potential), and the latter explains the high rates of deaths from opioid overdoses, caused by respiratory failure. See Volkow and McLellan (2016) for an overview of the pharmacologic properties and abuse-related risks related to opioids.

[2] This type of regulation has become widespread: by the end of 2018, 32 states had passed laws that impose limits on opioid prescribing for acute pain (Chua et al., 2019; Davis et al., 2019).

a veteran's care along with rich data that are rarely measured in other sources, such as indicators for homelessness and lab test results. Second, our within-hospital physician leniency design requires knowledge of the prescribing history of all physicians in a few hospitals; this is only possible when hospitals and providers operate in a single health care system.

Our empirical strategy utilizes variation in physician leniency in prescribing opioids—within the same hospital and controlling for the time of visit and patient observables— for identification. It proceeds in three steps. First, we evaluate the reduced form effects of ED provider opioid prescribing leniency on downstream drug use, morbidity, and mortality. We focus on long-term prescription opioid use, dependence, and misuse, on measures of physical and mental health, including mortality, and on transitions into illicit drug use. Second, we provide evidence for a causal channel that operates through opioid prescriptions, by comparing the aforementioned reduced form effects to those estimated on a "placebo" sample of patients who present with health conditions whose treatment never or rarely involves opioid prescriptions. Third, we report instrumental variable results estimating the causal impact of the opioid prescription itself on patient outcomes. For this exercise, we employ physician prescribing leniency as an instrumental variable for the initial opioid prescription in the ED, much like the judges instrument in applications related to judges and incarceration.[3]

We argue that relative to existing observational studies set outside of the ED, our instrumental variable research design more closely approximates a RCT that dispenses a single prescription: in the ED, patients have no discretion over choosing providers, and physicians' discretion over choosing patients is typically very 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

[3]This judge stringency instrument is commonly used in the economics of crime literature or any setting where an agent has discretion while making repeated decisions. Specifically, it is a residualized leave-out "jackknife" instrument. See, for example, (Doyle, 2007, entry into foster care) ; (Kling, 2006; Bhuller et al., 2019, 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., 2019, patent examiners), among others.

opioids.[4,5,6] 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.[7]

We establish four main results. First, practice variation as captured by physician opioid prescribing leniency has large and significant consequences for downstream patient opioid use and adverse opioid-related health behaviors and events. Being treated by a provider in the top decile of the leniency distribution, compared to being treated by someone in the bottom decile, increases long-term opioid use by 0.24pp (or 4%). It increases opioid seeking behavior, diagnoses of opioid use disorder, and opioid overdose mortality significantly, by 4%, 2% and 7.5%, respectively. We detect no statistically significant reduced form effects on non-opioid outcomes such as homelessness, suicide, and subsequent preventable hospitalizations.

Second, 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.

Third, we quantify the effect of an ED opioid prescription on long-term opioid use directly. Our IV results reveal that exposure to an opioid prescription via being quasi-randomly assigned a more lenient physician in an emergency department can induce a large increase

[4]Wide practice variation have been documented in a variety of healthcare settings: Barnett et al. (2017, 2019); Chan (2016); Chan et al. (2019); Currie and MacLeod (2017, 2020); Epstein and Nicholson (2009); Fadlon and Parys (2019); Finkelstein et al. (2016); Gowrisankaran et al. (2017); Molitor (2018); Silver (2019); Tu (2017); Tsugawa et al. (2017); Van Parys (2016).

[5]Specifically, this variation in clinical prescribing practice is due to differences in clinicians' interpretation of guidelines asking them to balance risk due to unmanaged pain and risk due to an opioid prescription.

[6]Sowicz et al. (2018) find that 66.4% of the variation in opioid prescribing in VHA EDs (our setting) in 2017 was due to within-facility, across-physician variation. Physician-level variation is even greater in earlier years, consistent with the decline in prescribing rates at the VHA (Sasson et al., 2019).

[7]Beyond identification, focusing on physician prescribing in EDs has broader practical implications. EDs are an important 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.

in long-term prescription opioid use. It increases the probability of long-term use of legal prescription opioids—defined as at least 180 days supply of new opioids filled in the first year following the initial ED visit—by 1.2 percentage points, on a base of 5.8%. This increased probability of opioid use following the initial ED prescription persists for over 24 months and never declines back to pre-ED opioid use levels.

Fourth, we find evidence that the increased long-term use of prescription opioids is accompanied by increases in the development of opioid use disorder, and thus consistent with misuse and dependence rather than appropriate medical care. IV results reveal that an opioid prescription originating from a ED visit is associated with a 0.34pp increase in the probability of developing an opioid use disorder within three years of the ED visit. By adopting standard proxies from the medical literature, we find that veterans begin exhibiting opioid-seeking behavior (overlapping opioid prescriptions, pharmacy shopping, and repeated visits for back pain and headaches increase by 1.9pp, 0.30pp, and 0.55pp respectively). Veterans' self-reported pain score also increases in the first year, despite the fact that opioids should be providing pain relief. Beyond increasing the likelihood of opioid dependence and misuse, we find evidence that the ED prescription has severe adverse effects on downstream health outcomes. Likelihood of recorded opioid overdoses, and accidental falls (an adverse outcome associated with opioid intoxication) increase, and deaths from opioid overdose increase 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 a (statistically insignificant) increase in positive heroin/fentanyl drug screens, hepatitis C diagnoses, and heroin/fentanyl overdose mortality.[8,9] Thus, gains in pain management would have to be very large to outweigh the significantly increased risk of severe adverse health and mortality outcomes that arises from prescription

[8]Based on data from patient assessment questionnaires conducted by clinicians, veterans also begin using cocaine more, while other substance use (e.g., marijuana, alcohol, sedatives, etc.) remains unchanged. The increase in cocaine use could be due to increased heroin use: cocaine is often combined with heroin, a combination called "speedball", in order to enhance the effects of each drug (Ives and Ghelani, 2006; Leri et al., 2003, 2005).

[9]Heroin and synthetic opioids (e.g., fentanyl and tramadol) account for 30-40% of the overall opioid mortality effect caused by exposure to a prescription opioid.

opioid exposure among the patients in our sample.

There is a large and growing literature on the health and economic consequences of opioid use; see Maclean et al. (2020) for a thorough review of the economics literature. The most closely related paper is Barnett, Olenski, and Jena (2017)—henceforth BOJ—who use data on a 20% random sample of Medicare Part D beneficiaries from 2008-2011 and find that patients treated by "high-intensity" prescribers are 30% more likely to be long-term users of legal prescription opioids compared to patients treated by "low-intensity" prescribers. In concurrent work, Barnett et al. (2019) replicate their 2017 study with VHA data 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, health outcomes, and 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.[10] Utilizing an instrumental variable strategy similar to ours, de Vaan and Stuart (2019) study within-household diffusion of opioids and find that spouses of patients who are prescribed opioids are more likely to also fill opioid prescriptions. By studying the family household, the authors focus on how the patient demand channel accelerated the opioid epidemic. 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. (2018). Both leverage patients who move and exploit geographic variation in opioid prescribing rates. The former is set in Denmark and find that a 10pp increase in primary care prescribing rates lead to a 4.5pp increase in opioid use and a significant decrease in labor income and employment. The latter find that a 20pp higher prescribing county increases the mover's propensity to abuse opioids by 6%. 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. We build on this work by exploiting

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

practice variation in ED settings, thereby moving closer to identifying the causal impact of opioid prescriptions on patient outcomes, since the short-lived nature of ED interactions, and quasi-random patient-physician assignment in this setting shuts down other (but not all) potential channels besides opioid prescribing that are present in other settings, correlate with opioid supply variation, and impact patient outcomes.

It is important to note that this paper studies the impact of an opioid prescription through a prescribing decision requiring clinical judgment rather than through specific VHA policies or differences in adherence to clinical practice guidelines. 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.[11] Opioid prescribing and care delivered at the VHA, and studied here, were within clinical guidelines during this period. Furthermore, it is not substandard care but rather practice variance *within practice norms* that provides us with the variation for our research design and the findings in this paper. Notably, this design strengthens the utility of the findings, providing data on outcomes within the range of clinical cases in which prescribing decisions are not straight-forward and clinician judgment varies. This may help to fill gaps in clinical practice recommendations.

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. The last section concludes.

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

[11] See Sandbrink et al. (2020) for a review of VHA opioid policies and strategies.

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 (CDW), which includes electronic health records (EHR), enrollment tables, and other health records. The EHR data includes standard inpatient, outpatient, and pharmacy data with variables for key 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, for example, federal supportive housing programs for veterans and the residential treatment programs for substance use disorders. ED records begin in 2006; however, other outpatient records are available starting in 1999.

Through the VHA health records, we only observe care that occurs at VHA facilities. However, 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 supplement our CDW data with VA/CMS data. These data include Medicare claims from 2011-2016 and Medicaid claims 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. In addition, the VHA reimburses community care when they cannot provide the services themselves (e.g., emergency care, nursing homes, inpatient hospice care). We also observe all medical and pharmaceutical claims that veterans file to get reimbursed. CMS and community records provide us with a more complete view of veterans' health events beyond care in the VHA.

Our last data sources provide us with date and cause of death for all veterans from VHA Vital Status files and CDC National Death Index (NDI) Plus files, respectively. The NDI Plus files include detailed information about cause of death, allowing us to differentiate between opioid drug overdose deaths from deaths by heart failure, 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, place of 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. ED encounters include both visits leading to inpatient hospital admissions and same-day discharge visits.

We make several restrictions to the sample of all emergency visits to improve power. First, we drop encounters with broad diagnosis categories that are prescribed an opioid less than 10% of the time, because 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.[12] 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 4.4 we present our main outcomes without excluding such conditions. Second, we exclude visits involving patients who are already very heavy prescription opioid users prior to the ED visit. Specifically, visits by patients for whom we observe opioid prescriptions totalling to over 3,150 milligrams of morphine equivalent in

[12]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.

the prior year are excluded (this is equivalent to 5 weeks of 90 daily MME, approximately the 84th percentile of prior year opioid usage). Including these visits would reduce power because an additional opioid prescription is unlikely to have an effect on the heaviest prior users. 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 take the first 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, $Prescribed_i$, is an indicator for whether patient i is prescribed—as opposed to administered—*any* prescription opioid at their ED encounter (see Appendix B for details on how prescriptions are located and assigned). Note that the patient could potentially not fill the prescription.[13] Below we detail our main outcomes, all of which are aggregated at the year-level, relative to the date of the ED visit.

Opioid use and opioid drug screens

To study long-term prescription opioid use, we follow the opioids literature and define long-term use as 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.[14] Since ED prescriptions

[13] We observe that approximately 3% of all VHA opioid prescriptions written by clinicians go unfilled.
[14] See Barnett et al. (2017, 2019); Jena et al. (2016); Dunn et al. (2010).

are non-refillable, this means a long-term user needs to fill 180 days supply of opioids from other physicians. We also study positive opioid drug screens (urine or blood) as a proxy for opioid use not captured by prescriptions. This variable is constructed unconditional on screening—patients who do not receive a screen receive a value of zero—because it has a more natural interpretation than conditional on screening. One limitation is that ED assignment may lead to differential screening rates.

Opioid-seeking behavior and pain scores

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 opioid literature and are strong predictors of overdose (e.g., Yang et al. (2015) and Finkelstein et al. (2018)) and the last proxy is based on conversations with VA physicians. We also construct an average self-reported pain scores (on a 0-10 scale) across all outpatient encounters. Since they are self-reported, the score can be exaggerated by the patient to obtain opioid prescriptions.

Opioid use disorder, overdose, and mortality

We study two clinical diagnoses captured by ICD codes that are directly related to opioid misuse and abuse: opioid use disorder (OUD) and opioid overdose. OUD is often referred to as opioid dependence and abuse.[15] Disjoint from OUDs, opioid overdose events are episodes when opioid receptors in the brain are blocked, potentially leading to unconsciousness, coma, and death. We define an opioid overdose —the vast majority of which are non-fatal with timely treatment—following CDC definitions and also investigate the *type* of opioid that caused the death: heroin; synthetic (excluding methadone; e.g., fentanyl); and natural and

[15]Diagnostic and Statistical Manual of Mental Disorders (DSM-V) defines OUDs as a "problematic pattern of opioid use leading to clinically significant impairment or distress."

semi-synthetic opioids (e.g., oxycodone).[16] Virtually all opioid overdose deaths are observed, regardless of place of death or type of overdose. However, OUDs and opioid overdoses that are not diagnosed by a clinician, or are treated for overdoses outside of the hospital are not captured in our data (e.g., naloxone rescue administered outside the VHA). It is likely that the fraction of such OUD and opioid overdose episodes not captured by our measures is high. For example, only 5.3% of all veterans for whom we observe an opioid overdose death had a recorded opioid overdose in the three years leading up to their death.

Health and non-opioid outcomes

We study two different mental health outcomes: diagnosed attempted suicide and self-harm (excluding suicide ideation), and major depressive disorder. Additionally, we use depression assessment results from mental health questionnaires, and we code response scores above the recommended thresholds as scoring positive. We also construct indicators for accidental falls as a proxy for impulsivity or sedation and are predictive of opioid overdose risk (Oliva et al., 2017). Finally, we proxy for homelessness from a variety of sources including diagnosis codes for lack of housing and outreach to VA homeless and/or shelter programs.[17]

Illicit drug use and other substance use

In addition to heroin and synthetic opioid deaths, we investigate illicit drug use via two proxies. The first 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.[18] Therefore, we code any test that mentions heroin, fentanyl or 6-MAM (the specific metabolite unique to heroin and

[16] ICD-9 codes for OUD: 304.0x, 304.5x, and 304.7x; opioid overdoses: 965.x, E850.0-E850.2, E935.0-E935.2, and E980.0. 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 classify deaths as: heroin (T40.1), synthetic opioids (T40.4), and natural and semi-synthetic opioids (T40.2 only).

[17] We borrow definitions from Peterson et al. (2015), which uses diagnosis codes, VA homeless service codes (HUD-VA, telephone, and community outreach for homeless veterans), and homeless inpatient treatment.

[18] 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).

fentanyl) in the order form, regardless of test result, as an intent to screen for heroin/fentanyl. Our second proxy for illicit drug use 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 (2016b) 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.[19] For both proxies, 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. Finally, we study results from a drug use questionnaire for 31,000 veterans on whether they have used marijuana, sedatives, cocaine/crack, and other stimulants in the past 30 days, and a screen for potentially hazardous alcohol use.

2.4 Summary Statistics

Table 1 describes relevant statistics of the opioid prescription, ED encounter, veteran, and their medical morbidities for our baseline sample. After excluding conditions that are rarely prescribed, in 26.1% of our baseline ED encounters, physicians prescribe an opioid.[20] The average ED prescription is for 8.2 days with 21 MME per day.[21] Figure 1 displays the empirical CDF of days supply and daily MME by year. While prescription length has declined over time, daily MME has remained unchanged. 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 2 shows the top 15 most common diagnosis categories and the share that are prescribed opioids in the ED for the baseline

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

[20] Across all ED visits, 12% are prescribed opioids, comparable to the 14.3% national average in 2011 (NCHS, 2016).

[21] Patients who are not being prescribed opioids are often prescribed nonsteroidal anti-inflammatory drugs (NSAID). In Table F.1 displays summary statistics of the average NSAID for patients who are not prescribed an opioid. There is no statistically significant relationship between opioid prescribing leniency and likelihood of prescribing NSAIDs.

sample.[22]

The average veteran is a middle-aged, white male (average age 55; 90% are male, and 69% are white), which also happens to be 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,[23] 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.[24]

3. Empirical Strategy

Our empirical strategy closely follows the literature that exploits quasi-random assignment of agents to cases—commonly known as the "judges design." Papers in this literature estimate the causal effect of different types of sentencing outcomes on defendants' outcomes, exploiting variation in the leniency of the assigned judge. In the case of emergency departments, the strategy rests on quasi-random assignment of patients to physicians who differ in their opioid prescribing rates.

The ED is a setting where patients have very little choice regarding which provider they see, alleviating much of the physician shopping and patient selection issues in traditional

[22]See Figure F.3 for the most common major diagnosis categories for all emergency visits.

[23]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.

[24]The 2018 National Survey on Drug Use and Health (NSDUH) supports this claim: 65.7% of survey respondents with opioid substance use disorder had some mental illness in the prior year, 28.4% had major depression and 3.2% attempted suicide.

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. Often, 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.

3.1 Leniency Instrument Construction

Our physician leniency instrument is constructed as a year-varying measure based on the following regression equation (following Doyle et al. (2015) and Dobbie et al. (2018)):

$$Prescribed_{ikt} = \alpha_0 + \alpha_{hym} + \alpha_{hdt} + \alpha_{diagnosis} + \alpha_{agebins} + \gamma W_{ik} + \epsilon_{ikt}. \quad (1)$$

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}.[25] Diagnosis fixed effects,[26] $\alpha_{diagnosis}$, are included to account for physician specialization with respect to diagnosis conditions and to increase precision.[27] As mentioned, these three set of controls are what is required for our quasi-random assignment assumption. Next, to improve statistical precision in our leniency measure, we

[25]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.

[26]We truncate ICD-9 diagnosis code to its three-digit value; all ICD-10 codes are crosswalked to ICD-9.

[27]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 9.

include controls for five-year age bin, $\alpha_{agebins}$, 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}.

Using *all* ED encounters—not just the baseline sample—for patient i's, encounter k, treated by physician j in year y, the instrumental variable:

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

where $\hat{e}_{ik'} = \hat{Prescribed}_{ik'} - Prescribed_{ik'}$, 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.[28] This leave-out mean eliminates the mechanical bias that stems from patient i's own case entering into the instrument. This residualized 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 C summarizes which physicians are more likely to be lenient prescribers by our measure. The appendix also correlates prescribing leniency with other dimensions that physicians can vary. In subsection 4.4, 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.

As in Barnett et al. (2017), VA ED physicians exhibit wide systemic variation in their propensity to prescribe opioids. Figure 3 graphs the histogram of the instrumental variable 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 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 roughly 17pp increase in the likelihood of being

[28] This leave-out mean measure is algebraically equivalent to a physician fixed effect estimated in separate leave-out regressions (Dobbie et al., 2018).

prescribed an opioid in the ED. The first stage F-statistic is 25 when all controls and fixed effects are included. The first stage coefficient is greater than 1 because *all* emergency visits are used to construct the leniency instrument, however, the first stage is calculated using the baseline sample, which excludes the rarely prescribed diagnoses.

3.2 Empirical Specification

To study the effects of being prescribed an opioid in the ED on a outcome Y, we estimate the following model using our baseline sample of only first visits:

$$Y_i = \beta_1 Prescribed_i + \theta X_i + \epsilon_i \tag{3}$$

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), gender, race, and the set of fixed effects from Equation 1. The $Prescribed_i$ 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. To circumvent this, 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.

3.3 Identifying Assumptions

Under several identifying assumptions, the 2SLS estimates the local average treatment effect (LATE) of being prescribed an opioid in the ED. First, we require that assignment of patients to ED physicians be random, conditional on seasonality, shift, and diagnosis ("conditional independence"). One can test for whether lenient physicians are routinely assigned patients with particular characteristics. Figure 5 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. For example, a unit increase in the patient's self-reported pain score is associated with a 5.9pp increase in the probability of being prescribed. 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 (e.g., pain score is now negatively correlated with leniency).

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 F.4 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.[29] 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 4.4.

To move from interpreting our estimate as the causal effect of being treated by a *lenient physician* (subsection 4.1) to one of the causal effect of being prescribed *an opioid* (subsection 4.3), we require the exclusion restriction to be satisfied. Specifically, the instrument must influence the outcome of interest only through its effect on initial opioid consumption. This is perhaps our strongest assumption and is at its core, untestable. However, there are reasons to believe its impact may be small and 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

[29]Looking ahead, the various levels of residualization have some effects on our baseline 2SLS causal estimates. Table F.2 presents the effects of an opioid prescription on our main outcomes with instrumental variables that are constructed with varying levels of residualization. The main estimates are attenuated 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.

medium-term outcomes is limited. 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 4.4.

Finally, the monotonicity assumption is necessary for interpreting the coefficient estimates obtained from the IV approach as 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). To test this, we run the first stage regression separately by gender, age, education, and diagnosis groups, among others. 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., 2019).[30] Table F.5 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 4.4.

4. Results

4.1 Reduced Form Results

Veterans who are treated by lenient opioid prescribing physicians in EDs are more likely to begin using prescription opioids long-term. Panel A of Figure 4 plots the reduced form event-study analysis of new monthly opioid prescriptions filled for patients who see lenient

[30] 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.

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 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 approaching up to their ED encounter.[31] The lenient and strict groups are on parallel paths of opioid use prior to the visit. However, after the visit, there is an increase in the probability of filling an opioid prescription for patients treated in the ED by lenient physicians relative to their strict counterparts.

This level shift in opioid use probability following the ED is on the order of approximately 0.25 percentage points (average difference between the two groups in months 2-24) and can be attributed to a 27.4pp (42.3% - 14.9%) difference in opioid prescription rates between the lenient and strict prescribers.

We report the reduced form results for opioid use, morbidity, and mortality in table form in Table 3, column 1 for our baseline sample. Estimated coefficients are scaled by the difference between average leniency between the top and bottom deciles for interpretability. We find statistically significant and large effects of assignment to a leniently prescribing physician on four out of the five opioid use and misuse related outcomes - highlighting the substantial role ED physicians can play in putting patients on a path of continued opioid use. We do not detect statistically significant effects of leniency on non-opioid outcomes (homelessness, suicide, preventable hospitalizations,[32] and heart disease mortality).

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

[31] This rise in opioid use prior to the ED visit and the findings on subsequent use cannot be explained by patients shopping for opioids. In Figure F.5 and Figure F.6, 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. These two groups of patients are unlikely to be ED shopping for opioids.

[32] Preventable hospitalizations are proxied by ambulatory care sensitive conditions (ACSC): hospitalizations due to conditions such as diabetes, asthma, hypertension and pneumonia that can largely be avoided with timely, effective, and continued primary care. We exclude any immediate hospitalizations from the initial ED visit.

being the underlying mechanism behind the effects. However, physicians could differ in 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 Investigating Mechanisms Behind the Reduced Form Effects

To test whether the reduced form effects observed in Section 4.1 are due to differences in opioid prescription rates across providers, we start with studying reduced form effects among patients with diagnosis conditions that are never prescribed opioids, as a "placebo 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 reduced form effect of leniency on long-term use and on health outcomes if and 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 categorize all emergency visits into four quartiles based on the unconditional probability of being prescribed an opioid for their diagnosis. The bins are 0-3%, 3-10%, 10-20%, and 20%+ (recall that our baseline sample only includes the latter two groups). The diagnoses that are never prescribed (0-3%) serve as a "placebo" test. We then run a reduced form regression of an outcome of interest on physician prescribing leniency, separately for each of the four sub-samples. The results of this exercise are displayed in in columns 2 and 3 of Table 3. Panel (A) of Table 3 shows that while there may be a statistically significant and large association between physician leniency and a given opioid-related outcome in our baseline sample, this association 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 2) or rarely (column 3) prescribed opioids. In Panel B of Table 3, columns 2-3, we perform the placebo exercise

for our main non-opioid outcomes. We find a sizeable and statistically significant effect of being assigned to a high prescribing physician on preventable hospitalizations 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%.[33]

The failure of our placebo check for the preventable hospitalization outcome suggests that there may be dimensions of care correlated with opioid prescribing tendency that vary across ED physicians and may impact 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 taking account of 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 "propensity" along all three of the above-mentioned dimensions, and including those proxies as controls in the IV analysis (details are provided in Section 4.4).

4.3 Instrumental Variable Results

Long-term prescription opioid use

To estimate the causal effect of the initial ED opioid, we run separate monthly regressions of Equation 3 with opioid use dummy as the outcome, instrumenting for $Prescribed_i$ with

[33]Note that, in contrast, the associated coefficient estimate for our baseline sample is much smaller and statistically indistinguishable from zero. This difference in effects across our baseline and placebo samples could be explained by increased subsequent health care visits for opioid seeking purposes among patients assigned to high prescribing providers in the baseline sample relative to the placebo samples. More encounters with medical providers may lead to better treatment of conditions that could otherwise lead to preventable hospitalizations, thereby off-setting initial differences in care across high- and low-leniency ED physicians that would result in different rates of preventable hospitalizations.

$Leniency_i$. Panel B of Figure 4 plots the 2SLS analog of Panel A. The causal effect of an initial ED opioid on opioid use in subsequent months is highest for the first month at 5%, and settles in at around 1.5% at around the six-month period, over an average opioid use rate of 9.5%.

In Table 4, the naïve OLS estimates that on average, an opioid prescription in the ED is associated with a 2.85 pp increase in the probability of long-term use (180 days supply in the first year), on a base of 5.8%. This number is relatively robust to controlling for baseline characteristics. However, with the IV model, the effect is more than halved to 1.17 pp in our preferred specification, a 20% increase on the overall average long-term use rate. 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. The remaining rows of Table 4 present the total milligrams of morphine equivalent filled by the patient. Reassuringly, there are no large differences in pre-period total morphine (column 4 is statistically significant albeit a small effect size relative to the base rate and the dynamic reduced form plot shows no movement along the extensive margin). The effect of an opioid prescription in the ED increases three-year morphine equivalent filled by 468 mg, excluding the initial ED prescription. Including the ED prescription, this is equivalent to over 40 tablets of Oxycontin 10 mg—approximately a 25% increase in total observed mg of morphine filled in the two-year period after an average emergency visit (20% increase conditional on being prescribed). The analyses and research design control for diagnosis condition; therefore, the increase in subsequent opioid use is because a veteran saw a more lenient prescriber at the ED, and not driven by differences in conditions. In subsection 4.4 we find that for our sample of non-heavy prior opioid users, the differences in outcomes stemming from prescribing differences along the intensive margin (i.e., drug dosage) are minimal relative to the extensive margin.

The bottom panel of Table 4 reports the regression coefficient for positive urine drug screen for opioids (veterans who are not tested receive a value of zero). Veterans who are prescribed opioids in the ED are 2.0pp (24%) more likely to test positive for opioids within

three years following their ED visit.

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. Laird and Nielsen (2016) and Finkelstein et al. (2018) also leverage variation in supply-side factors and find that when patients move to regions with a 10pp higher prescribing rate, subsequent prescription opioid use increases by 4.5pp and rates of opioid abuse increase by 3pp, respectively. In ongoing work (Eichmeyer and Zhang, 2020), we find that the patients quasi-randomly assigned to a 90th percentile primary care opioid-prescriber is associated with roughly a 0.72pp increase in long-term use (32% on the base), relative to the 0.24pp (4%) we find in the ED setting. The large differences in effect sizes further highlight the advantages in focusing on emergency departments, a setting where, while not perfect, it isolates the impact of prescription opioids more precisely.

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. There are few reasons for a single (non-cancer, non-end-of-life) ED encounter to cause a sustained increase in the probability of opioid use for 24 months, especially if the increase is because the patient was treated by a different physician.

Opioid-seeking behavior and opioid misuse

An opioid prescription stemming from assignment to different providers in the ED can lead to long-term prescription opioid use. However, it is unclear whether this long-term use is through appropriate medical care or active opioid seeking, suggestive of inappropriate use and patient misuse. We study three proxies for opioid-seeking behavior and opioid misuse along with average self-reported pain score. The regression outcomes in Table 5 indicate that an opioid 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%. In this regression we exclude patients for whom the ED diagnoses were back pain and headaches to avoid the concern that increases in this proxy could be due to their condition being treated poorly in the ED. Lastly, veterans' average self-reported pain score increases by 0.08 on a base of 2.81 (0-10 point scale). While there is no significant difference in any of the four measures in the year prior to the emergency encounter, the coefficients are positive. This raises a potential concern that some of our outcomes may be driven by some violations in quasi-random assignment. In subsection 4.4 we utilize a team-based research design and arrive at qualitatively similar estimates but with less precision.

The unanimous increase across all four measures provides evidence that an ED prescription increases not only long-term opioid use, but also unnecessary prescriptions, inappropriate opioid misuse, 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, veterans exhibit opioid-seeking behavior that is consistent with patients having excess demand for opioids. The findings on back pain and headaches, and pain scores especially point in this direction as they go against standard medical intuition: Being prescribed opioids should decrease experienced pain, rather than increase it.

Adverse opioid outcomes: OUD, opioid overdose, and mortality

Thus far, we've shown that some of the long-term opioid use induced by ED prescribers is inconsistent with appropriate medical care and can be explained by opioid-seeking behavior, suggesting some veterans are misusing opioids. Next, we investigate whether this misuse can lead to severe adverse outcomes for the veteran.

Table 6 reports the OLS and 2SLS regression output for OUDs, opioid overdoses, and opioid overdose mortality. For both OUD and opioid overdoses, there is no significant difference in the prior year for patients who are prescribed in the ED because of the leniency

of their physician. However, 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. One of the most severe outcomes associated with patients with OUDs is an opioid overdose. An ED prescription increases the three-year probability of overdosing by 0.039pp, on a base of 0.23%. This coefficient is not statistically significant; as previously mentioned non-fatal overdoses that are not treated at VHA facilities are not captured in the data.

Fortunately, the vast majority of opioid overdoses are nonfatal. By far the worst outcome for an opioid abuser is an opioid overdose death. The next row presents the results of a regression with opioid overdose mortality as the dependent variable. There is a positive effect effect of 0.075pp on opioid overdose mortality caused by an ED opioid prescription, on a base of 0.167%.[34] Conditional on a fatal opioid overdose, the median time from the ED prescription to death in our sample is three years and five months. The final two rows distinguishes between the type of opioid in opioid deaths, specifically, heroin or synthetic opioids, and natural/semi-synthetic opioids. Although statistical power suffers with such mortality splits, some overall patterns arise. An (non-heroin, non-fentanyl) opioid prescription may cause illicit opioid deaths. In fact, a substantial fraction (30-40%) of the total mortality effect is accounted by heroin and synthetic opioid related mortality.

Across all three outcomes, the OLS estimates are biased downward relative to the 2SLS estimates and the 2SLS estimates represent a large effect size over baseline means. This is consistent with the negative correlation between prior mental health and homelessness and being prescribed in Figure 5. In subsection 4.5, we find that LATE compliers affected by the instrument are veterans with particularly high overdose mortality risk that physicians take different approaches to treating. This means that physicians are good at recognizing patients who are at obvious risk of opioid overdoses, and the marginal patients that physicians exhibit varying clinical (prescribing) judgment on, are at significantly higher risk to be put on a path

[34] 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.

of dependence. In subsection 4.6, we benchmark these magnitudes to the greater epidemic, and implied opioid benefits that would be required to offset our estimated costs.

Health and non-opioid outcomes

In addition to the chain of opioid-related outcomes, the emergency opioid prescription could have downstream effects on a variety of health outcomes. For instance, Scherrer et al. (2016) find that long-term use of opioids can lead to new onset of depression. Along with depression, we also study attempted suicide and self-harm, accidental falls, and homelessness.

Table 7 reports the regression output. Consistent with the other prior-year characteristics, the veterans with prior health morbidities are less likely to be prescribed opioids. In general, veterans who receive opioids were in better health prior to the ED visit. Opioids increase the probability of accidental falls—a proxy for sedation—within three years by 0.5pp. The preferred 2SLS coefficient for depression is positive but not significant. We are able to rule out effects on depression diagnosis larger than 1.35pp, on a base of 35.5%, at the 95% level. The coefficients for attempted suicide and homeless episodes are not statistically significant from zero, or the prior-year effect. The coefficient for homelessness in the prior year is positively and weakly significant, presumably due to sampling variation given the balance of other prior morbidities and patient characteristics.

Transition into other substances and illicit drugs

The majority of veterans who die from an opioid overdose do not fill a prescription for opioids in the prior year (Lin et al., 2019) and 80% of heroin users began by abusing prescription opioids (Muhuri et al., 2013). Alpert et al. (2019) and Evans et al. (2019) find that heroin-related overdose deaths increased after Oxycontin's reformation to become abuse-deterrent. These facts, along with our estimates on heroin and synthetic opioid mortality suggest that some veterans are transitioning into illicit opioids/drugs.

By analyzing the Brief Addiction Monitor questionnaire responses of roughly 31,000 veterans in the three years after the initial ED encounter, we can break down the type

of drug use stemming from the initial encounter (Table 8). Self reported illicit/illegal or prescription drug abuse increases by 0.16pp, again roughly 20%. There is evidence of increased cocaine/crack use, suggesting that prescription opioids may be a gateway for other illicit drugs or that veterans are using mixed/laced illicit substances. Interestingly, use of marijuana or hazardous alcohol consumption do not increase, perhaps because these substances are more common and less stigmatized.

To research the effects on injection drug use, we also report the output of regressions of two different proxies: i) physician intent to screen for heroin/fentanyl and ii) positive hepatitis C diagnosis, on an indicator for any ED prescription in the last two rows of Table 8. The 2SLS coefficients for both proxies are positive, but statistically insignificant. This is not surprising given over 50% people living with hepatitis C are undiagnosed (CDC, 2016a). We fail to rule out effect sizes for intended heroin screens as large as 0.16pp, and 0.67pp for new HCV diagnoses.

4.4 Alternate Specifications and Robustness Checks

Section 3.3 summarizes the assumptions to interpret our findings as the causal effect of an opioid prescription. There are potentially multiple ways in which they could be violated.

Addressing threats to conditional independence

Previously, we showed the balance of patient observables with respect to physician prescribing leniency. However, there might be selection along unobservable margins. One way this may occur is via physician coding of diagnosis codes, which we use both to construct our sample and instrument. In column 2 of Table 9, we no longer drop any diagnosis codes—while still taking first visits—and construct leniency using the most recent outpatient diagnosis code prior to the ED visit. Another situation where conditional independence may be violated is nonrandom triaging based on unobserved severity. Our main remedy for this is to leverage variation in the composition of physicians working on a particular shift (on average 2-3 physicians per shift). Even if there is within-shift selection, patients are constrained by

the physicians on shift at the time of arrival. We leverage this *across-shift* variation in the physicians working when a patient arrives in the ED and construct a team prescribing leniency instrument (the details are in Appendix E). Note that when there is only a single physician working, this team instrument is equivalent to the baseline physician instrument and as the number of physicians working increases, the team instrument approaches the average prescribing rate in that ED, and loses first stage power. The results of Equation 3, where we now instrument $Prescribed_i$ with team leniency, are displayed in column 3 of Table 9. Apart from coefficient on HCV which is attenuated and noisier, our estimates remain largely unchanged. The main takeaway is that that patient-physician selection is unlikely to be driving our findings. Finally, we also study patients visiting the ED for injuries and poisonings, which are more likely to be random health shocks. The coefficients are reported in Table F.7 and the findings are qualitatively robust, thereby rejecting the notion that patients only become opioid dependent from continued opioid prescriptions for chronic conditions like chronic back pain; acute injuries are not immune to the adverse effects caused by opioids.

Addressing threats to the exclusion restriction

In Section 4.2, we perform placebo checks indicating that the reduced form results observed for opioid-related outcomes operate through an opioid prescription channel, as opposed to non-opioid related channels. However, we also document a failure of our placebo check for one outcome, namely preventable hospitalizations, highlighting a potential threat to interpreting our IV findings as causal effects from opioids rather than from lenient physicians: 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 is 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. Two of these decisions include admission of a patient to an inpatient hospital (hospitalization) and the intensity of the procedures performed on the patient. We explicitly model these two decisions as endogenous decisions as in Mueller-Smith (2015) and Bhuller et al. (2019). We include hospital admission as a dummy

variable and proxy for intensity of procedures with work-Relative Value Units (w-RVU), which is the part of the CMS fee schedule that converts procedure codes to a payment amount.[35] We construct instruments for admission and procedure propensities analogous to prescribing leniency by replacing *Prescribed* in Equation 1 with admission dummy and total w-RVU.

We include predicted admission and total w-RVU as controls in the baseline 2SLS regression. This approach follows Mueller-Smith (2015) and is an indirect least squares version of three endogenous variables and three instruments. Column 4 of Table 9 reports the 2SLS estimates. 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.

Beyond admission and intensity of procedures, overall physician quality can bias our estimated coefficients through a violation of the exclusion restriction. One proxy for physician quality is the effect on immediate patient mortality: what is the physician's effect on a veteran dying within one month of visiting the ED? With quasi-random assignment of patients to physicians, we can attribute any death that happens within a short period of one month to their practice in the ED. Such immediate mortality will not be caused by the physician's prescribing decision, but rather by other dimensions of physician care (the dimensions that may violate our exclusion restriction). 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 are nearly identical and our main findings are robust.

Finally, one important potential violation of the exclusion restriction relates to the intensive margin decision of the *amount* of opioids to prescribe. We have modeled opioid prescriptions as a binary extensive margin decision; however, physicians are also deciding on prescription length and dosage. If our extensive margin measure of prescribing leniency masks differences in the intensive margin, our 2SLS estimates will be biased. The bias will be

[35]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.

upward (downward) if extensive margin lenient physicians are also lenient (relatively stricter) along the intensive margin. Panel D of Figure F.8 plots the relationship between total MME and our extensive margin physician leniency and finds a small positive relationship: the average physician in the top decile of extensive margin prescribing leniency prescribes ca. 6 mg of morphine more than average physician in the bottom decile, conditional on being prescribed. This effect is not large since the average ED prescription has a morphine equivalent of 152 mg and the relationship is non-monotonic. Nevertheless, we adopt the standard approach in accounting for the intensive margin in the judges design (Bhuller et al., 2019). We include a variable for the total milligrams of morphine prescribed (including zero if the patient is not prescribed) in Equation 3, construct an intensive margin instrument (just like the propensities above and our baseline instrument), and run a 2SLS regression with two endogenous variables and two instruments. 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 9, 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 on long-term use and opioid overdoses are slightly lower; for example, the effect on long-term opioid use is now 1.09pp as opposed to 1.17pp. The estimates of the other outcomes are unchanged. This suggests that our main findings are driven by physician variation in the decision to prescribe *any* vs. *no* opioid, and not by variation in the decision of the amount of opioids prescribed. This finding has important implications for optimal regulation of opioid prescribing: It suggests that for a sample of relatively opioid-naïve patients, such as the ones in our sample, 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.

Addressing threats to monotonicity

The monotonicity assumption requires lenient physicians to be consistently lenient. 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 9 reports 2SLS estimates with these mutually exclusive instruments. Our main estimates are robust; the slight increases in magnitude is due to the fact that the sample is restricted to diagnoses with sufficiently numerous cases to measure leniency, and thus biased toward conditions such as injuries that experience a larger causal effect.

4.5 Complier Analysis

We have identified 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 (following Abadie, 2003; Dahl et al., 2014, see Appendix D for more details).

Approximately 39.4% of our baseline veteran 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.[36] In comparison, 7.6% are always-takers and 53% are never-takers. 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.

Table F.6 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

[36] 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. (2019). For Norwegians applying for disability insurance, 25% are compliers (Dahl et al., 2014). We find a larger share of compliers in our setting potentially due to the relatively low *perceived* risk from physicians (Hwang et al., 2016) compared to the immediate life-altering decisions of judge. Patients may also influence physician decisions by demanding opioids, a channel that largely shut off in the judicial system.

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.[37] 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.

4.6 Benchmarking Our Estimates

With our research design, we can quantify the role of ED physician prescribing in terms of the total number of deaths from the opioid epidemic. In our 11 year study period, approximately 9,200 veterans died of an opioid overdose. Of these 9,200 veterans, at total of 3,077 visited a VA ED. Of these, an estimated 738 are compliers (from subsection 4.5, 24% complier share for the universe of ED visits). Finally, we scale the number of compliers by our overdose relative effect size of 0.075/0.167. Our final estimate is that 3.6% of veterans who died from an opioid overdose between 2006 and 2016 experienced the event because of exposure to prescription opioids originating from their ED physician.

Returning to the trade-offs in balancing the risk between opioid addiction and pain management, we can provide some bounds on the benefits of pain management—for the marginal patient's first ED visit—that could plausibly justify the downstream costs. To this end, we calculate the expected decline in life-expectancy from the initial prescription. With multiple conservative assumptions, we arrive at a lower bound of an average life-expectancy

[37] 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.

loss of 0.54-0.66 years by prescribing an opioid to the average veteran visiting an ED.[38] This estimate implies that the incremental benefit of prescribing an opioid needs to relieve acute pain by an equivalent of 0.54-0.66 quality-adjusted life years relative to the alternative of non-opioid pain therapy. This time-frame is already much longer than what is considered acute pain. In a large randomized controlled trial, 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). These studies strongly reject our implied lower-bound estimate of an improved 0.54-0.66 life years.

5. 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 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

[38]There is a 39% chance that a prescribed veteran is a complier and the complier overdose mortality rate increases by 0.075pp. Next, we make the very conservative assumption that the effect on mortality is *zero* for non-compliers, arriving at a lower-bound increased likelihood of mortality of 0.029%. Next, with observed overdose deaths, we obtain the probability of dying each year (we only have 11 years of data, so we make another conservative assumption that the mortality effect is zero after year 11). The average male veteran of age 55 (our ED average) is expected to live an additional 22-26 years (VA, 2017). Summing over the 11 yearly probabilities, we arrive at an expected life-expectancy loss of 0.54-0.66 years.

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. For these patients, avoiding the prescription of opioids in the ED altogether, whenever possible, may be advisable. 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.

Alpert, Abby, William N. Evans, Michael M.J Lieber, and David Powell (2019), "Origins of the Opioid Crisis and Its Enduring Impacts." Working paper.

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 (2019), "Incarceration, Recidivism, and Employment." *Journal of Political Economy*, Forthcoming.

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 (2016a), "Hepatitis C a Silent Epidemic."

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

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

Chan, David C. (2016), "Informational Frictions and Practice Variation: Evidence from Physicians in Training." Nber working paper.

Chan, David C., Matthew Gentzkow, and Chuan Yu (2019), "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." Nber working paper 26467.

Chua, Kao-Ping, Chad M. Brummett, , and Jennifer F. Waljee (2019), "Opioid Prescribing Limits for Acute Pain Potential Problems With Design and Implementation." *The Journal of American Medical Association*, 321, 643–644.

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

Currie, Janet and W. Bentley MacLeod (2017), "Diagnosing Expertise:Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics*, 35.

Currie, Janet M. and W. Bentley MacLeod (2020), "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica*, 88, 847–878.

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

Davis, Corey S., Amy Judd Liebermana, Hector Hernandez-Delgadoa, and Carli Suba (2019), "Laws limiting the prescribing or dispensing of opioids for acute pain in the United States: A national systematic legal review." *Drug and Alcohol Dependence*, 194, 166–172.

de Vaan, Mathijs and Toby Stuart (2019), "Does Intra-household Contagion Cause an Increase in Prescription Opioid Use?" *American Sociological Review*, 18.

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.

Epstein, Andrew and Sean Nicholson (2009), "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections." *Journal of Health Economics*, 28, 1126–1140.

Evans, William N., Ethan M. J. Lieber, and Patrick Power (2019), "How the Reformulation of OxyContin Ignited the Heroin Epidemic." *The Review of Economics and Statistics*, 101, 1–15.

Fadlon, Itzik and Jessica N. Van Parys (2019), "Primary Care Physician Practice Styles and Patient Care: Evidence From Physician Exits in Medicare."

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist (2019), "What Is a Patent Worth? Evidence from the U.S. Patent 'Lottery'." *Journal of Finance*, Forthcoming.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2016), "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics*, 131, 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2018), "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.

Gowrisankaran, Gautam, Keith Joiner, and Pierre-Thomas Leger (2017), "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments."

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." Westat.

Hwang, Catherine, Lydia Turner, Stefan Kruszewski, Andrew Kolodny, and G. Alexander (2016), "Family Welfare Cultures." *The Clinical Journal of Pain*, 32, 279–284.

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.

Leri, Francesco, Jane Stewart, Benedikt Fischer, Jüurgen Rehm, David C Marsh, Suzanne Brissette, Julie Bruneau, Nady El-Guebaly, Lina Noël, Mark W Tyndall, and T Cameron Wild (2005), "Patterns of opioid and cocaine co-use: a descriptive study in a Canadian sample of untreated opioid-dependent individuals." *Experimental and Clinical Psychopharmacology*, 13, 303–10.

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.

Lowenstein, Margaret, David Grande, and M. Kit Delgado (2018), "Opioid Prescribing Limits for Acute Pain — Striking the Right Balance." *New England Journal of Medicine*, 379, 504–506.

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.

Molitor, David (2018), "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy*, 10, 326–356.

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*.

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.

Van Parys, Jessica (2016), "Variation in Physician Practice Styles Within and Across Emergency Departments." *Plos One*, 11, 1–19.

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.

Peterson, Rachel, Adi V. Gundlapalli, Stephen Metraux, Marjorie E. Carter, Miland Palmer, Andrew Red, Matthew H. Samore, and Jamison D. Fargo (2015), "Identifying Homelessness among Veterans Using VA Administrative Data: Opportunities to Expand Detection Criteria." *PLoS One*, 10.

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.

Ross, Edgar, Robert Jamison, and Robert Edwards (2011), "The Opioid Debate - Missing the Point." *The Journal of Pain*, 12, 508.

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.

Sasson, Comilla, Jason Smith, Chad Kessler, Jason Haukoos, Julianne Himstreet, Melissa Christopher, and Thomas Emmendorfer (2019), "Variability in opioid prescribing in veterans affairs emergency departments and urgent cares." *The American Journal of Emergency Medicine*, 37, 2044–1047.

Scherrer, Jeffrey F., Joanne Salas, Laurel A. Copeland, Eileen M. Stock, Brian K. Ahmedani, Mark D. Sullivan, Thomas Burroughs, David Schneider, Kathleen K. Bucholz, and Patrick J. Lustman (2016), "Prescription Opioid Duration, Dose, and Increased Risk of Depression in 3 Large Patient Populations." *Annals of Family Medicine*, 14, 54–62.

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., 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.

Silver, David (2019), "Haste or Waste? Peer Pressure and Productivity in the Emergency Department."

Sowicz, Timothy Joseph, Adam J. Gordon, Walid F. Gellad, Xinhua Zhao, Hongwei Zhang, Thomas Emmendorfer, and Chester B. Good (2018), "High Variability of Opioid Prescribing Within and Across Emergency Departments in the US Veterans Health Administration." *Journal of General Internal Medicine*, 33, 1831–1832.

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.

Tsugawa, Yusuke, Ashish K. Jha, Joseph P. Newhouse, Alan M. Zaslavsky, and Anupam B. Jena (2017), "Variation in Physician Spending and Association With Patient Outcomes." *JAMA Intern Med*, 177, 675–682.

Tu, Peter (2017), "Quasi-experimental Evidence of Physician Effects." Working paper.

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

Volkow, Nora D. and Thomas A. McLellan (2016), "Opioid Abuse in Chronic Pain — Misconceptions and Mitigation Strategies." *The New England Journal of Medicine*, 374, 1253–1263.

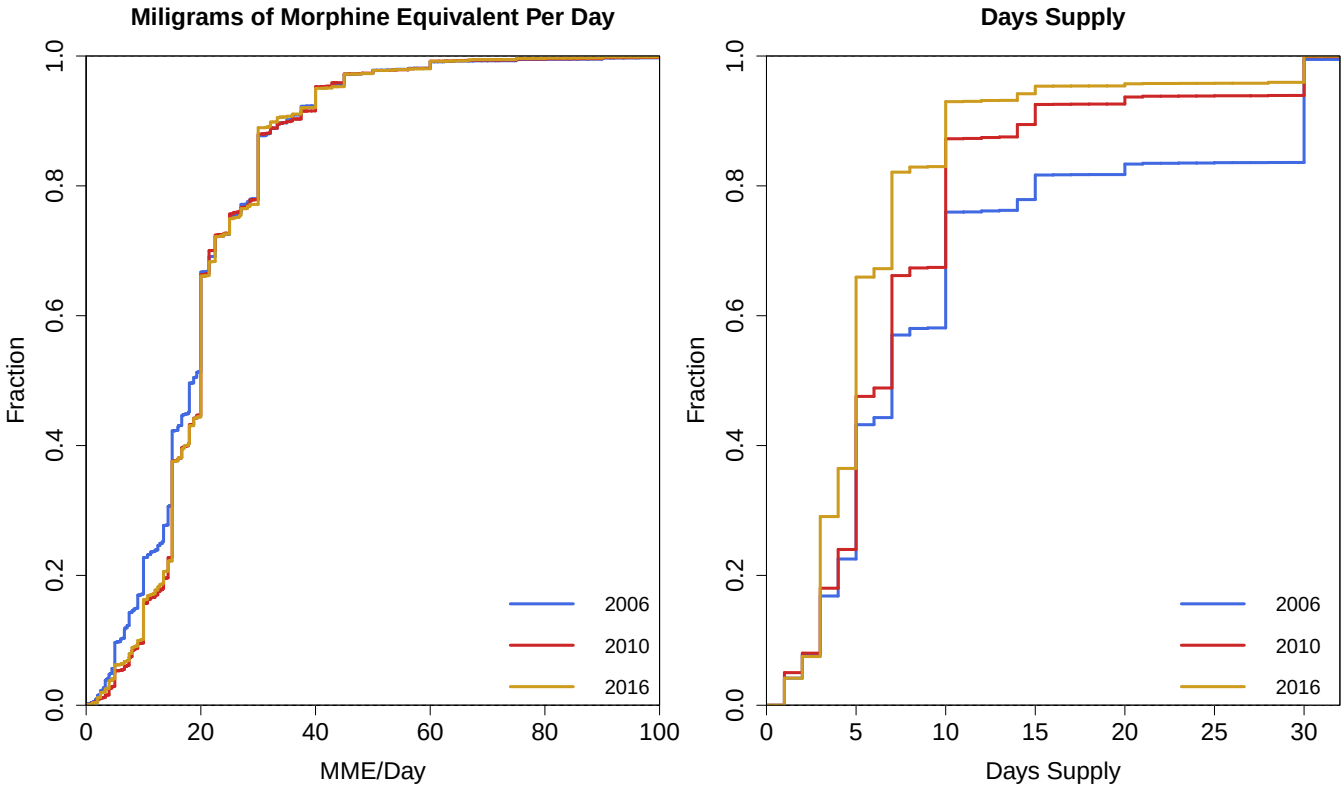
Yang, Zhuo, Barth Wilsey, Michele Bohm, Meghan Weyrich, Kakoli Roy, Dominique Ritley, Christopher Jones, and Joy Melnikowx (2015), "Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-Term Opioid Users in Medicaid." *The Journal of Pain*, 16, 445–453.

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.

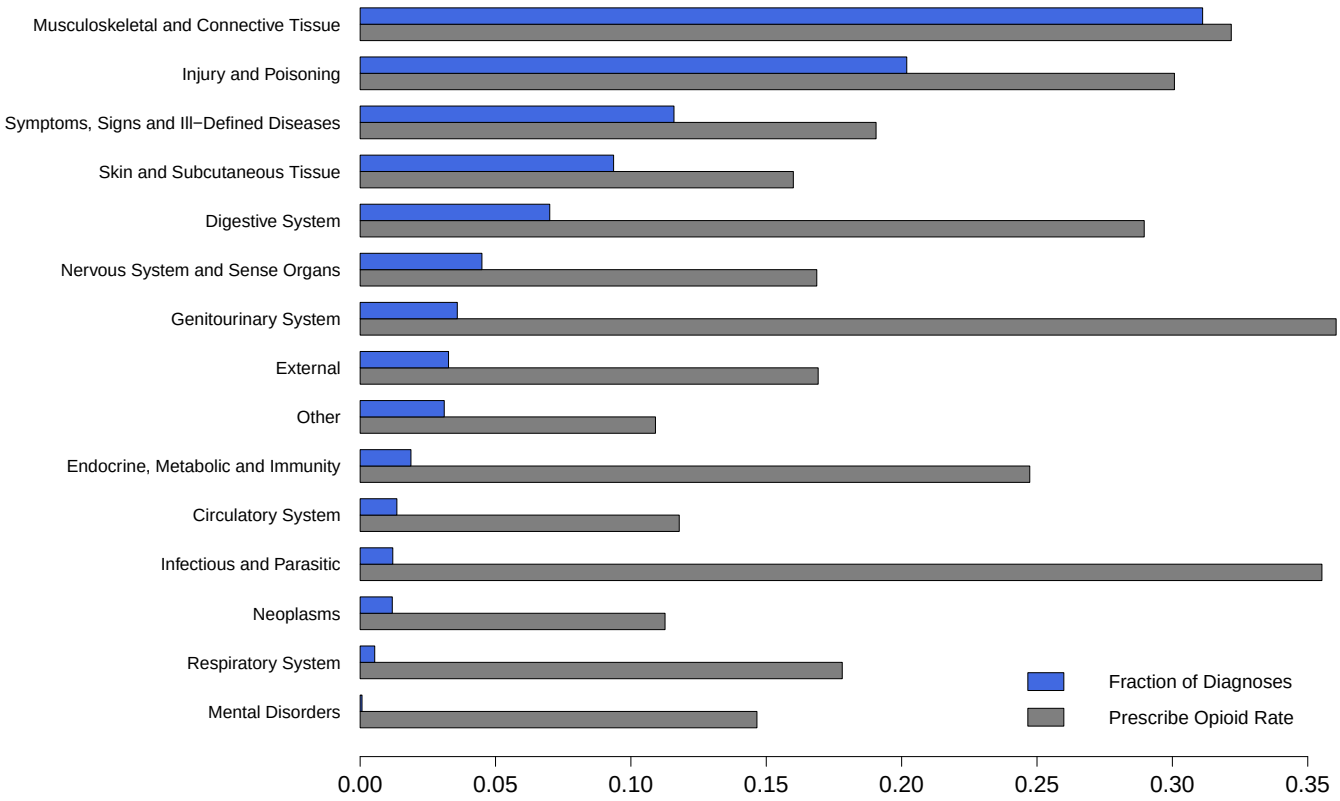
Figures

Figure 1: Empirical CDF of Opioid Prescription Characteristics



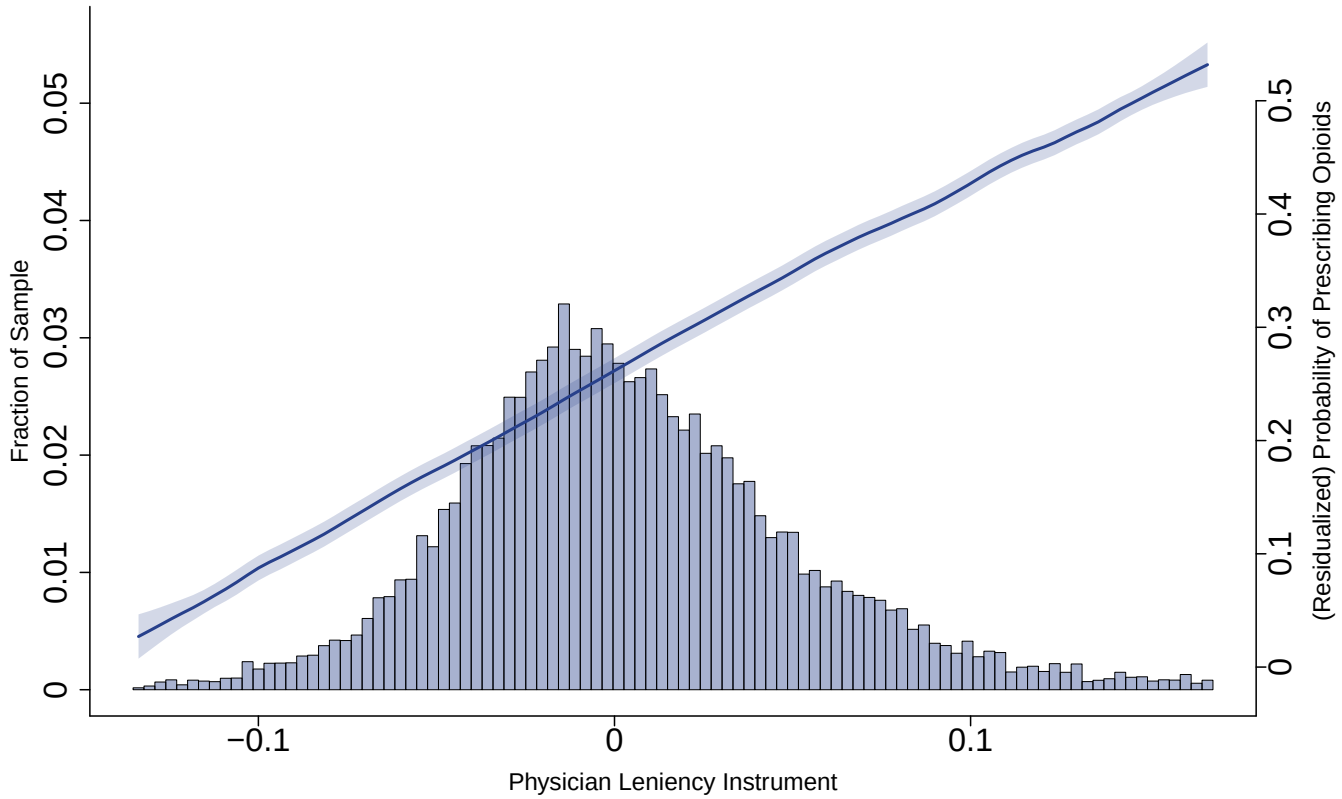
Notes: The empirical cumulative distribution function for days supply and milligrams of morphine equivalence per day, conditional on being prescribed, by year.

Figure 2: Frequent Diagnoses Occurring in Emergency Departments



Notes: The 15 most common major diagnosis categories (ICD-9 major chapters) for ED visits and the un-adjusted rate they are prescribed opioids in our baseline sample. Rarely-prescribed categories are excluded, see the text for details.

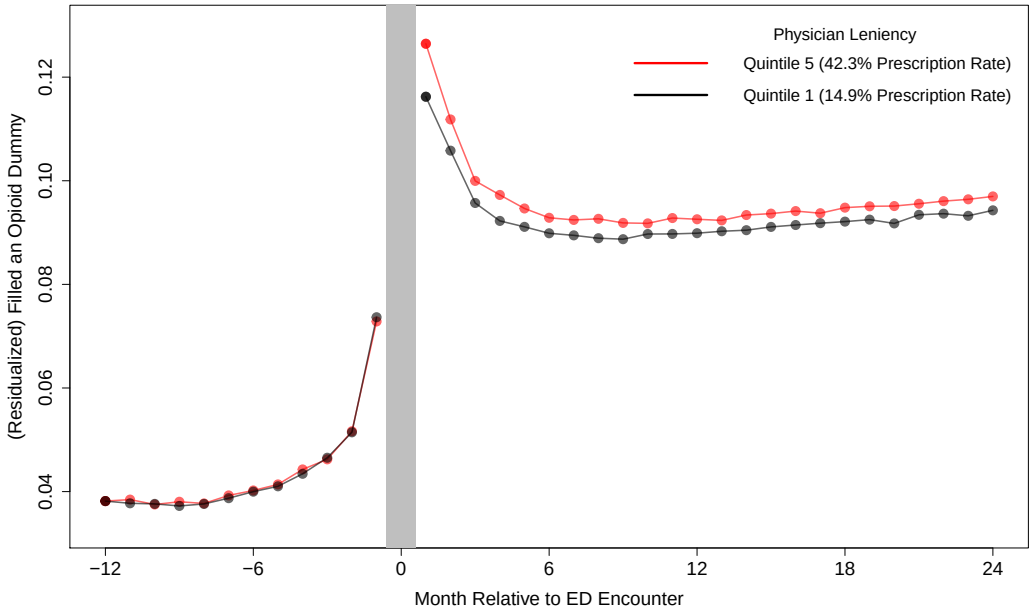
Figure 3: Distribution and First Stage of Instrument



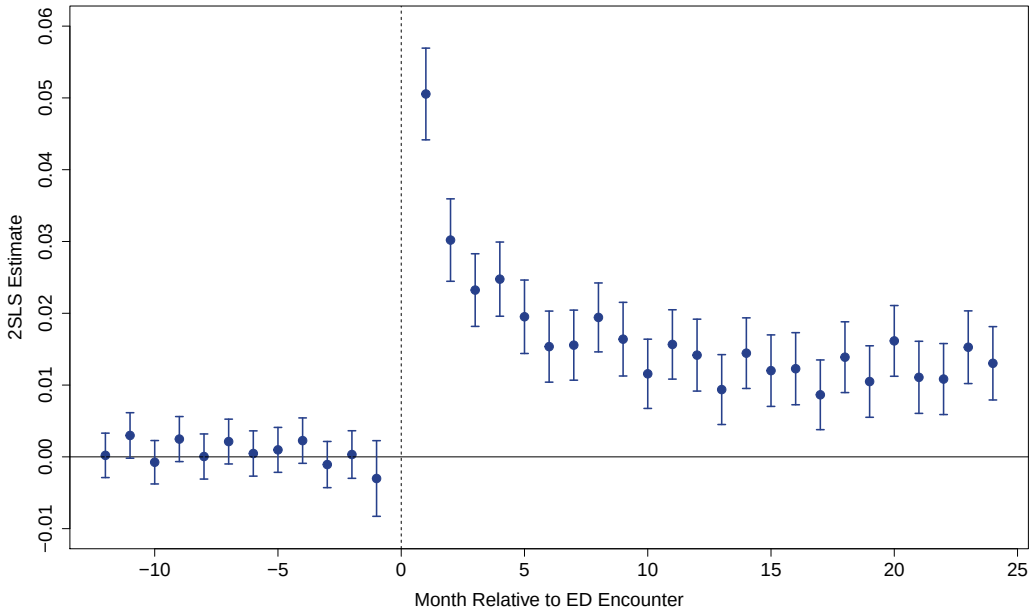
Notes: This figure plots the histogram of the baseline instrument along the x-axis and the left y-axis. A local-linear regression of the fitted probability of prescribed opioids on the instrument after residualizing (see text for baseline fixed effects and controls in residualization) is overlaid and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

Figure 4: Reduced Form and 2SLS Event Study for Prescription Opioid Use

Panel A. Reduced Form

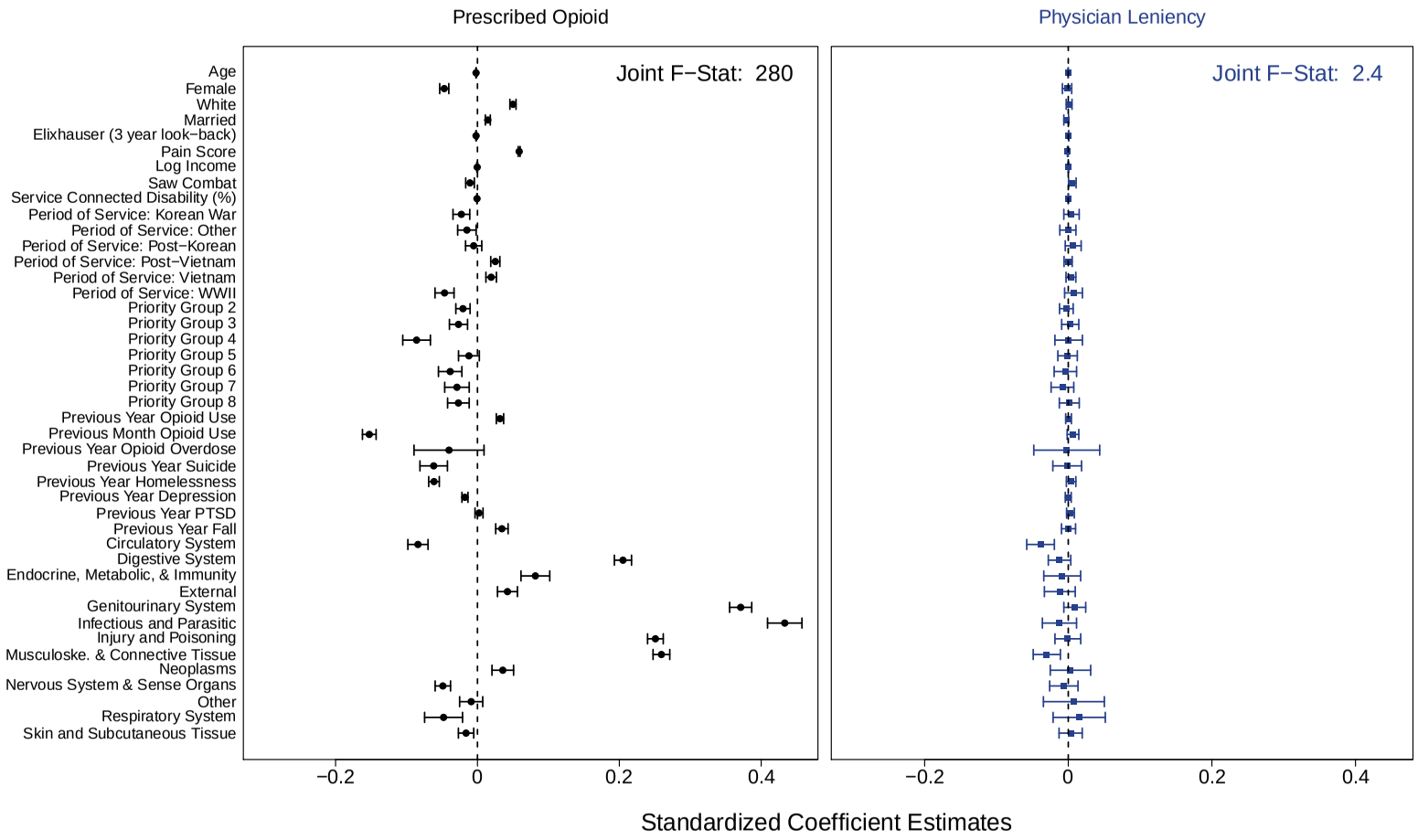


Panel B. 2-Stage Least Squares



Notes: Panel A plots the reduced form, event-study of opioid use on leniency of physician. Monthly residualized opioid use indicators (see text for baseline fixed effects and controls in residualization), for patients who see a strict and lenient physician are shown from one year pre-ED to 2 years post-ED. Panel B is the 2-stage least squares analog. Each point is an estimate from a separate regression of opioid use dummy that month on $Prescribed_i$ (the omitted opioid use dummy at time zero), instrumented with physician leniency. Only new filled opioid prescriptions contribute to each point and the initial ED prescription is omitted. 95% confidence intervals constructed using robust standard errors clustered at the physician level are also displayed.

Figure 5: 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, diagnosis, and age bins. 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. All variables include VHA and CMS data when available.

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, black, 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 Regressions on 90^{th}-10^{th} Percentile Physician Leniency

	Reduced Form Sample:		
	Baseline: Sample	P(*Prescribed*) $\in [0.03, 0.1)$	P(*Prescribed*) < 0.03
Dependent Variable (×100):	(1)	(2)	(3)
Panel A: Opioid-Related Outcomes			
Long-Term Use	0.237***	0.063*	0.065
	(0.041)	(0.035)	(0.088)
Mean Dep. Var. (×100)	5.8	4.5	4.0
Overlapping Prescriptions (Year 0-1)	0.389***	0.094*	0.081
	(0.053)	(0.052)	(0.062)
Mean Dep. Var. (×100)	9.9	7.8	7.9
Opioid Use Disorder (Year 0-3)	0.067*	0.020	0.035
	(0.038)	(0.034)	(0.049)
Mean Dep. Var. (×100)	3.3	2.9	4.6
Opioid Overdose (Year 0-3)	0.007	0.007	0.007
	(0.012)	(0.013)	(0.017)
Mean Dep. Var. (×100)	0.6	0.4	0.6
Opioid Overdose Mortality	0.015**	0.003	−0.004
	(0.007)	(0.007)	(0.010)
Mean Dep. Var. (×100)	0.2	0.1	0.2
Panel B: Non-Opioid Outcomes			
Homelessness (Year 0-3)	0.047	−0.059	−0.031
	(0.060)	(0.070)	(0.097)
Mean Dep. Var. (×100)	11.9	11.9	16.2
Suicide (Year 0-3)	−0.006	0.021	−0.009
	(0.022)	(0.023)	(0.039)
Mean Dep. Var. (×100)	1.6	1.3	2.2
Heart Disease Mortality	0.035	0.036	−0.076
	(0.033)	(0.046)	(0.059)
Mean Dep. Var. (×100)	4.0	5.4	7.4
Preventable Hospitalizations (Year 0-3)	0.040	0.165**	0.273**
	(0.059)	(0.076)	(0.112)
Mean Dep. Var. (×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 patient 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 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 4: Subsequent Opioid Use and Long-Term Use Outcomes

	Mean	OLS		2SLS		N=
Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Long-Term Use (×100)						
	5.8	2.85***	2.63***	1.32***	1.17***	1,879,150
		(0.070)	(0.061)	(0.209)	(0.202)	
Total Milligrams of Morphine						
Prior Year	214	0.39	−12.94***	12.35**	3.57	1,958,209
		(1.51)	(0.66)	(5.08)	(3.01)	
Year 0-3	2,353	969.1***	810.7***	500.8***	467.7***	1,775,800
		(22.0)	(19.3)	(61.9)	(59.5)	
Positive Opioid Urine Drug Screen (×100)						
Prior Year	2.0	−0.039	−0.227***	−0.189	−0.160	1,958,209
		(0.028)	(0.028)	(0.132)	(0.130)	
Year 0-3	8.2	2.87***	2.08***	2.05*	2.01**	1,775,800
		(0.070)	(0.065)	(0.257)	(0.253)	
Residualization FE?	-	Yes	Yes	Yes	Yes	-
Baseline Controls?	-	No	Yes	No	Yes	-

Notes: This table reports the effect of an opioid prescription on long-term use (180 days supply in the first 12 months), milligrams of morphine equivalent filled, and positive urine drug screens for opioids (patients who are not tested are coded as zero). Column 1 reports the mean dependent variable, columns 2 and 3 report ordinary least squares regression estimates, and columns 4 and 5 report the two-stage least squares IV regression estimates. All outcomes exclude 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, diagnosis, and age bins; 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 5: Proxies for Opioid-Seeking Behavior and Self-Reported Pain Score

		OLS		2SLS		N=
Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Overlapping Prescriptions (×100)						
Prior Year	2.9	−0.09**	0.07**	0.17	0.20	1,840,595
		(0.04)	(0.03)	(0.15)	(0.14)	
Year 0-1	9.9	4.4***	4.1***	1.9***	1.9***	1,840,595
		(0.1)	(0.1)	(0.2)	(0.2)	
Pharmacy Shopping (×100)						
Prior Year	0.2	0.07***	0.09***	0.06	0.06	1,840,595
		(0.01)	(0.01)	(0.04)	(0.04)	
Year 0-1	0.6	0.45***	0.43***	0.30***	0.30***	1,840,595
		(0.02)	(0.02)	(0.07)	(0.07)	
Repeated Back Pain & Headaches (×100)						
Prior Year	3.8	−0.24***	−0.53***	0.23	0.30	1,532,610
		(0.04)	(0.04)	(0.21)	(0.20)	
Year 0-1	6.2	0.83***	0.38***	0.47*	0.55***	1,532,610
		(0.06)	(0.05)	(0.27)	(0.27)	
Self-Reported Pain Score						
Prior Year	2.39	0.17***	0.18***	0.02	0.03	1,501,244
		(0.01)	(0.005)	(0.02)	(0.02)	
Year 0-1	2.81	0.36***	0.18***	0.05**	0.08***	1,682,968
		(0.01)	(0.005)	(0.02)	(0.02)	
Residualization FE?	-	Yes	Yes	Yes	Yes	-
Baseline Controls?	-	No	Yes	No	Yes	-

Notes: This table reports of effects of an opioid prescription on proxies for opioid-seeking behavior and self-reported pain scores for the prior year and the first year following the emergency visit. The proxies are defined in the text. Column 1 reports the mean dependent variable, columns 2 and 3 report OLS estimates, and columns 4 and 5 report the 2SLS IV regression estimates. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins; see text for baseline controls. The samples are constrained such that the patients are alive for the entire period the outcome is measured. Sample for repeated back pain and headaches exclude ED visits that were for back pain or headaches. 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 6: Opioid Use Disorder, Opioid Overdose, and Opioid Overdose Mortality

	Mean	OLS		2SLS		N=
Dependent Variable (×100):	(1)	(2)	(3)	(4)	(5)	(6)
Opioid Use Disorder (OUD)						
Prior Year	1.56	−0.338***	−0.493***	0.033	0.041	1,958,209
		(0.024)	(0.025)	(0.106)	(0.105)	
Year 0-3	3.27	0.318***	0.013	0.323*	0.335*	1,775,800
		(0.039)	(0.039)	(0.161)	(0.160)	
Opioid Overdose						
Prior Year	0.11	−0.008	−0.019***	0.015	0.013	1,958,209
		(0.006)	(0.006)	(0.030)	(0.030)	
Year 0-3	0.59	0.082***	0.026*	0.044	0.039	1,775,800
		(0.015)	(0.015)	(0.069)	(0.069)	
Opioid Overdose Death						
	0.167	0.065***	0.044***	0.074**	0.075**	1,846,133
		(0.008)	(0.008)	(0.034)	(0.034)	
Specific Cause of Death:						
Heroin or Synthetic Opioids	0.084	0.026***	0.017***	0.030	0.030	1,846,133
		(0.006)	(0.006)	(0.024)	(0.023)	
Non-Heroin Natural Opioids *Only*	0.050	0.027***	0.020***	0.034*	0.034*	1,846,133
		(0.005)	(0.005)	(0.020)	(0.020)	
Residualization FE?	-	Yes	Yes	Yes	Yes	-
Baseline Controls?	-	No	Yes	No	Yes	-

Notes: This table reports the effect of an opioid prescription on opioid use disorder (OUD) and opioid overdose at various points in time, eventual death from an opioid overdose, and specific type of opioid cause of death. ICD-10 mortality codes: heroin (T40.1); synthetic opioids excluding methadone (T40.4); natural opioids only (T40.2 and no other opioid type). Column 1 reports the mean dependent variable, columns 2 and 3 report ordinary least squares regression estimates, and columns 4 and 5 report the two-stage least squares IV regression estimates. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins; see text for baseline controls. The samples for the OUD and overdose regressions 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 7: Health and Non-Opioid Outcomes

	Mean	OLS		2SLS		N=
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Fall (×100)						
Prior Year	5.4	0.015	−0.042	0.137	0.120	1,958,209
		(0.041)	(0.041)	(0.221)	(0.218)	
Years 0-3	8.4	0.299***	0.052	0.552**	0.504**	1,775,800
		(0.052)	(0.051)	(0.262)	(0.257)	
Depression (×100)						
Prior Year	24.3	-0.143	−1.011***	0.268	0.433	1,958,209
		(0.087)	(0.083)	(0.422)	(0.411)	
Years 0-3	36.4	1.324***	0.195**	0.242	0.427	1,775,800
		(0.101)	(0.095)	(0.471)	(0.459)	
Homelessness (×100)						
Prior Year	6.2	−0.842***	−1.078***	0.374*	0.414*	1,958,209
		(0.048)	(0.050)	(0.224)	(0.221)	
Years 0-3	11.9	−0.408***	−0.798***	0.174	0.201	1,775,800
		(0.068)	(0.068)	(0.297)	(0.295)	
Suicide Attempt/Self-Harm (×100)						
Prior Year	0.65	−0.098***	−0.116***	−0.019	−0.019	1,958,209
		(0.014)	(0.015)	(0.079)	(0.079)	
Years 0-3	1.59	0.030	−0.038	−0.027	−0.026	1,775,800
		(0.024)	(0.024)	(0.117)	(0.117)	
Residualization FE?	-	Yes	Yes	Yes	Yes	-
Baseline Controls?	-	No	Yes	No	Yes	-

Notes: This table reports the effects of an opioid prescription on fall, depression, attempted suicide/self-harm and homeless dummies for the prior year and the 3-year period following the emergency visit. Column 1 reports the mean dependent variable, columns 2 and 3 report ordinary least squares regression estimates, and columns 4 and 5 report the two-stage least squares IV regression estimates. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins; see text for 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 8: Substance Use and Abuse Outcomes, Questionnaires, and Proxies

	Prior Year		Year 0-3	
	Coef	Mean Dep. Var.	Coef	Mean Dep. Var.
Dependent variable (×100)	(1)	(2)	(3)	(4)
Any drug *abuse* (illicit or prescription)†	0.063	0.224	0.156*	0.776
	(0.050)		(0.088)	
Opiate *use*†	0.003	0.077	0.066	0.274
	(0.028)		(0.054)	
Cocaine/Crack *use*†	0.026	0.088	0.143**	0.317
	(0.034)		(0.057)	
Sedatives *use*†	−0.016	0.049	0.061	0.172
(e.g., benzodiazepines)	(0.023)		(0.042)	
Other stimulant *use*†	−0.015	0.047	0.047	0.155
(e.g., amphetamines)	(0.021)		(0.040)	
Marijuana *use*†	0.023	0.134	0.004	0.471
	(0.038)		(0.068)	
Positive Alcohol screen	0.484**	9.46	0.214	20.5
	(0.217)		(0.355)	
Intended Heroin/Fentanyl Drug Screen	−0.039	0.10	0.015	0.070
	(0.029)		(0.072)	
Hepatitis C Diagnosis	0.177	4.3	0.259	5.9
	(0.175)		(0.209)	
N=	1,958,209		1,775,800	
†: in the past 30 days				

Notes: This table reports the effect of an opioid prescription on outcomes and proxies for substance abuse. 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 9: Alternate Specifications and Robustness Checks

	2SLS Estimates						
	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)
OUD	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)
Overdose	0.039	0.229***	0.087	0.032	0.035	0.032	0.139**
	(0.069)	(0.086)	(0.097)	(0.061)	(0.062)	(0.061)	(0.070)
Overdose Death	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)
Fall	0.504**	0.693**	0.525	0.468*	0.488*	0.521***	0.611**
	(0.257)	(0.034)	(0.393)	(0.255)	(0.258)	(0.256)	(0.281)
Hepatitis C	0.259	0.086	0.011	0.257	0.240	0.322	0.444*
	(0.207)	(0.320)	(0.321)	(0.207)	(0.209)	(0.205)	(0.244)
Observations	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. See subsection 4.4 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

For Online Publication

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, mortality, depression, homelessness, attempted suicide, and proxies for illicit opioid use. In addition, we supplement the legal observed VHA opioid prescriptions in Barnett et al. (2019) with Medicare and Medicaid claims and attempt to account for unobserved black market opioids.

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 3.1, 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 3.1, 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 F.8 reports the result of this exercise. The first three columns of row 1 are taken directly from Barnett et al. (2019); the Wald estimate[39] (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 differences in data definitions. Next, we make incremental changes to the sample restrictions

[39]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.

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[40] (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 F.8 (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 3.1, 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 "shift-level" variation in physician work schedule and patient arrival, diagnosis condition, and

[40]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.

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 F.7 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. 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 physican 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 F.9 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 4.4. Figure F.8 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 4.4 that they do not have significant effects on our findings.

D. 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 F.6.

E. Team (Across-Shift) Leniency

As mentioned in subsection 3.3, 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:

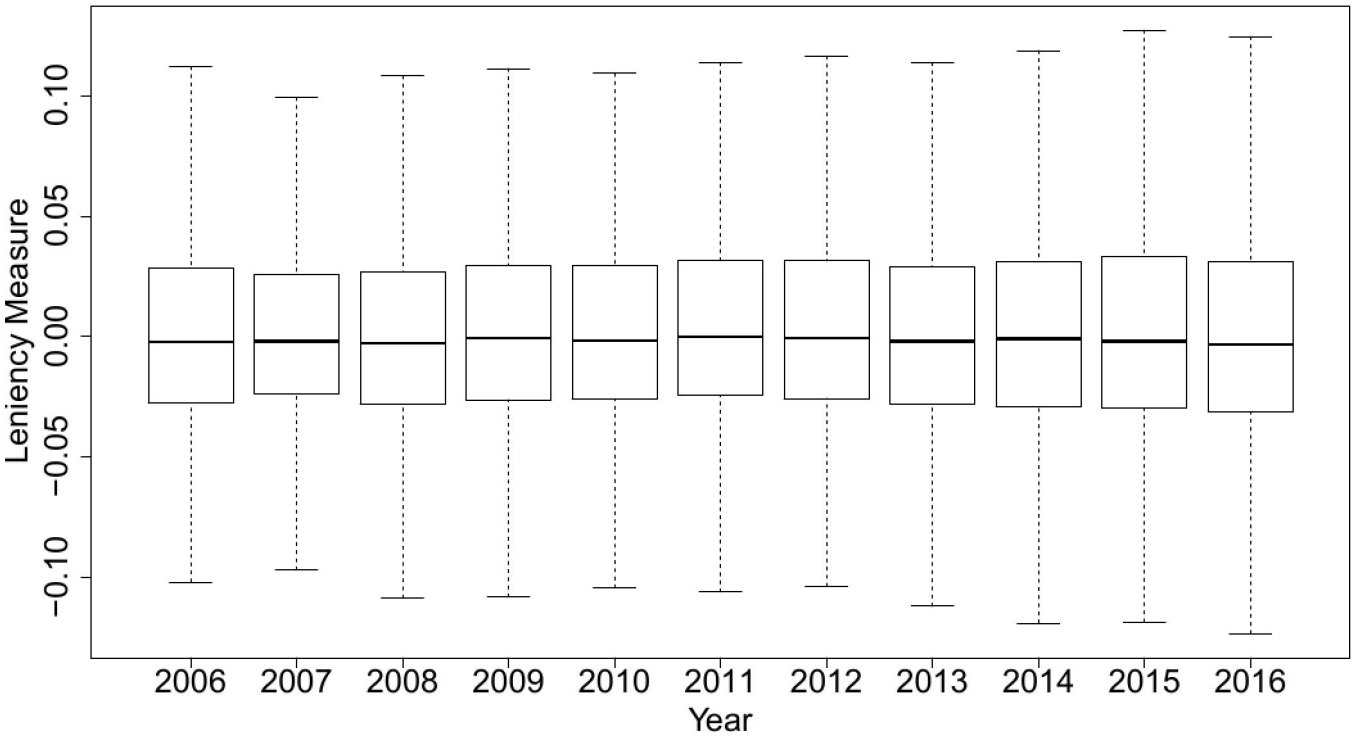
$$Leniency_s^{team} = \frac{1}{\sum_{j \in \mathbb{S}} N_{jy}} \left(\sum_{j \in \mathbb{S}} N_{jy} \times Leniency_{-s,jy}^{phys} \right) \tag{4}$$

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 F.9 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.

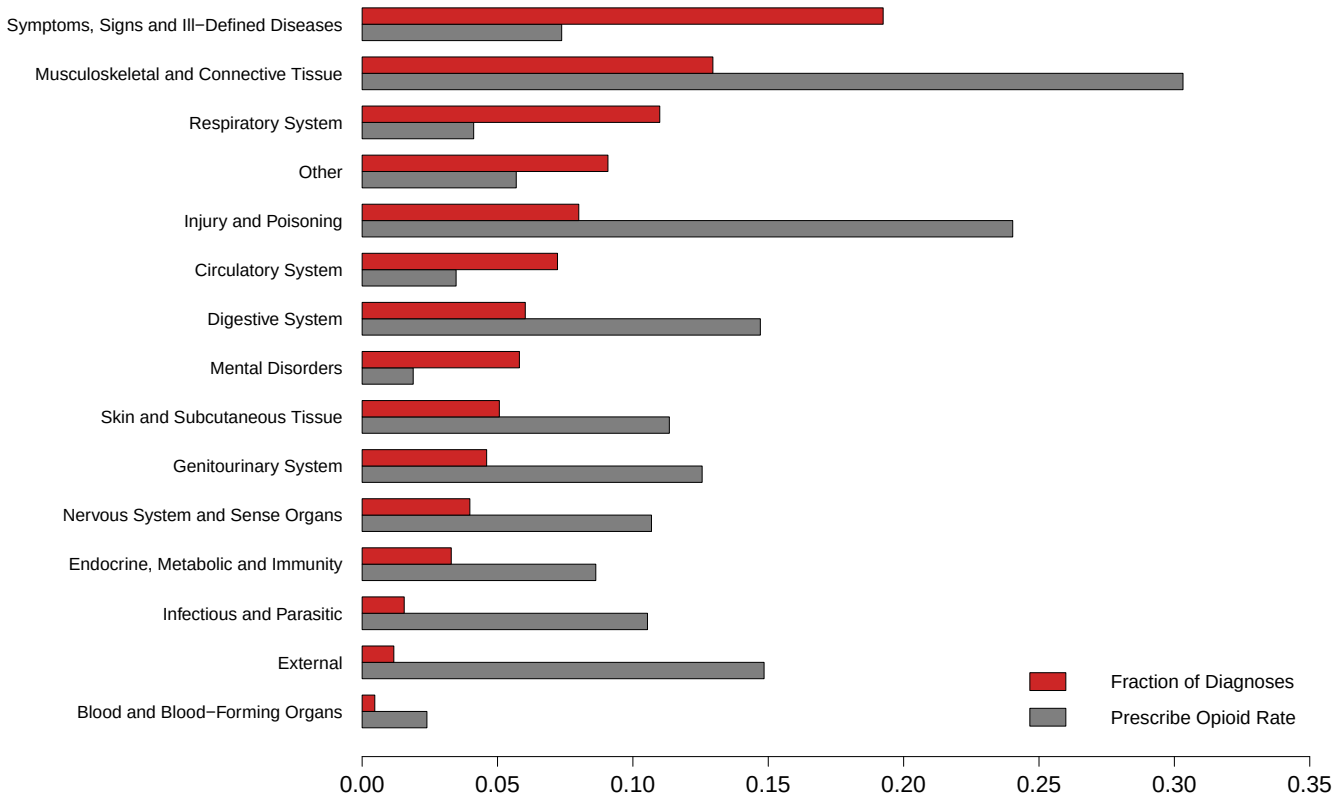
F. Additional Figures and Tables

Figure F.2: Boxplot of Residualized Prescribing Leniency Measure



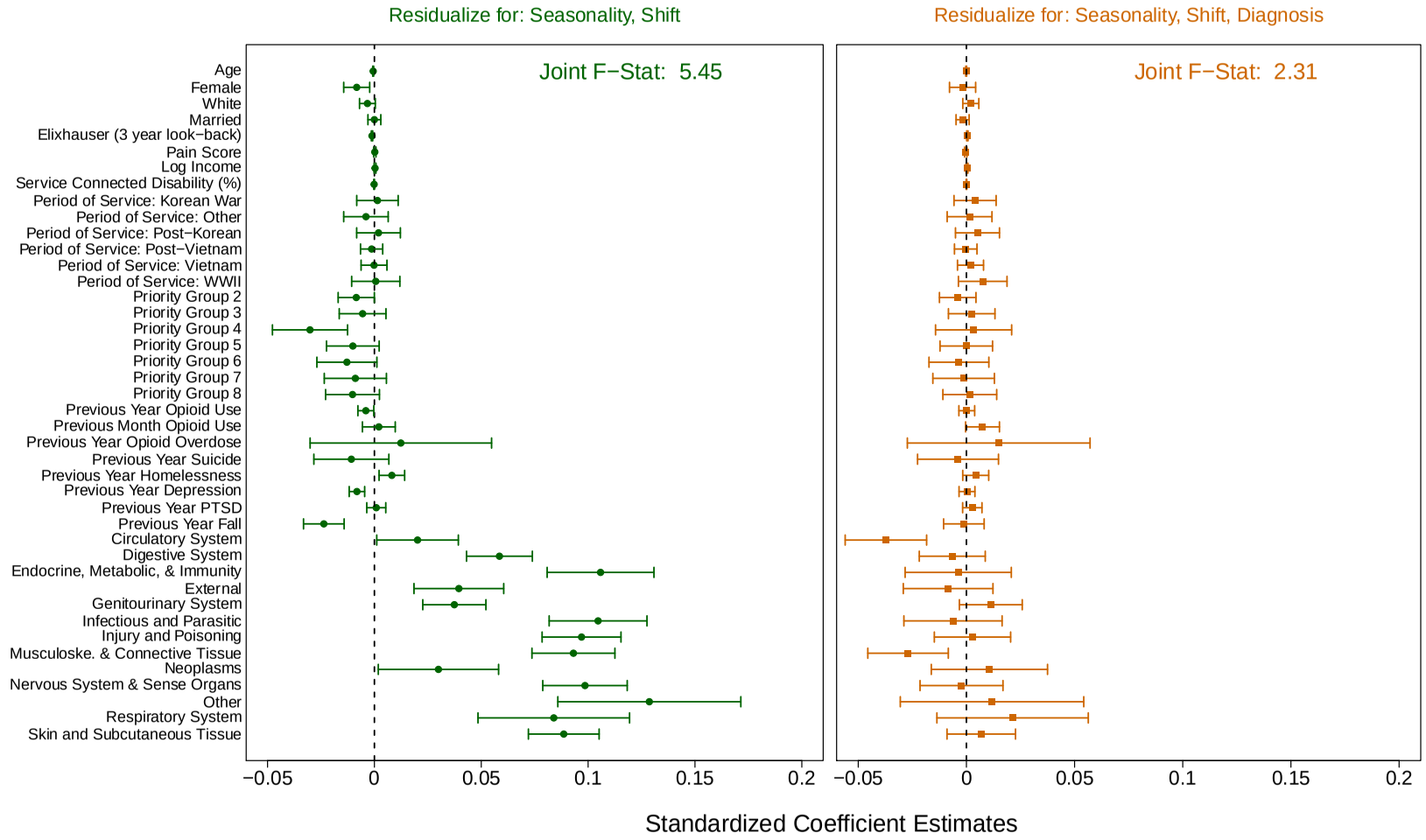
Notes: Boxplot of the residualized leniency measure for physicians in our baseline sample as defined in the text.

Figure F.3: Frequent Diagnoses Occurring in Emergency Departments



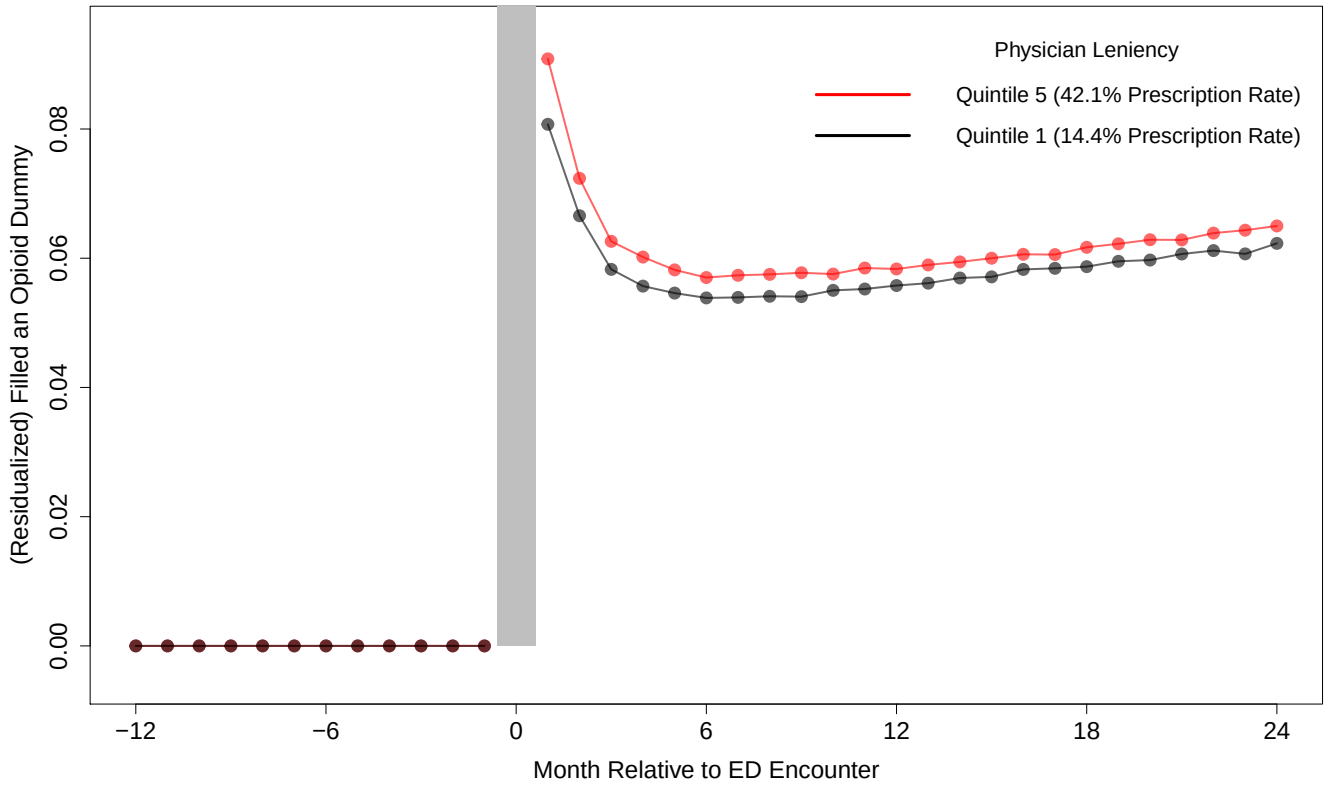
Notes: The 15 most common major diagnosis categories (ICD-9 major chapters) for *all* ED visits and the un-adjusted rate they are prescribed opioids

Figure F.4: Balance Test at Varying Levels of Residualization Controls



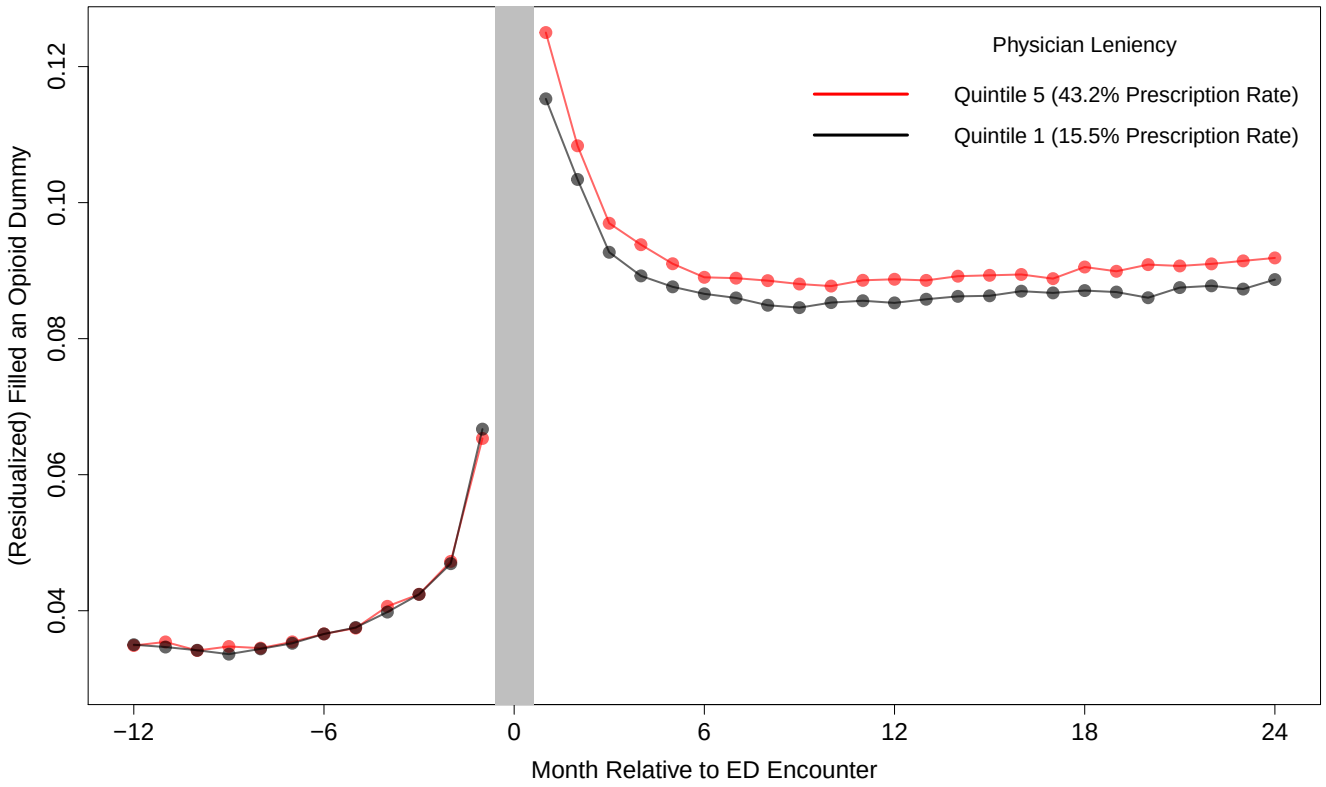
Notes: This figure displays the standard balance test (as seen previously) for the physician leniency measure constructed with different controls in the residualization of Equation 1, effectively, testing for different quasi-random assignment assumptions. In the left panel, the physician leniency instrument 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 F.5: Reduced Form: Subsequent Opioid use for Opioid-Naïve Patients



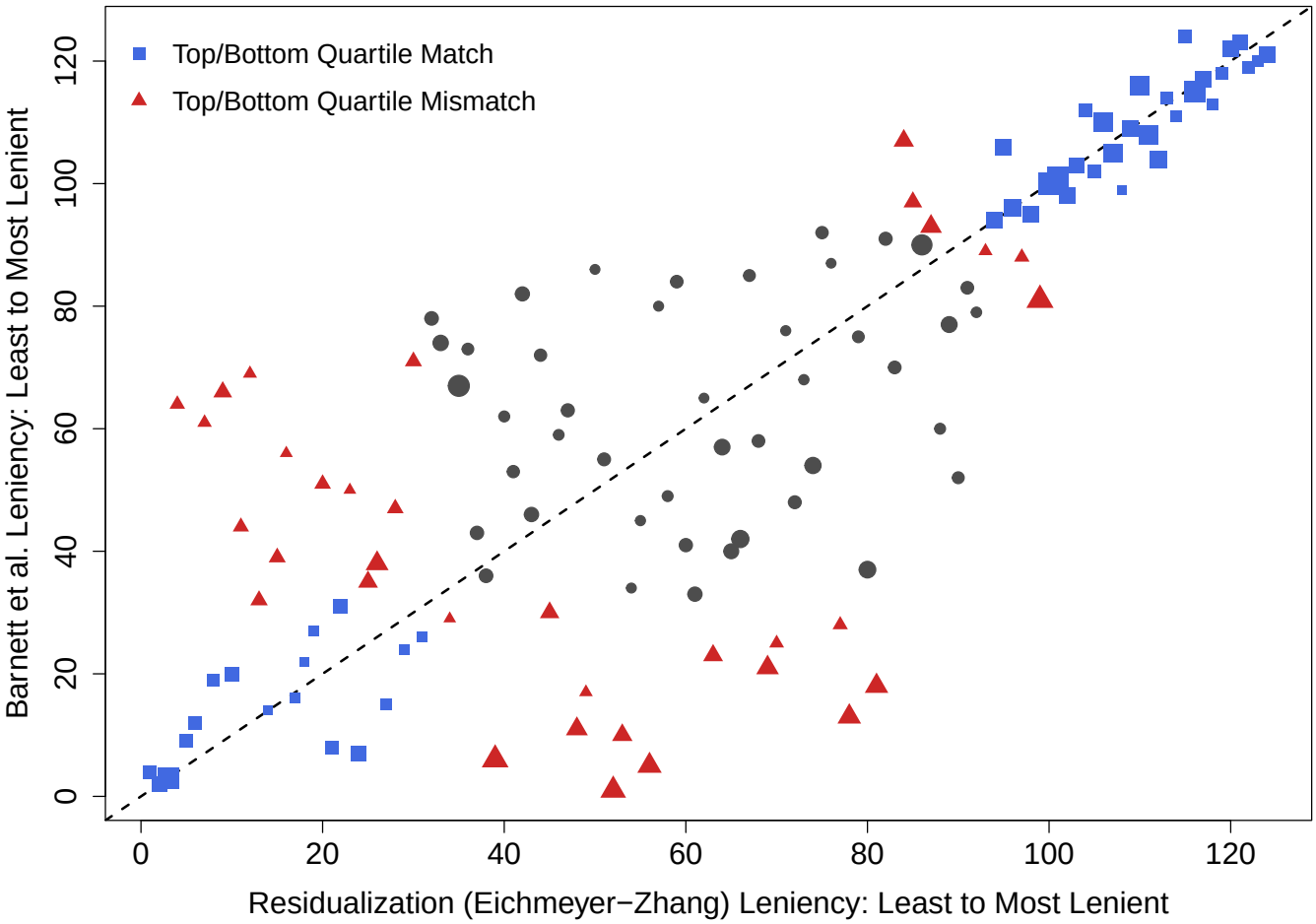
Notes: The reduced form event-study figure corresponding to Figure 4, but for opioid-naïve patients.

Figure F.6: Reduced Form: Subsequent Opioid use for Patients Without an ED Visit in the Prior Year



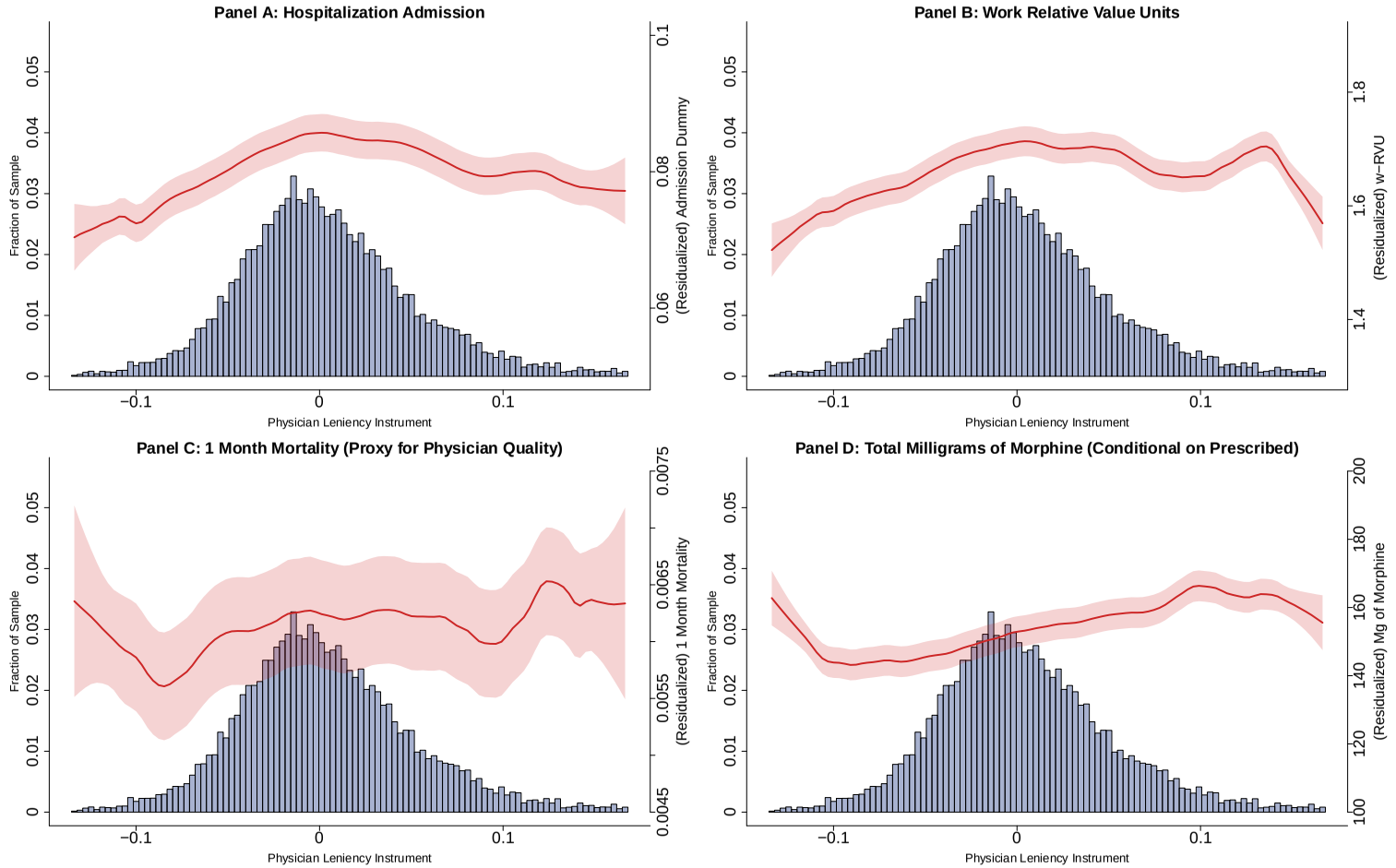
Notes: The reduced form event-study figure corresponding to Figure 4, but for patients who did not visit an ED in the prior year (for any condition), but did utilize VHA outpatient care. Presumably this group are not ED shopping for opioids.

Figure F.7: Ranking Physician Prescribing Leniency in Tampa Veteran Affairs Medical Center



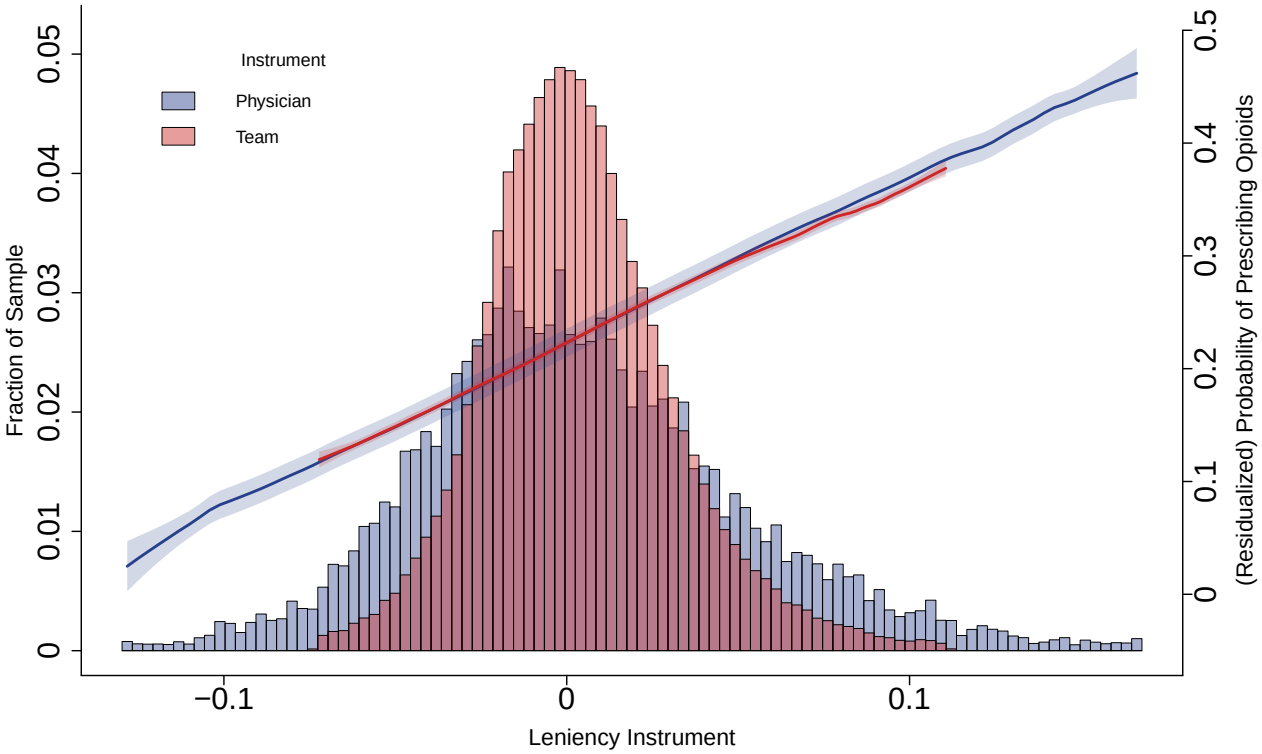
Notes: This graph shows the re-shuffling of physician ranking based on Barnett et al. method and our residualization method for physicians with at least 30 cases in Tampa Veteran Affairs Medical Center, the largest in the country by ED volume in 2012. Each point is a physician and the size of the point 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 F.8: 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 F.9: 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 F.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 F.2: 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)
OUD (Year 0-3)	−0.133	0.300*	0.335*
	(0.186)	(0.175)	(0.160)
Opioid Overdose (Year 0-1)	0.014	0.068	0.088*
	(0.048)	(0.044)	(0.043)
Opioid Overdose Death	0.068*	0.074**	0.075**
	(0.037)	(0.034)	(0.034)
Falls (Year 0-3)	−0.604**	0.461	0.504**
	(0.275)	(0.256)	(0.257)
Hepatitis C (Year 0-3)	0.080	0.250	0.259
	(0.218)	(0.206)	(0.209)

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 F.4: Effect of an ED Opioid on Outcomes by Veteran Combat Status

	Dependent variable:			
	Long-Term Use	Opioid Use Disorder	Opioid Overdose Mortality	Hepatitis C
	(1)	(2)	(3)	(4)
Combat	0.459	0.429	−0.076	0.233
	(0.666)	(0.564)	(0.122)	(0.585)
Non-Combat	1.315***	0.482**	0.123***	0.370
	(0.246)	(0.217)	(0.041)	(0.256)

Notes: This table reports the effect of an opioid prescription on the main outcomes for combat and non-combat veterans. The 2SLS regressions are estimated on each sub-sample separately. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins; see text for 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 F.5: Testing the Monotonicity Assumption

First Stage:	*Dependent variable:* Prescribed Opioid	
	Baseline Leniency	Reverse-Sample Leniency
Sub-sample:	(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, and displays the coefficient of the first stage regression back on that sub-sample. Hence the name "reverse-sample". Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01

Table F.6: 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 F.7: Opioid Prescription on Outcomes for ED Conditions Pertaining to Injuries and Poisonings

	Dependent variable (×100):								
	Long-Term Use	OUD Year 0-3	Overdose Year 0-1	Overdose Death	Fall Year 0-3	Depression Year 0-3	Suicide Year 0-3	Homeless Year 0-3	Hep C Year 0-3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Prescribed	1.438***	0.255	0.166*	0.031	0.340	0.624	−0.211	−1.36**	0.788*
	(0.391)	(0.360)	(0.098)	(0.070)	(0.619)	(0.969)	(0.283)	(0.634)	(0.422)
N=	382,034	382,034	382,034	373,178	361,090	361,090	361,090	361,090	361,090

Notes: This table reports 2SLS regression coefficients of *Prescribed* on the main outcomes the subset of the baseline sample with ED diagnosis of injury and poisoning (ICD-9 800-999, excluding drug poisonings). Residualization fixed effects include hospital-year-month, hospital-day of week-time of day, diagnosis, and age bins; see text for 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 F.8: 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 F.9: 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.