Physician Regulation: Evidence from the Opioid Epidemic *

Colleen Carey† Jetson Leder-Luis‡ Chloe Smith§

October 2026

Abstract

Physicians have wide latitude in the way they practice medicine, which can result in harmful outcomes. We examine the performance of the principal physician regulatory institution, state medical boards, during a recent episode of dangerous medical treatment: high-risk opioid prescribing in the United States. Using a novel dataset of nationwide disciplinary records, we assess the intensity of discipline on physicians with high-risk prescribing patterns. We document that disciplinary action for high-risk prescribing was rare and frequently occurred after many years of high-risk prescribing. We show that when boards chose to act, their disciplinary tools were successful in substantially reducing opioid prescribing. Turning to patient outcomes, we demonstrate that disciplinary action dramatically improved the health of the patients of high-risk prescribers, as compared to the patients of never-disciplined high-risk prescribers, without impairing their access to non-opioid services. Patients whose physicians faced discipline began to consume fewer opioids, received appropriate care for opioid use disorder, and were slightly less likely to die.

*We thank participants of the ASHEcon conference, the Empirical Health Law Conference, and the Junior Health Economics Conference for feedback. We thank Derrick Bobb, David Cameron, Ruhi Datar, Jason Gordon, Grace Kim, William Li, Sahaj Satani, Gabrielle Sorresso, Samuel Stille, and Victor Verma for research assistance. Amanda Agan contributed to text analysis. We thank Elizabeth Munnich for help in accessing data for Kentucky. This research was supported by a Cornell Center for Social Sciences Grant.

†Cornell University and NBER

‡Boston University and NBER

§Cornell University

1 Introduction

In the United States healthcare system, physicians are afforded wide latitude in their practice choices. While extensive licensure requirements govern the training of physicians, the actual day-to-day practice of medicine is largely left to the physician's choice. Insurance rules govern physician billing, and malpractice litigation corrects demonstrated harm to specific patients; but outside of these situations, physicians are largely able to practice medicine as they see fit.

The freedom afforded physician practice style has social costs and benefits. For example, off-label prescription of pharmaceuticals for uses not approved by the FDA has potential benefits in complicated areas of medicine like cancer care (Liu et al., 2024). However, unrestricted practice can also enable inappropriate treatment choices, whose harms may not be amenable to malpractice litigation if they aren't observable by patients or can't be clearly linked to physician treatment.

In this paper, we examine the economics of physician regulation through studying state medical boards, the institution that regulates doctors. State medical boards are the only regulatory entity that can restrict or prohibit physician treatment. Physicians must attain a state medical license via training and maintain it through appropriate behavior. Medical boards are state agencies that have the unique authority to discipline physicians for low-quality medical care, whether or not it has resulted in legally-demonstrable patient harm. Due to historical data limitations, the effects of these institutions has produced limited empirical evidence.

We analyze the performance of state medical boards in a recent episode of low-quality medical care: high-risk opioid prescribing. Beginning in the early 2000s, changes in medical guidelines encouraged a wider availability of opioid prescriptions for long-term at-home use (Baker (2017); Rummans, Burton and Dawson (2018)). This wider availability created a set of dependent users with increasing tolerance, and supported a black market for purely recreational users (Schnell, forthcoming). While guidelines required physicians to establish a *bona fide* medical relationship and to keep proper records, a small number of physicians were involved in "pill mill" practices that prescribed large volumes without regard to medical need (Kennedy-Hendricks et al. (2016)). For institutional reasons described below, neither insurance rules nor malpractice litigation were directly applicable to inappropriate opioid prescribing. But state medical boards were uniquely situated to address it: state medical boards could discipline opioid prescribing as poor medical care *per se*, and could directly restrict controlled substance prescribing.

Using a novel dataset of medical board disciplinary actions covering 48 states, we examine physician discipline for high-risk opioid prescribing. We match physician disciplinary records into the universe of Medicare claims data from 2008 to 2020 and examine the relationship between physician behavior and regulatory actions. First, we present descriptive statistics on the nature of the harm and lack of regulation. We show that a small share of high prescribers produce a substantial volume of opioid prescriptions. Focusing on "high-risk" prescribers who exceed a volume threshold, we show that fewer than 10,000 physicians supplied more than a quarter of all opioid prescriptions in Medicare. These high-risk prescribers also accounted for an even larger share of high-risk opioid prescribing patterns such as high-MEDs per patient or a large quantity of

opioids dispensed in a single date. The concentrated and observable nature of high-risk prescribing shows the potential value of using regulatory enforcement to limit the behavior of these providers.

Using our unique linkage between administrative medical claims and physician disciplinary records, we report the first estimate of state medical board disciplinary actions taken on high-risk prescribers. Despite the ease of identifying these prescribers, we show that 81% of high-prescribing physicians never faced disciplinary action by any state medical board. Even conditional upon taking any action, state medical boards were generally slow to act, disciplining physicians many years into their harmful behavior. Finally, we apply a large language model to case documents for disciplinary actions targeting high-risk prescribers; we confirm that improper prescribing (commonly, specifically opioid prescribing) was the basis for taking action in a majority of cases. We also determine that fewer than one-fifth of disciplinary actions targeting high-risk prescribers are associated with a law enforcement action – in most cases, SMBs were acting under their own authority.

We next show that state medical boards do have the power to limit harmful behavior. When state medical boards did issue disciplinary actions, they succeeded in changing physicians' opioid prescribing patterns. Using an event study design, we examine how prescribing changed for high-risk prescribers after disciplinary action. We find that even mild punishments (fines, required education, public reprimands) reduced opioid prescribing among high-risk prescribers by two-thirds, with more severe punishments nearly eliminating opioid prescribing.

Lastly, we demonstrate how disciplinary action affected the patients of high-risk opioid prescribers. We compare outcomes among patients who obtained opioids from high-prescribing physicians who did and did not face disciplinary action. Patients who obtained opioids from high-prescribing physicians were substantially less likely to receive opioids following their physician's discipline, showing that they did not successfully pursue prescriptions elsewhere and that regulatory enforcement was effective at eliminating use. Disciplinary action did not interfere with the delivery of standard medical care – patients of disciplined high-risk prescribers show no disruptions in the receipt of non-opioid prescriptions or primary care services. In fact, patients of disciplined high-risk prescribers show clear improvements in health relative to patients whose high-risk prescriber was never disciplined. Just three years after disciplinary action, cumulative mortality among patients whose doctors face enforcement is lower by a full percentage point. Our analysis also clarifies how this improvement is achieved: after discipline, patients are more likely to receive appropriate care. They are twice as likely to have a formal diagnosis of opioid use disorder, and are one-third more likely to initiate medication-assisted treatment for opioid use (i.e., buprenorphine).

Taken together, our results suggest that increased medical board oversight of physician practice would have positive social benefits. State medical boards can use their authority to curtail socially harmful behavior done by a small share of opportunistic physicians, like overprescription. This extends beyond opioids into concerns like the inappropriate use of antipsychotics (Sacarny et al., 2018), or the tradeoff in experimentation vs. adherence to guidelines in medical care (Currie and MacLeod, 2020).

Our work breaks new ground in its analysis of the preeminent physician regulatory institution:

state medical boards. Previous work on state medical boards typically addressed a single state and no prior work linked physician disciplinary records to large-scale administrative data on physician activities (Liu and Hyman, 2021; Khaliq et al., 2005). This linkage enables a new estimate of disciplinary intensity for dangerous care; our findings on the presence and timeliness of discipline for opioid prescribing suggests that high-risk medical care is incompletely regulated by state medical board discipline. These findings contribute to a larger literature on the performance of other regulatory curbs on physician behavior, such as for fraud and malpractice (Eliason et al., 2025; Carroll, Cutler and Jena, 2021; Fang, Li and Xiang, 2025).

Our analysis of patient outcomes provides a framework for assessing the quality–access tradeoff inherent in physician discipline. Among patients who obtained opioids from a high-risk prescriber, disciplinary action appears to set the patient on a new trajectory, in which they consume fewer opioids, are more likely to receive appropriate care for dependence, and have lower mortality rates. In our setting, we find no detectable effects of impaired access, suggesting that boards faced no quality–access tradeoff when considering discipline for high-risk prescribing. Our paper thus contrasts with Atal et al. (2025), who delineate a clear quality-access tradeoff on the physician training margin.

Our work lives in the context of the opioid epidemic. First, by demonstrating the concentration of prescribing in a small number of physicians, we join research stressing the supply-side sources of the opioid epidemic as opposed to changes in consumer demand (Currie and Schwandt, 2021; Ruhm, 2019). Moreover, within the supply side, we focus on a small number of physicians, as opposed to large-scale marketing interventions or broad changes in physician views (Alpert et al., 2022; Wilson et al., 2013). We highlight that state medical boards largely chose not to enforce quality standards on physicians with high-risk prescribing patterns, even though we confirm they had effective tools to do so.

Lastly, our work relates to a broader literature on regulation and occupational licensing, not specific to healthcare. Shavell (1984) considers when society should impose regulations that constrain behavior vs liability regimes that punish actual harms. As in our scenario, Shavell shows regulation is optimal when harms are difficult to correct ex-post, as is the case with the downstream effects of opioid overprescription. Kleiner (2000) highlights the role of occupational licensing in ensuring quality production of professions, but without a specific focus on physicians. Kleiner and Soltas (2023) revisit these questions with regard to welfare, but specifically exclude physicians due to their universal licensure and corresponding lack of variation. Kugler and Sauer (2005) consider the labor economics of physician licensure on employment and earnings.

This paper proceeds as follows. Section 2 provides institutional background on state medical boards as well as the opioid epidemic. Section 3 describes our data collection and matching, and Section 4 presents descriptive statistics on regulatory enforcement. Section 5 describes our methods and results for the effect on physicians, and Section 6 exhibits the methods and results on patients. Section 7 concludes.

2 Background: Physician Regulation and the Opioid Epidemic

2.1 Physician Regulatory Institutions

State medical boards (SMBs) license and regulate physicians in each state. Upon completion of medical education, including both medical school and residency, physicians must pass licensure examinations, regulated by states, to be allowed to practice. SMBs also generally require continuing medical education to maintain licensure (cite: FSMB licensure page) While SMBs vary across states, they typically comprise an appointed membership of physicians, lawyers, and public advocates.

State medical boards' ability to issue disciplinary action gives them unique authority over the practice of medicine. SMBs are the only regulatory institution that can incapacitate medical providers via practice restrictions or delicensure. While SMBs can and do issue disciplinary action for poor medical care, they can take action in the absence of a patient complaint (i.e., on the basis of a coworker report) and can act before patient harm is realized (i.e., in cases of physician infirmity). In addition, state medical boards can evaluate the quality of medical care *per se*, taking action for failures of professional judgment. The disciplinary actions available to state medical boards include fines, mandatory education, or public reprimands; or, in more extreme cases, SMBs can impose restrictions, like a loss of prescribing privileges, or revocation of the physician's license.

In addition to the requirement to maintain SMB licensure, physicians also face two major constraints on their practice: malpractice liability and insurance billing rules.

Physicians who cause demonstrable patient harm face malpractice liability, wherein they can be sued for damages by harmed patients. Generally, the standard is that the harm must be demonstrated and that the doctor is shown to have behaved recklessly and/or in a way that other physicians would not have done. Malpractice rules vary by state. Most physicians carry malpractice insurance, but rules vary by state and by affiliated institution. Kessler (2011) surveys economic research on malpractice liability. While SMBs and malpractice liability both address instances of poor medical care, malpractice liability is limited to *ex post* restitution for patient harms, and relies on patients to recognize those harms.

Billing rules, including fraud liability, further constrain physician behavior. In order to bill insurance for patient care, physicians must abide by coverage rules, and must accurately report patient conditions as well as the procedures or services they conduct. Different insurance plans cover different services, and reimburse at different rates, producing incentives for physicians to misreport; however, doing so may constitute fraud, depending on the nature of the misreporting and the effect on physician payments. These billing and fraud rules ultimately constrain physician's ability to receive payment for a service, not to perform it in the first place; insurance rules do not constrain a physician who chooses to perform a medical service for non-insurance reimbursement.

Lastly, in the context of prescription pharmaceuticals, the abuse of prescribing privileges can constitute drug crimes, investigated and prosecuted by the Drug Enforcement Administration and Federal law enforcement under criminal statutes. This litigation regime is in contrast to regulation, as it generally seeks to hold liable bad behavior as opposed to constrain it from happening in the

first place (Shavell, 1984). Due to high evidentiary standards and high investigation costs, litigation is difficult to scale, leading to regulation to be more effective for broad-based harms (Eliason et al., 2025). As discussed below, several thousand physicians prescribed high volumes of opioids, and while there were dozens of DEA and other criminal actions, they were unable to stem the flow of prescribing due to the high resource costs associated with criminal investigation and prosecution (Soliman, 2025).

2.2 The Opioid Epidemic

The opioid crisis was a major public harm. Beginning around the late 1990s, a small number of physicians began prescribing large numbers of opioids. Over the period 2003 through 2017, nearly half of opioids were supplied by the top 1% of providers (Kiang et al., 2020). This pattern was distinct from the contemporaneous broad-based reorientation towards pain that caused many physicians to increase opioid prescribing somewhat (Alpert et al., 2022). These cases were characterized by a large quantity and variety of opioid ingredients prescribed at each visit. High-volume prescriptions facilitated high-risk consumption patterns among patients, or enabled black market sales for recreation. High-risk consumption patterns and black markets created an addiction and overdose crisis that continues today (Maclean et al., 2022).

The harms caused by high-risk opioid prescribing can be seen as a failure to regulate physician behavior. Why were physicians permitted to use their prescription authority – fundamentally a licensed and regulated authority – to dispense opioids far in excess of patient need? There are institutional reasons why malpractice liability and insurers' anti-fraud rules were not effective against high-risk opioid prescribing. First, physician malpractice liability requires that the harmed individual be the patient of the doctor that treated them. In the case of opioid overprescription and pill mill behavior, some harmed individuals who purchased the pills on the black market did not have a *bona fide* doctor-patient relationship, and were not covered by malpractice liability. For the patients of the high-risk prescribers themselves, case law limited patients' ability to sue prescribers (Fresher, 2018). Anti-fraud investigators typically target higher dollar transactions and cases in which a medical service was not rendered, or was overbilled, rather than being subjectively inappropriate.

2.3 Regulatory Response to the Opioid Epidemic

Our study joins a large literature considering the role of physicians in the opioid epidemic. These papers commonly studied state policies that targeted prescribing, and are reviewed in Maclean et al. (2022). First, "pill mill" laws set minimum standards for pain clinics, which had emerged as frequent prescribers of opioids that were misused or diverted for black-market resale. Second, prescription drug monitoring programs are state databases designed to inform physicians about controlled substance prescriptions a patient may already be receiving from another prescriber. Both state policies were successful in deterring risky opioid prescribing (Buchmueller and Carey, 2018; Kaestner and Ziedan, 2023). Some states instituted direct restrictions on opioid prescriptions, e.g.,

restrictions on the strength and duration of first-time or acute-care prescriptions (Sacks et al., 2021). These laws are noteworthy in that they represent an unusual intervention by state policymakers into the direct practice of medicine.

However, we find only three small studies that address the performance of the SMB disciplinary system for risky opioid prescribing; none linked disciplinary actions with opioid prescribing records nor considered the effects of discipline in a causal framework. Galletly et al. (2024) summarizes 140 opioid-related disciplinary actions. Davis and Carr (2017) find that Florida physicians indicted for controlled substance prescribing typically maintain their license, although they lose it upon conviction. McDonald (2016) describes the administrative process of SMB discipline for a few opioid-related cases in Rhode Island.

The failure of regulatory administration to address harms from opioids leads to questions about the role of state medical boards, which this paper seeks to fill. In the subsequent sections, we examine if, when and how state medical boards addressed physician overprescribing, and their effectiveness when they did so.

3 Data

3.1 State Medical Board Disciplinary Actions

We leverage a novel dataset of all public state medical board actions for 48 states over 2008–22. This dataset reflects collection from 59 unique SMBs, since certain states (e.g., California, Maine, and Oklahoma) license doctors of osteopathy (DOs) separately from doctors of medicine (MDs). For 42 boards we collect actions by digitizing the public announcement issued after each board meeting; for 9 boards we request these reports directly from the board; and for 8 boards we use a license-search web tool to identify licensees with a disciplinary history. We were unable to obtain data from 4 boards (AZ DOs, IA, MT, and SD). For each action we obtain the physician's name and license number, the year and month of the action, and the action taken. In total these data include more than 75,000 actions. We provide further detail about our disciplinary data, including validation of state-year action counts against the National Practitioner Data Bank public use file, in Appendix Appendix B.

We categorize each action using keyword-based string tagging of the disciplinary-action text, augmented by comprehensive manual review, as described in Appendix Appendix C. For example, we identify delicensures by strings matching "revok*", "revocat*", or "surren*" plus phrases such as "refusal to renew" or "precluded from registration". We categorize "stayed revocation" as "probation" since it is analogous to a suspended sentence in criminal justice. We categorize revocations of controlled substance licenses as "restrictions". We differentiate between "adverse actions" and "non-adverse actions", where non-adverse actions include reinstatements after delicensure or suspension or "end/withdrawal" actions that close a case. We collapse adverse action categories into three tiers: (1) non-limiting actions, such as fines, reprimands, or required continuing medical education; (2) practice-restricting actions, such as probation, monitoring, prescribing restrictions,

or suspension; and (3) delicensure actions, such as license revocation or surrender. When an action includes multiple sanctions, we classify it by the most severe. Most actions (96%) contain a string description, and we parse a category for 76%. Uncategorized actions are typically blank or nonspecific (e.g., "Final Order", "Decision").

Our claims datasets record physician services using NPIs, but SMBs record only state license number and name. We therefore crosswalk SMB actions to NPIs through an algorithm detailed in Appendix Appendix B. We fail to match 7.5% of actions to a unique NPI, and thus cannot link these actions to claims-based prescribing records. If any of these unmatched actions should be matched to physicians we flag as high-risk prescribers, we will understate disciplinary intensity.

For approximately 80% of disciplinary actions, we also collect a case document provided by the Board that provides the legal basis for the action: what law was violated, and the evidence that it was violated. In Appendix D we provide a methodology for categorization of the *basis* for SMB action and validate it against a corpus of human coded cases.

3.2 Medicare Claims Datasets for Opioid Prescribing and Patient Outcomes

We use Medicare administrative claims for a random 20% sample for the years 2008-2020. For measures of prescribing, the Part D Event (PDE) dataset records claims for Part D enrollees (approximately 160 million patient-years). We observe physician services (e.g., Evaluation and Management services) for fee-for-service beneficiaries (approximately 100 million patient-years). We also use the Master Beneficiary Summary File (MBSF) and its Chronic Conditions and Cost and Use segments to characterize patients to implement exclusions for opioids prescribed for cancer or end-of-life. Prior to 2013, some PDE records use an alternative physician identifier; a validated harmonization algorithm is presented in Appendix E and then used to impute National Provider Identifiers to these claims. We exclude patients who reside in IA, MT, or SD, the states whose disciplinary records were uncollectable.

Medicare data are ideally suited to this study for two reasons. First, to our knowledge Medicare is the only large-scale dataset of linked medical-prescription records that covers 2008-2013, which are important years for measuring high-risk opioid prescribing given the 2011 opioid prescribing peak (IQVIA Institute for Human Data Science, 2020). In addition, our analysis of patient outcomes after physician discipline benefits from long periods of continuous enrollment, which is common in the Medicare setting.

3.3 Measuring High-Risk Prescribing and Patient Health Outcomes

High-Risk Prescribing We measure high-risk opioid prescribing patterns using prescribing records for Medicare beneficiaries in the PDE. To compare prescriptions of different ingredients and strengths, we convert all opioid prescriptions to morphine-equivalent dosage (MED) using standard methods: crosswalking from ingredient by route of administration for one unit, multiplying by prescription strength, and multiplying by quantity dispensed (Centers for Disease Control and Prevention, 2016).

High doses of opioids are clinically justified for cancer care or end-of-life, and short courses of opioids are a common standard of care after surgeries. To focus attention on high levels of prescribing with poor clinical justification, we exclude prescriptions related to cancer, end-of-life, and post-surgical care prior to forming our physician-year aggregates. We exclude all prescriptions for individual-years flagged with cancer in the MBSF Chronic Conditions segment. Similarly, we exclude all prescriptions for individual-years with any hospice claims identified in the MBSF Cost and Use segment. Following Crespin et al. (2022) to identify post-surgical periods, we exclude prescriptions issued in the 10- or 90-day window after receipt of a Current Procedural Terminology surgical code with a 10- or 90-day global period.

Thus, we form a physician-level[1] measure of high-risk prescribing that focuses on the volume of MEDs supplied outside of cancer, end-of-life, and post surgical indications. A physician's annual MED supplied equals the number of claims times the average MED per claim. As calibration, under our 20% sample a physician supplying 400+ claims is prescribing to one Medicare beneficiary per working hour (i.e., 2000 per year), and an average MED per claim above 1,500 reflects consistently strong prescriptions (i.e., 10 days at 150 MED/day or 30 days at 50 MED/day). Using these thresholds as a guide, we define a "high-risk prescriber" as a physician who supplies more than 600,000 total MEDs in any year 2008-20, approximately 9,000 unique physicians or half a percent of physician-years.

Our data show that opioid prescribing was heavily concentrated. Figure 1 shows the cumulative distribution function of opioid MEDs supplied by physicians in Medicare Part D. Similar to findings in Kiang et al. (2020) for a broader sample, we find that the top percentile of physicians supplied 38% of all opioid MEDs in Medicare Part D over the sample period. While our definition of high-risk prescribing is based on objective volume rather than relative rank, our high-risk definition identifies a subset of the top percentile: approximately 9000 physicians. These high-risk prescribers supplied 26% of the opioid MEDs in Medicare Part D over the sample period. From a regulatory perspective, this indicates that an enforcement regime narrowly targeting these physicians – such as restricting their prescribing authority – could have had large impacts on total prescriptions.

We construct additional measures that capture other dimensions of risk. Several disciplinary case documents describe "extreme dates" on which a physician prescribed amounts inconsistent with clinical motivations; we flag patient-dates on which a patient filled prescriptions totaling 10,000 MEDs from a single physician, e.g., more than 100 days of 100 MEDs per day prescribed in a single encounter. Figure A2 shows that this behavior is vanishingly rare among physicians below our high-risk threshold, but our high-risk physicians average dozens of such instances over the sample period.

Patient Health Outcomes For each patient in the cohort, we collect all utilization from any provider—not only the high-risk prescriber—and construct patient-year outcomes. First, we measure opioid use using indicators for any opioid claim, total opioid claims, and an indicator for

[1]Since our disciplinary data reflects discipline for physicians only, we drop mid-level prescribers from our analysis, e.g. nurse practitioners.

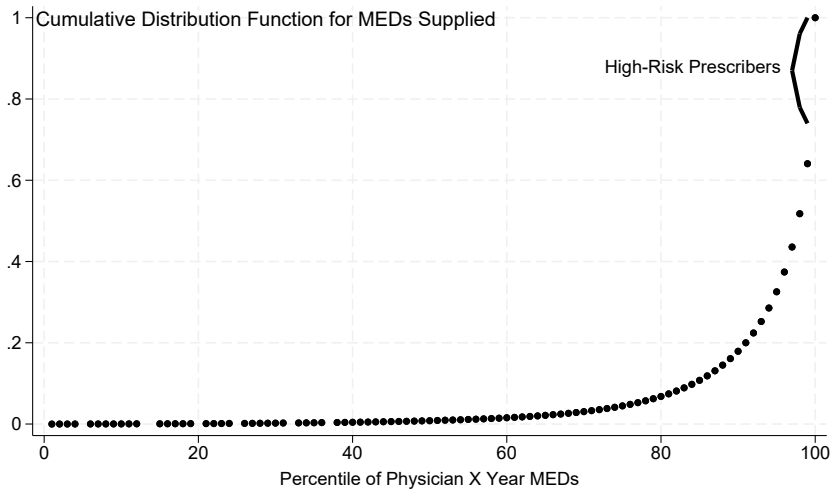


Figure 1: Opioid Prescription Concentration

Note: This figure reports the cumulative distribution function of the share of MEDs supplied by physician-year percentiles. Opioid prescribing is observed in Medicare Part D 2008–2020. Physician-years are ranked by annual MEDs and grouped by percentiles.

heavy opioid use defined as 36,500 MEDs/year (a rough annualization of the measure from Bohnert et al. (2011)). These measures capture whether discipline reduced the opioid supply reaching the prescriber's patients. Second, we measure opioid use disorder (OUD) diagnosis and treatment using OUD diagnoses constructed following Moore and Barrett (2017) as well as an indicator for any buprenorphine claim and total buprenorphine claims; these measures capture whether patients were identified and treated for opioid use disorder following discipline. Next, we measure access to care using indicators for receipt of any evaluation-and-management (E&M) service and counts of non-opioid prescriptions, which would decline if patients were unable to replace the disciplined physician. Finally, we measure all-cause mortality, which captures a well-observed outcome strongly associated with high-risk opioid use (Carey, Jena and Barnett, 2018).

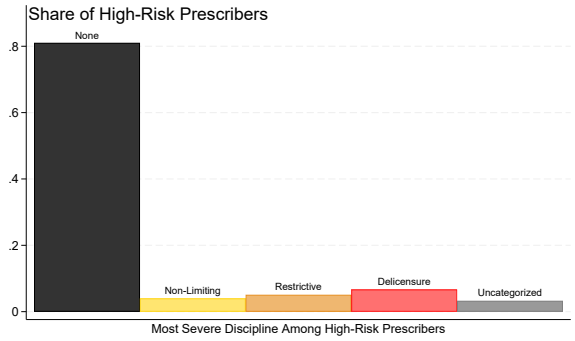
Our measures of OUD diagnosis and E&M services are limited to individuals enrolled in fee-for-service Medicare; all other measures are defined for both fee-for-service and Medicare Advantage enrollees.

4 Descriptive Statistics on Physician Discipline

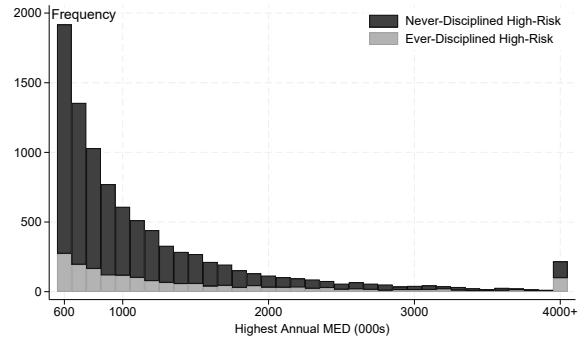
Using our novel linkage between the prescribing behavior of physicians and their state medical board disciplinary records, we present descriptive statistics on the rate and nature of physician discipline by SMBs, the relevant form of regulatory enforcement.

Figure 2 shows the actual response of state medical boards to high-risk prescribing. In Panel A, we report that 81% of high-risk prescribers never faced any disciplinary action. In Panel B, we report the distribution of the highest MEDs supplied in any year among our high-risk prescribers.

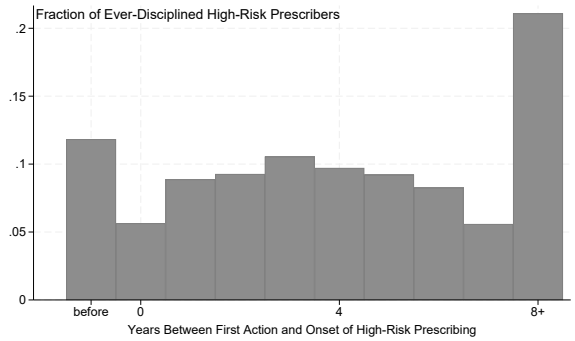
The share of high-risk prescribers who ever faced any disciplinary action is shaded. Enforcement was heavily incomplete, and also not well-targeted, with most enforcement occurring in the lower range of prescribing, and many super-top prescribers unaffected. Panel C shows the histogram of the number of years between the onset of high-risk prescribing behavior (e.g., the calendar year when more than 600,000 MEDs were supplied) and the year of enforcement, if any. The distribution is right-tailed, with most enforced prescribers receiving actions many years after the onset of high-risk prescribing.[2] This shows that, even among physicians who faced some enforcement, it was rarely timely, with the behavior persisting for years before action was taken. Taken together, these figures show that the SMB response was inadequate to stem the flow of opioid harm.



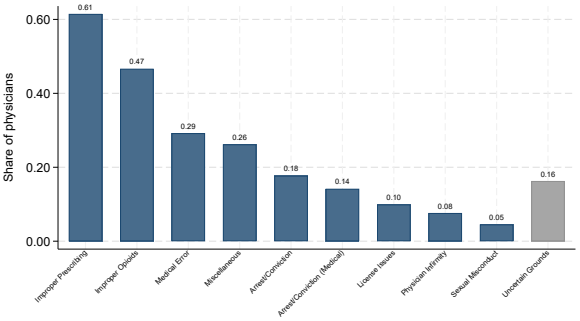
(a) High-Risk Prescribers and Disciplinary Action



(b) Targeting of Enforcement by MED Volume



(c) Years to Enforcement



(d) Basis for Discipline Among High-Risk Prescribers

Figure 2: Incompleteness of Opioid Enforcement

Note: Panel (a) reports the presence and severity of disciplinary actions taken on high-risk physicians (as defined in Section 3.3). Severity defined as most severe discipline observed. Panel (b) reports a histogram of the highest-annual MEDs ever supplied by each high-risk physician; the height of the gray region denotes the share of that MED level that was ever-disciplined. Panel (c) reports a histogram of the years between the first year of high-risk prescribing and the first disciplinary action observed among ever-disciplined high-risk physicians. Panel (d) reports the bases of disciplinary actions taken against high-risk physicians, using the methodology described in Appendix D.

Drawing on our LLM-assigned basis classification, we can further characterize the disciplinary actions taken on high-risk physicians (Figure 2d). Note that these categories are not mutually exclusive. 61% of high-risk physicians faced discipline for improper prescribing; in most such cases

[2]Some prescribers faced discipline before the onset of high-risk prescribing, either for more minor drug-related offenses or, potentially, other forms of misconduct.

the case specified improper prescribing of opioids. The remainder primarily faced a disciplinary action for medical error, a basis that may arise when the state medical board establishes that the physician did not have a *bona fide* doctor-patient relationship or did not keep proper medical records for the patient. The basis classification also demonstrates that only 14 to 18 percent of actions taken against high-risk opioid prescribers were predicated on an arrest or conviction (either in general or specifically for a crime related to medical practice). Thus, SMB disciplinary enforcement was ∼5 times more common than criminal enforcement for high-risk opioid prescribing.

5 Effects of Enforcement on Physicians

5.1 Methods

We now turn to examining the causal effect of disciplinary actions on physician behavior and the health of their patients.

We estimate an event study among high-risk opioid prescribers, comparing those who faced disciplinary action by state medical boards to those who did not. This design allows us to identify the effects of regulatory enforcement on physicians' opioid prescribing behavior.

Specifically, we estimate the dynamic difference in difference model:

$$y_{it} = \sum_{\tau \neq -2} \beta_\tau \mathbb{I}_{it}(\tau) + \delta_t + \zeta_i + \epsilon_{it} \tag{1}$$

For outcome variable y, physician i in year t, where τ is the years relative to enforcement. $\mathbb{I}_{it}(\tau)$ is an indicator for falling in relative year τ. To ensure sufficient balance in pre- and post-period event coefficients, we limit to high-risk prescribers whose first action occurred in the years 2012 to 2018. We cluster ϵ at the prescriber level. Figure A3 uses the Sun and Abraham (2021) estimator instead of two-way fixed effects, returning nearly identical results.

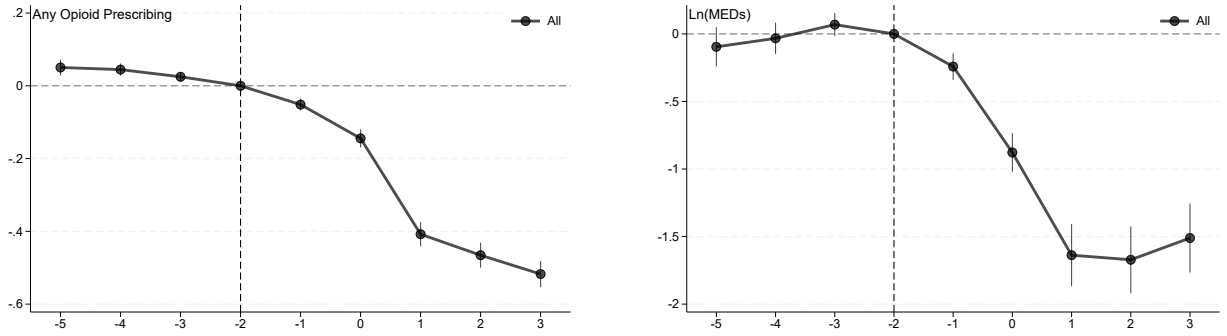
In our enforcement event studies, we use year -2 as the omitted period. Our disciplinary action data record the data of the official action taken by the board, not the date an investigation was initiated or the date when a physician would have become aware of that investigation. The timing of the investigation onset (or when the physician learned of it) would be the ideal treatment timing definition, since that is when the physician may begin responding to discipline or the potential for it. As we are unable to observe investigation onset, we use 2 years before the final agency action as our baseline year, to account for any treatment that may actually begin in year -1 relative to the documents we observe. For two states, California and Florida, we are able to observe investigation onset, which allows us to empirically validate the difference between investigation onset and official action, i.e. an estimate of that gap in other states. Appendix Figure A1 shows the distribution of this time gap and validates our use of baseline year -2. In California and Florida, we define treatment timing to be the first action other than investigation onset.

We examine the effect of a disciplinary action of any severity and examine outcomes by severity level: non-limiting (reprimand, fine, education), restrictive (suspension, probation, or restrictions

on practice), and delicensure. When considering severity, we define treatment timing using the first action taken against the physician, while defining the severity level as the most severe action experienced.

5.2 Results

Figure 3 shows the event study plot of the β_τ coefficients from estimating Equation 1. This shows the treatment effect of state medical board actions on the prescribing behavior of physicians. Figure Figure 3a (left) shows the extensive margin effect on any opioid prescription, which exhibits a roughly 50% reduction in prescription of any opioid. Figure Figure 3b shows the effect on log(MED), i.e. morphine equivalent doses, and shows a roughly 78% reduction in dosage by a year following regulatory action. Neither event study shows evidence of non-parallel pre-trends.



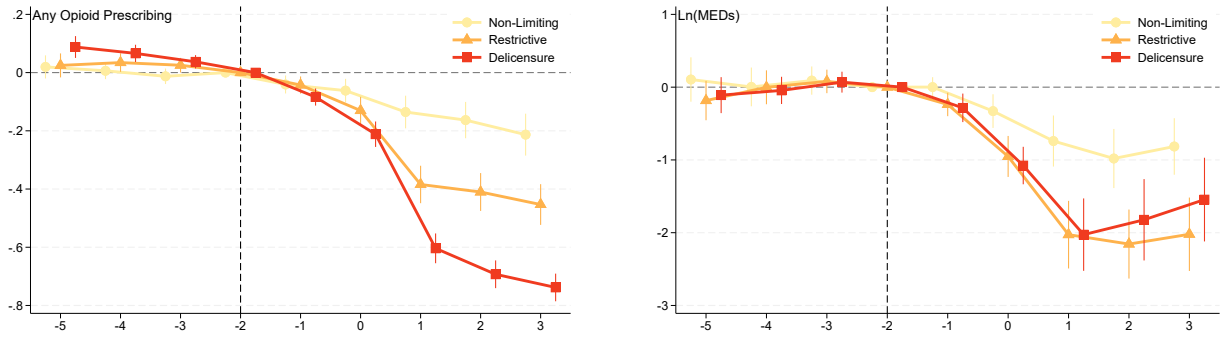
(a) Effect on Prescription of Any Opioid

(b) Effect on Log MED | Any

Figure 3: Effects of Physician Discipline

Note: This figure reports the β_τ event coefficients estimated in Eq. 1. The estimation sample reflects high-risk physicians with first disciplinary actions in the years 2012 to 2018 or never disciplined. The outcomes, measured in the years 2008 to 2020 in Medicare Part D, are an indicator for any opioid prescribing and the log of the MEDs supplied by the physician (in which zero MED observations are left missing). Standard errors are clustered at the physician level.

Next, we consider the heterogeneous effects of different forms of SMB actions. Figure 4 compares the effects on physicians who face non-limiting prescriptions, restrictive prescriptions, and full delicensure. As expected, stricter regulatory action produces greater effects on prescriptions, both on the extensive and intensive margins. Even a mild non-limiting action reduces the volume of opioid MEDs by two-thirds.



(a) Effect on Prescription of Any Opioid

(b) Effect on Log MED

Figure 4: Heterogeneous Effect of Disciplinary Actions by Severity

Note: This figure reports the β_τ event coefficients estimated in Eq. 1 separately by the severity of disciplinary actions. Disciplined physicians are categorized by the most severe action. Each set of coefficients reflects high-risk physicians disciplined with that severity in the years 2012 to 2018 or never disciplined. The outcomes, measured in the years 2008 to 2020 in Medicare Part D, are an indicator for any opioid prescribing and the log of the MEDs supplied by the physician (in which zero MED observations are left missing). Standard errors are clustered at the physician level.

Overall, these results indicate that SMBs were able to limit the harmful behavior of physicians they targeted, indicating the limited effectiveness of SMBs overall was not due to an inability to combat this behavior, but instead a lack of expansive regulatory action.

6 Effects of Enforcement on Patient Health & Health Care

Next, we consider the effect of physician discipline by an SMB on the patients seen by that physician. Mapping from physician outcomes to patient outcomes is necessary for a normative analysis, because the removal of harmful physicians from the system may not have the desired effect if patients either lose access to other valuable care, or if patients substitute to similarly harmful physicians. Overall, we are able to reject these concerns and find that enforcement against physicians is protective for their patients.

6.1 Methods

First, for each high-risk opioid prescriber ever disciplined by a state medical board, we form a patient cohort c comprising the patients who had received an opioid prescription from this physician in year -2 before discipline. This allows us to examine a stable cohort of patients defined by pre-treatment physician-patient matching, and to follow those patients after their prescriber receives discipline.

Then, as a comparison group, we consider a similar cohort for never-disciplined high-risk prescribers. Because never-treated physicians do not have a treatment year, we assign a counterfactual enforcement year for all control physicians (such that the empirical distribution of treatment years matches that of ever-disciplined), and similarly fix a cohort of patients receiving opioids 2 years before this counterfactual treatment year.

Table 1: Patient Characteristics by Disciplinary Status of High-Risk Physician

	Never	Pre	Post
Unique Patients	136,372	38,763	38,763
Unique Physicians	3,276	873	873
Disability Entitlement	.679	.73	.769
Dual Enrollment	.485	.55	.599
Age less than 65	.502	.62	.563
Age 65 to 79	.384	.31	.34
Age 80 plus	.114	.0695	.0966
Female	.629	.583	.58
$\mathbb{1}$(Any Opioid Claims in a Year)	.867	.907	.817
Average Opioid Claims in a Year	11.6	13.4	11
$\mathbb{1}$(Heavy Opioid Use)	.173	.247	.185

Note: This table reports characteristics of patients studied in the estimation of Eq. 2: individuals who obtained opioids from a high-risk physician two years before true or counterfactual disciplinary action, restricting to true or counterfactual disciplinary actions in the years 2012 to 2018.

Then, at the patient level, we estimate

$$y_{ict} = \sum_{\tau \neq -2} \beta_\tau \mathbb{I}_{it}(\tau) + \delta_{ct} + \zeta_i + \epsilon_{ict} \tag{2}$$

for outcome variable y, patient i who belongs to cohort c in year t, where τ is the years relative to that patient's physician receiving disciplinary action. $\mathbb{I}_{it}(\tau)$ is an indicator for falling in relative year τ. We included fixed effects by cohort$\times$year to ensure comparison within treated and control providers whose cohorts are defined at the same time. This within-cohort comparison ensures our treatment effects are not biased by attrition, i.e., patients who are present in event time -2 but depart Medicare after that time.

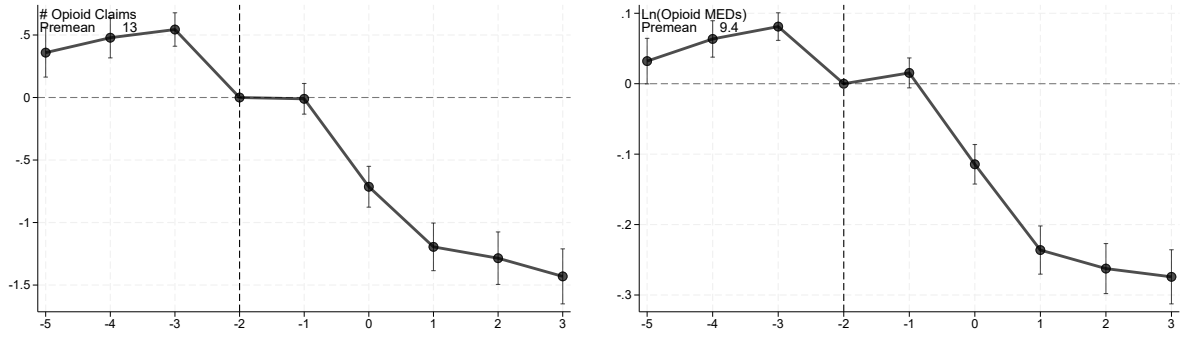
Table 1 describes the patients we use in estimation of Eq. 2. When restricting to physicians who received (true or counterfactual) discipline in the years 2012 to 2018, we study 3075 high-risk physicians who were never disciplined and 822 who were. These physicians distributed opioids to more than 160,000 patients in the two years prior to (true or counterfactual) discipline. About half of the patients were originally entitled to Medicare by reason of disability, and about half are dually enrolled in Medicaid, suggesting a high-morbidity low-income population.

6.2 Results

Figure 5 shows the patient-level effects of physician discipline on patient opioid prescription, as measured in number of annual opioid claims. In Panel A, we find a sharp decline in opioid receipt, about 1.5 claims on a baseline mean of 14, or about a 10% reduction. In Panel B, Log(MED), we find an even stronger reduction of about 20-25%. [3]. Note that this is not an obvious result,

[3]By construction, all patients included in this panel received opioids in year -2, and therefore we cannot use this design to examine the extensive margin pretrends

because it indicates patients are not simply switching to another physician and receiving the same amount of opioids, but rather are receiving fewer total opioids after their previous prescriber was disciplined.

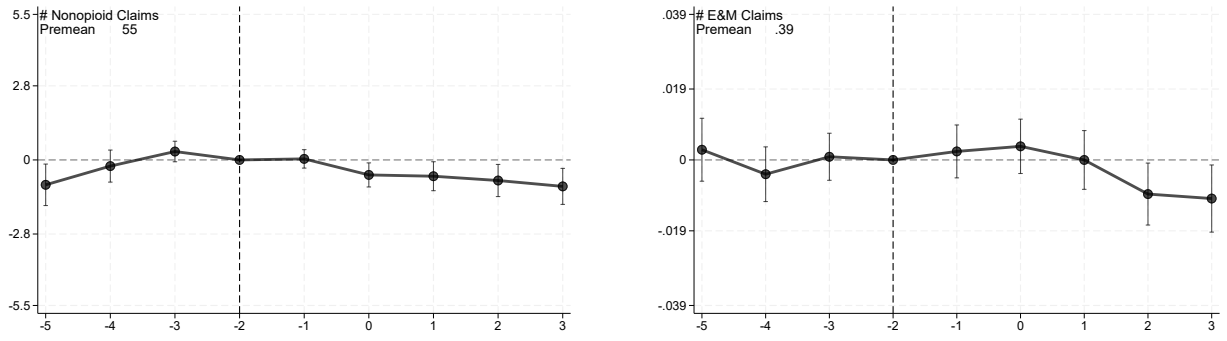


(a) # Opioid Claims **(b)** Ln(MEDs) | Any

Figure 5: Patients Obtained Fewer Opioids Following Physician Discipline

Note: This figure reports the β_τ event coefficients estimated in Eq. 2. The estimation sample reflects patients who obtained opioids from high-risk physicians two years prior to discipline, or two years prior to a randomly-assigned counterfactual year for never-disciplined high-risk physicians. The outcomes are measured at the patient level for the years 2008-2020 and reflect all care provided (not just that of the high-risk physicians). Panel (a) reports on the number of opioid claims filled by the patient and panel (b) reports the log of the MEDs equivalent (in which zero MED observations are left missing). Standard errors are clustered at the physician level.

Next, we consider whether patients lose access to valuable non-opioid care. Figure 6 plots the receipt of non-opioid claims (left) and total evaluation and management events (right), which reflect basic primary care visits. Both show no effect, indicating that patients, although they receive markedly fewer opioids, continue to receive other forms of care and are not separated from the healthcare system despite their prescribing physician facing discipline, including in many cases loss of license.

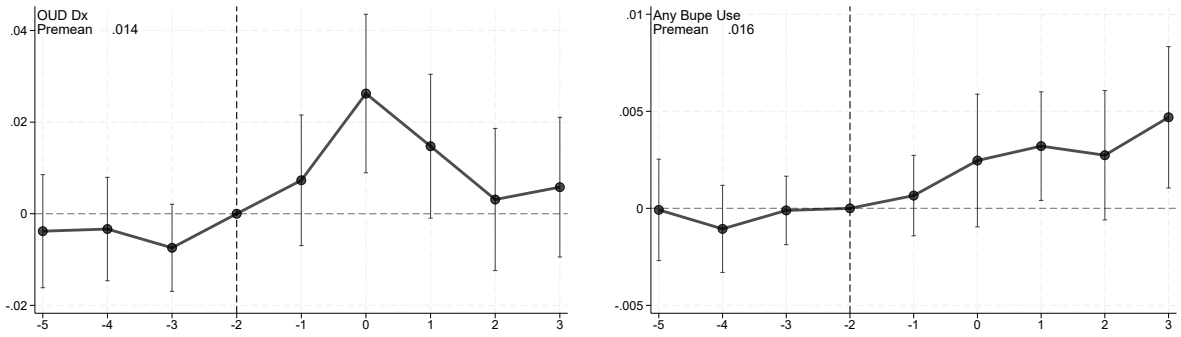


(a) Non-Opioid Claims **(b)** E&M Events

Figure 6: Patient Access to Non-Opioid Care was Unaffected by Physician Discipline

Note: This figure reports the β_τ event coefficients estimated in Eq. 2. The estimation sample reflects patients who obtained opioids from high-risk physicians two years prior to discipline, or two years prior to a randomly-assigned counterfactual year for never-disciplined high-risk physicians. The outcomes are measured at the patient level for the years 2008-2020 and reflect all care provided (not just that of the high-risk physicians). Panel (a) reports on the number of non-opioid claims filled by the patient and panel (b) reports the number of E&M claims (among full-year enrollees in fee-for-service Medicare). Standard errors are clustered at the physician level.

Next, Figure 7 shows that these patients begin to receive appropriate care for the harms of opioids. Figure 7a shows that patients are about 2 percentage points more likely to receive a diagnosis of opioid use disorder, on a baseline mean of 1.4%, or nearly double. Similarly, Figure 7b shows that patients are substantially more likely to receive a prescription for buprenorphine, a medication used to treat opioid disorder, about a one-third increase over the pre-period mean. Overall, it appears that following physician discipline, the patients of those physicians begin to receive more appropriate care.



(a) Opioid Use Disorder Diagnosis **(b)** Any Buprenorphine Claims

Figure 7: Patients Receive Addiction Care

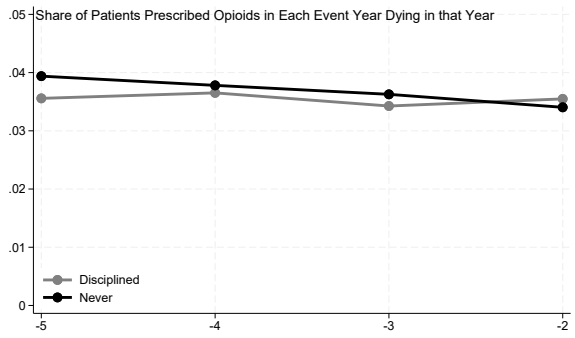
Note: This figure reports the β_τ event coefficients estimated in Eq. 2. The estimation sample reflects patients who obtained opioids from high-risk physicians two years prior to discipline, or two years prior to a randomly-assigned counterfactual year for never-disciplined high-risk physicians. The outcomes are measured at the patient level for the years 2008-2020 and reflect all care provided (not just that of the high-risk physicians). Panel (a) reports on the presence of a diagnosis code for opioid use disorder (among full-year enrollees in fee-for-service Medicare) and panel (b) reports on filling any claims for buprenorphine. Standard errors are clustered at the physician level.

Lastly, we examine the effects of physician discipline on patient mortality. Substantial work on the opioid epidemic has considered that reduction in opioid prescription could harm patients, if they turn instead toward illegal alternatives such as heroin, which could have higher likelihoods of overdose or other harm. In contrast, it is possible that patients who lose access to high-opioid-prescribing physicians and receive other forms of care are protected and therefore experience lower mortality. We empirically examine this question, which bears on the welfare consequences of SMB physician discipline.

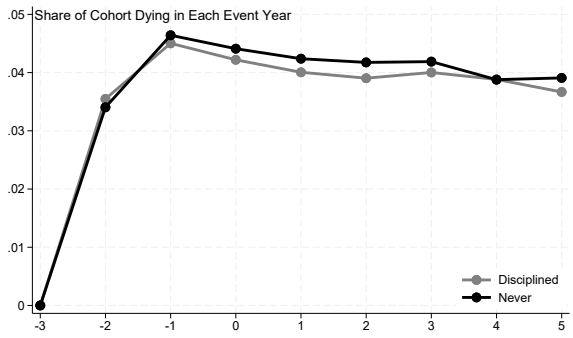
Figure 8 shows the effects of physician discipline on patient mortality. Importantly, traditional difference-in-differences is challenging for assessing mortality, because death is an absorbing state. By construction, patients in our cohort, who obtain opioids from a high-risk prescriber in year -2, are not yet dead in years -5, -4, and -3. To address these questions, in Panel A we leave aside the cohort construction and instead ask, between years -6 and -2 whether patients who received opioids in each given year died that same year, comparing will-be-disciplined physicians to their never-disciplined controls. We find no evidence of difference between these groups, suggesting parallel paths in the mortality of patients seeing will-be-disciplined and never-disciplined high-risk prescribers prior to discipline.

Next, we consider the treatment effect of discipline on mortality. Figure 8b shows the share of the cohort – that is, once again patients who received opioids in year -2 – dying in each event year. By construction, mortality in year -3 is 0 because these patients are alive in period -2. In year -2 and -1, treatment vs control mortality are nearly identical. Beginning in year 0, a gap opens, with disciplined prescribers' patients showing noticeably lower mortality in every period from 0 to +6. Figure 8c shows this in cumulative mortality, where the gap is widening over time. At 5 years beyond treatment, treated physicians have 1.1% lower cumulative mortality than untreated physcians (31.7% vs 32.8%). Notably, this is a patient population with high baseline mortality, at around 4-5% annually. Finally, Figure 8d presents the cohort annual mortality as an event study, comparing treated to controls, i.e. the specification in panel (b) but as a difference-in-differences. Note that this specification omits physician fixed effects.[4] Recognizing the limitations on pre-trend analysis (addressed by panel a), the event study figure shows that physician regulatory discipline causes a reduction in annual mortality of about 0.5%.

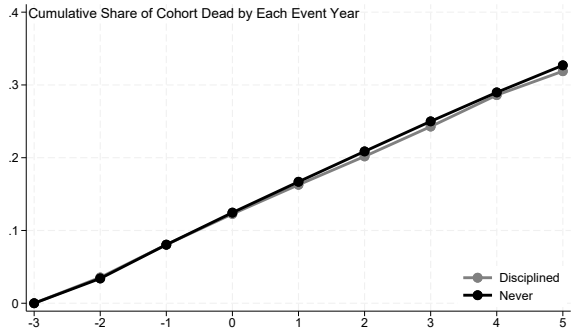
[4]Physician fixed effects are not appropriate for this analysis because of the use of patient cohorts. Since patient cohorts are defined at event year -2, mortality is zero by construction for every physician at every earlier event time. In this context, there is no baseline physician heterogeneity in mortality rates that a physician fixed effect might address – instead, baseline mortality rates are zero by construction for all physicians. A within-physician mean over the full sample would average over preperiod zeroes and postperiod realized mortality rates, and incorrectly redistribute preperiod "deviations" from those means into preperiod event coefficients. Thus, we omit physician fixed effects, retain the cohort-year × calendar-year fixed effect, and report the disciplined/never-disciplined difference by event year.



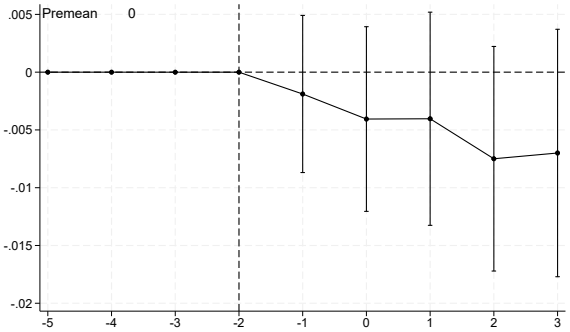
(a) Preperiod: Share of High-Risk Physicians' Patients Dying in Each Event Time



(b) Postperiod: Share of High-Risk Physicians' Patient Cohort Dying in Each Event Time



(c) Postperiod: Cumulative Share of High-Risk Physicians' Patient Cohort Dying by Each Event Time



(d) Event Study: Physician-Level Cumulative Mortality

Figure 8: Effect of Physician Discipline on Patient Mortality

Note: This figure reports mortality outcomes for patients of high-risk physicians. Panel (a) reports the same-year mortality rates of patients who obtained opioids from a never- (black) or ever-disciplined (gray) high-risk prescriber in the years prior to true or counterfactual discipline. Panel (b) reports the mortality rates of the fixed cohort of patients who obtained opioids from a high-risk physician two years prior to true or counterfactual discipline in the event year measured on the x-axis. Panel (c) reports on the cumulative mortality rate of the same fixed patient cohort by the event year measured on the x-axis. Panel (d) estimates an event study on same outcome depicted in Panel (c), with standard errors clustered on the physician.

As a final point, we consider the total welfare effect of this reduction in mortality. Never-disciplined top prescribers treated 228,740 patients in counterfactual event year -2. Therefore, if each of these providers had been treated, the roughly 1% reduction in 5-year cumulative total mortality would correspond to roughly 2,200 lives saved. At a value of a statistical life of roughly \$1 Million, this implies a potential welfare gain of \$2.2 Billion. This figure underestimates the total benefit of greater regulatory intervention, including the potential life years saved if treated physicians had been treated more promptly; the greater quality of life for those who remain alive but receive addiction care; as well as the positive externality of fewer diverted opioids from patients who sold, rather than consumed, the drugs prescribed.

7 Conclusion

The wide latitude afforded physicians has permitted unregulated prescribing of harmful opioids, a major contributor to the opioid crisis. We show that this behavior was concentrated among a small portion of exceptionally harmful physicians, and that regulatory enforcement was underutilized. However, that was not due to ineffectiveness, because physician regulation shows a strong positive benefit in reducing opioid utilization while promoting appropriate patient care and reducing mortality.

Our results speak to the benefit of potentially constraining physician practice choices, but in the prescribing of harmful substances but also in other potentially harmful activities, like risky surgical interventions. Our work does not imply that all such regulations would be valuable, but rather that, directionally, greater ex-ante oversight into harmful medical practices can improve patient welfare.

References

Alpert, Abby, William N. Evans, Ethan M. J. Lieber, and David Powell. 2022. "Origins of the Opioid Crisis and Its Enduring Impacts." *The Quarterly Journal of Economics*, 137(2): 1139–1179.

Atal, Juan Pablo, Tomás Larroucau, Pablo Muñoz, and Cristóbal Otero. 2025. "Physicians' Occupational Licensing and the Quantity-Quality Trade-off." Working paper.

Baker, David W. 2017. "History of The Joint Commission's Pain Standards: Lessons for Today's Prescription Opioid Epidemic." *JAMA*, 317(11): 1117.

Bohnert, Amy S. B., Marcia Valenstein, 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." *JAMA*, 305(13): 1315–1321.

Buchmueller, Thomas C., and Colleen Carey. 2018. "The Effect of Prescription Drug Monitoring Programs on Opioid Utilization in Medicare." *American Economic Journal: Economic Policy*, 10(1): 77–112.

Carey, Colleen M., Anupam B. Jena, and Michael L. Barnett. 2018. "Patterns of Potential Opioid Misuse and Subsequent Adverse Outcomes in Medicare, 2008 to 2012." *Annals of Internal Medicine*, 169(3): 181–190.

Carroll, Caitlin, David M. Cutler, and Anupam Jena. 2021. "How do Physicians Respond to Malpractice Allegations? Evidence from Florida Emergency Departments."

Centers for Disease Control and Prevention. 2016. "Calculating Total Daily Dose of Opioids for Safer Dosage." Accessed June 30, 2026.

Crespin, Daniel J., Ashley M. Kranz, Teague Ruder, Ateev Mehrotra, and Andrew W. Mulcahy. 2022. "Claims-Based Reporting of Post-Operative Visits for Procedures with 10- or 90-Day Global Periods: Updated Results Using Calendar Year 2019 Data." *Rand Health Quarterly*, 9(4): 7.

Currie, Janet, and Hannes Schwandt. 2021. "The Opioid Epidemic Was Not Caused by Economic Distress but by Factors That Could Be More Rapidly Addressed." *The ANNALS of the American Academy of Political and Social Science*, 695(1): 276–291.

Currie, Janet M., and W. Bentley MacLeod. 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica*, 88(3): 847–878.

Davis, Corey S., and Derek H. Carr. 2017. "Self-Regulating Profession? Administrative Discipline of "Pill Mill" Physicians in Florida." *Substance Abuse*, 38(3): 265–268.

Eliason, Paul, Riley League, Jetson Leder-Luis, Ryan C. McDevitt, and James W. Roberts. 2025. "Ambulance Taxis: The Impact of Regulation and Litigation on Health-Care Fraud." *Journal of Political Economy*, 133(5): 1661–1702.

Fang, Hanming, Ming Li, and Jia Xiang. 2025. "Fear and Risk Perception: Understanding Physicians' Dynamic Responses to Malpractice Lawsuits." National Bureau of Economic Research Working Paper 34115.

Fresher, Samuel. 2018. "Opioid Addiction Litigation and the Wrongful Conduct Rule." *University of Colorado Law Review*, 89(4): 1311.

Galletly, Carol L., Elsa A. Christenson, Jenna Ohlrich, and Julia Dickson-Gomez. 2024. "Out of Bounds: Physician Licensing Board Disciplinary Cases Related to Opioid Prescribing." *Journal of Law, Medicine & Ethics*, 52(3): 679–689.

IQVIA Institute for Human Data Science. 2020. "Prescription Opioid Trends in the United States: Measuring and Understanding Progress in the Opioid Crisis." IQVIA Institute for Human Data Science. Accessed July 6, 2026.

Kaestner, Robert, and Engy Ziedan. 2023. "Effects of Prescription Opioids on Employment, Earnings, Marriage, Disability and Mortality: Evidence from State Opioid Control Policies." *Labour Economics*, 82: 102358.

Kennedy-Hendricks, Alene, Matthew Richey, Emma E. McGinty, Elizabeth A. Stuart, Colleen L. Barry, and Daniel W. Webster. 2016. "Opioid Overdose Deaths and Florida's Crackdown on Pill Mills." *American Journal of Public Health*, 106(2): 291–297.

Kessler, Daniel P. 2011. "Evaluating the Medical Malpractice System and Options for Reform." *Journal of Economic Perspectives*, 25(2): 93–110.

Khaliq, Amir A., Hani Dimassi, Chiung-Yu Huang, Lutchmie Narine, and Raymond A. Smego. 2005. "Disciplinary action against physicians: who is likely to get disciplined?" *The American Journal of Medicine*, 118(7): 773–777.

Kiang, Mathew V., Keith Humphreys, Mark R. Cullen, and Sanjay Basu. 2020. "Opioid prescribing patterns among medical providers in the United States, 2003–17: a retrospective, observational study." *BMJ*, 368: l6968.

Kleiner, Morris M. 2000. "Occupational Licensing." *Journal of Economic Perspectives*, 14(4): 189–202.

Kleiner, Morris M, and Evan J Soltas. 2023. "A Welfare Analysis of Occupational Licensing in U.S. States." *The Review of Economic Studies*, 90(5): 2481–2516.

Kugler, Adriana D., and Robert M. Sauer. 2005. "Doctors without Borders? Relicensing Requirements and Negative Selection in the Market for Physicians." *Journal of Labor Economics*, 23(3): 437–465.

Liu, Jing, and David A. Hyman. 2021. "Physician Licensing and Discipline: Lessons From Indiana." *Journal of Empirical Legal Studies*, 18(3): 629–659.

Liu, Ruishan, Lisa Wang, Shemra Rizzo, Marius Rene Garmhausen, Navdeep Pal, Sarah Waliany, Sarah McGough, Yvonne G. Lin, Zhi Huang, Joel Neal, Ryan Copping, and James Zou. 2024. "Systematic analysis of off-label and off-guideline cancer therapy usage in a real-world cohort of 165,912 US patients." *Cell Reports Medicine*, 5(3): 101444.

Maclean, Johanna Catherine, Justine Mallatt, Christopher J. Ruhm, and Kosali Simon. 2022. "The Opioid Crisis, Health, Healthcare, and Crime: A Review of Quasi-Experimental Economic Studies." *The ANNALS of the American Academy of Political and Social Science*, 703(1): 15–49.

McDonald, James V. 2016. "Rhode Island Board of Medical Licensure and Discipline: Illustration of the Disciplinary Process as It Pertains to Cases Involving Opioid Prescribing." *Journal of Opioid Management*, 12(5): 355–359.

Moore, Brian J., and Marguerite L. Barrett. 2017. "Case Study: Exploring How Opioid-Related Diagnosis Codes Translate From ICD-9-CM to ICD-10-CM." Agency for Healthcare Research and Quality, Rockville, MD.

NBER. 2024. "National Provider Identifier (NPI) to State License Crosswalk." *https://www.nber.org/research/data/national-provider-identifier-npi-state-license-crosswalk*, Accessed March 10, 20234.

Ruhm, Christopher J. 2019. "Drivers of the Fatal Drug Epidemic." *Journal of Health Economics*, 64: 25–42.

Rummans, Teresa A., M. Caroline Burton, and Nancy L. Dawson. 2018. "How Good Intentions Contributed to Bad Outcomes: The Opioid Crisis." *Mayo Clinic Proceedings*, 93(3): 344–350.

Sacarny, Adam, Michael L. Barnett, Jackson Le, Frank Tetkoski, David Yokum, and Shantanu Agrawal. 2018. "Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial." *JAMA Psychiatry*, 75(10): 1003–1011.

Sacks, Daniel W., Alex Hollingsworth, Thuy Nguyen, and Kosali Simon. 2021. "Can Policy Affect Initiation of Addictive Substance Use? Evidence from Opioid Prescribing." *Journal of Health Economics*, 76: 102397.

Schnell, Molly. Forthcoming. "Physician Behavior in the Presence of a Secondary Market: The Case of Prescription Opioids." *Econometrica*.

Shavell, Steven. 1984. "Liability for Harm versus Regulation of Safety." *The Journal of Legal Studies*, 13(2): 357–374.

Soliman, Adam. 2025. "Disrupting Drug Markets: The Effects of Crackdowns on Rogue Opioid Suppliers." *American Economic Journal: Economic Policy*, 17(4): 165–191.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

Wilson, Hilary D., Elizabeth J. Dansie, Myoung S. Kim, Bruce L. Moskovitz, Wing Chow, and Dennis C. Turk. 2013. "Clinicians' Attitudes and Beliefs About Opioids Survey (CAOS): Instrument Development and Results of a National Physician Survey." *The Journal of Pain*, 14(6): 613–627.

Physician Regulation: Evidence from the Opioid Epidemic

by Colleen Carey, Jetson Leder-Luis, and Chloe Smith

APPENDIX

A Supplemental Figures and Tables **26**

B Data Appendix **29**
B.1 Collection of State Medical Board Disciplinary Data 29
B.2 Validation of Data Collection Methods 31
B.3 Linking Disciplinary Records to National Provider Identifiers 32

C Disciplinary Action Categorization **37**
C.1 Action Type Classification 37
C.2 Severity Classification 37
C.3 Corrections and Refinements 40
C.4 Clean Delicensures 41

D Text Analysis to Categorize *Basis* for SMB Actions **41**
D.1 Preparing Documents and Data for Coding 43
D.2 The Potential Bases 43
D.3 Human Coding by Undergraduate Research Assistants 44
D.4 Preparing for LLM Text Analysis 45
D.5 Prompt Engineering 46
D.6 Validation 48
D.7 Full Run 49

E Hamonizing Prescriber Identifiers in Medicare Part D **49**
E.1 Prescriber Identifiers in the Prescription Drug Event Dataset 49
E.2 Matching Algorithm Derived From Fully-Harmonized Part D Claims 50

A Supplemental Figures and Tables

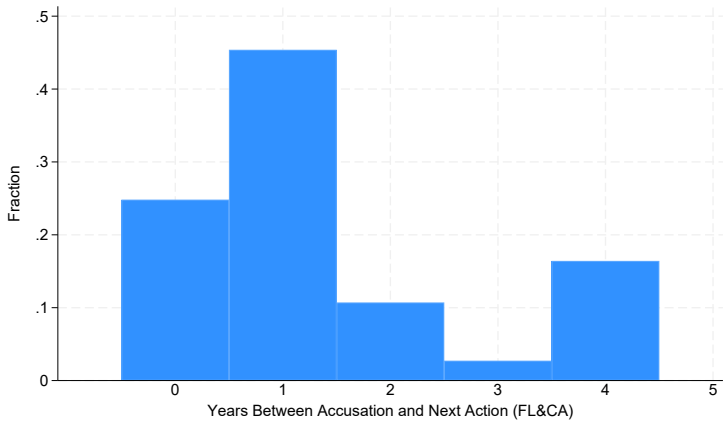


Figure A1: Years Between Investigation Onset and Adverse Action, CA&FL

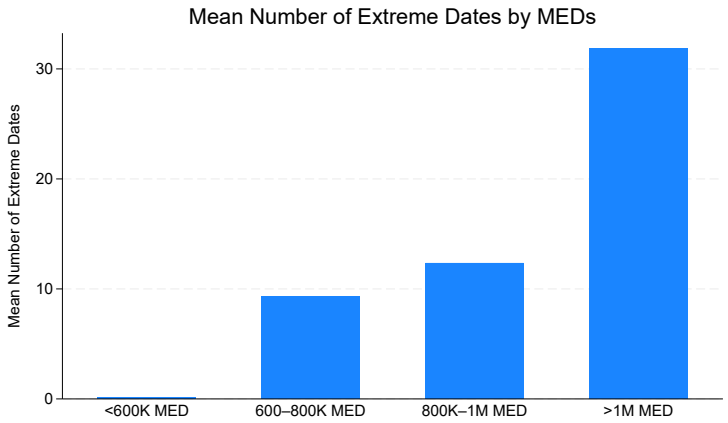
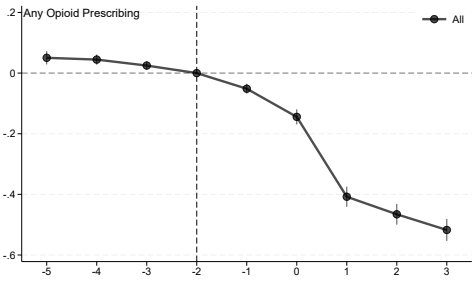
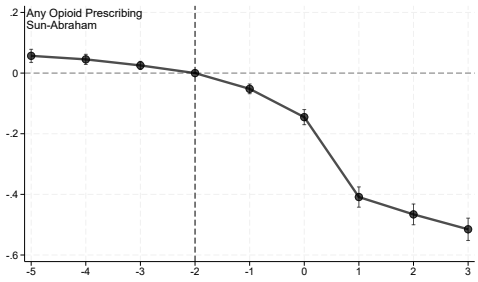


Figure A2: Average Number of "Extreme Dates" Recorded for Physician-Years by MEDs Supplied

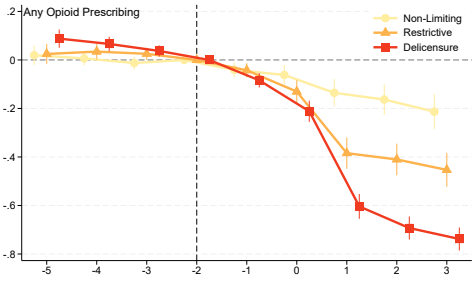
Note: "Extreme Date" defined as prescribing more than 10000 MEDs on a single date to a single beneficiary. Y-axis reports average number of such instances for physician-years by MEDs supplied by that physician-year,



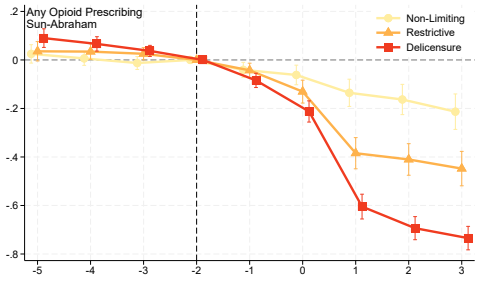
(a) TWFE: 1(opioid prescribing)



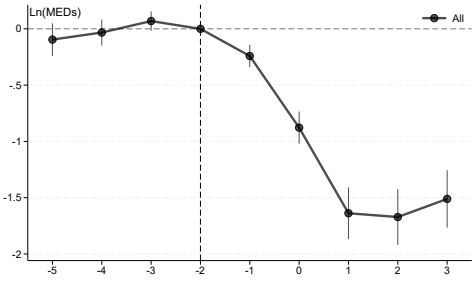
(b) Sun & Abraham: 1(opioid prescribing)



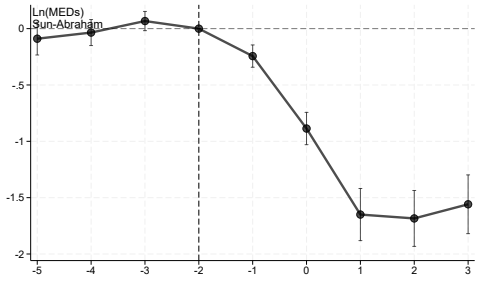
(c) TWFE: 1(opioid prescribing)



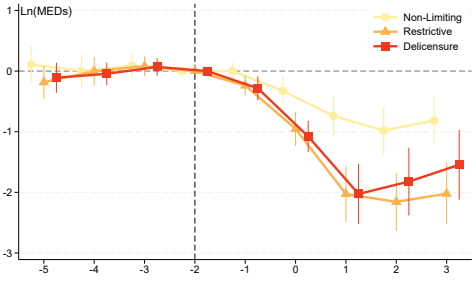
(d) Sun & Abraham: 1(opioid prescribing)



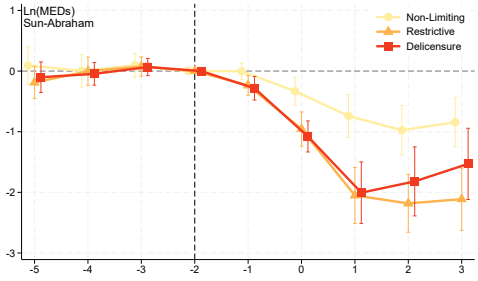
(e) TWFE: ln(MED) | any



(f) Sun & Abraham: ln(MED) | any



(g) TWFE: ln(MED) | any



(h) Sun & Abraham: ln(MED) | any

Figure A3: Effects of Physician Discipline: Estimation by TWFE vs. Sun & Abraham

Note: This figure reports the β_τ event coefficients reported in Figures Figure 3 and Figure 4 on the left and analogous coefficients from the differences-in-differences estimator in Sun and Abraham (2021).

B Data Appendix

B.1 Collection of State Medical Board Disciplinary Data

Licensure disciplinary actions are public records, and we formed a "basic" dataset recording the physician's name, license number, year and month of action, and action taken using these records. States commonly provide these records in two ways. The first, which was our preferred source, is Board Action Reports (BARs): newsletters or announcements published after each board meeting that describe actions taken at that meeting. The second method is a public-facing license search tool that can sometimes be filtered to physicians with prior disciplinary actions. In some cases, records were obtained through a formal public records request. Table A1 details our collection method for each of the 63 boards that license MDs or DOs in the United States (excluding territories). Of the 59 boards from which we successfully collected data, 42 used Board Action Reports, 9 used a license search tool, and 8 sets of records were obtained through a public records request. We obtained records back to 2008 with exceptions noted in Table A1; prior records were generally not maintained in electronic form. In every state, we standardized the records to report the "basic" information in a standardized way. In some states, we used the license search tool as a second step to e.g., obtain license numbers when only names were provided in the (BARs).

We have two forms of information about the basis for disciplinary action. "Detail" is a short phrase included in some Board Action Reports that summarizes the basis for action. "Text" refers to full case PDFs collected for 32 states. These are administrative law documents that describe the specific statute violated and the evidence of the violation; our text analysis process is described in Appendix D.

A small number of boards were uncollectable; our efforts are described in the final rows of Table A1.

Table A1: Data Collection by State Medical Board

State	Detail	Text	Method	Notes
AL		X	License search tool	Most recent action only
AR		X	Board action reports	
AK	X		Board action reports	
AZ MDs			Request for board action reports	
CA MDs		X	Board action reports	
CA DOs			Board action reports	
CO			License search tool	
CT		X	Board action reports	Reports 2008–2012 requested.
DC			Board action reports	Basic data years 2009–2022.
DE		X	Board action reports	
FL MDs		X	Board action reports	
FL DOs		X	Board action reports	

Continued on next page

Table A1 *(continued)*

State	Detail	Text	Method	Notes
GA			Board action reports	
HI	X	X	Board action reports	
ID			Board action reports	
IL	X		Board action reports	
IN		X	Board action reports	
KS		X	Board action reports	
KY		X	Board action reports	Reports 2008–2013 requested. Detail often missing. Detail data years 2014–2022.
LA		X	Board action reports	
MA		X	Board action reports	
MD		X	Board action reports	Basic data years 2010–2022. Detail data years 2010–2022.
ME MDs		X	Board action reports	
ME DOs			Board action reports	
MI MDs		X	Request for board action reports	
MI DOs		X	Request for board action reports	
MN		X	Board action reports	
MO		X	Board action reports	
MS			Board action reports	
NC		X	Request for board action reports	
ND			Request for board action reports	
NE			Request for board action reports	
NH		X	Board action reports	Reports 2008–2016 requested.
NM		X	Request for board action reports	
NV MDs		X	Board action reports	
NV DOs			Board action reports	
NJ		X	Board action reports	
NY		X	Board action reports	
OH	X		Board action reports	Reports 2008–2012 requested.
OK MDs		X	License search tool	
OK DOs			License search tool	
OR		X	Board action reports	Reports 2008–2015 requested.
RI		X	Board action reports	
PA MDs		X	License search tool	
PA DOs		X	License search tool	
SC			Board action reports	
TN MDs	X		Board action reports	
TN DOs	X		Board action reports	
TX	X		Board action reports	

Continued on next page

Table A1 *(continued)*

State	Detail	Text	Method	Notes
UT	X		Board action reports	Public records law covers last 10 years. Basic data years 2015–2022.
VA		X	Board action reports	
VT MDs		X	Board action reports	Reports 2008–2013 requested.
VT DOs			Board action reports	
WA MDs	X		Board action reports	Reports 2011–2022 requested. Basic data years 2011–2022.
WA DOs	X		Request for board action reports	
WI		X	Board action reports	
WV MDs		X	License search tool	
WV DOs		X	License search tool	
WY		X	License search tool	
Uncollected Boards				
AZ DOs	Records request returned ever-disciplined DOs without timing or action taken.			
IA	Records request answered; page fees were cost-prohibitive.			
MT	No Board action reports, no possibility to subset to disciplined physicians in search.			
SD	No Board action reports, no possibility to subset to disciplined physicians in search.			

Notes: Table reports data collection method or reason for failure for 63 boards operating in US states and DC. "X" in the detail column indicates that the board supplied a short phrase describing the reason disciplinary action was taken. "X" in the text column indicates that we obtained full case documents for actions taken in this state.

B.2 Validation of Data Collection Methods

We validate our data by comparing our state-year action counts against the National Practitioner Data Bank (NPDB), which collects reports directly from licensing boards and releases them publicly without physician identifiers. We filter the NPDB to licensure actions (report types 301 and 302, excluding malpractice and peer review records) and restrict to MD and DO license types. We collapse the NPDB to state-year counts using the variables *licnstat* and *aayear*. We delete state-year combinations that were not collected according to Table A1. We then adjust our data to align it with the set of actions collated by the NPDB. We drop CA "accusation filed" and FL "administrative complaint filed" entries, which are not in scope for NPDB. Deviations in either direction are expected. Our count may exceed NPDB's for two reasons: first, we cannot always identify and exclude accusations in states where they are not clearly labeled; second, in three states our collection methods were not restricted to MDs and DOs (CT includes Physician Assistants, HI includes Emergency Medical Technicians and Physician Assistants in certain years, and MA includes Acupuncturists). NPDB's count may exceed ours where our collection methods did not return a complete set of actions for a given state-year. Figure A4 presents the comparison on a log scale and shows strong overall agreement between the two sources.

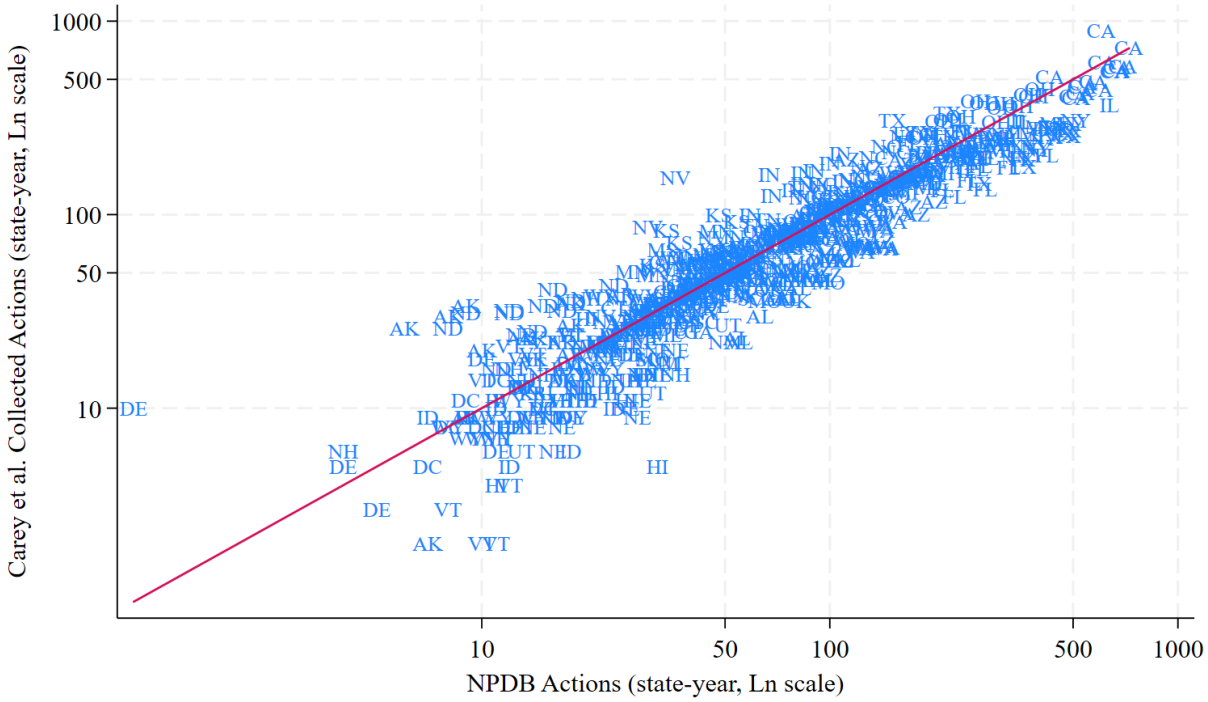


Figure A4: Comparison of State-Year Action Counts: Collected Data vs. NPDB

B.3 Linking Disciplinary Records to National Provider Identifiers

We link each disciplinary action record to a National Provider Identifier (NPI) from the National Plan and Provider Enumeration System's (NPPES) Physician and Non-Physician Practitioner database. NPIs are intended to be a lifelong identifier for physicians. The NPPES covers all practitioners with billing authority. Practitioners self-report name (as first, middle, and last), credential, and license numbers. We use the January release of the NPPES for the years 2008 through 2022.

There are several limitations to the NPPES. Most importantly, there is no auditing to ensure all active license numbers are reported in each year, and the set of license numbers is known to be incomplete (NBER, 2024). Secondly, since all information is reported by practitioners, there are typos and inadvertent variations in each record. Most importantly, while license numbers are reported in a state-specific format in the disciplinary data, NPPES records may neglect certain features such as an "MD" prefix or leading zeroes. In addition, we find variation over time in names within an NPI that sometimes reflects name changes (e.g., upon marriage) but is commonly due to the addition of middle names or suffixes or differences in the rendering of names (e.g., Mary Ann rendered as first name Mary Ann or first name Mary and middle name of Ann).

To address variation in the NPPES data, we begin with a cleaning step. We create a cleaned license field that standardizes NPPES license numbers to match the format recorded in each state's board disciplinary data, by extracting the alphabetic prefix and numeric portions separately and

recombining them—with appropriate zero-padding, suffix letters, or separators—according to each state's known license number structure. NPPES names are stripped of non-alphabetic characters and honorific suffixes, and middle-name conflicts within the same NPI are resolved in favor of the longer name.

On the disciplinary side, we clean names in the same way as the NPPES records. In addition, names are reversed from "*last, first middle*" where needed, and parsed into first, middle, and last components.

The matching pipeline proceeds through five step in sequence; each step attempts a match on the records that the previous step could not resolve. Table A2 reports matches for each step.

Step 1: License number match with name confirmation. Ninety-seven percent of our disciplinary records include a state license number. We join each record against NPPES on the combination of licensing state and license number. Because NPPES license numbers are self-reported by physicians and can differ in format from the board's record (e.g., a prefix added or dropped, a leading zero present in one source but not the other), we also attempt near-miss matches on the numeric portion of the license number alone. All license-number matches — exact and near-miss — are confirmed by Jaro-Winkler similarity of the first and last name: matches with similarity at or above 0.85 are classified as Step 1 (high confidence); those between 0.75 and 0.85 as Step 2 (confirmed with lower name similarity); and those below 0.75 are treated as disconfirmed and forwarded to later steps. Near-miss matches require a higher name similarity threshold of 0.92. TX is excluded from near-miss matching because its alpha prefix is identity-bearing (e.g., A12345 $\neq$ B12345). The states CT and MA are processed separately at this step because their boards also license non-physician practitioners; their records are matched into a NPPES subset that includes the extra credentials potentially included in those records (with non-physicians then excluded).

Three percent of records lack a license number; these include three states (HI, DC, MS) have no license numbers in the board records and enter the pipeline directly at Step 2.

Step 2: Cross-state propagation within disciplinary records. The previous step created a set of NPI-disciplinary action matches confirmed by both a matching license number and a matching name. Since it is common for the same individual to be disciplined by more than one state, we use this "oracle" to resolve unmatched records. Specifically, if an unmatched record's name matches a name in the oracle, we propagate that NPI to the unmatched record.

Middle names are used to disconfirm spurious matches under two rules. First, if both records carry a full middle name, the match is rejected when the two names share no word tokens. Second, if both records carry at least a middle initial, the match is rejected when the initials differ; this rule applies even when one side has a full middle name and the other only an initial, in which case the initial is compared to the first letter of the full name. Matches where either side has no middle name information are left unaffected. For example, if a disciplinee has a name recorded as Colleen M. Carey, the names Colleen Carey, Colleen M. Carey, and Colleen Marie Carey would all be "non-disconfirming", while names Colleen Catherine Carey or Colleen C. Carey would be disconfirming. However, all five names would be "non-disconfirming" if the disciplinee's name were

recorded as Colleen Carey alone.

When an unmatched record's name maps to more than one NPI in the oracle after middle-name disambiguation, the record is passed forward to Step 3.

Step 3: Exact name match. Records remaining after Step 2 are matched against the full national universe of MDs, DOs, and missing-credential NPIs. The match is on exact first-and-last name, without restricting to a particular state. The national universe is used in this state given the known incompleteness of state license recorded in the NPPES. Middle-name disconfirmation is again conducted as in Step 2.

After middle-name disconfirmation, there remain a non-trivial number of disciplinees with multiple potential NPI matches of equal credibility (exact first+last, nondisconfirming middle name). These are "final failures" – we cannot be certain which NPI was disciplined. Records without any exact first+last matches are forwarded to Step 4.

Step 4: Fuzzy name match. The final step "fuzzy matches" unmatched names from Step 3, calculating Jaro-Winkler similarity against the full national universe of MDs, DOs, and missing-credential NPIs. This step is designed to catch minor typos or parsing issues (i.e., apellidos rendered as a two-word last name vs. one apellido appearing in the middle name field). Three scores are computed per candidate pair — first-last, full name including middle name information, and a middle-initial-adjusted score that uses only initials if either name reports only a middle initial. The best score is taken across these three Jaro-Winkler similarity scores. A match is accepted only if the best score exceeds 0.97 and exceeds the second-best candidate by at least 0.03, ensuring near-uniqueness. Records with no candidate meeting this threshold, or with multiple tied candidates, are left unmatched.

Outcome of Matching of Disciplinary Records to NPIs. Table A2 reports the outcome of all four steps. First, of 75,592 actions, 66% match in step 1 with two pieces of confirming information: license number and name. Nearly all of the confirming names are in the top tier of Jaro-Winkler similarity, and nearly all match exactly on license number. 95% of names have a Jaro-Winkler score of 1 (exact match). Upon inspection, cases with a matching license number and lower Jaro-Winkler score were overwhelmingly last name differences consistent with marriage, which is why we allowed a lower Jaro-Winkler score among these license-number matches.

The cross-state propagation step takes the 49,534 action records matched in Step 1, and then propagates the NPI to other-state disciplinary records in cases of exact-matching first and last name and non-disconfirming middle name information. Of 25,304 disciplinary records entering this step, 41% are matchable, meaning they share a first and last name (and non-disconfirming middle name) with a disciplinee who was matched in Step 1.

The third step works with 15,695 records that were unmatched in either of the preceding steps. We match these individuals on exact first and last name to the NPPES universe, and find a unique match for 62% of those records, meaning there was only one individual in the NPPES universe who (1) has the same first and last name and (2) has non-disconfirming middle name information. There

Table A2: NPI Matching: Four-Step Linkage Procedure

NPPES universe (unique NPI–name pairs, national): 1,726,788	
Entering action records	75,592
Step 1: License Number Match	
Linkage variable: State license number	
Name confirmation thresholds: Exact plicnum: JW $\geq$ 0.75; Near-miss: JW $\geq$ 0.92	
of which: with a usable license number	73,285
Matched — exact plicnum (JW $\geq$ 0.75)	47,821
Matched — near-miss (JW $\geq$ 0.92)	1,713
First+last JW score among matches: mean (p10, p50, p90)	0.996 (1.000, 1.000, 1.000)
Total Step 1 matches	49,534
Unmatched — no sufficient match	22,998
Unmatched — nonunique plicnum	725
Missing license number (enter Step 2 directly)	2,306
Step 2: Cross-State Propagation Within Disciplinary Records	
Linkage variable: First + last name exact (within disciplinees alone)	
Middle name disconfirmation applied	
Entering records (no sufficient match from Step 1)	25,304
Matched — unique propagated NPI	10,336
Unmatched — nonunique after disambiguation (to Step 3)	238
Unmatched — no oracle candidate (to Step 3)	14,730
Step 1 nonunique failures also passing to Step 3	725
Total passing to Step 3	15,695
Step 3: Exact Name Match	
Linkage variable: First + last name exact (within national NPPES universe)	
Middle name disconfirmation applied	
Entering records	15,695
Matched (unique exact name, mname-consistent)	9,756
Final failure Multiple matches (exact name, mname-consistent)	1,467
Unmatched — no exact NPPES name match (to Step 4)	4,475
Step 4: Fuzzy Name Match	
Linkage variable: JW similarity	
Acceptance criterion: Best JW score > 0.97; margin over 2nd-best $\geq$ 0.03	
Entering records	4,475
FL name score among matches: mean (p10, p50, p90)	0.985 (0.975, 0.985, 1.000)
Matched	245
Final failure Unmatched — nonunique (best > 0.97 but gap < 0.03)	440
Final failure Unmatched — no sufficient match (best $\leq$ 0.97)	3,790

Notes: JW = Jaro-Winkler similarity score. All counts (entering, matched, unmatched) are at the action-record level.

are 1,467 records entering this step that cannot be resolved because there are multiple potential NPPES matches of the same credibility. We do not forward these to the fuzzy-match step since they will again result in failures due to nonuniqueness.

The final step conducts the fuzzy-name match. Only 245 disciplinary actions are matched in this step. Upon inspection, these appear to be typos, name spacing variations, and differences in first/middle/last parsing. There are 440 cases where there are multiple credible fuzzy matches, which we record as final failures. Most records entering this step do not find a credible fuzzy match.

Overall, of 75,592 entering action records, we fail to obtain an NPI for only 5,697 records, or

7.5% of the total entering records. A third of our final failures are failures to uniquely assign an NPI—the disciplinee exists in the NPPES, but we cannot be sure of the correct NPI match. The remaining set—only 5% of the disciplinees—do not appear to be reported in the NPPES at all over the sample period. We are unsure whether these disciplinees never held an NPI, are excluded from the NPPES universe for some reason, or are simply not found due to the challenges of name matching.

The NPI matching step varies across states, with the lowest matching percentage in AL at 37.6% and the highest in OK at 97.8%.

C Disciplinary Action Categorization

We categorize each disciplinary action by the type of action the state medical board took against the physician, using keyword-based pattern matching applied to the action text in board records. Adverse actions range from a public reprimand to the revocation or surrender of a license. Some actions are non-adverse, such as a reinstatement or a board's announcement that an earlier punishment, like a probationary period, has ended. We aggregate action types into severity levels: adverse actions fall into three levels of increasing severity, which we call non-limiting, restricting, and delicensure actions. Some actions cannot be categorized, but we still treat them as adverse, for reasons described below. After cleaning, our action data contain 14,107 distinct action strings, and 4.9% of the 69,871 actions have blank action text. Table A3 shows the three most frequent action strings in each action category.

C.1 Action Type Classification

Before matching, we clean each action string by converting it to lowercase and removing periods, hyphens, colons, semicolons, and ampersands. We then use regular expression pattern matching to identify action types from keywords commonly used in state medical board records. To give a sense of this process, table A4 lists the main keywords and phrases associated with each action category, along with the number of strings and actions assigned to it. A single action string often describes more than one action (e.g., a fine imposed together with probation), so action categories are not mutually exclusive.

Action strings that match none of the keywords, as well as actions with blank text, are uncategorized. We still treat these actions as adverse: the text, together with the context in which boards produce these records, generally indicates that the board took action against the physician, even when the text does not say what that action was. Table A5 shows the twenty most common uncategorized action strings, with the 3,400 actions with blank text ranked first among them. Apart from the blanks, the list is dominated by generic descriptions such as "decision" and "consent order."

C.2 Severity Classification

We aggregate categorized adverse actions into three severity levels, based on the degree to which they restrict practice. *Non-limiting* actions (accusations, reprimands, continuing medical education requirements, and fines) do not limit where or how a physician may practice. *Restricting* actions (restrictions, probation, and suspensions) limit practice within the state but do not permanently remove the license. *Delicensure* actions (surrenders and revocations) prohibit practice in the state, at least temporarily. Actions are classified as *non-adverse* if they indicate a reinstatement, withdrawal, or dismissal without any concurrent adverse action. When a single action string contains categories at more than one severity level, we assign the highest. Of the 69,871 actions in our data, 29.7% are non-limiting, 23.3% are restricting, 13.7% are delicensures, 24.1% are uncategorized, and 9.2% are non-adverse.

Table A3: Most Frequent Action Strings by Action Category

Action Category	Action String	Unique Actions
Non-limiting		
Accused	accusation filed	3,679
	administrative complaint filed	2,700
	first amended accusation filed	574
Reprimand	reprimand	1,124
	public reprimand	859
	public letter of concern	724
Continuing Education	advisory letter w nondisciplinary cme	245
and Continuing Medical	reprimand/fine (cme)	57
Education (CME)	fine, mandatory continuing education	39
Fine	fine imposed probation	539
	fine imposed	436
	fine	217
Restricting		
Restrictions	reprimand terms imposed other	266
	temporary suspension/restriction	223
	obligations imposed	205
Probation	probationary requests and modifications	1,095
	probation	573
	fine imposed probation	539
Suspended	suspension	1,215
	suspended	583
	summary suspension	520
Delicensure		
Surrendered	surrendered	1,085
	voluntary surrender	826
	license surrender	706
Revoked	revoked	737
	revocation	655
	license revocation	361
Non-adverse		
Ending or withdrawal	obligation(s) satisfied	1,219
	terms terminated	431
	probation terminated	337
Reinstated	reinstatement granted	423
	unconditional license	174
	mongfl monorder granting full licensure	159

Notes. Strings are shown after cleaning, since categories are assigned to the cleaned text and cleaning combines raw strings that differ only in capitalization, spacing, or the punctuation.

Table A4: Action Categories and Severity

Action Category	Keywords and Phrases	Strings	Unique Actions
Non-limiting		3,950	20,748
Accused	*accus-, filed, referred to formal hearing, open formal complaint, notice of proposed disciplinary action*	308	8,699
Reprimand	*reprimand-, citation-, admonish-, censure-, public letter of concern, advisory letter, civil penalty, cease and desist, penalty letter*	2,430	10,668
Continuing education	*cme, education-, class-, course-*	1,816	2,455
Fine	*fine-, monetary, costs*, or a dollar sign	3,872	7,140
Restricting		4,821	16,270
Restrictions	*restrict-, terms imposed, limitation, obligations imposed, conditioned license*	1,295	3,554
Probation	*probat-, monitor-*	2,455	6,487
Suspended	*suspen-, cease practice, inactive with cause, inactivated for*	2,182	8,393
Delicensure		1,942	9,570
Surrendered	*surren-, volun-, resignation, agreement not to practice, cessation of practice agreement, retired in lieu of discipline*, or retirement of a license	949	6,104
Revoked	*revo-, refuse(d) to renew, refusal to renew, deemed a revocation, precluded from*	1,058	3,797
Uncategorized		1,893	16,843
No keyword matched		1,893	13,443
Blank action text		–	3,400
Non-adverse		1,501	6,440
Ending or withdrawal	*withdraw-, dism-, termin-, end-, no action, satis-, rescind-, lifted, release-, complete-, vacated order, exonerate*	1,339	5,197
Reinstated	*reinsta-, returned to, unrestrict-, reentry agreement, granting full licensure, unconditional license*	446	1,809

Notes. A *string* is a distinct action text after cleaning, which lowercases words and removes periods, hyphens, colons, semicolons, and ampersands.

Table A5: Twenty Most Common Uncategorized Action Strings

Action String	Unique Actions	Share of All Actions (%)
Blank action text	3,400	4.9
decision	1,884	2.7
fdo legacy	769	1.1
consent order	695	1.0
quality of care	646	0.9
consent agreements	578	0.8
final orders	400	0.6
court actions	378	0.5
cls stipulation	367	0.5
agreed order	334	0.5
unprofessional conduct	321	0.5
approve consent order	262	0.4
inadequate medical records	235	0.3
quality of care violations	216	0.3
other states' actions	144	0.2
mon legacy	137	0.2
final disciplinary orders	134	0.2
agreed order entered into	114	0.2
interim consent order	110	0.2
order terminating agreed order	108	0.2

Notes. Shares are of all 69,871 actions, including actions with blank action text. Blank action text comes almost entirely from three states: 1,797 from Indiana, 993 from Louisiana, and 589 from South Carolina.

C.3 Corrections and Refinements

We apply several correction rules after keyword matching to improve classification accuracy. An accusation overrides every other adverse category, since a filed accusation is not yet a punishment. An ending, withdrawal, or dismissal overrides an accusation, reprimand, restriction, probation, suspension, or revocation in the same string, while a fine, education requirement, or surrender in the same string instead overrides the ending. Terminating a consent, agreed, or stipulated order or a corrective action agreement, and completing a required course, are not treated as endings. A stayed revocation is not a revocation, so the action takes whatever other category the string names. Revocations of a controlled substance or pain management registration, which do not remove the medical license, are classified as restrictions rather than delicensures, and a voluntary suspension is classified as a suspension rather than a surrender.

Keyword matching generally works well for short strings that boards use repeatedly, but rarely used strings are often long and may combine several different terms. We therefore reviewed action strings by hand, focusing on delicensures and reinstatements. Delicensure is the sanction intended to remove a physician from practice in the state, and reinstatement is what allows the physician to return, so classification errors in these categories matter most for our analysis. Every delicensure string that appears fewer than three times in our data was hand coded. In total, we reviewed 2,917 distinct strings covering 8,007 actions, including 96.0% of all delicensure strings. In addition, 747 strings were coded independently by more than one coder, with disagreements resolved through dis-

cussion. Hand coding changed the categories of 1,160 strings, moving 518 strings out of delicensure and 39 strings into it; for these strings, the hand-coded categories replace the keyword assignment.

C.4 Clean Delicensures

Delicensure strings differ in how clearly they describe a loss of license. Many are unambiguous (e.g., "revoked" or "voluntary surrender"), while others combine a delicensure with another sanction or describe it within a longer narrative. For use as a robustness check, we define a narrower set of *clean delicensures*: actions whose text unambiguously describes a revocation or surrender of the license. The clean delicensure indicator is a hand-picked list of 37 action strings, which together account for 65.9% of all delicensure actions. Table A6 lists these strings, the number of actions each covers, and each string's share of all delicensure actions.

Although clean delicensures make up only 37 of the 1,942 distinct delicensure strings, they account for 6,311 of the 9,570 delicensure actions. The remaining 1,905 delicensure strings are mostly rare: 1,727 of them appear only once in our data.

D Text Analysis to Categorize *Basis* for SMB Actions

Coding what action a state medical board took against a physician was relatively easy: board tend to report this directly, either in Board Action reports or through a license search tool, as we documented in Appendix Section B. Coding *why* the board took action is not as straightforward. A short phrase detailing the basis was included directly in Board Action Reports for a handful of states (see the "Detail" column of Table A1), but even where present this was at best a short label, not a full description of what actually happened. For most boards, the reason behind disciplinary action is not in the structured data at all, instead it is in the underlying case documents themselves—administrative-law style PDFs, which we obtained for 32 of the 48 states in our data (see the "Text" column of Table A1). These documents vary in length and detail depending on the state and case type but usually include details about specific wrongdoing and evidence brought before the state medical board.[1]

This Appendix section describes how we used these case documents, together with large language models and human coding, to code the bases for each disciplinary action—that is, what did the doctor do wrong?

One problem we face is that physicians can have multiple actions[2], but that does not mean that they were charged with multiple instances of wrongdoing. Multiple actions could stem from the same underlying behavior. For example, one New York physician's 2015 record shows the SMB applied "an additional year of probation from the probation imposed in the disciplinary action

[1] Case documents vary substantially in length: in a random sample of 1,500 documents, page counts ranged from 1 to 180, with a median of 8 pages and a mean of 14. These were sampled from the 77,394 case-document PDFs collected across all 32 states with detailed records. 99 of the 1,500 sampled files could not be parsed for a page count (corrupted or malformed scans) and are excluded from these statistics.

[2] Mean 3.9, median 2, Range 0-176

Table A6: Action Strings Counted as Clean Delicensures

Action String	Unique Actions	Share of All Delicensure Actions (%)
surrendered	1,085	11.3
voluntary surrender	826	8.6
revoked	737	7.7
license surrender	706	7.4
revocation	655	6.8
voluntary surrender of licensure	364	3.8
license revocation	361	3.8
voluntary surrender/revocation	166	1.7
surrender	159	1.7
voluntarily surrendered	149	1.6
surrenders	117	1.2
voluntary surrenders/retirements	111	1.2
license surrendered	100	1.0
surrender of license	93	1.0
voluntary surrender of license	90	0.9
resignation	83	0.9
revocation of license	53	0.6
agreed order of surrender entered into	51	0.5
revoked license	45	0.5
revocation of inchoate right to renew	42	0.4
license revoked	40	0.4
surrendered license	38	0.4
voluntary surrender in lieu of discipline	37	0.4
consent order of revocation	35	0.4
order of revocation	30	0.3
agreed order of surrender	27	0.3
voluntary surrenders	20	0.2
order of revocation issued	19	0.2
agreed order of permanent surrender	11	0.1
surrender of medical license	11	0.1
voluntary revocation	10	0.1
license voluntarily surrendered	8	0.1
revocations/surrenders	8	0.1
voluntary surrender of medical license	7	0.1
memorandum of decision revocation	6	0.1
oas order accepting surrender	6	0.1
approve voluntary surrender	5	0.1
All clean delicensure strings (37)	6,311	65.9

effective September 5, 2013 and an additional fine of $10,000"; the physician satisfied these terms as indicated in an action dated September 2017, and then a 2022 modification order then terminated the one remaining condition (a chaperone requirement) from the original order — three actions across nine years, all pertaining to a single underlying basis/disciplinary matter. In addition, a physician disciplined in one state may then be disciplined in another state where he/she holds a license for the same action (or for failing to report their discipline in the other state). The documents and the data do not usually include a "case number" or anything linking a particular action to a particular basis, and while some of the examples above are quite linear in nature others are not.

Given this difficulty, we choose to code the bases for discipline at the *physician* level (by NPI): for each physician, considering every action and document we have across every state that disciplined them, what is the full set of reasons they were disciplined that are uncovered within these documents.

D.1 Preparing Documents and Data for Coding

We collected 77,394 raw pdf files.

The first step is to convert each PDF file to a .txt file. We turn every page of a PDF into an image, and then use Optical Character Recognition (OCR) to read the text in the images. Tesseract OCR was the specific engine used. This process takes pictures of text and makes digital text data that is editable and searchable. Because many of our documents are scans, we could not use commonly used Python libraries for PDF text extraction and necessitated the OCR technology.

The individual txt files are then compiled into a single CSV file. We then construct a CSV file where every row has an identifier state x license number x year-month identifer and text data from a given document. If a doctor has two distinct documents in a single year-month, text is appended.

We use the row identifiers in the CSV to match to physicians' disciplinary records. Because document dates do not match 1:1 with action dates, we attach all physician text records onto the file with all of the same physician's disciplinary actions.

Our final data include 63,555 total case documents across 30,432 physicians.

D.2 The Potential Bases

We chose to code the basis for discipline using a fixed set of binary categories rather than free-form text, since we wanted to be able to run heterogeneity analyses by basis type, which requires a consistent, structured set of categories rather than open-ended descriptions. We formed this list of categories from our own reading of a wide range of case documents and from our sense of the kinds of heterogeneity analysis we wanted to be able to run. The final categories and definitions were an iterative process of learning from the RAs' experience coding actual cases and from our later prompt engineering experience with Claude. Table A7 lists these categories and brief versions of their definitions; full prompt language is available in our replication package.

Table A7: Coded Bases for Disciplinary Action

Category	Definition
Sexual misconduct	Sexual misconduct involving a patient or coworker in a medical-practice context.
Improper prescribing	Improperly prescribing, distributing, or acquiring drugs with recreational or abuse potential (opioids, benzodiazepines, stimulants, medical marijuana, etc.) for nontherapeutic reasons.
Improper opioids	Subset of improper prescribing specifically involving opioids or narcotics.
Physician disability	The licensee's own substance use, mental health condition, or debility was the grounds for discipline.
Law enforcement	An external arrest, indictment, or conviction was the grounds (including DUI or failure to pay taxes).
Law enforcement (medical)	Subset of law enforcement where the offense was medically related (e.g., billing fraud, patient assault, drug trafficking).
Medical error	A diagnostic or treatment mistake, including one made by a supervised NP/PA.
License issues	Administrative licensing problems: failure to complete required CME, pay fees, or disclose required information on an application.
Miscellaneous violation	A clear grounds for discipline that does not fit any other category.
Uncertain basis	The reason for discipline genuinely cannot be determined from the available documents.

Notes: Categories are not mutually exclusive; a physician can have more than one basis coded as 1. Improper opioids and law enforcement (medical) can only be coded 1 if their parent category is also 1.

D.3 Human Coding by Undergraduate Research Assistants

In January 2026 we hired 4 undergraduate Research Assistants (RAs) at Cornell University to hand code some of these documents. Their goal was to try to categorize the 'basis' of the discipline: what did the doctor do wrong (or what were they accused of doing wrong). This was the main task. Each week we would randomly select 11 physicians per coder, export their actions and text data from our core database into an Excel sheet, and additionally copy corresponding raw PDF files into folders for the RAs.

The RAs were then asked for each physician to look at the PDFs and other data and try to determine what they were accused of doing wrong. The potential categories were: sexual misconduct, improper prescribing, improper opioid prescribing, physician disability, law enforcement, law enforcement medical, medical error, license issues, miscellaneous violation, or uncertain basis.They could mark a 0 or 1 for any of these categories, and multiple categories could get a 1; we also allowed the students to enter -1 if they were not sure. Recall Table A7 provides brief descriptions of each of these types of actions (while these mirror what was given to the students they were changed and

updated overtime through our conversations with them and our later prompt engineering).

We met with the students each week so they could ask questions about the cases they were looking at or ask for clarification on the definitions of various categories. This led us to making some updates to the category definitions or even expanding categories over time. Note the students did not always backtrack and update past weeks coding for later weeks category or definition changes. This process went on for 10 weeks, and in the end they hand-coded categorizations for 391 physicians.

We applied a small number of hand-coded corrections to the RAs' raw responses before using them for LLM text analysis task. Most were systematic recoding rules rather than case-by-case judgment calls: when a coder marked every basis category -1 (uncertain) and also flagged uncertain basis, we zeroed the redundant -1s and let uncertain basis stand alone as the response (and we applied this whenever any -1 co-occurred with uncertain basis, not only when every category was -1). Rows left entirely blank across all categories were recoded to uncertain basis=1. We enforced logical hierarchy between related categories—e.g. law enforcement should always be 1 when law enforcement medical is 1.[3] We also decided later in the process that law enforcement medical and improper opioids could not be -1, so we recoded these -1 to 0.[4] A small number of additional fixes addressed data-entry artifacts specific to individual coders or coding weeks (e.g., free text entered in an answer cell, a parsing error, a field not yet defined in an early week of coding). We view these changes as innocuous and not impacting the intended response of the RA.

During this process, 24 physicians were double coded (each coded by 2 RAs) to better understand inter-code reliability. Of those 24, on 15 the pair of RAs agreed completely on the basis.[5] On the remaining 9, disagreement centered on whether a described action was a medical error (4 of 9), where the underlying judgment call was genuinely borderline, and on confusion between an administrative license issue, a miscellaneous violation, and a reciprocal action taken solely because another state board had already acted (4 of 9, spread across those three categories). Sexual misconduct, improper prescribing/opioids, physician disability, and uncertain basis had no or almost no disagreement among the double-coded pairs. For the final human coded dataset one of the PIs adjudicated between the two RAs responses if they disagreed and chose the one that seemed most appropriate for the final human coding.

D.4 Preparing for LLM Text Analysis

With these hand-coded (and PI cleaned) categorizations for 391 physicians we then prepared this set for LLM text analysis. We decided ex ante to set aside 150 of these for prompt engineering and the remaining 241 for validation. We made sure the original 10 that we did the prompt engineering test on were included in the 150 for prompt engineering (those, recall, had been chosen randomly).

On May 20, 2026 we randomly selected 10 NPIs from the human-coded data (stratified by basis:

[3] This was a rule and hierarchy we clarified later in the process.

[4] Same for patient mentioned and multiple distinct cases though again we are not using these in our analysis.

[5] Ignoring patient mentioned and multiple distinct violations which we do not end up using.

2 with no basis category identified (uncertain), 4 with a single basis chosen, and 4 with multiple bases chosen by the human coders) to test the prompt-engineering pipeline. This was a first end-to-end pilot of the Claude coder against the human-coded data, built with the assistance of Claude Code: we ran these 10 physicians' documents through an early draft of the prompt and schema and compared the output field-by-field against the human coding to surface obvious problems before investing in the full iterative prompt-engineering process. This pilot round informed the initial prompt design rather than producing a formal accuracy statistic, and all 10 NPIs were used in the prompt-engineering set (below) so that this early work would not be wasted.

We randomized in Stata: each NPI was given a random number drawn from a uniform distribution (`uniform()`), fixing the random-number seed (664313) for reproducibility, and NPIs were then ranked on this draw. The first 140 by this ranking, plus the 10 original pilot NPIs (assigned first regardless of rank), formed a pool of 150 NPIs that could be used for prompt engineering; the remaining 241 were locked into a separate validation folder not to be touched until prompt engineering was complete.

All coding was performed with a single fixed model, **Claude Opus 4.7 (Anthropic)**, from the first prompt-engineering run through the final validation and production runs. We logged the exact model version returned by the API for each run to confirm no silent model changes occurred over the course of the project. Our prompt reset the context window for each physician, and we helped ensure straightforward instructions to the LLM using a Pydantic schema.

D.5 Prompt Engineering

With the help of Claude Code we wrote a workflow for prompt engineering: one component drew a fresh batch of 10 NPIs from the prompt-engineering pool and tracked which NPIs had already been used in a ledger (so batches were never accidentally reused), a second sent each physician's full case history to Claude via the API and parsed the structured response against our Pydantic schema, and a third automatically scored Claude's output against the human coding and compiled every disagreement into a spreadsheet for us to review.

Importantly, we also asked Claude to give us the reasoning behind it's choices, including quotes from the underlying text data.[6]

Each batch thus had 10 NPIs. For a given batch, we used the workflow to code all 10 NPIs with the current version of the prompt, then compared Claude's output to the human coding and reviewed every disagreement together. If a disagreement revealed that our prompt's instructions were unclear or incomplete, we revised the prompt and re-ran the same batch of 10 NPIs under the new version – an *iteration* of that batch – to check whether the revision resolved the issue. We repeated this within a batch until an iteration required no further prompt changes, at which point we moved on to a fresh batch of 10 NPIs. Note that requiring no further prompt changes does

[6] For example, for a physician indicted and later convicted for conspiracy to distribute controlled substances via 'pill mills,' Claude coded `law enforcement medical=1` and `improper prescribing=1` with the reasoning: "Dr. [X] was indicted (and later convicted/imprisoned) for conspiracy to unlawfully distribute controlled substances via 'pill mills' — improper prescribing plus a medically-related law enforcement action."

not necessarily imply 100% agreement with the human coders, as we describe below. Table A8's *Iterations* column reports how many times each batch was coded under this process.

Even during the first of these prompt engineering runs, we started to realize that rather than treating the human coded data as the "gold standard" against which we would compare the LLM coded output, we should instead think of both the human coders and the LLM as potentially noisy measures of the underlying truth. As we compared the LLM coded output to what the students we would sometimes realize that Claude was closer to what we as PIs would consider the actual basis than the RAs. For example, some disagreements seemed to stem from the humans not fully understanding the definitions. For example, one physician was disciplined by their board via a consent order that "charged [him] with negligence on more than one occasion," for prescribing high doses of controlled substances despite a known alcohol-abuse history and failing to monitor lab values for another patient – ordinary board adjudicative language. The RA coded this as `law_enforcement_medical=1`, apparently reading "charged with negligence" as an external law enforcement matter, when it in fact described the board's own administrative proceeding; Claude consistently coded this `0` across three separate coding rounds on this physician (2026-05-28, 05-29, and 07-10), and we did not treat this as a Claude error requiring a prompt fix. There were several examples like this.

When we did find issues where a disagreement between the human coded basis and the Claude coded basis was due to unclear prompting on our side, we would edit the prompt to make this point clearer, and then re-run the same batch of NPIs to see if this "corrected" the issue. For example, one physician's state medical board record included, alongside a substantive misconduct finding, a separate administrative order for failing to timely complete a death certificate. Claude originally coded this as a license issue. It was not obvious from our original instructions whether this counted as a `license_issues` administrative matter or `misc_violation`; in the week this case came up in our prompt-engineering rounds (2026-07-17), we added an explicit instruction to the prompt Claude was using that "administrative failures related to medical recordkeeping or death certificates are not license issues," along with a companion clarification that failing to disclose required information on a license application *does* count as a license issue - distinguishing two cases that had been implicitly conflated.

Often we were finding disagreements that were not particularly material to the task. Many of these included Claude adding a "miscellaneous violation" to a case. That is the doctor mainly got in trouble for one of the other indicated bases (say improper opioid prescribing), but Claude, reading much more carefully than a human, also noted that the doctor additionally got in trouble for improper record keeping and would add a "miscellaneous violation" to the basis coding. This is a case where Claude was clearly following our prompt but our human RAs appear to have instead coded the case based on the "gist" of the basis. We did not attempt to alter the prompt to resolve these disagreements.

Each batch of prompt engineering would take us around one-hour as we would need to read through the disagreements and decide whether this was a moment we needed to update to the

prompt or whether the Claude's coding of the basis was valid and we did not need to update the prompt.

After doing this process for 5 or so batches we started to realizing we were not making many more major changes to the prompt, it started to be the case that we weren't really changing the prompt each time. On July 17, 2026 we decided on a stopping rule: if we go through two full rounds of prompt engineering without updating the prompt then we will stop. At that time we will move the rest of the prompt engineering set (which was randomly ordered) into validation.

This stopping rule led us to actually engineer for 80 NPIs. The rest were moved to validation. Table A8 summarizes the batches used.

Table A8: Prompt-Engineering Batches

Batch	Dates	Iterations	Agreement (first → last)
1	2026-05-28 – 07-10	3	90.0% → 90.0%
2	2026-05-29 – 07-10	10	88.2% → 91.8%
3	2026-06-19	1 (discarded)	–
4	2026-07-03 – 07-10	2	82.7% → 90.0%
5	2026-07-10	1	87.3%
6	2026-07-17	3	87.3% → 91.8%
7	2026-07-24	3	90.0% → 89.1%
8	2026-08-14	1	83.6%
9	2026-08-19	1	79.1%

Run once, found to contain data errors, and excluded from the prompt-engineering record.

D.6 Validation

With the prompt frozen after prompt engineering, we ran the finalized prompt on the 311 NPI validation sample (these are the initial 241 set aside for validation, plus the 70 NPIs that had been originally reserved for prompt engineering but were never actually used once we hit our stopping rule). The validation run was performed on September 8, 2026 using the same Claude Opus 4.7 model as every prior prompt engineering run, and under the prompt exactly as finalized in the last prompt-engineering iteration.

Table A9 reports on agreements between this Claude's validation run and the human coding on this held out sample. Overall we had 10 fields we compared × 311 NPIs= 3,110 field-level comparisons. Overall Claude's coding agreed with the human coding in 90.5% of these field-level comparisons. Miscellaneous violation was the weakest individual field, which is consistent with what we described above—the modal disagreement in this category was Claude identifying an additional miscellaneous violation that the human coder had not flagged, rather than a substantive disagreement about the physician's primary basis for discipline.

Table A9: Claude vs. Human Agreement, Validation Sample (N=311)

Field	N	Agreement
Sexual misconduct	311	97.4%
Law enforcement (medical)	311	95.5%
Law enforcement	311	94.5%
Physician disability	311	93.9%
Uncertain basis	311	93.6%
Improper opioids	311	92.3%
Improper prescribing	311	91.6%
License issues	311	91.3%
Medical error	311	82.6%
Miscellaneous violation	311	72.0%
Overall	**3,110**	**90.5%**

Notes: Excludes `patient_mentioned`, which was collected but is not used elsewhere in this paper. Agreement is the share of the 311 physicians for which Claude's and the human coder's 0/1/uncertain value matched exactly on that field.

D.7 Full Run

With prompt engineering complete and the validation run confirming the prompt's performance out of sample, we then had Claude code the full physician population. On September 22, 2026 we submitted the data for all of our physicians, 30,432 NPIs, via the Anthropic Batch API, using the same frozen prompt and again **Claude Opus 4.7** that was used throughout prompt engineering and validation. 45,571 distinct case documents were processed across 30,432 these physicians.[7]

As a final check, we also included our validation sample in this final run and re-examined agreement against the human-coded cample with this full production run. Of the 311 validation NPIs, 297 were in the full production population.[8] Agreement on this subset was **90.1%** (Table A10), closely matching the validation run's agreement of 90.5%.

E Hamonizing Prescriber Identifiers in Medicare Part D

E.1 Prescriber Identifiers in the Prescription Drug Event Dataset

The PDE dataset identifies the prescriber for each prescription fill using one of two identification systems: the National Provider Identifier (NPI), which was required from 2013 onward, or the prescriber's Drug Enforcement Administration (DEA) registration number. The DEA number is used in 55.3% of all fills in 2008, declining to nearly zero by 2013 (Table A11). No administrative key has good coverage of these DEA numbers, so a prescriber appearing under a DEA number in 2010 and an NPI in 2014 is observationally two distinct prescribers, and prescribing histories are

[7] Documents per physician vary: some physicians have a single disciplinary document, others have many across multiple states and actions.

[8] The remaining 14 validation NPIs were not present in the full production population; some were disciplined after 2022, while others failed our final name-matching procedure and were excluded from the final dataset.

Table A10: Claude vs. Human Agreement, Full Production Run (N=297)

Field	N	Agreement
Sexual misconduct	297	97.0%
Law enforcement (medical)	297	95.3%
Law enforcement	297	93.9%
Physician disability	297	93.6%
Uncertain basis	297	93.6%
License issues	297	91.9%
Improper prescribing	297	91.2%
Improper opioids	297	90.2%
Medical error	297	81.5%
Miscellaneous violation	297	72.7%
Overall	**2,970**	**90.1%**

Notes: Excludes `patient_mentioned`, per the same convention as Table A9. This is a subset of the 311-NPI validation sample (Table A9) restricted to the 297 physicians present in the full production population.

Table A11: Prescriber identifier type in Part D claims, by year

	2008	2009	2010	2011	2012	2013
DEA-coded claims (millions)	116.1	64.2	41.5	22.3	13.0	0.0
NPI-coded claims (millions)	93.7	153.8	183.4	212.1	233.5	277.8
DEA share (%)	55.3	29.5	18.5	9.5	5.3	0.0

Notes: Claim counts in millions of Part D prescription drug events. DEA share is the percent of events whose prescriber identifier is a DEA registration number rather than an NPI.

truncated at the coding transition. However, since the transition was sharp in 2013, we know that DEA numbers map many-to-one[9] and stably onto NPIs. Thus, we construct a many-to-one DEA-to-NPI crosswalk via a supervised matching. Our process develops a linkage rule that is learned and validated rather than assumed: since most prescriptions even in 2008–2013 are NPI-coded, we can mask the identities of NPI-coded prescribers, ask the algorithm to recover them from the same information available in the DEA case, and measure accuracy against the known answer.

E.2 Matching Algorithm Derived From Fully-Harmonized Part D Claims

We pool each DEA identifier d's beneficiary panel over 2008–2013, restricting to identifiers with at least 100 claims (all drug classes) pooled across years; below this volume, patient-sharing information is too sparse for reliable matching, and such prescribers account for 7.5% of DEA-coded claims. All panels are restricted to Traditional Medicare beneficiaries, for whom Part B claims are observed. We leverage the Part B claims since physician services are always identified by NPIs and many prescriptions recorded under a DEA number would have originated from an NPI-coded

[9]DEA numbers are specific to a state of practice.

physician service. In addition, we observe that prescriber-patient pairs are highly persistent, meaning we can leverage information from after the transition to NPIs to determine the identity of a DEA-coded prescriber.

Thus, for each NPI j and observed DEA number d, we compute two overlap measures: *service overlap* PB_{dj}, the share of d's beneficiary-years in which j billed a Part B service for that beneficiary in that year; and *prescribing continuation* $CONT_{dj}$, the share in which j has NPI-coded prescribing for the same beneficiary in the same or following year extending past d's last observed fill for that beneficiary. The timing restriction screens out predecessor physicians the beneficiary had left, whose prescribing ends before the DEA-coded prescribing does. We restrict to *candidate* NPIs for each DEA number d, which are the set with nonzero overlap on either measure.

Scoring proceeds in two stages. A *ranking* logit of the true-match indicator on PB_{dj}, $CONT_{dj}$, and a nurse practitioner/physician assistant (NP/PA) indicator produces a score s_{dj}, and the top-scoring candidate is selected for each DEA number. An *acceptance* logit, estimated on the top-scoring candidates only, maps the winner's score, its gap over the runner-up, and the NP/PA indicator into an acceptance probability a_d; the match is accepted if and only if $a_d \geq \tau$, and the DEA number is otherwise left unmatched. The gap screens the leading failure mode—physicians in the same practice sharing much of a patient panel—and can enter only at this stage, where a large gap signals that the top scoring candidate is sufficiently differentiated from the runner-up. Evidence for each DEA number is pooled across all years before the single assignment is made, imposing the many-to-one structure by construction.

E.2.1 Training and Performance for Matching Algorithm

We calibrate and test our algorithm in-sample using a 1.0% sample of NPI-coded prescribers active in 2008–2012 ($N = 5,561$ after the volume floor). These prescribers function as "labeled data" since NPI is observed, but we apply our algorithm to them as though NPI is unobserved and use the results to tune our acceptance threshold τ.

The true NPI is in the candidate set for 98.7% of masked prescribers, and the unthresholded top candidate is correct for 80.2%. Precision among accepted matches rises monotonically in τ—from roughly 94% to 99.5% claims-weighted as τ ranges from 0.50 to 0.95—so the threshold transparently trades coverage against precision. At the selected $\tau = 0.60$, 80.6% of masked prescribers (90.7% of their claims) receive a match, with precision among accepted matches of 89.2% unweighted and 96.8% claims-weighted; prescribers below the threshold are left unmatched rather than assigned a low-confidence NPI.

Thus, using in-sample labeled data, we generate a validated algorithm to apply to DEA-coded prescribers 2008–2013. When applied to the 1,470,106 distinct DEA identifiers observed in 2008–2013, the algorithm accepts 248,726 matches, covering 79.5% of DEA-coded claims.

One limitation of our algorithm arises from "incident-to" billing, in which E&M services are delivered by an NP/PA but billed under a supervising physician's NPI. Prescriptions written at this encounter have the NP/PA's identifier. Since office visit evidence is less definitive for non-physicians,

we include the NP/PA indicator within each model stage. We note that the continuation analysis is robust to "incident-to" billing, since NP/PAs identified under a DEA number will transition to an NPI-identifier in 2013 in the same way as physicians. A second limitation is that we assign NPIs using only Traditional Medicare beneficiaries, since we only observe E&M services for those patients. The certainty of our NPI assignment, then, increases with the share of the prescriber's patients who are Traditional Medicare.