Can AI Distort Human Capital?*

Xuelin Li† **Meizi Zhou**‡

This version: April 2024

Abstract

We document that interactions with manipulated AI can distort the development of human capital in opioid prescription contexts. Physicians in our sample adopted electronic health record software from a list of federally certified companies in 2011. Between 2016 and spring 2019, one company secretly embedded a biased AI reminder system to promote extended-release opioid sales. Affected physicians not only increased opioid claims relative to the control group during the treatment window but also maintained a higher propensity for prescriptions even after the removal of the biased function. This long-term distortion of human capital relies on the unconsciousness of AI biases and does not occur following other explicit promotions, such as pharmaceutical detailing payments. Using machine-learning algorithms, we quantify that human capital distortion explains 54% of the treatment effects in a physician decision model with dynamic learning. Experience with opioids, along with caution regarding elder patients, mitigates the distortion.

Keywords: artificial intelligence, human capital, clinical decision support system, opioid, multi-arm causal forest

*We are grateful to Tania Babina, Carri Chan, Hong Luo, Laura Veldkamp for helpful comments and suggestions.

†Columbia University, Graduate School of Business - xuelin.li@columbia.edu

‡Boston University, Questrom School of Business - mzzhou@bu.edu

1 Introduction

The rapid advancement of artificial intelligence (AI) has sparked a growing interest in understanding and mitigating potential risks associated with this innovative technology. Existing literature highlights concerns such as labor displacement (Acemoglu and Restrepo, 2018, 2020; Babina et al., 2022), product market concentration (Babina et al., 2024; Calvano et al., 2020), and social inequalities (Bartlett et al., 2022; Fuster et al., 2022; Howell et al., 2024). In response to these challenges, the European Union (EU) has introduced the AI Act, heralded as the first-ever comprehensive legal framework to harness the power of AI systems in a safe and ethical manner. This risk-based regulatory approach calls for further efforts in assessing other risk areas in AI usage.

In this paper, we uncover a unique risk aspect and argue that human capital can be distorted through interactions with AI systems. With recent advancements in machine learning techniques, AI predictions now commonly assist human decision-making processes (Agrawal et al., 2019). We hypothesize that human capital will adjust through collaboration with AI: these predictions will not only affect the current decision, but also alter workers' habits when similar tasks emerge in the future. However, this "learning-from-AI" phenomenon is dangerous as predictions may have errors and biases. In a worse scenario, our empirical setting illustrates a situation where developers deliberately manipulate these predictions. We show that not only professionals with expertise are vulnerable to these manipulations, but these AI-generated mistakes are also reinforced by user habit formation. As a result, merely removing or fixing the biased algorithms cannot eliminate the biased practices by the agents.

Empirically documenting this effect presents several challenges. First, pinning down the specific influences of AI on human behaviors is difficult due to substantial heterogeneity in human-AI interactions across tasks, agents, and algorithms (e.g Kleinberg et al., 2018; Angelova et al., 2023). Even in similar tasks, a single worker may face various recommended decisions across scenarios, making it challenging to identify alterations in habits consistently. Second, an essential step in this study is to obtain the long-term counterfactual of worker behaviors without further AI influence after their initial interaction. However, workers endogenously adopt and abandon AIs based on their recent experiences. As AI becomes a

"general-purpose technology" (Bresnahan and Trajtenberg, 1995), AI adoption is almost an irreversible decision and workers rarely suspend usage, particularly due to exogenous reasons.

We address these challenges in the context of opioid prescriptions by physicians assisted with clinical decision support systems (CDSSs). CDSS, integrated within EHR systems, uses algorithms on patient data to generate personalized treatment guidance. CDSS can remind clinicians of medical decisions such as prescribing, diagnosing, and testing. This context is economically important for understanding the risks associated with AI. Most healthcare AI systems, including CDSS, are categorized as "high-risk" under the EU AI Act, requiring additional oversight for transparency and robustness. Besides, CDSS is widely used in healthcare provider settings. For example, Jing et al. (2019) show that it is utilized between 71.8% and 100% in primary care settings.

The background of our study is that the Affordable Care Act created financial incentives for physicians to use electronic health record (EHR) software from a list of certified companies in 2011. However, from July 2016 to the Spring of 2019, one major EHR company, Practice Fusion, secretly embedded a pain clinical decision support ("Pain CDS") functionality in its EHR to promote the sales of Purdue Pharma's extended-release opioids (EROs) in an "unbranded effort." This setting provides the following empirical advantages. First, the AI assistance (i.e., Pain CDS) was manipulated to unanimously recommend opioid prescriptions regardless of medical necessity and potential abuse risks but it appeared to healthcare providers as unbiased medical information. This homogeneity allows us to measure an intuitive and single-direction impact on human behaviors. Second, our identification strategy follows a simple difference-in-differences (DID) design, leveraging physician-level annual opioid prescription data from 2013 to 2021. Treated physicians adopted Practice Fusion in 2011, and we can compare their behaviors to other incentive program participants after the shock. Most importantly, our sample spans the period when the bias alert was removed (i.e., 2019 – 2021) so that we can evaluate the long-term distortions on human capital.

We begin with a conceptual framework to guide our quantitative analysis. In the model, physicians make decisions on opioid prescriptions by forming beliefs about the underlying necessity. This belief is jointly decided by a prior formed through observing patient characteristics and an update due to potential pain alerts. Furthermore, we introduce a dynamic

learning mechanism where the periodic prior belief gradually leans towards more opioid prescriptions, i.e., a long-term distortion channel, after each interaction with the manipulated CDSS. The main prediction is that the treatment group will prescribe more LA opioids relative to the control group from 2016 to 2021. For example, in the year of 2018 (the third treatment period), a treated physician suffers both direct manipulation and accumulated biases in his prior due to two periods of learning from AIs. Empirically, we estimate the average treatment effects using regression methods to confirm the existence of both channels. This stylized model also allows us to quantitatively decompose short-term and long-term distortions using a machine-learning algorithm.

We first confirm that in the treatment window from 2016 to 2018, physicians having adopted Practice Fusion exhibit a significantly higher propensity of prescribing long-acting (LA) opioids, i.e., the category encompassing all Purdue Pharma's EROs, compared to the control group. We characterize this propensity through various measures, including number of claims, costs of drugs, days of supply and the fraction of LA opioids over all types of opioid claims. The coefficients are not only highly statistically significant (t-statistics from 2.6 to 4.5) but also economically substantial. Affected physicians increase their annual number of LA claims by 5.9%, generating an aggregated cost of approximately $2.5 million every year in our sample. These results imply that even professionals with expertise follow biased AI recommendations and cannot distinguish intentional manipulations embedded in the system.

If the findings were purely driven by the transient manipulation, we would not observe any significant differences between the treatment and control groups after the removal of the AI manipulation. However, in the post-treatment window from 2019 to 2021, we document evidence of human capital distortion across all different measures: affected physicians continue to prescribe 11.6% more LA opioid claims relative to the control group. This long-term distortion costs Medicare Part D $5.0 million every year in our sample, but also generate long-term real healthcare consequences for beneficiaries. For example, in each 3-digit zip code area, a 1% increase in the number of Practice fusion adopters is associated with 0.48% more Medicare beneficiaries requiring treatment for opioid abuse from 2020 to 2021. To establish the generality of our results, we also measure the number of hospital visits due to opioid overdose using a database covering 43 million privately-insured individuals. Consis-

tent with the findings from Medicare beneficiaries, a 1% increase in the number of Practice Fusion adopters in an MSA is associated with 0.131% more overdose payments in the outpatient setting and 0.133% in the inpatient setting by private payers. Equivalently, this implies that one additional increase in the number of treated physicians from the mean leads to an increase of $40,000 outpatient payments and $88,000 inpatient payments.

We confirm the result robustness by using alternative empirical specifications and placebo tests with regulated medications not promoted by Practice Fusion. We also estimate the coefficient dynamics across all calendar years to characterize the time series of impacts. Supporting the parallel trend assumption, the two groups exhibit no significant differences before 2015. When the pain alert was active, affected physicians gradually increased LA opioid usage over the years. After 2019, this trend flattens, but the treatment group still maintains a significantly higher propensity for opioid prescriptions. This pattern reinforces the key assumption in our conceptual framework that each periodic interaction with manipulated CDSS will repeatedly accumulate belief biases.

Furthermore, AI manipulation has the potential to distort human capital beyond the scale of traditional financial incentives. Given that Purdue Pharma's public image became increasingly negative in the period of our analysis, we argue that the unawareness of AI manipulation plays a unique role in habit formation. To show this, we focus on the alternative group of physicians having financial relationships with Purdue Pharma, who explicitly received in-kind payments from the company. Purdue Pharma suspended these detailing activities also in 2019, coinciding with the removal of the pain alert. In stark contrast, recipients exhibit a significantly lower volume of LA claims compared to the control group post-payment suspension, displaying no long-lasting persistence in prescriptions. This finding aligns with Purdue Pharma's motivation for contracting with Practice Fusion, claiming that direct EROs promotions were hindered by providers' political pressure. Instead, physicians are unconscious of the bias in AI predictions and perceive the recommendations as optimally generated from real data. Through collaborating with the EHR, physicians gradually update more favorable beliefs about the benefit-risk profile of EROs over time, leading to an inclination towards long-term prescriptions.

Lastly, we use a machine-learning algorithm built upon Nie and Wager (2021) to esti-

mate the economic magnitude of long-term belief distortion within our conceptual framework. This algorithm uses the prescription probability of LA opioids over all claims as the outcome variable. During 2016 to 2018, it can decompose the aggregate treatment effects for each affected physician into components attributed to direct CDSS manipulation and long-term belief distortions. In the data, a typical treated physician has a higher prescription probability of 2.3 bps, or 9.5% of the unconditional average. The predicted prescription probability by our algorithm closely matches this magnitude, with an average treatment effect of 2.4 bps. We generate counterfactual prescription probabilities assuming no long-term distortion or that belief distortion does not accumulate after more interactions with biased AI. The treatment effect will decrease by 54% (to 1.1 bps) in the first case and by 17% (to 2.0 bps) in the second case. These magnitudes underscore the importance of human capital distortion in explaining the observed prescription changes.

Our algorithm also allows us to study the heterogeneity of treatment effects by estimating them as a function of observed characteristics. We hypothesize and test two channels through which affected physicians are more robust to AI biases. First, physicians with more previous experience in LA opioid usage are more likely to make decisions based on their expertise, and less likely to be influenced by AI recommendations. For instance, our algorithm predicts that a treated physician without any previous LA opioid usage will have larger responses (0.023% v.s. 0.017%) to direct manipulation in 2016. Second, our algorithm shows that average patient age of each physician has the highest feature importance in determining treatment effect heterogeneity. Intuitively, physicians who interact with senior patients are more cautious about the potential opioid side-effects. In terms of magnitudes, the high patient age group has a net treatment effect of 0.9 bps based on prescription probability, roughly 30% of the effect of the low-age group (3.2 bps).

2 Related Work

This paper closely relates to the literature on the potential risks of AI adoption. Mostly related is the literature on the labor market implications of artificial intelligence, including but not limited to Acemoglu and Restrepo (2018), Autor and Salomons (2018), Acemoglu and

Restrepo (2020), Acemoglu and Restrepo (2022), and Babina et al. (2022). Instead of focusing on labor share, composition or wage, we study the development of human capital through AI interactions. On this end, existing literature focuses on worker skills across various tasks that are exposed to AI replacements or compatible with AI capacities (Brynjolfsson et al., 2018; Felten et al., 2018; Webb, 2019; Acemoglu et al., 2022). Our paper differs as a "within-task" analysis that studies how working with AI will change the long-term performance of professionals with specific skills. Our empirical results echo the theoretical prediction in Acemoglu (2021). He argues that AI assistance will diminish workers' understanding of automated tasks, which in turn reduces the productivity in their specialized fields requiring human judgements. Instead, we document that agents will reinforce the biases and errors in algorithms in other settings even without further AI recommendation.

Also related is the impact of AI adoption on consumer welfare. Algorithms may unintentionally amplify biases embedded in unrepresentative training samples, creating fairness concerns (Bartlett et al., 2022; Cowgill and Tucker, 2019; Fuster et al., 2022; Howell et al., 2024). Moreover, AI technology may give rise to superstar firms (Babina et al., 2024) and allow sellers to implicitly coordinate in price settings (Calvano et al., 2020), generating concerns over the monopoly power of companies. We document the costs of AI from a different perspective: firms can deliberately manipulate the predictions, exploiting humans' trust on AI, to achieve specific goals. Even professionals with expertise will have a hard time figuring out the manipulation. This manipulation ultimately hurts the welfare of customers and even endangers their lives.

Our work also adds to the literature of understanding the role of information technologies in the healthcare industry. For discussion and overview, see Bronsoler et al. (2022) and Dranove and Garthwaite (2022). Examples of CDSS benefits include reducing adverse medical events (Hydari et al., 2019), attenuating racial biases in healthcare delivery (Ganju et al., 2020), and avoiding high-cost orders (Doyle et al., 2019). Closely related is Agarwal et al. (2023). The authors document that professional radiologists do not fully capitalize on the potential gains from AI assistance due to deviations from Bayesian belief updating. Radiologists may underweight the AI's information relative to their own signals. In their work, the provision of AI assistance will not gradually change the acquisition and interpretation of

these prior signals by the professionals. Our work complements theirs by documenting how experience with AI can change the long-term beliefs of professionals, which can dynamically affect the deviations documented in their paper. Lastly, this paper falls into the broad literature comparing human decisions to algorithm-based recommendations in tasks such as diagnosing (Mullainathan and Obermeyer, 2022), judging (Kleinberg et al., 2018; Angelova et al., 2023), asset management (D'Acunto et al., 2019; Rossi and Utkus, 2020), and financial analysts (Cao et al., 2021; Coleman et al., 2022; Grennan and Michaely, 2020).

3 Institutional Details and Data

3.1 Clinical Decision Support Systems

Clinical Decision Support Systems (CDSS) are AI tools that can be either rule-based expert systems or machine learning (ML)-based systems. Rule-based AI replicates the decision-making of domain experts in various fields. For example, physicians specializing in knee replacement may develop CDSS with customized "order sets" to optimize treatment plans, including lab tests, medication, patient instructions, surgical instruments, and home-care procedures. Each aspect of the order set may include optional or alternative steps to address expected care variations. In practice, users (e.g., physicians and nurses) will receive reminders and recommendations during the whole patient care procedure (Sloane and Silva, 2020). Instead of relying on expert-generated rules, ML-based systems utilize statistical algorithms to generate strategies, by processing data such as doctor notes and scan images. For example, radiologists at Geisinger Healthcare in Pennsylvania used a convolutional neural network to analyze head CT scans. This algorithm greatly improves the accuracy and reduces the handling time for detecting intracranial hemorrhage (Arbabshirani et al., 2018).

Both rule-based and ML-based AIs are complex. It is challenging to verify whether they operate as intended and ensure their functions are safe and optimal for current medical care (Sloane and Silva, 2020). Consequently, fraud and misinformation in AI-enabled healthcare software are significant concerns. For instance, software developers may exaggerate the algorithm's capabilities and overstate its clinical effectiveness. IBM claims that Watson for

Oncology can recommend cancer treatments to doctors based on algorithms trained by vast clinical data. Instead, the recommendations are actually derived from laborious human input (Ross and Swetlitz, 2017). Alternatively, physicians may deliberately choose one synonym over another in AI-based systems for more reimbursement or prescribing regulated medications (Cerrato and Halamka, 2020; Finlayson et al., 2019). As a result, government agencies are cautiously developing regulatory frameworks to promote innovation in this space. In the U.S., before the Food and Drug Administration approved the marketing of the first-ever autonomous AI-based diagnostic system in 2018, it established the Digital Health Software Precertification program, which assesses a firm's underlying quality in ensuring software products meet safety and effectiveness standards. In Europe, the EU AI Act categorizes medical software used to diagnose patients, make therapeutic decisions, and monitor physiological processes as high-risk AIs, subjecting them to additional regulatory requirements.

Our empirical setting focuses Practice Fusion's CDSS, which is a rule-based AI, allowing us to directly identify the manipulation generated by the system. Practice Fusion was an EHR company founded in 2005. At one point, it was named the top EHR for customer satisfaction among primary care providers and ranked No. 1 for value among ambulatory professionals. Investors including Kleiner Perkins Caufield & Byers made a $70 million Series D investment in September 2013. This new round valued the company at around $700 million, making it one of the largest digital health startups at the time.[1] Practice Fusion's EHR provides clinical decision support from various aspects, including generating differential diagnoses based on patient symptoms, sending notifications for necessary tests during workflows, and issuing prescribing alerts for drug history, drug-drug interactions, allergies, compliance, and drug dosage. Practice Fusion claimed that the EHR continuously updates with regulatory changes through policy exports and provider feedback, and clinicians do not need to take any action as the EHR automatically receives updates.

Beginning in Fall 2013, Practice Fusion solicited remuneration and negotiated with a pharmaceutical company "Pharma Co.X," later identified as Purdue Pharma, to create and embed a pain clinical decision support ("Pain CDS") functionality in its EHR.[2] The pur-

[1] "Practice Fusion Lands A Whopping $70M To Bring A Big Data Cure To The Healthcare Crisis," TechCrunch, September 2013.

[2] All details and quotes in this section, unless otherwise specified, are from Exhibit C "Statement of

pose is to increase prescriptions for Purdue Pharma's three ERO products. During their communication, Purdue Pharma expressed concern that "providers are hesitant about using high dosages to combat pain for a variety of reasons, mostly political pressure." In response, Practice Fusion proposed developing a Pain CDS to initiate ERO products and embedding Purdue Pharma's ERO products as treatment options in an "unbranded effort."

The two companies signed the official contracts on March 1, 2016, in which Purdue Pharma agreed to pay Practice Fusion approximately $1 million. The Pain CDS functionality went live on July 6, 2016, providing a series of alerts and recommendations for both diagnosing and treatments. First, it recommended providers to document a pain score for the patient. Second, it displayed a "Brief Pain Inventory" with the patient's previous pain scores within the previous three months. Then the provider would go through a list of questions on the severity and impact of patient's pain, summarizing the patient's current pain as "worst," "on average," and "least" in the previous 24 hours. If the patient reported a pain score of four or higher twice within four months, or the patient was diagnosed with chronic pain after completing the Brief Pain Inventory, the CDS utilized a drop-down menu of options for pain management. Practice Fusion added an "Opioid Therapy" treatment option without considering the patient's condition. In particular, ERO products are listed for patients with less than severe or non-chronic pain, and even in cases where the pain could be treated with non-ERO options. This treatment option was combined with other options in the list such as non-opioid analgesics and pain-management, which were sourced from a 2016 New England Journal of Medicine article titled "Opioid Abuse in Chronic Pain — Misconceptions and Mitigation Strategies" (Volkow and McLellan, 2016), although the intention of the paper is not to provide a clinical instruction in chronic pain management.

On December 14, 2016, Practice Fusion conducted a presentation at Purdue Pharma's headquarter. The meeting revealed that through November 30, 2016, the CDSS had alerted during 21 million visits involving 7.5 million patients, generating a general shift toward EROs particularly from immediate-release opioids. The CDSS alerted more than 230,000,000 times from July 6, 2016 to Spring 2019, in which it was suspended right before Purdue Pharma filed for bankruptcy.

Facts" of Case 2:20-cr-00011-wks, United States Attorney's Office for the District of Vermont.

3.2 EHR Adoption

Our data construction starts with a granular record of eligible physicians adopting different EHR products since April 2011, downloaded from the HealthIT.gov website. In 2009, Congress passed and President Obama signed the Health Information Technology for Economic and Clinical Health (HITECH) Act, encouraging the adoption of health IT to improve the quality and efficiency of care. The Act created incentive programs that offer physicians financial benefits for the meaningful use of EHRs. To start, the CMS and the Office of the National Coordinator for Health Information Technology (ONC) have established standards for EHRs to become certified. Practice Fusion's EHRs were certified in 2011. Eligible professionals (EPs), including most U.S. physicians, can adopt certified EHR technology and attest to fulfilling meaningful use criteria to receive incentive payments. Each physician must achieve three stages with different criteria to receive full payments.[3]

We focus on the Medicare EHR Incentive Program administered by the CMS from 2011 to 2016, providing a maximum of $44,000 in total payments across years.[4] A participating EP must demonstrate meaningful use every year to receive payments, and the CMS publicly releases a complete record of attestation information, enabling researchers to track each EP's EHR products over time. The HealthIT.gov streamlines and disambiguates the raw data into a single "EHR Products Used for Meaningful Use Attestation" file. In each attestation year, this file lists all the certified EHR products attested by an EP (identified by the NPI). For each record, it supplements the provider zip code, specialty, vendor name, product version and product setting. The initial sample contains over 1.8 million records for almost 360,000 NPIs using 1,232 EHR products owned by 730 vendors.

We impose the following major restrictions on the initial sample. First, certified EHR products span various application categories, including ambulatory, cardiology imaging, financial decision support, human resources and IS security, etc. All Practice Fusion products are ambulatory EHRs. To maintain comparability, we drop all the EHR products without

[3]These stages include data capturing and sharing (stage 1), advanced clinical process (stage 2), and improved outcomes (stage 3).

[4]Alternatively, EPs can participate in the Medicaid EHR Incentive Program run by every state. However, there is no centralized disclosure platform for Medicaid programs. But they can only participate in one program.

any ambulatory applications. There are over 1 million adoption records of the remaining 190 products by roughly 210,000 physicians. Second, we focus on the regions (3-digit zip code areas) with non-zero physicians adopting Practice Fusion products to alleviate the confounding effects due to heterogeneous healthcare conditions across regions. In these regions, we include only physicians with the same specialty as the Practice Fusion adopters in the control group to account for practice differences across various medical settings. Lastly, we drop physicians using four vendors with other misbehaviors in the Incentive Program, including Athenahealth, Nextgen Healthcare, Modernizing Medicine (ModMed), and Greenway Health. Their violations include missing the required functionality to become certified and paying physicians to falsely attest.[5] None of these vendors promoted opioids or received kickbacks from opioid drug makers. The rationale is that physician practices may be distorted by these EHR products as well, questioning their validity as clean control groups. The final sample has roughly 27% (57,138) of the remaining physicians following the first step. To sum up, these physicians participate in the Medicare EHR Incentive Program and use EHRs with ambulatory functions. They belong to either the treatment group, adopting Practice Fusion products, or the control group, comprising comparable neighbors of Practice Fusion users.

3.3 Opioid Prescription

We then link the above sample to the "Medicare Part D Prescribers by Provider" dataset by the CMS, containing annual prescription information by individual physicians under the Medicare Part D Prescription Drug Program from 2013 to 2021. Relevant to our study, it provides the total number of claims, costs, and days of supply of opioid drugs each year. Besides, it provides the same information for long-acting opioids. These are drugs formulated to release more gradually into the bloodstream with a longer duration of analgesic action. Purdue Pharma's dominant ERO products, e.g., OxyContin, are all long-acting. We then back out the short-acting (SA) opioid claim amount and also utilize antibiotic and antipsychotic

[5]The only exception is that ModMed solicited and received kickbacks from Miraca Life Sciences Inc. (Miraca) in exchange for recommending Miraca's pathology lab services. However, we cannot perform the analysis based on this case since we do not have utilization data of Miraca's services.

drugs for placebo tests. As these drugs are not promoted by Practice Fusion, the treated physicians should not exhibit significant increases in SA opioid prescriptions throughout the sample.

We include a few patient characteristics as control variables. These include the average age of beneficiaries, the percentage of male, African American, and Hispanic beneficiaries, the percentage of beneficiaries qualified for both Medicare and Medicaid benefits, and the average risk score of beneficiaries. The last variable is a hierarchical condition category developed by the CMS based health-influencing factors.[6]

3.4 Purdue Pharma Detailing

Apart from manipulating the CDSS, Purdue Pharma's primary marketing strategy involves directly detailing to physicians. Indeed, since Purdue Pharma introduced OxyContin in 1996, it aggressively engaged in detailing to promote the product, leading to the decades-long opioid crisis (Van Zee, 2009; Alpert et al., 2022). We retrieve Purdue Pharma's detailing information from the CMS Open Payments database. This database collects and reports financial relationships between drug and medical device companies and providers since 2013, involving detailing payments such as research, meals, travel, gifts or speaking fees. We record whether physicians in our sample receive payments from Purdue Pharma in a particular year. Section 5.2 compares the long-term effects of this direct marketing strategy with the indirect manipulation of the decision support system on prescription behavior.

4 Research Design

4.1 Conceptual Framework

Our conceptual framework features a physician's decision making process by maximizing the chance of correct prescription in a setting of dynamic learning. For physicians i facing case j at year t, they take a binary decision $a_{ij} \in \{0, 1\}$, where $a_{ij} = 1$ indicates prescribing

[6]These factors include the beneficiary's age, sex, eligibility for Medicaid, initial reason for Medicare qualification, residence in an institution such as a long-term care facility, and the diagnoses assigned to the beneficiary in inpatient, outpatient and office-based settings during a base year.

LA opioids. The underlying state $\omega_j \in \{0,1\}$, unobserved to the physician, indicates the true fitness of opioid usage. Physicians have disutility from both unnecessary prescription (choosing $a = 1$ when $\omega = 0$) and under-treatment (choosing $a = 0$ when $\omega = 1$), with payoffs modeled as

$$u_i = -\mathbf{I}\{a = 1, w = 0\}c_1 - \mathbf{I}\{a = 0, w = 1\}c_2. \tag{1}$$

In each case j, physicians observe patient characteristics Z_j and possibly a Practice Fusion pain alert, indicated by $s \in \{0,1\}$. They then form a belief $p_{it}(\omega|Z_j, s)$ and generate the likelihood ratio as

$$l_{it}(Z_j, s) = \frac{p_{it}(\omega = 1|Z_j, s)}{p_{it}(\omega = 0|Z_j, s)} = \underbrace{\frac{p_{it}(\omega = 1|Z_j)}{p_{it}(\omega = 0|Z_j)}}_{q_{it}(Z_j)} \times \underbrace{\frac{p_i(s|Z_j, \omega = 1)}{p_i(s|Z_j, \omega = 0)}}_{\delta_i(Z_j, s)}.$$

The decomposition follows from the Bayes rule and has an intuitive interpretation. $q_{it}(Z_j)$ is the prior belief of physician i observing Z_j in case j at year t. $\delta_i(Z_j, s) \geq 1$ represents the periodic belief update due to CDSS alerts. For simplicity, we assume δ_i is a static function (independent of t) and physicians do not change their beliefs without seeing the alert, i.e $\delta_i(Z_j, 0) = 1$. To model the long-term belief distortion, denote k by the number of years that a physician j has observed positive CDSS alerts by t. We assume that every additional period of exposure to CDSS manipulation will bias the likelihood upward by a certain degree:

$$q_{it}(Z_j) = q_{i0}(Z_j) \prod_{m=1}^{k} \gamma_i^m.$$

$q_{i0}(Z_j)$ is the initial belief without learning. $\gamma_i^m \geq 1$ is the belief bias due to the m^{th} interaction with manipulated CDSS. Note that γ_i^m is implicitly a function of patient characteristics in that interaction. Finally, the payoff function in Equation (1) implies that physicians follow a simple cut-off strategy by

$$a_{it}(Z_j, s) = \mathbf{1}\left\{ q_{i0}(Z_j) \times \left(\prod_{m=1}^{k} \gamma_i^m \right) \times \delta_i(Z_j, s) \geq \frac{c_1}{c_2} \right\}. \tag{2}$$

The main prediction of this conceptual framework is that the treatment group will prescribe more LA opioids relative to the control group from 2016 to 2021, if $\delta_i(Z_j, s) \geq 1$ and $\gamma_i^m \geq 1$ (and these inequalities strictly hold with non-zero probabilities). To illustrate, consider two otherwise similar physicians but one uses Practice Fusion (treated) and one does not (control) in the following three cases. First, in the year of 2016 (the initial treatment period), both physicians have not interacted with manipulated CDSS, and thus the second component in Equation (2) is one. The treated physician has a higher prescription volume only due to the instantaneous belief update $\delta_i(Z_j, s)$. Second, in the year of 2020 (a post-treatment period), the last component equals one since the alerts are removed. But the treated physicians still have higher chances of prescription due to $\prod_{m=1}^{3} \gamma_i^m$ (three years of previous interaction). Lastly, in the year of 2017 (the second treatment period), the treatment group suffers both manipulation $\delta_i(Z_j, s)$ and belief distortion γ_i^1 after one period of learning. Empirically, we will first estimate the average treatment effects using regression methods to confirm the existence of both channels, as detailed in the next section. We then quantitatively decompose the effects from δ and γ using machine-learning algorithms in Section 6.

Model Discussion We interpret physician human capital or skills as the precision of belief $l_{it}(Z_j, s)$. This definition of skill is consistent with Chan et al. (2022) and Agarwal et al. (2023), or noted as decision making quality in Currie and MacLeod (2017). There are two alternative ways to interpret γ. One is through physician preference (e.g. Abaluck et al., 2016): previous LA opioids prescription makes physicians less concerned about the risk of overdosing (a smaller c_1 over time). This interpretation is equivalent to moving $\prod_{m=1}^{k} \gamma_i^m$ to the right-hand-side of Equation (2). The other one is through mis-belief in their treatment skills when prescribing opioids (e.g. Chandra and Staiger, 2020). For example, previous LA opioids prescription makes physicians more confident in using EROs to treat any type of patients. This interpretation is equivalent to adding a positive utility of unconditionally prescribing opioids ($a = 1$) in Equation (1). Both interpretations will generate long-term reductions in the cutoff threshold. Given our data limitation, we cannot distinguish these interpretations (see Chan et al., 2022) but the predictions are all consistent.

4.2 Empirical Strategy

Our main sample includes annual observations of physician EHR usage and opioid prescriptions from 2013 to 2021. This sample spans three phases: pre-treatment (2013 – 2015), treatment (2016 – 2018), and post-treatment (2019 – 2021). Our primary specification is essentially a simple DID design:

$$Y_{i,k,t} = \alpha + \beta PF_i \times Post_t + \gamma' X_{i,t} + \delta_i + \mu_{k,t} + \varepsilon_{i,k,t}. \tag{3}$$

In the above equation, $Y_{i,k,t}$ denotes the opioid prescription made by physician i in area k during year t. PF_i indicates whether physician i adopted Practice Fusion products in the Medicare EHR Incentive Program. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $X_{i,t}$ indicates the set of control variables capturing patient characteristics, detailed in Section 3.3. We include two-way fixed effects (FEs) — the physician FE, δ_i, and the area-year FE, $\mu_{k,t}$ — which will absorb the standalone effects of PF_i and $Post_t$. This specification effectively compares the opioid prescription of two otherwise similar physicians in the same area and year, where one is exposed to the Practice Fusion manipulation, and the other is not.

To estimate both the short-term and long-term effects, we estimate Equation (3) separately in the following two samples. The first sample is truncated at 2018, including only the pre-treatment and treatment phases, and demonstrating the direct effects of CDSS manipulation. Using the notations from our conceptual framework, the coefficients estimated from this sample represent the joint effects from both manipulation $\delta_i(Z_j, s)$ and belief distortion γ_i^m. The second sample drops the yearly observations from 2016 to 2018, consisting of the pre-treatment and post-treatment phases, and exhibiting the persistent impacts on human capital. As previously explained, any significant differences between the treatment and control groups are fully attributed to $\prod_{m=1}^{k} \gamma_i^m$ in this sample.

Table 1 Panel A summarizes the main variables in the whole sample (all three phases). 6,244 unique physicians, i.e. more than 10% of the sample, belong to the treatment group. On average, each physician prescribes 13.35 annual LA claims, supplying 382 days of usage and costing around $2,600 for all the beneficiaries. LA claims account for 4.26% of all the

opioid claims. Since our sample focuses on the Medicare Part D prescription, the average beneficiary in our sample is above 70 years old and more risky (139%) than the national average (normalized as 1). Unconditionally, each physician has a 4% chance of receiving promotional payments from Purdue Pharma.

Our identification relies on the parallel trend assumption, which requires that the treatment and control groups exhibit similar tendencies in LA opioid prescriptions without Practice Fusion's impacts. This assumption likely holds in our setting for the following reasons. First, the self-selection concern is mitigated by the timing difference. Treated physicians participated in the EHR incentive program and adopted Practice Fusion's products in 2011. The actual manipulation occurred in 2016. It is almost implausible that physicians could predict the manipulation or intentionally adopted a product generating biased recommendations. Second, there exists an alternative explanation that the treatment group had a stronger preference for LA opioids. They insisted on prescribing Oxycontin after Purdue Pharma suffered from litigation and reputation losses in recent years, while the control group was more sensitive to the negative news. Contrary to this hypothesis, Table 1 Panel B shows that, if anything, the treatment group were more conservative in LA opioids. Compared to the control group, they supplied significantly 51 fewer days of drugs, and their prescription rate was also 0.8% lower. More importantly, the two groups exhibit no significant differences in receiving financial payments from Purdue Pharma. Lastly, we will further visualize the coefficient dynamics, demonstrating no significant trends before the shock. Besides, we perform a placebo test that shows the null effects on SA opioids or other regulated medicines, i.e., products not recommended in Practice Fusion's CDSS.

5 Regression Results

5.1 Opioid Prescription

We start by estimating Equation (3) in the initial sample from 2013 to 2018 in Table 2 to investigate whether the treatment group significantly prescribed more LA opioids when Practice Fusion manipulated the CDSS. This analysis provides a first-stage result to confirm

the existence of direct impacts in the short term. One might argue that as experienced professionals, affected physicians will use their discretion to reject the manipulated recommendation and avoid unnecessary opioid prescriptions. Instead, Table 2 indicates that physicians utilizing Practice Fusion EHRs experienced a significant 5.9% increase in annual LA claims compared to the control group, equivalent to approximately 0.8 additional claims each year, from 2016 to 2018. Meanwhile, their total LA opioid costs surged by 15.4%, accompanied by a 11.4% increase in days of supply. A ballpark estimate of the total additional costs for Medicare Part D in our sample is approximately $2.5 million annually.[7] This magnitude closely aligns with Practice Fusion's own prediction, ranging between $8,458,232 and $11,277,643, in opioid revenue driven by the CDSS.

Table 2 also includes a placebo test to validate our identification assumption. As we explained earlier, Purdue Pharma's leading products are extended-release opioids utilized for chronic pain management, suggesting the manipulation should have minimal effects on SA opioids. Column (4) confirms that the proportion of LA claims to total opioid claims rises by 0.24 percentage points compared to the control group, corresponding to a 6% increase from the sample average (4.26%). Lastly, column (5) shows that affected physicians do not exhibit any significant differences in SA claims after the shock, confirming that the changes were concentrated in LA opioids. The negative coefficient, though insignificant, is in line with Practice Fusion's claim that the CDSS alert generated a shift from SA opioids to ERO products.

Next, we estimate Equation (3) in the sample consisting of both pre-treatment and post-treatment phases. If the distortion in opioid prescription was only transient, then affected physicians would not exhibit any significant changes in LA claims from 2019 to 2021. However, Table 3 demonstrates that affected physicians continue to prescribe more LA opioid claims even after the CDSS stopped making recommendation after the Spring of 2019. The economic magnitudes are even larger. For example, column (1) implies a 11.6% increase in LA claims, nearly double the corresponding coefficient in Table 2. Rises in LA opioid costs for affected physicians become 30.8%, implying a $5.0 million aggregated additional payments

[7]The calculation is based on the sample average annual cost $2,623, multiplied by the coefficient and the number of physicians in the treatment group (6,244).

by Medicare Part D every year in our sample. The larger magnitude is consistent with our conceptual framework, as long-term bias repeatedly accumulated during the treatment window. This result implies that, through learning with AI, the short-term manipulation will generate persistent long-term impacts on human capital. Affected physicians develop a habit of opioid usage and ultimately maintain the tendency to prescribe opioids, regardless of the CDSS recommendation.

Tables 2 and 3 estimate the average treatment effects in different phases. To visualize the coefficient dynamics in the whole sample, we estimate and plot the coefficients β_c from the following equation in Figure 1:

$$Y_{i,k,t} = \alpha + \sum_{\substack{c\in[2013,2021] \\ c\neq 2015}} \beta_c PF_i \times \mathbf{I}(t=c) + \gamma' X_{i,t} + \delta_i + \mu_{k,t} + \varepsilon_{i,k,t}. \tag{4}$$

$\mathbf{I}(t=c)$ indicates whether year t is calendar year c. β_c estimates the difference between the treatment and control group at calendar time c. Following the convention, we drop the year 2015, the year before the treatment, making it as the reference year in the above equation. There are two main takeaways from the figure. First, there exist no significant differences between the two groups in the pre-treatment phase (before the red vertical line), supporting the parallel-trend assumption. Second, while Practice Fusion manipulated the CDSS (between the two vertical lines), affected physicians gradually increased LA opioid usage over the years due to increased belief distortions. While this momentum stopped after the suspension of biased recommendation in 2019, the treatment group maintained a constantly high frequency of prescription. These dynamics explain the magnitude differences in Tables 2 and 3. This pattern also supports the key assumption in the conceptual framework: each periodic interaction with the CDSS will lead to more long-term biases.

We perform a few robustness checks of Tables 2 and 3 in the Appendix. First, we show that our results hold with a more restricted geographic boundary (5 digit zip code areas) for the control group in Table B.1. This restriction will generate fewer observations in the sample, as expected. But the coefficient magnitudes and statistical significance become even larger. Second, we show that imposing granular $Area \times Year$ fixed effects is not necessary for our results. Table B.2 shows that our results are robust to using the separated $Area$ fixed

effects and $Year$ fixed effects. Lastly, we consider other regulated medicines in Medicare Part D, which are antibiotics (due to concerns of antimicrobial resistance) and antipsychotics (due to side effects such as obesity and metabolic disorder). We do not find that treated physicians significantly increase prescriptions of these drugs in Table B.3. Consistent with column (5) of Table 2, the estimated coefficients are all negative, suggesting a potential crowd-out effect of long-acting opioid prescription.

We supplement additional evidence on the long-term impacts from the perspective of real costs on beneficiaries. Starting from 2020, Medicare covers Opioid Treatment Programs (OTPs) that provide medication-assisted treatment for people diagnosed with an opioid use disorder (OUD) in Medicare Part B. Through the "Medicare Physician & Other Practitioners" file, we can aggregate the area-level Medicare utilization of OTPs based on the HCPCS codes and use it as a proxy for local opioid addictions among the elderly group. We prefer this measure over other public data, such as mortality due to drug overdose, for two reasons. First, although we cannot track the actual OUD conditions of the identified beneficiaries who received LA opioids from the treatment group, the aggregated utilization is a fuzzy but consistent measure from the same Medicare group. Second, overdose mortality is an extreme situation that occurs with a lower frequency, leaving us with smaller variation in the data.

Because the coverage began in recent years and did not span the pre-treatment phase, we acknowledge the limitation of this analysis as correlational evidence and interpret the results with caution. In Table 4, our focal regressor is the number of unique physicians who adopted Practice Fusion in the Medicare Incentive Program in each 3-digit zip code area. The outcome variables include the number of beneficiaries, number of services, bill charges, and actual Medicare payments for OTPs. Control variables include local demographic characteristics such as logged population, logged income per capital, male percentage, African American percentage, Hispanic percentage, average age, unemployment rate, and health insurance coverage percentage. We find that Practice Fusion usage among providers indeed correlates with long-term OUDs among the Medicare beneficiaries. For example, a 1% increase in the number of Practice Fusion adopters is associated with 0.48% higher Medicare beneficiaries and 1.245% higher Medicare payments for OTPs.

In this paper, we focus on Medicare because we have physician-level identifiers and consis-

tent measures of opioid prescription and OTP for the same group of beneficiaries. However, the concern is that almost all Medicare beneficiaries are the elder population, and our results cannot be generalized due to heterogeneous patient demographics in other insured populations. To address this concern, we provide consistent evidence of long-term real costs using the Marketscan database. The data provider sources commercial insurance claim information for employees, retirees, and dependents from over 260 medium and large employers and 40 health plans. The initial database covers over 43 million privately-insured individuals with employment-based health plans, representing roughly 14% of all insured. Due to privacy concerns, the data provider de-identifies the physician information so we cannot link Practice Fusion adoption to individual service providers in this database. Moreover, granular patient geographic identifiers, such as the zip code or the FIPS code, is also redacted and we can only rely on the MSA information. As a result, our estimation effectively concentrates on the metropolitan population.

We follow a similar strategy as Table 4 by defining the number of unique physicians that adopted Practice Fusion in the Medicare Incentive Program in each MSA area. Notice that this is still a valid measure in Marketscan because physicians testify meaningful use of the EHR with patients insured by all health plans, not just Medicare. The word "Medicare" in the program name is due to the federal sponsorship (CMS, rather than state medical boards). In Marketscan, we have both inpatient (overnight hospital stays) and outpatient (non-staying) visits of each patient. Based on the diagnosis code, we can infer whether the patient has symptoms of drug overdose. Then for each MSA, we aggregate the number of visits due to drug overdose and the total payments (insurance reimbursement and patient copay) associated with these visits.

Table 5 shows consistent evidence that higher Practice Fusion adoption in a given area leads to long-term overdose problems in the same place. A 1% increase in the number of Practice Fusion adopters is associated with 0.13% more overdose payments in both the outpatient and the inpatient setting. Putting these coefficients into economic magnitudes, a typical MSA has on average 7 Practice Fusion adopters, and one additional treated physician represents a 5 percentage point increase in $Log(NumPF)$ from the mean. The average annual total payment due to drug overdose is around \$600,000 for outpatient services and

$1.3 million for inpatient services in each MSA. So these coefficients imply an estimated increase of $40,000 outpatient payments and $88,000 inpatient payments associated with that marginal treated unit.

5.2 Comparison to Detailing

In this section, we demonstrate that human capital distortion exists only when physicians are unaware of the biased recommendation. To establish, we study an alternative group of physicians that receive explicit in-kind payments from Purdue Pharma and evaluate their LA opioid prescription behaviors over time. As shown in Figure A.1, Purdue Pharma detailing activities peaked in 2014, where almost 4,000 physicians in our sample received promotion payments. Recipient numbers steadily decreased, leaving only one physician with a financial tie in 2019. There exist two potential reasons for this trend. First, the CMS Open Payments system was implemented online in 2014, retrospectively revealing all the in-kind payments from August 2013. Physicians might worry about their public images being damaged by disclosed payments. Second, after Purdue Pharma's 2015 settlement with Kentucky, other states began suing rapidly. By January 2019, 36 states had sued Purdue Pharma. The potential litigation risk might make the company more cautious with promotion.

Recall that our treatment group, i.e., Practice Fusion adopters, continue increasing LA opioid prescription even in the background of Purdue Pharma's negative publicity in this window. Table 6 Panel A replicates Table 3 by comparing prescription behaviors before (i.e., 2013 to 2015) and after (i.e., 2019 to 2021) treatment. $Detail_i$ indicates whether physician i has received in-kind payments before 2019. Note that Purdue Pharma effectively suspended detailing activities after filing for bankruptcy in 2019. Therefore, Table 6 allows us to evaluate the possibility that detailing recipients maintain high opioid prescription rates without further financial incentives.

However, former payment recipients substantially reduce LA opioid prescriptions after 2019. Their LA claim amount on average drops by 43% compared to the pre-treatment period before 2015. Indeed, Figure 2 shows that their prescription frequency steadily decreased since 2015, consistent with the trend that Purdue Pharma detailing became less popular in the same window. This result highlights that even though financial incentives can also contribute

to transient jumps in LA opioid usage, they cannot foster a long-term preference. Physicians may simply decide on prescriptions as a function of financial benefits and even worry about litigation and reputation when lawsuits started to target Purdue Pharma. Therefore, we argue that a key feature of artificial intelligence on human capital is the unconsciousness of behavior changes. Our results mirror the literature studying how financial incentives affect prosocial behaviors (Bénabou and Tirole, 2006). This literature argues that these rewards may crowd out prosocial behaviors because receiving payments will damage a person's social reputation. Instead, this paper shows that rewards are necessary to compensate agents to take antisocial behaviors at the cost of negative publicity.

Figure A.2 shows that the geographic distribution of the two physician groups is highly uncorrelated ($\rho = -0.19$), further confirming our argument that Practice Fusion adopters did not have other incentives to prescribe LA opioids. Table 6 Panel B, as a robustness check, confirms our main findings from Table 3 remain valid after accounting for physicians receiving detailing payments. From Purdue Pharma's perspective, manipulating Practice Fusion's CDSS was a substitute for detailing in promotion, partially offsetting its revenue losses.

6 Magnitude of Belief Distortion

6.1 Method

Table 3 suggests the existence of long-term belief distortion. We now quantify the economic magnitude of this effect based on the conceptual framework. Recall that during the treatment window from 2016 to 2018, an affected physician suffers both direct manipulation $\delta_i(Z_j, s)$ and belief distortion γ_i^m. We will separate out the impacts from these two channels using a machine-learning algorithm built upon the multi-arm causal forest (Nie and Wager, 2021), allowing us to estimate the conditional average treatment effect (CATE) as a function of observed characteristics.

Our data are aggregated at the physician-year level without the granular information of each visit j's decision a_{ij} and patient characteristics Z_j. Therefore, we need to estimate the

expected probability of prescription based on Equation (2):

$$Pr(a_{it}=1|\bar{X}_j)=\int \mathbf{1}\left\{q_{i0}(z)\times\left(\prod_{m=1}^{k}\gamma_i^m\right)\times\delta_i(z,s)\geq\frac{c_1}{c_2}\right\}f\left(z|\bar{X}_j\right)dz. \tag{5}$$

In the above equation, $f\left(z|\bar{X}_j\right)$ is the conditional distribution of patient characteristics. We assume this distribution is characterized by physician-level variables $\bar{X}_j$, including the average beneficiary age, risk score, male fraction, African American fraction, Hispanic fraction, dual qualification (Medicaid and Medicare) fraction, physician seniority (years since graduation) and physician gender. Denote $p^T(\bar{X}_j)$ and $p^C(\bar{X}_j)$ as the equilibrium prescription probability for the treatment group and control group respectively. Following Equation (5), we can decompose $p^T(\bar{X}_j)$ into

$$\begin{aligned}
& p^T(\bar{X}_j) \\
= & \int \mathbf{1}\left\{q_{i0}(z)\geq\frac{c_1}{c_2}\right\}f\left(z|\bar{X}_j\right)dz \\
+ & \int \mathbf{1}\left\{\frac{c_1}{c_2}>q_{i0}(z)\geq\frac{c_1}{c_2\delta_i(z,s)}\right\}f\left(z|\bar{X}_j\right)dz \\
+ & \sum_{l=1}^{k}\int \mathbf{1}\left\{\frac{c_1}{c_2\delta_i(z,s)\times\left(\prod_{m=1}^{l-1}\gamma_i^m\right)}>q_{i0}(z)\geq\frac{c_1}{c_2\delta_i(z,s)\times\left(\prod_{m=1}^{l}\gamma_i^m\right)}\right\}f\left(z|\bar{X}_j\right)dz.
\end{aligned}$$

Each line in the above equation has an intuitive interpretation. The second line represents the expected probability for a similar physician in the control group, i.e. $p^C(\bar{X}_j)$. The third line represents the marginal treatment effect due to direct manipulation, denoted by $\tau_0(\bar{X}_j)$. In the last line, each summation term represents the additional long-term impact due to the lth interaction, denoted by $\tau_l(\bar{X}_j)$. Formally, we will estimate the following equation using the sample from 2013 to 2018:

$$p(\bar{X}_j)=p^C(\bar{X}_j)+\mathbf{I}\{j\in\mathbf{T},t\geq 2016\}\left(\sum_{l=0}^{t-2016}\tau_l(\bar{X}_j)\right). \tag{6}$$

$\mathbf{I}\{j\in\mathbf{T},t\geq 2016\}$ indicates that the physician belongs to the treatment group and the current year belongs to the treatment window. We also calculate the number of previous

interactions k in Equation (5) by $t - 2016$. In the data, for each physician i at year t, we quantify the prescription probability $p_{i,t}$ as the total number of LA opioid claims divided by the total number of all claims, multiplied by 100. For clarity, $p_{i,t}$ is expressed in percentage points due to low opioid prescription rates. As Table 1 suggests, in our sample, a typical physician has 13.35 LA opioid annual claims on average, relative to 4,144 total annual claims. Therefore, the unconditional average prescription probability is around 0.225%.[8]

The logic behind the multi-arm causal forest, or broadly the generalized random forest, involves constructing a collection of decision trees. Each tree is built from a random subsample of the data as the initial node, and splits it into child nodes recursively to form leaves. Each node is split based on a random subset of variables using threshold strategies. This threshold strategy aims to maximize the difference in estimated treatment effects post-split. Once the trees are constructed, the final treatment effect for a given data point is estimated by comparing outcomes between treated and control units, weighted by how many times they belong to the same bottom leaf.

We estimate $\tau_l(\bar{X}_j)$ by adjusting the multi-arm causal forest by Nie and Wager (2021) as follows. In the original algorithm, each unit i belongs to one of the mutually exclusive treatment arms, or the control arm. They then estimate the CATE of each treatment arm, without any boundary conditions. In our setting, there is only one treatment arm, but treated units gradually receive heterogeneous CATE $\tau_l(\bar{X}_j)$ depending on the number of previous treatment periods. We also impose additional restrictions on regulating the values of $\hat{\tau}_l(\bar{X}_j)$ to be non-negative. An important caveat is that $\tau_1(\cdot)$ and $\tau_2(\cdot)$ are estimated conditional on $\tau_0(\cdot)$. Therefore, we cannot extrapolate the estimated $\hat{\tau}_1(\cdot)$ and $\hat{\tau}_2(\cdot)$ to the 2019 – 2021 sample without an independence assumption.

For each observation in the data (both treated and control), this algorithm generates the predicted treatment effects $\hat{\tau}_l(\bar{X}_j)$ respectively, along with the baseline counterfactual prescription probability $\hat{p}^C(\bar{X}_j)$ in the control group. We derive the predicted probability

[8]Due to the extreme values of the right-tail outliers, we further truncate the sample at the 95th percentile.

using Equation (6)

$$\hat{p}(\bar{X}_j) = \hat{p}^C(\bar{X}_j) + \mathbf{I}\{j \in \mathbf{T}, t \geq 2016\} \left(\sum_{l=0}^{t-2016} \hat{\tau}_l(\bar{X}_j) \right). \tag{7}$$

6.2 Results

Table 7 Panel A summarizes our estimation results. In the whole sample, the average prescription probability is around 0.242%, whereas our predicted values have a similar mean of 0.241%. It is worth nothing that the standard deviation of the predicted value is much smaller. There exists a big spike of zero prescriptions in the data whereas predictions smooth this distribution around zero.. We summarize the predicted $\hat{\tau}_l$ from 2016 to 2018 (the treatment window) in the remaining rows as a reference. The average $\hat{\tau}_l$ ranges from 1.2 to 1.5 basis points (bps), or roughly 5% to 6% of the unconditional average of $\hat{p}$.

We cannot directly infer the economic impacts of long-term distortion from these average values for the following two reasons. First, these treatment effects start at different calendar years for a given treated physician. Second, $\hat{\tau}_l$ is a function of $\bar{X}_j$ so the actual magnitudes depend on the conditional distribution of $\bar{X}_j$ among treatment units. Instead, we perform a similar regression as Equation (3) in Panel B, except that we drop the control variables to avoid bad control problems since they are a subset of $\bar{X}_j$ used for prediction. The results are quantitatively close if we keep the control variables.

As a benchmark, we first compare the raw prescription probability in data to the estimated $\hat{p}$. Column (1) implies that a typical treated physician has a higher chance of LA opioid prescription with 2.3 bps, equivalent to 9.5% of the unconditional average. "Assuming the average claim count (4,144), our estimates suggest a 0.95 increase in annual LA claims, aligning with Table 2's implied magnitudes. To assess the fitness of our predicted probability, using $\hat{p}$ as the outcome yields a 2.4 bps coefficient, matching the ground truth closely.

We next assume no long-term distortion by setting $\hat{\tau}_1 = \hat{\tau}_2 = 0$ in Equation (7). Under this assumption, Column (3) shows a 54% reduction in the average treatment effect to 1.1 bps. Column (4) studies the possibility that there exists only a one-time belief distortion and this bias does not reinforce with additional interactions. We consider this possibility by

only setting $\hat{\tau}_2 = 0$ in Equation (7). The coefficient becomes 2.0 bps, which is 17% smaller than column (2).

Table B.4 examines the cross-sectional and time-series correlations among $\hat{\tau}_l$. The first three columns imply that the three treatment effects are contemporaneously correlated. When an affected physician has a stronger response to direct manipulation, he also suffers larger belief distortion at the same time. For example, a one basis point increase in $\hat{\tau}_0$ is associated with 0.17 – 0.20 bps higher $\hat{\tau}_1$ and $\hat{\tau}_2$. Physicians also exhibit persistent patterns in their behavior changes, documented in the remaining columns. For example, column (4) implies that a one basis point increase in $\hat{\tau}_0$ in the previous quarter correlates with a 0.79 bps increase in the current quarter as well. This simple regression has a R^2 of 0.69. A higher reaction to the direct manipulation will also imply larger long-term belief distortions in the next period. This large correlation can be due to persistent patient characteristics or inherent physician preferences.

6.3 Heterogeneity

Which types of affected physicians are less sensitive to the implicit impacts from the Pain CDS? In this section, we explore several heterogeneity tests based on ex-ante characteristics to better understand the interaction between human and artificial intelligence. We first guide our analysis with anecdotal facts about Practice Fusion's strategy. During its communication with Purdue Pharma, it believed the CDSS would target "opioid naive" users and utilize their limited knowledge on potential addiction risks. Besides, if physicians had previously prescribed opioids, they might be more attentive to the coverage of ERO's risks and litigation of Purdue Pharma, thereby reluctant to accept opioids as a treatment option. So we first hypothesize that the experience of previous opioid usage for affected physicians will mitigate the treatment effects.

We start with supporting evidence from our quantitative estimates. In Panel A of Figure 3, we separate the sample into two groups based on whether the physician had LA prescription in 2015. We then plot aggregated treatment effects for each year within the treatment window. For example, the difference between the treatment and control group is purely driven by $\hat{\tau}_0$ in 2016. So the first comparison implies that if a treated physician has

no previous LA opioid usage, then his average response to the direct manipulation is much larger (0.023% v.s. 0.017%). Adding long-term belief distortions in subsequent years yields consistent results.

To confirm this hypothesis in our empirical study, we interact the treatment variable with the pre-shock (2015) level of LA opioid claims for each physician in Table 8. The coefficients of all interaction terms are negative with substantial statistical significance. To understand the economic magnitudes, Panel A column (2) implies that a typical treated physician with a pre-existing median-level LA prescription amount (0) will have a treatment effect of 18.1% in the short-term. This impact will reduce to zero if that physician had around three LA claims before the shock. Although our quantitative estimation cannot directly back out the effect magnitudes from 2019 to 2021, Panel B documents consistent heterogeneity in this alternative sample.

Next, we guide the heterogeneity test based on the feature importance of variables. Recall that leaves are split by cut-off thresholds based on a subset of $\bar{X}_j$. For each variable x_j, feature importance measures the fraction of leaves split using thresholds related to that particular variable. Logically, this measure tells how important each variable is in determining the treatment effect heterogeneity. Table B.5 shows that $AvgAge$ is the dominant splitting variable, accounting for 48% of splits.. The following two are dramatically less important: $DualPct$ (17.5%) and $AvgRisk$ (15.5%). We also believe there exists an intuitive logic for patient age to generate substantial heterogeneity. Physicians may exercise additional caution and invest more effort in considering the side effects of extended-release opioids when facing older patients. Figure 3 Panel B, supports this hypothesis with the quantitative estimates. As the previous figure, we plot the aggregated treatment effects corresponding to each year with subsamples based on the median level of $AvgAge$ in 2015. It is evident that the low-age group has a substantially large treatment effect. Note that feature importance does not necessarily suggest a monotonic heterogeneity pattern. Indeed, we find that higher treatment effects tend to concentrate in middle-level $DualPct$ and $AvgRisk$ groups.

The regression results in both panels of Table 9 support this hypothesis as all interaction terms are significantly negative. In terms of magnitudes, the high-age group has a net treatment effect of 0.9 bps in column (1) of Panel A, which is roughly 30% of the low-age

group's effect (3.2 bps). Similarly, the high-age treated group will only increase the number of LA claims by 2.3% in the short-term, which is 39% of the baseline effect and 23% of the low-age group's effect.

7 Conclusion

This paper sheds light on the lasting impacts of AI interactions on human capital in the context of opioid prescriptions. A biased pain alert promoting the sales of extended-release opioids resulted in a significant increase in opioid claims by affected physicians during the treatment window (2016–2018). This effect persisted even after the removal of the biased alert, indicating a long-term distortion in human capital development. Notably, this distortion was not observed in physicians receiving explicit payments from Purdue Pharma, highlighting the unique role of AI in shaping human behavior beyond traditional financial incentives.

The study contributes to understanding the risks of AI biases on professionals' long-term beliefs and habits. The unconscious nature of AI biases makes them undetectable even by professionals with expertise. Instead, physicians may perceive the recommendations as optimally generated from real data and update more favorable beliefs regarding the benefit-risk profile of opioids over time. Consistent with the recent EU AI Act, our study serves as a caveat for the potential risks of AI manipulation and the need for greater awareness and accountability in AI-driven decision-making processes. In light of AI's expanding influence across diverse sectors, managing its effects on human capital is essential for ensuring responsible and ethical technological integration.

Lastly, this study has policy implications for the usage of artificial intelligence in the healthcare sector. Closely related is the digitization of Prescription Drug Monitoring Program (PDMP). PDMPs are state-administered registries of prescription data for controlled substances (including opioids), assisting clinicians in inform prescribing decisions. While PDMPs are considered as a powerful tool to stop the opioid epidemic, accessing the PDMP requires the provider to use a separate record system and identify the patient. Many states, such as Wisconsin, proposed to integrate their PDMPs in providers' EHRs and even provide

automatic alerts to facilitate clinical workflows. While we agree with the importance of integration, our works suggests that the alert system has to be cautiously designed to avoid patients from mistakenly mimicking the recommendations.

References

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh, 2016, The determinants of productivity in medical testing: Intensity and allocation of care, *American Economic Review* 106, 3730–3764.

Acemoglu, Daron, 2021, Harms of ai, Technical report, National Bureau of Economic Research.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo, 2022, Artificial intelligence and jobs: evidence from online vacancies, *Journal of Labor Economics* 40, S293–S340.

Acemoglu, Daron, and Pascual Restrepo, 2018, The race between man and machine: Implications of technology for growth, factor shares, and employment, *American economic review* 108, 1488–1542.

Acemoglu, Daron, and Pascual Restrepo, 2020, Robots and jobs: Evidence from us labor markets, *Journal of political economy* 128, 2188–2244.

Acemoglu, Daron, and Pascual Restrepo, 2022, Tasks, automation, and the rise in us wage inequality, *Econometrica* 90, 1973–2016.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, 2023, Combining human expertise with artificial intelligence: Experimental evidence from radiology, Technical report, National Bureau of Economic Research.

Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb, 2019, Artificial intelligence: the ambiguous labor market impact of automating prediction, *Journal of Economic Perspectives* 33, 31–50.

Alpert, Abby, William N Evans, Ethan MJ Lieber, and David Powell, 2022, Origins of the opioid crisis and its enduring impacts, *The Quarterly Journal of Economics* 137, 1139–1179.

Angelova, Victoria, Will S Dobbie, and Crystal Yang, 2023, Algorithmic recommendations and human discretion, Technical report, National Bureau of Economic Research.

Arbabshirani, Mohammad R, Brandon K Fornwalt, Gino J Mongelluzzo, Jonathan D Suever, Brandon D Geise, Aalpen A Patel, and Gregory J Moore, 2018, Advanced machine learning

in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration, *NPJ digital medicine* 1, 9.

Autor, David, and Anna Salomons, 2018, Is automation labor-displacing? productivity growth, employment, and the labor share, Technical report, National Bureau of Economic Research.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson, 2024, Artificial intelligence, firm growth, and product innovation, *Journal of Financial Economics* 151, 103745.

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson, 2022, Firm investments in artificial intelligence technologies and changes in workforce composition, *Available at SSRN 4060233* .

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, 2022, Consumer-lending discrimination in the fintech era, *Journal of Financial Economics* 143, 30–56.

Bénabou, Roland, and Jean Tirole, 2006, Incentives and prosocial behavior, *American economic review* 96, 1652–1678.

Bresnahan, Timothy F, and Manuel Trajtenberg, 1995, General purpose technologies 'engines of growth'?, *Journal of econometrics* 65, 83–108.

Bronsoler, Ari, Joseph Doyle, and John Van Reenen, 2022, The impact of health information and communication technology on clinical quality, productivity, and workers, *Annual Review of Economics* 14, 23–46.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock, 2018, What can machines learn and what does it mean for occupations and the economy?, in *AEA papers and proceedings*, volume 108, 43–47, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello, 2020, Artificial intelligence, algorithmic pricing, and collusion, *American Economic Review* 110, 3267–3297.

Cao, Sean, Wei Jiang, Junbo L Wang, and Baozhong Yang, 2021, From man vs. machine to man+ machine: The art and ai of stock analyses, Technical report, National Bureau of Economic Research.

Cerrato, Paul, and John Halamka, 2020, *Reinventing clinical decision support: data analyt-*

ics, artificial intelligence, and diagnostic reasoning, chapter 3 (Taylor & Francis).

Chan, David C, Matthew Gentzkow, and Chuan Yu, 2022, Selection with variation in diagnostic skill: Evidence from radiologists, *The Quarterly Journal of Economics* 137, 729–783.

Chandra, Amitabh, and Douglas O Staiger, 2020, Identifying sources of inefficiency in healthcare, *The quarterly journal of economics* 135, 785–843.

Coleman, Braiden, Kenneth Merkley, and Joseph Pacelli, 2022, Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations, *The Accounting Review* 97, 221–244.

Cowgill, Bo, and Catherine E Tucker, 2019, Economics, fairness and algorithmic bias, *preparation for: Journal of Economic Perspectives* .

Currie, Janet, and W Bentley MacLeod, 2017, Diagnosing expertise: Human capital, decision making, and performance among physicians, *Journal of labor economics* 35, 1–43.

Doyle, Joseph, Sarah Abraham, Laura Feeney, Sarah Reimer, and Amy Finkelstein, 2019, Clinical decision support for high-cost imaging: a randomized clinical trial, *PLoS One* 14, e0213373.

Dranove, David, and Craig Garthwaite, 2022, Artificial intelligence, the evolution of the healthcare value chain, and the future of the physician, Technical report, National Bureau of Economic Research.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G Rossi, 2019, The promises and pitfalls of robo-advising, *The Review of Financial Studies* 32, 1983–2020.

Felten, Edward W, Manav Raj, and Robert Seamans, 2018, A method to link advances in artificial intelligence to occupational abilities, in *AEA Papers and Proceedings*, volume 108, 54–57, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Finlayson, Samuel G, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane, 2019, Adversarial attacks on medical machine learning, *Science* 363, 1287–1289.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, 2022, Predictably unequal? the effects of machine learning on credit markets, *The Journal of Finance* 77, 5–47.

Ganju, Kartik K, Hilal Atasoy, Jeffery McCullough, and Brad Greenwood, 2020, The role of decision support systems in attenuating racial biases in healthcare delivery, *Management science* 66, 5171–5181.

Grennan, Jillian, and Roni Michaely, 2020, Artificial intelligence and high-skilled work: Evidence from analysts, *Swiss Finance Institute Research Paper* .

Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong, 2024, Lender automation and racial disparities in credit access, *The Journal of Finance* 79, 1457–1512.

Hydari, Muhammad Zia, Rahul Telang, and William M Marella, 2019, Saving patient ryan—can advanced electronic medical records make patient care safer?, *Management Science* 65, 2041–2059.

Jing, Xia, Lina Himawan, and Timothy Law, 2019, Availability and usage of clinical decision support systems (cdsss) in office-based primary care settings in the usa, *BMJ Health & Care Informatics* 26.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, 2018, Human decisions and machine predictions, *The quarterly journal of economics* 133, 237–293.

Mullainathan, Sendhil, and Ziad Obermeyer, 2022, Diagnosing physician error: A machine learning approach to low-value health care, *The Quarterly Journal of Economics* 137, 679–727.

Nie, Xinkun, and Stefan Wager, 2021, Quasi-oracle estimation of heterogeneous treatment effects, *Biometrika* 108, 299–319.

Ross, Casey, and Ike Swetlitz, 2017, Ibm pitched its watson supercomputer as a revolution in cancer care. it's nowhere close, *Stat* .

Rossi, Alberto G, and Stephen P Utkus, 2020, Who benefits from robo-advising? evidence from machine learning, *Evidence from Machine Learning (March 10, 2020)* .

Sloane, Elliot B, and Ricardo J Silva, 2020, Artificial intelligence in medical devices and clinical decision support systems, in *Clinical engineering handbook*, 556–568 (Elsevier).

Van Zee, Art, 2009, The promotion and marketing of oxycontin: commercial triumph, public health tragedy, *American Journal of Public Health* 99, 221–227.

Volkow, Nora D, and A Thomas McLellan, 2016, Opioid abuse in chronic pain—misconceptions and mitigation strategies, *New England Journal of Medicine* 374, 1253–1263.

Webb, Michael, 2019, The impact of artificial intelligence on the labor market, *Available at SSRN 3482150* .

Figure 1: Treatment Dynamics of Practice Fusion's Manipulation

This figure plots the treatment dynamics with respect to LA opioid prescription. We plot the β_es estimated from Equation (4). x-axis indicates the calendar years. The base year is 2015 (indicated by the red vertical line), that is, the year prior to the shock. The green vertical indicates the suspension of Practice Fusion manipulation (2019). 95% confidence intervals are indicated by the solid lines.

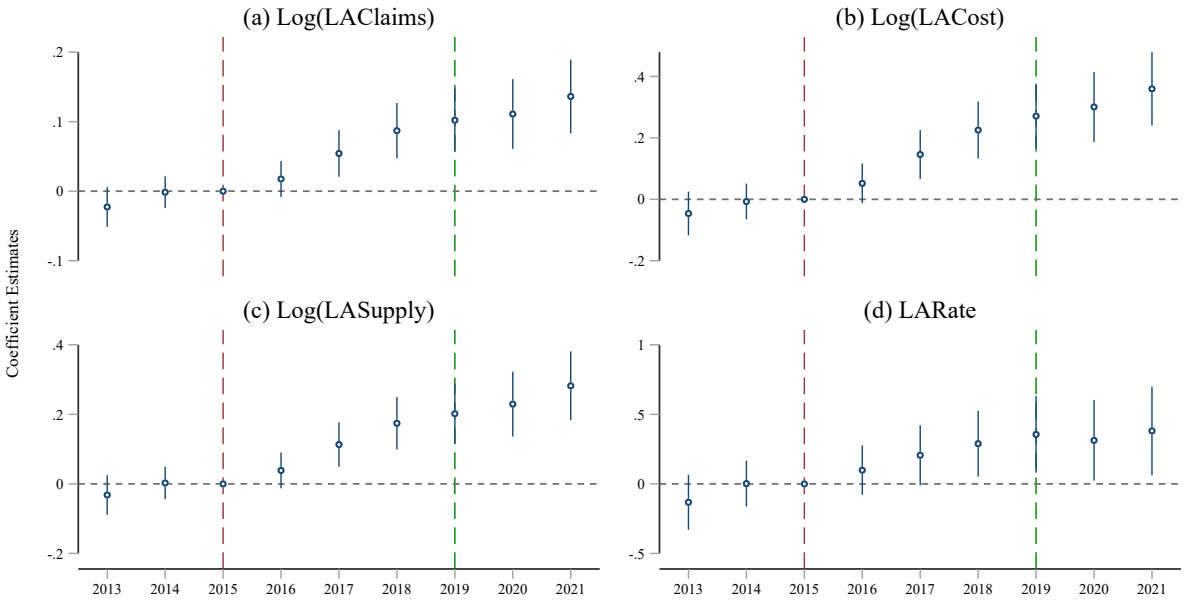


Figure 2: Treatment Dynamics of Purdue Pharma Detailing

This figure plots the treatment dynamics with respect to LA opioid prescription due to Purdue Pharma detailing. As before, we estimate β_cs in Equation (4), except that we replace PF_i with $Detail_i$. The base year is 2015 (indicated by the red vertical line), that is, the year prior to the shock. The green vertical indicates the suspension of in-kind payments (2019). 95% confidence intervals are indicated by the solid lines.

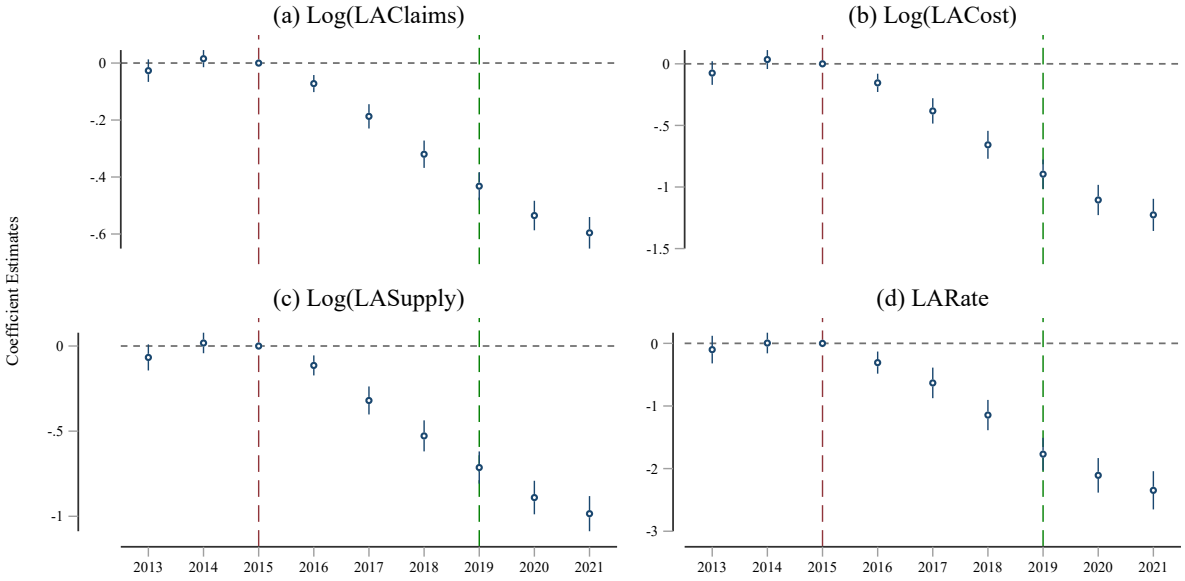
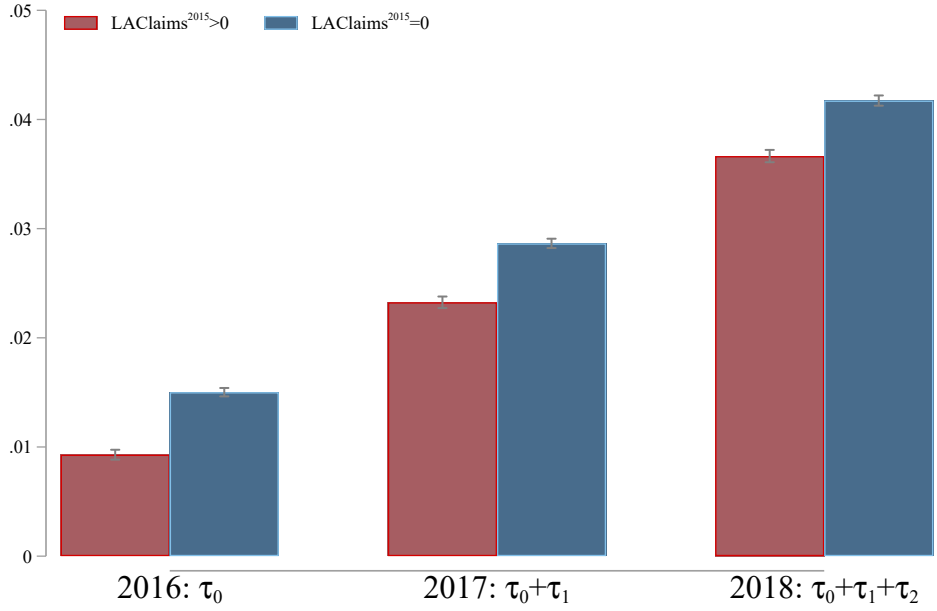


Figure 3: Treatment Effect Heterogeneity

This figure plots the treatment effect heterogeneity based on ex-ante opioid usage and patient age. Panel A splits the sample based on whether the physician has used LA opioid before 2015. Panel B splits the sample based on the median level of patient age. In each figure, we plot the average cumulative treatment effects in the corresponding year, represented by the bar height. 95% confidence interval is also included.

Panel A: LA Opioid Usage



Panel B: Patient Age

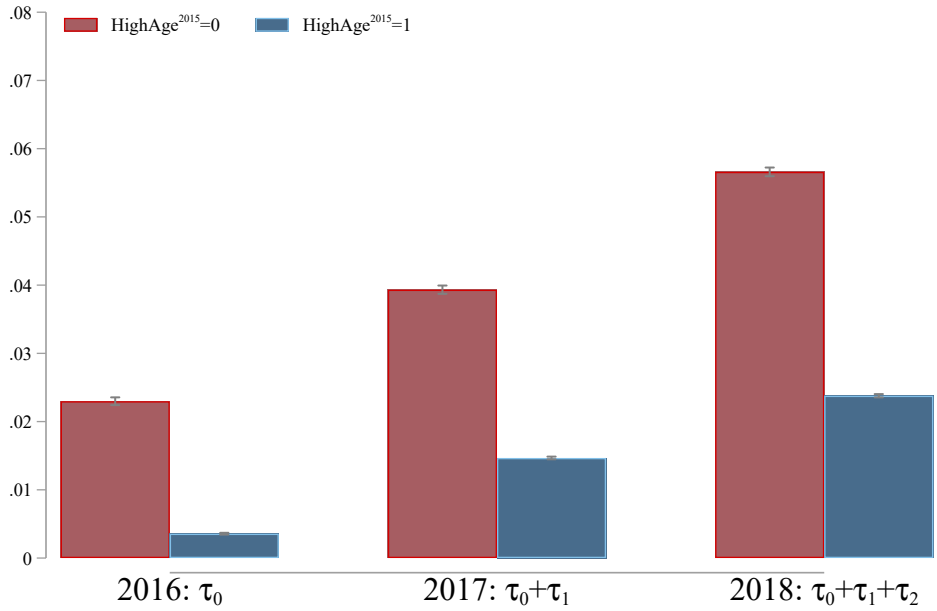


Table 1: Summary Statistics

This table provides the summary statistics of the main variables in the paper. Panel A summarizes the annual observations of physicians from 2013 to 2021. PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $LAClaims_{i,t}$ is the number of long-acting opioid claims by physician i in year t. $LACost_{i,t}$ is the dollar amount of long-acting opioid prescription costs by physician i in year t. $LASupply_{i,t}$ is the days of supply for long-acting opioids by physician i in year t. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. $TotClaims_{i,t}$ is the number of all claims by physician i in year t. $AvgAge_{i,t}$ is the average age of beneficiaries visiting physician i in year t. $MalePct_{i,t}$ is the fraction of male beneficiaries visiting physician i in year t. $BlackPct_{i,t}$ is the fraction of African American beneficiaries visiting physician i in year t. $HispanicPct_{i,t}$ is the fraction of Hispanic beneficiaries visiting physician i in year t. $AvgRisk_{i,t}$ is the average risk score of beneficiaries visiting physician i in year t. $DualPct_{i,t}$ is the fraction of beneficiaries qualified to receive both Medicare and Medicaid benefits. $Detailing_{i,t}$ is one physician i receives in-kind payments from Purdue Pharma in year t, and zero otherwise. Panel B compares the treatment and control groups in 2015, the year before the treatment. t-statistics are calculated based on standard errors clustered at the 3-digit zip code level.

Panel A: Full Sample

Variable	N	Mean	Std	p25	Median	p75
$PF \times Post$	443,286	0.060	0.237	0.000	0.000	0.000
$LAClaims$	443,286	13.546	33.311	0.000	0.000	12.000
$LACost$	443,286	2,623.096	7,691.018	0.000	0.000	788.490
$LASupply$	443,286	382.032	928.782	0.000	0.000	360.000
$LARate$	443,286	4.255	8.898	0.000	0.000	4.628
$TotClaims$	443,286	4,144.498	4,876.197	754.000	2,616.000	5,884.000
$AvgAge$	443,286	71.925	3.930	70.090	72.415	74.381
$MalePct$	443,286	0.398	0.125	0.321	0.409	0.476
$BlackPct$	443,286	0.107	0.178	0.000	0.029	0.140
$HispanicPct$	443,286	0.072	0.154	0.000	0.000	0.077
$AvgRisk$	443,286	1.399	0.527	1.076	1.261	1.553
$DualPct$	443,286	0.223	0.184	0.085	0.181	0.321
$Detailing$	443,286	0.039	0.194	0.000	0.000	0.000

Panel B: Ex-ante Difference in 2015

	Control Obs.	*Treat Obs.*	*Control Mean*	*Treat Mean*	*Diff.*	*t-stat*	*p-value*
$LAClaims$	47,258	4,871	16.88	15.86	1.02	1.14	0.25
$LACost$	47,258	4,871	3,210.51	3,025.45	185.06	1.01	0.31
$LASupply$	47,258	4,871	479.65	428.36	51.29	2.08	0.04
$LARate$	47,258	4,871	5.15	4.35	0.81	3.65	0.00
$Detailing$	47,258	4,871	0.07	0.07	0.00	0.70	0.48

Table 2: The Short-term Impacts on Opioid Prescription

This table provides the results on the short-term impacts of Practice Fusion's CDSS on opioid prescription. The sample consists of annual observations at the physician level from 2013 to 2018, i.e. the pre-treatment and treatment phases. PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. $Log(SAClaims)_{i,t}$ is the logarithm of one plus number of short-acting opioid claims. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*	(5) *Log(SAClaims)*
$PF \times Post$	0.059***	0.154***	0.114***	0.236**	−0.025
	(3.943)	(4.474)	(4.117)	(2.568)	(−1.447)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	305,082	305,082	305,082	305,082	305,082
R^2	0.79	0.76	0.76	0.70	0.87

Table 3: The Long-term Impacts on Opioid Prescription

This table provides the results on the long-term impacts of Practice Fusion's CDSS on opioid prescription. The sample consists of annual observations at the physician level from 2013 to 2015, and 2019 to 2021, i.e. the pre-treatment and post-treatment phases. PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*
$PF \times Post$	0.116***	0.308***	0.231***	0.343**
	(4.826)	(5.694)	(5.199)	(2.476)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	292,094	292,094	292,094	292,094
R^2	0.72	0.69	0.69	0.62

Table 4: Long-term Opioid Therapy Utilization

This table provides the cross-sectional evidence on the long-term costs for Medicare beneficiaries due to Practice Fusion adoption. The sample consists of annual observations at the 3-digit zip code level in 2020 and 2021. $Log(NumPF)_k$ is number of physicians having adopted Practice Fusion in area k. $OTPBenes_{k,t}$, $OTPSvcs_{k,t}$, $OTPCharge_{k,t}$, and $OTPPay_{k,t}$ are the number of beneficiaries, number of services, bill charges and actual Medicare payments for Opioid Treatment Programs in area k and year t, respectively. Control variables include local demographic characteristics such as logged population, logged income per capital, male percentage, African American percentage, Hispanic percentage, average age, unemployment rate and health insurance coverage percentage. Year fixed effects are included. Standard errors are clustered at the area level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1) *Log(OTPBenes)*	(2) *Log(OTPSvcs)*	(3) *Log(OTPCharge)*	(4) *Log(OTPPay)*
$Log(NumPF)$	0.475***	0.763***	1.269***	1.245***
	(4.966)	(4.984)	(4.922)	(4.895)
Controls	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
N	1,777	1,777	1,777	1,777
R^2	0.28	0.28	0.27	0.27

Table 5: Long-term Commercial Insurance Drug Overdose Visits

This table provides the cross-sectional evidence on the long-term costs for commercial insurance enrollees due to Practice Fusion adoption. The sample consists of annual observations at the MSA level in from 2019 to 2022. $Log(NumPF)_k$ is number of physicians having adopted Practice Fusion in area k. $ODVisit_{k,t}$ and $ODPayk, t$ are the number of visits due to drug overdose and the associated payments in area k and year t, respectively. Control variables include local demographic characteristics such as logged population, logged income per capital, male percentage, African American percentage, Hispanic percentage, average age, unemployment rate and health insurance coverage percentage. Year fixed effects are included. Standard errors are clustered at the area level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)
	Outpatient		**Inpatient**	
	Log(ODVisit)	*Log(ODPay)*	*Log(ODVisit)*	*Log(ODPay)*
$Log(NumPF)$	0.123***	0.131***	0.138***	0.133***
	(2.739)	(2.864)	(3.457)	(2.848)
Controls	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
N	1,300	1,300	1,280	1,280
R^2	0.73	0.71	0.78	0.72

Table 6: Comparison of Long-term Effects: Explicit Detailing v.s. Implicit Manipulation

This table provides the results on the long-term impacts of Purdue Pharma's detailing on opioid prescription. The sample consists of annual observations at the physician level from 2013 to 2015, and 2019 to 2021, i.e. the pre-treatment and post-treatment phases. $Detail_i$ is one if physician i had received in-kind payments from Purdue Pharma by 2019, and zero otherwise. PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Effects of Detailing

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*
$Detail \times Post$	−0.425***	−0.805***	−0.661***	−2.191***
	(−5.606)	(−4.866)	(−5.026)	(−4.969)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	283,358	283,358	283,358	283,358
R^2	0.71	0.68	0.68	0.60

Panel B: Comparison of Effects

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*
$PF \times Post$	0.180***	0.435***	0.337***	0.698***
	(7.153)	(7.614)	(7.165)	(5.199)
$Detail \times Post$	−0.583***	−1.188***	−0.958***	−2.805***
	(−7.470)	(−6.943)	(−7.038)	(−6.444)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	283,358	283,358	283,358	283,358
R^2	0.71	0.68	0.68	0.60

Table 7: Economic Magnitudes of Long-term Belief Distortion

This table exhibits the economic magnitudes of long-term belief distortion. Panel A summarizes the raw prescription probability $p_{i,t}$ in the predicting sample, the predicted probability $\hat{p}_{i,t}$, and the estimated treatment effects τ_l. We summarize the treatment effects in the treatment windows from 2016 to 2018. Panel B estimates the average treatment effects using the probability in the data in column (1), predicted probability in column (2), the counterfactual probabilities without any long-term belief distortion (without both τ_1 and τ_2) in column (3), and the counterfactual probabilities with only one-time belief distortion (without τ_2) in column (4). No control variables are included. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Summary Statistics

Variable	N	Mean	Std.
p	314,643	0.242	0.530
$\hat{p}$	314,643	0.241	0.156
$\hat{\tau}_0$	155,766	0.012	0.034
$\hat{\tau}_1$	155,766	0.015	0.017
$\hat{\tau}_2$	155,766	0.015	0.018

Panel B: Counterfactual Analysis

	(1)	(2)	(3)	(4)
			$\hat{p}$	
	p	*Original*	$\hat{\tau}_1 = \hat{\tau}_2 = 0$	$\hat{\tau}_2 = 0$
$PF \times Post$	0.023***	0.024***	0.011***	0.020***
	(4.821)	(14.173)	(7.029)	(11.857)
Controls	N	N	N	N
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	314,145	314,145	314,145	314,145
R^2	0.76	0.69	0.69	0.69

Table 8: Heterogeneous Effects of Impacts: Long-acting Opioid Experience

This table provides the results on the heterogeneous impacts due to long-acting opioid prescription experience before the shock. Panel A estimates the effects in the short-term sample (2013 – 2018) and Panel B in the long-term sample (2013 – 2015 & 2019 – 2021). PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_i^{2015}$ is the logarithm of one plus long-acting opioid claims for physicians i in year 2015, i.e. the pre-shock level of LA opioid usage. $p_{i,t}$ is the prescription probability. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Short-term Effects

	(1) p	(2) *Log(LAClaims)*	(3) *Log(LACost)*	(4) *Log(LASupply)*	(5) *LARate*
$PF \times Post$	0.057***	0.181***	0.403***	0.319***	0.878***
	(12.997)	(14.967)	(13.878)	(13.743)	(12.135)
$PF \times Post$	−0.039***	−0.130***	−0.265***	−0.218***	−0.682***
$\times Log(LAClaims)^{2015}$	(−10.691)	(−12.481)	(−11.631)	(−12.001)	(−10.077)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	304,420	304,420	304,420	304,420	304,420
R^2	0.77	0.80	0.76	0.76	0.70

Panel B: Long-term Effects

	(1) p	(2) *Log(LAClaims)*	(3) *Log(LACost)*	(4) *Log(LASupply)*	(5) *LARate*
$PF \times Post$	0.127***	0.445***	0.978***	0.788***	2.043***
	(19.868)	(21.536)	(21.125)	(21.006)	(19.406)
$PF \times Post$	−0.093***	−0.342***	−0.697***	−0.580***	−1.770***
$\times Log(LAClaims)^{2015}$	(−18.384)	(−21.458)	(−19.238)	(−19.939)	(−19.600)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	304,420	304,420	304,420	304,420	304,420
R^2	0.77	0.80	0.76	0.76	0.70

Table 9: Heterogeneous Effects of Impacts: Patient Age

This table provides the results on the heterogeneous impacts due to patient age. Panel A estimates the effects in the short-term sample (2013 – 2018) and Panel B in the long-term sample (2013 – 2015 & 2019 – 2021). PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $HighAge_i^{2015}$ is one if average beneficiary age for physicians i in year 2015 is above the sample median, and zero otherwise. $p_{i,t}$ is the prescription probability. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Short-term Effects

	(1) p	(2) *Log(LAClaims)*	(3) *Log(LACost)*	(4) *Log(LASupply)*	(5) *LARate*
$PF \times Post$	0.032***	0.099***	0.241***	0.190***	0.394***
	(4.383)	(4.785)	(5.033)	(4.914)	(3.207)
$PF \times Post$	−0.023***	−0.076***	−0.167***	−0.143***	−0.299*
$\times HighAge^{2015}$	(−2.696)	(−3.019)	(−2.755)	(−2.967)	(−1.877)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	304,420	304,420	304,420	304,420	304,420
R^2	0.77	0.79	0.76	0.76	0.70

Panel B: Long-term Effects

	(1) p	(2) *Log(LAClaims)*	(3) *Log(LACost)*	(4) *Log(LASupply)*	(5) *LARate*
$PF \times Post$	0.053***	0.198***	0.481***	0.378***	0.598***
	(5.433)	(6.346)	(6.851)	(6.743)	(3.312)
$PF \times Post$	−0.030**	−0.156***	−0.331***	−0.280***	−0.488**
$\times HighAge^{2015}$	(−2.384)	(−4.124)	(−3.849)	(−4.091)	(−2.029)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	304,420	304,420	304,420	304,420	304,420
R^2	0.70	0.72	0.69	0.69	0.62

Appendix

A Appendix Figures

Figure A.1: Number of Recipients of Purdue Pharma Detailing

This figure plots the yearly number of recipients of Purdue Pharma detailing in our sample. We stopped at 2019 since in-kind payments to physicians were effectively suspended after Purdue Pharma filed for bankruptcy in 2019.

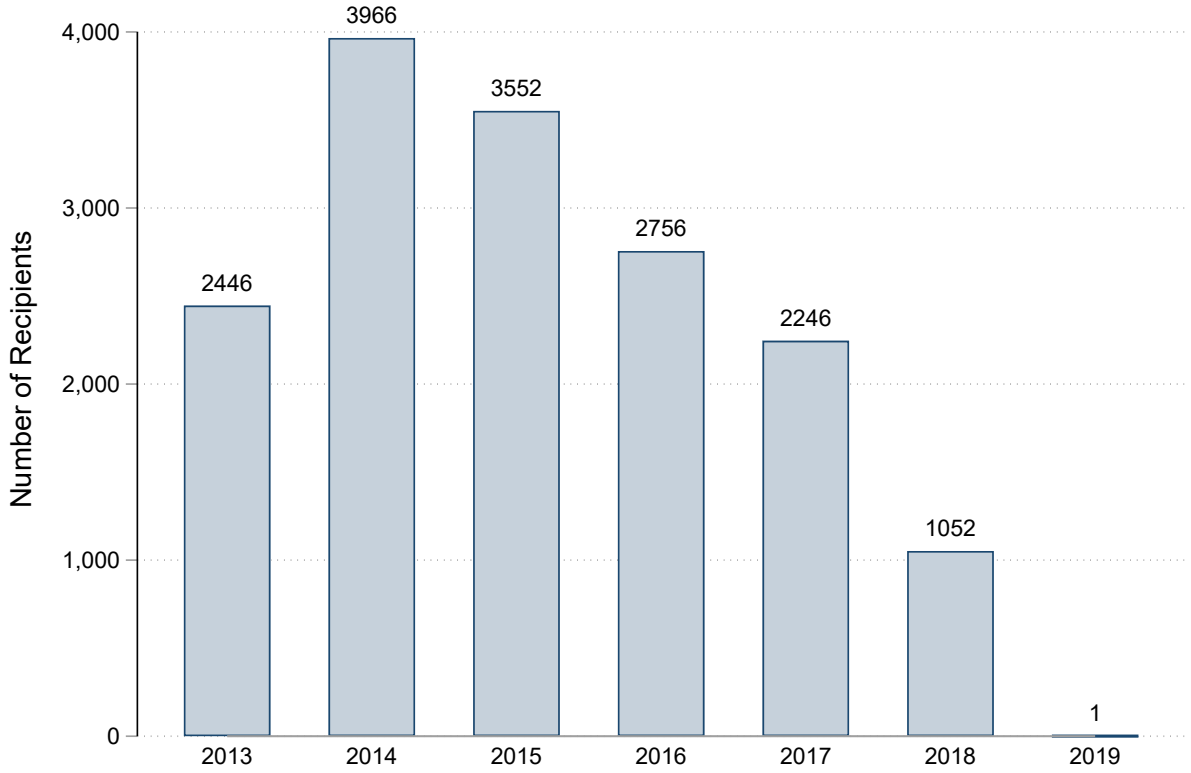
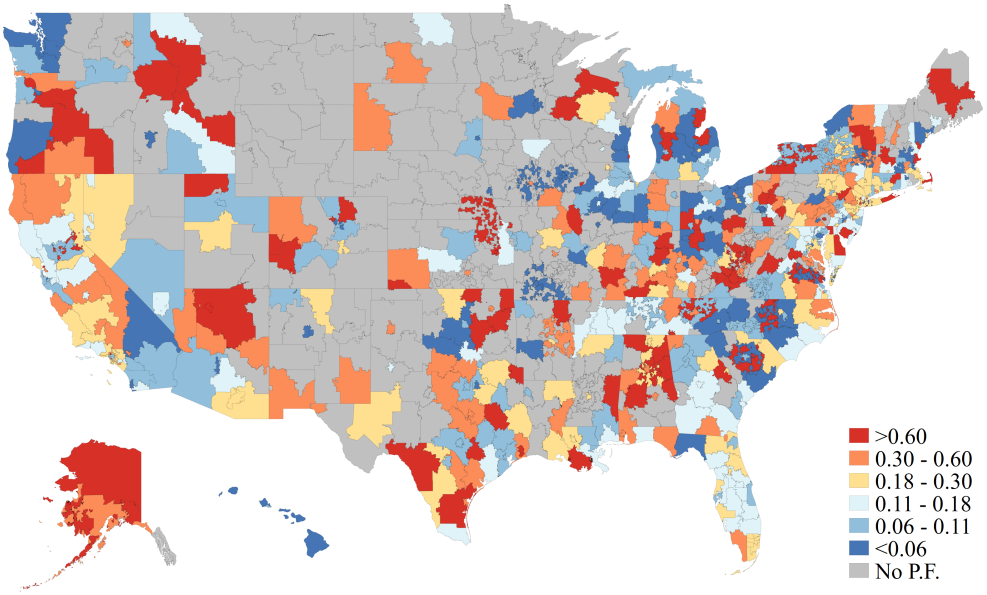


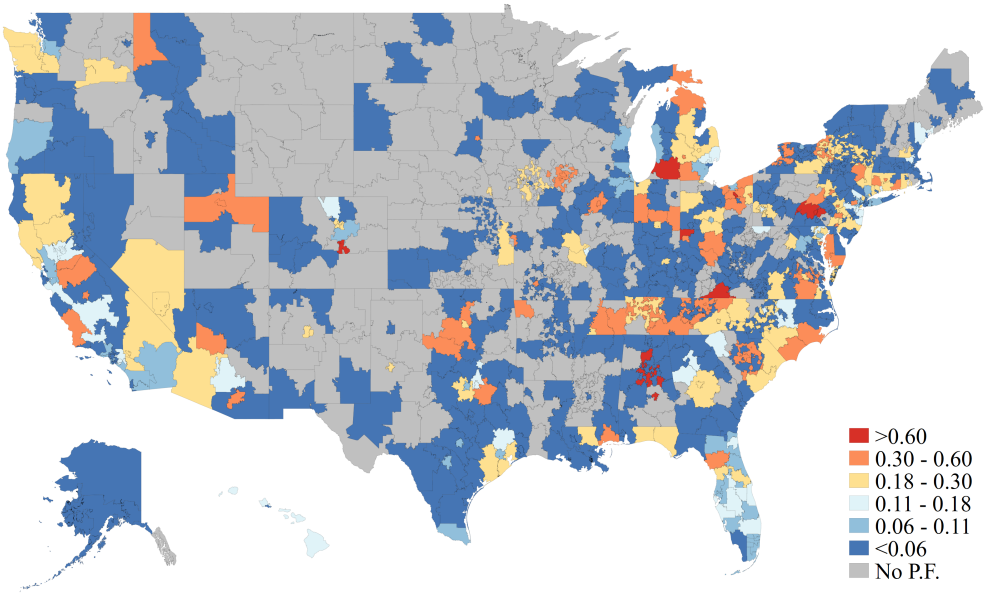
Figure A.2: Number of Recipients of Purdue Pharma Detailing

This figure plots the geogrphaci distributioon of Practice Fusion adopters and detailing recipients in our sample. Each area is a 3-digit zip code area. In each area, we calculate the fraction of Practice Fusion adopters and detailing recipients over the total number of physicians in that area from our sample. Areas without Practice Fusion adopters are marked in gray and dropped from the sample in our analysis.

Panel A: Distribution of Practice Fusion Adopters



Panel B: Distribution of Detailing Recipients



B Appendix Tables

Table B.1: Robustness Check: 5-digit Zip Code Areas

This table shows the robustness of Tables 2 and 3 by requiring the control group to be in the same 5-digit zip code areas and with the same specialty of the treatment group. Panel A estimates the effects in the short-term sample (2013 – 2018) and Panel B in the long-term sample (2013 – 2015 & 2019 – 2021). PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. $Log(SAClaims)_{i,t}$ is the logarithm of one plus number of short-acting opioid claims. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Short-term Effects

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*	(5) *Log(SAClaims)*
$PF \times Post$	0.070***	0.172***	0.133***	0.403***	−0.004
	(3.895)	(4.209)	(4.010)	(3.741)	(−0.191)
Controls	Y	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	86,934	86,934	86,934	86,934	86,934
R^2	0.79	0.75	0.76	0.71	0.87

Panel B: Long-term Effects

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*
$PF \times Post$	0.145***	0.376***	0.291***	0.500***
	(5.305)	(6.121)	(5.783)	(3.249)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	83,059	83,059	83,059	83,059
R^2	0.72	0.69	0.69	0.63

Table B.2: Robustness Check: Area Fixed Effects and Year Fixed Effects

This table shows the robustness of Tables 2 and 3 by specifying non-interacting area and year fixed effects. Panel A estimates the effects in the short-term sample (2013 – 2018) and Panel B in the long-term sample (2013 – 2015 & 2019 – 2021). PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. $Log(LAClaims)_{i,t}$ is the logarithm of one plus $LAClaims_{i,t}$. $Log(LACost)_{i,t}$ is the logarithm of one plus $LACost_{i,t}$. $Log(LASupply)_{i,t}$ is the logarithm of one plus $LASupply_{i,t}$. $LARate_{i,t}$ is the percentage of long-acting opioid claims out of all opioid claims by physician i in year t. $Log(SAClaims)_{i,t}$ is the logarithm of one plus number of short-acting opioid claims. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects, area fixed effects and year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Short-term Effects

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*	(5) *Log(SAClaims)*
$PF \times Post$	0.076***	0.192***	0.143***	0.329***	-0.023
	(4.925)	(5.448)	(5.044)	(3.507)	(-1.398)
Controls	Y	Y	Y	Y	Y
Area FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y	Y
N	305,440	305,440	305,440	305,440	305,440
R^2	0.79	0.76	0.76	0.70	0.86

Panel B: Long-term Effects

	(1) *Log(LAClaims)*	(2) *Log(LACost)*	(3) *Log(LASupply)*	(4) *LARate*
$PF \times Post$	0.151***	0.386***	0.292***	0.529***
	(5.684)	(6.485)	(6.036)	(3.637)
Controls	Y	Y	Y	Y
Area FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	292,450	292,450	292,450	292,450
R^2	0.72	0.68	0.69	0.61

Table B.3: Placebo Test: Antibiotics and Antipsychotic

This table shows the placebo test of Tables 2 and 3 using Antibiotics and Antipsychotic. Panel A estimates the effects in the short-term sample (2013 – 2018) and Panel B in the long-term sample (2013 – 2015 & 2019 – 2021). PF_i is one if physician i adopts Practice Fusion's EHR, and zero otherwise. $Post_t$ is one if year t is greater or equal to 2016, and zero otherwise. Columns (1) and (2) are for Antibiotics and columns (3) and (4) are Antipsychotic. $Claims_{i,t}$ is the number of claims by by physician i in year t for the corresponding drug. $DrugCost_{i,t}$ is the dollar amount of drug costs by physician i in year t for the corresponding drug. Control variables include $AvgAge$, $MalePct$, $BlackPct$, $HispanicPct$, $DualPct$, and $AvgRisk$. Physician fixed effects, area-specialty fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Short-term Effects

	(1)	(2)	(3)	(4)
	Antibiotics		**Antipsychotic**	
	Log(Claims)	*Log(DrugCost)*	*Log(Claims)*	*Log(DrugCost)*
$PF \times Post$	−0.025*	−0.012	−0.018	−0.023
	(−1.780)	(−0.515)	(−1.466)	(−0.785)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	305,082	305,082	223,193	223,193
R^2	0.86	0.81	0.91	0.89

Panel B: Long-term Effects

	(1)	(2)	(3)	(4)
	Antibiotics		**Antipsychotic**	
	Log(Claims)	*Log(DrugCosts)*	*Log(Claims)*	*Log(DrugCosts)*
$PF \times Post$	−0.032	−0.018	−0.034*	−0.016
	(−1.478)	(−0.484)	(−1.911)	(−0.395)
Controls	Y	Y	Y	Y
Area×Year FEs	Y	Y	Y	Y
Physician FEs	Y	Y	Y	Y
N	292,094	292,094	211,315	211,315
R^2	0.82	0.77	0.88	0.86

Table B.4: Correlation between Direct Manipulation and Long-term Belief Distortion

This table exhibits the correlation between effects of direct manipulation and long-term belief distortion. Columns (1) to (3) are estimated in sample years 2016 to 2018. The remaining columns are estimated in sample years 2017 and 2018 due to the lagged explanatory variables. No control variables and fixed effects are included. Physician fixed effects and area-year fixed effects are included. Standard errors are clustered at the 3-digit zip code level, and t-statistics are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1) $\hat{\tau}_1$	(2) $\hat{\tau}_2$	(3) $\hat{\tau}_2$	(4) $\hat{\tau}_0$	(5) $\hat{\tau}_1$	(6) $\hat{\tau}_2$	(7) $\hat{\tau}_2$
$\hat{\tau}_0$	0.201***	0.174***	0.106***				
	(35.47)	(38.16)	(19.86)				
$\hat{\tau}_1$			0.336***				
			(50.58)				
$\hat{\tau}_{0,t-1}$				0.785***	0.212***	0.172***	0.091***
				(106.10)	(35.02)	(43.54)	(20.59)
$\hat{\tau}_{1,t-1}$							0.403***
							(66.65)
N	155,766	155,766	155,766	102,658	102,658	102,658	102,658
R^2	0.10	0.07	0.17	0.69	0.12	0.08	0.23

Table B.5: Variable Importance

This table exhibits variable importance in estimating the causal forest. Importance is calculated as a weighted sum of how many times feature j was split on at each depth in the forest, defined as

$$imp(x_j) = \frac{\sum_{k=1}^{4} \left(\frac{\text{\# Tree splits on } x_j \text{ at depth } k}{\text{\# All tree splits at depth } k} \right) k^{-2}}{\sum_{k=1}^{4} k^{-2}}.$$

Seniority is the number of years since physician's graduation in medical school. *PhysicanGender* is one if physician is female and zero otherwise. Other variables are defined in Table 1.

(1) *Variable*	(2) *Importance*
AvgAge	0.481
DualPct	0.175
AvgRisk	0.150
MalePct	0.059
Seniority	0.055
BlackPct	0.046
HispanicPct	0.030
PhysicanGender	0.002