Quotas and Opioid Crisis: An Experimental Analysis*

Yue Deng† Daniel Houser‡ Joachim Winter§

September, 2022

Abstract

The opioid crisis has claimed hundreds of thousands of lives and cost trillions of dollars. Over-prescription of opioids contributes substantially to those numbers by making such drugs overly available and overly accessible. Responding to the current trend of decreasing prescriptions each year, the Drug Enforcement Administration (DEA) embarked on a campaign to reduce the Aggregate Production Quota. However, certain aspects of DEA quotes remain unclear, including: i) how to best set efficient quotas; ii) the mechanisms behind quotas' potential effects, and iii) the influence of quotas on the relevant players in the opioid crisis. Drawing on Schnell's (2017) model of secondary markets, and based upon our previous findings of over-prescription in the presence of secondary markets (Deng and Houser 2022), we design a laboratory experiment to investigate the effects of quotas. We find that when quotas are in effect, patients visit physicians less frequently and physicians prescribe more strictly. Consequently, we find that introducing quotas reduces the total consumption of opioids and positively impacts overall health outcomes. Our results provide rigorous evidence that quota policies can reduce the total supply of opioids and the frequency of drug diversions; as a result, quotas have the potential to mitigate the opioid crisis significantly.

Keywords: Opioid crisis, Prescription Opioids, Quota

JEL: C91, I11, I12, I18, L10.

*We thank Johanna Mollerstrom, Cesar Martinelli, Thomas Stratmann, Alexander Tabarrok, Kevin McCabe, and participants at the ICES Brown Bag Lecture, APEE 46th International Conference and 2022 ESA Global Meetings for helpful comments. This research was financed by the Interdisciplinary Center for Economic Science (ICES) at George Mason University.

†Department of Economics and Interdisciplinary Center for Economic Science, George Mason University, Fairfax, VA, United States. Email: ydeng9@gmu.edu

‡Department of Economics and Interdisciplinary Center for Economic Science, George Mason University, Fairfax, VA, United States. Email: dhouser@gmu.edu

§Department for Empirical Economic Research, Ludwig Maximilian University of Munich, Munich, Germany. Email: Winter@lmu.de

1 Introduction

Since 1999, the ongoing opioid crisis has claimed more than 600,000 lives in the US and Canada. Among opioid related overdose deaths, nearly 247,000 have resulted from overdoses of prescription opioids[1]. A recent report on Lancet (Humphreys et al. 2022) predicts a staggering 1.2 million more opioid overdose deaths by 2029. Over-prescription of prescription opioids has served as a driving force in the crisis[2]. In 2016, to curb the crisis and reduce over-prescription, the DEA proposed reductions on five categories of opioids[3].(Linn et al. 2020).

While the number of manufactured prescription opioids has been declining, it remains unclear what role quotas play in the opioid crisis, specifically, the mechanisms through which quotas can be effective or ineffective. Past studies have investigated both demand side and supply side policies (Maclean et al. 2020; Catherine Maclean et al. 2022). While the evidence suggests that supply side policies, like quotas, out-perform demand side policies (Deiana and Giua 2021), no single policy alone is expected to solve the opioid crisis (Humphreys et al. 2022).

This paper uses a controlled laboratory experiment to address how quotas affect the opioid crisis. Specifically, we study how quotas influence the behavior of patients and physicians and the impact quotas have on overall health.

Given that less prescription opioids are being manufactured, one top concern for quota policies is protecting the availability and accessibility of the drugs for legitimate users. The DEA sets quotas annually, taking into consideration the legitimate needs of patients. However, if doctor do not alter their prescribing behavior, it follows that the same number of eligible patients will be competing for a reduced number of available drugs. Consequently, quotas could negatively affect the patients with the greatest pain relief demand. Therefore,

[1]In addition to prescription opioids, heroin and other synthetic opioids are driving the opioid crisis. The number of deaths involving prescription opioids declined from 2017 (17,029) to 2019 (14,139) but increased again in 2020 with 16,416 lives lost due to prescription opioid overdose.

[2]Past studies on the opioid crisis can be divided into two branches, one focusing on the causes and consequences of the opioid crisis (Alpert et al. 2022; Iversen and Lurås 2006; Lusted et al. 2013; Powell et al. 2020; Schnell and Currie 2018; Sullivan et al. 2010; Thombs et al. 2020; Maclean et al. 2020) and the other analyzing the effectiveness of ongoing policies aimed at curbing this crisis (Abouk and Powell 2021; Alpert et al. 2018; Arora and Bencsik 2021; Alexeev and Weatherburn 2022; Buchmueller and Carey 2018; Doleac and Mukherjee 2019; Meinhofer 2015; Meinhofer and Witman 2018; Mulligan 2020).

[3]The DEA proposes to reduce the amount of fentanyl produced by 31%, hydrocodone by 19 percent, hydromorphone by 25 percent, oxycodone by nine percent and oxymorphone by 55 percent. Combined with morphine, the proposed quota would amount to a 53 percent decrease in the amount of allowable production of these opioids since 2016. From 2016 to 2020, the proposed quota has reduced the amount of allowable production by 53%.

the first key question to answer is how quotas influence physicians' behavior[4].

Additionally, quotas are set and adjusted using data that reflects already-declining prescriptions rates. As a result, they do not effectively consider how quotas affect physicians' and patients' behavior. This could lead to inaccurate estimates about quotas' effectiveness, which could lead to inaccurate quotas being set.

Quotas can potentially affect both the legal market for prescription opioids and the illegal secondary market for the drugs. This effect on the secondary market is nonnegligible, as it influences quotas' overall health impact. If the causal effect of drug availability and drug diversions found in prior research (Powell et al. 2020) also applies here, we should expect more stringent quota policies to result in fewer drug diversions. However, in practice, this also depends on physicians' prescribing behavior. If, under the quota system, the same number of potential sellers receive prescriptions, we would not see a reduction in drug diversion activities.

There is limited evidence on how quotas affect the behavior of patients and physicians as it relates to the opioid crisis. Similarly, there is limited information about the unintended effects quotas could have on population health. The limited studies on quotas (Enzinger et al. 2021; Schatman and Wegrzyn 2020) have identified the potential issue of drug shortages and attributed the behavior of oncologists in prescribing less opioids to annual drug production quotas (Haider et al. 2019). These studies, however, lack systematic review. Nor do they analyze the susceptible patient population afflicted by shortage issues due to quotas or physician behavior. Our study contributes to the literature by providing causal evidence on how quotas affect physicians' prescribing behavior and the impact of behavior on different types of patients.

Our paper also contributes to the literature regarding quotas. Quotas are an important topic in environmental economics, gender economics, and political economics and thus have been studied as a valuable tool to improve efficiency in fisheries (Essington 2010; Heal and Schlenker 2008; Natividad 2016; Newell et al. 2005); increase the representation of women in political activities, and raise the leadership competence of men (Besley et al. 2017). Quotas may function effectively in these different fields by increasing the value of each unit subject to quota restrictions, thereby nudging people to use the quotas more efficiently. If this view is valid, we should expect more cautious behavior associated with prescription opioid use (reflected in consumption and prescription behavior) when quotas are limited. While the role quotas play in the opioid crisis lacks empirical study thus far, a recent paper on cannabis

[4]A similar supply-side policy that limits the length of physicians prescribing drugs, although shortened the length of doctors prescribing from 30 days to 7 days, leads to more patients being prescribed and an overall increase in number of drugs flowing to the secondary market (Sacks et al. 2021).

use proposed that user-set quotas could serve as a commitment device to nudge people away from excess use (Iwry and Kleiman 2017). As the first paper to empirically examine the role of quotas in the opioid crisis, our paper contributes to the field of quotas and tentatively unravels the mechanisms behind their effectiveness in the opioid crisis.

Given the various goals the DEA attempts to balance when setting quotas, it remains unknown whether the quotas effectively serve the population's health goals. As Jeffrey A. Singer (2019) noted, "DEA is tasked with the impossible assignment of determining just how many opioids, of all types, are needed to treat pain or provide anesthesia to roughly 325 million Americans in any given year." We add to the literature by examining two quotas, each focusing on one goal in the experiment. We thereby investigate the prioritized rule to follow when the DEA faces tradeoffs between multiple goals.

It is difficult to find natural experiments that allow for flexibly adjusting a quota[5] while controlling other variables each year. To solve this problem, we use a lab experiment to circumvent obstacles in the natural environment[6], where physicians' decisions can be influenced by factors like the physician's personal traits (Schnell and Currie 2018)) and environmental variations[7]. Using a lab experiment also helps overcome the data availability problem regarding tracking drug diversions in empirical literature. Additionally, it helps present a rather complete story about how quotas impact both the legal market and the illegal secondary market.

The theoretical framework models physicians' payoffs as being partially comprised of both patients' health outcomes and the constituted partially of monetary payoff (Ellis and McGuire 1986;Mcguire 2000; Schnell 2017). Drawing on the framework of Deng and Houser(2022), this paper introduces two quotas to answer the main questions about the role quotas play in the opioid crisis and an efficient quota strategy to follow. As in the base model, physicians are partially altruistic[8], and patients can divert prescribed drugs to the secondary market. Patient health outcomes partially influence physicians' payoffs. Health outcomes also depend on whether prescribed patients actually consume the drugs and whether non-prescribed patients buy the drugs from prescribed patients who sell them. Therefore, the key factors driving prescription health outcomes are the prescription standard set by the physician and drug diversion activities on the secondary market. Compared to the base model,

[5]Controlled substance quotas are established annually and adjusted only once during the year.

[6]In the natural environment, quotas are set annually and updated only once during the year.

[7]The environmental differences include, but are not limited to policy, heterogenous patients' profile distribution, norm, and culture. As shown by previous literatures (Molitor 2018; Phelps 1992; Phelps 2000; Bardey et al. 2021), there is vast heterogeneity in prescription behavior across physicians , even without the discussion of quota.

[8]The physician's utility is the weighted sum of revenue from prescribed visiting patients and the health impact conferred by their prescriptions.

the only variation in the quota cases are two quotas that limit the number of prescriptions each physician can dispense. Each quota aligns to one goal of DEA's quota policies. That is, one quota ensures a sufficient supply of drugs for all legitimate patients and the other serves to reduce drug diversion to the greatest extent possible. In theory, quotas with non-binding[9] features should not change physicians' or patients' behavior. All equilibria remain the same in quota cases as compared to the benchmark case without quotas.

While the theory predictions remain the same, quotas could nonetheless shift the behavior of patients and physicians by providing more clues to them. Indeed, DEA quotas can play a critical role in nudging players in the opioid crisis, if both patients and physicians respond well to the clues quotas serve. Likewise, clues can either benefit or mislead depending on whether they reduce unnecessary prescriptions and diversions and make prescriptions more approachable to legitimate users. By varying the quotas (clues) flexibly in the lab while maintaining the uncertain[10] nature of the game (given the presence of the secondary market), this paper provides evidence regarding whether the magnitude of quotas influences their effectiveness, and whether effectiveness is related to the clues quotas provide to the subjects in the game.

Our main findings are as follows: First, quotas help significantly reduce the number of prescriptions. This is evident from the fact that when quotas are in effect, physicians, on average, set higher prescription standards, and patients, on average, visit physicians significantly less. Second we find that quotas help significantly reduce the number of drug diversions when tentatively over-prescribing behavior occurs. Thirdly, we find that while quotas, regardless of their magnitude[11], generally help improve health outcomes, they are most effective when they correspond to the equilibrium number of prescriptions (i.e., aim to eliminate drug diversions). Finally, this paper validates the role risk attitudes play in influencing physicians' prescribing behavior when the secondary market is present. We further find that very stringent quotas reinforce the role risk attitudes play in differentiating physicians' prescribing behavior.

The remainder of the paper is organized as follows. Section 2 sets forth the procedure for setting quotas in practice. Section 3 reviews the model and shows the equilibrium and predictions. Section 4 describes the experimental design. Section 5 presents the results. Section 6 draws conclusions and discusses related issues and policy implications. Instructions for the experiment and proofs are in the Appendix.

[9]The non-binding quotas are the quotas that allow for no less than the equilibrium number of prescriptions, which maximizes the physicians' utility.

[10]The uncertainty is embedded in the presence of the secondary market where drug diversion exists.

[11]This only applies to the quotas we introduced in this paper. If the quota reduced further to an extent that is below the equilibrium number of prescriptions, the health impacts will fall.

2 Quotas in practice

The DEA sets quotas[12] for each basic class of controlled substances annually. The quotas are adjusted once a year. The Aggregate Production Quotas (APQ) can be viewed as a pie regulating the total amount of drugs all registered manufacturers are allowed to produce. The decision to implement quotas is a balancing act to limit the number of opioids while ensuring that the country's basic needs are met.

According to the Federal Register (Drug Enforcement Administration 2021), the DEA's first goal in setting quotas is to satisfy legitimate demand[13]. As mentioned in the Federal Register (Drug Enforcement Administration 2021), the DEA adjusts quotas for Schedule I and II controlled substance annually by assessing the annual needs for those drugs. Normally, the DEA enacts regulations and adjusts them in cooperation with other agencies[14] (e.g., the FDA) that provide useful consultations and information for the DEA's assessment. For example, confronting COVID–19 in 2020, the DEA increased the aggregate production quota (APQ) for drug products containing fentanyl, hydromorphone, morphine, and codeine after assessing the increased demands for those substances.

In setting quotas, the DEA considers both the legitimate demand for a particular controlled substance and the extent to which the substance is diverted. Table B1 in Appendix B shows DEA diversion estimates for the supplies of five covered controlled substances.[15] This data provides an empirical basis for the DEA to reduce the quotas for these five categories.

The DEA's consideration of the two factors noted above (legitimate demand and diversion) should, in theory, lend a fair amount of credibility to the APQ's annual implementation. However, the DEA may not be considering one potential conflict between these two goals: eliminating drug diversion will lead to a shortage of supply to the patients with legitimate medical demand for the drug[16]. Similarly, the goal of satisfying every single medical legitimate patient's demand will inevitably result in spillover effect of the prescribed drugs,

[12]There are 3 types of quotas: Aggregate Production Quotas, Individual Manufacturing Quotas, and Procurement Quotas. They rely on one another.

[13]Responding to comments regarding shortage, the DEA claims to be "committed to ensuring the adequate and uninterrupted supply of controlled substances to meet the estimated legitimate medical, scientific, research, and industrial needs of the United States, for lawful export requirements, and for the establishment and maintenance of reserve stock." Annual Production Quotas (APQs) set by the DEA provide for all legitimate medical purposes.

[14]The DEA mentioned that their actions were taken based on consultations with federal partners at the Department of Health and Human Services (HHS), drug manufacturers, drug distributors, and hospital associations. Similarly, in 2018, the DEA cooperated with the FDA to increase the quotas for injectable hydromorphone in response to the domestic shortage of injectable hydromorphone at that time.

[15]The estimates were calculated by combining the diversion estimates from the state PDMP (Prescription Drug Monitoring Program) data and the supply chain diversion data.

[16]The lucrativeness of the drugs' profit could outweigh some legitimate drug users' benefits by consumption.

i.e., drug diversions from some legitimate medical users to illegitimate users. Therefore, a rational quota should also attempt to balance the key goals (among many other goals). Considering the DEA's inflexibility in annually adjusting quotas and the delayed response time of manufacturers, shortages and drug diversions could easily occur if quotas are not set up and implemented correctly.

Using the information the DEA has at its disposal when regulating APQ, this paper models an environment where the regulator knows the number of patients with legitimate medical demand for a drug, as well as the extent of drug diversions. More importantly, when setting quotas, the regulator (the experimenter) understands how serious drug diversion would be. Our experiment provides a simple setup and ensures flexibility in shifting quotas. As a result, it provides a clean environment in which to closely monitor the effects of different quotas, i.e., the tradeoff between ensuring supply to legitimate users and mitigating drug diversion activities. Our results regarding quotas achieving the greatest population health outcome could also shed light on the relative importance of the two factors in setting quotas (whether a quota should put more weight on ensuring sufficient supply to legitimate users or on reducing drug diversions).

3 Decision Problem

While quotas represent a limit on the total number of drugs manufactured, our model assumes that quotas are shared equally among all physicians such that each physician's prescribing capacity is represented by a quota $\bar{q}$ in the model. Given that each patient's prescription result is a binary outcome of $\rho_i = [0,1]$ with $\rho_i = 1$ associated with one unit of prescription, quota $\bar{q}$ is also the maximal number of patients to whom a physician can prescribe. The key question is the pain level of patients who consume the drugs under the quota system (as opposed to when there are no quotas) as its answer determines the health outcome of quotas.

Following Deng and Houser(2022), we construct a model of patients' and physicians' behavior that incorporates quotas. However, the quota we introduce, $\bar{q}$, is non-binding[17] to the equilibrium number of prescriptions q^* that maximizes the utility of the physician. With $\bar{q} \geq q^*$, the equilibrium number of prescriptions q^* is unaffected by the quota's presence. The reason is that the equilibrium behavior of the key players (physicians and patients) determines the number of prescriptions, and this remains the same as in the benchmark case without quotas.

[17]The quota we introduce is either equal to or larger to the theoretical optimal number of prescriptions.

Given that non-binding quotas are set based upon q^* derived in the benchmark case without quotas, we first introduce the benchmark case with only the presence of the secondary market. We show the equilibrium number q^* after deriving the equilibrium behavior of the physician and the patients.

In the benchmark model, there is one physician j and I patients assigned to the physician. The number of prescriptions a physician prescribes (q) is mutually determined by the patients i ($i = 1, 2, \ldots I$) and the physician j. The physician j sets a prescription threshold (measured in pain), κ_j, such that patients with pain levels, κ_i reaching this threshold, κ_j, are eligible for the prescription. Those eligible patients who choose *visit* from the action set α_i ={*visit*, *not visit*} each receives one unit of prescription. Therefore, the physician's threshold and the eligibles' visiting decisions determine the number of prescriptions dispensed (N_{ij}). Without quotas, $q = N_{ij}$; that is, all the pain eligible patients who visit will receive a prescription; with quotas $\bar{q}$, $q = min\ \{N_{ij},\ \bar{q}\}$, such that the total number of prescriptions cannot exceed the quota. Given that visiting is costly and receiving a prescription provides a profitable resale opportunity[18], under equilibrium, all patients who will receive a prescription will choose to visit and those who cannot be prescribed will not visit the physician to avoid any unnecessary cost[19]. Given that all patients who can receive a prescription will visit under equilibrium, the number of prescriptions under the no quota case, q, equals the number of pain-eligible patients, $N_j(\kappa_j)$, which is completely determined by the physician's threshold decision κ_j[20]; when there are non-binding quotas, under equilibrium, the number of prescriptions, q, is also determined by the quota $\bar{q}$ as q = min $\{N_j,\bar{q}\}$.

Based on their prescription result, patients decide whether to consume or sell on the secondary market. That is, if prescribed ($\rho = 1$), the patient decides whether to *consume* or *sell* ($\alpha_i(\rho{=}1)$ = visit $\times$ {*consume*, *sell*}); if not prescribed ($\rho{=}0$), the patient decides whether to *consume by buy* or *do nothing* ($\alpha_i(\rho{=}0)$ = {not visit}$\times${*do nothing*, *consume by buy*})[21]. The market clears such that the number of drug diversions, $m = min$ {number of 'buy', number of 'sell'}[22]. Like in the benchmark case, patients on the secondary market are price takers such that no bidding process is involved. This design of price being set ex-ante aligns

[18]The prescription is valuable either from a consumption perspective or selling perspective, as the price on the secondary market always exceeds the cost of being prescribed.

[19]Under equilibrium, patients know whether they can be prescribed; therefore, the number of visitors equals the number of prescriptions made as only prescription-eligible ones will visit.

[20]Under equilibrium in the no-quota case (SM), all the patients with $\kappa_i \geq \kappa_j$ will visit; therefore N_j does not depend on patient i, but only on physician j's threshold.

[21]Theoretically, visit $\times$ {*do nothing*, *consume by buy*} is not considered, as the patients know whether they can be prescribed under equilibrium. Given that a visit is costly, patients who cannot be prescribed would choose "not visit."

[22]All tentative selling decisions with unmatched buyers due to "more 'sell' than 'buy'" are consumed by the patients and all unavailable tentative buying decisions end.

with the simplicity of the experiment design and is ecologically valid, given that a single physician cannot influence the market price.

We retrospect the generalized model in Appendix B and show how to derive the equilibrium number of prescriptions, q^*, after explaining patients' and physicians' problems and how their equilibrium decisions are derived. Given q^*, any quotas greater than or equal to q^* are non-binding to the equilibrium number of prescriptions.

In summary, for patients, their choice sets, incentives, and assumptions about whether they can receive a prescription under equilibrium[23] are all the same as in the benchmark case (SM). The sole difference is that the quota restricts the availability of the drug in question, limiting the physician's ability to prescribe. For physician j, the only variation in their utility is the upper limit on the number of prescriptions they can dispense. For this reason, the number of prescriptions is denoted in q instead of the number of eligible patients, N_j under threshold κ_j. Under equilibrium, all patients know whether they can receive a prescription, and at most $\bar{q}$ number of patients with $\rho_i = 1$ will find visiting optimal. Given the incentive of physicians to provide prescriptions still be the revenue R_j plus β_j times welfare of the health impacts received by "consuming" patient i [24], and the number of drug diversions being $m(\kappa_j)$, with the variation in the number of prescriptions dispensed being $q = min\ \{N_j,\ \bar{q}\}$, the physician's utility with quota becomes:

$$u_j(\kappa_j, \alpha_i^*\ ) = q(\alpha_i^* = visit|(\kappa_j, \bar{q})) \cdot R_j + \beta_j(\sum_{i=1}^{q-m(\kappa_j)} h(k_i) + m(\kappa_j) \cdot \bar{h}^{SM}) \tag{1}$$

$$\text{s.t. } q = min\ \{N_j, \bar{q}\}$$

3.1 Parameters and theory predictions

3.1.1 Parameters

Like the setup in Deng and Houser(2022), there are four sickness levels and four euphoria levels. From lowest to highest, the four sickness levels are: sick0, sick1, sick2, sick3 and the four euphoria levels are: enjoy0, enjoy1, enjoy2, enjoy3. Given that a patient's complete profile is only fully revealed to themselves, as well as the fact that euphoria level is private information, we build four different profiles of patients (shown in Table 1), which are sufficient to ensure the heterogeneous and privacy of the visit incentives of the patients while

[23]In theory, all patients with $\rho_i = 1$ visit such that the number of visitors $N_{ij} = min\ \{N_j,\ \bar{q}\} \leq \bar{q}$ and all visitors are prescribed. In the experiment, without knowing whether a patient can be prescribed, N_{ij} could exceed $\bar{q}$ and lead to the process of random assigning $\bar{q}$ out of N_{ij} patients being prescribed.

[24]Consuming patient i can either be the prescribed patient who receives health impact $h(\kappa_i)$ when consuming or the non-eligible patient who buys on the secondary market and receives the average health impact of the secondary market buyers ($\bar{h}^{SM}$).

simplifying the experiment. Knowing the four sickness levels and the four enjoyment levels, the belief of each participant in the experiment is that there are T types of patients, with T in the range of [4, 16].

Table 1: Experiment parameters of the patients: sickness levels and the associated health impacts, enjoyment levels and the number of patients of each type

Number of patients with each profile ($I = 24$)	Pain levels	Health impact	Enjoyment levels
I_{sick0}	κ_{sick0}	h_{sick0}	γ_{enjoy3}
8	0.94	-314	1000
I_{sick1}	κ_{sick1}	h_{sick1}	γ_{enjoy1}
4	1.4	-45	165
I_{sick2}	κ_{sick2}	h_{sick2}	γ_{enjoy0}
4	1.7	86	50
I_{sick3}	κ_{sick3}	h_{sick3}	γ_{enjoy2}
8	2.5	348	250

Notes: Each row represents the profile (κ_i, γ_i) of the patients with identical sickness level and enjoyment level. The same sickness level patients, if consuming, receives identical points which are $h(\kappa_i) + \gamma_i$.

Table 2: Experiment Parameters of the physician j

physician's revenue for prescribing to each patient (R_j)	103
physician's altruistic level (β_j)	1.1

Notes: the payoff of the physician with parameters is: $u_j(\kappa_j)$=103 $\times$ $q(\kappa_j)$ + 1.1 $\times$ $\sum_{i=1}^{q(\kappa_j)} h(\kappa_i)$ such that $q(\kappa_j) = min\ (N_j,\ \bar{q})$. The parameters of the patients and physician apply to all three cases in the model.

Table 3: Parameter values of the markets

cost parameters	c^v	c^d
on the primary market	103	15
cost parameter (value to sell)	p^{SM}	
on the secondary market	550	

Notes: Each of the 24 patients is given an endowment of 653 such that no one is having a budget constraint problem.

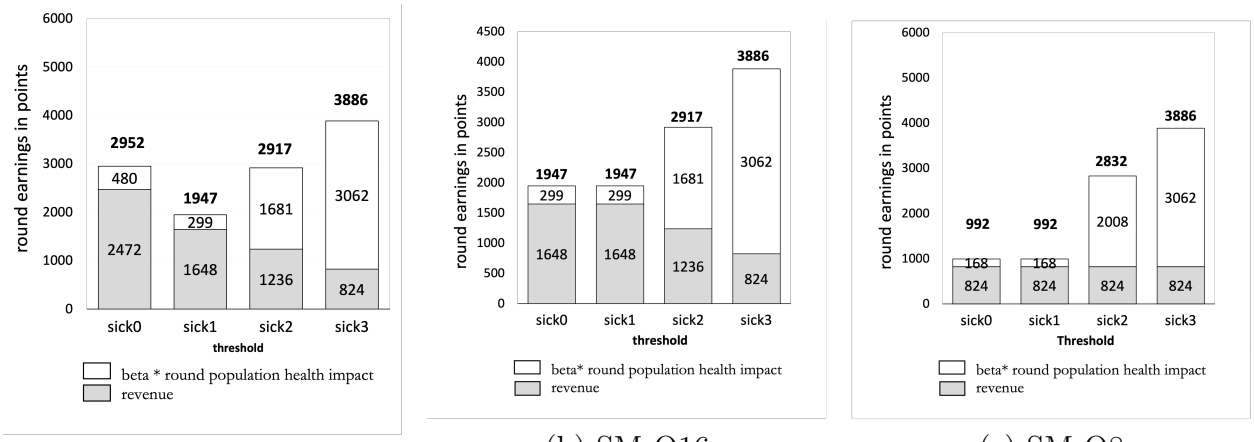
In Appendix B, we show how the equilibrium number of prescriptions, q^*, is derived under our parameterized setup. $q^* = 8$ is derived upon knowing the equilibrium prescription threshold of the physician ($sick3$), the optimal behavior of sick0 and sick3 patients being "consume" and the optimal behavior of sick1 and sick2 patients being "not consume"[25]. We further introduce two quotas with $\bar{q} \geq q^* = 8$.[26] The quota cases are denoted, respectively, as SM_Q16 and SM_Q8, where SM_Q16 allows for up to $\bar{q} = 16 (> q^* = 8)$ number of prescriptions and SM_Q8 allows for $\bar{q} = 8$ $(= q^*)$ number of prescriptions (exactly the equilibrium number of prescriptions). Table 4 summarizes the equilibrium in the benchmark case SM. We further illustrate that it is also the equilibrium in the two non-binding quota cases. As shown in Figure 1, the expected payoffs of the physician when choosing different thresholds in the three cases (SM, SM_Q16, and SM_Q8) always have $sick3$ as the equilibrium threshold achieving the highest payoff. The only difference between Figure 1 (b) and Figure 1(a) is the "shortened" revenue part payoff denoted in the grey areas in Figure 1 (b) when the threshold is $sick0$. Similarly, the only difference between Figure 1 (c) and Figure 1(a) is the "shortened" revenue part payoff denoted in grey areas in Figure 1(c) when the threshold is $sick0$, $sick1$, or $sick2$. The reason is that quota 16 is binding to the number of prescriptions induced by threshold $sick0$ ($N_j(\kappa_j = sick0) = 24 > 16$), and quota eight is binding to the number of prescriptions induced by thresholds $sick0$, $sick1$, and $sick2$ ($N_j(\kappa_j = sick0) = 24 > N_j(\kappa_j = sick1) = 16 > N_j(\kappa_j = sick2) = 12 > 8$). While the reduced cap on revenue payoff[27] due to the quota provides less incentive for the physician to set the threshold low, it does not change the equilibrium threshold from $sick3$.

The reasons for setting those two quotas are: (1) they are non-binding to $q^* = 8$, and (2) each quota satisfies one goal DEA has when setting quotas. Due to the non-negative health impacts received by the four sick2 and eight sick3 patients when consuming the drugs, 12 drug users have legitimate pain relief demands such that a quota of 16 satisfies their demand and provides a buffer for extra needs. By contrast, a quota of eight would theoretically eliminate drug diversions given that sick1 and sick2 patients who receive a prescription would always $sell$.

[25] In theory, given that a prescribed patient's choice set is $\alpha_i\,(\rho_i = 1)$=visit $\times$ $\{consume, sell\}$ and a non-prescribed patient's choice set is $\alpha_i(\rho_i = 0)$ = not visit $\times$ $\{$**c**$consume\ by\ buy,\ do\ nothing\}$, "not consume" refers to "sell" if $\rho_i = 1$ and "do nothing" if $\rho_i = 0$. Regardless of whether being prescribed, "not consume' is the optimal behavior of sick1 and sick2 patients.

[26] Given that the equilibrium threshold sick3 is robust to the patients deviating from the optimal visiting behavior, even if a patient of any sickness level deviates from the optimal visiting behavior due to not knowing whether they can be prescribed, this does not change the fact that any quota not smaller than 8 is non-binding to the number of prescriptions under the equilibrium threshold $sick3$. The reason is that under threshold of $sick3$, the maximal number of prescriptions can only be 8.

[27] The maximal revenue part payoff when the quota is 16 is 16*103 = 1648; The maximal revenue part payoff when the quota is eight is 8*103 = 824.



(a) SM (b) SM_Q16 (c) SM_Q8

Notes: Given that only sick0 and sick3 patients' optimal behavior is "consume" regardless of whether they can be prescribed, when the quota is 16 (shown in Figure 1(b)) and the threshold is sick0 or sick1, the expected payoff of the physician is 16 times 103 plus the health impact of 16 patients of sick0 and sick3.; consuming the drugs. When the quota is (shown in Figure 1(c)) and the threshold is sick0 or sick1, the expected payoff of the physician is eight times 103, plus the expected health impact of any eight patients among sick0 and sick3 patients consuming the drugs

Figure 1: Expected payoffs of the physician when choosing different thresholds in (a) the benchmark case, SM; (b) the case with quota of 16; and (c) the case with quota of 8)

Table 4: Equilibrium in SM case and non-binding quota cases

α^*_{sick0}	not visit $\times$ *consume by buy*
α^*_{sick1}	not visit $\times$ do nothing
α^*_{sick2}	not visit $\times$ do nothing
α^*_{sick3}	visit $\times$ *consume*
κ_j^{*SM}	$sick3$
N_j^{*SM}	8
m^*	0

Notes: Given that the decision of *consume by buy* gives positive utility to sick0 patients when buying successfully and incurs no cost to the tentative buyer who does not buy due to the unavailability of the drugs on the secondary market, the decision of *consume by buy* weakly dominate the decision of *do nothing* for sick0 patients. Given that none of the prescribed drugs to sick3 patients will be diverted to the secondary market, although the sick0 patients' weakly dominant strategy is to choose *consume by buy* when being non-eligibility, no drugs will be available on the secondary market nor be bought by sick0 patients under equilibrium.

3.2 Hypotheses

According to the theoretical payoffs of the physician setting different thresholds shown in Figure 1, our first hypothesis regarding the prescription threshold decisions of the physician is:

Hypothesis 1 *(Effect on physician's behavior)*
A non-binding quota does not change the physician's prescription threshold. In all three cases, the threshold sick3 is set most frequently by the physician.

Our second hypothesis concerns patient behavior. Under complete information, non-binding quotas should have no impact on the visiting behavior of the patients, as the theory-predicted threshold *sick*3 does not change in all three cases. The threshold, if known by patients, would result in eight sick3 patients[28] whose optimal behavior is α^*_{sick3}=visit $\times$ *consume* throughout the three cases.

While the equilibrium in all three cases is the same, the experiment with unrevealed threshold but revealed quota could change the rational expectations[29] of prescribed patients with the same pain level. This kind of behavior change is captured in Hypothesis 2 below. Note that sick3 patients' visit behavior should not be influenced by the presence of quotas, as these patients are aware of their sickness level's hierarchy among the four pain levels. Our second hypothesis is therefore:

Hypothesis 2 *(Effect on patient's visit behavior)*

Hypothesis 2a *With quotas, patients visit the physician less often. The smaller the quota, the lower the number of visitors.*

Hypothesis 2b *sick3 patients do not alter their visiting behavior regardless of the presence of quotas. The patients visit the physician more frequently in the SM case than in the NSM case.*

Our third hypothesis concerns the effect of quotas on the number of drug diversions. Given the optimal behavior of the patients at each threshold, the number of transactions at each threshold in SM are, respectively: $m^{SM}(\kappa_j = \ sick1) = 8$, $m^{SM}(\kappa_j = \ sick2) = 4$,

[28] If the quota reduces the visit rate of sick3 patients, this can be attributed to a side effect of quotas, as it would mean they are discouraging even the most pain-ridden patients.

[29] With the rational expectations being that each sickness level has an equal number of patients (6 for each sickness level patients), the quotas of 16 and 8 both serve as signals to reduce the chance of patients with lower sickness levels receiving a prescription. The lower the quota, the less patients should deem "visit" as optimal.

$m^{SM}(\kappa_j = \ sick0) = m^{SM}(\kappa_j = \ sick3) = 0$. Under a quota of 16, aside from the potential elevated number of drug diversions under threshold sick0 ($m^{SM_Q16}(\kappa_j = \ sick0) = [0, 8] > m^{SM}(\kappa_j = \ sick0) = 0$)), the number of drug diversions under the threshold of sick1, sick2 and sick3 i the same as the number of drug diversions under these three thresholds in SM. As shown in Table 5, $m^{SM}(\kappa_j = \ sick1) = m^{SM\text{-}Q16}(\kappa_j = \ sick1) = 8$; $m^{SM}(\kappa_j = \ sick2) = m^{SM\text{-}Q16}(\kappa_j = \ sick2) = 4$; $m^{SM}(\kappa_j = \ sick3) = m^{SM\text{-}Q16}(\kappa_j = \ sick3) = 0$. That is, when the physician's prescription standard makes everyone eligible ($\kappa_j = \ sick0$), a quota could instead increase the number of drug diversions by depriving some "consuming type" patients (sick0 and sick3 patients with $\alpha_i^* = consume$) of the opportunity to receive the prescription even if they are eligible; thus, they may become buyers on the secondary market.

When the quota is eight, however, the number of drug diversions is reduced under the previous "drug diversion triggering" thresholds $sick1$ and $sick2$. The reason is that the quota of eight is binding to the number of prescriptions induced by the threshold of $sick1$ ($N_j = 16 \ > 8$) or $sick2$ ($N_j = 12 \ > 8$) in SM. That is, the quota of eight increases competition among eligible patients and reduces the probability of selling type patients (sick1 and sick2) receiving prescriptions, even if thresholds (sick1 or sick2) make them eligible. As more sick3 patients receive prescriptions, less "selling" type patients sick1 and sick2 patients should be prescribed, and fewer drug diversions should be expected. For example, if the threshold is $sick2$, prescribed patients will be the randomly chosen eight patients among the eligible sick2 and sick3 patients who visit. Provided more than four patients of sick3 are prescribed, less than four of the "selling patients" sick2 will be able to receive the drug from the physician. Therefore, the theoretical number of drug diversions under threshold $sick2$ in SM_Q8 becomes any integer in the range of [0, 4]. Similarly, if the threshold is sick1, even if the eligible patients (sick1, sick2 and sick3) all visit, only eight of these 16 patients will be prescribed. Provided there is one sick3 patient being prescribed, the number of diversions cannot reach eight. Therefore, theoretically, the number of drug diversions in SM_Q8 when the threshold sick1 is chosen is $m^{SM\text{-}Q8}(\kappa_j = \ sick1) \in [0, 8]$[30]. By comparing the last two columns in Table 5, we further predict that the number of drug diversions in SM_Q16 is greater than or equal to the number of drug diversions in SM˙Q8. The third hypothesis is therefore:

Hypothesis 3 *(Effect on drug diversions)*

Hypothesis 3a *The number of drug diversions differs at different prescription thresholds, with sick1 triggering the maximal number of drug diversions. Sick2 is expected to trigger*

[30]If the prescribed patients are eight sick3 patients, then $m^{SM\text{-}Q8} = 0$.

the second highest number of drug diversions. And thresholds of sick3 and sick0 result in no drug diversion.

Hypothesis 3b *A quota of eight is a weakly dominant strategy for reducing the number of drug diversions, as compared to quota of 16.*

Table 5: The theoretically predicted number of drug diversions $m_j^*(k_j)$ under each threshold given the optimal decision of patients.

Threshold	SM	SM_Q16	SM_Q8
*sick*0	0	[0,8]	[0,8]
*sick*1	8	8	[0,8]
*sick*2	4	4	[0.4]
*sick*3	0	0	0

Notes: When the threshold is *sick*0, under SM_Q16 or SM_Q8, not all patients can be prescribed; therefore, the number of drug diversions is greater than in SM, as the unprescribed sick0 or sick3 patients turn to buyers. In both quota cases, if the prescribed patients due to randomization are sick1 and sick2, the number of buyers (16 patients of sick0 and sick3) exceeds the number of sellers (8 patients of sick1 and sick2), resulting in eight drug diversions.

Given that the physician's equilibrium threshold is *sick*3 and that sick3 patients are predicted to behave optimally to "visit" the physician in all three cases, the final drug consumers constituting the population health impact will be the eight sick3 patients across the three cases. Non-binding quotas should therefore not influence the population health impact. The fourth hypothesis relating to the population health impact of non-binding quotas is therefore:

Hypothesis 4 *(Effect on health outcomes)*
The population health impact is the same with and without non-binding quotas.

Finally, as both quota cases are extensions of the case with the secondary market, we should expect the risk attitude distribution of the low threshold-makers to be consistent with the distribution in SM. Similarly, we should expect the distributions of high threshold-makers' risk attitudes in the two quota cases not to differ significantly from the benchmark case. The scarcity of drugs due to quotas, however, could make physicians more cautious in prescribing when the quota drops. Therefore, our fifth hypothesis is:

Hypothesis 5 *High threshold-makers exhibit similar risk aversion distributions in SM, SM_Q16 and SM_Q8; low threshold-makers also show similar risk aversion distributions in all three cases. The presence of quotas might reinforce the role risk attitudes play in influencing prescription behavior.*

4 Experimental Setting

4.1 Experimental Design

We test our hypotheses using an incentivized computer-based lab experiment with 250 students (47% female, average age 23). The experiment took place at the laboratory at George Mason University using a program based on O-tree (Chen et al. 2016). Experimental sessions were conducted from October 2019 to Oct 2021 at George Mason University. The participants were invited through the university's economics research participant pool (via SONA System). The experiments lasted for about two and a half hours. Participants could earn $31.6 on average.

We employ a 3×1 experimental design, where the treatment variable is the number of prescriptions a physician can make (or "quotas"), $\bar{q}$. When the quota is present, it is binding to the I number of patients assigned to the physician, that is $\bar{q} < I = 24$ but it is not binding to the equilibrium number of prescriptions q^*, that is $\bar{q} \geq q^* = 8$. The experiment was developed following a between-subjects design. Each subject only participates in one environment, either $\bar{q} = I = 24$ (without quota, SM), $\bar{q} = 16 < I$ (SM_Q16) or $\bar{q} = 8 < I$ (SM_Q8). The sample characteristics in the three treatments are similar, see Table 6.

Table 6: Sample size of the key variables

	subjects in *SM_Q8* (N = 75)		subjects in *SM_Q16* (N = 75)		subjects in *SM* (N = 100)	
	Mean.	s.d.	Mean.	s.d.	Mean.	s.d.
Age	24.48	5.37	23.07	5.12	21.84	4.04
Female	0.50	0.50	0.43	0.50	0.52	0.50
Loss aversion	3.16	1.56	2.30	1.60	2.74	1.73
Risk aversion	5.32	2.25	4.80	2.49	5.36	2.31

4.2 Experiment procedure

The drug prescription game of all four treatments uses a 30 round 25 subject (one physician, 24 patients) design, with role randomly assigned at the beginning of each round, such that any subject could be a sick0, sick1, sick2 or sick3 patient or the physician in any round. The game tree at the end of the instructions in Appendix A takes SM_Q8 as an example to show the game.

Patients were each provided an endowment of 653 points. Subjects observed their profile and all possible payoffs at the beginning of the round[31]. Each subject as patient decided whether to visit the physician after viewing the physician's profile and their six possible outcomes in the game, depending on their choices[32]. Then, each patient answered a 7-likert question regarding their likelihood of being prescribed given their knowledge as shown in Table B1. While the patients were deciding, the physician chose the threshold from sick0, sick1, sick2, sick3, more severe than sick3 where more severe than sick3 is the same as sick3+ in the model, suggesting that none of the patients is eligible to be prescribed. In the quota cases, given that the threshold of sick0 in SM_Q16 and the threshold of sick0, sick1, and sick2 in SM_8 could induce more than the number of prescriptions allowed by the quota[33], a random selection mechanism was implemented by the computer to select $min\{N_{ij},\ \bar{q}\}$ number of prescription receivers, where N_{ij} is the number of prescriptions dispensed if without quota[34].

The physician's threshold decision was made after observing the sickness level distribution of the I patients, the physician's own profile R_j, β_j, and the different set of possible payoffs[35] linked to each threshold decision. In contrast to the theoretical framework, the patients making the visiting decision knew nothing about the prescription threshold in Round t, or the subject identity of the physician.

The patients who did not visit knew they were not prescribed without knowing whether

[31]Each patient was given a table of possible payoffs. The table in display reflected patient i's six possible payoffs when the patient chose their sickness level from a dropdown menu of sick0, sick1, sick2, sick3 and their enjoyment level from another dropdown menu of enjoy0, enjoy1, enjoy2, enjoy3. The table changed dynamically when a different profile was chosen. Learning the payoff tables of other possible profile patients (15 other tables) did not help the decision making of patient i.

[32]The six possible outcomes of a patient are as shown in the tree figure in instructions in Appendix A

[33]Under these thresholds in each of the two cases, the number of eligible patients exceeds the quota. Among the eligible patients, if more than $\bar{q}$ number of them visit, the quota is binding to the number of prescriptions the physician would otherwise make.

[34]N_{ij} is the number of eligible visiting patients given the threshold, κ_j, and N_{ij} number of pain eligible patients' visiting decisions. That is, N_{ij} is the number of patients with $\alpha_i(\kappa_i \geq \kappa_j) =$ visit.

[35]The physician tentatively chose thresholds from a drop-down menu to learn the set of possible payoffs when choosing each threshold. The table changed dynamically when different thresholds were chosen (five tables in total: table of *sick*0 as the threshold, table of *sick*1 as the threshold, table of *sick*2 as the threshold, table of *sick*3 as the threshold, and table of *more severe than sick*3 as the threshold).

they were eligible, and the patients who visited knew whether they were eligible. In SM, all eligible patients who chose to visit were prescribed. In SM_Q16 or SM_Q8, an eligible visiting patient could potentially not be prescribed if $\bar{q} < N_{ij}$. An eligible visiting patient who was not prescribed would be notified that they were eligible and in the queue to be prescribed, were not prescribed due to the quota. The physician was informed of the number of "visit" patients at each sickness level, the sickness levels of the prescribed patients[36], and the patients who eventually consumed the drugs.

In SM and both quota treatments, the experiment proceeded after the revelation of the prescription result to the patients (and the eligibility result to the visiting patients). Based on the prescription result, the prescribed patients were given the option to "consume" or "sell," and the patients without prescriptions were given the option to "buy" or "do nothing" on the secondary market. The submitted "buy" and "sell" orders would all succeed only if an equal number of buyers and sellers were present[37]. The physician was in the waiting page while the patients were making their decisions on the secondary market. After all the patient subjects made their choices, the physician was notified of the sickness levels of the final prescription receivers and the physician's round utility. The physician was also informed of the initially visiting patients' sickness levels and the initially prescribed patients' sickness levels, so that the physician could track drug diversions following their prescription decision. The patients were notified of the transaction result and their round end utility.

In all treatments, participants engaged in all 30 rounds (four practice rounds, followed by 26 real rounds) following the procedure mentioned above.

After all subjects completed the prescription game, they were asked to complete a loss-aversion task (Gächter et al. 2007) and a risk-aversion task (Holt and Laury 2002), followed by a short demographic questionnaire. When all subjects finished these parts, they were paid in cash privately. The payment is a random-chosen-round's (5th-30th rounds) payment of the drug prescription game and the earnings from the loss aversion task and the risk aversion task.

Table 7 summarizes the sample size of the key variables, which are the prescription thresholds the physician made, the visiting decisions made by all types of patients, and the round population health impacts in all the non-practice rounds.

[36]In SM and two quota cases with the number of eligible visiting patients not exceeding $\bar{q}$, the prescribed patients are the visiting patients with $\kappa_i \geq \kappa_{jt}$ (Threshold made by physician j at round t).; In SM_Q16 and SM_Q8 when the number of eligible visiting patients exceeds $\bar{q}$, the prescribed patients are the randomly selected $\bar{q}$ patients from the eligible visiting patients with $\kappa_i \geq \kappa_{jt}$.

[37]If the number of the buyers and sellers on the secondary market is not equal, then not all buyers and sellers will be able to transact successfully. For example, if there are five sellers and three buyers, then all three buyers can purchase the drug, while two sellers would not be able to sell, and, similarly, if there were more buyers than sellers.

Table 7: Sample size of the key variables

Treatment	*SM*	*SM_Q16*	*SM_Q8*
Participants	100	75	75
Patients	sick0: 100 sick1: 99 sick2: 99 sick3: 100	sick0: 75 sick1: 75 sick2: 75 sick3: 75	sick0: 75 sick1: 75 sick2: 75 sick0: 75
Physicians	72	75	75
Incentivized rounds	26 × 4 sessions = 104	26 × 3 sessions = 78	26 × 3 sessions = 78

Note: the key variables are the physicians' threshold decisions, patients' decisions and round population health impacts. Due to complete randomization of role updating at each round in SM, only 72 of the 100 participants were physician at least once, only 99 subjects played the role of sick1 and sick2 in SM.

5 Results

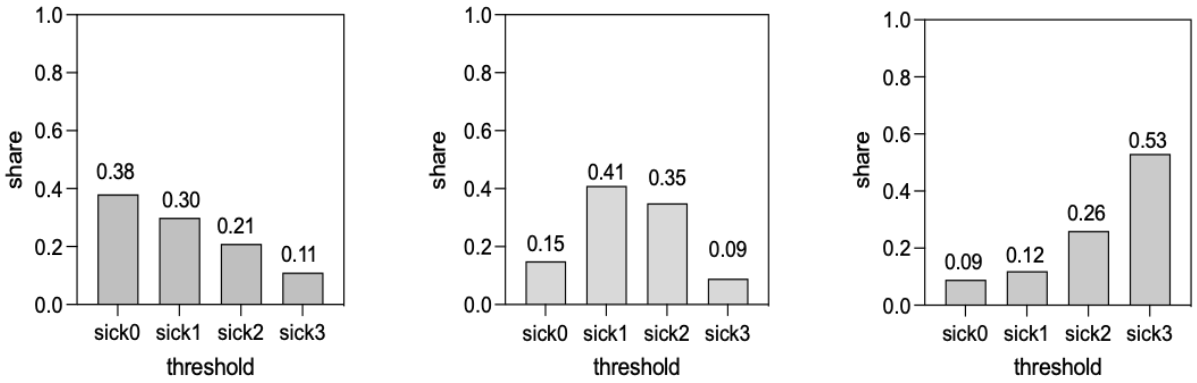
5.1 The physicians' prescription decisions

We begin the empirical analysis by plotting the share of physician subjects choosing each threshold. To analyze the effects of quotas on physicians' prescription decisions, we compare the physicians' prescription decisions in the two quota cases with the benchmark case, SM.

In SM, which had no quota, the physician was allowed to prescribe to all the patients with I = 24. As shown in Figure 2 (a), the most frequent chosen threshold was *sick*0, such that all patients were eligible to be prescribed. We reject our hypothesized threshold of *sick*3 in SM.

In SM_Q16, the number of prescriptions allowed to be prescribed reduces from 24 to 16. Accordingly, the highest monetary incentive[38] reduces in SM_Q16, as the maximal portion of visit fees collected by the physician drops from 24 to 16. While this change does not alter the equilibrium threshold, it does change the behavior of the physician such that they chose *sick*0 instead of *sick*1. As shown in Figure 2(b), the threshold of *sick*1 was chosen most frequently by 41% of the physicians in SM_Q16. The share of 41% is significantly higher than the 25% suggested by randomization (binomial test, $p = 0.001$). Similarly, in SM_Q8, due to the quota of eight, the monetary incentive reduction when choosing the threshold of *sick*0, *sick*1 and *sick*2 changes the behavior of the physician in SM_Q8 and ensures that equilibrium threshold of *sick*3 is chosen most frequently. With 53% significantly exceeding

[38] As shown in Figure 1(a), the highest monetary incentive (revenue part payoff) is 2472 and is achieved when the threshold is *sick*0 in SM without quota.



(a) SM (b) SM_Q16 (c) SM_Q8

Notes: N = 71 in SM, N = 75 in SM_Q16 and N = 75 in SM_Q8. There was one subject in SM who chose "more severe than sick3" when playing the physician for the first time. Since the threshold of "more severe than sick3" is known by all subjects as one that will lead to zero payments for both patients and physician, it is deemed as a mistake and is not counted in Figure 2. Figure 2 includes 71 subjects. In total, there were 72 subjects who were physician at least once.

Figure 2: Expected payoffs of the physician when choosing different thresholds in (a) the benchmark case, SM; (b) the case with quota of 16; and (c) the case with quota of 8

25% (binomial test, $p < 0.001$), the equilibrium threshold *sick*3 was chosen most often, as shown in Figure 2(c).

The behavior change of physicians in reacting to different quotas can be explained by nudges, which make some thresholds more prominent than others. Viewing the sickness levels of the 24 patients, the dominant chosen threshold in each treatment (*sick*0 in SM, *sick*1 in SM_Q16 and *sick*3 in SM_Q8), respectively, correspond to prescribing to the quota number of ($\bar{q} = 24$ in SM, $\bar{q} = 16$ in SM_Q16 and $\bar{q} = 8$ in SM_Q8) most severely sick patients given that all eligible patients visit. Quotas therefore raise the "competence"[39] of the initial drug receivers. This result is consistent with findings in the gender and political economics literature regarding quotas' effect on raising leadership competence (Bardey et al. 2021). Such nudging effect in raising "competence" of the prescribed patients is seen especially in Figure 2(b), where the threshold of *sick*2, even stricter than quota 16's suggested threshold[40], was chosen by 35% of the physicians. Given that 35% is also significantly higher than 25% and that the thresholds of sick2 and sick3 both represent non-over-prescribing behavior, the quota of 16 significantly reduces the over-prescription problem with just a small reduction

[39]The competence here is manifested in the receivers' sickness levels.

[40]Quota 16's corresponding threshold is sick1 as that threshold corresponds to prescribing to the 16 most heavily sick patients (8 sick3 + 4 sick2 + 4 sick1).

in the number of allowed prescriptions (from 24 to 16)[41].

Therefore, we reject Hypothesis 1, as the theory predicted threshold *sick*3 in SM was not dominantly chosen in SM or SM_Q16. Only in SM_Q8 did the physicians choose *sick*3 most frequently. This is our first result:

Result 1 *(Prescription behavior of physician)*
- The theory predicted threshold sick3 is only dominantly chosen in SM_Q8.
- The physicians respond well to the cues in the form of quotas. Sick0 giving all the patients the eligibility was chosen most frequently in SM without quotas; sick1 giving 16 patients the eligibility was chosen most frequently in SM_Q16; and sick3 that gave eight patients the eligibility was chosen most frequently in SM_Q8.
- Relative to the prescription thresholds in SM, physicians set higher thresholds in cases with quota. The lower the quota, the higher the prescription threshold the physicians set.

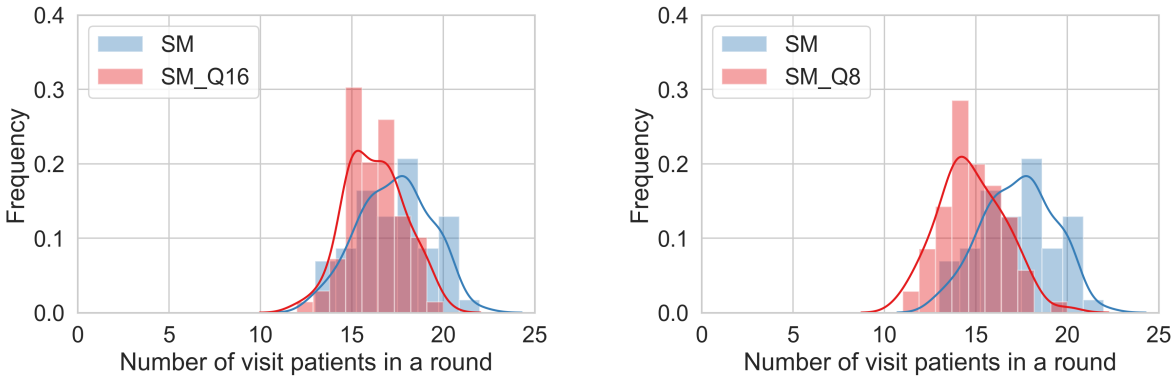
5.2 Patients

In this section, we analyze the effect of quotas on visit decisions of patients in the legal market and the number of drug diversions caused by their second decisions on the secondary market. These two decisions imply the effect of quotas on the demand side.

5.2.1 Visit behavior of the patients

Figure 3 shows the comparisons of the histogram of the number of visitors in a round in SM with each of the two quota cases. With 78 incentivized rounds in SM_Q16, 78 incentivized rounds in SM_Q8, and 104 incentivized rounds in SM, the number of round visitors in SM_Q16 and SM_Q8 are both significantly less than that in the benchmark case, SM.

[41] The quota of eight, while also achieving the effect of curbing the over-prescribing behavior, suffers from the problem of over-cutting and under-prescribing. Under a quota of eight, sick2 patients with legitimate health demand are not among the most severe eight-sickness-level patients. Therefore, given the dominant threshold of *sick*3, they are being deprived of the opportunity to access prescription opioids through legal channels.



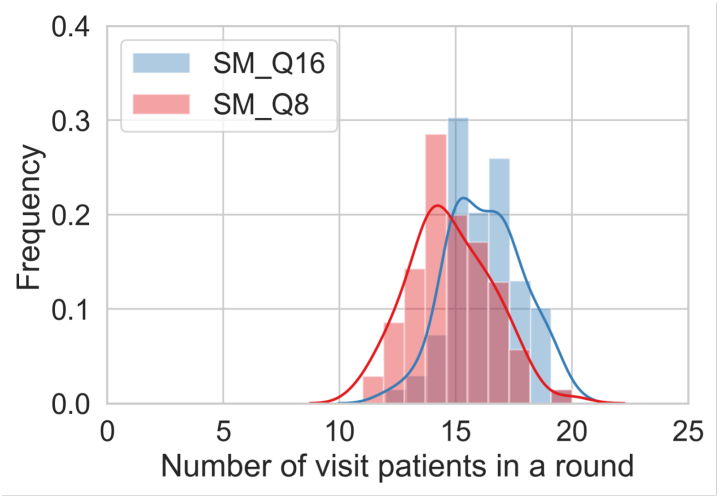
(a) SM_Q16 and SM

(b) SM_Q8 and SM

Notes: $N^{SM} = 104$, $N^{SM\_Q16} = 78$, $N^{SM\_Q8} = 78$

Figure 3: The comparison of each quota case with the benchmark case, SM with respect to the histogram of the number of round visitors.

To show the effect of the magnitude of quotas on patients' visiting behavior, we further compare the histogram of round visitors in SM_Q16 and SM_Q8 (shown in Figure 4). The figure shows that stipulating a stricter quota significantly reduces the number of visitors (t-value $= 5.46$, $p < 0.0001$).



Notes: $N^{SM\_Q16} = 78$, $N^{SM\_Q8} = 78$

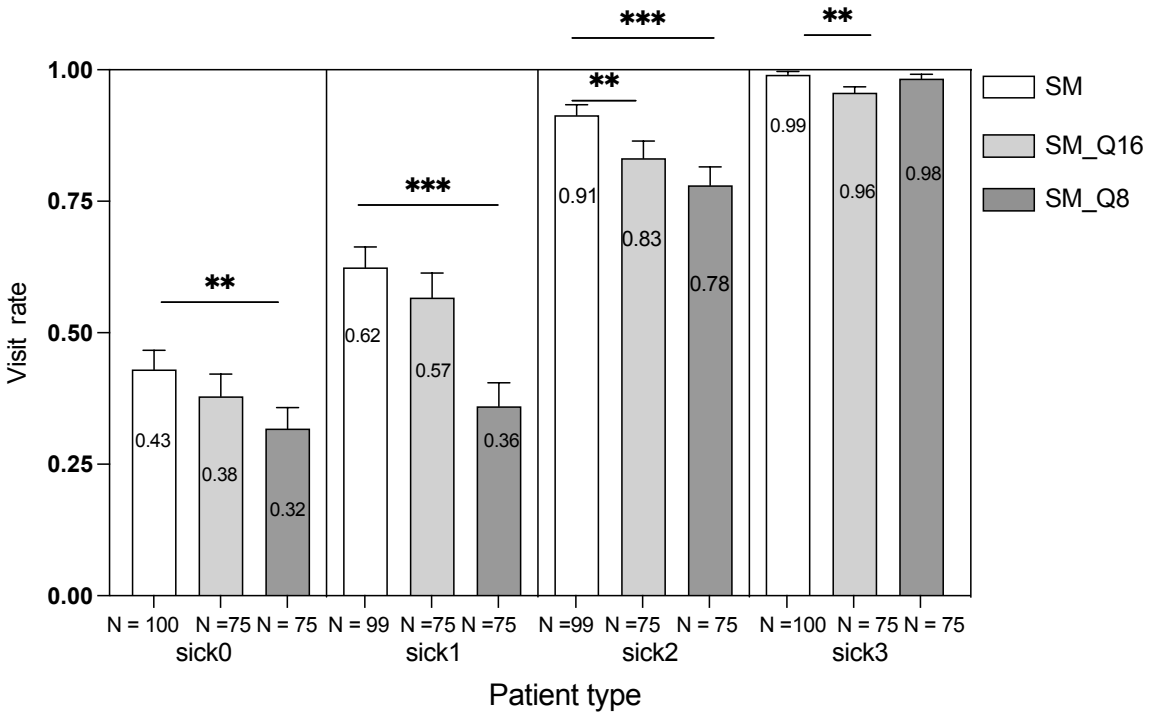
Figure 4: Histogram of number of visitors in a round in the two quota cases

Our second result regarding the visiting behavior of the patient therefore accepts the hypothesis under the incomplete information. Furthermore, from Figure 5, we show that even the most severely sick patient type (sick3) reduced their visit rate under the presence of quota.

Result 2 *(visit behavior of the patients)*
- Quotas change patients' visit decisions such that the stricter the quota, the greater the decline in the number of visitors.
- In contrast to Hypothesis 2, sick3 patients reduced their visit rate under the quota of 16 as well.

The theory-predicted hypothesis in both the complete and incomplete information settings suggests that the sickest patients (sick3 patients in our setting) should not change their visiting decisions, regardless of the presence of quotas. Below, we summarize the visiting behavior of each sickness level of patient, so we can conclude whether quotas have heterogeneous effect on patients of different sickness levels.

To check whether the effect of quotas in lowering visit rates is consistent for all sickness level patients, we calculate the visit rate of each sickness level patient type and compare the average visit rate of one-sickness-level patients in the quota cases with the benchmark case.



Notes: With role randomization at each round, each subject played each of the four sickness level patient types at least once. If a subject played one type of sickness level patient (e.g. sick0) for G rounds, and chose "visit" in this role for D rounds, the visit rate for him as this sickness level patient is D/G. Vertical black bars represent standard error of the mean (SEM). Each data point in each column is the average visit rate of a subject playing this patient type. * $p < 0.1$, ** $p < 0.05$, ***, $p < 0.01$.

Figure 5: Average visit rate of each sickness level patient

By comparing the same color bars in Figure 5, the within treatment comparison shows that the average visit rate is increasing with the subject's assigned sickness level, such that the more severe the sickness level, the more frequent the visit rate. This finding is consistent across three treatments.

Comparing the patient subjects' behavior between the benchmark case (SM) and the two quota cases (SM_Q16 and SM_Q8), we find that quotas of the same magnitude disparately affect as eight reduces the visit rate of patients with low sickness levels (sick0, sick1, sick2). By contrast, a less strict quota, like 16, reduces the visit rate of high sickness level patients (sick2 and sick3). As shown in Figure 5 (the left six bars), sick0 and sick1 subjects do not respond to the quota of 16, and we do not find significant differences in the average visit rate of sick0 or sick1 sickness level patients in SM_Q16 as compared to SM. sick2 sickness level patients, however, respond particularly well to both quotas, such that both the quota of 16 (t-value = 2.25, $p = 0.013 < 0.05$) and the quota of eight (t-value = 3.53, $p < 0.001$) significantly reduces the average visit rate of such subjects. However, sick3 subjects only reduce their visit rate in SM_Q16 (t-value = 2.75, $p = 0.003 < 0.001$), but not in SM_Q8 (t-value = 0.68. $p = 0.25$).

An unexpected finding regarding how a quota of eight affects patients' visiting behavior is that, compared to the benchmark case (SM), a quota of eight only reduces the visit rate of patients of sick0, sick1 and sick2, but not that of sick3. One possible explanation is that a quota of eight enhances the confidence of sick3 patients, such that they deem themselves more "qualified" to be prescribed compared to other sickness level patients under such a tight quota.

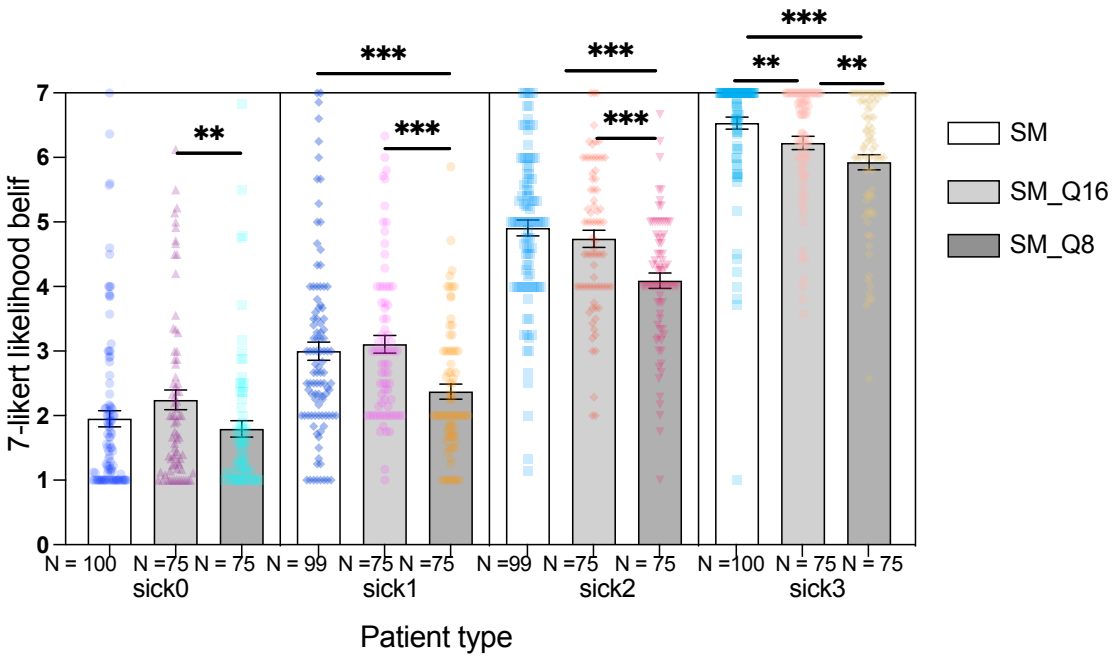
The reduced visit rate in quota treatments as compared to the benchmark case can be explained by the declining beliefs of the patients being prescribed in quota cases (shown in Figure 6), except for the fact that many subjects commented at the end of the experiment that they believed there were six patients for each sickness level[42]. Therefore, the quota, combined with patients' beliefs about the number of qualified patients and the demand of other patients (represented by their beliefs of the distribution of the 24 patients' sickness levels) influence their visiting behavior.

[42]For quota of 16, under the belief of six patients for each sickness level patients, sick2 and sick3 are the thresholds allowing for a non-exceeding-quota number of eligible prescriptions. Some sick1 patients would also find themselves potentially being prescribed. Similarly, under the quota of 8, the threshold allowing for a non-exceeding-quota number of eligible prescriptions is sick3 and some sick2 patients would still find them potentially being prescribed.

5.2.2 Belief of each type of patient

Figure 6 presents our findings after analyzing subjects' answers in the 7-likert questionnaire regarding their beliefs about the likelihood of being prescribed as each type of sickness level patient. As shown in Figure 6, within each treatment, the beliefs about the likelihood of being prescribed increase with the severity of the assigned patient's sickness levels. This finding is consistent with the increasing visit rate of the more severely sick patients shown in Figure 5.

By comparing subjects' beliefs as each type of sickness level patient across the three treatments, we find that a quota of eight reduces the beliefs of all the sickness level patients either as a comparison to SM or SM_Q16. However, compared with the benchmark case SM, a quota of 16 only reduces the beliefs of sick3 patients (t-value = 2.19, $p = 0.015$).



Notes: Vertical black bars represent standard error of the mean (SEM). Each dot in each treatment represents the average belief of one subject as one sickness level (regarding the likelihood of being prescribed).
* $p < 0.1$, ** $p < 0.05$, ***, $p < 0.01$.

Figure 6: Average beliefs of the likelihood of being prescribed as each type of patient

5.2.3 Drug diversions on the secondary market

Theoretically, the effect of quotas in reducing the number of drug diversions depends on the threshold set by the physician. The quotas' presence should not reduce, but actually increase the number of drug diversions when the threshold is $sick0$. Thresholds of $sick1$ and

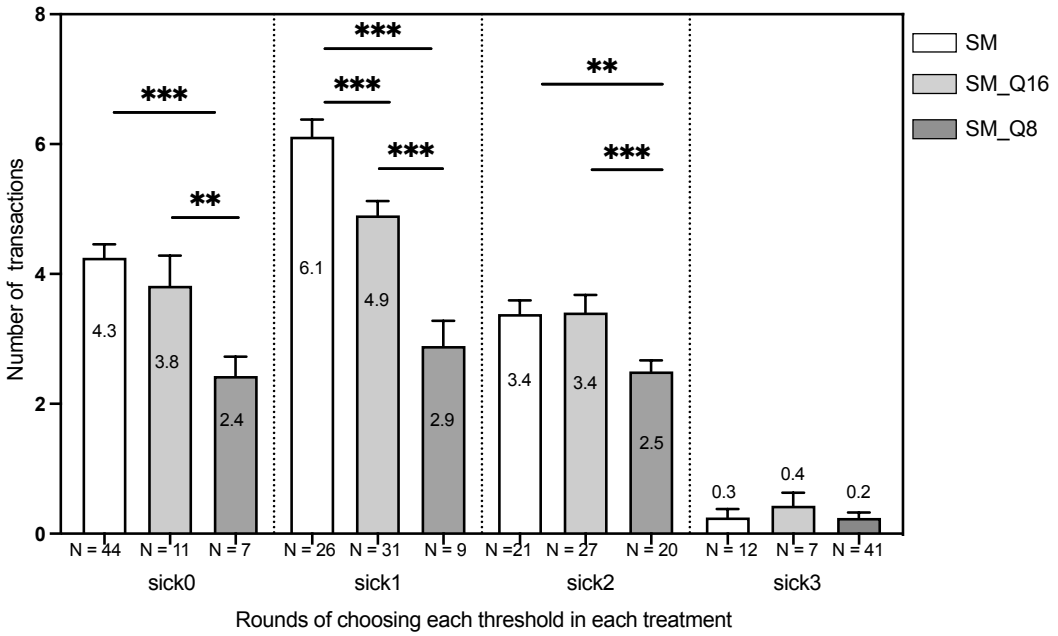
*sick*2, on the other hand, should make the quota effective in reducing the number of drug diversions (Table 5).

As shown in Figure 7 below, we find that the equilibrium threshold *sick*3 results in approximately zero drug diversions in all three cases. Unlike the theory predictions, however, all the other thresholds (including threshold of *sick*0) resulted in less drug diversions in the quota cases than in the benchmark case of SM. In the rounds when the threshold was *sick*0, the number of drug diversions in SM_Q8 was significantly lower than in SM_Q16 (Mann-Whitney U Test, U-value = 16, $p < 0.05$) and SM (t-value = - 3.38, $p < 0.001$). However, the number of drug diversions in SM_Q16 was not significantly lower than in SM under the threshold of sick0 (t-value = - 0.91, $p = 0.18$). This indicates that a quota aiming to satisfy all the legitimate medical demands, even at the direction of reducing the total supply, would not necessarily be able to reduce the number of drug diversions.

Besides, the presence of quota weakens the role physician's threshold decision plays in influencing the number of drug diversions. In both the benchmark case and SM_16, the threshold decision plays key role in influencing the drug diversions under different thresholds[43] while there exhibits similarly low level of drug diversions in SM_Q8 across thresholds ($sick0, sick1, sick2$)[44]. This indicates the low relevance of physician's over-prescribing behavior and the number of drug diversions when the quota is extremely stringent. That is, when the quota is low enough, the responsibility of reducing drug diversions shifts from the physician to DEA such that the physician could be less concerned about the drug diversion issues.

[43] As shown in Figure 7, in SM_16, the threshold is still highly influential for the number of drug diversions such that the effectiveness of quota 16 in reducing drug diversions depends on its impact on physicians' prescription decision.

[44] In rounds when the threshold was *sick*1, SM_Q16 resulted in significantly less drug diversions compared to SM (t-value = - 3.58, $p < 0.001$); SM_Q8 also had significantly less drug diversions than SM_Q16 (t-value = - 4.40, $p < 0.001$) or SM (t-value = -6.43, $p < 0.001$). In rounds where the threshold was *sick*2, SM_Q8 resulted in significantly less drug diversions than SM_Q16 (t-value - -2.63, $p = 0.006$) or SM (t-value = -3.22, $p = 0.001$).



Notes: Vertical black bars represent the standard error of the mean (SEM). Each data point in each column is the number of drug diversions under this threshold.
* $p < 0.1$, ** $p < 0.05$, ***, $p < 0.01$.

Figure 7: Average number of drug diversions under each threshold in the three cases

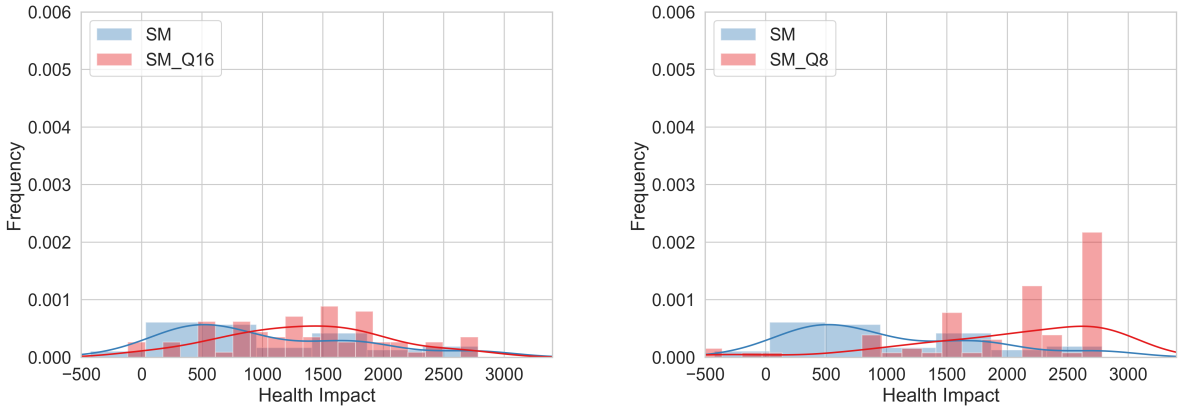
Corresponding to Hypothesis 3, the result regarding the effect of quotas on the number of drug diversions can be summarized as follows:

Result 3 *While the number of drug diversions depends on the prescription threshold, the presence of quota weakens the role physician's threshold decision plays in influencing the number of drug diversions. The more stringent the quota, the less contingent the threshold would influence the number of drug diversions.*
- In contrast to predictions in Hypothesis 3, the presence of quota eight effectively reduces the number of drug diversions under all the non-equilibrium thresholds (sick0, sick1 and sick2).
- In contrast to predictions in Hypothesis 3, the presence of quota 16 only effectively reduces the number of drug diversions under the thresholds of sick1.
- Like the predictions in Hypothesis 3, the equilibrium threshold sick3 achieves the lowest number of drug diversions such that the presence of quota is not necessary.

5.3 Population health impact

5.3.1 Treatment comparison

We next discuss population health outcomes, which are outcomes mutually driven by the behavior of patients and physicians. Figure 8 shows that the treatment difference in population health outcomes is significant between the benchmark case and either of the quota cases ($p < 0.01$).



(a) SM_Q16 and SM

(b) SM_Q8 and SM

Notes: $N^{SM} = 104$, $N^{SM\_Q16} = 78$, $N^{SM\_Q8} = 78$

Figure 8: The histogram of the population health impact of (a) the benchmark case (SM) and the case of quota 16 (SM_Q16); (b) the benchmark case (SM) and the case of quota 8 (SM_Q8)

Further, by comparing the histogram of the two quota cases (shown in Figure 9), we find that a quota of eight achieves a significantly higher population health outcome than a quota of 16 ($p < 0.001$).

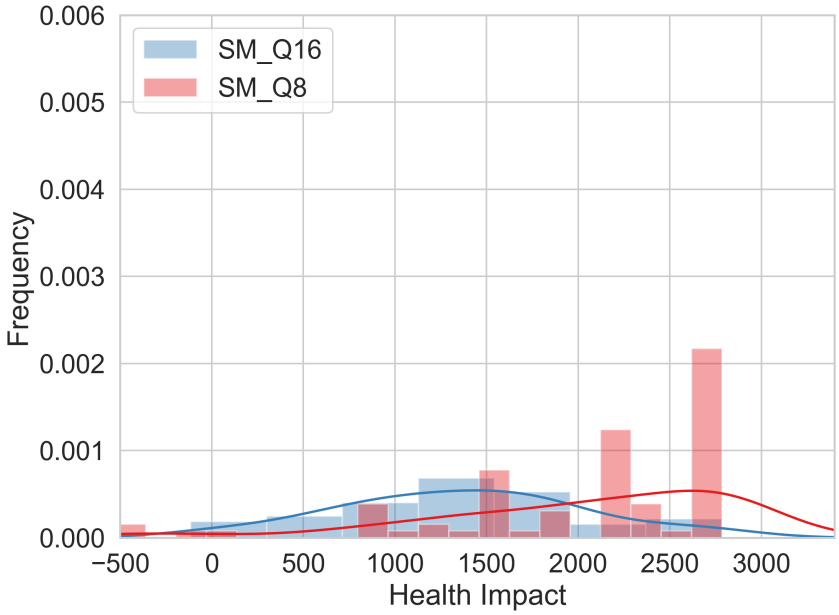


Figure 9: The histogram of the population health impact of SM_Q16 and SM_Q8

5.3.2 The population health effect of quota under different threshold decisions

As shown in Figure 8, the effect of quotas on improving population health impacts (as compared to the benchmark case) is significant. However, the positive impact of quotas on population health outcome could be driven either by the reduced frequency of over-prescription behavior of the physician[45] or the effect of quota itself on the public health outcome given the same prescription behavior. To explore the mechanism driving quotas' positive impacts on the population health, we plot the round population health impact under each threshold in the three cases[46] (shown in Figure 10).

Contrary to the predictions shown in Figure B2, we find that quotas effectively increase the population health impact when thresholds are low (corresponding to over-prescription behavior). Aligned to the predictions shown in Figure B2, Figure 10 shows that when the physicians set thresholds that do not represent "over-prescription," i.e., when the thresholds are *sick*2 or *sick*3, neither quota significantly impacts population health as compared to

[45] As shown in Figure 2, the dominate threshold was *sick*0 in SM, *sick*1 in SM_Q16 and *sick*3 in SM_Q8, so it could be that the reduced number of prescriptions drives the improved population health impact rather than other channels.

[46] Theoretically, the effect of the quota on the population health impact can vary under different thresholds (see Figure B2 in Appendix B).

the population health impact in the benchmark case, SM[47]. Comparing population health impacts across the two quota cases, we find that the magnitude of the quota (either eight or 16) has no influence on population health impact under the same prescription threshold[48].

Aside from the mean, quotas also influence variance in population health impact under certain thresholds. As shown in Figure 10 (see the vertical lines representing the standard errors to the mean), under the threshold of *sick*0, quotas significantly increase variance in population health impact in both SM_Q16 (F = 0.38, $p = 0.03$) and SM_Q8 (F = 0.072, $p < 0.001$), as compared to SM. Under the threshold of sick1, only the quota of eight (SM_Q8) significantly increases variance in population health impact as compared to SM (F = 0.31, $p = 0.02$). This suggests that physicians' over-prescription attempts could result in more unpredictive population health outcomes (could be linked to extremely harmful ones) under the presence of quotas[49]. The stricter the quota (e.g., quota of eight), the more care physicians should take to prescribe cautiously and avoid extremely harmful public health outcomes. Fortunately, as summarized in Result 1, we find that quotas quota indeed nudge cautious prescription behavior (as represented by higher prescription standards), and greatly improve the population health impact overall. Based on our findings regarding population health impact, our fourth Result is:

Result 4 *(Population health outcome)*
- Generally, the presence of quotas improves the population health impact, such that a quota of eight has a more positive impact than a quota of 16.
- The difference between a quota of eight and a quota of 16 in improving the population health impact is entirely due to the stricter prescription thresholds set by physician subjects in SM_Q8. Given the same threshold, the two cases do not have significantly different population health impacts.
- Under thresholds that are linked to "over -prescription," quotas (either quota) improve the average population health impact compared to SM.

[47]When the threshold was *sick*2, the average population health impacts in SM and SM_Q16 are not significantly different (t-value = 0.86, $p = 0.20$); nor do the two values in SM and SM_Q8 significantly differ (t-value = 0.94, $p = 0.18$). When the threshold was *sick*3, the average round population health impact in SM is also not significantly different to SM_Q16 (Mann Whitney U test, $p > 0.05$) or SM_Q8 (t-value = -0.41, $p = 0.34$).

[48]We do not find significant difference in the population health impact between SM_Q8 and SM_Q16 when the threshold is *sick*0 (Mann Whitney U test, U-value = 25, $p > 0.05$) or *sick*1 (t value = - 0.98, $p = 0.17$).

[49]Theoretically, under a quota of eight, the final drug consumers will be any eight of the 16 patients of sick0 and sick3. The probability that the final drug consumers will represent the worst case scenario (8 sick0 patients) is 1/9 given that the possible combination will be eight sick0; 7 sick0 + 1 sick3; 6 sick0 + 2 sick3; .....;1 sick0 + 7 sick3, 8 sick3. Without quotas, under the threshold of sick0, the theoretical final drug consumers can only represent one scenario, that is, all 24 patients.

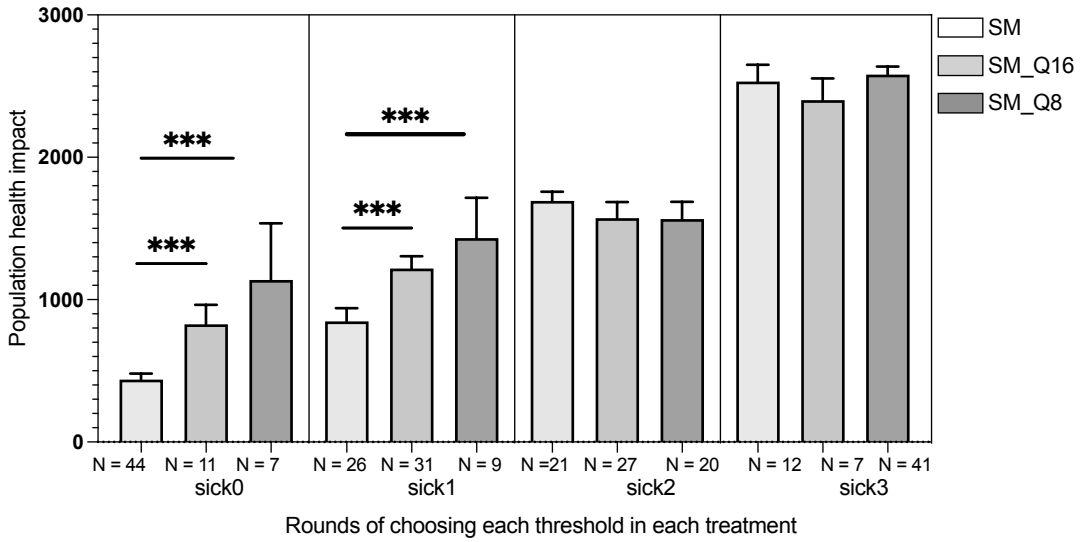


Figure 10: Population health impact under each threshold in the three cases

5.4 Risk aversion and physician's behavior

Previously in Deng and Houser(2022), we found that subjects' risk attitudes influence the prescription behavior of physicians when the secondary market is present, with over-prescribers exhibiting lower risk aversion levels, and non-over-prescribers exhibiting greater risk aversion levels. In this paper, to verify whether the presence of quotas weakens or intensifies the role risk attitudes plays in influencing physicians' over-prescription behavior, we plot the risk aversion attitudes of physician subjects choosing each threshold.

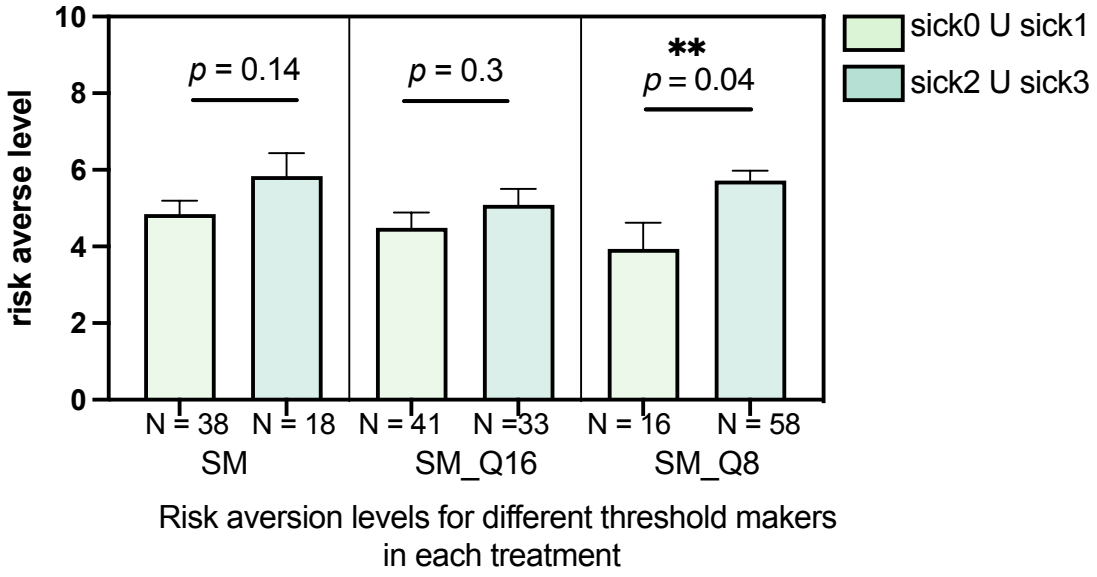
By summarizing the risk attitudes of subjects who over-prescribed (set threshold as $sick0 \cup sick1$) and comparing their risk attitudes with those non-over prescribers' risk attitudes (set threshold as $sick2 \cup sick3$), we find that the average risk aversion level is significantly lower for over-prescribers than non-over prescribers (t-value = - 2.94, $p = 0.002$) in SM_Q8. The left four bars in Figure 11 show that in SM_Q8, the risk aversion level is increasing with the physician's prescription threshold decision; that is, the lower the threshold the physician set (overprescribing more seriously) in SM_Q8, the lower the risk aversion level the physician subject might have. Comparing to the case in SM, the quota of eight intensifies the role risk attitudes play in influencing physicians' decisions. As shown in Figure 11, the subjects choosing different thresholds exhibited a more divergent risk attitude distribution compared to other cases. This is driven particularly by the low risk aversion levels of the sick0 threshold-makers in SM_Q8.

However, in SM_Q16, the average risk aversion levels of over-prescribers and non-over-prescribers does not differ significantly (t-value = -1.04, $p = 0.15$).

We further plot the risk aversion distributions of over-prescribers and non-over-prescribers in the three cases. By comparing SM_Q16 with SM (Figure B4) and comparing SM_Q8 with SM (Figure B5), we find that risk level distributions of low threshold-makers (over-prescribers) in SM and SM_Q16 (SM_Q8) are similar. The risk aversion distributions of high threshold-makers (non-over-prescribers) in SM and SM_Q16 (SM_Q8) also do not exhibit any significant difference[50].

Our fifth result regarding the role risk attitudes plays in influencing physicians' prescription behavior under quotas is as follows:

Result 5 *Consistent with Hypothesis 5, the distributions of the risk attitudes of the low threshold-makers do not differ across treatments. The same is true for the distributions of the high threshold-makers across treatments. Inconsistent with Hypothesis 5, only a quota as strict as eight intensifies the role risk attitudes play in influencing prescription behavior.*



Notes: Risk aversion level is the number of non-risky choice(s) subjects made in the risk aversion task.

Figure 11: Average risk aversion level of the physician subjects choosing the same threshold

[50]As shown in Figure B3 (a) in Appendix B, the distribution of risk aversion levels of the over-prescribers ($sick0 \cup sick1$) in SM and SM_Q16 does not differ significantly (two-sided Kolmogorov-Smirnov test, $p = 0.44$). The same is true for non-over-prescribers' risk aversion distributions (see Figure B3(b)) in the two treatments (two-sided Kolmogorov-Smirnov test, $p = 0.40$). As shown in Figure B4 (a), the distribution of risk aversion levels of over-prescribers (sick0 ∪ sick1) in SM and SM_Q8 also does not differ significantly (two-sided Kolmogorov-Smirnov test, $p = 0.81$). The same is true for non-over-prescribers' risk aversion distributions (see Figure B4(b)) in the two treatments (two-sided Kolmogorov-Smirnov test, $p = 0.99$).

6 Conclusion

In our laboratory experiment, we study whether quotas influence the behavior of the key players in the opioid crisis and whether they can improve public health outcomes inflicted by the opioid crisis. We find that quotas can greatly improve public health outcomes by effectively reducing the number of opioid prescriptions. Under quotas, physicians' prescription standards increase and patients' willingness to visit declines. The stricter the quota, the higher the prescription standard and the lower the frequency of patient visits. From a health perspective, a quota of eight, which corresponds to the number of prescriptions that should be prescribed to eliminate the drug diversions, outperforms a quota of 16, which aims to satisfy the medical demands of legitimate drug users. However, when controlling for the prescription behavior of the physician, the additional population health gains a quota of eight provides relative to a quota of 16 disappear. Our findings therefore suggest the importance of quotas' behavior perspective influence. Either approach to setting quotas—reducing drug diversions or ensuring sufficient supply to legitimate users—should work, provided they raise prescription standard the physician previously set. However, the effectiveness of quotas in improving health outcomes by raising prescription standards is a long-term goal. In the short term, policies aimed at reducing the availability and diversion of certain drugs can help reduce supply, as well as "switching" behavior, where drug-seekers switch to more addictive drugs like heroin. Therefore, to avoid side-effects caused by shortages in the short term, Medication-Assisted Treatments (MAT) should be implemented alongside quotas.

Additionally, we find that both quotas effectively reduce drug diversions and its presence weakens the role physician's prescription decision has in determining the number of drug diversions. Therefore, physician could be less concerned about the presence of the secondary market when an extremely low quota is present. Our study further validates the role risk attitudes plays in over-prescription behavior when the secondary market is present. Quotas even reinforce the role risk attitudes play in over-prescription behavior.

Given that this paper is based upon a controlled laboratory environment, the price on the secondary market is constant across treatments. The potential effect of quotas on the secondary-market-price of prescription opioids and the subsequent effect on total opioid consumptions is not discussed in this paper and could be an important topic in future research[51].

Furthermore, scarcity arising from a quota system could induce more cheating behavior,

[51] By incorporating the potential switching effect to illicit opioids when the price of the prescription opioids increase, (Mulligan 2020) discussed potential outcomes of both increased total consumption of opioids and decreased consumption of opioids depending on the marginal prices of illicitly manufactured opioids (e.g., heroin).

which could increase the burden and difficulty of physicians in distinguishing truly pain-stricken patients from disguised ones.

Finally, several behavioral explanations that have not been empirically scrutinized could be important contributing factors to the prescription response documented in this paper. For example, in each of the quota cases, the most frequent chosen prescription threshold is always same as the threshold allowing for the quota number of prescriptions. Environmental clues in the form of quotas nudge stricter prescription behavior, regardless of whether the clue is strong (e.g., quota of eight) or weak (e.g., quota of 16). Understanding the importance of these behavioral effects and why they happened represent promising avenues towards a brighter future.

A Appendix: Experimental instructions (Use *SM_Q8* as examples)

Thank you for agreeing to participate in today's experiment. You are about to participate in a decision-making experiment and at the end of the session you will be paid in cash based on your performance. By showing up, you have already earned $10. If you finish the experiment, at the end of this session, you will earn an additional $5 participation fee.

Today's experiment consists of 3 parts. At the beginning of each part, you will receive new instructions. You will spend most time on first part. Your decisions in one part have no influence on the proceedings or earnings of the other parts.

Your decisions and those of other participants will determine your earnings. Your earnings will be paid to you privately at the end of today's session. Your earnings in Part 1 will be denoted in points. At the end of the experiment, each point that you earned will be converted into 1 US cents (1 point = 0.01 US dollar).

Part 1: Decisions and Payoffs
This part consists of 30 rounds. In each round of this experiment, only 1 participant will be randomly chosen as a physician, the rest 24 participants are patients. At the beginning of each round, the role of each participant could be updated. There will be 4 practice rounds, the final payment is a random draw from the 5th to the 30th rounds. Thus, your role, decision and other participants' decisions in that round determine your final payment.

Initial endowment:
Each **patient** player is endowed with 653 points at the beginning of each round. The **physician** player of each round has no endowment.

Role Introduction:
• All the patients in this experiment are sick. But sicknesses can differ in their severity level. Each patient's sickness level determines how much "health impact" she/he could get from consuming a drug. The more severe the sickness, the more benefit the drug can bring to the patient. Because the drug can have negative side-effects, patients who are not very sick could be harmed, overall, by taking the drug. The drug can also bring different patient different levels of enjoyment. The enjoyment level has no relationship to a patient's sickness level. Thus, it is possible that a patient who is not very sick could enjoy taking the drug a

lot; whereas a patient who is very heavily sick might only receive little enjoyment from the drug.

For sickness levels, we differentiate the patients by 4 levels, from lowest to highest: sick, sick*, sick**, sick***. The different levels of enjoyment are, from lowest to highest: enjoy, enjoy*, enjoy**, enjoy*** (4 levels). As sickness levels are unrelated with enjoyment levels, there are 16 different combinations, and each patient only knows her/his combination but not the combination of others. Each sickness level is associated with a health impact number, and each enjoyment level is associated with a specific number. The sum of the "health impact" and the "enjoyment level" is the drug's value to the patients. The greater those two numbers are, the more the drug can contribute to the patient's payoff in that round.

• The physician in the market can observe all patients' sickness levels, but not the enjoyment levels. The physician's payoff is based on the health impact of the patients who eventually consumed the drug.

Table 1: possible sickness levels and enjoyment levels and their associated numbers

Sickness levels (x-axis number on the figure of *welcome page*)	Sick (x = 0.94)	Sick* (x = 1.4)	Sick** (x = 1.7)	Sick*** (x = 2.5)
Health Impact (y-axis number on the figure of *welcome page*)	-314	-45	86	348

Enjoyment levels	Enjoy	Enjoy*	Enjoy **	Enjoy ***
Number	50	165	250	1000

Environment Introduction:

Please note, there is a primary market, where the physician can prescribe a drug to the patients who visit. Also, there is a secondary market, where the patients can sell (buy) the

drug which they previously received (not received) from a physician. The physician has no control over secondary market activity. However, the physician's payoff in a round is determined by the secondary market's transaction results: total health impacts of those patients who eventually take the prescription in that round.

Your decision (as patient player) & Payoffs

(Please look at the figure at the end of this instruction for possible outcomes info):
As a patient, your decision consists of 2 components (I and II):

I. **Decision in the primary market**:
Visit the physician or not
A. **Visit** (653 points endowment - 103 Points visit fee = 550 Points)
(You need to pay 103 points visit fee **regardless** the prescription decision of the physician. Once you are prescribed, you pay an additional 15 points drug fee to get the drug)
B. **Not visit** (653 points endowment)

II. **Decision in the secondary market (associated payoff)**
Your available choices depend on the outcome at the end of the primary market phase.

- Possible Outcome 1: the patient visited and then got the drug from the primary market

(653 points - 103 points - 15 points drug fee)
A. **Sell** (+550 points)
B. **Consume** (+Health impact +Enjoy)

- Possible Outcome 2 & 3: the patient did not get the drug from the primary market

$$\begin{cases} Route\ of\ outcome\ 2: the\ patient\ visited\ and\ did\ NOT\ get\ the\ drug\ from\ the\ primary\ market \\ Route\ of\ outcome\ 3: the\ patient\ did\ not\ visit \end{cases}$$

A. **Buy** (-550 points)
   – if visited (653 – 103 – 550 + Health impact + Enjoyment level)
   – if not visited (653 – 550 + Health impact + Enjoyment level)
B. **Not get the drug**
   – if visited (653 -103 = 550 Points)
   – if not visited (653 Points)

*$Note$: If the number of the buyers and sellers on the secondary market is not equal then not all buyers and sellers will be able to transact successfully. For example, if there are 5 sellers and 3 buyers then all 3 buyers can purchase the drug while 2 sellers would not be able to sell,

and similarly if there are more buyers than sellers. This example also indicates that "buy" or "sell" decision will not necessarily lead to a successful transaction which involves 550 points (revenue for 'sell' patients AND expenditure for 'buy' patients are thus just pending if you hit "sell" or "buy" button, the transaction will not necessarily be executed).

Decision of physician:

As the patients decide whether to visit, the physician sets a threshold sick level for a prescription. The patient needs to be at least as sick as the threshold level to get the prescription from the physician. As the physician can see the profile of every patient (sickness level and the "health impact" bring by the drug), the physician is actually deciding who to prescribe in the primary market by setting this threshold sick level. The physician knows nothing about each patient's enjoyment level throughout the experiment.

The physician can prescribe to at most 8 patients.

- If the physician sets a threshold that leaves more than 8 visiting patients eligible for a prescription, only 8 of the sickness eligible visiting patients will receive the prescription from the physician. Among eligible visitors, the computer will randomly choose 8 patients to receive the prescription.

- If there are fewer than 8 sickness eligible visiting patients, all the eligible visiting patients will receive a prescription. The number of patients receiving a prescription will be equal to the number of eligible visiting patients.

Physician Payoff

Once all the patients have made their visiting decisions and the physician has made the threshold sickness level decision, the patients know whether or not she/he gets the prescription at the end of the "primary market" phase. Then all the patients enter the "secondary market" to make transactions. Once the transactions have completed, each patient knows their own transaction result. And the physician knows the patients to whom he/she prescribed the drug, as well all the people who ultimately consume the drug based on secondary market transactions.

Unlike the patient who cares about self-received-health-impact AND enjoyment level, the physician cares about the health impact on all patients who consume the drug.

◊ The payoff of the physician is the sum of the 2 parts below:

– *Visiting fee*:

$$\boxed{Number\ of\ visitors\ who\ reached\ the\ threshold\ sickness\ level \times (103\ points)}$$

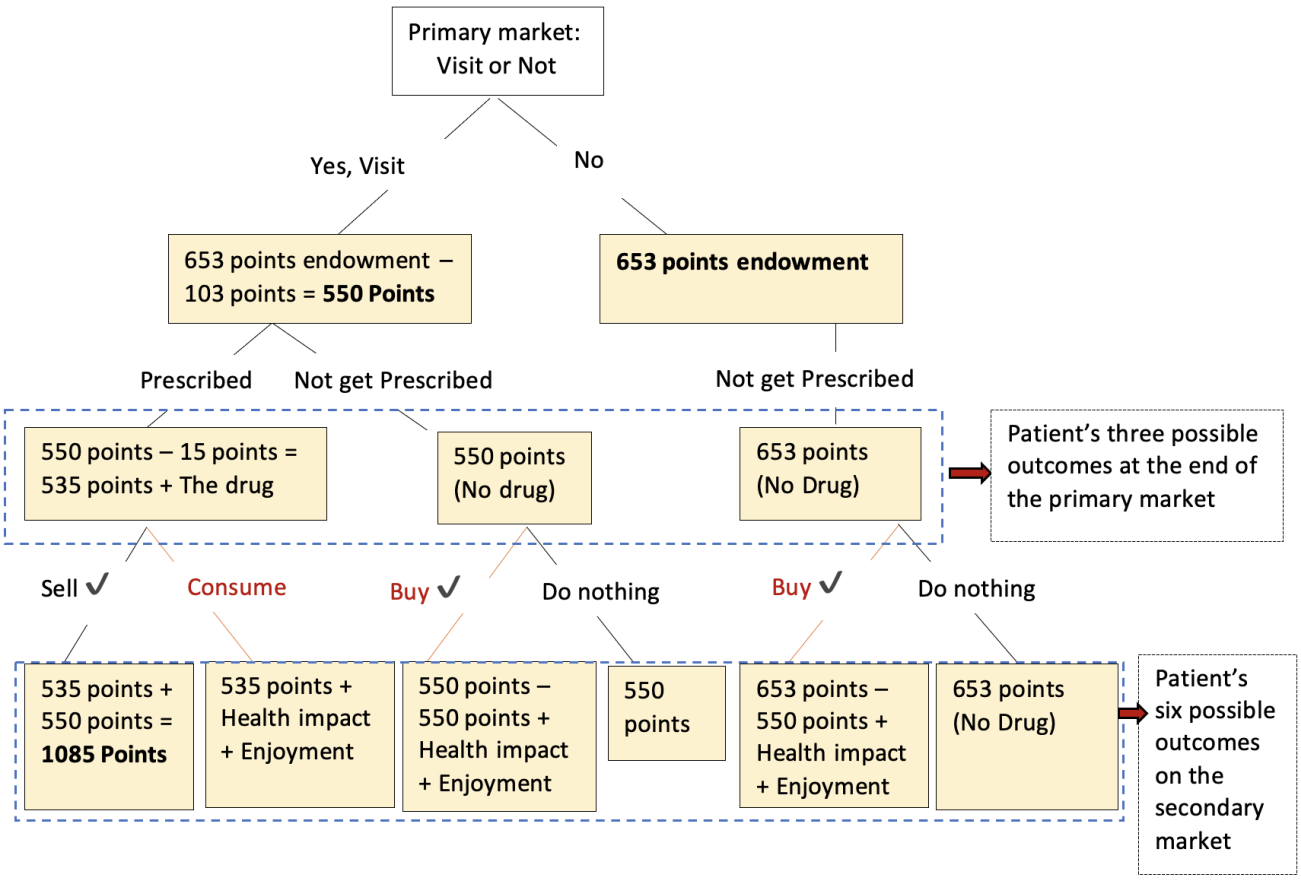
– *Health impact Part*:

$1.1 \times$ *Sum of the health impact of* **those patients who consumed the drug**

(*successful secondary market buyers AND prescribed patients who consume*)

**Note*: the highest possible payoff of a physician is when 8 prescriptions are distributed (visit fee part = 8×103 points) and the 8 prescriptions are eventually being consumed by the 8 most heavily sick patients (among the 24) on the secondary market (health impact part = 1.1 $\times$ sum of the health impacts of the 8 most heavily sick patients).

For detailed information about physician's payoffs, how the payoffs would change when making different threshold sickness level decisions AND how the payoffs would be impacted by different sickness level patients' decisions, please look at the dynamic table on your computer before choosing a formal threshold sickness level to submit.
This is the end of the instructions. You will be given a short quiz to ensure that you understand the instructions. Once you complete the quiz successfully, you'll proceed to the experiment.



Among the 6 final results (last row of the figure above). Result 2, 3, 5 (last step route marked red) are the 3 cases that you can get the drug after going through secondary market, the routes are:

- ◊ *Result 2: 'Yes, visit'→'Get Prescribed' → 'consume'*
- ◊ *Result 3: 'Yes, visit'→ 'Not Get the Prescription'→ 'Buy successfully'*
- ◊ *Result 5: 'No'→ 'Not Get the Prescription'→ 'Buy successfully'*

B Appendix: Supplementary analysis

B.1 DEA's empirical evidence in setting quotas

As shown in the table below, all categories have serious drug diversion problems; therefore, the DEA reduced the aggregate production quotas for each covered controlled substance listed in the table. On average, 18.8% decrease in quota numbers are implemented in 2022 for Schedule II opioids.

TABLE B1—Diversion estimates based on supply chain diversion data for covered controlled substance (g)

Controlled substance	(g)
Fentanyl	76
Hydrocodone	19,325
Hydromorphone	896
Oxycodone	45,368
Oxymorphone	524

Notes: The estimate does not contain any loss reported due to fire, weather, or other disaster.

Source: Federal Register / Vol. 86, No. 229 /December 2, 2021 / Notices

B.2 Model

B.2.1 Theoretical framework

The number of prescriptions distributed is $min\ \{N_{ij},\ \bar{q}\}$ where N_{ij} is the number of prescriptions prescribed to the pain eligible patients given the physician's pain threshold decision κ_j and the pain eligible patients' visiting decision which is $\alpha_i(\kappa_i \geq \kappa_j) = visit$.

To derive the equilibrium number of prescriptions q^*, we first introduce below the setup of the model without quota, the optimal behavior of the patients $\alpha_i^*(\kappa_j)$ under the physician's prescription standard (threshold), κ_j ,and the equilibrium threshold of the physician k_j^* given $\alpha_i^*(\kappa_j)$. We then show the cases with non-binding quotas and the theory predictions in these cases.

The case without quota is exactly the case of SM (with secondary markets) described in Deng and Houser 2022. As the name suggested, there are two markets: (1) a legal market where prescriptions are made; and (2) an illegal, secondary market where prescribed drugs are reallocated. On the legal market, the physician is assigned I patients who might need drugs from the physician. The patients are all painful, $\kappa_i \in \mathbb{R}^+$, and have a privately

known euphoria level, $\gamma_i \in \mathbb{R}$. The pain level and the euphoria level are assumed to be independent. After observing the I patients' pain levels, the physician j sets a prescription threshold (measured in pain): κ_j, such that all the patients with pain levels reaching this threshold $\kappa_i \geq \kappa_j$) are eligible to be prescribed. Given that the profit for selling the drugs is positive, all the patients with $\kappa_i \geq \kappa_j$ will seek the drugs from the physician (either to consume or sell) by taking the action of $\alpha i^*(\kappa i \geq \kappa j) = visit$. With each pain eligible "visit" patient receiving one unit of the prescriptions from the physician, the total number of prescriptions given the pain threshold κ_j and the optimal visiting decisions $\alpha i^*(\kappa i \geq \kappa j) = visit$ of the patients is $N_j(\alpha_i^*)$. On the secondary market, some of the prescribed patients divert the drugs from the legal market to non-prescribed buyers. This process of drug diversion changes the initial allocation of drugs made by the physician and might result in harmful population health impacts that concern the physician . The presence of quota, $\bar{q}$, restricts the number of prescriptions distributed such that $N_j = min(N_j(\alpha_i^*), \bar{q})$, but does not change the drug diversion process which is crucial for the opioid crisis.

To mathematically describe the decision problem in our setting, consider that a patient, i, if consuming the drugs, receives $v_i(consume)$ which represents the sum of the health impact of pain relief (denoted by the monetized health impact: $h(\kappa_i) \in \mathbb{R}$) and the euphoria level, γ_i, that is $v_i(consume) = h(\kappa_i) + \gamma_i$. By selling the prescribed drugs, the patient receives $v_i(sell)$, which equals the price of the drugs on the secondary market: p^{SM}. The incentive of the patients getting the drugs from the physician is therefore (1) to reduce pain, if $h(\kappa_i) > 0$, $\gamma_i \leq 0$; (2) to gain euphoria (even at the expense of hurting health), if $\gamma_i > 0$, $h(\kappa_i) \leq 0$; (3) to both reduce pain and gain euphoria, if $h(\kappa_i) > 0$, and $\gamma_i > 0$; (4) to gain profit, that is, $v_i(sell) - c(sell) > 0$, as $v_i(sell) = p^{SM}$ exceeds the cost of getting prescribed from the legal market, $c(sell)$ = visit cost c^v + drug fee c^d.

Knowing the prescription standard κ_j set by the physician, a pain eligible patient's problem given κ_j is an action $\alpha_i(\kappa_i \geq \kappa_j) \in A_i(\kappa_i \geq \kappa_j) = visit \times \{consume, sell\}$ so that $u_i(\alpha_i)$ is maximized. Given that $u_i(\alpha_i = visit \times consume) = h(\kappa_i) + \gamma_i - c^v - c^d$; $u_i(\alpha_i = visit \times sell) = p^{SM} - c^v - c^d$, a pain eligible patient's optimal decisions are described in Equation (2) and (3):

$$\alpha_i^{*SM}(\kappa_i \geq \kappa_j) = \arg\max(h(\kappa_i) + \gamma_i - c^v - c^d, p^{SM} - c^v - c^d) = \{visit \times consume\}, \quad (2)$$

$$\text{if } h(\kappa_i) + \gamma_i \geq p^{SM}$$

$$\alpha_i^{*SM}(\kappa_i \geq \kappa_j) = \arg\max(h(\kappa_i) + \gamma_i - c^v - c^d, p^{SM} - c^v - c^d) = \{visit \times sell\}, \quad (3)$$

$$\text{if } h(\kappa_i) + \gamma_i < p^{SM}$$

Given that visiting is costly, a pain non-eligible patient's (with $\kappa_i < \kappa_j$) problem is an action $\alpha_i(\kappa_i < \kappa_j) \in A_i(\kappa_i < \kappa_j) = not\ visit \times \{consume\ by\ buy, do\ nothing\}$ such that $u_i(\alpha_i)$ is maximized. Since the cost of choosing "*consume by buy*" on the secondary market is the price of the drugs on the secondary market, p^{SM}, $u_i(\alpha_i = not\ visit \times consume\ by\ buy) = h(\kappa_i) + \gamma_i - p^{SM}$; $u_i(\alpha_i = not\ visit \times do\ nothing) = 0$, a pain non-eligible patient's optimal decisions are described in Equations (4) and (5):

$$\alpha_i^{*SM}(\kappa_i < \kappa_j) = \arg\max(h(\kappa_i) + \gamma_i - p^{SM}, 0) = \{not\ visit \times consume\ by\ buy\}, \quad (4)$$
$$\text{if } h(\kappa_i) + \gamma_i \geq p^{SM}$$

$$\alpha_i^{*SM}(\kappa_i < \kappa_j) = \arg\max(h(\kappa_i) + \gamma_i - p^{SM}, 0) = \{not\ visit \times do\ nothing\}, \quad (5)$$
$$\text{if } h(\kappa_i) + \gamma_i < p^{SM}$$

From Equation (2) to (5), we can summarize that all patients, regardless of prescription eligibility, would optimally consume the drugs if $h(\kappa_i) + \gamma_i \geq p^{SM}$, and would optimally choose the alternative [52] if $h(\kappa_i) + \gamma_i < p^{SM}$. The first decision of whether to visit the physician, however, does rely on the eligibility determined by the relative severity of the sickness level of patient i, κ_i, and the physician's prescription standard, κ_j.

The physician, after viewing the I patients' pain levels, sets a pain threshold such that all patients reaching this threshold are eligible. Given the utility of the physician in Equation (6), the physician's payoff consists of two parts: the revenue and the population health impact that their prescriptions have bestowed. Concerning the revenue and the population health outcome, the physician's prescription decision is influenced by their revenue of prescribing to each patient $R_j \in \mathbb{R}^+$and the physician's preference denoted by β_j over the population health impact compared to the revenue.

$$u_j(\kappa_j, \alpha_i^*) = N_j(k_j, \alpha_i^*) \cdot R_j + \beta_j \cdot \sum_{i=1}^{N_j} h(\kappa_i) \quad (6)$$

Given that the patients who are pain eligible would optimally visit the physician, and that the patients with $h(\kappa_i) + \gamma_i \geq p^{SM}$ would optimally consume the drugs regardless of the eligibility, the utility of the physician at each threshold can be calculated. As shown in Equation (7), the threshold that achieves the highest utility given the optimal decisions of

[52] The alternative choice is sell for i with $\kappa_i \geq \kappa_j$; *do nothing* for i with $\kappa_i < \kappa_j$)

the patients under each threshold is the equilibrium threshold.

$$\kappa_j^* = arg\ max\{U_j(\kappa_j, \alpha_i^*)\} \tag{7}$$

The solution for Equation (7) is the threshold κ_j such that the expected marginal utility of prescribing to the last and least painful patient with κ_i=κ_j is zero. With the marginal utility of prescribing to a patient with κ_i=κ_j being one piece of revenue, R_j, plus the expected health impact with m/N_j probability of the prescribed patient selling to the patient on the secondary market, the equilibrium threshold κ_j^* satisfies:

$$R_j + \beta_j \left(\frac{N_j - m}{N_j}\ h(\kappa_j^{*SM}) + \ \frac{m}{N_j}\ \bar{h}^{SM} \right) = 0 \tag{8}$$

In Equation (8), the expected health impact is the weighted average of the health impact received by the prescribed patient with pain level κ_j^* and the average health impact of a patient buying on the secondary market, $h^{\bar{S}M}$. This generalized format of the solution is not as important as Equations (6) and (7), as our experimental design uses a discrete pain level setting.

With the optimal number of prescriptions being q^* under κ_j^* and α_i^*, any quota greater or equal to q^* is theoretically non-binding and would not influence the solution of κ_j^* or the optimal behavior of the patients given k_j^*.

Following the task of each agent and the basic model setup, we summarize the assumptions regarding the information set of each agent in Table B1 below.

Table B1: key assumptions of the model about the information set of the two agent types

Agent	Information
All (Common knowledge)	• I Patients with four discrete pain levels and four discrete euphoria levels $\{\kappa_i, \gamma_i\} = \{(\kappa_{sick0}, \kappa_{sick1}, \kappa_{sick2}, \kappa_{sick3}) \times (\gamma_{enjoy0}, \gamma_{enjoy1}, \gamma_{enjoy2}, \gamma_{enjoy3})\}$ • The value of consuming the drug, $v_i(\alpha_i = consume)$: Pain relief + euphoria level = $h(\kappa_i) + \gamma_i$ • The value to sell the drugs, $v_i(\alpha_i = sell)$: p^{SM} • The cost to consume the drugs ◊ on the primary market: the visit cost + the drug fee = c^v+c^d ◊ on the secondary market: the price of the drugs on the secondary market, p^{SM} • Physician's profile $\{R_j, \beta_j\}$ • Physician's threshold decision κ_j (only known under equilibrium)
Patient (Private information)	• Patient's euphoria level γ_i
Physician	• I Patients' pain level distribution: $\{\kappa_i$: i = 1,2,3...$I\}$

B.2.2 Equilibrium Analysis

Given that the equilibrium number of prescriptions q^* is a mutual outcome derived from the equilibrium behavior of the physician and the patients, we first derive the equilibrium behavior of the patients under parameterization in SM. Given the visit fee c^v being 103, drug fee c^d equals 15, and the value (cost) of selling (buying) the drugs on the secondary market being 550 (p^{SM}), all the patients in Table 1, if eligible for prescription, would visit the physician, due to the profitable selling opportunity. That is, $\alpha_i^*(\kappa_i \geq \kappa_j) = visit \times \{sell, consume\}$. Under equilibrium, the patients know the threshold the physician sets and can make optimal visiting decisions. The knowledge of the threshold, κ_j, however, is absent in the experiment and therefore can make the patients make non-optimal visiting decisions depending on their beliefs regarding the threshold.

As shown in Equation (1) to (4), the second decision about whether to consume on the secondary market depends solely on the relative magnitude of $h(\kappa_i) + \gamma_i$ and p^{SM}, therefore, the patients' optimal decision of whether to consume does not depend on the knowledge of the threshold and should be consistent in the theory and in the lab where the threshold is not ex-ante revealed. With $h(\kappa_i) + \gamma_i \geq p^{SM}$, sick0 are sick3 patients' optimal behavior would be to consume the drugs rather than the alternative, as the value to consume exceeds the

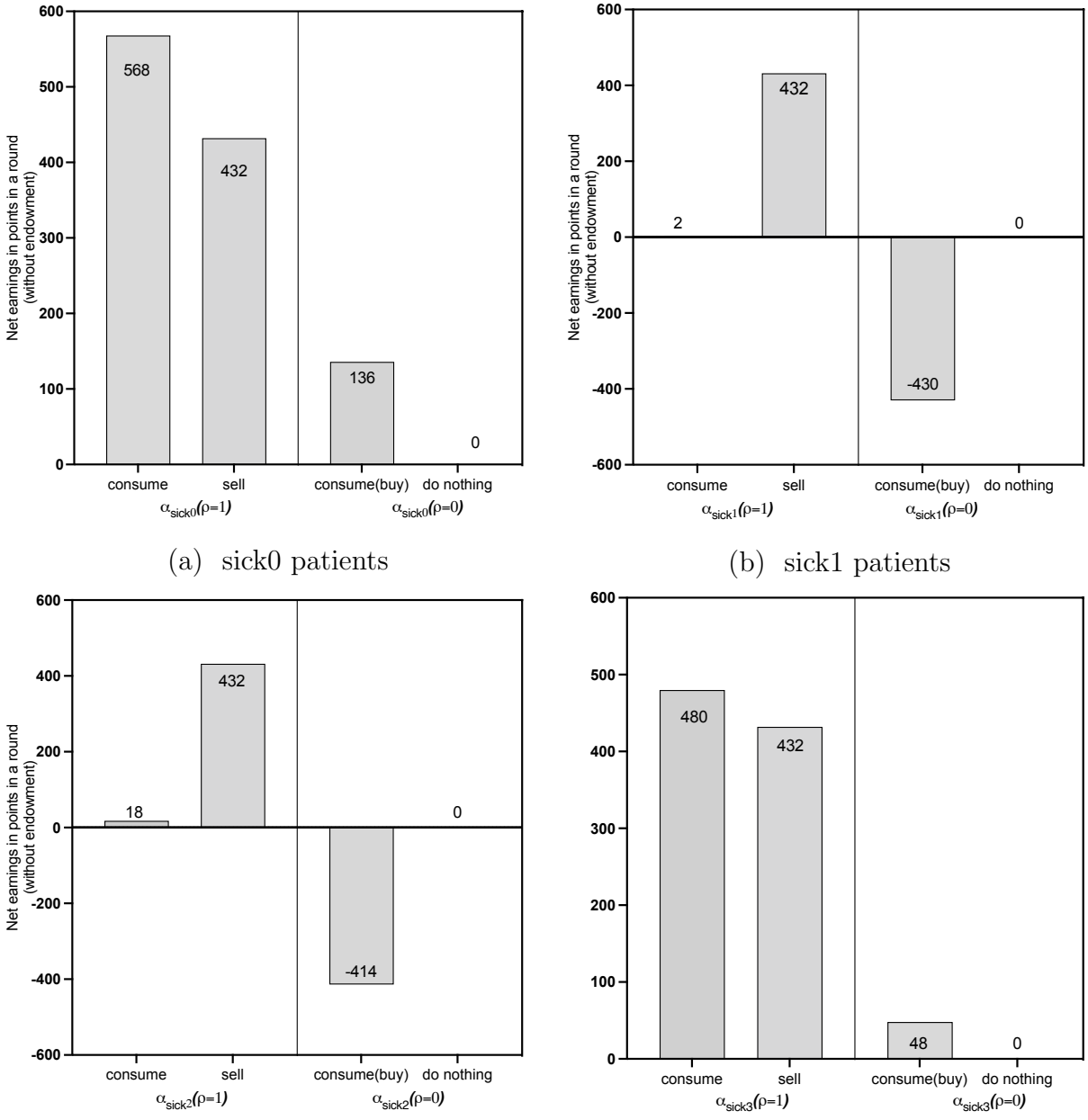
value to sell if being eligible and the utility to consume by buy also exceeds the payoff of do nothing if not being eligible to be prescribed. Specifically, sick0 patients are incentivized by the euphoria level($\gamma_{enjoy3} = 1000$) to have the optimal behavior, α^*_{sick0}, as consume, whereas sick3 patients are incentivized by the pain relief demand ($h_{sick3} = 348$) to have the optimal behavior, α^*_{sick3}, as consume. That said, if eligible, sick0 and sick3 patients' optimal strategy is $\alpha^*_{sick0}(\kappa_i \geq \kappa_j) = \alpha^*_{sick3}(\kappa_i \geq \kappa_j)$ = visit $\times$ *consume*; if non-eligible, sick0 and sick3 patients' optimal strategy is $\alpha^*_{sick0}(\kappa_i < \kappa_j) = \alpha^*_{sick3}(\kappa_i < \kappa_j)$= not visit $\times$ *consume by buy*. With $h(\kappa_i) + \gamma_i < p^{SM}$, patients of sick1 and sick2, however, would visit the physician driven by the incentive to sell on the secondary market. That is, if eligible, they would both have the optimal strategy as: $\alpha^*_{sick1}(\kappa_i \geq \kappa_j) = \alpha^*_{sick2}(\kappa_i \geq \kappa_j)$ =visit $\times$*sell* , and if non-eligible , they would both have the optimal strategy as : $\alpha^*_{sick1}(\kappa_i \geq \kappa_j) = \alpha^*_{sick2}(\kappa_i \geq \kappa_j)$= not visit $\times$*do nothing*. That said, they would always prefer to NOT consume regardless of whether they are eligible for the prescription.

Given the optimal decisions of the patients under each threshold decision of the physician, the physician's task is to choose a threshold from $\{sick0, sick1, sick2, sick3, sick3+\}$ such that sick0 represents the lowest pain threshold offering every patient eligibility, and sick3+ represents the strictest threshold, under which no patients are eligible for prescriptions. Given the revenue of the physician prescribing to one visiting patient being $R_j = 103$, and the preference of the physician on health impact as compared to the revenue being 1.1, the parameterized utility function of the physician is: $U_j = N_j(\alpha^*_i = visit\ |\kappa_j) \times 103 + 1.1\sum_{i=0}^{N_j(\alpha^*_i)} h(k_i|\alpha^*_i = consume,\ \kappa_j)$. Under this setup, the equilibrium threshold κ^*_j is characterized by Equation (9) below where 0 is the payoff of setting the threshold as sick3+:

$$\kappa^*_j = \arg\max \Big(u_j(\kappa_j = \kappa_{sick0}, \alpha^*_i), u_j(\kappa_j = \kappa_{sick1}, \alpha^*_i), u_j(\kappa_j = \kappa_{sick2}, \alpha^*_i), u_j(\kappa_j = \kappa_{sick3}, \alpha^*_i), 0 \Big) \tag{9}$$

B.3 Theoretical earnings of the patients behind each choice and his optimal decisions

Figure B2 below shows the payoffs of the prescribed patients (if being prescribed, denoted as $\rho = 1$) choosing consume or sell and the payoffs of the non-prescribed patients ((if not being prescribed, denoted as $\rho = 0$) choosing *consume* (*buy*) or *do nothing*. The figure illustrates that the optimal choice of the sick0 and sick3 patient is consume (regardless whether $\rho = 1$), and the optimal choice of sick1 and sick2 patients is not consume (either *sell* or *do nothing*).



(a) sick0 patients

(b) sick1 patients

(c) sick2 patients

(d) sick3 patients

Notes: $\rho = 1$ when the patients can be prescribed, and $\rho = 0$ when the patients cannot be prescribed
Figure B1: payoffs of the four sickness level patients when choosing different choices $\alpha_i(\rho)$ conditional on knowing whether can be prescribed

B.4 Theoretical population health impact

Figure B2 below shows that, theoretically, the population health impact is smaller in SM_Q8 than in the other two cases when the threshold is not *sick*3. This makes more sense if we take price into consideration, as the scarcity of drugs flowing to the secondary market makes

the incentive to sell higher and should encourage more drug diversions and induce a lesser population health impact as compared to no quota cases.

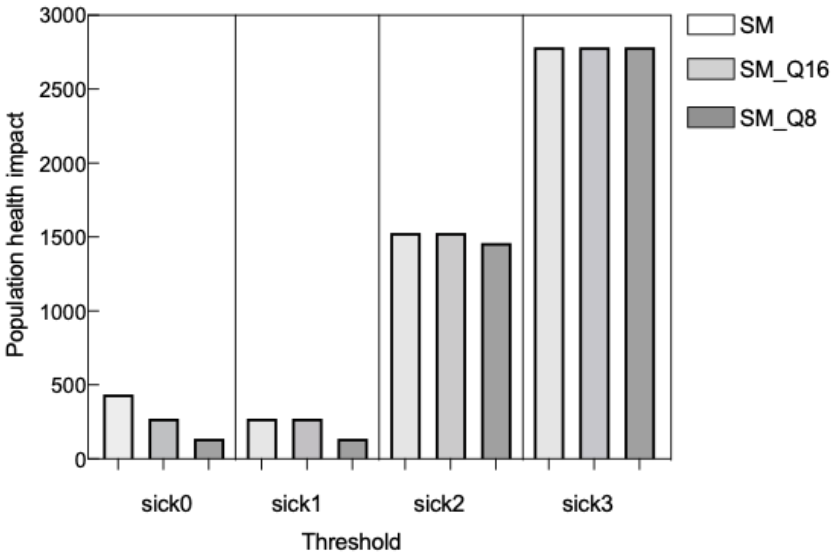
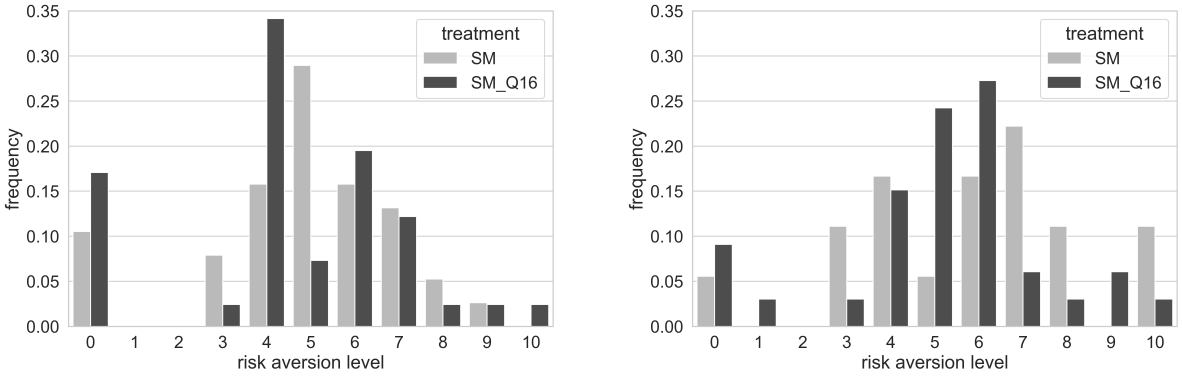


Figure B2: Theoretical population health impact under each threshold in the three cases

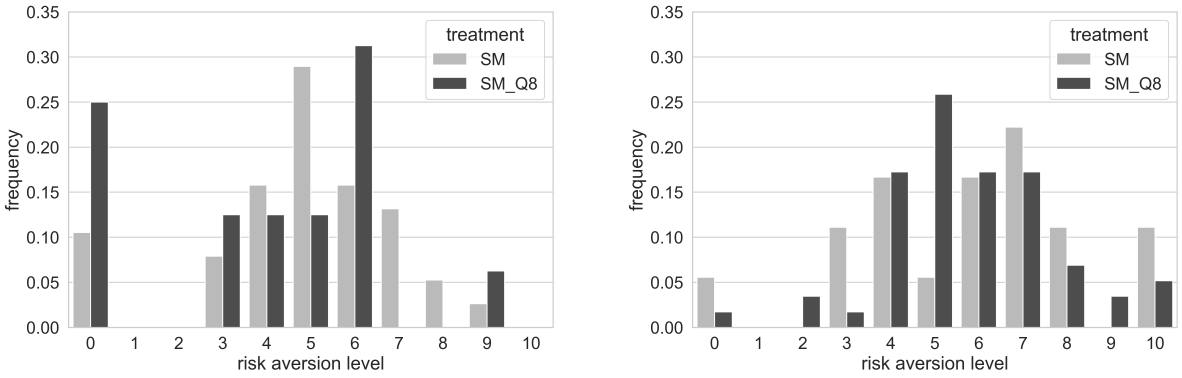
B.5 Risk attitudes distributions



(a)over-prescribers (b)non-over-prescribers

Notes: Risk aversion level is the number of non-risky choice(s) subjects made in the risk aversion task.

Figure B3: Distribution of (a) over-prescribers' and (b) non-over prescribers' risk aversion levels in SM and SM_Q16



(a)over-prescribers (b)non-over-prescribers

Notes: Risk aversion level is the number of non-risky choice(s) subjects made in the risk aversion task.

Figure B4: Distribution of (a) over-prescribers' and (b) non-over prescribers' risk aversion levels in SM and SM_Q8

References

Abouk, R. and D. Powell (2021). Can electronic prescribing mandates reduce opioid-related overdoses? *Economics & Human Biology 42*, 101000.

Alexeev, S. and D. Weatherburn (2022, 8). Fines for illicit drug use do not prevent future crime: evidence from randomly assigned judges. *Journal of Economic Behavior & Organization 200*, 555–575.

Alpert, A., W. N. Evans, E. M. Lieber, and D. Powell (2022, 4). Origins of the Opioid Crisis and its Enduring Impacts. *The Quarterly Journal of Economics 137*(2), 1139–1179.

Alpert, A., D. Powell, and R. L. Pacula (2018, 11). Supply-Side Drug Policy in the Presence of Substitutes: Evidence from the Introduction of Abuse-Deterrent Opioids. *American Economic Journal: Economic Policy 10*(4), 1–35.

Arora, A. and P. Bencsik (2021). Policing Substance Use: Chicago's Treatment Program for Narcotics Arrests.

Bardey, D., S. Kembou, and B. Ventelou (2021, 11). Physicians' incentives to adopt personalised medicine: Experimental evidence. *Journal of Economic Behavior & Organization 191*, 686–713.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017, 8). Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden. *American Economic Review 107*(8), 2204–2242.

Buchmueller, T. C. and C. Carey (2018). The effect of prescription drug monitoring programs on opioid utilization in medicare. *American Economic Journal: Economic Policy 10*(1), 77–112.

Catherine Maclean, J., J. Mallatt, C. J. Ruhm, K. I. Simon, and b. J. Christopher Ruhm Frank Batten (2022, 4). The Opioid Crisis, Health, Healthcare, and Crime: A Review Of Quasi-Experimental Economic Studies. *NBER Working Paper 29983*.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Deiana, C. and L. Giua (2021, 4). The Intended and Unintended Effects of Opioid Policies on Prescription Opioids and Crime. *B.E. Journal of Economic Analysis and Policy 21*(2), 751–792.

Deng, Y. and D. Houser (2022, 2). The Opioid Crisis and Secondary Markets: Evidence from a Laboratory Experiment. *SSRN Electronic Journal*.

Doleac, J. L. and A. Mukherjee (2019, 8). The Effects of Naloxone Access Laws on Opioid Abuse, Mortality, and Crime. *Opioid Abuse, and Crime*.

Drug Enforcement Administration (2021, 12). Established Aggregate Production Quotas for Schedule I and II Controlled Substances and Assessment of Annual Needs for the List I Chemicals Ephedrine, Pseudoephedrine, and Phenylpropanolamine for 2022. *Federal Register (The Daily Journal of the United States Government)*, 68513–68523.

Ellis, R. P. and T. G. McGuire (1986, 6). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics 5*(2), 129–151.

Enzinger, A. C., K. Ghosh, N. L. Keating, D. M. Cutler, M. B. Landrum, and A. A. Wright (2021). US Trends in Opioid Access Among Patients With Poor Prognosis Cancer Near the End-of-Life. *Journal of Clinical Oncology 39*(26), 2948–2958.

Essington, T. E. (2010, 1). Ecological indicators display reduced variation in North American catch share fisheries. *Proceedings of the National Academy of Sciences of the United States of America 107*(2), 754.

Gächter, S., E. J. Johnson, and A. Herrmann (2007). Individual-Level Loss Aversion in Riskless and Risky Choices.

Haider, A., Y. Qian, Z. Lu, S. Naqvi, A. Zhuang, A. Reddy, S. Dalal, J. Arthur, K. Tanco, R. Dev, J. Williams, J. Wu, D. Liu, and E. Bruera (2019, 6). Implications of the Parenteral Opioid Shortage for Prescription Patterns and Pain Control Among Hospitalized Patients With Cancer Referred to Palliative Care. *JAMA oncology 5*(6), 841–846.

Heal, G. and W. Schlenker (2008). Sustainable fisheries. *Nature 455*(7216), 1044–1045.

Holt, C. A. and S. K. Laury (2002). Risk Aversion and Incentive Effects. *American Economic Review 92*(5), 1644–1655.

Humphreys, K., C. L. Shover, C. M. Andrews, A. S. Bohnert, M. L. Brandeau, J. P. Caulkins, J. H. Chen, M. F. Cuéllar, Y. L. Hurd, D. N. Juurlink, H. K. Koh, E. E. Krebs, A. Lembke, S. C. Mackey, L. Larrimore Ouellette, B. Suffoletto, and C. Timko (2022, 2). Responding to the opioid crisis in North America and beyond: recommendations of the Stanford–Lancet Commission. *The Lancet 399*(10324), 555–604.

Iversen, T. and H. Lurås (2006). Capitation and Incentives in Primary Care. In *Chapter 25 in The Elgar Companion to Health Economics*. Edward Elgar Publishing.

Iwry, J. and M. A. Kleiman (2017, 10). A Nudge Toward Temperance: User-Set Consumption Limits as an Element of Cannabis Policy. *SSRN Electronic Journal*.

Linn, B. S., T. Mahvan, B. E. Y. Smith, A. B. Oung, H. Aschenbrenner, and J. M. Berg (2020, 7). Tips and tools for safe opioid prescribing. *MDedge Family Medicine 69*(6), 280–292.

Lusted, A., M. Roerecke, E. Goldner, J. Rehm, and B. Fischer (2013). Prevalence of pain among nonmedical prescription opioid users in substance use treatment populations: Systematic review and meta-analyses. *Pain Physician 16*(6), 671–684.

Maclean, C., J. Mallatt, C. J. Ruhm, and K. Simon (2020). Economics Studies on the Opioid Crisis: A Review. *NBER 6(11), 951–952. 6(11), 951–952.*, 5–24.

Mcguire, T. (2000). Chapter 9 – Physician Agency*. In *Handbook of Health Economics*, Volume 1, pp. 461–536.

Meinhofer, A. (2015). Prescription Drug Monitoring Programs: The Role of Asymmetric Information on Drug Availability and Abuse. *American Journal of Health Economics 27*(5), 976–980.

Meinhofer, A. and A. E. Witman (2018, 7). The role of health insurance on treatment for opioid use disorders: Evidence from the Affordable Care Act Medicaid expansion. *Journal of Health Economics 60*, 177–197.

Molitor, D. (2018, 2). The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration. *American Economic Journal: Economic Policy 10*(1), 326–356.

Mulligan, C. B. (2020). Prices and Federal Policies in Opioid Markets. *NBER Working Papers*.

Natividad, G. (2016, 3). Quotas, Productivity, and Prices: The Case of Anchovy Fishing. *Journal of Economics & Management Strategy 25*(1), 220–257.

Newell, R. G., J. N. Sanchirico, and S. Kerr (2005, 5). Fishing quota markets. *Journal of Environmental Economics and Management 49*(3), 437–462.

Phelps, C. E. (1992, 9). Diffusion of Information in Medical Care. *Journal of Economic Perspectives 6*(3), 23–42.

Phelps, C. E. (2000, 1). Information Diffusion and Best Practice Adoption. *Handbook of Health Economics 1*(PART A), 223–264.

Powell, D., R. L. Pacula, and E. Taylor (2020). How increasing medical access to opioids contributes to the opioid epidemic: Evidence from Medicare Part D. *Journal of Health Economics 71*, 102286.

Sacks, D. W., A. Hollingsworth, T. Nguyen, and K. Simon (2021, 3). Can policy affect initiation of addictive substance use? Evidence from opioid prescribing. *Journal of health economics 76*.

Schatman, M. E. and E. L. Wegrzyn (2020). The United States Drug Enforcement Administration and Prescription Opioid Production Quotas: An End Game of Eradication? *Journal of pain research 13*, 2629–2631.

Schnell, M. (2017). Physician Behavior in the Presence of a Secondary Market: The Case of Prescription Opioids. *Princeton University Working Paper*.

Schnell, M. and J. Currie (2018). Addressing the opioid epidemic: Is there a role for physician education? *American Journal of Health Economics 4*(3), 383–410.

Sullivan, M. D., M. J. Edlund, M. Y. Fan, A. Devries, J. Brennan Braden, and B. C. Martin (2010). Risks for possible and probable opioid misuse among recipients of chronic opioid therapy in commercial and medicaid insurance plans: The TROUP Study. *Pain 150*(2), 332–339.

Thombs, R. P., D. L. Thombs, A. K. Jorgenson, and T. H. Braswell (2020). What Is Driving the Drug Overdose Epidemic in the United States? *Journal of Health and Social Behavior 61*(September), 2467.