PHYSICIAN-INDUCED DEMAND:
EVIDENCE FROM CHINA'S DRUG PRICE ZERO-MARKUP POLICY

Hanming Fang
Xiaoyan Lei
Julie Shi
Xuejie Yi

ABSTRACT

We use a natural experiment in China's health care reform, in which the government mandated zero markups for drugs sold in public hospitals, to study physician-induced demand for medical care. We develop a theoretical model that robustly predicts that the drug price zero markup policy (ZMP) will lead to a reduction in the demand for drugs but an increase in the demand for non-drug services, only when the demands for drugs and non-drug services are at least partially physician induced. Our model also yields testable predictions of the impact of the ZMP regarding the equilibrium number of treated patients. Exploiting the staggered rollout of the ZMP and a unique claims data set, we find that, at the admission level, the reform decreases drug expenses by 63.4 log points (47.0 percent), but the reduction in drug expenses is almost fully substituted by the increases in expenses for non-drug services. Quantitatively similar results are also obtained at the physician and at the hospital levels. We also show that the ZMP reduces the number of patients in the treated hospitals, as predicted by the model, and that the reform has little impact on the quality of health care proxied by the readmission rate.

Hanming Fang
Department of Economics
Ronald O. Perelman Center
for Political Science and Economics
133 South 36th Street
Suite 150
Philadelphia, PA 19104
and NBER
hanming.fang@econ.upenn.edu

Xiaoyan Lei
CCER
National School of Development
Peking University
Beijing 100871, CHINA
xylei@nsd.pku.edu.cn

Julie Shi
School of Economics
Peking University
Beijing 100871, CHINA
jshi@pku.edu.cn

Xuejie Yi
Department of Economics
 Stanford University
Stanford, CA 94305
xuejieyi@stanford.edu

1 Introduction

Health care expenditures are growing rapidly in all countries around the world (World Health Organization 2018). A large empirical literature provides explanations for and policy recommendations to curb the rising health care expenditure. The literature can be divided into two strands. The first strand focuses on demand-side explanations, including the moral hazard in the demand for medical care resulting from third-party (i.e., insurance) payments. This strand includes the RAND Health Insurance Experiment (Newhouse et al. 1993) and the Oregon Health Insurance Experiment (Finkelstein et al. 2012).[1] The second strand focuses on supply-side behavior. For example, Finkelstein et al. (2016) use patient migrations to show that about half of the variation in medical expenditures across the regions in the United States was driven by physician behavior. An important literature in the second strand aims to document empirically how health care providers respond to financial incentives. The literature examines how changes in reimbursement rules, or medical prices more generally, affect physicians' (or providers') diagnostic, prescription, and treatment choices (see, e.g., Gruber and Owings 1996; Yip 1998; Dafny 2005; Clemens and Gottlieb 2014; Alexander 2020, among others).[2] Although the contexts of the studies differ in the sources of price changes or the margin of the physicians' behavioral response, the literature has found that, somewhat paradoxically, physicians tend to prescribe more services in response to *either* an increase *or* a decrease in the related reimbursement rates. The typical explanation is that when the price of a service is lowered, physicians will prescribe more of the service in order to make up for the lower unit price to maintain an income target; when the price of a service is raised, physicians will have an incentive to prescribe more of that service to increase his/her income. However, as is well known, empirical tests for models of physician-induced demand are remarkably challenging because it is difficult to separate the demand and the supply sides; overuse of medical procedures or drugs could also reflect consumer demand and/or physicians' belief about consumer demand (see Currie et al. 2014).

In this paper, we exploit a drug pricing reform in China, called *zero-markup policy* (ZMP), which had a staggered rollout starting in 2009, to provide evidence of physician induced demand. Prior to the ZMP, hospitals in China were allowed to mark up drug prices by up to 15 percent of the wholesale prices, and the profits from drug sales were commonly viewed as one of hospitals' key profit generators. The ZMP mandates that hospitals can only charge drug prices at their procurement cost with zero markup (i.e., no profit is allowed from drug sales). We first develop a theoretical model to analyze how the ZMP may affect physicians' choices of drugs and non-drug services when treating patients. The model robustly predicts that the ZMP will lead to a reduction in the demand for drugs but an increase in the demand for non-drug services, *only when* the demands for drugs and non-drug services are at least partially physician induced. Our model also yields additional testable predictions of the impacts of the ZMP on the equilibrium number of treated patients.

The ZMP was first rolled out in basic medical facilities, including township or community health centers;

---

[1] Hall and Jones (2007) explain the rising health spending as the rational response to the growth of income per person.

[2] A detailed but selective review of the related literature is provided below.

a few years later, it was implemented in higher-tier hospitals located in the county seats. We exploit the staggered rollout of the ZMP and a unique comprehensive claims data set to empirically test the predictions of the model. Using a difference-in-differences (DID) empirical strategy, we find that the drug expenses of township health centers (the treatment group) *dropped* by 63 log points (47 percent) compared with those of county hospitals (the control group); at the same time, we find that expenses for non-drug services were 28 log points (32 percent) *higher* in the township health centers than in the county hospitals. Interestingly, we find that although the expenses increased for all categories of non-drug services in the township health centers (the treatment group), the relative magnitude of the increases depends on how discretionary the non-drug services are: the increase is the largest (around 30 log points or 35 percent) for spending on tests and examinations, bed fees, and nursing services where physicians are likely to have more discretion on the amount to be ordered, while the increases are smaller (about 7 log points or 7 percent) for less discretionary medical materials. Overall, we find that the reduction in drug expenses is almost fully compensated by the increases in expenses for non-drug services, and thus the patients' total medical spending does not change significantly. We also find that the ZMP reduces the number of patients in the treated hospitals relative to that in the control hospitals. The empirical evidence is consistent with the model's prediction *only when* the demands for drugs and non-drug services are at least partially physician induced.

Our paper contributes to the literature on physician-induced demand in several ways. Relative to the existing literature, we analyze a unique empirical setting of the effect of the ZMP to test for physician-induced demand. The literature tends to find that, in general physicians prescribe more in response to a decrease in the price to maintain an income target. However, this chain of reasoning has some caveats. An increase in prescriptions following a decrease in price is also consistent with patient-driven demand, as long as the patient's out-of-pocket expense is positively linked to the price; thus, in general it is impossible to distinguish patient-driven demand from physician-induced demand. However, more prescriptions raise physicians' income *only if* each prescription results in a positive profit, a condition that is precisely violated in our case under the ZMP. Indeed, our theoretical model yields robust predictions of the impact of the ZMP on physicians' choices between drugs and non-drug services, and highlight the stark contrasts of the impact on ZMP with or without physician-induced demand: if drug demand is completely patient driven, it will never be lower following the ZMP; however, the demand for drugs may be lower only if it is at least partially physician induced. Moreover, we also examine the impact of the ZMP on the demand for non-drug services, and test additional predictions from the model.

Related Literature. First and foremost, our paper contributes to the literature on how providers may adjust their behaviors in response to financial incentives following price or income shocks, which is known as the "*physician-induced demand*" literature. Since the seminal contribution by Arrow (1963), it has been theoretically and empirically recognized that a central feature of the health care market is the agency relationship between physicians and their patients, where physicians may exploit this agency relationship by providing excessive care to maintain their incomes. McGuire (2000) provides a comprehensive review of the literature on physician agency. Yip (1998) uses the Medicare fee reduction for coronary artery bypass grafting

(CABG) to test for physician-induced demand, and finds that physicians whose incomes were reduced the most by the Medicare fee cuts performed higher volumes of CABGs. The estimates suggest that thoracic surgeons recoup about 70 percent of their revenue loss by increasing the volume and intensity of CABGs. Gruber and Owings (1996) use the cross-state variations in the decline in fertility in the United States in the 1970s as a plausibly exogenous change in the financial environment facing obstetrician/gynecologists to test for physician-induced demand for more profitable cesarean deliveries. They find that a 10 percent decline in the fertility rate is associated with a 0.97 percentage point increase in the cesarean delivery rate. Dafny (2005) studies hospital responses to changes in diagnosis-specific prices by exploiting a 1988 policy reform that generated large price changes for 43 percent of Medicare admissions. She finds that hospitals, particularly for-profit hospitals, responded primarily by upcoding patients to diagnosis codes with the largest price increases. Similarly, Clemens and Gottlieb (2014) study the impact of Medicare's area-specific shocks on payment rates for physician services in 1997. They find that areas with higher payment shocks experience significant increases in health care supply; specifically, a 2 percent increase in payment rates led to a 5 percent increase in care provision per patient. The effects were twice as strong in elective procedures, such as cataract surgery, as in less discretionary services, like dialysis. Fang and Gong (2017) study potential Medicare overbilling based on the hours worked implied by the service codes physicians submit to Medicare, and find suggestive evidence that the coding patterns of the flagged physicians seem to be responsive to financial incentives: within code clusters that have different levels of service intensity, the physicians tend to submit more higher-intensity service codes, compared with the unflagged physicians. Geruso and Layton (2020) exploit large and geographically heterogeneous increases in Medicare Advantage (MA) enrollment and variations in the financial incentives for physicians within county markets to identify upcoding. They find that a 10 percentage point increase in MA penetration leads to a 0.64 percentage point increase in the reported average risk score in a county. Alexander (2020) studies the effect of a pilot payment reform program in which hospitals paid doctors bonuses for reducing the total costs of treatment for each Medicare admission. She identifies the effects of the bonus program by comparing the behavior of a given doctor who works at multiple hospitals, some of which participate in the program and others that do not. She finds that doctors respond to the bonuses by changing the composition of admitted patients and by sorting healthier patients to participating hospitals.

Our paper contributes to the above literature by analyzing the ZMP, which is a unique policy experiment. The ZMP essentially renders a previously "profitable" (for physicians and hospitals) segment of the heath care sector completely unprofitable. This results in clean theoretical predictions of physicians' health care choices with and without financial incentives. We highlight these sharp theoretical predictions through the detailed analysis of a theoretical model. In this regard, our analysis complements the literature on physician-induced demand.

Our paper is also related to a small literature on the role of financial incentives in physicians' drug prescription behavior. In the United States, pharmacies instead of hospitals dispense drugs; thus, hospitals and physicians can only benefit from prescribing drugs indirectly if they receive payments from pharmaceutical companies (see, e.g., Grennan et al. 2020). The literature on the effects of physicians' financial incentives

on drug prescription behavior mostly appears in contexts outside the United States. For example, Iizuka (2012) shows that in Japan, where physicians can legally make profits by prescribing and dispensing drugs, physicians' financial incentives can explain why cheaper generics are infrequently adopted. Similarly, Liu et al. (2009) study how physicians' financial incentives affect their prescription decision on the choice of generic versus brand-name drugs in Taiwan, where physicians prescribe and dispense drugs. There are also several studies on China. Currie et al. (2014) conduct an audit study of physicians' decisions to prescribe antibiotics by varying physicians' financial incentives, beliefs about what patients want, and desires to reciprocate for a small gift. Their findings suggest that antibiotics abuse in China is largely driven by physicians' financial incentives. Lu (2014) conducts a field audit experiment in which patients' ability to pay and intention to purchase drugs from the doctors' hospitals were randomly assigned. Their findings also suggest the role of physicians' financial incentives in their prescription behavior. Wu (2019) studies the impact of a reform called the "drug percentage incentive scheme," which caps the percentage of revenue that a hospital can receive from drug sales at 45 percent for all tertiary hospitals, on physician behavior. She finds that under this policy the physicians increased non-drug expenditures and decreased drug expenditures, keeping the total expenditures almost unchanged.

Our paper extends our understanding of the health care system in China, and we believe that the results are also relevant for other developing countries where controlling health expenditures is likely to be an important policy objective. Various health care reforms were implemented in China in recent decades; however, detailed evaluations of the effects of the reforms, particularly the ZMP, are scarce.[3] Two closely related papers are Yi et al. (2015) and Chen et al. (2017). Using survey data on 43 township health centers, Yi et al. (2015) exploit the variations in township health centers' reliance on revenue from drug sales, and investigate whether the centers that relied more on revenue from drug sales under the ZMP had greater reductions in revenue share from drug sales and greater increases in the supply of non-drug services, compared with centers that relied less on revenue from drug sales. Chen et al. (2017) empirically examine Shandong province's initial implementation of China's 2009 Essential Medications List policy. Using a panel of 1,466 patients with hypertension and/or diabetes, they find that the policy reduced drug markups by providers, but the lost drug revenue was compensated by the increased office visits, keeping the total medical expenditure unchanged. We contribute to this literature by presenting a theoretical framework of physician-induced demand, and deriving and testing the theoretical predictions of the effects of the ZMP. We also employ a cleaner difference-in-differences identification strategy on a rich claims-level data set.

The remainder of the paper is structured as follows. In Section 2, we provide the institutional background related to the ZMP in China; in Section 3 we present a theoretical model that illustrates how physician-induced demand can be inferred from the changes in physicians' choice of drugs and non-drug services in response to the effects of the ZMP; in Section 4 we describe the data and present the summary statistics; in Section 5 we present our empirical results on the impacts of the ZMP regarding drug, non-drug, and total expenditures, as well as the quantities of drugs and patients' health outcomes; finally, in Section 6 we

---

[3]See Wagstaff et al. 2009; Yip et al. 2012; Yip et al. 2019; and Burns and Liu 2017, among others, for an extensive review of China's health care reforms.

conclude.

2   Institutional Background

Public hospitals in China deliver more than 90% of the country's inpatient and outpatient services (Yip et al., 2012). The prices of the medical services in the public hospitals are regulated by the government. Physicians are employed by the public hospitals to provide diagnostic services, conduct tests, and dispense drugs to patients. It is commonly known that the price schedule was distorted, in particular, labor-related services were heavily under-priced;[4] in contrast, hospitals were allowed to charge up to a 15 percent markup over the drug procurement prices when selling drugs to patients. As the physicians are employed by hospitals, their income (including bonuses) is highly dependent on hospitals' revenues. Although hospitals are public facilities, they received very limited government funding. This led to severe over-prescription of drugs in health care sector in China.[5]

Although pharmacies are common in China, public hospitals account for around 80% of drug sales in China. Prescription drugs are usually purchased in hospitals instead of from pharmacies for two reasons. First, physicians encourage patients to fill prescriptions at the hospitals in order to increase hospital revenue. Second, some pharmacies are not included in the insurance network from which prescriptions purchased are eligible for reimbursement. For our analysis, we focus on inpatient visits, for which the medications are always directly filled in the hospitals.

China started to introduce universal national health insurance to its citizens, called the New Rural Cooperative Medical Scheme (NRCMS) for rural residents and the Urban Resident Basic Medical Insurance (URBMI) for urban residents in 2003 and 2007, respectively, in addition to the original Urban Employee Medical Insurance (UEMI). To a large extent, the high drug expenses incurred in Chinese hospitals became the financial burden of the national insurance system. To curb the unsustainable drug expenses and to improve access to affordable, safe, and effective drugs, the Chinese government introduced an important drug reform in 2009. The reform has three components: the most important component is called the *zero-markup policy* (ZMP), and it is accompanied by two supplement policies, the *Essential Medicines List* (EML) and *centralized drug procurement*. We now describe each component in more detail.

Zero-Markup Policy (ZMP).   The most important component of the drug reform initiated in 2009 is the ZMP, which prohibits hospitals from charging any markup on drug price, i.e., all drugs must be sold at the procurement price with zero markup. In contrast, prior to the reform, hospitals were allowed to charge up to 15 percent markup over the procurement price when selling drugs, which introduced perverse incentives

---

[4]For example, the diagnostic fee in public hospitals was only 5 to 12 RMB, or about $0.8 to $1.8, according to the exchange rate of 6.8 RMB for 1 US dollar in 2010.

[5]For example, Li et al. (2012) find that antibiotics were prescribed about twice as frequently in China as that recommended by the World Health Organization. See also Wagstaff et al. (2009) and Yip et al. (2012) for descriptions of these problems.

for hospitals both in the number of drug prescriptions and in the negotiation of procurement prices.[6] The ZMP has a significant effect on hospitals' revenues because, prior to the ZMP, a large fraction of hospitals' revenue came from the profit from selling drugs.[7]

National Essential Medicine List (EML).   The 2009 drug reform also issued the national EML, which is a list of 307 types of medicines that were initially selected by the central government based on financial cost, safety, clinical efficacy, and availability. In addition, provincial governments can extend the list according to the medical needs of the local population. The drugs on the EML can be updated once every three years. Hospitals and physicians must prioritize the use of the drugs on the EML. By contrast, prior to the reform, hospitals and physicians could freely determine which drugs to prescribe to their patients. The government provides guidelines on the share of the essential drugs in the total drugs used for different levels of hospitals: the share is 100 percent for township hospitals in rural areas and community hospitals in urban regions, 40 percent for county hospitals, and 25-30 percent for comprehensive hospitals in cities. These EML restrictions were not binding for the hospitals in our sample.

Provincial Centralized Procurement of Drugs.   In the 2009 drug reform, the provincial centralized procurement changed the way drugs are procured. Before the reform, wholesale drug prices were negotiated between hospitals and pharmaceutical firms; thus, the wholesale price for the same drug could vary across different hospitals. The 2009 drug reform requires each province to establish a central platform where all the essential medicines are purchased at prices determined by the bidding system in the platform, thus equalizing the wholesale prices across hospitals for the same drug (in terms of brand/producer, packages, and dosage units).[8] Provincial centralized procurement authorizes the winning bidders to sell essential drugs to the hospitals. But it is up to the hospitals' discretion on which brand to purchase as there are usually multiple winning producers for the same drug. Therefore, the price dispersion within each molecule still exists after the reform.

Time Line of the Drug Reform.   The Chinese central government issued the drug reform policy in August 2009. The provincial governments then established their detailed plans for local implementation of the drug reform. It was decided that the ZMP would be rolled out in a staggered fashion. First, it would apply to the basic medical facilities, namely, the village clinics, township health centers in rural areas, and community health centers in urban areas. In addition, one-third of the counties would pilot the reform from November 2009, and the remaining counties would enact the ZMP from April 2010. Second, the ZMP would gradually extend to upper-level hospitals. From 2012, it would apply to the county hospitals; and beginning

---

[6]It was noted that the markup allowance before the ZMP led the hospitals to prefer higher priced formulary equivalents because they offered larger profit opportunities.

[7]According to the National Health Ministry of China (2011), the profit from drug sales accounted for 51 percent of community hospitals' total revenue in 2009.

[8]The platform set bidding prices no higher than the wholesale price in the past three years. More information could be found at http://www.gov.cn/gongbao/content/2010/content_1765276.htm.

in 2016, all public hospitals were required to implement the ZMP, including the most comprehensive general hospitals. As public hospitals are the major provider of health care services in China, the ZMP has influenced almost the entire Chinese population.

Study Region. In this paper we use data from a county in a western Chinese province. The sample county had a population size of 422,000 in 2010, which was slightly larger than a median size county (with population of 380,000 according to the 2010 Chinese Population Census). The county's per capita income in 2009 for rural residents was 7,324 RMB, slightly higher than the national average of 5,919 RMB (National Bureau of Statistics of China, 2010). Overall our sample county is comparable to a median county in China in terms of population demographics and other social and economic characteristics.

The sample county implemented the ZMP in a staggered rollout: from October 1, 2010, all of the basic medical facilities implemented the ZMP; however, the county hospitals were not required to do so until October 2013. That is, in the period between October 1, 2010 and September 30, 2013, the county hospitals could still charge up to 15% markup from their drug sale while the basic medical facilities were subject to the ZMP.

Both the basic medical facilities and the county hospitals in our sample county began to purchase from the central procurement platform since the implementation of ZMP in the township centers. Our model would take the potential change in wholesale price into consideration. Besides ZMP, the basic medical hospitals were also required to prescribe 100 percent of the drugs from the EML, while county hospitals were not restricted to only use the drugs on the EML. As will be discussed later, the EML prescription regulation did not bind the county hospitals' prescription choice at all and only had a small impact on the increasing share of EML in township hospitals. Thus, our results are mainly affected by the ZMP rather than the EML prescription rule.

3 Model

In this section, we present a model of physician behavior to illustrate how the ZMP can be an ideal policy experiment to test for physician-induced demand.

A physician observes a patient whose vector of characteristics $\mathbf{X}_i$ is summarized by index $\theta \in (0,1)$, where a higher value of θ indicates a more severely ill patient. The physician makes quantity choices on two items that may improve the patient's health, drugs and non-drug medical services. In what follows, we denote the quantity of drugs by d, and the quantity of non-drug services by s. Suppose that the health of the patient is given by $H(d, s; \theta)$ and we assume that $H(\cdot, \cdot; \theta)$ has the following properties:

Assumption 1 *For all $\theta \in (0,1)$, $H(\cdot, \cdot; \theta)$ is concave and satisfies*

$$
\begin{aligned}
\lim_{d \to 0} H_1(d, s; \theta) &= +\infty; & \lim_{s \to 0} H_2(d, s; \theta) &= +\infty; \\
\lim_{d \to +\infty} H_1(d, s; \theta) &< 0; & \lim_{s \to +\infty} H_2(d, s; \theta) &< 0.
\end{aligned}
$$

Assumption 2 (Drugs and non-drug services are substitutes) *For all $\theta \in (0,1)$, $H_{12} < 0$.*

Assumption 1 consists of two parts. The concavity assumption ensures that the first-order conditions for the maximization problem below are sufficient; the "Inada-type" conditions at 0 and $+\infty$ ensure that the optimal solution will be interior. Assumption 2 says that drugs and non-drug services are substitutes. This assumption may seem strong, as in some cases drugs and non-drug services may be complements. For example, the service itself may require some drugs to be successful. However, this assumption is without loss of generality: if some minimum level of drug use, say $\underline{d}$, and some service, say $\underline{s}$, are essential for a patient's recovery, we can interpret d and s as levels of drugs and non-drug services that exceed $(\underline{d}, \underline{s})$.

Let r_d^h and r_s^h denote the *retail* prices of drugs and non-drug services charged by the hospital. Let c denote the *wholesale* cost of drugs paid by the hospital. Chinese hospitals were allowed to charge a high markup over the wholesale cost c before the ZMP, that is,

$$r_d^h > c.$$

Suppose that, as a normalization, non-drug services are costly for the physician in terms of effort cost, $e(s)$, but otherwise they do not have a monetary cost.[9]

The patient's *out-of-pocket* prices for drugs and non-drug services are, respectively, r_d^p and r_s^p. Below we will consider several cases of the relationship between patients' out-of-pocket price and hospital prices.

We assume that the physician's maximization problem is given by

$$V(\theta) \equiv \max_{\{d,s\}} \overbrace{[H(d,s;\theta) - r_d^p d - r_s^p s]}^{\text{Patient Surplus}} + \alpha \left[\overbrace{(r_d^h - c)d + r_s^h s}^{\text{Physician Net Income}} \right] - \overbrace{e(s)}^{\text{Effort Cost of Service}}, \tag{1}$$

where $\alpha \geq 0$ represents the weight of the physician's net income in the choice of (d, s) for the patient.

Remark 1 *There are multiple interpretations of $\alpha \geq 0$ in Problem (1). First, although the physicians in Chinese hospitals are salaried workers, they may receive a bonus that is tied to the net revenues they generate for the hospital (Yip and Hsiao, 2008). Second, even if the physicians are not directly paid a bonus, the hospital may have an income target that needs to be met by all the employed physicians since the local government only provides a limited subsidy toward the hospitals' regular operation needs. If the hospital's income target is a constraint on the physicians' choices of drugs and non-drug services for the patients, a positive α may reflect the Lagrangian multiplier of the hospital's income target constraint.*

Definition 1 *We say that the demands for drugs and non-drug services are **completely patient driven** when $\alpha = 0$.*

Definition 2 *We say that the demands for drugs and non-drug services are **partially physician induced** when $\alpha > 0$.*

---

[9]For example, to prescribe a test, the physician may need to spend more time with the patient to wait for and check the results. We can also think of such effort cost as the opportunity cost, since providing some patients with non-drug services would crowd out other patients from getting similar services.

For a given $\alpha \geq 0$, the first-order conditions for the optimal d and s are given by

$$H_1(d, s; \theta) = r_d^p - \alpha(r_d^h - c), \tag{2}$$

$$H_2(d, s; \theta) - e'(s) = r_s^p - \alpha r_s^h. \tag{3}$$

Definition 3 *A **zero-markup policy** (ZMP) for drug prices is a policy that requires the hospital to set* $\widehat{r_d^h} = c$.

We are interested in analyzing how the ZMP affects the quantity choices of d and s in Problem (1). For now, fix the procurement wholesale price c. The ZMP reduces the hospital's drug prices from $r_d^h > c$ before the policy change to $\widehat{r_d^h} = c$ after the policy change. However, the ZMP may also change the patient's out-of-pocket price r_d^p, if r_d^p is tied to the hospital price via the co-insurance rate, for example. We will consider different cases of how r_d^p and r_d^h are related. There are two interesting cases to consider:

Case 1 (Constant Co-Pay) *The patient's out-of-pocket price r_d^p is a constant:* $r_d^p = r_d \geq 0$.

Case 2 (Constant Coinsurance) *The patient's out-of-pocket price r_d^p is proportional to the drug price charged by the hospital,* $r_d^h : r_d^p = \gamma r_d^h$, *where* $\gamma \in (0, 1]$ *is the co-insurance rate.*

Note that Case 1 covers the case if the patient has zero co-pay, which will be the case when $r_d = 0$; Case 2 covers the case where the patient is uninsured, which will be the case when $\gamma = 1$.[10]

How the physician's quantity choices for drugs and non-drug services respond to the ZMP is a comparative statics question for the maximization problem (1). We can express the first-order conditions (2)-(3) for maximization problem (1) before and after the ZMP as:

$$H_1(d, s; \theta) = \beta_d(\alpha), \tag{4}$$

$$H_2(d, s; \theta) - e'(s) = \beta_s(\alpha) \tag{5}$$

where $\beta_d(\alpha)$ and $\beta_s(\alpha)$ are parameters that multiply d and s, respectively in maximization problem (refP). Specifically, using subscript 0 and 1 to denote the pre-ZMP and the post-ZMP, respectively, the first-order conditions before and after the ZMP can be written as:

$$\beta_{d0}(\alpha) = r_d^p - \alpha(r_d^h - c) \tag{6a}$$

$$\beta_{d1}(\alpha) = \widehat{r_d^p} \tag{6b}$$

$$\beta_{s0}(\alpha) = \beta_{s1}(\alpha) = r_s^p - \alpha r_s^h \tag{6c}$$

where $\widehat{r_d^p}$ is the patient's out-of-pocket price after the ZMP, whose value will depend on whether it is Case (1) or Case (2). Since the ZMP does not change the prices of non-drug services, the comparative statics is with respect to β_d only.

---

[10]In China, patients need to pay out of pocket for their inpatient expenditures until a deductible is reached, after which there is a constant coinsurance rate. Therefore, the actual situation might lie between Case 1 and Case 2.

Applying the Implicit Function Theorem to the first order conditions as β_d changes before and after the ZMP, we have

$$\frac{\partial d^*}{\partial \beta_d} = \frac{H_{22} - e''}{H_{11}(H_{22} - e'') - H_{12}^2} < 0, \tag{7}$$

$$\frac{\partial s^*}{\partial \beta_d} = -\frac{H_{12}}{H_{11}(H_{22} - e'') - H_{12}^2} > 0. \tag{8}$$

Proposition 1 *If the demands for drugs and non-drug services are completely patient driven, i.e., if $\alpha = 0$, then ZMP will*

- *in Case (1), have no effect on the demand for drugs and non-drug services;*

- *in Case (2), increase the demand for drugs and decrease the demand for non-drug services.*

Proof. Let $\alpha = 0$. First, consider Case (1). The parameters of the maximization problem before and after the ZMP do not change:

$$\beta_{d0}(0) = \beta_{d1}(0) = r_d,$$
$$\beta_{s0}(0) = \beta_{s1}(0) = r_s^p.$$

Thus, the ZMP has no effect: $d_0^*(\beta_{d0}(0);\theta) = d_1^*(\beta_{d1}(0);\theta)$ and $s_0^*(\beta_{d0}(0);\theta) = s_1^*(\beta_{d1}(0);\theta)$.

Next, consider Case (2). The parameters of the maximization problem before and after the ZMP are

$$\beta_{d0}(0) = \gamma r_d^h$$
$$\beta_{d1}(0) = \gamma c.$$

Since $r_d^h > c$, we have $\gamma r_d^h > \gamma c$. Thus, the ZMP leads to a decrease in $\beta_d : \Delta \beta_d < 0$. Applying $\Delta \beta_d < 0$ to (7) and (8), we obtain that in Case (2), $d_1^*(\beta_{d1}(0);\theta) > d_0^*(\beta_{d0}(0);\theta)$ and $s_1^*(\beta_{d1}(0);\theta) < s_0^*(\beta_{d0}(0);\theta)$. ∎

Proposition 2 *If the demands for drugs and non-drug services are partially physician induced, i.e., if $\alpha > 0$, then the ZMP will*

- *in Case (1), decrease the demand for drugs, and increase the demand for non-drug services;*

- *in Case (2), decrease the demand for drugs, and increase the demand for non-drug services if and only if $\alpha > \gamma$.*

Proof. Now let $\alpha > 0$. In Case (1), we have

$$\beta_{d0}(\alpha) = r_d - \alpha(r_d^h - c)$$
$$\beta_{d1}(\alpha) = r_d,$$

thus, the ZMP leads to an increase in $\beta_d : \Delta \beta_d \equiv \beta_{d1}(\alpha) - \beta_{d0}(\alpha) = \alpha(r_d^h - c) > 0$. Applying $\Delta \beta_d > 0$ to (7) and (8), we obtain that in Case (1), $d_1^*(\beta_{d1}(0);\theta) < d_0^*(\beta_{d0}(0);\theta)$ and $s_1^*(\beta_{d1}(0);\theta) > s_0^*(\beta_{d0}(0);\theta)$.

10

In Case (2), we have

$$\begin{aligned}
\beta_{d0}(\alpha) &= \gamma r_d^h - \alpha\left(r_d^h - c\right) \\
\beta_{d1}(\alpha) &= \gamma c.
\end{aligned}$$

Hence

$$\Delta\beta_d = \gamma c - \left[\gamma r_d^h - \alpha\left(r_d^h - c\right)\right] = (\alpha - \gamma)\left(r_d^h - c\right).$$

Since $r_d^h > c$, we have $\Delta\beta_d > 0$ if and only if $\alpha > \gamma$. The claim follows by applying this result to (7) and (8).
■

Propositions 1 and 2 clearly indicate that a positive α is a *necessary* condition for the demand for drugs to decrease and the demand for non-drug services to increase, following the ZMP. Therefore, empirical evidence that the demand for drugs decreases and the demand for non-drug services increases following the ZMP suggests that the demands are partially physician induced, according to Definition 2. Proposition 2 indicates that in Case (2), α must be larger than the patient's coinsurance rate γ for the ZMP to reduce the demand for drugs and increase the demand for non-drug services. If and only if the physician cares more about his/her own net income than about the patient's surplus, that is, if and only if $\alpha > \gamma$, the ZMP could decrease the quantity demanded for drugs and increase the quantity demanded for non-drug services.

For patients who do not have insurance, that is, those with $\gamma = 1$, it must be that $\alpha > 1$; that is, the physician must put a higher weight on the implications for his/her net income of the choices of d and s than that on patient surplus. Furthermore, to the extent that patients have health insurance that reduces γ to be less than 1, for any physician payoff function, it is more likely for the condition $\alpha > \gamma$ to be satisfied and, according to Proposition 2, for the ZMP to induce lower demand for drugs and higher demand for non-drug services.

3.1 Endogenous Wholesale Price

Propositions 1 and 2 are robust to the possibility that wholesale price c may change due to the establishment of the procurement platform after the ZMP. In this subsection, we suppose that the new wholesale price after the introduction of the ZMP is $\hat{c}$.

It is possible that the wholesale price $\hat{c}$ after the ZMP will be higher than c, due to potential kickback schemes between the procurer and the drug company, for example. However, there are scenarios under which the new wholesale price $\hat{c}$ could be lower than c. For example, drug procurement was moved to the provincial level; this could potentially increase the bargaining position of the government procurement and result in a lower wholesale price.

We have the following result:

Proposition 3 *If the wholesale drug price responds to the ZMP, then*

1. *When $\alpha = 0$, the ZMP will*

(a) in Case (1), have no effect on the demand for drugs and non-drug services;

(b) in Case (2), increase the demand for drugs and decrease the demand for non-drug services if $\hat{c} > r_d^h$.

2. When $\alpha > 0$, the ZMP will

(a) in Case (1), decrease the demand for drugs, and increase the demand for non-drug services.

(b) in Case (2), decrease the demand for drugs, and increase the demand for non-drug services if and only if $\alpha/\gamma > \left(r_d^h - \hat{c}\right) / \left(r_d^h - c\right)$. Moreover, the higher $\hat{c}$ is, the larger the reduction in the demand for drugs and the rise in the demand for non-drug services.

Proof. Claim 1(a) follows from applying (7) and (8), noting that in Case (1), $\beta_{d0}(0) = \beta_{d1}(0) = r_d$; thus $\Delta\beta_d = 0$. Similarly, Claim 1(b) follows from the fact that $\beta_{d1}(0) = \gamma\hat{c} > \beta_{d0}(0) = \gamma r_d^h$ and thus $\Delta\beta_d > 0$ if $\hat{c} > r_d^h$.

Claim 2(a) follows from the fact that $\beta_{d1}(\alpha) = r_d > \beta_{d0}(\alpha) = r_d - \alpha\left(r_d^h - c\right)$. For Claim 2(b), note that

$$
\begin{aligned}
\beta_{d0}(\alpha) &= \gamma r_d^h - \alpha\left(r_d^h - c\right) \\
\beta_{d1}(\alpha) &= \gamma\hat{c}.
\end{aligned}
$$

Thus,

$$\Delta\beta_d = \gamma\hat{c} - \left[\gamma r_d^h - \alpha\left(r_d^h - c\right)\right] = (\alpha - \gamma)\left(r_d^h - c\right) + \gamma\left(\hat{c} - c\right).$$

Simple algebra shows that $\Delta\beta_d > 0$ if and only if

$$\frac{\alpha}{\gamma} > \frac{r_d^h - \hat{c}}{r_d^h - c}.$$

In addition, note that $\Delta\beta_d$ is increasing in $\hat{c}$. Hence, following (7) and (8), the higher $\hat{c}$ is, the larger is the reduction in the demand for drugs and the rise in the demand for non-drug services. Claim 2(b) follows. ∎

3.2 Discussion

In this subsection, we discuss some additional potential implications of the ZMP. We will first consider the implications for the set of treated patients. Considering maximization problem (1), the set of treated patients can be described as:

$$\Theta(\boldsymbol{\beta}(\alpha)) \equiv \{\theta : V(\theta; \boldsymbol{\beta}(\alpha)) \geq 0\}$$

where $\boldsymbol{\beta}(\alpha) \equiv (\beta_d(\alpha), \beta_s(\alpha))$, as described by Eq. ($6$). The envelope condition yields the following result:

$$\frac{\partial V(\theta; \boldsymbol{\beta}(\alpha))}{\partial\beta_d} = -d^* < 0. \tag{9}$$

Thus, if the ZMP results in a positive change in β_d, we would expect that the set of treated patients will shrink, at least weakly. Combining Eq. (9) with our previous analysis, we have the following prediction:

Proposition 4 *Suppose the wholesale drug price responds to the ZMP. Then the ZMP will affect the set of treated patients as follows:*

1. *When $\alpha = 0$, the ZMP will*

   (a) *in Case (1), have no effect on the set of treated patients Θ;*

   (b) *in Case (2), expand the set of treated patients Θ if $\hat{c} > r_d^h$.*

2. *when $\alpha > 0$, the ZMP will:*

   (a) *in Case (1), shrink the set of treated patients Θ;*

   (b) *in Case (2), shrink the set of treated patients Θ if and only if $\alpha/\gamma > \left(r_d^h - \hat{c}\right) / \left(r_d^h - c\right).$*

Proof. The results follow from Eq. (9) and the characterized values of $\Delta\beta_d$ for different cases as described in the proof of Proposition 3. ∎

So far, we have assumed that the weight α in the physician's maximization problem (1) stays constant, and our interpretation of α is that it is a measure of the physician's financial incentives. To the extent that the ZMP may affect the physician's (or the hospital's) ability to raise revenues, α may be affected by the ZMP. We now consider this possibility. The physician's maximization problem after the ZMP is now:

$$\hat{V}(\theta) \equiv \max_{\{d,s\}} \overbrace{\left[H(d,s;\theta) - \hat{r}_d^p d - r_s^p s\right]}^{\text{Patient Surplus}} + \hat{\alpha} \left[\overbrace{\underbrace{\left(\hat{r}_d^h - \hat{c}\right)}_{\equiv 0 \text{ due to ZMP}} d + r_s^h s}^{\text{Physician Net Income}} \right] - \overbrace{e(s)}^{\text{Effort Cost of Service}}, \tag{10}$$

where $\hat{\alpha}$ denotes the new weight on the net income. We make the following assumption to reflect the fact that the ZMP intensifies the revenue needs of the physician (or the hospital):

Assumption 3 $\hat{\alpha} > \alpha.$

The first-order conditions for the optimal solutions to maximization problem (10) are similar to those for Problem (1), and they can be written as

$$H_1(d,s;\theta) = \beta_d(\hat{\alpha}), \tag{11}$$

$$H_2(d,s;\theta) - e'(s) = \beta_s(\hat{\alpha}) \tag{12}$$

where $\beta_d(\hat{\alpha})$ and $\beta_s(\hat{\alpha})$ are parameters that multiply d and s, respectively, in Problem (10); specifically, they are given by:

$$\beta_{d1} = \beta_d(\hat{\alpha}) = \widehat{r_d^p} \tag{13a}$$

$$\beta_{s1} = \beta_s(\hat{\alpha}) = r_s^p - \hat{\alpha} r_s^h. \tag{13b}$$

Comparing (13) and (6), we see that the possible changes in α after the ZMP will now lead to a negative change in β_s :

$$\Delta\beta_s = \beta_{s1} - \beta_{s0} = \beta_s(\hat{\alpha}) - \beta_s(\alpha)$$
$$= (\alpha - \hat{\alpha}) r_s^h < 0 \tag{14}$$

13

where the inequality follows from Assumption 3. As a result, the total impacts of the ZMP on the demands for drugs and non-drug services are given by:

$$\Delta d^* = \frac{\partial d^*}{\partial \beta_d} \times \Delta \beta_d + \frac{\partial d^*}{\partial \beta_s} \times \Delta \beta_s$$

$$= \overbrace{\frac{H_{22} - e''}{H_{11}(H_{22} - e'') - H_{12}^2}}^{<0} \times \Delta \beta_d + \overbrace{\left[-\frac{H_{12}}{H_{11}(H_{22} - e'') - H_{12}^2} \right]}^{>0} \times \Delta \beta_s, \qquad (15)$$

$$\Delta s^* = \frac{\partial s^*}{\partial \beta_d} \times \Delta \beta_d + \frac{\partial s^*}{\partial \beta_s} \times \Delta \beta_s$$

$$= \overbrace{\left[-\frac{H_{12}}{H_{11}(H_{22} - e'') - H_{12}^2} \right]}^{>0} \times \Delta \beta_d + \overbrace{\frac{H_{11}}{H_{11}(H_{22} - e'') - H_{12}^2}}^{<0} \times \Delta \beta_s. \qquad (16)$$

Proposition 5 *Suppose that the ZMP changes the wholesale drug price from c to $\hat{c}$ and the weight on net revenues from α to $\hat{\alpha} > \alpha$. Then,*

1. *in Case (1), it will decrease the demand for drugs, and increase the demand for non-drug services.*

2. *in Case (2), it will decrease the demand for drugs, and increase the demand for non-drug services if $\alpha/\gamma > \left(r_d^h - \hat{c} \right) / \left(r_d^h - c \right)$. Moreover, the higher $\hat{c}$ is, the larger is the reduction in the demand for drugs and the larger is the rise in the demand for non-drug services.*

Proof. Claim 1 follows from the fact that, in Case (1),

$$\Delta \beta_d = \beta_{d1}(\hat{\alpha}) - \beta_{d0}(\alpha) = r_d - \left[r_d - \alpha \left(r_d^h - c \right) \right] > 0,$$

and $\Delta \beta_s < 0$, as seen from (14).

For Claim 2, the same calculations in the proof of Proposition 3 yield that $\Delta \beta_d > 0$ if and only if $\alpha/\gamma > \left(r_d^h - \hat{c} \right) / \left(r_d^h - c \right)$. Applying this, together with the fact that $\Delta \beta_s < 0$ (as seen from (14)), to (15) and (16), we obtain the first statement in Claim 2. The second statement in Claim 2 follows from the fact that $\Delta \beta_d$ is increasing in $\hat{c}$. ∎

Finally, we discuss the impact of the ZMP on the patients' health outcomes:

$$H\left(d^*\left(\beta_d(\alpha); \theta \right), s^*\left(\beta_d(\alpha); \theta \right); \theta \right),$$

or the patients' welfare:

$$H\left(d^*\left(\beta_d(\alpha); \theta \right), s^*\left(\beta_d(\alpha); \theta \right); \theta \right) - r_d^p d^*\left(\beta_d(\alpha); \theta \right) - r_s^p s^*\left(\beta_d(\alpha); \theta \right)$$

where $d^*\left(\beta_d(\alpha); \theta \right)$ and $s^*\left(\beta_d(\alpha); \theta \right)$ are the drug and non-drug service quantities received by patients with characteristics θ if the physician faces parameter $\beta_d(\alpha)$, which is affected by the ZMP. Since $d^*\left(\beta_d(\alpha); \theta \right)$ and $s^*\left(\beta_d(\alpha); \theta \right)$ are chosen to maximize the objective function of the physicians, in general the impact of the ZMP on the patients' health and welfare outcomes is ambiguous.

4 Data and Descriptive Statistics

4.1 Data Sets

Our main data set is the medical claims data for the population covered by the NRCMS in the sample county, from January 1, 2009 to September 30, 2013. The NRCMS is the major source of heavily subsidized public health insurance for rural residents in China. It was introduced in 2003, and by 2009 it covered more than 95 percent of the rural population. Thus, our data set includes almost all the rural residents in the county. All the basic medical facilities (including the village clinics and township health centers) in the county started implementing the ZMP and EML on October 1, 2010, but the county hospitals were not required to implement the ZMP and EML until October 2013.

Although the data set contains claims for visits to all medical facilities, we only use the data for visits to township health centers and county hospitals. The reason is that we focus on inpatient visits only in this paper, and the village clinics do not provide inpatient services. We also exclude visits to the two largest county hospitals and the Center for Disease Control (CDC), as patients in these hospitals were in general different from the patients in township centers. These sample restrictions leave us with patients who visited the 18 township centers (the treatment group) and 13 county hospitals (the control group); and we have 7 and 12 quarters of data, respectively, before and after the implementation of the ZMP in the township centers. With the sample restrictions, we end up with 82,356 admissions and 51,061 unique patients.

The claims data contains rich information on each inpatient visit, including the patient's ID, service date, length of stay, diagnosis, and detailed expenditures on drugs and non-drug medical services, if any. For each inpatient claim, we also observe the attending physician's name and their affiliated hospital, and the details of the prescriptions for drugs and non-drug services (including tests and physical examinations). As we mentioned, we focus on inpatient visits in this paper, thus the rich information provides a particular advantage in investigating the physicians' behavioral changes across drug prescriptions and non-drug services.

We also use two supplemental data sets. The first is the enrollment information of the NRCMS, in which we observe the enrollment status of each individual each year, as well as the demographic information on each enrollee, including age, gender, marital status, and education. The second is the reimbursement policy of the NRCMS, which could vary annually.

4.2 Descriptive Statistics

In Table 1, we present the demographic characteristics of patients with inpatient visits, which we use in our analysis. We present the demographic distributions for the full sample, and separately for the subsamples of the township centers and county hospitals. For each sample, we also present the claim level patient statistics separately for the periods before and after the ZMP was implemented in the township centers on October 1, 2010. In Column (1), we see that among our full sample of 82,356 inpatient visits, more than 80 percent are aged 40 years or older, 41.5 percent are male, about 72.0 percent are married, and around 30.2 percent and 41.0 percent are illiterate and attend at most primary school, respectively. In Columns (2) and (3) where we

break the full sample into those with inpatient visits before and after October 1, 2010, we see that overall the demographic distributions are quite stable, except for the age distribution. In Columns (4), we see that patients in the township centers tend to be very similar to the full sample. Comparing Columns (4) and (7), we see that patients in county hospitals tend to be younger than those in the township centers, and there are higher shares of males and more educated patients. Comparing Columns (5) and (6), we see that in the township centers the patients tend to be older after the reform, similar to what happened in the county hospitals. This could be a result of the large-scale flows of younger individuals to work as migrant workers in this period.[11] Importantly, the differences between county and township hospital samples tend to be quite stable over time. In our claims-level analysis below, we will control for patients' characteristics, including demographic characteristics.

[Table 1 About Here]

In Table 2, we describe the basic inpatient statistics, including total medical expenditures, drug expenditures, non-drug expenditures, inpatient stay duration (days), 30-day readmission rate, as well as insurance deductible and reimbursement rate.[12] The total expenditures include all payments received by the health facilities, including the insurance payments and the patients' out-of-pocket expenses. The total expenditures are further divided into two mutually exclusive categories, drug expenditures and non-drug expenditures, the latter of which include the diagnostic fee, bed fee, and fees for physical examinations and tests.

In Panels A and B of Table 2, we show the statistics for patients at the township health centers and county hospitals, respectively. In Columns (1) and (2), we show the mean and standard deviation of the relevant variables for the whole sample from January 1, 2009 to September 30, 2013. In Columns (3)-(6), respectively, we present the statistics for the periods before and after the ZMP was implemented in the township centers on October 1, 2010. Column (7) presents the differences between the post-ZMP and the pre-ZMP periods. In the full sample, the average total expenditures in the township centers were $1,080$ RMB, slightly more than one-third of the level in county hospitals, which was $3,133$ RMB. Similarly, the average inpatient duration was 6.9 days in the township centers versus 12.3 days in the county hospitals. The patterns were similar for periods before and after the ZMP reform, although the magnitudes of the differences vary. These facts imply that patients who are admitted to county hospitals are more likely to have more severe health conditions requiring higher expenditure. We take this difference into consideration in our empirical analysis below by controlling for detailed patient characteristics, including their diagnosis codes. We will also show that the differences between patients in township health centers and patients in county hospitals were stable over time prior to the implementation of the ZMP at the township centers. Table 2 also shows that the average reimbursement rate was higher in township health centers than that in county hospitals, but the reimbursement rate in county hospitals improved more compared with that in

---

[11]See a related report on this phenomenon by the National Bureau of Statistics of China at http://www.stats.gov.cn/ztjc/ztfx/fxbg/201103/t20110310_16148.html.

[12]The inpatient reimbursement is only qualified for expenditures exceeding a certain threshold, which is known as the "insurance deductible." In China, both the insurance deductible and the reimbursement rate might change annually according to local government policies.

township health centers. Table 2 also shows that average expenditure on drugs decreased dramatically in the township health centers, from 582 RMB before the ZMP reform to 419 RMB after the reform, which is in stark contrast with the increase from $1,058$ RMB to $1,169$ RMB in the county hospitals during the same period. The non-drug expenses in the township centers increased by around 50 percent, from 456.5 RMB before the ZMP to 687.3 RMB after the ZMP in the township centers; however, the non-drug expenses only increased by around 21 percent in the county hospitals, from $1,752.8$ RMB to $2,122.7$ RMB over the same period.

[Table 2 About Here]

Figure 1 plots the mean log drug expenditures per visit (Panel (a)), mean log non-drug expenditures per visit (Panel (b)), and the mean log total expenditures per visit (Panel (c)), by quarter, separately for the township centers and county hospitals. The vertical solid line represents October 1, 2010 when the ZMP was implemented in the township centers. The averages were higher in the county hospitals than in the township centers, which is consistent with Table 2, but the differences between the two types of facilities were quite stable before the reform. Interestingly, the average drug and non-drug expenditures started to diverge right after the ZMP reform in the township centers: Panel (a) shows that drug expenditures decreased in the township centers, while they remained relatively stable in the county hospitals; Panel (b) shows that the non-drug expenditures increased more in the township centers.[13] As is shown in Panel (c), the patterns of total spending are similar in both types of facilities. Of course, the raw average before/after differences presented in Table 2 and Figure 1 are only suggestive, as they do not yet control for important differences across the patients of the two types of health facilities.

[Figure 1 About Here]

5 Empirical Results

Propositions 2 and 3 in Section 3 show that following the ZMP, drug expenditures would decrease and non-drug expenditures would increase after the reform *only if* medical demands were partially physician induced. In this section, we empirically test these predictions. We first directly explore how drug, non-drug, and total expenditures are affected by the ZMP. Sections 5.1 and 5.2 present the admission-level results and physician- and hospital-level results, respectively. In Section 5.3, we decompose the change in drug expenditures by price and quantity. In Section 5.4 we examine the impact of the ZMP on patient health outcomes.

5.1 Admission-Level Results

Empirical Specification and Main Results. We employ a difference-in-differences (DID) estimation strategy to analyze the quantitative impact of the ZMP on drug, non-drug, and total expenditures. The

---

[13]Many factors could potentially drive the variations in medical expenditures over time, such as increases in household income and/or the reimbursement rates of the NRCMS.

township health centers. Table 2 also shows that average expenditure on drugs decreased dramatically in the township health centers, from 582 RMB before the ZMP reform to 419 RMB after the reform, which is in stark contrast with the increase from $1,058$ RMB to $1,169$ RMB in the county hospitals during the same period. The non-drug expenses in the township centers increased by around 50 percent, from 456.5 RMB before the ZMP to 687.3 RMB after the ZMP in the township centers; however, the non-drug expenses only increased by around 21 percent in the county hospitals, from $1,752.8$ RMB to $2,122.7$ RMB over the same period.

[Table 2 About Here]

Figure 1 plots the mean log drug expenditures per visit (Panel (a)), mean log non-drug expenditures per visit (Panel (b)), and the mean log total expenditures per visit (Panel (c)), by quarter, separately for the township centers and county hospitals. The vertical solid line represents October 1, 2010 when the ZMP was implemented in the township centers. The averages were higher in the county hospitals than in the township centers, which is consistent with Table 2, but the differences between the two types of facilities were quite stable before the reform. Interestingly, the average drug and non-drug expenditures started to diverge right after the ZMP reform in the township centers: Panel (a) shows that drug expenditures decreased in the township centers, while they remained relatively stable in the county hospitals; Panel (b) shows that the non-drug expenditures increased more in the township centers.[13] As is shown in Panel (c), the patterns of total spending are similar in both types of facilities. Of course, the raw average before/after differences presented in Table 2 and Figure 1 are only suggestive, as they do not yet control for important differences across the patients of the two types of health facilities.

[Figure 1 About Here]

5 Empirical Results

Propositions 2 and 3 in Section 3 show that following the ZMP, drug expenditures would decrease and non-drug expenditures would increase after the reform *only if* medical demands were partially physician induced. In this section, we empirically test these predictions. We first directly explore how drug, non-drug, and total expenditures are affected by the ZMP. Sections 5.1 and 5.2 present the admission-level results and physician- and hospital-level results, respectively. In Section 5.3, we decompose the change in drug expenditures by price and quantity. In Section 5.4 we examine the impact of the ZMP on patient health outcomes.

5.1 Admission-Level Results

Empirical Specification and Main Results. We employ a difference-in-differences (DID) estimation strategy to analyze the quantitative impact of the ZMP on drug, non-drug, and total expenditures. The

---

[13]Many factors could potentially drive the variations in medical expenditures over time, such as increases in household income and/or the reimbursement rates of the NRCMS.

17

treatment and control groups are patients of the township centers and county hospitals, respectively. The basic estimation equation is

$$y_{iht} = \alpha + \beta Town_{ih} \times Post_t + \boldsymbol{\delta}\mathbf{X}_{it} + \tau Z_{ht} + \gamma_t + \gamma_h + \varepsilon_{iht}, \tag{17}$$

where:

- y_{iht} is the outcome variables of interest for patient i, who had admissions in hospital h in time t (measured by quarter).[14] The outcome variables include drug expenditures, non-drug expenditures, and total spending, all in logs.

- The variable $Town_{ih}$ is a dummy variable that takes the value 1 if admission i was at a township center.

- The variable $Post_t$ is a dummy variable that takes the value 1 if the admission was on or after October 1, 2010, the starting date of the ZMP in the township centers.

- The vector $\mathbf{X}_{it}$ consists of the patient demographic characteristics, including age, gender, education, marital status, and his/her diagnosis codes. In the data, a patient's diagnosis is coded by the International Classification of Diseases (10th revision, ICD10). The ICD10 system includes more than 20,000 diagnosis codes, which are automatically classified into 20 Major Diagnosis Categories (MDCs). We include dummies for the 20 MDCs in $\mathbf{X}_{it}$.

- The variable Z_{ht} contains information on the reimbursement policy for hospital h at time t, including the insurance deductible and the coinsurance rate, which could change annually at a given hospital. We summarize the generosity measure of the insurance by calculating the expected out-of-pocket expenditure of a typical enrollee based on the population's average inpatient expenditures in the previous year according to the current year's policy in hospital h at time t (Finkelstein and Poterba, 2004).

- γ_t captures the quarter fixed effects, and γ_h captures the hospital fixed effects.

- ε_{iht} is an i.i.d. shock.

The key coefficient of interest is β, which captures the causal effect of the ZMP on outcome y_{iht} in the township centers where the ZMP was implemented relative to that in the county hospitals where the ZMP was not yet implemented.

Table 3 reports the estimation results from the admission-level analysis using Eq. (17), where we report the main coefficient of interest β. Robust standard errors clustered at hospital level are in parentheses. In Column (2), we show that drug expenditures were reduced by 63.4 log points (about 47.0 percent) after the ZMP was implemented in the township centers relative to that in the county hospitals where the ZMP was not yet implemented. The estimate is almost identical in Column (1) where we did not control for patient demographics and diagnosis codes, suggesting that the impact of the ZMP is almost orthogonal to

---

[14]We also ran the same regressions controlling for month fixed effects, instead of quarter fixed effects. The results are quantitatively and qualitatively almost identical. See Appendix Table A4.

the patient characteristics. In Columns (3) and (4), we show that compared with the county hospitals, non-drug expenditures in the township centers increased by 27.88 log points (about 32.3 percent) after the implementation of the ZMP. All the coefficient estimates are statistically significant at the 1 percent level. Finally, in Columns (5) and (6), we show that the ZMP reduced total expenditures by about 11.5 log points (about 12.2 percent) in the township centers relative to those in the county hospitals; however, the effect is not statistically significant. This suggests that following the ZMP, the increase in non-drug expenditures compensated almost all the reduction in drug expenditures.

[Table 3 About Here]

To summarize, the results reported in Table 3 suggest that under the ZMP, patients' drug expenditures decreased but their non-drug expenditures increased, leaving the total expenditures almost unchanged. These findings are consistent with the theoretical predictions of Propositions 2 and 3 when $\alpha > 0$, that is, when the medical demand is partially physician induced. A question might be that if non-drug expenditures were profitable for the physicians/hospitals, why did they wait until the ZMP went into effect to increase the prescription of non-drug expenses? In our theoretical model, we capture this by assuming that physicians will incur effort cost $e(s)$ for prescribing non-drug services s. This is reasonable, because non-drug services, whether they are diagnostic service or testing, would require time input from the physician. For example, to prescribe a test, the physician may need to spend more time with the patient to wait for and check the results. In contrast, prescribing drugs does not require such time input. Thus, in the absence of the ZMP, physicians' incentives to over-prescribe non-drug services were moderated by the effort costs.

It is useful to examine the heterogeneous impacts of the ZMP on drug expenditures for the 20 MDC categories. We run the regressions according to Eq. (17) separately for each MDC category. The β coefficient estimates from these regressions are shown in Figure 2 by MDC category. It can be seen that drug expenditures decreased following the ZMP in all the categories, although some of the estimates are not statistically significant. The magnitudes of the reductions vary. Drug expenditures decreased the most for mental illness, and they decreased the least for ear- and eye-related diseases.[15]

[Figure 2 About Here]

In Table 4, we examine the impact of the ZMP on non-drug expenditures by subcategories, specifically, infusion, bed, nursing care, tests and examinations, and medical materials. We run the regressions according to Eq. (17) using expenditures on each of the non-drug categories as the outcome variable, separately. The β coefficients in Table 4 show that the ZMP increased non-drug expenditures the most for more discretionary services, such as tests and examinations, bed stay, and nursing care, with increases ranging from 22 to 35 log points (25 to 42 percent). Expenditures on less discretionary services, e.g., medical materials, increased the least (about 3.5 log points) and the estimated effect is statistically insignificant.

[Table 4 About Here]

---

[15]The results are intuitive. Treatment for mental diseases relies heavily on medications, so the ZMP policy would influence the revenue the most; in contrast, treatment for ear or eye diseases always applies equipment and tests and uses relatively fewer medications, so the policy would have less impact on drug expenditures in the treatment of these diseases.

Event Study and Parallel Pre-Trend Assumption. For the DID empirical specification to provide an estimate of the causal impact of the ZMP, we need to ensure that the control and treatment groups exhibit parallel pre-trends on the outcome variables of interest. To test the parallel pre-trend assumption, as well as to explore the possible time-varying effects of the ZMP, we estimate the following regression specification by creating a set of quarter dummies and interacting them with whether the health facility is a township center:

$$y_{iht} = \alpha + \sum_{\kappa=-7}^{11} \beta_\kappa Town_{ih} \times \mathbb{1}\{t = \kappa\} + \boldsymbol{\delta}\mathbf{X}_{it} + \tau Z_{ht} + \gamma_t + \gamma_h + \varepsilon_{iht}, \tag{18}$$

where $\mathbb{1}\{t = \kappa\}$ is a group of dummy variables indicating in which quarter κ admission i occurred, so β_κ captures the difference between township centers and county hospitals in quarter κ. The rest of the variables are defined in the same way as those in Eq. (17).

Figure 2 displays the full set of estimated β_κ coefficients, for $\kappa \in \{-7, -6, ..., -1, 0, ..., 11\}$ where we normalize $\beta_{-1} = 0$ for $\kappa = -1$, the quarter before the implementation of the ZMP in the township centers on October 1, 2010. Panel (a) presents the coefficients for drug expenditures, Panel (b) for non-drug expenditures, and Panel (c) for total expenditures. The 95 and 99 percent confidence intervals of the β_κ are displayed as whiskers. Panel (a) of Figure 3 indicates that before the ZMP, the time patterns in drug expenditures were statistically similar between the township centers and the county hospitals; however, after the ZMP reform was implemented in the township centers, there was a significant and widening reduction in drug expenditures in the township centers relative to the county hospitals. Similarly, Panel (b) shows that for non-drug expenses, the differences between the township centers and county hospitals were statistically insignificant before the ZMP reform, but they widened after the reform. Panel (c) shows that the time patterns in the differences in total expenditures remained insignificant before and after the reform. Figure 3 suggests that the parallel pre-trend assumption for the DID analysis is valid.

[Figure 3 About Here]

5.2 Physician- and Hospital-Level Results

Our analysis so far has been at the admission level. Now we conduct a similar analysis at physician level to directly detect the change in physician behavior. Specifically, the estimating equation is as follows:

$$y_{jt} = \alpha + \beta Town_j \times Post_t + \tau Z_{jt} + \gamma_t + \gamma_j + \varepsilon_{jt}, \tag{19}$$

where now y_{jt} is the log of the average outcomes of patients treated by physician j in quarter t; and $Town_j$ equals 1 if physician j works at a township hospital.[16] Z_{jt} controls for the average reimbursement amount across the patients who are treated by physician j in quarter t. In addition, we control for the physician fixed effects γ_j (which is more general than hospital fixed effects), and the quarter fixed effects γ_t. The

---

[16]In our data, we observe the physicians' names and their affiliated hospitals on the prescription data set. In China, a physician is rarely affiliated with multiple hospitals. Therefore, we identify a unique physician as a physician name-hospital pair.

coefficient β captures the impact of the ZMP on the behavior of physicians in the township centers relative to physicians in the county hospitals. We cluster the robust standard errors at the hospital level.

To ensure that our results are not driven by newly entering or the exiting physicians, we create a balanced panel of physicians who were present in our analysis data before and after the ZMP. We have a total of 535 unique physicians and 4,060 physician-quarter observations in our sample.

[Table 5 About Here]

We report the physician-level estimation results using Eq. (19) in Panel A of Table 5. In Column (1), the outcome variable is the log of average drug expenditures per admission prescribed by the physician, and it shows that following the implementation of the ZMP in township centers, physicians in township centers reduced their average drug expenditures per admission by 63.4 log points (47.0 percent), and it is statistically significant at the 1 percent level. In Column (2), the dependent variable is the log of average non-drug expenditures per admission prescribed by the physician, and it shows that following the implementation of the ZMP in township centers, physicians in township centers increased their average non-drug expenditures per admission by 23.0 log points (25.9 percent), and it is statistically significant at the 5 percent level. In Column (3), the outcome variable is the log of average total expenditures per admission prescribed by the physician, and it shows that following the implementation of the ZMP in township centers, physicians in township centers reduced their total expenditures per admission by 13.5 log points (12.6 percent), and it is statistically significant at the 5 percent level. The physician-level results on the effect of the ZMP on drug and non-drug expenditures reported in Columns (1) and (2) of Table 5 are similar in magnitude to the admission-level results reported in Table 3. The physician-level results on the impact of the ZMP on average total expenditure per admission are statistically significant at the 5 percent level; this is in contrast to the admission-level results in Table 3, which show a similar negative effect of about -11.4 log points (10.8 percent), but it is statistically insignificant.

Finally, in Column (4) of Table 5, we report the impact of the ZMP on the log quarterly number of inpatient admissions of the physician. It shows that following the implementation of the ZMP in township centers, physicians in township centers decreased their number of admissions by 35.9 log points (30.2 percent) relative to those of the county hospitals, and it is statistically significant at the 10 percent level. This result confirms the predictions in Part (2) of Proposition 4, which states that if the medical demand is partially physician induced, then the set of treated patients will shrink after the ZMP. To the best of our knowledge, this is a novel test of the theory of physician-induced demand.

We can also summarize the data at the hospital-quarter level; that is, we calculate the average of drug, non-drug, and total expenditures per admission, together with the total number of admissions, by quarter for each hospital. We then run the regressions at the hospital-quarter level using a specification similar to Eq. (20) with the physician fixed effects γ_j replaced by the hospital fixed effects.

The hospital-level results are reported in Panel B of Table 5. The results show that at the hospital level, following the implementation of the ZMP in the township centers, drug expenditures per admission decreased by 54.9 log points (about 42.2 percent) as shown in Column (1), but non-drug expenditures increased by 35.3 log points (about 42.3 percent) as shown in Column (2), and total expenditures decreased by 5.81 log

21

points (about 0.6 percent) as shown in Column (3) in the township centers relative to the county hospitals. The estimated effects for drug and non-drug expenditures are statistically significant at the 1 percent level, but the estimated effect for total expenditures is not statistically significant. Column (4) shows that the implementation of the ZMP in the township centers reduced the average total quarterly admissions in the township centers by 48.7 log points (38.5 percent) relative to those in the county hospitals, and the estimate is statistically significant at the 1 percent level. The hospital-level results are consistent, both qualitatively and quantitatively, with the physician-level results (Panel A) and the admission-level results reported in Table 3.

We now employ an event study framework to explore the variation in the effects over time, by adding a set of quarter dummies interacting with the facility dummy:

$$y_{jt} = \alpha + \sum_{\kappa=-7}^{11} \beta_\kappa Town_j \times \mathbb{1}\{t = \kappa\} + \tau Z_{jt} + \gamma_t + \gamma_s + \gamma_j + \varepsilon_{iht}, \tag{20}$$

where j represents a specific physician or hospital, and $\mathbb{1}\{t = \kappa\}$ is a group of dummy variables that take the value 1 if the outcome variable of interest occurred in quarter $\kappa \in \{-7, ..., 11\}$; thus, β_κ captures the difference between the township centers and the county hospitals in quarter κ. γ_t and γ_s represent the year fixed effects and season fixed effects respectively. The other variables are defined the same way as those in Eq. (19).

Figure 4 plots the coefficient estimates $\beta_\kappa, \kappa \in \{-7, ..., 11\}$, for drug expenditures (Panel (a)), non-drug expenditures (Panel (b)), total expenditures (Panel (c)), and the number of inpatient admissions (Panel (d)), at the physician level. The physician-level results are consistent with the admission-level results, as shown in Figure 2. The overall pattern supports the parallel pre-trend assumption, and we find that drug expenditures decreased, non-drug expenditures increased, and total expenditures decreased somewhat. Finally, Panel (d) of Figure 4 shows that the time patterns in the total number of inpatient admissions for physicians before the ZMP are similar between the township centers and the county hospitals, but after the ZMP, the number of admissions for physicians in the township centers experienced a reduction relative to that for physicians in the county hospitals.

[Figure 4 About Here]

Taking Stocks. So far, the results in Tables 3 and 5 show that, following the implementation of the ZMP in the township centers, drug expenditures in the township centers decreased by 60 log points at the admission level, 63 log points at the physician level, and 55 log points at the hospital level, per admission. The price differences between the county hospitals and township centers were about 15 percent (translating into $100 * \ln(1 - 0.15) \approx -16$ log points) conditional on the same drug; the reason is that the wholesale prices for the same drug were the same for the township centers and the county hospitals under the central procurement platform, but the county hospitals were still permitted to charge a 15 percent markup over the drug wholesale price in the study period. Thus, we can conclude that the estimated effects of the ZMP on drug expenditures in the township centers translate into an effect on the quantity of drugs demanded by about 44 log points (admission level), 47 log point (physician level), and 39 log points (hospital level), per

admission. Similarly, since the prices for non-drug services were not changed by the reform, the estimated effects of the ZMP on the non-drug expenditure suggest that the quantity of non-drug services increased by 27 log points at the admission level, 23 log points at the physician level and 35 log points at the hospital level.

5.3 Decomposing the Drug Expenditure Effect of the ZMP

Our main finding in Section 5.1 is that the ZMP reduced drug expenditures by about 60 log points (around 45 percent) per admission in the treated township centers relative to those in the untreated county hospitals. In this subsection, we decompose the drug expenditure effect of the ZMP into the quantity effect and the retail price effect, and further examine whether the physicians in the treated township health centers prescribed fewer types of drug (extensive margin) and fewer drugs for each type (intensive margin).

To conduct this decomposition, we need to create a panel of drugs. In the prescription data set, we observe each drug by its chemical name (chemical compound), package format, prescribed unit, and expenditure. For example, suppose that a drug was 250ml (package) 10 percent glucose (chemical name) injection, two units were prescribed, and the bill for this prescription was 12.5 RMB. The retail price of the drug was not directly listed in the data set; instead, it is deduced from the expenditure and quantity. Before the ZMP was implemented, the retail price was equal to the wholesale price plus the markup, which was in general 15 percent of the wholesale price. Hence, the wholesale price can be backed out if we divide the deduced retail price by 1.15 in the pre-ZMP period. In the post-ZMP period, the retail price equals to the wholesale price in the township health centers.

Drugs with the same chemical name are considered to be the same type of drugs, and we standardize the drug quantity by multiplying the prescription unit and the package. We also calculate the standardized price as the total prescription expenditure divided by the standardized total prescription quantity. The quality or efficiency of drugs that were categorized as the same type could still vary, as they may be provided by different pharmaceutical firms with different levels of chemical strength or daily dosage requirement. To deal with this issue, we therefore further group the drugs of the same chemical name into up to 4 bins based on their wholesale price. In summary, we classify each drug jointly by its chemical name and price bins. This classification is not perfect, but it makes the best use of the information in the data, and the drugs in the same bin are now more comparable.

We then disentangle the changes in drug expenditures into changes in prices and quantities after the ZMP. In this analysis, the identification comes from the prescribed drugs that appear both in the township centers and the county hospitals before and after the ZMP. Therefore, we run the regressions based on the balanced sample, including only drugs that were provided both before and after the reform. Our balanced panel of drugs accounts for around 76.7 percent of the total prescriptions. As mentioned in Section 2, the township centers were required to prescribe 100 percent from the EML, and county hospitals were required to prescribe at least 40 percent from the EML after ZMP. Such prescription regulation was not binding for county hospitals since the EML shares were 72.3 percent and 72.6 percent, respectively, before and after the

ZMP. And the regulation slightly increases the EML share for township hospitals from 89.4 percent to 94.2 percent. In our balanced sample, the EML accounted for 96.88% before the ZMP and 98.35% after the ZMP. The consistently high share of EML prescription indicates that the EML prescription regulation should have a negligible impact on our empirical results.

We run the following regressions:

$$y_{diht} = \alpha + \beta Town_{dih} \times Post_t + \boldsymbol{\delta} \mathbf{X}_{it} + \gamma_{ds} + \gamma_t + \gamma_h + \varepsilon_{diht}, \tag{21}$$

where each observation is at the drug-visit level, and the dependent variable y_{diht} is either log of prescription quantity, or standardized drug price, or the expenditure of each type of drug for each visit. Similar to Eq. (17), insurance policy, individual demographics, and diagnosis are controlled in the regression. To deal with potential confounding factors that may affect the price and quantity, we need to control for hospital fixed effects γ_h and quarter fixed effects γ_t. We also control for drug-season level fixed effects γ_{ds} to allow the effects varying across seasons within the same drug.

[Table 6 About Here]

The first three columns in Table 6 display the results. It shows that, among the balanced panel of drugs, the quantity of drugs prescribed per visit decreased by about 17.8 log points (16.3 percent), the price decreased by about 17.1 log points (15.7 percent, as expected), and the total expenditures decreased by 35 log points (29.4 percent). That is, half of the reduction in drug expenditure is contributed by quantity reduction, and the other half is contributed by price reduction.

Remark 2 *Recall that, as reported in Table 3, the reduction in all drug expenditure following the implementation of the ZMP in the township centers is* 47.0 *percent. Column (3) of Table 6 indicates that about three quarters of the drug expenditure reductions were from this balanced drug panel. This is consistent with the fact, which we previously mentioned, that the balanced panel of drugs accounts for around 76.7 percent of the total prescriptions.*

We also report the results from some additional outcomes of interest. In Column (4) of Table 6, we use the total number of different types of prescribed drugs in an inpatient visit as the outcome variable to explore the extensive margin impact, using the same regression specification of Eq. (17). The result in Column (4) indicates that the ZMP does not seem to significantly change the prescription behavior in the extensive margin.

Proposition 5 predicts that in post-ZMP period, township centers may decrease the prescription for drugs with high-price level. We examine the prediction in the last two columns of Table (6). To check the heterogeneous impacts across drugs, we first define "Expensive" as drugs whose average wholesale price is above the median wholesale price of drugs with the same chemical name during our sample period. Then we calculate the share of "Expensive" drug expenditure out of the total drug expenditure, as well as the share of "Expensive" drug prescription quantity out of the total drug prescription quantity per visit. Then we set the share of "Expensive" drug expenditure and the share of "Expensive" drug prescription quantity as the outcome variables, and apply the framework of Eq. (17). The last two columns of Table (6) show that both

the expenditure share and the quantity share of "Expensive" drugs decreased slightly after the ZMP reform in the township centers, although the results are not statistically significant. Therefore, the empirical results at least weakly support Case (2) in Proposition 5.

5.4 How Does the ZMP Affect Patient Outcomes?

So far, we have established that the ZMP reform reduced the price and quantity of drugs and increased the provision of non-drug services. An important question is, how did this compositional change in expenditures of drugs and non-drug services affect the quality of care and other related outcomes for the patients? We only have limited outcome measures in the claims data unfortunately; so we use the *readmission*, a dummy variable that takes the value 1 if the patient was readmitted to any medical facility, regardless of whether the readmission occurred inside or outside our study county, for the same diagnosis within two weeks or a month of discharge respectively, to proxy for the quality of care. Since we have data on the universe of all medical claims, we tracked each patient's inpatient records across all medical facilities to construct this measure. The average two-week readmission rate in our sample is 1.6 percent, and the monthly readmission rate is 2.6 percent. Table 7 reports the regression results using the same specifications as in Eq.(17) at the admissions level by setting the readmission dummy as the outcome variable. It shows that the reform did not significantly change the readmission pattern. Thus it is likely that patients' health outcomes were not affected by the changes in the drug and non-drug service composition following the ZMP.

[Table 7 About Here]

As shown in Column (4) of Table 5, the overall utilization has been decreasing since the number of admissions decreased. The last two columns in Table 7 provides another measure of utilization using the length of stay per inpatient visit. The average length of stay is 8.18 days, and the reform did not change it significantly either.

5.5 Robustness Checks

Although the empirical results are consistent with the model predictions, there is still a possibility that the findings were induced by the switch of patients between the different levels of facilities. Specifically, patients with more severe illnesses may have switched to county hospitals to receive care after the reform. If that is the case, patients in county hospitals would have worse health status than before, and those in township centers would be healthier. As patients with different disease types should involve different treatments, the resulting change in service provision might be induced by the change in patients' sorting. To address this concern, we compare the diagnoses of patients treated in the township centers and those treated in the county hospitals before and after the ZMP reform. The results are presented in Appendix Table A1. It shows that the majority of the diagnosis distributions were stable after the reform.

We also calculate the patients' risk scores, and explore the pattern of the risk score distributions before and after the ZMP reform. The risk score is a measure predicting the risk of future medical expenditures for an individual based on his/her characteristics, such as age, gender, and diseases. We follow the same

specifications as in Eq.(17) at the admission level by setting the patients' risk score as the outcome variable. As shown in Appendix Table A2, patients' risk score did not vary significantly across the township centers and the county hospitals before and after the ZMP reform, suggesting that the extent of patient switching is limited. This result is robust as we add more demographic characteristics in the regression.

In Appendix Table A3, we show that our main results are also robust to different estimation strategies and sample selections. In our main analysis, we showed the raw patterns as well as regression results with different controls on fixed effects, and the results were all consistent. Three additional robustness checks are reported in Appendix Table A3. First, we include the two largest hospitals as well as the county's disease control center in the sample. Second, we exclude the inpatient records that took place within six months of the ZMP reform in the township centers (i.e., between April 1, 2010 and September 30, 2010) to alleviate the concerns that patients'/physicians'/hospitals' expectations of the reform may have played a role in our findings. Third, we exclude the four private hospitals from the sample. Our results are robust to all of these sample restrictions. In Appendix Table A4, we also show that our results are robust to controls for month fixed effects and quarter fixed effects.

6   Conclusion

We use a natural experiment in China's health care reform, in which the government mandated zero markups for drugs sold in public hospitals, to study physician-induced demand for medical care. While there is large empirical literature on physician-induced demand, we argue through the lens of a theoretical model that the ZMP provides a unique setting to detect the presence of physician-induced demand. The model robustly predicts that the ZMP will lead to a reduction in the demand for drugs but an increase in the demand for non-drug services, *only when* the demands for drugs and non-drug services are at least partially physician induced. Our model also yields testable predictions of the impact of the ZMP regarding the equilibrium number of treated patients.

Exploiting the staggered rollout of the ZMP and a unique claims data set, we find that at the admission level, the reform decreases drug expenses by 63.4 log points (47.0 percent), but the reduction in drug expenses is almost fully substituted by the increases in expenses for non-drug services. Quantitatively similar results are also obtained at the physician and at the hospital levels. We also show that the ZMP reduces the number of patients in the treated hospitals, as predicted by the model, and that the reform has little impact on the quality of health care proxied by the readmission rate.

There are some limitations of our research. First, our results are based on data collected from township centers and county hospitals. Drugs purchased from other facilities, such as pharmacies, were not included in our analysis. However, this should not be a major concern, as the analysis focused only on inpatient hospital stays. It is uncommon for patients to purchase medications from pharmacies and use them during their hospital stays. Second, our sample only included the rural population in a single county. There is likely great heterogeneity between the rural and urban populations and across geographic areas. Therefore our empirical finding regarding physician-induced demand is only suggestive for other settings.

References

Alexander, Diane, "How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs," *Journal of Political Economy*, November 2020, *128*.

Arrow, Kenneth J, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 1963, *53*, 941–973.

Burns, Lawton Robert and Gordon G Liu, *China's Healthcare System and Reform*, Cambridge University Press, 2017.

Chen, Brian K, Y Tony Yang, and Karen Eggleston, "Patient Copayments, Provider Incentives, and Income Effects: Theory and Evidence from the Essential Medications List under China's 2009 Healthcare Reform," *World Medical & Health Policy*, 2017, *9* (1), 24–44.

Clemens, Jeffrey and Joshua D Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–49.

Currie, Janet, Wanchuan Lin, and Juanjuan Meng, "Using Audit Studies to Test for Physician Induced Demand: The Case of Antibiotic Abuse in China," *Journal of Development Economics*, 2014, *110*, 39–51.

Dafny, Leemore S, "How Do Hospitals Respond to Price Changes?," *American Economic Review*, 2005, *95* (5), 1525–1547.

Fang, Hanming and Qing Gong, "Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked," *American Economic Review*, 2017, *107* (2), 562–591.

Finkelstein, Amy and James Poterba, "Adverse Selection in Insurance Markets: Policyholder Evidence from the UK Annuity Market," *Journal of Political Economy*, 2004, *112* (1), 183–208.

_ , **Matthew Gentzkow, and Heidi Williams**, "Sources of Geographic Variation in Health Care: Evidence from Patient Migration," *Quarterly Journal of Economics*, 2016, *131* (4), 1681–1726.

_ , **Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group**, "The Oregon Health Insurance Experiment: Evidence from the First Year," *Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

Geruso, Michael and Timothy Layton, "Upcoding: Evidence from Medicare on Squishy Risk Adjustment," *Journal of Political Economy*, 2020, *128* (3).

Grennan, Matthew, Kyle Myers, Ashley Swanson, and Aaron Chatterji, "No Free Lunch? Welfare Analysis of Firms Selling through Expert Intermediaries," Working Paper, 2020, *Available at SSRN: https://ssrn.com/abstract=3216172*.

Gruber, Jonathan and Maria Owings, "Physician Financial Incentives and Cesarean Section Delivery," *Rand Journal of Economics*, 1996, *27* (1), 99.

Hall, Robert and Charles I. Jones, "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, 2007, *122* (1), 39–72.

Iizuka, Toshiaki, "Physician Agency and Adoption of Generic Pharmaceuticals," *American Economic Review*, 2012, *102* (6), 2826–58.

Li, Yongbin, Jing Xu, Fang Wang, Bin Wang, Liqun Liu, Wanli Hou, Hong Fan, Yeqing Tong, Juan Zhang, and Zuxun Lu, "Overprescribing in China, Driven by Financial Incentives, Results in Very High Use of Antibiotics, Injections, and Corticosteroids," *Health Affairs*, 2012, *31* (5), 1075–1082.

Liu, Ya-Ming, Yea-Huei Kao Yang, and Chee-Ruey Hsieh, "Financial Incentives and Physicians' Prescription Decisions on the Choice between Brand-name and Generic Drugs: Evidence from Taiwan," *Journal of Health Economics*, 2009, *28* (2), 341–349.

Lu, Fangwen, "Insurance Coverage and Agency Problems in Doctor Prescriptions: Evidence from a Field Experiment in China," *Journal of Development Economics*, 2014, *106*, 156–167.

McGuire, Thomas G, "Physician Agency," in "*Handbook of Health Economics*," Vol. 1, Elsevier, 2000, pp. 461–536.

National Bureau of Statistics of China, *China Statistical Yearbook 2010*, Javvin Technologies Incorporated, 2010.

National Health Ministry of China, *China Public Health Statistical Yearbook 2012*, Peking Union Medical College Publishing House, Beijing, 2011.

Newhouse, Joseph P, Insurance Experiment Group Rand Corporation, Insurance Experiment Group Staff et al., *Free for All? Lessons from the Rand Health Insurance Experiment*, Harvard University Press, 1993.

Wagstaff, Adam, Winnie Yip, Magnus Lindelow, and William C Hsiao, "China's Health System and Its Reform: A Review of Recent Studies," *Health Economics*, 2009, *18* (S2), S7–S23.

World Health Organization, "Public Spending on Health: A Closer Look at Global Trends," Technical Report 2018.

Wu, Bingxiao, "Physician Agency in China: Evidence from a Drug-Percentage Incentive Scheme," *Journal of Development Economics*, 2019, *140*, 72–89.

Yi, Hongmei, Grant Miller, Linxiu Zhang, Shaoping Li, and Scott Rozelle, "Intended and Unintended Consequences of China's Zero Markup Drug Policy," *Health Affairs*, 2015, *34* (8), 1391–1398.

Yip, Winnie and William C Hsiao, "The Chinese Health System at a Crossroads," *Health affairs*, 2008, *27* (2), 460–468.

Yip, Winnie C, "Physician Response to Medicare Fee Reductions: Changes in the Volume of Coronary Artery Bypass Graft (CABG) Surgeries in the Medicare and Private Sectors," *Journal of Health Economics*, 1998, *17* (6), 675–699.

Yip, Winnie Chi-Man, William C Hsiao, Wen Chen, Shanlian Hu, Jin Ma, and Alan Maynard, "Early Appraisal of China's Huge and Complex Health-care Reforms," *The Lancet*, 2012, *379* (9818), 833–842.

Yip, Winnie, Hongqiao Fu, Angela T Chen, Tiemin Zhai, Weiyan Jian, Roman Xu, Jay Pan, Min Hu, Zhongliang Zhou, Qiulin Chen et al., "10 Years of Health-care Reform in China: Progress and Gaps in Universal Health Coverage," *The Lancet*, 2019, *394* (10204), 1192–1204.

Figure 1: Quarterly Patterns of Expenditures in Township Centers and County Hospitals

hh

(a) Log Drug Expenditures

(b) Log Non-Drug Expenditures

(c) Log Total Expenditures
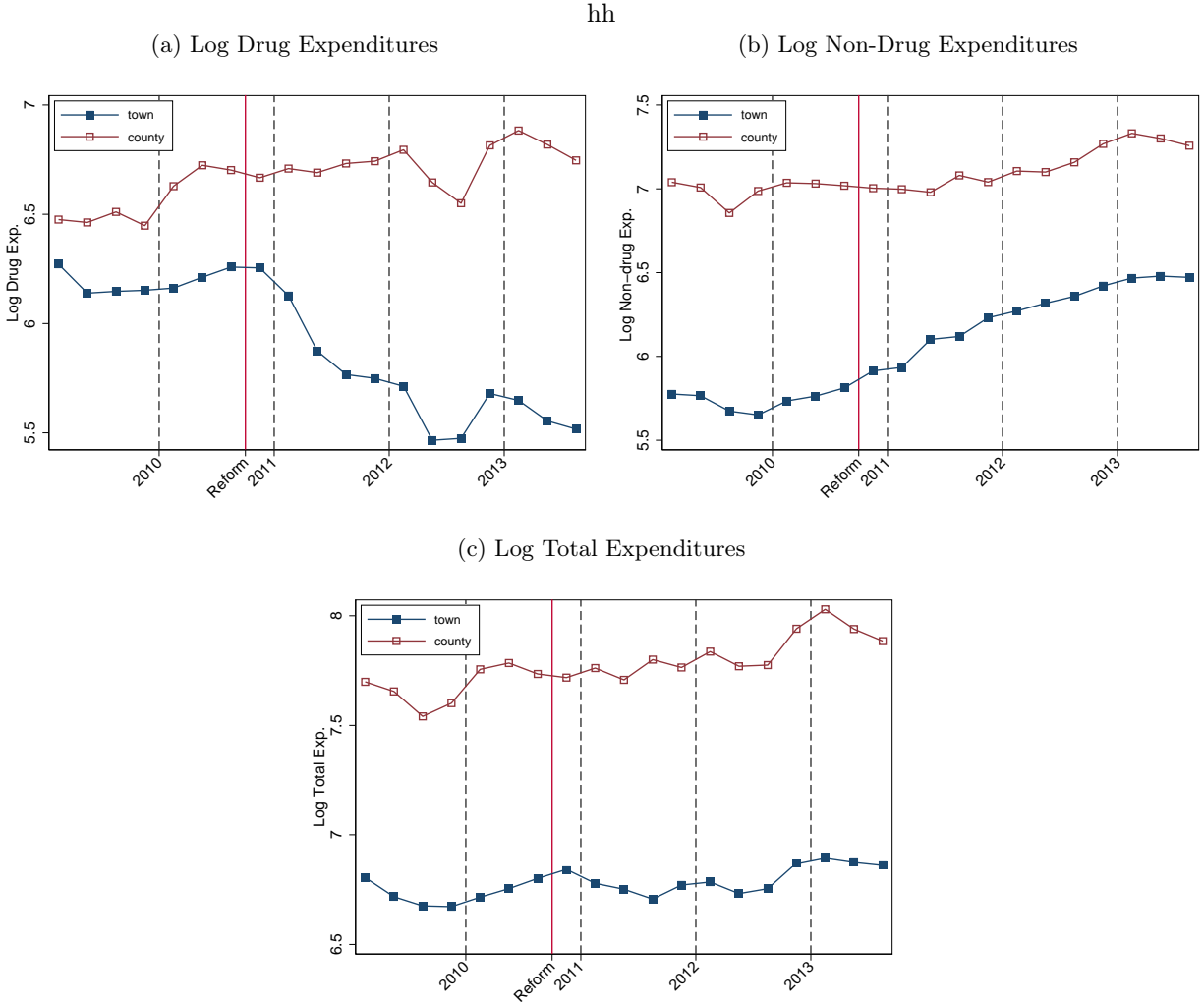
Notes: This figure plots the time trend of the medical spending for township health centers and the county hospitals separately. In Panel (a), we calculate the average drug expenditure (in log) per visit for each hospital by quarter, then we average across the township health centers and county hospitals separately. We define the non-drug expenditure as the total expenditure minus the drug expenditure. In Panel (b), we calculate the average non-drug expenditure (in log) per visit for each hospital by quarter, then we average across the township health centers and county hospitals separately. In Panel (c), we calculate the average total medical expenditure (in log) per visit for each hospital by quarter, then we average across the township health centers and county hospitals separately.

Figure 2: Coefficients on Drug Expenditures by 20 Disease Categories
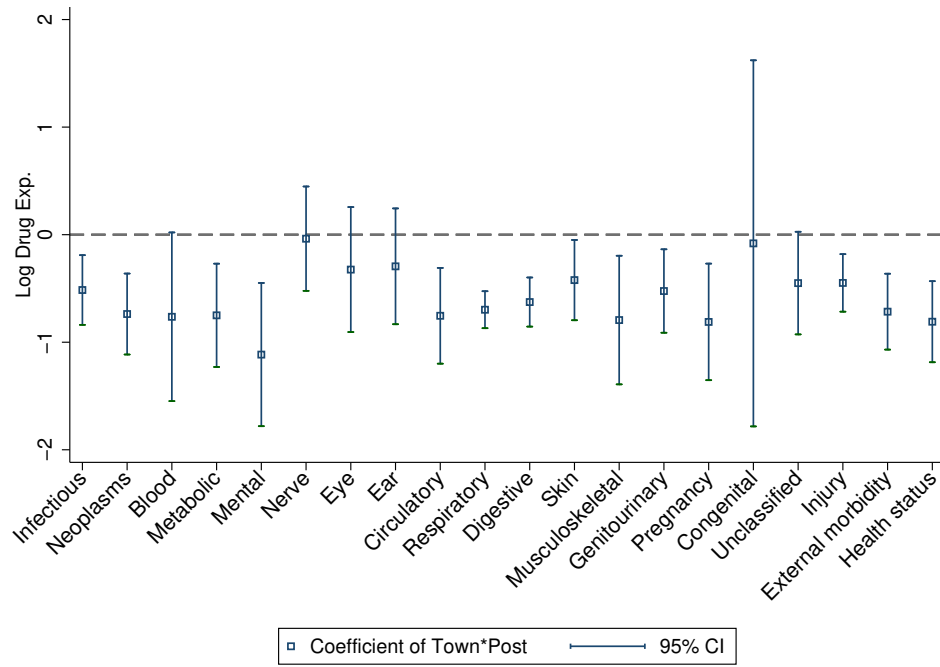
Notes: We categorize the sample into 20 disease categories, and estimate a diff-in-diff type of analysis in Eq. (17) for each disease category separately. This figure plots the coefficients of the interaction term β for each disease category. In the regression, we control for the patients' demographic characteristics, and medical insurance reimbursement generosity. The standard errors are clustered at the hospital level.

Figure 3: Quarterly Differences in Drug, Non-Drug, and Total Expenditures between Township Centers and County Hospitals

(a) Drug Expenditure

(b) Non-Drug Expenditure

(c) Total Expenditure

Notes: This figure plots the β_κ coefficients of Eq. (18). In Panel (a), we regress the log of the average drug expenditure on the dynamic time indicator such that the effects of the reform would vary across quarters. Panel (b) shows the coefficients when we set the log of average non-drug expenditure as the dependent variable. Panel (c) shows the coefficient plot when the dependent variable is the log of the average total medical expenditure. In the regression, we control for the patients' disease type, demographic characteristics, and medical insurance reimbursement generosity. The standard errors are clustered at the hospital level.

Figure 4: Physician-Level Results: Difference between Township Centers and County Hospitals

(a) Drug Expenditure

(b) Non-Drug Expenditure

(c) Total Expenditure

(d) Total Number of Visits
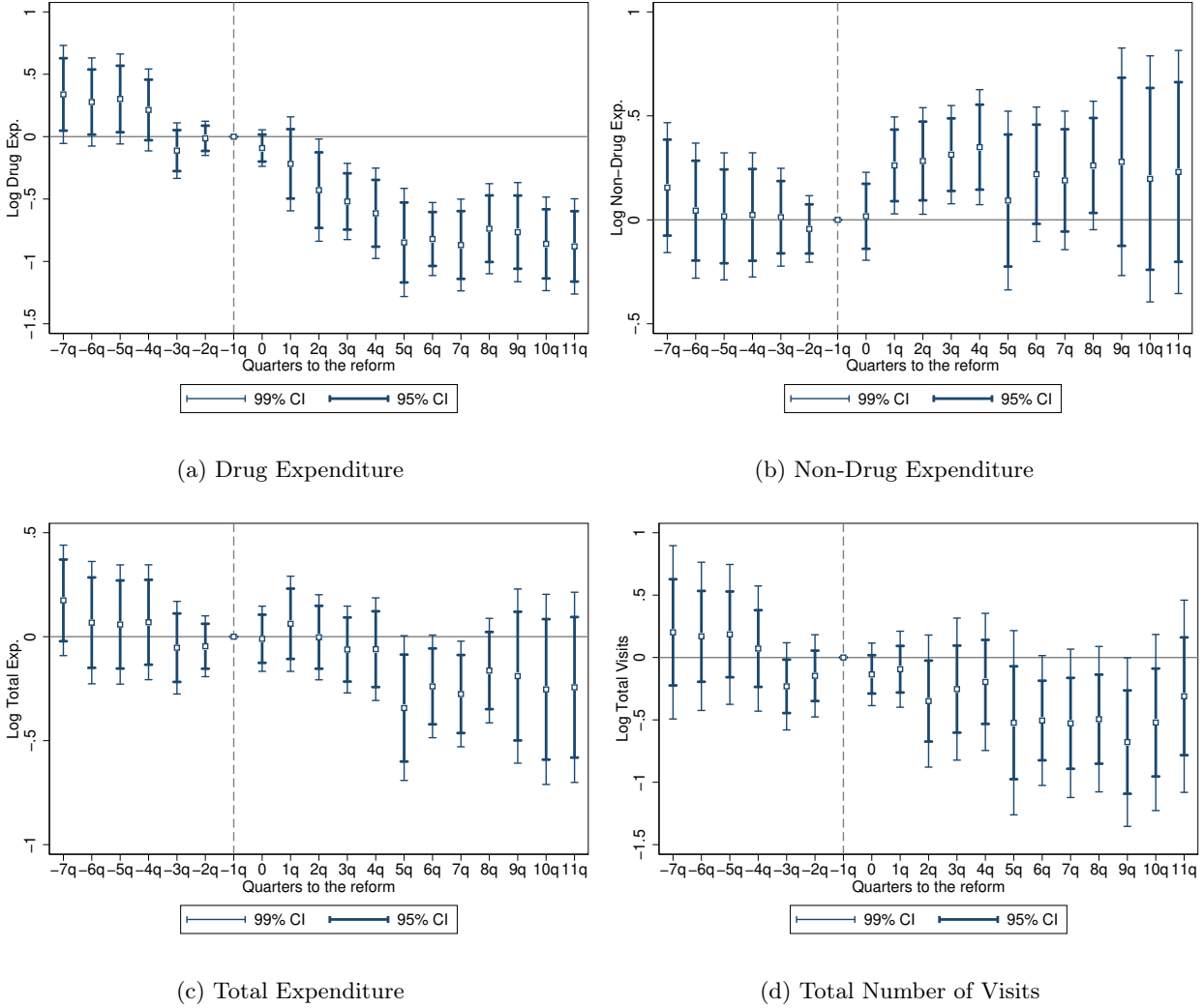
Notes: This figure plots the event study coefficients based on Eq. (20). We calculate the average drug expenditure, non-drug expenditure, total medical expenditure, and total number of visits by quarter for each physician. In Panel (a), we regress the average drug expenditure in logs on the dynamic time indicator such that the effects of the reform would vary across quarters. Panel (b) shows the coefficients when we set the average non-drug expenditure in logs as the dependent variable. Panel (c) shows the coefficient plot of the average total medical expenditure in logs. In Panel (d), we calculate the total number of visits treated by the physician per quarter and then plot the coefficients of the corresponding event study. In the regression, we control for the average medical insurance reimbursement generosity of patients treated by the physician, year fixed effects, season fixed effects, and the physician fixed effects. The standard errors are clustered at the hospital level.

Table 1: Demographic Statistics of Inpatient Patients

	Full Sample			Township Centers			County Hospitals		
	All	Before	After	All	Before	After	All	Before	After
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age (%)									
0-19	5.07	7.70	3.47	4.90	7.61	3.14	5.61	8.05	4.41
20-40	14.43	18.71	11.83	13.36	17.74	10.52	17.89	22.45	15.65
41-64	46.02	45.03	46.63	45.18	44.44	45.66	48.72	47.27	49.44
65+	34.48	28.56	38.07	36.57	30.22	40.67	27.77	22.23	30.5
Gender (%)									
Male	41.49	40.70	41.97	40.99	40.06	41.59	43.10	43.15	43.08
Marital Status (%)									
Never married	8.41	10.20	7.33	7.67	9.92	6.21	10.8	11.27	10.57
Married	71.96	71.09	72.49	71.86	70.61	72.66	72.29	72.91	71.99
Other	19.63	18.72	20.19	20.48	19.47	21.13	16.91	15.81	17.45
Education (%)									
Illiterate	30.16	28.35	31.26	31.91	29.55	33.44	24.54	23.78	24.91
Primary school	40.98	39.28	42.01	41.23	39.5	42.35	40.16	38.42	41.02
Junior middle school	24.92	27.54	23.33	23.36	26.42	21.38	29.93	31.8	29.01
High school	3.31	3.99	2.90	2.95	3.74	2.44	4.47	4.96	4.23
College and above	0.63	0.85	0.50	0.55	0.79	0.39	0.89	1.04	0.82
Obs.	82,356	31,125	51,231	62,791	24,675	38,116	19,565	6,450	13,115

Notes: We define the admissions from January 1, 2009 to September 30, 2010 as occurring before the reform, and the admissions from October 1, 2010 to September 30, 2013 as occurring after the reform. We define *College and above* to include 2-year college, 3-year college, 4-year college, and graduate education. Column (1) summarizes the patients' demographic characteristics for inpatient visits. In Columns (2) and (3), we break down the inpatient sample into pre- and post-reform periods. In Column (4), we summarize the patients' demographic statistics for inpatient visits in township health centers. In Columns (5) and (6), we decompose the township inpatient sample into pre- and post- reform periods. In Column (7), we summarize the patients' demographic statistics for inpatient visits in county hospitals. In Columns (8) and (9), we decompose the county hospital inpatient sample into pre- and post-reform periods.

Table 2: Average Inpatient Experiences in the Township Centers and County Hospitals

Variables	All		Before		After		Diff.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Township Centers							
Total Exp. (RMB)	1,079.66	1,114.29	1,039.13	997.42	1,105.89	1,183.07	66.76***
Drug Exp. (RMB)	482.94	477.16	582.32	464.50	418.60	474.21	-163.72***
Non-Drug Exp. (RMB)	596.59	809.28	456.50	664.46	687.29	878.74	230.79***
Inpatient Duration (Day)	6.90	8.17	6.33	5.80	7.27	9.37	0.93***
Insurance Deductible (RMB)	102.98	28.93	100.50	7.07	104.59	36.60	4.08***
Reimbursement Rate (%)	79.77	7.40	72.71	3.26	84.35	5.49	12.42***
30-day Readmission Rate (%)	2.51	15.65	2.08	14.27	2.79	16.48	0.72***
Panel B: County Hospitals							
Total Exp. (RMB)	3,133.28	4,694.14	2,810.57	4,300.37	3,291.99	4,868.47	481.41***
Drug Exp. (RMB)	1,132.50	1,493.84	1,057.73	1,264.63	1,169.27	1,593.28	111.53***
Non-Drug Exp. (RMB)	2,000.76	3,723.79	1,752.77	3,580.97	2,122.72	3,786.24	369.95***
Inpatient Duration (Day)	12.30	25.41	11.94	21.84	12.47	27.00	0.54*
Insurance Deductible (RMB)	274.89	63.88	283.07	42.83	270.87	71.67	-12.20***
Reimbursement Rate (%)	65.47	12.45	53.11	6.42	71.55	9.95	18.43***
30-day Readmission Rate (%)	2.77	16.41	2.06	14.21	3.12	17.38	1.06***

Notes: We define the admissions from January 1, 2009 to September 30, 2010 as occurring before the reform, and define the admissions from October 1, 2010 to September 30, 2013 as occurring after the reform. The inpatient reimbursement is only qualified for expenditures exceeding a certain threshold. Here we denote that threshold as *insurance deductible,* which might change annually according to the government policies. The reimbursement rate would also change annually according to the local government policies. 30-day readmission is a dummy that takes value 1 if the patient incurred revisits to any medical facility for the same diagnosis within a month of discharge.

Table 3: Impact of the ZMP on Drug, Non-Drug and Total Expenditures: Admission-Level Analysis

	Drug Expenditures		Non-Drug Expenditures		Total Expenditures	
	(1)	(2)	(3)	(4)	(5)	(6)
Town × Post	-0.626***	-0.634***	0.278***	0.272***	-0.108	-0.114
	(0.111)	(0.107)	(0.099)	(0.095)	(0.090)	(0.088)
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Insurance Policy	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	No	Yes	No	Yes
Diagnosis	No	Yes	No	Yes	No	Yes
Adj. R^2	0.278	0.328	0.467	0.521	0.393	0.444
Obs.	82,356	82,356	82,356	82,356	82,356	82,356

Notes: In Columns (1) and (2), the dependent variables are the drug expenditure per visit; in Columns (3) and (4), the dependent variables are the non-drug expenditure per visit; and in Columns (5) and (6), the dependent variables are the total medical expenditure per visit. All dependent variables are in log values. We estimate the regression based on Eq. (17). *Insurance policy* measures the generosity of the insurance, and is calculated as the out-of-pocket expenditures of a typical enrollee based on the population's average inpatient expense last year according to the current year's policy. *Demographics* include age, gender, education and marital status. *Diagnosis* includes a group of dummy variables indicating to which Major Disease Categories the diagnosis belongs. ***, ** and * represent statistical significance at 0.01, 0.05 and 0.1 level, respectively. Standard errors are clustered at the hospital level.

Table 4: Impact of the ZMP on Non-Drug Expenditures by Subcategories

	Infusion	Bed	Nurse	Exam.	Material
	(1)	(2)	(3)	(4)	(5)
Town × Post	0.166	0.261**	0.223	0.352*	0.035
	(0.120)	(0.116)	(0.142)	(0.195)	(0.153)
Hospital FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Insurance Policy	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.245	0.330	0.515	0.398	0.378
Obs.	82,356	82,356	82,356	82,356	82,356

Note: We extend our analysis based on Eq. (17) to the subcategories of examination and service expenditure. We check the total infusion expenditure, bed expenditure, nursing care expenditure, examination and service expenditure and medical device material expenditures separately. The outcome variables are all in log values. In all the regressions, we control for the generosity of the insurance policy. We also control for the demographic characteristics for each patient and the diagnosis in 20 Major Disease Categories.***, ** and * represent represent statistical significance at 0.01, 0.05 and 0.1 level, respectively. Standard errors are clustered at the hospital level.

Table 5: Impact of the ZMP on Drug, Non-Drug, Total Expenditures and Number of Admissions: Physician-Level and Hospital-Level Analysis

	Drug Exp.	Non-Drug Exp.	Total Exp.	Num. Admissions
	(1)	(2)	(3)	(4)
A. Physician-Level Results				
Town × Post	-0.634***	0.230**	-0.135**	-0.359*
	(0.0863)	(0.0915)	(0.0640)	(0.177)
Quarter FE	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes
Adj. R^2	0.703	0.791	0.744	0.608
Obs.	4,060	4,060	4,060	4,060
B. Hospital-Level Results				
Town × Post	-0.549***	0.353***	-0.0581	-0.487***
	(0.107)	(0.113)	(0.0862)	(0.174)
Hospital FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Adj. R^2	0.849	0.904	0.896	0.844
Obs.	487	487	487	487

Note: The upper panel of this table analyzes the physician-level outcome change based on Eq. (19). The dependent variables are calculated at the physician-quarter level. We first calculate the total drug/non-drug/total expenditure and the total number of inpatient admissions treated by that physician in the quarter. Then we divide the total expenditure by the total number of inpatient admissions to calculate the average drug/non-drug expenditure/total expenditure incurred by a specific physician. Then we take the log value of the average expenditures and the total number of admissions as the outcome variable. The lower Panel of this table analyzes the hospital-level outcome change. Similar to the upper panel, we construct the average hospital-quarter level drug, non-drug and total expenditure, as well as the total number of admissions, and take the log values for all the outcome variables. ***, ** and * represent represent statistical significance at 0.01, 0.05 and 0.1 level, respectively. Standard errors are clustered at the hospital level.

Table 6: Impact of the ZMP on Drug Variety, Quantity, Retail Price and Expenditure

	Quant.	Price	Exp.	Variety	Expensive Drug Exp. Share	Expensive Drug Quant. Share
	(1)	(2)	(3)	(4)	(5)	(6)
Town×Post	-0.178**	-0.171**	-0.348***	-0.024	-0.085	-0.061
	(0.081)	(0.081)	(0.055)	(0.062)	(0.059)	(0.052)
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Insurance Policy	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes	Yes	Yes	Yes
Drug-Season FE	Yes	Yes	Yes	-	-	-
Adj. R^2	0.879	0.955	0.810	0.181	0.335	0.177
Obs.	862,402	862,402	862,402	82,356	82,274	82,274

Note: In Columns (1) to (3), each observation is at the visit-drug level. In Column (1) drug quantity is the standardized quantity defined as the total units prescribed in the visit times the package volume for a certain type of drug. In Column (2), the drug retail price is defined as the drug expenditure divided by the standardized prescription quantity in the visit. In Column (3), we calculate the total expenditure for each type of drug during the inpatient visit. In Column (4), each observation is at the inpatient visit level, and the dependent variable is the total number of different types of drugs prescribed in the visit. All outcome variables in Columns (1)-(4) are in logs. In Columns (5) and (6), we define the "Expensive" drug as drugs whose average wholesale price is above the median wholesale price of drugs with the same chemical name during our sample period. In Column (5), we calculate the share of "Expensive" drug expenditure out of the total drug expenditure in the visit. In Column (6), we calculate the share of "Expensive" drug prescription quantity of the total drug prescription quantity in the visit. In all regressions, we control for quarter fixed effects, hospital fixed effects and patients demographics and disease codes, and Columns (1)-(3) control for the drug-season fixed effects as well. ***, ** and * represent statistical significance at the 0.01, 0.05 and 0.1 levels, respectively. Standard errors are two-way clustered at the hospital-drug level for Columns (1)-(3) and are clustered at the hospital level for Columns (4)-(6).

Table 7: Readmission Rate and Inpatient Duration

	Revisit in 2 Weeks		Revisit in a Month		Inpatient Duration	
Mean value	0.016		0.026		8.18	
	(1)	(2)	(3)	(4)	(5)	(6)
Town × Post	-0.006	-0.007	-0.005	-0.005	-0.033	-0.107
	(0.005)	(0.005)	(0.004)	(0.004)	(1.186)	(1.198)
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Payment Policy	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	No	Yes	No	Yes
Diagnosis	No	Yes	No	Yes	No	Yes
Adj. R^2	0.009	0.028	0.011	0.022	0.135	0.148
Obs.	82,356	82,356	82,356	82,356	82,356	82,356

Notes: For each inpatient admission record, we calculate whether the patient incurs another inpatient visit within a certain period of time due to the same disease. In Columns (1) and (2), the dependent variable is a dummy indicating whether the same patient revisited the hospitals for the same diagnosis within 2 weeks, and in Columns (3) and (4), the dependent variable is a dummy indicating whether the same patient revisited hospitals for the same diagnosis within a month. In Columns (5) and (6), the dependent variable is the inpatient duration in logs. The regressions all follow the same specification as Eq. (17). ***, ** and * represent statistical significance at the 0.01, 0.05 and 0.1 levels, respectively. Standard errors are clustered at the hospital level.

Table A1: Disease Distributions in Township Centers and County Hospitals before and after the ZMP

	Before ZMP			After ZMP			ΔDifference
	County	Township	Difference	County	Township	Difference	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Certain infectious and parasitic diseases	0.213	0.014	0.200***	0.224	0.010	0.213***	0.013
Neoplasms	0.042	0.031	0.011***	0.043	0.035	0.008***	-0.003
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	0.002	0.002	0.000	0.004	0.004	0.000	0.000
Endocrine, nutritional, and metabolic diseases	0.008	0.005	0.002*	0.010	0.010	-0.001	-0.003
Mental and behavioral disorders	0.025	0.000	0.025***	0.026	0.001	0.025***	0.000
Diseases of the nervous system	0.004	0.002	0.001	0.003	0.003	0.000	-0.001
Diseases of the eye and adnexa	0.023	0.001	0.022***	0.030	0.001	0.029***	0.007**
Diseases of the ear and mastoid process	0.009	0.015	-0.005***	0.011	0.018	-0.007***	-0.001
Diseases of the circulatory system	0.057	0.056	0.000	0.055	0.062	-0.006*	-0.007
Diseases of the respiratory system	0.177	0.439	-0.262***	0.182	0.457	-0.275***	-0.013
Diseases of the digestive system	0.118	0.175	-0.057***	0.111	0.160	-0.049***	0.008
Diseases of the skin and subcutaneous tissue	0.010	0.010	0.000	0.010	0.009	0.001	0.001
Diseases of the musculoskeletal system and connective tissue	0.023	0.025	-0.002	0.030	0.028	0.001	0.003
Diseases of the genitourinary system	0.106	0.121	-0.015**	0.086	0.107	-0.022***	-0.007
Pregnancy, childbirth, and the puerperium	0.006	0.002	0.004***	0.005	0.002	0.003***	0.000
Congenital malformations, deformations, and chromosomal abnormalities	0.001	0.001	0.000	0.002	0.000	0.002***	0.002**
Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	0.011	0.023	-0.012***	0.015	0.021	-0.006***	0.005*
Injury, poisoning, and certain other consequences of external causes	0.140	0.066	0.074***	0.131	0.057	0.074***	0.000
External causes of morbidity and mortality	0.013	0.004	0.009***	0.012	0.006	0.007***	-0.003
Factors influencing health status and contact with health services	0.010	0.008	0.002	0.010	0.010	0.000	-0.002

Note: We define January 1, 2009 - September 30, 2010 as before ZMP, and define October 1, 2010 - September 30, 2013 as after the ZMP. We first calculate the mean disease rate for 20 disease categories. Then we calculate the difference between township center and county hospitals both before and after the ZMP. Finally we check whether such difference has been consistent over time.

Table A2: Robust Check: Patients' Risk Score

	Patients' Risk Score		
	(1)	(2)	(3)
Town × Post	0.516	0.188	-0.122
	(0.434)	(0.542)	(0.413)
Hospital FE	No	Yes	Yes
Quarter FE	No	Yes	Yes
Insurance Policy	No	Yes	Yes
Demographics	No	No	Yes
Diagnosis	No	No	Yes
Adj. R^2	0.010	0.058	0.494
Obs.	82,356	82,356	82,356

Note: We analyze Eq. (17) by setting patient's risk score as the outcome variable to check the robustness of the sample distribution before and after the ZMP. To calculate the patients' risk score, we project the patients' total medical expenditure per year on their age, gender and disease codes. Then we normalize their expenditure by the average annual medical expenditure across all patients in that county as the risk score. ***, ** and * represent statistical significant at the 0.01, 0.05 and 0.1 levels, respectively. The standard errors are clustered at the hospital level.

Table A3: Robustness Checks: Alternative Sample Restrictions

	Including the Two Largest County Hospitals			Excluding 6 Months Prior to the Reform			Excluding Private Hospitals		
	Drug Exp.	Non-drug Exp.	Total Exp.	Drug Exp.	Non-drug Exp.	Total Exp.	Drug Exp.	Non-drug Exp.	Total Exp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Town × Post	-0.514***	0.173**	-0.201***	-0.629***	0.281**	-0.119	-0.666***	0.263*	-0.173
	(0.069)	(0.076)	(0.061)	(0.094)	(0.118)	(0.101)	(0.068)	(0.143)	(0.103)
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Insurance Policy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.957	0.581	0.536	0.358	0.524	0.453	0.294	0.504	0.392
Obs.	127,249	127,249	127,249	72,801	72,801	72,801	72,642	72,642	72,642

Note: We analyze Eq. (17) using different sample specifications to check the robustness of the estimation results. In Columns (1)-(3), we analyze the sample results by including the two largest county hospitals in our sample. In Columns (4)-(6), we analyze the sample results excluding observations that took place within 6 months prior to the reform. In Columns (7)-(9), we analyze the sample excluding the four private hospitals. In all the regressions, we set the log value of drug expenditure, non-drug expenditure, and the total medical expenditure as the dependent variable, and include the patients' demographic characteristics, disease code, and reimbursement rate as the independent variable. ***, ** and * represent statistical significance at the 0.01, 0.05 and 0.1 levels, respectively. The standard errors are clustered at the hospital level.

Table A4: Robustness Check: Month Fixed Effects

	Drug Exp.		Non-Drug Exp.		Total Exp.	
	(1)	(2)	(3)	(4)	(5)	(6)
Town × Post	-0.634***	-0.631***	0.272***	0.270***	-0.114	-0.115
	(0.107)	(0.106)	(0.095)	(0.095)	(0.088)	(0.088)
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	-	Yes	-	Yes	-
Month FE	-	Yes	-	Yes	-	Yes
Insurance Policy	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.328	0.329	0.521	0.522	0.444	0.445
Obs.	82,356	82,356	82,356	82,356	82,356	82,356

Note: We analyze Eq. (17) by including month fixed effects to check the robustness of the estimation results. In Columns (1), (3), and (5), we show our baseline regression results. In Columns (2), (4), and (6), we include month fixed effects to control for more cyclical changes. ***, ** and * represent statistical significance at the 0.01, 0.05 and 0.1 levels, respectively. The standard errors are clustered at the hospital level.