From Gray Income to Health Gains: The Impact of Anti-Corruption Inspections in Chinese Hospitals

Ruiqi Ding, Aljoscha Janssen, Jessica Ya Sun, Xuan Zhang*

July 2025

Abstract

We investigate how anti-corruption enforcement affects physician agency and patient welfare. Exploiting the staggered rollout of hospital inspections in China, we analyze 758,972 obstetrics and gynecology admission records from a central province between 2015 and 2017. We find that inspections led to an increase in aggregate healthcare spending, driven by an expansion in the quantity of admissions rather than rising costs per case. Crucially, this expansion was accompanied by significant improvements in neonatal and maternal health outcomes. Mechanism analyses reveal that inspections corrected incentive distortions: providers substituted away from pharmaceutical rents toward service provision, reduced length of stay, and increased prenatal admissions. Furthermore, the enforcement induced the exit of rent-seeking physicians. A welfare calculation suggests that the monetary value of the health gains substantially exceeds the marginal financial costs, indicating that anti-corruption oversight in healthcare can generate large net social benefits. (JEL I11, I18, D73, J13)

Keywords: Anti-corruption enforcement; Physician incentives; Informal payments; Maternal and neonatal outcomes; Health care utilization.

*Ding: Singapore Management University (email: ruiqi.ding.2024@phdecons.smu.edu.sg); Janssen: Singapore Management Univeristy (email: ajanssen@smu.edu.sg); Sun: Huazhong University of Science and Technology (email: jessicasunya@hust.edu.cn); Zhang: Singapore Management University (email: xuanzhang@smu.edu.sg)

1 Introduction

Informal patient payments—ranging from ”red envelopes” and kickbacks to other off-schedule transfers—remain pervasive in hospital care and are widely viewed as corrosive to equity and efficiency in low- and middle-income countries (Lewis, 2007; Chereches̡ et al., 2013). Global health commentary emphasizes that such corruption absorbs substantial resources and impedes progress toward universal coverage, implying that effective control measures could yield significant welfare gains (García, 2019). This phenomenon is also salient in China, where ”red packets” and kickbacks from pharmaceutical firms to physicians persist despite regulatory prohibitions (Liu et al., 2020; Xu & Yuan, 2022; Fu et al., 2023). Yet, despite strong reasons to expect that gray income distorts provider effort and access, especially for maternal and neonatal care that is time-sensitive and capacity-constrained, there is little causal evidence on policies that remove providers' informal revenue streams within hospitals at scale, or on the downstream effects on utilization and clinical outcomes (Vian, 2020). Existing economic work has primarily evaluated formal payment reforms and audit programs—often in high-income settings and focused on billed (official) revenues—rather than the curtailment of informal earnings (Dafny, 2005; Clemens & Gottlieb, 2014; Fang et al., 2021; Gupta, 2021; M. Shi, 2024). We fill this gap by analyzing a large-scale anti-corruption campaign targeting gray income in a developing economy and tracing its downstream consequences on hospital behavior and patient health.

In this paper, we evaluate a large anti-corruption inspection campaign designed to dismantle gray income channels and enforce compliance within Chinese hospitals. To identify the causal effect of these inspections on physician behavior and health outcomes, we compile a novel administrative dataset on inpatient obstetrics and gynecology (OBGYN) care, linking monthly hospital-level aggregates with individual admission records. We exploit the staggered rollout of inspections across hospitals to estimate event-study and difference-in-differences models. Our identification strategy relies on the variation in the timing of inspections, which is uncorrelated with short-run OBGYN demand shocks. This setting allows us to compare changes in hospital outcomes (admissions, intensive vs. extensive margins of spending) and clinical outcomes (neonatal and maternal health) in inspected hospitals relative to those not yet inspected. Our study offers two distinct innovations: first, we treat the campaign—which targeted commercial bribery and off-schedule payments—as a shock to the informal incentive structure itself, rather than a change to official price schedules; second, we comprehensively trace the effects from provider supply responses to health outcomes.

We document that inspections led to a substantial increase in OBGYN outcomes. Following an inspection, total monthly OBGYN expenditure per hospital rose by approximately RMB 295,000 [1] (a 14% increase over the pre-inspection mean of RMB 2.1 million). This aggregate growth was

[1] All magnitudes are derived from administrative data. Monetary values are expressed in constant 2015 RMB, deflated using the Consumer Price Index (CPI).

driven predominantly by prenatal (ante-partum) care, where spending surged by 68%. Crucially, this expansion occurred almost entirely at the extensive margin: the number of prenatal admissions increased by approximately 17 patients per month (a 70% rise from a baseline of 24), while expenditure per admission remained statistically unchanged. Furthermore, we find that this expansion in access translated into tangible improvements in health outcomes. For newborns, the rate of NICU admissions fell by 6.2 cases per 1,000 deliveries, while the incidence of low birth weight and very low birth weight declined by 6.4 and 2.3 cases per 1,000 deliveries, respectively. We also observed a modest increase in average birth weight (approximately 13 grams). For mothers, the inspections led to a reduction in delivery complications, including significant declines in perineal lacerations, cervical lacerations, puerperal sepsis, and pregnancy-induced hypertension.

We investigate the mechanisms driving these health improvements and isolate two primary channels: behavioral substitution by incumbents and compositional shifts via selective attrition. First, the crackdown on illicit 'gray income' (e.g., pharmaceutical kickbacks) induced physicians to substitute toward legitimate, service-intensive provision—specifically prenatal care—to offset income shocks. This interpretation is supported by three findings: (i) treatment effects are concentrated in hospitals with high ex-ante rent-seeking norms (proxied by baseline C-section rates) and among senior physicians with greater discretionary power; (ii) hospital expenditures exhibit a structural break, characterized by a sharp decline in drug spending and a compensatory rise in service fees; and (iii) hospitals accommodated supply expansion through intensive-margin efficiency gains, significantly reducing average length of stay. Second, we document a selective attrition channel wherein inspections precipitated the exit of 'low-quality' types. Physicians with historically high drug intensities and C-section rates self-selected out of the public sector prior to the policy shock.

Additionally, We confront and rule out two competing mechanisms. First, the results are not driven by strategic patient sorting; contrary to a 'cherry-picking' hypothesis, treated hospitals increased admissions of high-risk pregnancies. Second, the effects do not reflect a mechanical upgrade in monitoring technology. We find that treatment effects load primarily on hospitals that possessed high-quality monitoring systems prior to the inspection. Together, these findings attribute welfare gains to corrected financial incentives and the market exit of rent-seeking agents."

Finally, to benchmark the welfare implications of the policy, we perform a conservative back-of-the-envelope cost–benefit analysis. We weigh the aggregate increase in medical expenditures against the social value of improved neonatal and maternal health outcomes. Our calculations yield a net social benefit of approximately RMB 43.4 million (USD 6.16 million), driven primarily by the short-term medical savings from averted very low birth weight cases and the long-term human capital returns associated with increased birth weight. These estimates likely represent a lower bound on the true welfare surplus, as they exclude the economic value of reductions in infant mortality and long-term disability.

Our findings contribute to three strands of literature. First, we extend the literature on physician agency and the supply-side response to financial incentives. A rich body of empirical work documents that providers adjust treatment intensity and case mix to offset income shocks or exploit payment weights (Gruber & Owings, 1996; Dafny, 2005; Clemens & Gottlieb, 2014). In the context of China, recent studies on the Zero Mark-Up Drug Policy (ZMDP) highlight a "whac-a-mole" phenomenon: restricting drug revenues leads physicians to substitute toward diagnostic services to stabilize total income (Wu, 2019; Fang et al., 2021; J. Shi et al., 2023). However, these studies predominantly focus on changes to formal price schedules or reimbursement rates. We depart from this literature by examining a shock to informal (gray) income constraints. We show that, unlike price adjustments—which typically generate intensive-margin responses (e.g., more tests per patient)—constraining gray income induces an extensive-margin adjustment. Providers respond to the crackdown not only by reallocating billing, but by expanding inpatient admissions, effectively substituting illicit rents with formal clinical effort.

Second, we contribute to the debate on the welfare implications of supply-induced demand. A central question in health economics is whether provider-driven volume expansions represent wasteful induced demand or the fulfillment of unmet need. Evidence remains mixed: while some reimbursement-driven expansions yield marginal health benefits (Mohanan et al., 2021), others suggest that removing profit distortions can curb over-treatment without compromising safety (Currie et al., 2014). Leveraging rich administrative data on obstetrics—a high-stakes setting for health capital formation—we show that the supply response induced by anti-corruption efforts is welfare-enhancing. By effectively reallocating physician effort from rent-seeking to risk monitoring and patient care, the policy generates significant returns in neonatal and maternal health, mirroring the benefits seen in programs that expand access to underserved populations (Currie & Gruber, 1996; Björkman & Svensson, 2009).

Finally, we speak to the political economy of state capacity and corruption control in public service delivery. The efficacy of monitoring is theoretically ambiguous: while it can reduce leakage (Olken, 2007), strict enforcement may also backfire if it causes agents to retreat into inefficiency or reduce effort (Gerardino et al., 2024; M. Shi, 2024). Our study identifies a setting where anti-corruption enforcement succeeds by realigning incentives rather than merely paralyzing the bureaucracy. We provide evidence that cracking down on gray income pushes providers toward formal markets and measurable service improvements, offering a "proof of concept" that state capacity can be strengthened without sacrificing service quantity or quality.

The paper proceeds as follows. Section 2 details the institutional background and the inspection campaign. Sections 3 and 4 describe the data and identification strategy, respectively. Section 5 presents the baseline estimates, while Section 6 explores the mechanisms driving these results. Section 7 offers a cost-benefit analysis, and Section 8 concludes.

2 Policy Context

2.1 Informal Payments, Hospital Governance, and Physician Incentives

Since the 1980s, China's healthcare system has undergone profound reforms paralleling the nation's broader transition from a planned to a market-oriented economy. During this period, the government systematically reduced direct fiscal subsidies, compelling public hospitals to become financially self-sustaining. To reconcile this fiscal retrenchment with the goal of affordable access, policymakers maintained administratively suppressed prices for routine services and procedures while permitting hospitals to levy a 15 percent markup on pharmaceutical sales. This asymmetric pricing structure introduced a classic principal–agent problem: under conditions of information asymmetry, physicians—seeking to maximize personal income—faced strong incentives to induce demand through excessive drug prescriptions and redundant diagnostics. This arrangement, while balancing hospital budgets, undermined allocative efficiency and patient welfare.

The comprehensive healthcare reform launched in 2009 sought to correct these distorted incentives. The centerpiece of this effort was the Zero Mark-up Drug Policy (ZMDP), which abolished the pharmaceutical markup, tightened regulations on drug expenditure shares, and standardized reimbursement lists. These measures were designed to sever the direct link between provider compensation and drug sales. A growing body of empirical evidence suggests the reforms successfully lowered drug prices and weakened the "drug-financed" hospital model (Yip et al., 2010; Wu, 2019; Fang et al., 2021). However, structural imbalances persisted. Pharmaceutical prices remained heavily regulated, intensifying volume-based competition among firms, while physician base salaries remained suppressed relative to their workload and human capital. This "high-intensity, low-reward" equilibrium preserved the high marginal utility of supplementary income, creating a fertile environment for unintended behavioral responses (Xu & Yuan, 2022).

Following the elimination of the legally sanctioned markup, the market for pharmaceutical rents reorganized. Firms adapted by substituting institutional-level profits with direct, illicit financial relationships with individual physicians. Evidence indicates that in the absence of formal profit-sharing mechanisms, pharmaceutical companies increasingly resorted to kickbacks to incentivize prescriptions (Yip et al., 2010). Consequently, rents previously captured formally by the hospital were transformed into physician-level "gray income," often structured as volume-based commissions. Simultaneously, on the demand side, patients facing severe information frictions and quality uncertainty turned to informal payments—commonly known as "red pockets"—to secure priority access or perceived higher effort. These payments, typically exchanged prior to surgery or during treatment, further entrenched informal financial flows within the clinical encounter.

Ultimately, the Chinese experience illustrates the hydraulic nature of financial incentives in regulated markets. Eliminating a single source of "legal but distortionary" revenue does not ensure incentive alignment if the underlying participation constraint—adequate physician compensation—is not met. Instead of disappearing, rents were reallocated from compliant, hospital-level profits to illicit, informal physician income. This shift not only pushed economic activity outside legal boundaries but also heightened the agency costs associated with corruption, with ambiguous implications for healthcare quality and social welfare.

2.2 Anti-corruption and the Turn toward Active Hospital Governance

At the 18th National Congress of the Chinese Communist Party (CPC) in November 2012, the incoming leadership identified pervasive official corruption as a fundamental threat to the Party's legitimacy and state capacity, elevating anti-corruption to a core governing priority. This commitment was formalized in December 2012 with the promulgation of the ”Eight-Point Regulation” (*baxiang guiding*), which marked the launch of a sweeping nationwide anti-corruption campaign. Distinguished by its scale, intensity, and sustained enforcement, the initiative is widely regarded as the most comprehensive anti-corruption effort in the history of the People's Republic of China.

The campaign was unprecedented in both its reach and targeting strategy. Explicitly encompassing both ”tigers” (senior officials) and ”flies” (grassroots cadres), the enforcement regime signaled that no administrative rank conferred immunity. Crucially, scrutiny extended beyond party and government organs to include state-owned enterprises, the military, and key public service sectors such as healthcare and education. This broad scope reflected an objective surpassing the mere discipline of political elites: the campaign aimed to dismantle the rent-seeking mechanisms embedded in the operational fabric of economic and social institutions—mechanisms that had proliferated during decades of rapid, decentralized growth.

Implementation was characterized by rigorous, top-down enforcement. Spearheaded by the Central Commission for Discipline Inspection (CCDI) under a mandate of zero tolerance, the strategy utilized repeated inspection waves, unannounced audits, and expanded investigative powers. Since 2013, hundreds of thousands of officials—ranging from ministerial leaders to local functionaries—have been investigated or disciplined. Unlike previous episodic crackdowns, this effort has been institutionalized over more than a decade, reshaping political incentives, bureaucratic norms, and governance structures, with profound implications for the delivery of public services.

Notably, a defining characteristic of the post-2012 campaign, relative to prior initiatives, was its intense and specific targeting of the healthcare sector. Data from publicly disclosed criminal court verdicts corroborate this shift: the aggregate volume of first-instance bribery convictions surged sharply following 2012. Moreover, conditional on a corruption case being prosecuted, the specific share of cases involving medical corruption has exhibited a sustained year-on-year upward trend since the campaign's inception, increased from 2% in 2011 to 10% in 2018, as shown in Figure 1.

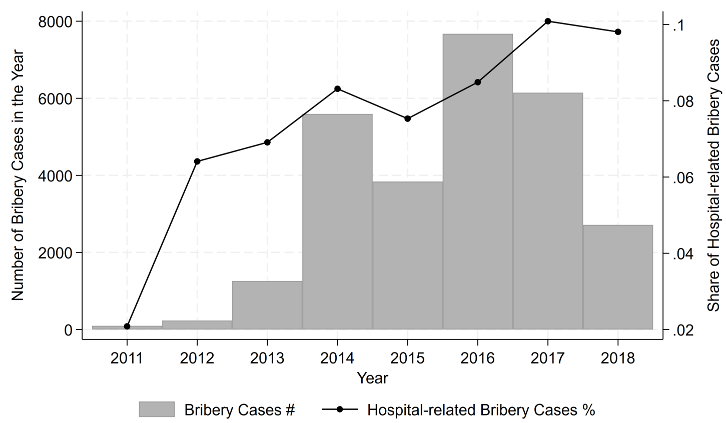


Figure 1: Annual Trends in Corruption Cases and the Hospital-Related Share

Notes: This figure plots the annual evolution of bribery litigation in China from 2011 to 2018 using data sourced from China Judgements Online. The gray histogram (left vertical axis) displays the total number of bribery cases per year, while the solid line (right vertical axis) indicates the annual share of these cases that are hospital-related. A case is defined as "hospital-related" if textual analysis of the case description identifies the explicit involvement of a hospital or physicians. The figure highlights the temporal relationship between aggregate corruption cases and the relative prominence of healthcare-sector corruption.

Within the healthcare sector, enforcement specifically targeted the "gray income" networks linking hospitals, physicians, and pharmaceutical firms. Regulatory oversight was intensified across the entire delivery chain, spanning drug procurement, prescribing behavior, and patient interactions. This marked a pivotal transition from de facto tolerance to active prohibition: practices previously regarded as customary were reclassified as actionable disciplinary violations or criminal offenses. Consequently, hospital inspections were transformed into credible, high-stakes events. For administrators and physicians long accustomed to supplementing their earnings through these informal channels, the campaign introduced severe downside risks, as inspection outcomes could now dictate career trajectories, stall promotions, or trigger legal sanctions.

2.3 The Large Hospital Inspection Program as a Governance Shock

In 2005, the Chinese government introduced the *Large Hospital Inspection Work Plan*[2] (hereafter, Hospital Inspection), marking the first nationwide effort to institutionalize systematic oversight of public hospitals. This initiative emerged in response to mounting concerns regarding hospital governance, medical ethics, and the efficiency of public healthcare delivery amidst a period of rapid system expansion and deteriorating public trust.

Since its inception, the program has undergone five distinct rounds of implementation. The first two rounds (2005 and 2010) were managed primarily by the Ministry of Health. The third round, launched in 2015, represented a significant institutional departure. It established a dual-level

[2]Formally titled *daxing yiyuan xuncha gongzuo fang'an*, jointly issued by the National Health and Family Planning Commission (NHFPC, the predecessor of the National Health Commission) and associated ministries.

structure jointly organized by central and provincial authorities: the National Health Commission (NHC) selected a subset of hospitals for direct inspection, while provincial health commissions were mandated to inspect the remaining large hospitals within their jurisdictions. Each inspection typically spanned 5–10 days, was led by senior officials, and targeted a rotating set of hospitals annually. This round formally entrenched the Hospital Inspection as a regular monitoring mechanism, a design subsequently retained in the fourth (2019–2022) and fifth (2023–2026) rounds.

The focus and enforcement intensity of the program have evolved substantially. Early inspections prioritized administrative compliance—specifically financial management, personnel administration, and procurement protocols. Sanctions during these initial phases were relatively weak: in 2005, noncompliant hospitals received only rectification recommendations without formal penalties. Even by 2010, repercussions were largely limited to internal warnings or duty adjustments, with public reprimands or removals remaining rare. By contrast, the third round marked a sharp discontinuity, coinciding with the nationwide anti-corruption campaign launched in 2013. Beginning in 2015, inspections explicitly targeted the "gray income" ecosystem, including commercial kickbacks between pharmaceutical firms and providers, informal patient payments (*hongbao*), and other rent-seeking practices. Enforcement became markedly stricter. Serious violations faced public disclosure; hospital leaders were subject to Party disciplinary sanctions or removal, and egregious cases were referred for criminal prosecution. Furthermore, hospitals with systemic governance failures faced coordinated cross-agency penalties, such as intensified reimbursement audits by the National Healthcare Security Administration. As inspection design and enforcement capacity matured, the program exerted increasingly visible and consequential constraints on hospital behavior.

Following two pilot phases, the 2015 inspection formally institutionalized anti-corruption oversight through a standardized three-step protocol. First, hospitals were required to conduct internal self-assessments and submit reports detailing existing problems and corrective actions. Second, inspection teams—comprising government officials, medical experts, and Party discipline inspectors—conducted on-site evaluations via document reviews, staff interviews, and random audits of medical and financial records. Third, results were publicly released, and hospitals were mandated to implement rectification measures subject to follow-up verification. This hybrid design combined top-down administrative enforcement with bottom-up self-discipline. The 2015 protocol emphasized three core dimensions: (i) the implementation of "Party Committee leadership under the responsibility of the hospital president," which strengthened political oversight over strategic decisions and personnel; (ii) strict enforcement of the "Nine Prohibitions" (*jiu bu zhun*), which explicitly banned kickbacks, improper billing, and supply-induced demand; and (iii) rigorous compliance with budgetary discipline and procurement regulations.

This study primarily focuses on the anti–informal income component of the 2015 Hospital

Inspection as implemented in a central Chinese province. Our sample comprises 44 large public hospitals—primarily tertiary and key secondary institutions—subject to a staggered rollout. Fifteen hospitals were inspected in November 2015, fifteen in September 2016, and the remaining fourteen in September 2017. Three features make this round particularly salient. First, occurring in the wake of the 2012 national anti-corruption campaign, this round subjected hospitals to an unprecedented intensity of scrutiny and political pressure. Second, it institutionalized external oversight in a sector traditionally characterized by high autonomy and strong revenue-driven incentives. Third, by explicitly linking inspection outcomes to hospital reputation, leadership evaluation, and resource allocation, the program fundamentally altered the incentive structure for both administrators and physicians. From an economic perspective, these inspections constitute a regulatory shock that sharply increased the cost of informal income generation while raising the marginal returns to compliance with formal standards of care.

3 Data

3.1 Data Source

Our analysis relies on a novel database constructed by linking two administrative datasets provided by the Health Commission of a central Chinese province. The merged dataset covers the universe of public secondary and tertiary hospitals in the province from 2014 to 2017.[3]

The first dataset is the Health Statistics Yearbooks (HSY), a hospital-level panel containing detailed information on hospital operations and capacity. Key variables include total revenue and its composition (disaggregated by outpatient vs. inpatient care, and by revenue source such as pharmaceuticals, consumables, and medical services), as well as capacity measures like bed count, staffing levels (physicians and nurses), and IT infrastructure. We utilize these variables as baseline controls to account for hospital heterogeneity in scale and quality.

The second and primary data source consists of patient-level discharge records from the Obstetrics and Gynecology (OBGYN) departments of 164 hospitals. This administrative microdata spans the universe of OBGYN inpatient admissions in the province during the sample period, totaling approximately 1.09 million observations. For each admission, we observe unique hospital identifiers, admission and discharge timing, patient demographics, and itemized expenditures (e.g., pharmaceuticals, diagnostic tests, and consumables).

A key feature of discharge records is the granularity of the clinical information. The records contain up to nine diagnostic codes (ICD-10) and seven procedure codes (ICD-9-CM3), capturing

[3]China classifies hospitals into three tiers based on capacity and function. We restrict our sample to Tier 2 (secondary) and Tier 3 (tertiary) hospitals, as Tier 1 (primary) facilities serve mainly as community health centers with limited inpatient capacity and rarely perform the complex surgical procedures relevant to our study.

the full diagnostic trajectory from the initial outpatient assessment to admission and final discharge and allowing us to track the evolution of the patient's condition. For delivery admissions, the data additionally report neonatal birth weight and specific diagnoses for newborns. This rich clinical detail allows us to precisely categorize admission types and construct measures of patient health outcomes, which we describe in detail in the following subsection.

3.2 Outcome Construction

We construct two primary sets of outcome variables to capture the impacts of inspections on hospital supply behavior and patient welfare: hospital outcomes metrics and patient health outcomes.

Hospital outcomes. To measure hospital outcomes, we aggregate administrative records to the hospital-month level. Our primary outcome is total monthly inpatient expenditure within the OBGYN department. We further decompose supply responses along two margins to identify the mechanisms underlying aggregate expenditure changes: the extensive margin (monthly admission volume) and the intensive margin (average expenditure per admission). Given the specialized nature of obstetric care, we disaggregate these measures by clinical function into pregnancy-related and non-pregnancy-related admissions. Pregnancy-related admissions are further partitioned into four mutually exclusive categories: pregnancy loss, prenatal care, delivery, and postnatal care[4]. This disaggregation enables us to assess whether supply responses are uniform across service types or concentrated in areas characterized by greater physician discretion.

Patient Health Outcomes. To evaluate whether changes in service provision translate into welfare effects, we construct a vector of health indicators using patient-level delivery records. For neonatal health, we employ both continuous and binary metrics. These include birth weight (in grams), indicators for Low Birth Weight (LBW) and Very Low Birth Weight (VLBW), and an indicator for Neonatal Intensive Care Unit (NICU) admission[5]. Following Best et al. (2017), we also generate a distributional measure indicating whether birth weight falls within one standard deviation of the sample mean, capturing convergence toward optimal health. For maternal health, we identify

[4]We employ a hierarchical classification algorithm to define mutually exclusive admission types. First, we classify an admission as *pregnancy-related* if any diagnosis code falls within ICD-10 Chapter XV (codes beginning with 'O'); otherwise, it is classified as *non-pregnancy-related*. Second, within the pregnancy-related sample, we identify *delivery admissions* based on the presence of a recorded neonatal birth weight, which serves as a definitive marker for a birth event. Third, for the remaining non-delivery pregnancy admissions, we assign categories based on diagnosis codes: (i) *pregnancy loss* includes admissions with codes O00–O08 (ectopic pregnancy, hydatidiform mole, and other abortive outcomes) or explicit diagnoses of abortion, ectopic gestation, or stillbirth; (ii) *prenatal admissions* are identified by codes O10–O48 (maternal disorders related to pregnancy) or Z32–Z36 (antenatal screening and supervision); and (iii) *postnatal admissions* comprise the residual category, primarily capturing complications of the puerperium.

[5]We define *Low Birth Weight* (LBW) and *Very Low Birth Weight* (VLBW) based on standard clinical thresholds of <2,500g and <1,500g, respectively. Lacking a direct administrative flag for intensive care utilization, we construct a proxy for *NICU admission* using a composite algorithm. Following established protocols, this measure identifies NICU admissions based on a birth weight threshold of <2,000g (Braun et al., 2020; Pursley & Zupancic, 2020) as well as the presence of high-severity diagnostic codes identified by Vance et al. (2023).

adverse outcomes based on discharge diagnoses. Key indicators include obstetric trauma (perineal and cervical lacerations), hypertensive disorders (including eclampsia), and severe complications such as sepsis and Acute Respiratory Distress Syndrome (ARDS)[6].

3.3 Sample Construction and Descriptive Statistics

We impose a series of restrictions to harmonize the data with the institutional context of the inspection campaign and to ensure measurement reliability. First, we restrict the study window to the period from January 2015 to December 2017. We exclude 2014 due to substantial missing values in key variables. This restriction retains 929,837 admission records from the original 164 hospitals, with approximately 47% of admissions occurring in hospitals eventually subject to inspection. Second, we exclude small-scale hospitals, defined as those with fewer than 1,000 OBGYN discharges over the three-year period or fewer than 100 beds across all departments. This restriction mitigates measurement error arising from idiosyncratic volatility in low-volume facilities and removes approximately 2% of observations. Third, to ensure data quality for our health outcomes, we drop hospitals that failed to report birth weight information for more than half of the study months. We also trim hospital-month observations with fewer than 10 admissions to avoid outliers driven by the "small denominator" problem. Our final analytical sample consists of 758,972 admission records from 107 hospitals spanning January 2015 to December 2017. Of these, 32 hospitals underwent inspection (accounting for 398,813 admissions), while the remaining 75 were never inspected (contributing 360,530 admissions).

Table A1 reports summary statistics for hospital outcomes and health outcomes. The first four columns stratify the sample by inspection timing (2015 cohort, 2016 cohort, 2017 cohort) and the control group (never inspected), while Column 5 presents full sample statistics. In the pooled sample (Column 5), the average Cesarean section rate is approximately 37%, the average birth weight is 3,327 grams, and the average length of stay is 6.2 days. Severe adverse events are rare: the VLBW rate is 0.57%, and maternal mortality is 0.033%. In terms of hospital outcomes, the average hospital admits 210 OBGYN patients per month, 82% of which are pregnancy-related. Monthly OBGYN revenue averages RMB 1.06 million, with drugs accounting for 17% of total costs. Comparing across columns reveals that inspected hospitals are generally larger than non-inspected hospitals, consistent with the policy's focus on major public facilities. Among the treated groups, the 2016 cohort is the largest, with an average of 452 monthly admissions.

Figure 2 illustrates the unconditional trends in our primary outcome—aggregated monthly

[6]We define adverse maternal outcomes based on ICD-10 diagnosis codes recorded during delivery admissions. Specifically, we construct indicators for: (i) *perineal lacerations*, defined by codes O70.1–O70.3; (ii) *cervical lacerations*, defined by code O71.3; (iii) *hypertensive disorders*, encompassing gestational hypertension and eclampsia (codes O14–O16 and related subcategories); and (iv) *severe maternal complications*, identified by the presence of specific diagnoses for sepsis or Acute Respiratory Distress Syndrome (ARDS).

OBGYN expenditure—for both inspected and uninspected hospitals. The x-axis denotes time relative to the inspection month ($t = 0$). For the control group, we randomly assign placebo inspection dates drawn from the distribution of actual inspection timings. Two distinct patterns emerge. First, in the pre-inspection period ($t < 0$), expenditure trends are stable and parallel across treated and control groups, lending credibility to the validity of our identification strategy. Second, immediately following the inspection ($t > 0$), the treated group exhibits a marked upward break in trend, diverging progressively from the control group. This persistence implies that the inspection induced a structural shift in hospital operations rather than a mere transitory shock. While these raw patterns do not adjust for patient case mix or time-varying shocks, they provide compelling motivating evidence for the formal econometric analysis that follows.

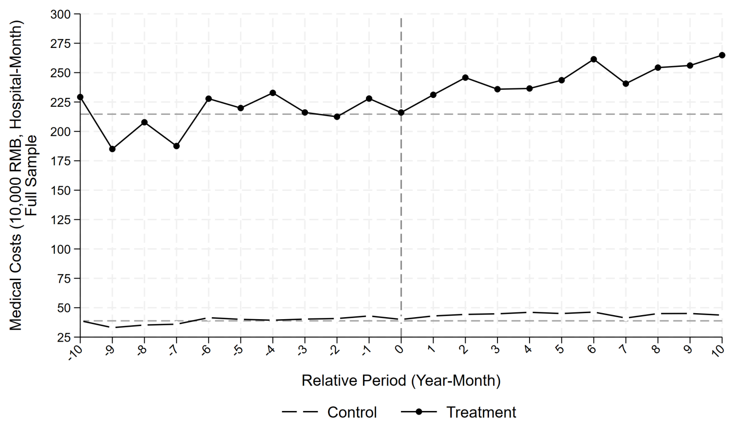


Figure 2: Raw Trend of Aggregated Expenditure between Inspected and Uninspected Hospitals

Notes: This figure illustrates the raw trends in aggregated medical expenditure at the hospital-month level for the full sample. The vertical axis measures expenditure in 10,000 RMB. The solid line depicts the trajectory for inspected hospitals (Treatment), while the dashed line represents uninspected hospitals (Control). The horizontal axis indicates the relative period (in months) centered on the inspection date ($t = 0$). For the control group, placebo inspection dates were randomly assigned based on the temporal distribution of actual inspections to allow for a direct comparison. The figure plots the evolution of costs from 10 months prior to 10 months following the inspection (or placebo) date.

4 Empirical Strategy

4.1 Identification Strategy

Given the institutional features of China's hospital inspection program, our identification strategy exploits the staggered rollout of province-level anti-corruption inspections as a quasi-natural experiment. This design generates plausibly exogenous variation in the timing of on-site scrutiny across hospitals within the province.

The central identifying assumption is that the timing of the inspection is uncorrelated with unobserved determinants of the outcome variables. In our context, the selection and scheduling of inspections were governed by top-down administrative mandates and broader political considerations, rather than being endogenous responses to hospital-specific short-term financial performance or idiosyncratic fluctuations in health outcomes. Consequently, the precise timing of the "treatment" is plausibly orthogonal to high-frequency changes in the monthly operations of OBGYN departments. This institutional setting supports the parallel trends assumption: in the absence of the inspection shock, outcomes in treated and control hospitals would have evolved along parallel trajectories.

Accordingly, we employ a difference-in-differences (DID) framework. We estimate the causal impact of inspections by comparing the changes in outcomes for treated hospitals before and after the inspection relative to the changes in hospitals that had not yet been inspected (or were never inspected) during the same period. To capture the multi-dimensional impact of the policy, we specify models at two distinct levels of aggregation: the hospital-month level for hospital outcomes and the patient level for health outcomes.

4.2 Hospital-Level and Patient-Level Specifications

To examine the impact on hospital outcomes (e.g., aggregate expenditure, admission volume), we estimate the following two-way fixed effects (TWFE) model using the hospital-month panel dataset:

$$Y_{ht} = \beta_0 + \beta_1 D_{ht} + \mu_h + \gamma_t + \varepsilon_{ht} \tag{1}$$

where Y_{ht} denotes outcomes for hospital h in month t. Depending on the analysis, these outcomes include department-level measures such as total OBGYN hospitalizations, aggregated inpatient expenditures, and average expenditure per admission. The treatment indicator D_{ht} is defined as the interaction between an indicator for whether hospital h ever appears on the inspection list and a post-inspection indicator that switches on after inspection for treated hospitals. The coefficient of interest, β_1, captures the average treatment effect of inspection on hospital outcomes and patient outcomes. We include hospital fixed effects μ_h to absorb time-invariant hospital characteristics (such as location, administrative rank, and baseline capacity) and month-by-year fixed effects γ_t to control for common temporal shocks (such as seasonal flu patterns or province-wide policy changes). Standard errors are clustered at the hospital level to account for serial correlation in outcomes within hospitals over time.

To assess the impact on neonatal and maternal health outcomes, which are binary or continuous variables defined at the individual admission level, we estimate the following patient-level specification:

$$Y_{iht} = \beta_0 + \beta_1 D_{ht} + \mu_h + \gamma_t + \varepsilon_{iht} \tag{2}$$

where Y_{iht} represents the health outcomes for patient i admitted to hospital h in month t. Maternal health outcomes include indicators for delivery-related trauma or severe obstetric complications, while neonatal outcomes include birth weight and admission to the neonatal intensive care unit (NICU). The treatment indicator D_{ht} remains defined at the hospital-month level, consistent with the institutional level of the intervention. As in the hospital-level analysis, we include hospital and time fixed effects and cluster standard errors at the hospital level.

4.3 Event Study Specification and Discussion

To scrutinize the dynamic effects of the inspection and assess the validity of the parallel trends assumption, we extend Equation (1) and Equation (2) into an event-study framework. Specifically, for the hospital-level specification, we replace the single post-treatment indicator with a series of leads and lags capturing the number of months before and after inspection for treated hospitals. The event-study specification can be written as:

$$Y_{ht} = \beta_0 + \sum_{k \neq -1} \beta_k \mathbf{1}\{t - T_h = k\} + \mu_h + \gamma_t + \varepsilon_{ht} \tag{3}$$

where T_h denotes the month in which hospital h is inspected, and $\mathbf{1}\{t - T_h = k\}$ is an indicator equal to one if month t is k periods away from the inspection event. The period immediately prior to inspection ($k = -1$) is omitted and serves as the reference category. The coefficients β_k trace the evolution of outcomes in treaed hospitals relative to control hospitals over event time. Under the identifying assumption of parallel trends, coefficients on the pre-inspection indicators ($k < 0$) should be statistically indistinguishable from zero. We present graphical evidence of these dynamics in the Results section, where we also discuss the timing and persistence of post-inspection effects.

Despite the strengths of our research design, several potential threats to identification warrant discussion. First, recent econometric literature highlights that conventional TWFE estimators may yield biased estimates in settings with staggered treatment adoption and heterogeneous treatment effects (De Chaisemartin & d'Haultfoeuille, 2020; Goodman-Bacon, 2021). To address this, we complement our baseline TWFE estimates with alternative estimators robust to these issues, specifically the group-time average treatment effect approach of Callaway and Sant'Anna (2021) and the interaction-weighted event-study estimator proposed by Sun and Abraham (2021). Results from these heterogeneity-robust methods are qualitatively similar to our baseline findings. Second, other contemporaneous healthcare policies could confound the estimated effects. Accordingly, we explicitly account for major reforms implemented during the study period, such as the Zero Markup Drug Policy (ZMDP), by controlling for their specific rollout timelines in robustness checks. Third, to address potential spillover effects where patients might substitute between inspected and non-inspected facilities, we re-estimate our models excluding control hospitals located within a 5-kilometer radius of any treated hospital. Finally, given the potential differences in scale between treated and control hospitals, we restrict the sample to a common-support region based on baseline hospital characteristics. We show that restricting the analysis to this comparable subsample yields estimates consistent with our main results.

5 Baseline Results

In this section, we present the baseline results regarding the impact of hospital inspections on hospital outcomes (resource utilization) and patient health. Regarding hospital outcomes, we examine the changes in monthly aggregated inpatient expenditure in treated hospitals relative to control hospitals following the inspection, while also exploring the mechanisms driving these changes at the extensive and intensive margin. Regarding patient health, given our focus on the OBGYN department, we primarily analyze the impact on neonatal and maternal health outcomes. Additionally, we investigate other health indicators, such as maternal mortality rates or incidence of medical malpractice, as a supplementary analysis.

5.1 Hospital Outcomes

We begin by examining the impact of hospital inspections on monthly aggregate inpatient expenditure using Equation (1). Panel A of Table 1 reports the baseline results. Column 1 reveals that following an inspection, total monthly expenditure in the OBGYN departments of treated hospitals increased significantly by RMB 295,405 relative to the control group—an approximate increase of 14.3%. Decomposing this aggregate effect into pregnancy-related and non-pregnancy-related admissions (Columns 2 and 3), we find that the surge in expenditure is driven almost entirely by pregnancy-related admissions. Further decomposing pregnancy-related admissions into four clinical subgroups—pregnancy loss, prenatal, delivery, and postnatal admissions (Columns 4–7)—reveals substantial heterogeneity. While expenditure increased across prenatal, delivery, and postnatal categories, the relative magnitudes are most pronounced for prenatal and postnatal admissions, which saw increases of 67.99% and 81.51% relative to their pre-inspection means, respectively.

Table 1: Effect of Hospital Inspection on Hospital Production

	Full	NonPreg	Preg	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Aggregated Expenditure (RMB)							
D	295,405***	42,971	252,435***	404	57,849**	171,899**	11,565*
	(104,647)	(40,924)	(84,385)	(6,004)	(22,206)	(68,001)	(6,732)
Obs. (Hospital-Month Pair)	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.948	0.950	0.916	0.945	0.683	0.900	0.505
Pre-treat. Mean (T)	2,104,839	884,272	1,220,566	177,080	85,088	928,911	14,187
Magnitude (%)	14.03	4.859	20.68	0.228	67.99	18.51	81.51
Panel B: Admission Number							
D	38.49***	6.20**	32.29***	-0.23	16.91***	9.01	1.82*
	(11.84)	(2.67)	(10.96)	(0.78)	(4.75)	(8.27)	(1.03)
Obs. (Hospital-Month Pair)	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.954	0.959	0.936	0.961	0.733	0.908	0.571
Pre-treat. Mean (T)	323.6	90.99	232.6	36.36	24.07	166.2	2.363
Magnitude (%)	11.89	6.812	13.88	-0.640	70.25	5.421	77
Panel C: Average Expenditure (RMB)							
D	-75.94	-208.25	54.80	34.24	131.50	-99.69	520.68
	(167.10)	(191.60)	(154.53)	(149.07)	(184.66)	(178.24)	(345.64)
Obs. (Hospital-Month Pair)	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.936	0.894	0.917	0.847	0.797	0.897	0.380
Weight	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	6,161	8,372	5,050	4,311	3,222	5,494	4,142
Magnitude (%)	-1.233	-2.487	1.085	0.794	4.082	-1.815	12.57
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table estimates the impact of hospital inspection on hospital production metrics at the hospital-month level. Panel A reports the effects on aggregated medical expenditure (RMB), Panel B on the number of admissions, and Panel C on the average expenditure per admission. Column (1) presents results for the full sample, which is decomposed into non-pregnancy (Col. 2) and pregnancy-related (Col. 3) admissions. Columns (4)–(7) further categorize pregnancy-related admissions into pregnancy loss, prenatal, delivery, and postnatal care. The regressions in Panel C are weighted by admission volume to account for hospital size variations. Row D presents the Difference-in-Differences coefficients. Magnitude (%) represents the percentage change relative to the pre-treatment mean of the treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To ensure the validity of these findings, we conduct a battery of robustness checks reported in Table A2. First, given that the implementation of the Zero-Markup Drug Policy (ZMDP) partially overlapped with our study period and is known to alter physician incentives (Fang et al., 2021), we explicitly control for the ZMDP rollout to rule out potential contamination (Panel A). Second, to address potential spillover effects where patients might switch from inspected hospitals to nearby non-inspected ones, we exclude control group hospitals located within a 5km radius of any treated hospital (Panel B). Third, to mitigate size differences between treated and control hospitals, we reconstruct the sample using a common support based on hospital characteristics[7] (Panel C). Across

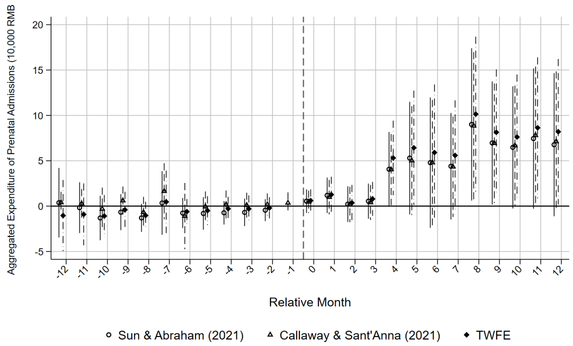
[7] We constructed a composite index by averaging the Z-scores of pre-inspection hospital characteristics (including full admission volumes, pregnancy admission volume, C-section rates, etc.). To ensure common support, we calculated quintiles within each group separately and retained quantiles 1–4 of the treatment group and quantiles 2–5 of the control group.

all specifications, our baseline finding remains robust: inspections led to a significant increase in aggregate OBGYN expenditure, primarily driven by prenatal and postnatal care.

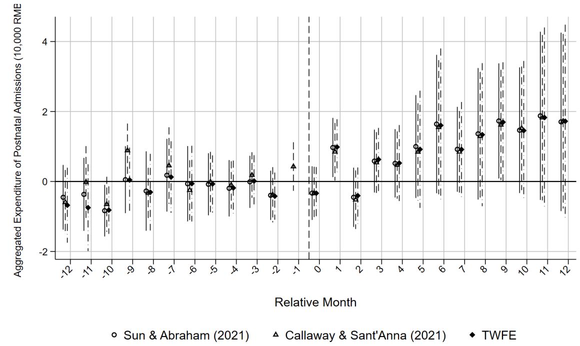
The observed increase in aggregate expenditure could theoretically stem from the extensive margin (an increase in the number of admissions) or the intensive margin (an increase in the average cost per admission). To identify the driver, we first examine the extensive margin by estimating the effect of inspections on monthly admission volume. Panel B of Table 1 reports these results. Consistent with the expenditure findings, Column 1 shows that the total number of OBGYN admissions in treated hospitals increased significantly by approximately 38 cases (12%) following inspection. This expansion is heavily concentrated in specific admission types. Column 5 indicates that prenatal admissions increased by approximately 17 cases—a dramatic 70% surge. Postnatal admissions also saw a statistically significant increase (Column 7), though the absolute magnitude (approximately 2 cases) is smaller. Notably, we find no significant changes in the volume of pregnancy loss or delivery admissions. This null result is intuitive: unlike prenatal checks or postnatal recovery, which can be discretionary, admissions for delivery or pregnancy loss are dictated by biological necessity and are thus difficult for physicians to manipulate. These findings are robust to the same battery of tests applied to the expenditure analysis, as reported in Table A3.

We next examine the intensive margin to determine whether the rise in aggregate expenditure was driven by an increase in the cost per patient. Panel C of Table 1 reports the results for the average cost per admission. Column 1 indicates no significant change in the average cost per stay for the full OBGYN sample. This null result holds across all subcategories; whether for pregnancy loss, prenatal care, delivery, or postnatal recovery (Columns 4–7), the cost per admission in treated hospitals remains statistically indistinguishable from that in control hospitals. Taken together, these results confirm that post-inspection revenue growth was driven entirely by the extensive margin—specifically, an increase in admission volume—rather than by an increase in the intensity or price of individual treatments.

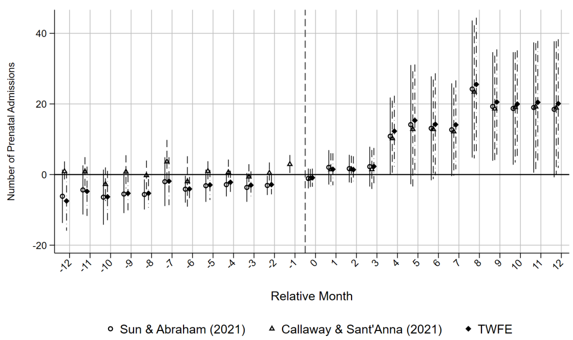
To validate the parallel trends assumption and explore the dynamic evolution of treatment effects, we estimate the event study specification in Equation (3). Figure 3 plots the dynamic coefficients for hospital outcomes—specifically aggregate expenditure and admission volume—for prenatal and postnatal care, as these categories exhibited the largest extensive margin responses. The top and bottom rows correspond to the static results reported in the first and second panels of Table 1, respectively. The x-axis denotes months relative to the start of the inspection, with the month immediately preceding the inspection ($k = -1$) serving as the reference category. To address potential biases inherent in two-way fixed effects (TWFE) models with staggered adoption, we also report estimates using the heterogeneity-robust estimators proposed by Sun and Abraham (2021) and Callaway and Sant'Anna (2021). Results for other admission types, such as pregnancy loss and delivery, are presented in Figures A1 and A2.



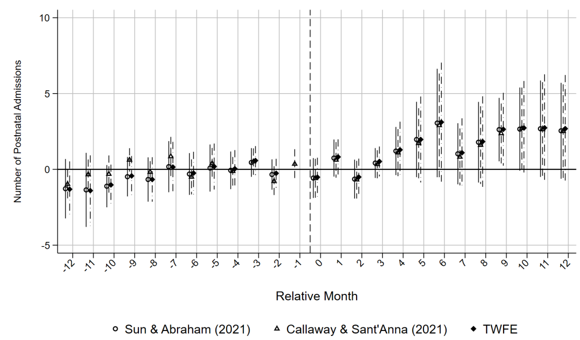
Panel A: Aggregated Expenditure of Prenatal Admissions



Panel B: Aggregated Expenditure of Postnatal Admissions



Panel C: Number of Prenatal Admissions



Panel D: Number of Postnatal Admissions

Figure 3: Event Study: Hospital Outcomes

Notes: This figure reports event study estimates for the dynamic impacts of hospital inspection on hospital outcomes metrics. Panels A and B plot the coefficients for aggregated medical expenditure (in 10,000 RMB) related to prenatal and postnatal admissions, respectively, while Panels C and D plot the corresponding effects on admission volume. To ensure robustness against potential biases in standard difference-in-differences models with staggered timing, each panel compares results from three distinct specifications: (1) the standard Two-Way Fixed Effects (TWFE) model (solid diamonds), (2) the interaction-weighted estimator by Sun and Abraham (2021) (hollow circles), and (3) the estimator by Callaway and Sant'Anna (2021) (triangles). The horizontal axis represents the month relative to the inspection month ($t = 0$), with $t = -1$ serving as the omitted reference period. Vertical bars represent 95% confidence intervals.

The results in Figure 3 support the validity of our research design. The coefficients for the pre-treatment periods are statistically indistinguishable from zero, indicating no differential trends between treated and control hospitals prior to the intervention. Following the inspection, we observe a marked and sustained increase in both expenditure and volume, confirming that the impact of the inspection on hospital outcomes was persistent rather than transitory.

We note that the confidence intervals widen toward the end of the post-treatment window. We attribute this loss of precision to sample attrition. For instance, hospitals inspected in September 2017 contribute only three post-treatment months before our data truncation in December 2017; as the sample size for these later lags decreases, the standard errors naturally increase. Furthermore, the treatment effect typically materializes with a lag of three to four months. This delay is consistent with the institutional design of the inspection process: following the initial audit, the inspection team issues a feedback report, to which the hospital must respond with a detailed ”rectification plan.” This administrative cycle typically spans a quarter, meaning that substantive operational changes are often implemented only in the third or fourth month post-inspection, coinciding with the timing of the follow-up ”re-inspection.”

To summarize, our analysis of hospital outcomes reveals a robust pattern: anti-corruption inspections led to a significant increase in aggregate revenue within OBGYN departments relative to non-inspected peers. This growth was driven primarily by an expansion in pregnancy-related admissions, with the most pronounced and robust increases observed in prenatal and postnatal care. In contrast, expenditure on delivery and pregnancy loss admissions remained largely unaffected. Further decomposition reveals that this expenditure growth was driven exclusively by the extensive margin (an increase in the number of admissions) rather than the intensive margin (changes in the cost per admission). Moreover, this expansion in admission volume was sustained and did not dissipate after the inspection period concluded, suggesting a structural shift in provider behavior.

These findings align with the ”income compensation hypothesis.” By severing the link between physicians and illicit ”gray income” (e.g., pharmaceutical kickbacks) without a concurrent increase in formal wages, the inspections induced a negative income shock. To recoup these losses, physicians increased the supply of legitimate, high-discretion medical services—such as prenatal screenings. Notably, this response operated entirely along the extensive margin, contrasting with the intensive-margin adjustments typically observed in response to changes in formal fee schedules. A critical normative question remains: Does this induced increase in service provision represent ”supply-induced demand” (wasteful over-provision of unnecessary care) or the fulfillment of previously ”unmet medical needs”? If the former, the policy may have reduced allocative efficiency. If the latter, the policy may have generated positive externalities for patient welfare. To resolve this ambiguity, the following section investigates the impact of inspections on neonatal and maternal health outcomes.

5.2 Patient Health Outcomes

In this section, we examine whether the observed expansion in service provision translates into tangible health benefits. We focus on two primary domains: neonatal health and maternal health. To maximize estimation precision and granularity, we employ patient-level data from delivery admissions and estimate the effects of hospital inspection on health using Equation (2).

We construct a comprehensive set of metrics to evaluate neonatal health: (1) admission to the Neonatal Intensive Care Unit (NICU), inferred from delivery diagnosis records; (2) continuous birth weight; (3) incidence of Low Birth Weight (LBW) and Very Low Birth Weight (VLBW); and (4) an indicator for whether birth weight falls within one standard deviation of the sample mean, following Best et al. (2017), to capture distributional improvements. Panel A of Table 2 reports the results. We find statistically significant improvements in neonatal health following inspections. Specifically, in treated hospitals, the rates of NICU admission and low birth weight decreased by approximately 6 cases per 1,000 births, while the incidence of very low birth weight decreased by approximately 2 cases per 1,000 births. Furthermore, the average birth weight in treated hospitals increased by 13.5 grams ($p < 0.05$). Crucially, the proportion of newborns with a birth weight within one standard deviation of the population mean increased by approximately 12 per 1,000. This distributional shift indicates that the increase in birth weight represents a convergence toward the optimal mean rather than a pathological shift toward macrosomia.

Table 2: Effect of Hospital Inspection on Maternal Health

	(1)	(2)	(3)	(4)	(5)
Panel A: Neonatal Health (expressed per 1,000)					
	NICU	LBW	VLBW	BirthWeight	BirthWeight_1SD
D	-6.23*	-6.38*	-2.33**	13.47**	11.92***
	(3.36)	(3.66)	(1.08)	(6.07)	(4.29)
Obs. (Delivery Admissions)	445,793	445,793	445,793	445,793	445,793
R-squared	0.036	0.039	0.015	0.032	0.013
Pre-treat. Mean (T)	51.90	84.03	17.89	3236	721.2
Magnitude (%)	-11.99	-7.597	-13.04	0.416	1.653
Panel B: Maternal Health (expressed per 1,000)					
	Perineal_Lac	Cervical_Lac	Sepsis	ARDS	HyperT_Disorder
D	-2.13**	-6.08**	-0.26**	-0.13***	-8.17**
	(0.90)	(2.74)	(0.11)	(0.05)	(3.88)
Obs. (Delivery Admissions)	445,793	445,793	445,793	445,793	445,793
R-squared	0.012	0.107	0.001	0.000	0.031
Pre-treat. Mean (T)	3.492	27.36	0.278	0.113	70.85
Magnitude (%)	-61.05	-22.21	-92.60	-115.6	-11.54
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: This table estimates the effects of hospital inspection on neonatal and maternal health outcomes, utilizing a sample restricted to delivery admissions. Panel A reports results for neonatal health. Columns (1)–(3) and (5) are binary indicators scaled per 1,000 admissions: NICU (Neonatal Intensive Care Unit admission), LBW (Low Birth Weight < 2,500g), VLBW (Very Low Birth Weight < 1,500g), and BirthWeight_1SD (birth weight within one standard deviation of the mean). Column (4) represents continuous Birth Weight in grams. Panel B reports maternal health outcomes, all scaled per 1,000 admissions, including indicators for perineal laceration, cervical laceration, sepsis, ARDS (Acute Respiratory Distress Syndrome), and hypertensive disorders. Row D reports the DID coefficients. Magnitude (%) indicates the percentage change relative to the pre-treatment mean of the treatment group (Pre-treat. Mean (T)). Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We observe parallel improvements in maternal health. Panel B of Table 2 presents results based on delivery diagnoses, focusing on the incidence of obstetric trauma (perineal and cervical lacerations), hypertensive disorders (including hypertension and eclampsia), and severe maternal complications (Puerperal Sepsis and Acute Respiratory Distress Syndrome [ARDS]). Relative to the control group, treated hospitals exhibited significant reductions in adverse maternal outcomes post-inspection. Specifically, per 1,000 deliveries, the incidence of perineal lacerations decreased by approximately 2 cases, cervical lacerations by 6 cases, sepsis by 0.26 cases, ARDS by 0.13 cases, and hypertensive disorders by 8 cases. All estimates are statistically significant at the 5% level or better, signaling a broad-based improvement in the safety and quality of obstetric care.

A potential concern is that the intense pressure of inspections might induce panic or distraction among medical staff, temporarily compromising patient safety. To address this, we examined extreme adverse outcomes and indicators of medical error. As reported in Table A4, we find no significant impact on maternal mortality, neonatal mortality, severe fetal conditions, major congenital anomalies, macrosomia, or medical malpractice (proxied by complications of surgical anesthesia). This null result suggests that the inspections did not generate harmful disruptions or ”defensive medicine” artifacts that compromised safety.

Collectively, these findings demonstrate that hospital inspections led to significant improvements in both neonatal and maternal health without inducing adverse side effects on mortality or malpractice rates. These results provide a preliminary answer to the question posed in the previous section: the additional prenatal services induced by the inspections were likely not wasteful ”supply-induced demand.” Instead, they appear to have addressed pre-existing unmet medical needs. By increasing the supply of prenatal care—likely to compensate for lost illicit income—physicians were able to identify and mitigate potential risks earlier in the pregnancy, thereby preventing complications during delivery. To further validate this causal chain, the following section presents a series of heterogeneity and mechanism analyses.

6 Mechanism Analysis

6.1 Heterogeneity in Health Improvement

In this section, we explore the heterogeneous impacts of hospital inspections on patient health outcomes. Our previous findings suggest that inspections improved health outcomes because physicians, having lost their sources of ”gray income,” compensated for this financial shock by increasing the supply of legitimate prenatal care. Following this logic, the health benefits should be more pronounced in hospitals where physicians were more dependent on gray income prior to the intervention.

Since gray income is illicit and inherently unobservable, we employ the pre-inspection Cesarean section (C-section) rate as a proxy for a hospital’s reliance on financial incentives. To avoid potential anticipatory effects, we utilize data from January to June 2015 to calculate these baseline rates. This proxy is grounded in extensive literature suggesting that, under conditions of information asymmetry, physicians possess strong financial incentives to induce demand for C-sections to maximize revenue (Gruber & Owings, 1996; Johnson & Rehavi, 2016). Furthermore, China has long exhibited high C-section utilization by international standards, with national estimates ranging from 35% to 40% in the mid-2010s and hospital-level data showing substantial variation in the mid-40% range (Li et al., 2017; Zhang et al., 2022). Consequently, we posit that hospitals with higher baseline C-section rates are likely those where physicians are more responsive to financial incentives and thus more reliant on gray income.

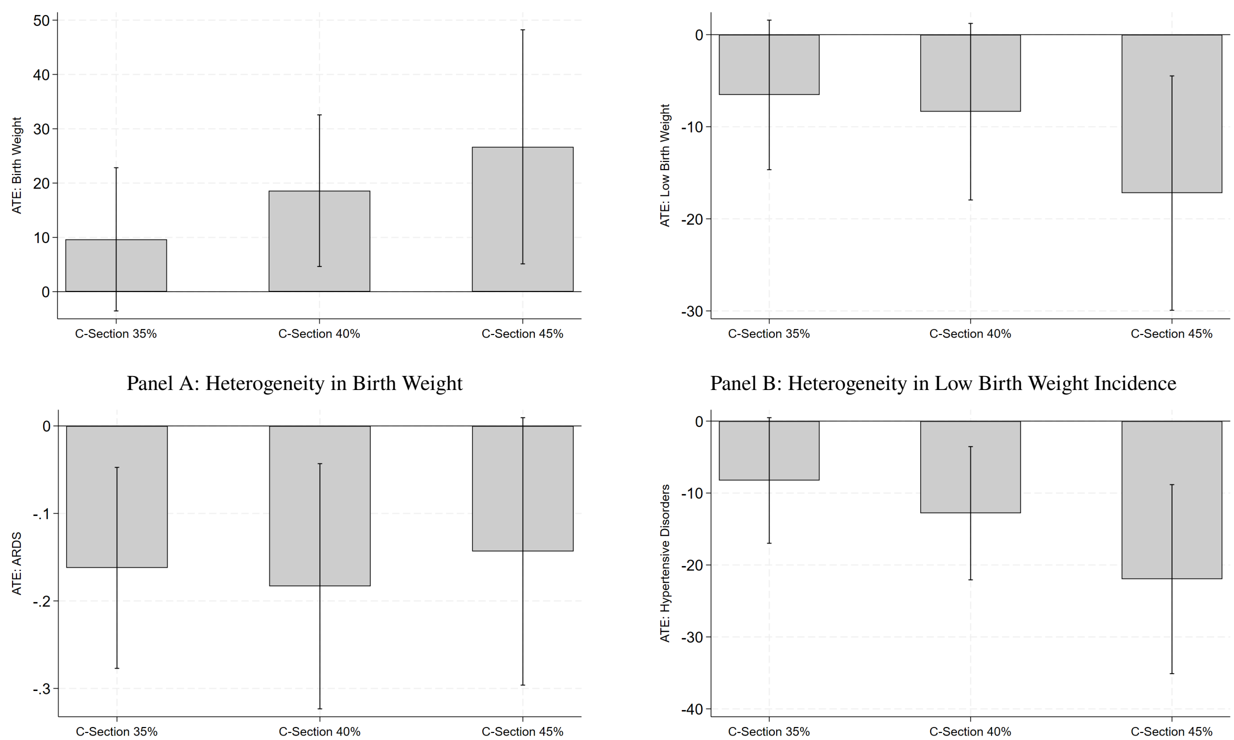
We stratified hospitals into two groups: "High C-section rate hospitals" (baseline rate > 40%) and "Low C-section rate hospitals" (baseline rate < 40%). We then estimated the impact of inspections on neonatal and maternal health separately for each group. The results, reported in Table 3 Panel A, reveal that the improvements in health outcomes are predominantly concentrated in the high C-section rate hospitals. Columns 1–2 indicate that in these high-rate hospitals, neonatal birth weight increased significantly by 18.6 grams post-inspection, whereas no significant change was observed in low-rate hospitals. Similarly, the incidence of low birth weight (LBW) decreased by approximately 8 cases per 1,000 births in the high-rate group. This pattern extends to maternal health: the incidence of ARDS and hypertensive disorders decreased by 0.18 and 13 cases per 1,000 deliveries, respectively, in high C-section rate hospitals. Conversely, we observed no comparable improvements in the low-rate group.

Table 3: Heterogeneity in Health Outcomes: by Pre-inspection C-section Rate

Health Outcomes	Birth Weight (gram)		Low-Birth-Weight		ARDS		Hypertensive Disorders	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: by Pre-inspection C-section Rate								
	H_Csec_R	L_Csec_R	H_Csec_R	L_Csec_R	H_Csec_R	L_Csec_R	H_Csec_R	L_Csec_R
D	18.60**	-0.07	-8.37*	1.25	-0.18**	-0.10	-12.80**	-0.41
	(7.12)	(8.90)	(4.89)	(2.49)	(0.07)	(0.07)	(4.72)	(4.85)
Observations	203,154	201,497	203,154	201,497	203,154	201,497	203,154	201,497
R-squared	0.038	0.025	0.042	0.031	0.000	0.000	0.028	0.033
Pre-treat. Mean (T)	3237	3234	86.37	76.29	0.107	0.133	71.20	69.72
Magnitude (%)	0.575	-0.00220	-9.688	1.634	-170.7	-72.94	-17.98	-0.586
Panel A: by Physician Seniority								
	Junior	Senior	Junior	Senior	Junior	Senior	Junior	Senior
D	5.27	14.08*	-0.08	-7.42*	-0.14	-0.14**	-0.29	-10.37**
	(10.07)	(7.19)	(5.45)	(3.99)	(0.11)	(0.06)	(8.40)	(4.56)
Observations	167,241	263,635	167,241	263,635	167,241	263,635	167,241	263,635
R-squared	0.030	0.035	0.040	0.039	0.003	0.000	0.033	0.030
Pre-treat. Mean (T)	3240	3234	81.88	85.12	0.0754	0.129	63.70	73.67
Magnitude (%)	0.163	0.435	-0.0992	-8.715	-189	-105.5	-0.457	-14.08
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the heterogeneous policy effects on neonatal (Birth Weight and Low-Birth-Weight) and maternal (ARDS and Hypertensive Disorders) health outcomes at the admission level. Panel A stratifies the sample by hospitals with high versus low pre-inspection C-section rates (H_Csec_R vs. L_Csec_R), while the lower panel (Panel B) stratifies by physician seniority (Senior vs. Junior). Row D reports the estimated coefficients from the Difference-in-Differences specification. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the treatment group (Pre-treat. Mean (T)). All regressions control for Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To further validate this gradient, we refined our categorization using three increasing thresholds for pre-inspection C-section rates: >35%, >40%, and >45%. Figure 4 plots the heterogeneity of treatment effects across these thresholds. Panels A and B demonstrate a clear positive correlation: the higher the hospital's initial C-section rate, the larger and more significant the post-inspection increase in birth weight and the reduction in LBW incidence. A similar pattern is evident for maternal health; Panel D shows that the reduction in hypertensive disorders is most pronounced in hospitals with the highest baseline C-section rates. These findings align closely with our theoretical framework: hospitals and physicians with a stronger ex-ante dependence on profit-driven behaviors (proxied by C-sections) experienced a larger behavioral shock from the anti-corruption inspections. Consequently, the shift toward providing necessary prenatal care—and the resulting health improvements—was most substantial in these institutions.



Panel C: Heterogeneity in ARDS Incidence

Panel D: Heterogeneity in Hypertensive Disorder Incidence

Figure 4: Heterogeneity: Health Outcome by Hospitals' Pre-inspection C-section

Notes: This figure presents the heterogeneity of the Average Treatment Effects (ATE) on neonatal and maternal health outcomes across hospitals with different pre-inspection C-section intensities. Panel A displays the impact on Birth Weight (in grams), while Panels B, C, and D report the effects on the incidence of Low Birth Weight, ARDS, and Maternal Hypertensive Disorders (scaled per 1,000 admissions). The horizontal axis defines the treatment group using three ascending thresholds of baseline C-section rates (>35%, >40%, and >45%). Vertical lines indicate 95% confidence intervals.

Finally, we investigate whether the improvements in maternal and neonatal health vary by physician seniority. Our hypothesis posits that because senior physicians typically possess greater decision-making authority over the treatment process, they are the primary targets for gray income (e.g., pharmaceutical kickbacks or direct bribes to promote specific drugs). Consequently, the health benefits resulting from anti-corruption inspections should be most pronounced among patients treated by these senior practitioners. In the Chinese context, while attending physicians generally hold primary responsibility for patient management, there exists a distinct hierarchy; attending physicians who also hold high-ranking titles (such as Chief Physician) or administrative roles (such as Department Head) possess higher professional status and are thus more prone to receiving illicit payments. To operationalize this distinction, we identify seniority based on surname matching: attending physicians are classified as ”senior attending physicians” if their surname matches that of a Chief Physician or Department Head within the unit; otherwise, they are classified as ”junior attending physicians.”

Table 3 Panel B reports the heterogeneous effects across these two groups. The results reveal that health improvements are significantly larger and more robust for patients under the care of senior physicians. For instance, Columns 1 and 2 indicate that among patients treated by senior physicians, neonatal birth weight increased significantly by approximately 14 grams following the inspection. Similarly, the final two columns show that the incidence of hypertensive disorders during delivery in the senior group decreased by approximately 10 cases per 1,000—a reduction of 14%. This effect magnitude substantially exceeds that observed in the junior physician subsample. Collectively, these results align with our intuition and further corroborate that the anti-corruption inspections successfully improved health outcomes by altering the behavior of the physician demographic most heavily implicated in rent-seeking activities.

While the previous section established that hospital inspections generated positive externalities for patient health, two critical questions remain unresolved. First, an economic puzzle arises regarding incentive compatibility: if increasing the supply of prenatal care simultaneously boosts hospital revenue and improves patient outcomes, why did physicians fail to optimize this supply prior to the inspections? Second, the specific channels driving these health improvements remain a ”black box.” Crucially, observed improvements in health outcomes do not necessarily equate to genuine enhancements in hospital performance or social welfare. For instance, in response to inspections, physicians might engage in defensive behavior or strategic patient selection (often termed ”cherry-picking”) to minimize costs or labor intensity. By preferentially admitting patients with lower disease severity, hospitals could generate an artificial ”improvement” in average health outcomes. However, such ”improvements,” driven by the exclusion of high-risk patients, would represent a distortion of medical access rather than a socially beneficial gain.

Therefore, to rigorously assess the true welfare impact of hospital inspections, this section dissects the specific mechanisms through which the observed improvements in neonatal and maternal health were realized.

6.2 Decreased Length of Stay and Increased Preventive Treatment

Our investigation into the mechanisms begins by examining Average Length of Stay (ALOS). Given our baseline finding that monthly admission volumes surged following inspections, a logistical puzzle arises: how did OBGYN departments accommodate this influx of patients within the short term, given that physical infrastructure (e.g., ward capacity) is fixed? Theoretically, hospitals can expand capacity through three channels: (1) ”ghost admissions” (admitting patients nominally for billing purposes without physical occupancy); (2) adding temporary corridor beds; or (3) accelerating bed turnover by reducing the length of stay per admission. Due to data limitations, we cannot directly verify the first two practices. Therefore, we focus our analysis on the third channel: checking whether inspections incentivized hospitals to improve bed turnover efficiency.

Table A5 reports the impact of inspections on ALOS. Column 1 of Panel A reveals that the average single-admission duration in treated OBGYN departments decreased significantly by 0.22 days. Decomposing admissions into pregnancy-related and non-pregnancy-related categories (Column 3), we find that this reduction is driven almost exclusively by pregnancy-related admissions. Further decomposition into the four specific subgroups (pregnancy loss, prenatal, delivery, and postnatal) indicates that the decline in ALOS is concentrated in delivery admissions, which saw a reduction of 0.26 days—a decrease of approximately 4%. To address potential bias arising from regression to the mean in hospitals of varying sizes (Baker et al., 2025), Panel B replicates the analysis weighting by monthly admission volume. The results remain robust, confirming a significant reduction in ALOS for treated hospitals.

We next investigate the structural nature of this ALOS reduction. Combined with the observed surge in prenatal admission volume, we hypothesize that the aggregate drop in ALOS might be mechanically driven by a compositional shift toward short-term preventive admissions (i.e., 1–2 day stays). To test this, we distinguish between admissions based on the ”splittability” of the medical process. The intuition is that non-delivery admissions (e.g., prenatal checks) are flexible and can be easily fragmented into multiple short stays to generate revenue. In contrast, delivery admissions follow strict, standardized clinical pathways and are difficult to split into artificial short stays.

Table A6 presents the results for ”short-term admissions” (defined as stays of less than 3 days). Consistent with our hypothesis, Columns 1–3 show that for non-delivery cases, the number of 2-day and 3-day admissions rose significantly by 60% and 32%, respectively. Conversely, as a placebo test, we examined short-term delivery admissions and found no significant change. These findings suggest that the reduction in ALOS and the increase in volume operate through complementary channels: for delivery admissions (which are hard to split), inspections drove legitimate efficiency gains (intensive margin), reducing stay duration without compromising outcomes. For prenatal admissions (which are easy to split), inspections incentivized an expansion of volume via short-term stays (extensive margin).

Finally, we examine whether these operational responses—reduced ALOS and increased prenatal supply—exhibit the same heterogeneity observed in our health outcome analysis. Table 4 reports these results. Panel A reveals that the reduction in ALOS and the expansion of prenatal care are predominantly concentrated in hospitals with high pre-inspection C-section rates. This aligns with our previous findings that hospitals more reliant on gray income were more responsive to the inspection shock. However, the physician-level heterogeneity in Panel B is more nuanced: while the reduction in ALOS is driven largely by junior attending physicians, the increase in prenatal care supply is evident across both junior and senior physician groups.

Table 4: Heterogeneous Effects on Efficiency and Service Volume by Hospital and Physician Characteristics

Outcomes	ALOS		Prenatal Admission Number	
	(1)	(2)	(3)	(4)
Panel A: by Pre-inspection C-section Rate				
	H_Csec_R	L_Csec_R	H_Csec_R	L_Csec_R
D	-0.23**	-0.19	17.35***	15.75*
	(0.09)	(0.15)	(5.68)	(9.37)
Observations	1,385	1,594	1,385	1,594
R-squared	0.886	0.909	0.652	0.846
Pre-treat. Mean (T)	7.613	7.538	24.79	22.02
Magnitude (%)	-3.080	-2.519	69.99	71.52
Panel B: by Physician Seniority				
	Junior	Senior	Junior	Senior
D	-0.47**	-0.15	7.36**	8.42***
	(0.22)	(0.14)	(3.39)	(2.91)
Observations	3,125	3,384	3,517	3,517
R-squared	0.622	0.715	0.788	0.706
Pre-treat. Mean (T)	7.878	7.640	6.644	17.12
Magnitude (%)	-5.985	-1.968	110.8	49.19
Hospital FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: This table presents the heterogeneous impacts of hospital inspection on Average Length of Stay (ALOS) (Columns 1–2) and the Number of Prenatal Admissions (Columns 3–4). Panel A stratifies the sample by hospital type: H_Csec_R and L_Csec_R denote hospitals with pre-inspection C-section rates above and below the sample median, respectively. Panel B stratifies the sample by physician seniority (Junior vs. Senior). Row D reports the Difference-in-Differences coefficients. Magnitude (%) represents the percentage change relative to the pre-treatment mean of treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In summary, we conclude that hospital inspections facilitated the expansion of prenatal care supply by improving bed turnover efficiency. This operational shift allowed hospitals to process more patients, thereby generating the health improvements observed in the previous section. Furthermore, this channel is primarily driven by hospitals with a higher ex-ante reliance on gray income, reinforcing the income compensation hypothesis.

6.3 Structural Changes in Expenditure (Drug-Service Substitution)

A primary objective of the anti-corruption inspections is to eliminate sources of "gray income," with pharmaceutical kickbacks being the most significant component. This raises a critical question: did the inspections alter the structural composition of medical service provision? While our baseline results indicated no significant change in total expenditure per admission, this aggregate null result may mask a compositional shift where a decrease in one expenditure category is offset by an increase in another.

To investigate this, we decomposed total inpatient expenditure into four distinct categories: drug, service, diagnostic, and consumable expenditure. Table A7 reports the impact of inspections on these components. Panel A reveals a clear substitution pattern: in treated hospitals, average drug expenditure per admission decreased significantly by 209 RMB, while medical service expenditure increased by 173 RMB. Diagnostic and consumable expenditures remained statistically unchanged. To account for hospital size heterogeneity, Panel B weights the regression by monthly admission volume; the results remain robust, confirming a negative impact on drug spending and a positive impact on service spending. This suggests that as inspections reduced the relative returns (or increased the risk) of drug prescriptions, physicians substituted away from pharmaceuticals toward other medical services.

However, before attributing the decline in drug expenditure to a behavioral reduction in prescribing (a quantity effect), we must rule out a mechanical "price effect." The implementation of the Zero-Markup Drug Policy during our study period lowered drug prices, which could theoretically drive the observed decrease in expenditure without any change in physician behavior.

To disentangle these mechanisms, we examined the frequency of admissions with near-zero drug expenditure. The intuition is straightforward: if physicians are genuinely reducing unnecessary prescriptions, we should observe an increase in the number of patients treated with little to no medication. Table A8 presents these results. Columns 1–4 report the impact on the volume of low-drug-cost admissions, while Columns 5–8 report the impact on their share of total admissions. We find that the number of admissions with drug expenditures $\leq$ 10 RMB, $\leq$ 50 RMB, and $\leq$ 100 RMB increased by 3.86, 6.74, and 10.32 cases, respectively—increases of over 20%. Similarly, the share of these low-drug-cost admissions rose significantly by 1–3 percentage points. This evidence strongly suggests that the decline in drug expenditure is not merely a price artifact but stems from a genuine behavioral shift where physicians reduced the quantity of drugs prescribed.

Having established that physicians substituted services for drugs, we next ask whether this trade-off persists across different clinical contexts. Table A9 subdivides OBGYN admissions into non-pregnancy and pregnancy-related categories (sub further into pregnancy loss, prenatal, delivery, and postnatal). Panel A reports effects on service expenditure, while Panel B reports effects on drug expenditure.

The results reveal an interesting dichotomy. For non-pregnancy admissions, we observe the classic substitution pattern: service expenditure increased significantly while drug expenditure fell. However, for pregnancy-related admissions—and specifically prenatal care (Column 4)—medical service fees increased significantly, but drug expenditure remained unchanged. This finding is clinically consistent: pharmacological interventions are inherently limited during prenatal care due to safety concerns for the fetus. Consequently, baseline drug use was likely already low and rational, leaving little room for reduction. Thus, for prenatal care, the income compensation occurred purely through the expansion of services without a corresponding cut in drugs.

Given that the reduction in drug expenditure was largely concentrated in non-pregnancy admissions (which are less relevant to our neonatal and maternal health outcomes), we focus our final heterogeneity analysis on the expansion of medical service expenditure. Table 5 presents these results, with Columns 1–2 covering all OBGYN admissions and Columns 3–4 focusing specifically on prenatal admissions. Panel A shows that the increase in service expenditure is driven predominantly by hospitals with high pre-inspection C-section rates. Moreover, within these hospitals, the magnitude of the increase is substantially larger for prenatal admissions. Panel B examines physician heterogeneity, revealing that senior physicians drove the increase in service expenditure, again with a more pronounced effect in prenatal care.

Table 5: Heterogeneous Effects on Service Expenditure by Hospital and Physician Characteristics

	Service Expenditure of Full Admission		Service Expenditure of Prenatal Admission	
	(1)	(2)	(3)	(4)
Panel A: by Pre-inspection C-section Rate				
	H_Csec_R	L_Csec_R	H_Csec_R	L_Csec_R
D	151.55**	128.76	245.67**	89.82
	(65.75)	(77.27)	(99.90)	(81.80)
Observations	1,385	1,594	1,360	1,495
R-squared	0.904	0.671	0.870	0.537
Pre-treat. Mean (T)	2701	2401	1390	1100
Magnitude (%)	5.610	5.363	17.67	8.167
Panel B: by Physician Seniority				
	Junior	Senior	Junior	Senior
D	-30.51	161.17***	145.22	235.50***
	(87.79)	(49.89)	(101.77)	(80.22)
Observations	3,125	3,385	2,406	2,849
R-squared	0.877	0.885	0.881	0.828
Pre-treat. Mean (T)	2684	2625	1266	1362
Magnitude (%)	-1.137	6.141	11.47	17.29
Hospital FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: This table reports the difference-in-differences estimates of the impact of hospital inspection on average service expenditure per admission across different subsamples. The unit of analysis is the hospital-month. Columns (1) and (2) use the full sample of admissions, while Columns (3) and (4) restrict the sample to prenatal admissions. Panel A stratifies hospitals based on their pre-inspection C-section rates: H_Csec_R (High C-section Rate) denotes hospitals with a baseline C-section rate above the sample median, while L_Csec_R (Low C-section Rate) denotes those below the median. Panel B stratifies admissions based on the seniority of the attending physician: Junior and Senior refer to admissions managed by junior and senior physicians, respectively. Row D reports the estimated coefficients on the interaction term between the treatment dummy and the post-inspection dummy. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In conclusion, our analysis indicates that hospital inspections altered the relative risk-return structure of medical inputs, prompting a strategic reallocation of resources. Physicians responded to the loss of gray income by substituting away from drug prescriptions (where kickbacks were curtailed) and toward legitimate medical services (to compensate for revenue loss).

Crucially, this substitution was most aggressive among the groups most dependent on gray income—specifically, high C-section rate hospitals and senior attending physicians. Furthermore, because physicians possess high discretion in prenatal care (and limited ability to prescribe drugs), the ”service expansion” channel was particularly strong in this category. This shift toward service-intensive, preventive prenatal care provides the mechanistic link explaining the improvements in neonatal and maternal health observed in our main results.

6.4 Selective Attrition and the Retention of High-Quality Personnel

As detailed in the background section, a defining feature of the 2015 hospital inspections was the severity of the punitive measures introduced. Physicians found accepting pharmaceutical kickbacks faced not only the revocation of their medical licenses but also potential criminal prosecution and imprisonment. Given these high stakes, we propose a ”selective attrition hypothesis”: physicians with a high dependence on gray income may have self-selected out of inspected public hospitals to avoid scrutiny. These physicians likely chose to retire early or move to institutions outside the inspection scope (e.g., private hospitals). Consequently, their departure would create vacancies filled by new entrants or leave the workload to remaining staff. If the exiting physicians were indeed the most aggressive rent-seekers, their replacement by less ”financially distorted” peers (or the purification of the remaining pool) would lead to more rational treatment plans, thereby improving neonatal and maternal health.

To empirically test this hypothesis, we tracked physician presence using surname data from monthly admission records. We classified attending physicians into three distinct groups: (1) Exit Physicians, whose names appeared before the inspection but permanently vanished afterward; (2) Entry Physicians, whose names appeared only after the inspection (likely new hires or promotions); and (3) Stay Physicians, who were present throughout the pre- and post-inspection periods.

Figure A3 illustrates the prevalence of physician turnover. Panel A shows the general distribution, indicating that treated hospitals tended to have a higher number of Exit Physicians prior to the inspection compared to control hospitals. Panel B narrows the window to the six months immediately preceding the inspection. The pattern remains consistent: treated hospitals exhibited a higher rate of physician exit just before the policy implementation, suggesting an anticipatory response to the looming threat of investigation.

If selective attrition is a key driver of health improvements, we should observe larger treatment effects in hospitals that actually experienced this turnover. To test this, we stratified hospitals into two groups based on whether they had any Exit Physicians prior to the inspection. The intuition is straightforward: if a hospital experienced physician exit, it suggests a ”cleaning out” of rent-seeking elements, and thus the post-inspection improvement should be more pronounced.

Table 6 (Panel A) confirms this heterogeneity. We find that the improvements in patient health are almost entirely concentrated in hospitals that experienced physician exit. In these hospitals, neonatal birth weight increased by approximately 19 grams, the incidence of low birth weight decreased by 10 cases per 1,000, and the incidence of maternal ARDS and hypertensive disorders dropped by 118% and 15.37%, respectively. In contrast, hospitals without selective attrition (no exits) showed no significant health improvements. To rule out normal personnel turnover, Panel B replicates this analysis defining ”Exit Hospitals” based only on departures within the six-month window prior to inspection. The results remain robust, confirming that the health gains are driven by hospitals where the inspection likely triggered the departure of specific staff members.

Table 6: Heterogeneous Effects on Health Outcomes by Pre-Inspection Physician Turnover

Health Outcomes	BW	LBW	ARDS	HyperT_Disorder	BW	LBW	ARDS	HyperT_Disorder
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Baseline								
	Hosp. W/ Exit Phys. before Inspection				Hosp. W/O Exit Phys. before Inspection			
D	18.80***	-10.00**	-0.10**	-11.06**	-6.20	4.49	-0.19	0.88
	(6.76)	(4.45)	(0.04)	(4.73)	(11.92)	(2.67)	(0.16)	(5.24)
Observations	321,265	321,265	321,265	321,265	109,611	109,611	109,611	109,611
R-squared	0.033	0.043	0.000	0.031	0.029	0.024	0.001	0.028
Pre-treat. Mean (T)	3231	87.09	0.0834	71.96	3269	64.76	0.324	63.87
Magnitude (%)	0.582	-11.48	-118.1	-15.37	-0.190	6.938	-58.49	1.377
Panel B: Robustness Check								
	Hosp. W/ Exit Phys. within 6 Months before Inspection				Hosp. W/O Exit Phys. within 6 Months before Inspection			
D	17.87**	-9.97*	-0.10*	-12.67**	6.10	0.28	-0.18*	-1.84
	(7.81)	(5.29)	(0.05)	(5.71)	(9.09)	(2.36)	(0.10)	(3.18)
Observations	245,831	245,831	245,831	245,831	185,045	185,045	185,045	185,045
R-squared	0.033	0.043	0.000	0.029	0.028	0.021	0.001	0.026
Pre-treat. Mean (T)	3221	94.29	0.0816	76.64	3284	51.59	0.220	52.38
Magnitude (%)	0.555	-10.58	-116.5	-16.53	0.186	0.535	-82.63	-3.517
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the difference-in-differences estimates of the impact of hospital inspection on neonatal and maternal health outcomes, stratified by the status of physician turnover prior to the inspection. Columns (1)–(4) restrict the sample to hospitals that experienced at least one physician exit before the inspection, while Columns (5)–(8) restrict the sample to hospitals with no physician exits during the same period. Panel A defines physician exits based on the entire pre-inspection period observed. Panel B conducts a robustness check by restricting the definition of physician exits to the six-month window immediately preceding the inspection date. The dependent variables are BW (Birth Weight in grams), LBW (Low Birth Weight indicator, ¡2,500g), ARDS (Acute Respiratory Distress Syndrome indicator), and HyperT_Disorder (Hypertensive Disorders of Pregnancy indicator). Row D reports the estimated coefficients on the interaction term between the treatment dummy and the post-inspection dummy. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Finally, to confirm that this attrition was indeed ”selective”—meaning the ”worse” doctors left—we must compare the clinical behavior of Exit Physicians against their peers. Did the exiting doctors actually exhibit more rent-seeking behavior? Figure 5 provides statistical evidence to answer this. Panels A and B compare the pre-inspection behavior of Exit Physicians versus Stay Physicians. We observe that Exit Physicians consistently had a distribution of drug expenditure shares skewed to the right (higher drug dependence) and significantly higher C-section rates compared to their staying colleagues. This suggests they were indeed more influenced by financial incentives. Conversely, Panels C and D compare the post-inspection behavior of Stay Physicians versus Entry Physicians. Here, we find no significant differences in drug shares or C-section rates.

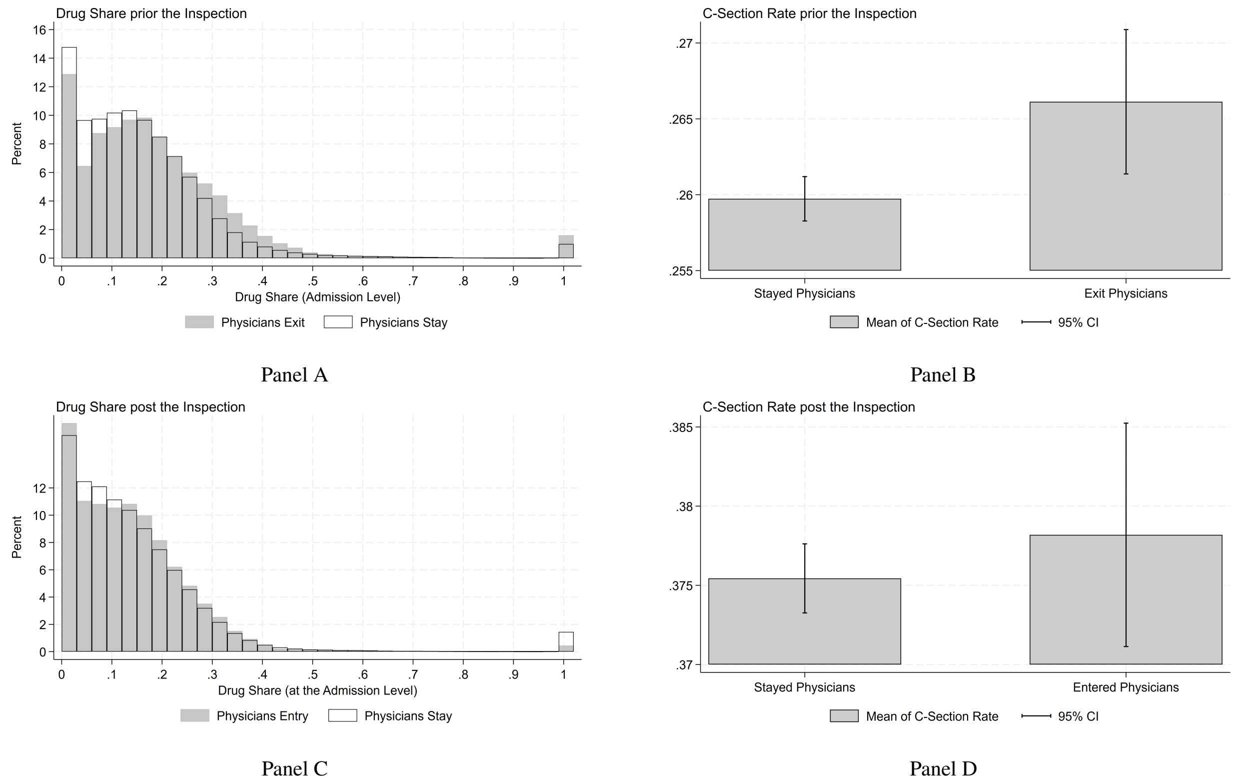


Figure 5: Comparison of Practice Styles: Exiting, Entering, and Incumbent Physicians

Notes: This figure examines the selection of physicians by comparing the practice patterns of those who exited or entered the hospital relative to those who remained (incumbent physicians). Panel A and Panel B utilize data from the pre-inspection period. Panel A plots the histogram of the drug share (defined as the ratio of drug expenditures to total expenditures per admission) for physicians who exited the hospital versus incumbent physicians. Panel B compares the average C-section rates between exiting and incumbent physicians. Panel C and Panel D utilize data from the post-inspection period. Panel C plots the histogram of the drug share for newly entering physicians versus incumbent physicians. Panel D compares the average C-section rates between entering and incumbent physicians. In Panels B and D, the height of the bars represents the mean value, and the vertical lines indicate the 95% confidence intervals.

Collectively, these findings validate the selective attrition channel. The inspections induced a compositional shift in the physician workforce: those with the highest propensity for rent-seeking (proxied by high drug shares and C-section rates) self-selected out of the hospital. They were replaced by, or left behind, a workforce that practiced less distorted medicine. This ”purification” of the physician pool contributed significantly to the observed improvements in neonatal and maternal health.

6.5 Rule Out Patient Sorting and Cherry Picking

Our preceding analyses have established that inspections incentivized physicians to increase the supply of prenatal care (via improved efficiency) and purged rent-seeking physicians (via selective attrition), thereby improving health outcomes. However, a critical threat to the validity of these findings remains: did the inspections alter physician preferences regarding which patients to admit?

Specifically, we must determine whether the observed health improvements are driven by genuine quality gains or merely by "cherry-picking"—a form of strategic patient sorting where hospitals preferentially admit lower-risk patients while turning away complex cases.

Theoretically, physicians facing inspections have dual incentives to engage in such behavior. First, avoiding high-risk patients reduces the likelihood of medical malpractice disputes during a sensitive period of scrutiny. Second, by selecting healthier patients, physicians can reduce labor intensity to compensate for the demoralizing loss of gray income. If such sorting occurred, the improvement in neonatal and maternal health would be a statistical artifact of a healthier patient pool rather than a social welfare gain; indeed, excluding high-risk patients would represent a net welfare loss. To test for this, we estimate Equation 2 at the patient level to verify whether the ex-ante risk profile of admitted patients in treated hospitals changed relative to control hospitals.

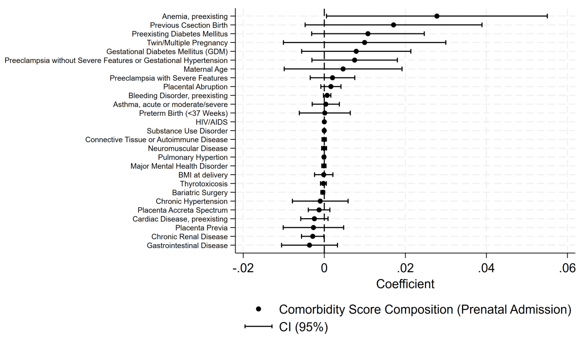
We begin by examining maternal age, a fundamental proxy for obstetric risk, as advanced maternal age is strongly associated with adverse outcomes (Flenady et al., 2011). If physicians were actively avoiding risk, we would expect a decrease in the average age of admitted patients. Table 7 presents these results. Columns 1–4 report the impact on average age across pregnancy loss, prenatal, delivery, and postnatal admissions. We find no significant decline in patient age in treated hospitals post-inspection. Furthermore, Columns 5–8 estimate the probability of admitting patients aged 35 or older (advanced maternal age). Contrary to the cherry-picking hypothesis, Column 7 reveals that for delivery admissions, treated hospitals actually admitted more high-risk mothers (aged $\geq$ 35) following the inspection.

Table 7: Impact of Hospital Inspection on Patient Age Composition

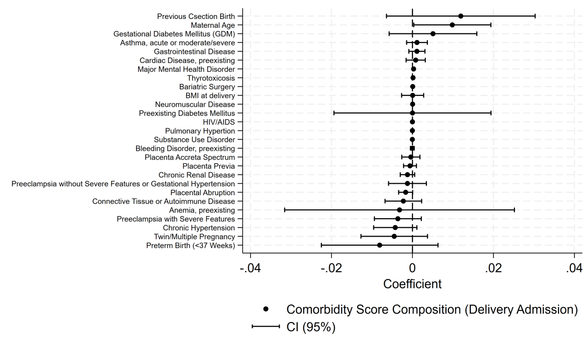
Outcome	Age				$\mathbf{1}[Age \geq 35]$			
Subgroup	PregLoss	Prenatal	Delivery	Postnatal	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D	0.12	0.05	0.04	-0.25	0.01	0.00	0.01**	-0.03
	(0.12)	(0.12)	(0.06)	(0.28)	(0.01)	(0.01)	(0.00)	(0.02)
Observations	63,154	76,785	445,793	5,775	63,154	76,785	445,793	5,775
R-squared	0.032	0.062	0.069	0.096	0.020	0.022	0.022	0.051
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	29.90	28.62	28.78	29.03	0.230	0.132	0.126	0.172
Magnitude (%)	0.393	0.170	0.125	-0.877	4.542	3.531	7.808	-17.10

Notes: This table examines potential changes in patient composition with respect to age following hospital inspections. The unit of analysis is the individual admission. Columns (1)–(4) use the patient's continuous age (in years) as the dependent variable. Columns (5)–(8) use a binary indicator equal to one if the patient is aged 35 or older as the dependent variable. The sample is stratified by admission type: Pregnancy Loss (Cols 1 & 5), Prenatal (Cols 2 & 6), Delivery (Cols 3 & 7), and Postnatal (Cols 4 & 8). Row D reports the estimated coefficients on the interaction term between the treatment dummy and the post-inspection dummy. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To capture clinical risk with greater precision than age alone, we further analyze diagnostic records to identify the presence of 27 specific obstetric comorbidities at the time of admission. Following Leonard et al. (2020), these comorbidities are established predictors of severe maternal morbidity. The regression results are illustrated in Figure 6[8]. Panel A, which focuses on prenatal admissions, reveals a crucial pattern: rather than avoiding sick patients, treated hospitals significantly increased their intake of high-risk expectant mothers. Specifically, the probability of admitting patients with conditions such as preexisting anemia, previous C-sections, and preexisting diabetes mellitus rose significantly post-inspection. Panel B, focusing on delivery admissions, shows no significant reduction in the prevalence of these comorbidities relative to the control group.



Panel A: Impact among Prenatal Admissions



Panel B: Impact among Delivery Admissions

Figure 6: Impact on the Composition of Obstetric Comorbidities

Notes: This figure plots the estimated regression coefficients for the likelihood of 27 specific obstetric comorbidities present at the time of admission. Panel A reports estimates for the sample of prenatal admissions, while Panel B reports estimates for the sample of delivery admissions. The dependent variable in each regression is a binary indicator equal to one if the patient possesses the specific comorbidity listed on the y-axis, and zero otherwise. The set of 27 comorbidities is selected following the clinical risk classification by Leonard et al. (2020) as predictors of severe maternal morbidity. The dots represent the point estimates of the treatment effect, and the horizontal lines indicate the 95% confidence intervals. Standard errors are clustered at the hospital level.

In summary, we find no evidence that treated hospitals engaged in cherry-picking or cream-skimming behavior. On the contrary, the data indicates that inspections led to an expansion of prenatal care specifically for higher-risk patients (e.g., those with preexisting conditions). This finding not only rules out patient sorting as a confounder but also strengthens our main conclusion: the observed improvements in neonatal and maternal health occurred despite an influx of riskier patients, suggesting that the true improvement in quality of care may be even larger than our baseline estimates suggest.

[8] We define obstetric comorbidity based on the 27 Comorbidity Diagnosis Groups outlined in Appendix 2 of Leonard et al. (2020). See https://cdn-links.lww.com/permalink/aog/b/aog_136_3_2020_07_02_leonard_20-592_sdc1.pdf for details.

6.6 Rule Out Quality Monitoring

Finally, we address an alternative explanation related to hospital infrastructure. The 2015 inspection guidelines included provisions for upgrading hospital information systems. It is therefore plausible that the observed health improvements were not driven by changes in financial incentives, but rather by enhanced quality monitoring (Miller & Tucker, 2011). Specifically, if inspections forced hospitals to upgrade their Medical Information Systems (MIS) and Electronic Medical Records (EMR), the resulting increase in data transparency could have subjected physicians to tighter scrutiny, deterring negligence and mechanically improving outcomes.

To test this channel, we employ a heterogeneity analysis based on the pre-inspection status of hospital IT infrastructure. Using the 2014 Hospital Statistical Yearbook (HSY), we stratified hospitals into two groups: (1) Established IT Hospitals, which already possessed functional MIS and EMR systems in 2014; and (2) Laggard IT Hospitals, which lacked these systems in 2014. We hypothesize that If the ”quality monitoring” channel is the primary driver, we should observe larger treatment effects in the ”Laggard IT” group. These hospitals had the most room for technological improvement and would have experienced the sharpest increase in monitoring capacity following the mandatory upgrades. Conversely, hospitals with established systems should see a smaller marginal effect.

Table 8 reports the results of this subsample analysis. Contrary to the quality monitoring hypothesis, Columns 1–4 reveal that the improvements in neonatal and maternal health are predominantly concentrated in the Established IT Hospitals—those that already had robust systems in place prior to the inspection. The ”Laggard IT” group showed no significant health gains. This finding suggests that the inspection's impact was not mediated by the introduction of monitoring technology. Consequently, we can effectively rule out improved quality monitoring as the primary mechanism driving the observed health outcomes.

Table 8: Heterogeneous Effects on Health Outcomes by Hospital Information Technology Status

	Hospital W/ MIS and EMR in 2014				Hospital W/O MIS or EMR in 2014			
Health Outcomes	BirthWeight	LBW	ARDS	HyperT_Disorder	BirthWeight	LBW	ARDS	HyperT_Disorder
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D	13.62**	-6.42	-0.17***	-9.34**	7.68	-4.87	0.07	-4.09
	(6.44)	(4.08)	(0.05)	(4.26)	(9.99)	(3.32)	(0.09)	(9.92)
Observations	329,901	329,901	329,901	329,901	115,892	115,892	115,892	115,892
R-squared	0.033	0.039	0.000	0.030	0.028	0.036	0.000	0.033
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	3240	81.39	0.122	67.06	3216	98.54	0.0670	91.77
Magnitude (%)	0.420	-7.882	-141.4	-13.93	0.239	-4.946	109.1	-4.462

Notes: This table reports the heterogeneous treatment effects on neonatal and maternal health outcomes, stratified by the hospital's adoption of information technology as of 2014. Columns (1)–(4) restrict the sample to hospitals equipped with both a Management Information System (MIS) and Electronic Medical Records (EMR) in 2014. Columns (5)–(8) correspond to the subsample of hospitals without these systems (W/O MIS or EMR). The dependent variables are defined as follows: BirthWeight denotes the birth weight of the newborn in grams; LBW serves as an indicator for low birth weight cases; ARDS indicates the presence of Acute Respiratory Distress Syndrome; and HyperT_Disorder indicates the presence of hypertensive disorders of pregnancy. The row labeled D represents the point estimates for the treatment effect. All regressions include hospital fixed effects and time fixed effects. "Pre-treat. Mean (T)" reports the mean of the dependent variable for the treatment group prior to the intervention. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7 Cost-Benefit Analysis

This section presents a back-of-the-envelope cost–benefit analysis of the hospital inspection policy, translating our estimated treatment effects into monetary terms. We adopt a deliberately conservative approach to assess whether the observed improvements in health outcomes plausibly justify the policy-induced increase in medical expenditures.

7.1 Policy Costs

We define the "cost" of the policy as the aggregate increase in medical expenditure observed in treated hospitals. Our sample includes 32 hospitals subject to inspection with varying exposure periods: 14 hospitals from the first round (25 post-treatment months each), 11 from the second round (15 months each), and 7 from the late rollout (3 months each). This yields a total of 536 treated hospital-months.

Consistent with our baseline results, total inpatient expenditures at treated hospitals increased by RMB 295,405 per hospital-month following the inspection. Aggregating this increase across all treated months yields a total policy cost of approximately RMB 158.34 million ($536 \times 295,405$). We interpret this figure as the fiscal and resource cost associated with tighter oversight, higher compliance, and the structural shift in clinical practice induced by the anti-corruption inspections.

7.2 Policy Benefit

We quantify policy benefits through two primary channels: improvements in neonatal health and improvements in maternal health.

Neonatal Health Benefits

The first source of benefits arises from improved neonatal outcomes, specifically the reduction in very low birth weight (VLBW) cases and the general rightward shift in the birth weight distribution.

First, regarding short-term medical savings, our estimates indicate that inspections reduce the probability of VLBW births by 0.233 percentage points. Given that 105,888 newborns were delivered in treated hospitals during the post-inspection period, this implies the avoidance of approximately 246 VLBW cases. Based on anecdotal and clinical evidence from China, the direct medical cost of treating a VLBW infant ranges between RMB 150,000 and RMB 500,000. Applying this range to the avoided cases yields short-term medical savings of RMB 36.9 million to RMB 123.0 million.

Second, regarding long-term productivity, improved neonatal health translates into life-long human capital gains (Black et al., 2007; Almond et al., 2018; Bharadwaj et al., 2018). Prior literature suggests that a one-standard-deviation increase in birth weight is associated with at least a 2.75 percent increase in adult annual income (Lambiris et al., 2022). Our estimates show a mean increase in birth weight of 13 grams. Assuming a discount rate of 3 percent, a real wage growth rate of 1 percent, and a 40-year working life, we calculate the present value of lifetime income gains. [9] Using the average annual wage in the non-private sector as a benchmark implies a per-child lifetime income gain of RMB 1,402, corresponding to an aggregate social benefit of RMB 148.45 million. Using the more conservative private-sector wage benchmark yields a per-child gain of RMB 863 and a total benefit of RMB 91.38 million.

Maternal Health Benefits

The second source of benefits stems from improved maternal health. Our results indicate a 0.608 percentage point reduction in the incidence of cervical lacerations, implying that approximately 643 women avoided this complication in the post-treatment period.

We quantify these benefits through reduced hospitalization costs and avoided productivity losses. Cervical lacerations typically prolong hospital stays and delay recovery. In our data, the average daily cost of a delivery-related stay is RMB 950. If avoiding a laceration reduces

[9] The present value is calculated as $PV = [p \times w_0] \times \sum_{t=1}^{T} \frac{(1+g)^{t-1}}{(1+r)^t}$, where $r = 3\%$, $g = 1\%$, $T = 40$, w_0 is the baseline annual wage (RMB 74,318 for urban non-private units; RMB 45,761 for private units in 2017), and p represents the percentage income increase, calculated as $2.75\% \times (13/515)$.

hospitalization by one day, the aggregate medical savings amount to RMB 0.61 million; if it reduces the stay by two days, savings rise to RMB 1.22 million. Furthermore, assuming that cervical laceration extends postpartum recovery (and thus delays return to work) by five days, and applying an average daily wage of RMB 297, the reduction in cases prevents approximately RMB 0.95 million in lost productivity.

7.3 Net Social Benefits

To construct a conservative estimate, we utilize the midpoints of the relevant ranges. The average short-term benefit from neonatal medical savings is RMB 79.95 million (= $\frac{36.9+123}{2}$), while the average long-term benefit from human capital accumulation is RMB 119.92 million (= $\frac{148.45+91.38}{2}$). For maternal health, the combined savings from medical costs and preserved productivity amount to approximately RMB 1.87 million. Summing these components yields a total social benefit of RMB 201.74 million.

Comparing these benefits against the total policy cost of RMB 158.34 million, we estimate a net social benefit of RMB 43.40 million (approximately USD 6.16 million). This suggests that even under conservative assumptions, the hospital inspection policy generates positive net returns to society.

7.4 Discussion and Conservativeness

This cost–benefit analysis is intentionally simplified and likely represents a lower bound of the true welfare gains. Unlike comprehensive frameworks such as Baran et al. (2025), our calculation excludes the economic value of preventing infant mortality, childhood disability, and adult disability associated with VLBW births—factors that constitute the majority of social costs in high-income settings (estimated at over $1.3 million per VLBW case for mortality risk alone). We exclude these components due to the scarcity of credible, context-specific parameter estimates for China. However, accounting for these omitted dimensions would substantially magnify the estimated benefits, further reinforcing our conclusion that hospital anti-corruption inspections yield significant social welfare gains.

8 Conclusion

This paper provides the comprehensive empirical evidence on the welfare consequences of anti-corruption enforcement within China's public hospital sector. Exploiting the staggered rollout of the 2015 hospital inspections, we document a robust causal link between tighter oversight and improved patient outcomes. Our analysis yields five primary conclusions.

First, regarding healthcare utilization, we find that inspections significantly increased total inpatient expenditures. Crucially, this rise was driven entirely by the extensive margin—a substantial expansion in admission volume—rather than the intensive margin of expenditure per admission. Second, this expansion in care translated into tangible improvements in neonatal and maternal health, manifested by significant reductions in low birth weight (LBW) incidence, fewer maternal complications (such as hypertensive disorders and lacerations), and a distributional shift of birth weights toward the optimal mean.

Third, our heterogeneity analysis reveals that these effects were not uniform; they were predominantly concentrated among hospitals and physicians with a higher ex-ante reliance on gray income—specifically, hospitals with high baseline Cesarean section rates and senior attending physicians. This suggests the intervention effectively targeted the margins where agency problems were most acute. Fourth, we identify the mechanisms driving these results as a combination of income compensation and efficiency gains. As inspections curtailed pharmaceutical kickbacks, physicians substituted away from drug-centric revenue models toward the provision of legitimate, labor-intensive prenatal services. This supply shock was facilitated by improved bed turnover efficiency (reduced ALOS for deliveries) and a ”cleansing” effect via the selective attrition of rent-seeking physicians. Finally, our cost–benefit analysis demonstrates that the policy generated a positive net social benefit: the long-term human capital and health gains from improved outcomes conservatively outweigh the fiscal costs associated with increased medical utilization.

Our findings offer two salient policy implications for healthcare governance in developing economies where state capacity is evolving and corruption remains a friction.

First, our results support the normalization and institutionalization of anti-corruption inspections. We find no evidence that the ”shock” of inspections compromised patient safety or induced harmful defensive medicine; instead, it improves hospital performance and health outcomes, and generates net social gains. Consequently, policymakers should transition from sporadic, campaign-style crackdowns to a permanent, rigorous audit regime. Establishing a credible, persistent threat of detection can permanently alter the risk-return calculus of rent-seeking, driving hospitals to compete on clinical efficiency and service quality rather than pharmaceutical arbitrage.

Second, and perhaps more fundamentally, our mechanism analysis highlights the necessity of increasing legitimate physician income. Our findings indicate that physicians are highly responsive to financial incentives: when illicit income channels were blocked, they aggressively expanded legitimate service supply to recover lost revenue. This implies that the pre-inspection prevalence of corruption was, in part, a symptom of misaligned pecuniary incentives where formal wages were below the market equilibrium. Therefore, successful anti-corruption efforts must be coupled with structural wage reforms. Policymakers should decouple physician income from drug sales while simultaneously raising service fees and base salaries to reflect the true value of clinical labor. By aligning financial rewards with patient welfare, the health system can sustain the quality improvements observed in this study without relying solely on the punitive threat of inspections.

In sum, this paper demonstrates that state-led anti-corruption interventions can succeed not only in curbing malfeasance but also in correcting the misallocation of medical resources, ultimately fostering a more efficient healthcare system.

References

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, *56*(4), 1360–1446.

Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant'Anna, P. H. (2025). Difference-in-differences designs: A practitioner's guide. *arXiv preprint arXiv:2503.13323*.

Baran, C., Currie, J., Dursun, B., & Tekin, E. (2025). *Clean rides, healthy lives: The impact of electric vehicle adoption on air quality and infant health* (tech. rep.). National Bureau of Economic Research.

Best, K. E., Tennant, P. W., & Rankin, J. (2017). Survival, by birth weight and gestational age, in individuals with congenital heart disease: A population-based study. *Journal of the American Heart Association*, *6*(7), e005213.

Bharadwaj, P., Lundborg, P., & Rooth, D.-O. (2018). Birth weight in the long run. *Journal of Human Resources*, *53*(1), 189–231.

Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in uganda. *The Quarterly Journal of Economics*, *124*(2), 735–769.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the cradle to the labor market? the effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, *122*(1), 409–439.

Braun, D., Braun, E., Chiu, V., Burgos, A. E., Gupta, M., Volodarskiy, M., & Getahun, D. (2020). Trends in neonatal intensive care unit utilization in a large integrated health care system. *JAMA network open*, *3*(6), e205239.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230.

Cherecheş, R. M., Ungureanu, M. I., Sandu, P., & Rus, I. A. (2013). Defining informal payments in healthcare: A systematic review. *Health Policy*, *110*(2–3), 105–114. https://doi.org/10.1016/j.healthpol.2013.01.010

Clemens, J., & Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, *104*(4), 1320–1349.

Currie, J., & Gruber, J. (1996). Saving babies: The efficacy and cost of recent changes in the medicaid eligibility of pregnant women. *Journal of political Economy*, *104*(6), 1263–1296.

Currie, J., Lin, W., & Meng, J. (2014). Addressing antibiotic abuse in china: An experimental audit study. *Journal of development economics*, *110*, 39–51.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, *95*(5), 1525–1547.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review*, *110*(9), 2964–2996.

Fang, H., Lei, X., Shi, J., & Yi, X. (2021). Physician-induced demand: Evidence from china's drug price zero-markup policy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3880231

Flenady, V., Koopmans, L., Middleton, P., Frøen, J. F., Smith, G. C., Gibbons, K., Coory, M., Gordon, A., Ellwood, D., McIntyre, H. D., et al. (2011). Major risk factors for stillbirth in high-income countries: A systematic review and meta-analysis. *The lancet*, *377*(9774), 1331–1340.

Fu, H., Lai, Y., Li, Y., Zhu, Y., & Yip, W. (2023). Understanding medical corruption in china: A mixed-methods study. *Health Policy and Planning*, *38*(4), 496–508. https://doi.org/10.1093/heapol/czad015

García, P. J. (2019). Corruption in global health: The open secret. *The Lancet*, *394*(10214), 2119–2124.

Gerardino, M. P., Litschig, S., & Pomeranz, D. (2024). Distortion by audit: Evidence from public procurement. *American Economic Journal: Applied Economics*, *16*(4), 71–108.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, *225*(2), 254–277.

Gruber, J., & Owings, M. (1996). Physician financial incentives and cesarean section delivery. *RAND Journal of Economics*, *27*(1), 99–123.

Gupta, A. (2021). Impacts of performance pay for hospitals: The readmissions reduction program. *American Economic Review*, *111*(4), 1241–1283.

Johnson, E. M., & Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, *8*(1), 115–141.

Lambiris, M. J., Blakstad, M. M., Perumal, N., Danaei, G., Bliznashka, L., Fink, G., & Sudfeld, C. R. (2022). Birth weight and adult earnings: A systematic review and meta-analysis. *Journal of Developmental Origins of Health and Disease*, *13*(3), 284–291.

Leonard, S. A., Kennedy, C. J., Carmichael, S. L., Lyell, D. J., & Main, E. K. (2020). An expanded obstetric comorbidity scoring system for predicting severe maternal morbidity. *Obstetrics & Gynecology*, *136*(3), 440–449.

Lewis, M. (2007). Informal payments and the financing of health care in developing and transition countries. *Health Affairs*, *26*(4), 984–997. https://doi.org/10.1377/hlthaff.26.4.984

Li, H.-T., Luo, S., Trasande, L., Hellerstein, S., Kang, C., Li, J.-X., Zhang, Y., Liu, J.-M., & Blustein, J. (2017). Geographic variations and temporal trends in cesarean delivery rates in china, 2008-2014. *Jama*, *317*(1), 69–76.

Liu, N., Bao, G., & He, A. J. (2020). Does health insurance coverage reduce informal payments? evidence from the "red envelopes" in china. *BMC Health Services Research*, *20*(1), 95. https://doi.org/10.1186/s12913-020-4955-7

Miller, A. R., & Tucker, C. E. (2011). Can health care information technology save babies? *Journal of Political Economy*, *119*(2), 289–324.

Mohanan, M., Donato, K., Miller, G., Truskinovsky, Y., & Vera-Hernández, M. (2021). Different strokes for different folks? experimental evidence on the effectiveness of input and output incentive contracts for health care providers with varying skills. *American Economic Journal: Applied Economics*, *13*(4), 34–69.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of political Economy*, *115*(2), 200–249.

Pursley, D. M., & Zupancic, J. A. (2020). Using neonatal intensive care units more wisely for at-risk newborns and their families. *JAMA Network Open*, *3*(6), e205693–e205693.

Shi, J., Liang, L., & Hou, M. (2023). The impacts of china's drug price zero-markup policy on medical expenditures and health outcomes. *China Economic Review*, *79*, 101949.

Shi, M. (2024). Monitoring for waste: Evidence from medicare audits. *The quarterly journal of economics*, *139*(2), 993–1049.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, *225*(2), 175–199.

Vance, A. J., Bell, S., Tilea, A., Beck, D., Tabb, K., & Zivin, K. (2023). Identifying neonatal intensive care (nicu) admissions using administrative claims data. *Journal of neonatal-perinatal medicine*, *16*(4), 709–716.

Vian, T. (2020). Anti-corruption, transparency and accountability in health: Concepts, frameworks, and approaches. *Global Health Action*, *13*(sup1), 1694744. https://doi.org/10.1080/16549716.2019.1694744

Wu, B. (2019). Physician agency in china: Evidence from a drug-percentage incentive scheme. *Journal of Development Economics*, *140*, 72–89.

Xu, H., & Yuan, M. (2022). The red packet phenomenon from the perspective of young chinese doctors: A questionnaire study. *BMC Medical Ethics*, *23*(1), 56. https://doi.org/10.1186/s12910-022-00793-w

Yip, W. C.-M., Hsiao, W., Meng, Q., Chen, W., & Sun, X. (2010). Realignment of incentives for health-care providers in china. *The Lancet*, *375*(9720), 1120–1130.

Zhang, Y., Betran, A., Li, X., Liu, D., Yuan, N., Shang, L., Lin, W., Tu, S., Wang, L., Wu, X., et al. (2022). What is an appropriate caesarean delivery rate for china: A multicentre survey. *BJOG: An International Journal of Obstetrics & Gynaecology*, *129*(1), 138–147.

A Appendix Tables and Figures

Table A1: Summary Statistics of Hospital Characteristics and Clinical Outcomes by Inspection Cohort

	(1) Inspected(2015)	(2) Inspected(2016)	(3) Inspected(2017)	(4) Control	(5) Full Sample
No. of Admissions	292.95	452.20	287.23	146.70	210.47
	(144.56)	(387.19)	(238.54)	(128.14)	(212.19)
No. of Pregnancy Admissions	208.47	336.52	208.45	135.85	173.09
	(108.23)	(303.21)	(174.61)	(122.38)	(167.82)
No. of Non-pregnancy Admissions	84.47	115.68	78.78	10.84	37.37
	(66.99)	(101.78)	(73.51)	(12.97)	(61.91)
No. of Medical Procedures	343.80	505.42	314.36	101.76	194.71
	(261.35)	(425.51)	(266.85)	(123.77)	(255.59)
Total Costs (RMB)	1,821,557	2,984,057	1,877,342	503,140	1,055,526
	(1,276,403)	(2,867,692)	(1,691,441)	(762,746)	(1,571,175)
Drug Costs (RMB)	342,225	520,010	402,315	78,858	186,645
	(247,027)	(461,918)	(395,770)	(107,690)	(278,252)
Diagnostic Costs (RMB)	406,897	668,025	344,755	72,348	203,467
	(312,426)	(615,275)	(298,903)	(85,500)	(329,928)
Surgical Costs (RMB)	325,214	491,774	382,191	110,089	201,027
	(230,391)	(391,824)	(366,200)	(109,799)	(247,225)
ALOS (Days)	7.30	7.47	7.51	5.59	6.17
	(1.16)	(1.31)	(0.96)	(1.37)	(1.56)
C-section Rate (%)	41.49	44.31	41.96	35.23	37.68
	(14.95)	(12.24)	(8.95)	(15.06)	(14.80)
VLBM Rate (%)	0.98	1.61	1.41	0.21	0.57
	(1.48)	(1.99)	(2.03)	(1.21)	(1.52)
Birth Weight (Grams)	3290.67	3259.43	3298.82	3350.52	3327.49
	(108.40)	(129.02)	(140.52)	(117.28)	(123.97)
Maternal Mortality Rate (/100,000)	1.5	4.6	3.3	49.1	33.7
	(34.1)	(66.0)	(30.7)	(814.0)	(664.8)
N	503	396	252	2455	3606

Notes: This table reports the means of key hospital-level variables, with standard deviations reported in parentheses. The unit of observation is the hospital-month. Columns (1)–(3) restrict the sample to the treatment groups, consisting of hospitals that underwent inspections in 2015, 2016, and 2017, respectively. Column (4) presents statistics for the control group, composed of hospitals that were not inspected. Column (5) reports summary statistics for the full sample. ”No. of Admissions” and subsequent admission counts represent the monthly volume of patients. Financial variables (”Total Costs,” ”Drug Costs,” ”Diagnostic Costs,” and ”Surgical Costs”) are expressed in RMB. ”ALOS” denotes the Average Length of Stay in days. ”C-section Rate” represents the percentage of deliveries performed via cesarean section. ”VLBM Rate” indicates the prevalence of very low birth mass (or weight) cases. ”Maternal Mortality Rate” is scaled as the number of maternal deaths per 100,000 admissions. *N* denotes the total number of hospital-month observations used in the analysis.

Table A2: Robustness Checks: The Effects of Hospital Inspections on Aggregated Expenditures

Aggregated Expenditure (RMB)	Full	NonPreg	Preg	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Zero Markup Drug Policy							
D	225,043**	25,193	199,850**	-637	48,831**	130,479*	9,695
	(99,965)	(36,524)	(81,091)	(5,891)	(19,409)	(68,652)	(5,859)
D_ZMDP	174,128***	43,997**	130,132***	2,575	22,318*	102,503**	4,625*
	(51,926)	(19,182)	(47,516)	(3,647)	(13,117)	(42,367)	(2,718)
Obs. (Hospital-Month Pair)	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.949	0.950	0.917	0.945	0.685	0.901	0.507
Pre-treat. Mean (T)	2,104,839	884,272	1,220,566	177,080	85,088	928,911	14,187
Magnitude (%)	10.69	2.849	16.37	-0.360	57.39	14.05	68.34
Panel B: Exclude neaby untreated hospitals							
D	330,088***	47,016	283,072***	1,567	60,671***	199,204***	11,620*
	(102,410)	(40,250)	(81,854)	(5,999)	(22,314)	(65,351)	(6,730)
Obs. (Hospital-Month Pair)	3,223	3,223	3,223	3,223	3,223	3,223	3,223
R-squared	0.961	0.954	0.937	0.954	0.689	0.922	0.498
Pre-treat. Mean (T)	2,104,839	884,272	1,220,566	177,080	85,088	928,911	14,187
Magnitude (%)	15.68	5.317	23.19	0.885	71.30	21.44	81.90
Panel C: Common Support							
D	235,301**	57,656	177,645**	4,980	45,244***	117,439	3,452**
	(100,102)	(42,381)	(76,868)	(3,758)	(14,760)	(73,686)	(1,684)
Obs. (Hospital-Month Pair)	2,763	2,763	2,763	2,763	2,763	2,763	2,763
R-squared	0.944	0.943	0.912	0.934	0.755	0.899	0.539
Pre-treat. Mean (T)	1,562,417	645,310	917,106	123,167	57,243	715,671	7,816
Magnitude (%)	15.06	8.935	19.37	4.043	79.04	16.41	44.17
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports robustness checks for the estimated effects of hospital inspections on aggregated hospital expenditures (measured in RMB). The unit of observation is the hospital-month. Columns (1)–(7) report estimates for different subsamples of admissions: the full sample (Full), non-pregnancy admissions (NonPreg), pregnancy-related admissions (Preg), pregnancy loss cases (PregLoss), prenatal care (Prenatal), delivery admissions (Delivery), and postnatal care (Postnatal). Panel A presents estimates controlling for the concurrent "Zero Markup Drug Policy." The variable D_ZMDP is an interaction term defined as $\mathbb{1}(\text{MunicipalHospitals}) \times \mathbb{1}(\text{PostDec2016})$, where $\mathbb{1}(\text{MunicipalHospitals})$ is an indicator for municipal-level hospitals and $\mathbb{1}(\text{PostDec2016})$ is an indicator for months after December 2016. Panel B tests for potential spillover effects by excluding uninspected (control) hospitals located within a 5-kilometer radius of any inspected hospital. Panel C restricts the sample to observations within the common support to ensure comparability between treatment and control groups. The row labeled D reports the coefficient on the main treatment term. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Robustness Checks: The Effects of Hospital Inspections on Admission Number

Admission Number	Full	NonPreg	Preg	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: ZMDP							
D	28.56**	3.52	25.04**	-0.42	13.84***	5.57	1.50*
	(11.76)	(2.34)	(10.97)	(0.87)	(4.29)	(9.10)	(0.89)
D_ZMDP	24.57***	6.62***	17.95**	0.46	7.60**	8.50	0.80*
	(8.44)	(1.91)	(7.83)	(0.68)	(3.39)	(7.79)	(0.42)
Obs. (Hospital-Month Pair)	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.954	0.959	0.936	0.961	0.736	0.908	0.574
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	323.6	90.99	232.6	36.36	24.07	166.2	2.363
Magnitude (%)	8.825	3.871	10.76	-1.154	57.50	3.353	63.29
Panel B: Exclude neaby untreated hospitals							
D	43.35***	6.46**	36.88***	-0.01	17.84***	12.68*	1.84*
	(11.20)	(2.62)	(10.32)	(0.77)	(4.83)	(7.62)	(1.03)
Obs. (Hospital-Month Pair)	3,223	3,223	3,223	3,223	3,223	3,223	3,223
R-squared	0.957	0.960	0.936	0.955	0.730	0.911	0.554
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	323.6	90.99	232.6	36.36	24.07	166.2	2.363
Magnitude (%)	13.40	7.104	15.86	-0.0329	74.12	7.628	77.70
Panel C: Common Support							
D	25.66**	5.47*	20.19**	-0.72	15.31***	1.58	0.71*
	(11.15)	(2.90)	(9.74)	(0.91)	(4.78)	(9.08)	(0.37)
Obs. (Hospital-Month Pair)	2,763	2,763	2,763	2,763	2,763	2,763	2,763
R-squared	0.952	0.955	0.937	0.958	0.758	0.911	0.617
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	257.3	69.77	187.5	28.90	19.58	134.2	1.576
Magnitude (%)	9.974	7.842	10.77	-2.481	78.19	1.176	45.08
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports robustness checks for the estimated effects of hospital inspections on the volume of admissions. The unit of observation is the hospital-month. Columns (1)–(7) report estimates for different subsamples of admissions: the full sample (Full), non-pregnancy admissions (NonPreg), pregnancy-related admissions (Preg), pregnancy loss cases (PregLoss), prenatal care (Prenatal), delivery admissions (Delivery), and postnatal care (Postnatal). Panel A presents estimates controlling for the concurrent "Zero Markup Drug Policy." The variable D_ZMDP is an interaction term defined as $\mathbb{1}(\text{MunicipalHospitals}) \times \mathbb{1}(\text{PostDec2016})$, where $\mathbb{1}(\text{MunicipalHospitals})$ is an indicator for municipal-level hospitals and $\mathbb{1}(\text{PostDec2016})$ is an indicator for months after December 2016. Panel B tests for potential spillover effects by excluding uninspected (control) hospitals located within a 5-kilometer radius of any inspected hospital. Panel C restricts the sample to observations within the common support to ensure comparability between treatment and control groups. The row labeled D reports the coefficient on the main treatment term. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: The Effects on Other Maternal and Fetal Health Outcomes

Other Health Outcomes	Maternal_Death	Fetal_Death	Severe_Fetal_Cond	Major_Cgam	Macrosomia	Malpractice
	(1)	(2)	(3)	(4)	(5)	(6)
D	-0.16	-0.70	-3.81	-0.10	0.91	0.14
	(0.13)	(0.65)	(5.92)	(0.43)	(1.96)	(0.26)
Obs. (Delivery Admissions)	445,793	445,793	445,793	445,793	445,793	445,793
R-squared	0.006	0.005	0.032	0.002	0.007	0.003
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	0.0412	7.344	121.7	2.359	75.44	0.546
Magnitude (%)	-384.5	-9.496	-3.130	-4.220	1.207	25.23

Notes: This table reports the estimated effects of hospital inspections on a set of additional maternal and fetal health outcomes. The sample is restricted to delivery hospitalizations. The dependent variables are binary indicators defined as follows: Maternal_Death takes the value of one if the mother died during childbirth; Fetal_Death takes the value of one if the case involved intrauterine fetal death or stillbirth; Severe_Fetal_Cond indicates the presence of severe fetal conditions, identified by ICD-10 codes O35 and O36; Major_Cgam indicates the presence of major congenital anomalies; Macrosomia indicates that the newborn was large for gestational age, defined as a birth weight exceeding 4,000 grams; and Malpractice is an indicator for cases involving complications of surgical anesthesia. The row labeled D reports the point estimates for the treatment effect. All regressions include hospital fixed effects and time fixed effects. "Pre-treat. Mean (T)" reports the mean of the dependent variable for the treatment group prior to the intervention. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: The Impact of Hospital Inspection on Average Length of Stay (ALOS)

Average Length of Stay	Full	NonPreg	Preg	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
D	-0.22***	-0.12	-0.23***	-0.06	0.08	-0.26**	-0.34
	(0.07)	(0.14)	(0.08)	(0.15)	(0.14)	(0.11)	(0.39)
Obs. (Hospital-Month Pair)	3,606	3,374	3,606	3,356	3,427	3,434	1,675
R-squared	0.911	0.594	0.886	0.484	0.435	0.770	0.276
Pre-treat. Mean (T)	7.594	10.21	10.21	7.230	6.364	6.632	7.653
Magnitude (%)	-2.935	-1.136	-2.255	-0.860	1.259	-3.977	-4.408
Panel B: With Weights							
D	-0.15*	-0.04	-0.16*	0.03	-0.06	-0.15*	0.10
	(0.08)	(0.12)	(0.09)	(0.14)	(0.16)	(0.09)	(0.32)
Obs. (Hospital-Month Pair)	3,606	3,374	3,606	3,356	3,427	3,434	1,675
R-squared	0.936	0.678	0.906	0.645	0.560	0.842	0.259
Weight	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	7.594	10.21	10.21	7.230	6.364	6.632	7.653
Magnitude (%)	-1.990	-0.426	-1.549	0.481	-0.911	-2.337	1.242
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table estimates the impact of hospital inspection on the Average Length of Stay (ALOS), measured in days, at the hospital-month level. Panel A reports the baseline unweighted estimates, while Panel B presents estimates weighted by the monthly volume of admissions to account for hospital size heterogeneity. Column (1) presents results for the full sample, which is decomposed into non-pregnancy (Col. 2) and pregnancy-related (Col. 3) admissions. Columns (4)–(7) further categorize pregnancy-related admissions into pregnancy loss, prenatal, delivery, and postnatal care. Row D reports the Difference-in-Differences coefficients. Magnitude (%) represents the percentage change relative to the pre-treatment mean of the treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: The Impact of Hospital Inspection on the Frequency of Short-Stay Admissions

	Non-delivery Admissions			Delivery Admissions		
Number of Short Admissions	#LOS1	#LOS2	#LOS3	#LOS1	#LOS2	#LOS3
	(1)	(2)	(3)	(4)	(5)	(6)
D	-0.33	3.45**	2.64***	-0.04	-1.06	-0.19
	(0.47)	(1.48)	(0.61)	(0.06)	(1.46)	(2.25)
Observations	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.817	0.863	0.680	0.471	0.877	0.770
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	0.841	5.733	8.334	0.0411	3.757	14.04
Magnitude (%)	-38.83	60.14	31.70	-106.3	-28.26	-1.366

Notes: This table estimates the effect of hospital inspection on the frequency of short-stay hospitalizations at the hospital-month level. Columns (1)–(3) focus on non-delivery admissions, while Columns (4)–(6) focus on delivery admissions. The dependent variables are the monthly counts of admissions with specific durations: #LOS1, #LOS2, and #LOS3 denote the number of admissions with lengths of stay of 1 day (same-day discharge), 2 days, and 3 days, respectively. Row D reports the Difference-in-Differences coefficients, indicating the change in the monthly volume of these specific cases. Magnitude (%) represents the percentage change relative to the pre-treatment mean of treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: The Impact of Hospital Inspection on Components of Average Expenditure per Admission

VARIABLES	Drug_C	Service_C	Diagnosis_C	Consumable_C
	(1)	(2)	(3)	(4)
Panel A: Baseline				
D	-209.19**	173.73***	-20.53	-5.63
	(93.58)	(49.28)	(31.04)	(33.94)
Observations	3,606	3,606	3,606	3,606
R-squared	0.846	0.892	0.941	0.922
Pre-treat. Mean (T)	1313	2623	1283	941.9
Magnitude (%)	-15.94	6.623	-1.599	-0.598
Panel B: With Weights				
D	-174.21*	103.93*	26.57	-32.23
	(101.47)	(59.20)	(46.57)	(39.89)
Observations	3,606	3,606	3,606	3,606
R-squared	0.853	0.900	0.951	0.936
Weight	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	1313	2623	1283	941.9
Magnitude (%)	-13.27	3.962	2.070	-3.421
Hospital FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: This table examines the effect of hospital inspection on the distinct components of average expenditure per admission. The dependent variables are the average costs associated with Drugs (Col. 1), Medical Services (Col. 2), Diagnostics (Col. 3), and Consumables (Col. 4). Panel A presents the baseline unweighted estimates, while Panel B weights regressions by monthly admission volume. Row D reports the Difference-in-Differences coefficients. Magnitude (%) represents the percentage change relative to the pre-treatment mean of treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Impact of Hospital Inspection on the Prevalence of Zero- and Low-Drug-Expenditure Admissions

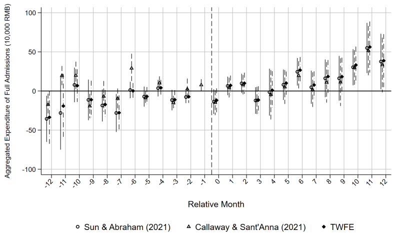
Outcomes	#Drug_C_0	#Drug_C_10	#Drug_C_50	#Drug_C_100	%Drug_C_0	%Drug_C_10	%Drug_C_50	%Drug_C_100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D	1.64	3.86*	6.74*	10.32**	0.00	0.01*	0.03**	0.03**
	(1.12)	(1.95)	(3.99)	(4.24)	(0.00)	(0.01)	(0.01)	(0.01)
Observations	3,606	3,606	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.711	0.762	0.823	0.856	0.785	0.787	0.840	0.868
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weight	No	No	No	No	Yes	Yes	Yes	Yes
Pre-treat. Mean (T)	8.873	12.66	27.98	39.01	0.0314	0.0453	0.100	0.138
Magnitude (%)	18.51	30.49	24.10	26.46	10.56	31.14	25.27	20.44

Notes: This table investigates the impact of hospital inspection on the prevalence of admissions with zero or very low drug expenditures at the hospital-month level. Columns (1)–(4) report the effects on the number (#) of admissions, while Columns (5)–(8) report the effects on the proportion (%) of admissions falling below specific drug cost thresholds. The suffixes _0, _10, _50, and _100 correspond to thresholds of 0 RMB, 10 RMB, 50 RMB, and 100 RMB, respectively. Regressions for counts (Cols 1–4) are unweighted, whereas regressions for proportions (Cols 5–8) are weighted by monthly admission volume. Row D reports the Difference-in-Differences coefficients.Magnitude (%) represents the percentage change relative to the pre-treatment mean of treatment group (Pre-treat. Mean (T)). All specifications include Hospital and Time Fixed Effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

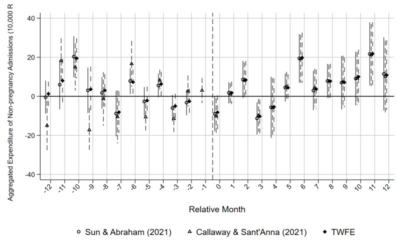
Table A9: Heterogeneous Effects on Service and Drug Expenditures by Admission Type

VARIABLES	NonPreg	Preg	PregLoss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Service Expenditure of Different Admission Groups						
D	221.40***	169.79***	160.57***	192.41**	-36.56	197.12
	(72.83)	(52.77)	(56.48)	(74.57)	(79.01)	(163.70)
Observations	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.862	0.879	0.844	0.811	0.856	0.293
Pre-treat. Mean (T)	2990	2379	1638	1289	2703	1618
Magnitude (%)	7.405	7.138	9.804	14.92	-1.353	12.19
Panel B: Drug Expenditure of Different Admission Groups						
D	-301.05**	-103.60	-85.44	-45.72	-138.83**	-104.20
	(118.37)	(73.05)	(86.17)	(66.81)	(68.95)	(98.23)
Observations	3,606	3,606	3,606	3,606	3,606	3,606
R-squared	0.744	0.809	0.598	0.534	0.783	0.306
Pre-treat. Mean (T)	2039	956.7	966.5	587.4	1019	1010
Magnitude (%)	-14.76	-10.83	-8.841	-7.784	-13.62	-10.31
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

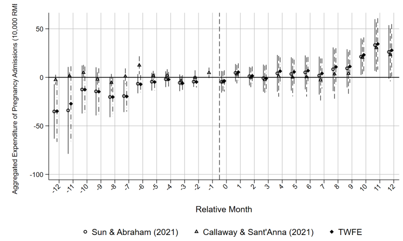
Notes: This table reports the difference-in-differences estimates of the impact of hospital inspection on specific components of medical expenditure across different admission categories. The unit of analysis is the hospital-month. Panel A presents estimates for average service expenditure per admission, while Panel B presents estimates for average drug expenditure per admission. Column (1) restricts the sample to non-pregnancy-related admissions. Column (2) restricts the sample to pregnancy-related admissions, which are further decomposed into pregnancy loss, prenatal care, delivery, and postnatal care in Columns (3)–(6), respectively. Row D reports the estimated coefficients on the interaction term between the treatment dummy and the post-inspection dummy. Magnitude (%) is calculated as the coefficient divided by the pre-treatment mean of the dependent variable for the treatment group (Pre-treat. Mean (T)). All regressions control for hospital fixed effects and time (month-by-year) fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.



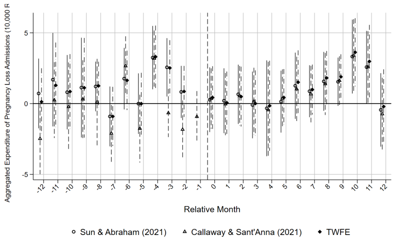
Panel A: Full Admissions



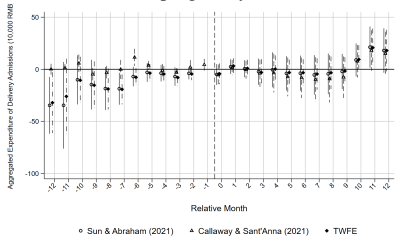
Panel B: Non-pregnancy Admissions



Panel C: Pregnancy Admissions



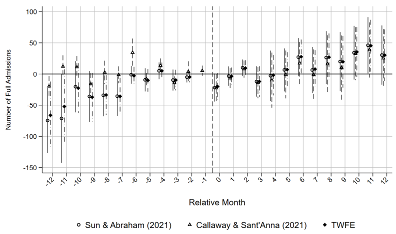
Panel D: Pregnancy Loss Admissions



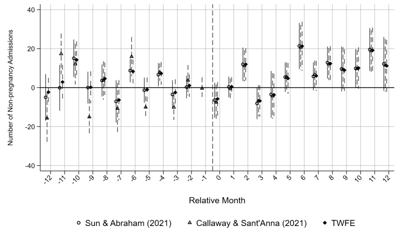
Panel E: Delivery Admissions

Figure A1: Event Study: Aggregated Expenditure

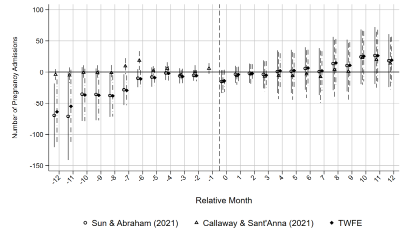
Notes: This figure plots the event study coefficients for the effect of hospital inspections on aggregated hospital expenditures at the hospital-month level. The dependent variable is measured in units of 10,000 RMB. Panel A reports estimates for the full sample of admissions. Panels B through E report estimates for specific subsamples: non-pregnancy admissions, pregnancy-related admissions, pregnancy loss cases, and delivery admissions, respectively. Within each panel, the figure compares point estimates and 95% confidence intervals derived from three distinct specifications: (i) the standard Two-Way Fixed Effects (TWFE) model (solid diamonds); (ii) the interaction-weighted estimator proposed by Sun and Abraham (2021) (hollow circles); and (iii) the estimator proposed by Callaway and Sant'Anna (2021) (solid triangles). The x-axis represents the time in months relative to the inspection date, with $k = -1$ serving as the omitted reference period. Standard errors are clustered at the hospital level.



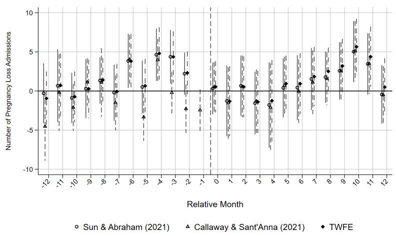
Panel A: Full Admissions



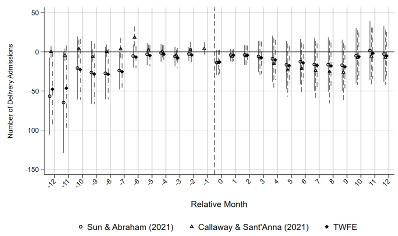
Panel B: Non-pregnancy Admissions



Panel C: Pregnancy Admissions



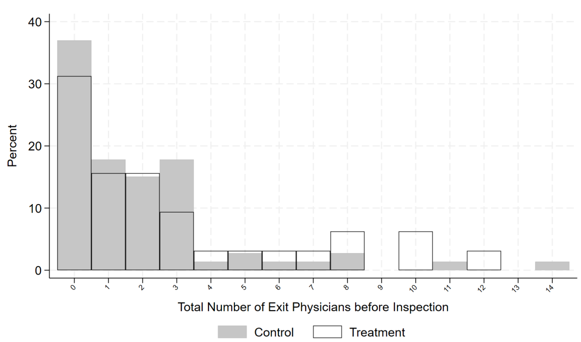
Panel D: Pregnancy Loss Admissions



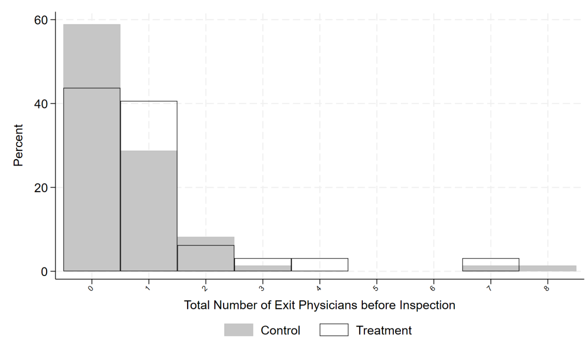
Panel E: Delivery Admissions

Figure A2: Event Study: Admission Number

Notes: This figure plots the event study coefficients for the effect of hospital inspections on the volume of admissions at the hospital-month level. Panel A reports estimates for the full sample of admissions. Panels B through E report estimates for specific subsamples: non-pregnancy admissions, pregnancy-related admissions, pregnancy loss cases, and delivery admissions, respectively. Within each panel, the figure compares point estimates and 95% confidence intervals derived from three distinct specifications: (i) the standard Two-Way Fixed Effects (TWFE) model (solid diamonds); (ii) the interaction-weighted estimator proposed by Sun and Abraham (2021) (hollow circles); and (iii) the estimator proposed by Callaway and Sant'Anna (2021) (solid triangles). The x-axis represents the time in months relative to the inspection date, with $k = -1$ serving as the omitted reference period. Standard errors are clustered at the hospital level.



Panel A: Without Time Restriction



Panel B: Within 6 Months before Inspection

Figure A3: Distribution of Pre-Inspection Physician Exits in Treated and Control Hospitals

Notes: This figure plots the distribution of the total number of physicians who exited the hospital prior to the inspection date. The sample includes both treated and control hospitals, distinguished by bar color (gray for Control, white for Treatment). Panel A reports the distribution of the cumulative number of physician exits calculated without any specific time window restriction (i.e., over the entire observed pre-inspection period). Panel B restricts the calculation to physician exits occurring within the six-month window immediately preceding the inspection date. The horizontal axis represents the count of exit physicians, and the vertical axis represents the percentage of hospitals within each bin.