Redirected Rents: Fraud Enforcement, Physician Behavior, and the Health Benefits of Anti-Corruption Inspections

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

July 31, 2026

Abstract

Enforcement against health-care fraud is designed to deter illicit provider behavior, but its full effect depends on what providers do with the margins that remain. We study staggered anti-corruption inspections of Chinese hospitals, an enforcement shock to physicians' illicit income from pharmaceutical kickbacks, and trace the response of care and health outcomes using all obstetrics and gynecology discharge records from a central province between 2015 and 2017. Using the plausibly exogenous inspection timing, we identify the enforcement effects on provider behavior and patient outcomes with a difference-in-differences design. Inspections improve maternal and infant health: mean birth weight rises by 18 grams and births cluster more tightly around the healthy norm, while genital tract lacerations and hypertensive disorders fall by about 11% and 14%. The mechanism is not deterrence alone but redirected demand inducement: drug use falls while fee-schedule service intensity and prenatal admissions for preventable conditions rise, and the increases are largest where baseline drug prescribing was most intensive. Because the next-best legitimate revenue channel in obstetrics, guideline-concordant prenatal care, is itself health-improving, the reallocation of physicians' rent-seeking effort is accidentally pro-health. The welfare effect of anti-fraud enforcement thus depends not only on the fraud it deters, but on the legitimate margins to which providers substitute.

Keywords: Anti-corruption inspection; informal payments; maternal and infant health; prenatal care; physician incentives; physician turnover

*Ding: Singapore Management University (email: ruiqi.ding.2024@phdecons.smu.edu.sg); Janssen: Singapore Management University (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).
Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare that they have no relevant or material financial interests, affiliations, or other relationships that relate to the research described in this paper. All remaining errors are our own.

1 Introduction

Corruption by agents is widespread, and it erodes both the efficiency and the equity of the services principals receive. The problem is especially consequential in health care, where physicians can exploit their informational and gatekeeping position to extract rents through manufacturer kickbacks, revenue-linked over-provision, and informal patient payments (Lewis, 2007; Chereches et al., 2013). Such rent extraction not only inflates spending but may also directly harm patient health. Governments have responded with a range of policies: disclosure mandates such as the U.S. Sunshine Act, procurement reforms, and anti-kickback enforcement (Pham-Kanter et al., 2012; Gerardino et al., 2024; Leder-Luis, 2025). Yet evidence on their effectiveness is mixed (Olken, 2007; Li et al., 2022; Shi, 2024), and the channels through which provider corruption affects patient health remain poorly understood.

China offers a setting where pharmaceutical kickbacks and informal payments persist despite explicit regulatory prohibitions (Xu & Yuan, 2022; Fu et al., 2023), and such gray income absorbs substantial resources and impedes progress toward universal coverage (García, 2019). Maternal and infant care is a domain where these distortions should matter most: delivery is time-sensitive and capacity-constrained, provider discretion is large, and severe maternal morbidity continues to rise (Chatterji et al., 2026). Yet there is little causal evidence on whether anti-corruption enforcement improves patient outcomes and through which channels (Vian, 2020).

Whether such enforcement helps or harms patients is far from obvious. Curtailing informal income may push providers to substitute toward legitimate, billable services; whether patients then gain depends on what those services happen to be, and the added official revenue it generates absorbs real resources whether or not it buys better health. The net social return is therefore unclear. The existing economics literature offers little direct guidance: it has evaluated formal payment reforms and audit programs in high-income settings, focusing on billed (official) revenues rather than the curtailment of informal earnings (Dafny, 2005; Clemens & Gottlieb, 2014; Fang et al., 2021; Gupta, 2021; Shi, 2024; Chen & Dillender, 2025). We take up this question by examining a large-scale anti-corruption campaign targeting gray income in one province of a large developing economy.

We study the campaign's instrument inside hospitals: the 2015 round of China's national hospital inspection program, the first round wielded as a punitive anti-corruption tool aimed at this gray income. To identify the causal effect of the inspections on physician behavior and patient health, we use administrative admission records from the obstetrics and gynecology (OBGYN) departments of hospitals across a central Chinese province. Because these records report both healthcare utilization and clinical outcomes for mothers and infants, we can follow the policy from provider supply responses through to health and a cost-benefit assessment. The inspections rolled

out on a staggered schedule, so we estimate event-study and difference-in-differences (DiD) models that compare inspected hospitals with not-yet-inspected ones; identification rests on the timing of each inspection, which we argue is plausibly uncorrelated with short-run OBGYN demand. Obstetrics is a deliberate choice: care is drug- and antibiotic-intensive, so the gray-income channel has room to contract; demand is near-exogenous to the inspection date; and outcomes arrive within months, inside our short panel. Conceptually, we treat the campaign as a shock to the *informal* incentive structure (commercial bribery and off-schedule payments) rather than a change in official price schedules, and we ask where the displaced effort goes. The answer hinges on a feature of obstetrics we document below: once the drug channel is compromised, the next-best legitimate way to defend revenue is to bill more services and admit more women for prenatal care, which is itself health-improving.

The inspections increased OBGYN activity. Total monthly OBGYN revenue per hospital rose by 18% following an inspection, or 401,940 RMB against a pre-inspection mean of 2.25 million.[1] The largest proportional gains are in prenatal (antenatal) and postnatal care, where revenue rose by 79% each, and the expansion runs along the extensive margin: prenatal admissions climbed by 77% while revenue per admission and average length of stay were unchanged. Health outcomes improved in step. Mean birth weight rose by about 18 grams, NICU admissions and the incidence of low and very low birth weight fell, and genital tract lacerations and hypertensive disorders of pregnancy declined by about 11% and 14%; rarer severe complications declined as well. These patterns suggest rents were redirected rather than eliminated, toward a legitimate channel that happens to be pro-health.

Why would tighter anti-corruption enforcement improve health? Our evidence points to income compensation: the campaign cut illicit gray income without raising formal wages, and physicians and hospitals, once the illicit pharmaceutical channel became more costly, reallocated activity toward legitimate revenue-generating services, many of which in obstetrics are health-improving. Three sets of findings support this mechanism. First, provider behavior moves as the hypothesis predicts. Per-admission revenue shifts almost one-for-one from drugs to services (a 227 RMB fall in drug revenue offset by a 229 RMB rise in service revenue), a reallocation visible even within individual physicians' own caseloads and across most hospital departments; physicians expand discretionary prenatal care for medium-risk pregnancies, such as those with anemia, a prior C-section, or gestational diabetes; and, once we adjust for case mix, they perform fewer C-sections, with the decline concentrated among older mothers. Second, the response follows a dose-response gradient. The behavioral shift is largest where the pre-inspection stakes were highest: among

[1] All magnitudes are derived from administrative data. Monetary values are reported in nominal RMB throughout the paper; consumer prices in the study province rose by under four percent in total across the 2015–2017 window, so deflating would not move any figure we report materially.

providers with the most intensive baseline drug prescribing, the highest baseline C-section rates, and the seniority to capture rents in the first place. For the infant-health outcomes and hypertensive disorders, so are the downstream health gains. Third, the inspections induced selective turnover: the physicians most likely to exit were those with the most rent-intensive practice styles, and the health gains concentrate in the hospitals that lost them. A battery of tests finds no evidence for the leading alternatives: the improvements are not driven by cherry-picking healthier patients, by patient socioeconomic status or a differential decline in "red-packet" payments, or by hospitals' baseline information-system capacity.

We close the analysis with a conservative cost–benefit calculation, weighing the added hospital revenue against the monetized value of the health gains: short-term medical savings on averted very-low-birth-weight cases and the long-term human-capital returns to higher birth weight. On a benchmark that projects real wages forward at two percent a year (a fraction of the eight to ten percent urban China recorded between 2000 and 2017), and omitting averted infant mortality and long-term disability entirely, the benefits exceed the added revenue by about 22 million RMB. We read this as indicating that the enforcement paid for itself, while treating the precise magnitude with caution.

Our findings contribute to three strands of literature. First, we speak to a growing but unsettled literature on whether policy tools aimed at curbing industry-driven rents in medicine actually change physician behavior. Evidence from transparency-based instruments is modest and heterogeneous: payment disclosure under the U.S. Physician Payments Sunshine Act and its state-level precursors, and institutional restrictions on detailing and gifts, shift prescribing only for particular drug classes or subsets of physicians (Pham-Kanter et al., 2012; Larkin et al., 2017; Li et al., 2022; Newham & Valente, 2024). We study a structurally different instrument — unannounced, sanction-backed inspection of an illegal rent — and find large behavioral responses. Comparing the two suggests three features that determine whether such tools bind: whether the tool attaches sanctions or merely information; whether the targeted transfer is legal-but-disclosed or illegal; and whether the rent-bearing activity is separable from the physician's other tasks, so that shutting it down redirects effort rather than suppressing supply. The comparison thus speaks less to whether anti-corruption tools work than to when they work, and points to the conditions under which enforcement rather than transparency is the binding instrument.

Second, we identify a cost of corruption that the literature on physician financial incentives leaves largely implicit. That literature documents distortions in the quantity of care: over-providing high-reimbursement procedures when income falls (Gruber & Owings, 1996; Yip, 1998; Wang, 2023), and expanding volume when payment rates rise (Grant, 2009; Clemens & Gottlieb, 2014), so that removing a rent should lower volume. We find the opposite sign on a different margin: eliminating prescribing-based gray income increases prenatal hospitalization, with no change in

per-admission revenue. Corruption's cost is thus not only the deadweight loss of excess drugs but the opportunity cost of physician effort crowded out of appropriate services that generate no rents. This differs from standard supplier-induced demand (Ellis & McGuire, 1986; McGuire, 2000; Eliason et al., 2018) in operating through an extralegal rent channel rather than reimbursement rates, and in being corrected by enforcement rather than payment reform.

Third, we trace this reallocation through to maternal and infant health outcomes. A central challenge in health economics is distinguishing whether provider-driven increases in volume reflect wasteful over-treatment or the fulfillment of previously unmet clinical need (Currie et al., 2014; Chandra & Staiger, 2020). Work on prenatal care focuses almost entirely on outpatient visit timing and intensity (Currie & Gruber, 1996; Corman et al., 2019; Cygan-Rehm & Karbownik, 2022; Kose et al., 2024), leaving inpatient monitoring during pregnancy largely unexamined as a distinct margin. The staggered rollout generates plausibly exogenous variation in prenatal hospitalization, and we find marked health gains that are not explained by hospital cherry-picking (Einav et al., 2018; Gupta, 2021; Savva et al., 2023; Shi, 2024). We recover the marginal value of this particular reallocation rather than the average value of prenatal hospitalization — but the sign of a compositional shift is a quantity this literature typically assumes rather than measures.

These results also bear on corruption control in public service delivery. Enhanced monitoring reduces leakage (Olken, 2007), but strict enforcement risks defensive behavior or bureaucratic retreat (Gerardino et al., 2024; Shi, 2024; Chen & Dillender, 2025). We find no such chilling response: curtailing gray income redirects effort toward formal, measurable services. Whether that helps patients depends on what the next-best legitimate activity is — in obstetrics it is pro-health, a scope condition we return to in the conclusion.

Section 2 describes the institutional background and the inspection campaign. Section 3 presents the data and identification strategy. Section 4 reports the baseline estimates, and Section 5 examines the mechanisms behind them. Section 6 presents the cost-benefit analysis, and Section 7 concludes.

2 Institutional Background

2.1 Informal Payments, Hospital Finance, and Physician Incentives

Since the 1980s, China's healthcare system has evolved alongside the country's broader shift to a market economy. As the government pared back direct fiscal subsidies, public hospitals had to become financially self-sustaining. To keep care affordable despite this retrenchment, policymakers held down the prices of routine medical services while letting hospitals earn a 15% markup on the drugs they sold. This pricing asymmetry created a classic principal-agent problem: shielded by information asymmetry, physicians had strong incentives to induce demand through excessive

prescribing and redundant diagnostics. The 2009 healthcare reform, chiefly the Zero-Markup Drug Policy (ZMDP), abolished the markup and weakened this "drug-financed" hospital model (Yip et al., 2010; Wu, 2019; Fang et al., 2021). But the structural imbalance persisted: base salaries stayed low relative to physicians' training and workload, so the marginal value of extra income remained high (Xu & Yuan, 2022).

With the legal markup gone, pharmaceutical rents reorganized rather than disappeared. Manufacturers replaced institution-level profit-sharing with direct, illicit ties to individual physicians, paying volume-based kickbacks to reward prescribing (Yip et al., 2010), while patients facing information frictions and quality uncertainty continued to offer informal "red packets" (*hongbao*) to secure priority access or better care. Shutting down one distortionary but legitimate revenue source thus did not realign behavior as long as the deeper constraint, low formal pay, went unaddressed: rents migrated from formal hospital revenue to physician-level gray income. It was against this backdrop of entrenched gray income that the state launched its sweeping anti-corruption campaign.

2.2 Anti-corruption Enforcement and the Hospital Inspection Program

In November 2012, the 18th National Congress of the Chinese Communist Party (CPC) identified corruption as a major threat to regime legitimacy, leading to the December 2012 launch of the "Eight-Point Regulation" (*baxiang guiding*) and a nationwide anti-corruption campaign. Led by the Central Commission for Discipline Inspection (CCDI), the campaign adopted a sustained, top-down enforcement strategy involving inspections, audits, and expanded investigative powers, targeting rent-seeking across government agencies, state-owned enterprises, and public service sectors. Unlike previous episodic crackdowns, this campaign reshaped bureaucratic incentives: hundreds of thousands of officials have been disciplined since 2012 (Wang, 2020), and bribery-related court judgments increased eightfold between 2011 and 2016.[2]

Healthcare became a prominent target of this post-2012 anti-corruption regime. Among prosecuted corruption cases, the share originating from the healthcare sector increased from 2% in 2011 to 10% in 2018, reflecting intensified scrutiny of informal income networks connecting hospitals, physicians, and pharmaceutical firms. By criminalizing these once-customary payments, the state moved from de facto tolerance to strict prohibition: routine inspections came to carry the threat of career termination, replacing a dependable stream of supplementary income with acute regulatory risk. The campaign reached into hospitals through a concrete administrative instrument: the national hospital inspection program.

[2]The number of bribery-related court judgments in China from 2011 to 2016 is sourced from China Judgements Online.

In 2005, the Chinese government launched the "Large Hospital Inspection Work Plan"[3] (hereafter, Hospital Inspection), establishing the first nationwide systematic oversight of public hospitals; the program has since undergone five rounds. Early rounds prioritized basic administrative compliance and imposed weak sanctions. The third round, launched in 2015, marked a sharp discontinuity. Institutionally, it introduced dual-level oversight (the National Health Commission directly inspected a subset of hospitals while provincial commissions managed the rest) and institutionalized the inspections as a permanent regulatory mechanism.[4] Substantively, aligning with the post-2012 anti-corruption campaign, it explicitly targeted the gray income ecosystem, including pharmaceutical kickbacks and *hongbao*, and enforcement became punitive: severe violations triggered public disclosure, administrative removal, license revocation, or criminal prosecution. Each inspection, led by senior officials, spanned 5 to 10 days and followed a standardized three-stage protocol of hospital self-assessment, on-site evaluation by multidisciplinary teams, and public release of results with verified rectification. Appendix B details the program's history, sanction regime, and protocol.

Our empirical analysis focuses on the 2015 round's anti-corruption mandate within a central Chinese province, where 44 large public hospitals experienced a staggered inspection rollout: 15 in November 2015, 15 in September 2016, and 14 in September 2017. Occurring amid the broader anti-corruption campaign, and with inspection outcomes explicitly linked to hospital resource allocation and leadership evaluations, the rollout acts as a distinct regulatory shock: it raised the expected costs of rent-seeking and the marginal returns to compliance with formal standards of care for administrators and physicians.

2.3 Revenue Channels in Obstetric Care and the Pro-Health Substitute

Obstetric care generates provider revenue through two broad channels. The first is pharmaceuticals: obstetrics and gynecology (OBGYN) admissions are drug-intensive, so drug sales long supplied a large share of departmental revenue and, through markup and kickbacks, the gray income described above (Kakolwa et al., 2021; Lian et al., 2022). The second is the regulated fee schedule, which prices delivery and surgical procedures, diagnostics, and prenatal (antenatal) admissions. Because service prices were held well below cost while drugs carried a markup, the marginal yuan of provider income came disproportionately from drugs (Yip et al., 2010). When enforcement raises the cost of the drug-and-kickback channel, the next-best legitimate way to defend revenue is to lean on the fee schedule: billing more services and admitting more women for prenatal care. This

[3]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.

[4]China's health administration was restructured when the Ministry of Health was reorganized into the National Health and Family Planning Commission in 2013 and later into the National Health Commission in 2018. For simplicity, we use "NHC" to refer to all three throughout the paper.

substitution margin is what makes obstetrics distinctive. Extra drug volume beyond clinical need is at best wasteful and at worst harmful, but extra guideline-concordant prenatal care is itself health-improving: prenatal admissions screen for and manage the conditions that drive severe maternal and neonatal morbidity, including hypertensive disorders, gestational diabetes, and anemia (Corman et al., 2018; Sandall et al., 2018; Leonard et al., 2020; Zhang et al., 2022). The legitimate next-best channel in obstetrics is therefore pro-health by institutional design: the same fee-schedule incentive that lets physicians recover lost drug income also rewards care that improves patient health.

Obstetrics is also an unusually clean setting in which to observe this reallocation. Demand is close to exogenous: the timing of pregnancy and delivery is independent of a hospital's inspection date, so treated and control hospitals face comparable patient inflows. Outcomes are realized quickly, within the months between a prenatal visit and delivery, so the health consequences of a change in practice appear within our short panel. And because baseline drug and antibiotic use is high, the drug channel has ample room to contract while the service channel expands (Currie et al., 2014). Together, these features let us trace, in near real time, how an enforcement shock to illicit drug income reshapes physician behavior and, through it, maternal and infant health.

3 Data and Design

3.1 Data Source

Our analysis uses a database built by linking two administrative datasets from the Health Commission of a central Chinese province. The merged sample covers the universe of public secondary and tertiary hospitals from 2014 to 2017.[5]

The first dataset, the Health Statistics Yearbooks, is a hospital-level panel capturing institutional operations and capacity. It records total revenue and its composition, disaggregated by care setting (outpatient versus inpatient) and source (pharmaceuticals, consumables, medical services), alongside capacity metrics such as bed counts, staffing levels, and IT infrastructure. We use these variables as baseline controls for hospital-level heterogeneity.

The primary data source consists of admission-level discharge records from the OBGYN departments of 164 hospitals. These administrative microdata capture all provincial OBGYN inpatient admissions during the 2015–2017 period, yielding approximately 930,000 observations. Each record contains a hospital identifier, the surname and professional title of the attending physician and of the supervising physician, admission and discharge dates, patient demographics, and itemized revenue breakdowns for pharmaceuticals, diagnostics, and consumables. The records carry no

[5]China classifies hospitals into three tiers. We restrict our sample to Tier 2 (secondary) and Tier 3 (tertiary) hospitals; Tier 1 (primary) hospitals function mainly as community clinics with limited inpatient capacity and rarely perform the complex procedures relevant to our study.

unique physician identifier. Providers can therefore be tracked only through the surname–title pair, which constrains the physician-level analysis in Section 5; we describe the resulting measurement error, and bound its consequences, where each measure is introduced. We link these discharge records to the hospital-level yearbook panel using hospital identification codes, matching 99.7% of admissions.

These discharge records provide granular clinical information. Each observation includes up to nine diagnoses (coded via ICD-10) and seven procedures (coded via ICD-9-CM3), tracking the patient's diagnostic trajectory from initial assessment to final discharge. For delivery admissions, the data further report birth weight and specific newborn diagnoses. These clinical variables let us categorize admission types and construct the maternal and infant health outcomes detailed in the subsequent section.

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 outcome metrics and patient health outcomes.

Hospital outcomes. To quantify medical resource utilization, we aggregate administrative discharge records to the hospital-month level. Our primary outcome is total monthly inpatient revenue within the OBGYN department. To understand what drives aggregate revenue dynamics, we decompose the provider supply response along two margins: the extensive margin (total monthly admission volume) and the intensive margin (average revenue per admission). Given the clinical heterogeneity inherent to OBGYN services, we further disaggregate these measures by clinical function, partitioning admissions into pregnancy-related and non-pregnancy-related cases. Pregnancy-related admissions are further partitioned into four mutually exclusive categories: pregnancy loss, prenatal care, delivery, and postnatal care admissions.[6] This disaggregation enables us to assess whether supply responses are uniform across service types or concentrated in areas characterized by greater physician discretion.

Maternal and Infant Health Outcomes. To evaluate whether changes in service provision translate into welfare effects, we construct a vector of health indicators using admission-level

[6]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') or a specific subset of Chapter XXI (codes Z32–Z39); 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 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.

delivery records. For infant 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.[7] 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 whether births cluster in a healthy band around the population norm. For maternal health, we identify adverse maternal outcomes based on discharge diagnoses as well. Key indicators include obstetric trauma, measured by genital tract laceration (Sentilhes et al., 2019); retained placenta; hypertensive disorders of pregnancy and eclampsia; severe maternal morbidity such as acute respiratory distress syndrome (ARDS) (Centers for Disease Control and Prevention, 2024); and surgical/anesthesia complications.[8]

3.3 Sample Construction and Descriptive Statistics

To align the sample with the institutional context of the campaign and ensure measurement reliability, we impose four restrictions; Table A1 reports the sample size and characteristics after each. First, we exclude small hospitals (fewer than 1,000 OBGYN discharges over 2015–2017 or fewer than 100 total beds), which removes idiosyncratically volatile low-volume facilities and approximately 2% of observations. Second, we drop hospitals that failed to report birth weight for more than half of the study months. Third, we remove hospital-month observations with fewer than 10 admissions to prevent small-denominator outlier bias. Fourth, we exclude the hospitals inspected in the third round (September 2017), which lack a sufficient post-treatment window.[9] Finally, because the campaign may also deter non-targeted hospitals in the same administrative jurisdiction, a naive

[7] 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 for infants identified by Vance et al. (2023).

[8] With the exception of *surgical and anesthesia complications*, which are directly observed in the raw administrative data, we construct all other adverse maternal outcome indicators using ICD-10 diagnosis codes recorded during the delivery admission. Specifically, we define binary indicators for: (i) *genital tract laceration*, equal to one if the delivery records an obstetric laceration of the perineum (second degree or above), the cervix, or the upper vagina (ICD-10 codes O70.1–O70.3, O71.3, and O71.4); (ii) *retained placenta*, equal to one if the delivery records retained placenta or retained membranes (ICD-10 code O73); (iii) *hypertensive disorders*, encompassing gestational hypertension and eclampsia (ICD-10 codes O14–O16 and related subcategories); and (iv) *acute respiratory distress syndrome* (ICD-10 codes J80, J95.1, J95.2, J95.3, J95.82x, J96.0x, J96.2x, J96.9x, R06.03, R09.2), which is defined in accordance with CDC guidelines. Grouping the perineal, cervical, and upper-vaginal sites into a single genital tract laceration measure is preferable to tracking each site separately: the sites are clinically adjacent, individually rare, and most importantly subject to substitution in how a given tear is coded, so the combined measure is far less sensitive to site-level coding conventions than its components. For brevity, later tables and figures label this measure simply *laceration*.

[9] This round comprised 14 hospitals province-wide, of which our data cover seven. The gap is largely mechanical rather than selective: the inspection roster was ordered with general hospitals first and specialty hospitals later, and six of the 14 third-round hospitals are eye, cardiovascular, occupational disease, infectious disease, and tuberculosis facilities that provide essentially no obstetric care.

control group would understate the true effect. We therefore split the non-inspected hospitals into a spillover control group (those in districts with at least one inspected hospital) and a clean control group (those in districts with none), retain the inspected hospitals and the clean controls in the main analytical sample, and set the spillover group aside for robustness checks. The final sample spans January 2015 to December 2017 and contains 585,639 admission records from 87 hospitals: 25 inspected hospitals (326,426 admissions) and 62 clean control hospitals (259,213 admissions).

Table A2 reports pre-inspection summary statistics for the treatment, spillover control, and clean control groups, together with the difference in means between the treatment and clean control groups. Before November 2015, the average treated hospital admits 293 OBGYN patients per month, 69% of them pregnancy-related, with mean inpatient revenue of 6,162 RMB per admission (22.6% from pharmaceuticals); the average C-section rate is around 44%, mean birth weight is 3,261 grams, and the LBW and VLBW rates are 7.15% and 1.39%. Inspected hospitals are larger than their non-inspected counterparts and exhibit higher C-section and (V)LBW rates, consistent with the policy's explicit targeting of major public hospitals, which handle a disproportionate share of complex pregnancies (Yin et al., 2023).[10] These baseline level differences are expected given the policy's targeting and do not by themselves threaten identification, which relies on parallel pre-inspection *trends* rather than level balance, an assumption we probe with the raw trends below and formalize in the event studies.

Figure A1 plots unconditional hospital-month trends in aggregate medical revenue by admission type for inspected and non-inspected hospitals.[11] Revenue in the two groups moves in parallel before inspection ($k < 0$) and diverges upward for the treated group afterward ($k > 0$), with the increase driven mainly by pregnancy-related admissions (Panel C), particularly prenatal care (Panel D), and no discernible change in non-pregnancy admissions (Panel B). Although these raw patterns do not adjust for patient case mix or unobserved time-varying shocks, they offer motivating evidence for the formal analysis.

3.4 Identification Strategy

To identify the causal effects of anti-corruption inspections, we employ a staggered DiD design that compares treated hospitals against the control group before and after inspection; after dropping the third-round hospitals, the identifying variation comes from the November 2015 and September

[10] Recorded maternal mortality runs the other way: no maternal death is recorded in any treated or spillover hospital in the pre-inspection period, against 43.89 per 100,000 admissions in the clean control group. A literal zero across the largest facilities in the province is more readily explained by where a death is recorded (a transfer to intensive care or to another department may sever the link to the original OBGYN admission) than by outcomes, so we do not use the mortality series to characterize baseline quality.

[11] The horizontal axis denotes relative time in months centered on the inspection date ($k = 0$). To ensure comparability, we randomly assign placebo inspection dates to the control group, drawn from the temporal distribution of actual inspection timings.

2016 cohorts. The design requires inspection timing to be exogenous to unobserved determinants of hospital outcomes, and two institutional features support this. First, wave assignment was not sequential: the provincial implementation plan, issued in April 2015, named all 44 hospitals and fixed each hospital's inspection year in a single document (before the first on-site inspection, within the first months of our sample window), so assignment cannot be a response to hospital performance observed during the study period.[12] Second, the ordering follows administrative rather than clinical criteria: each wave spans the full range of administrative levels and geographic regions of the province, and the only systematic gradient across waves is a rising share of specialty hospitals; Table A3 reports formal balance tests across the two estimation cohorts. One residual concern is anticipation, since the roster was public for all but the first three months of our sample. Appendix C exploits the three pre-publication months to test for an announcement response and finds none: across thirteen outcomes, only two indicate a level response at publication (with the opposite sign to its post-inspection effect), and the variables carrying our mechanism, prenatal revenue and admissions, exhibit neither a level response nor a pre-inspection drift. To the extent that some anticipatory response nonetheless occurred, it would attenuate our estimates, which are therefore best read as a lower bound on the total effect of the program.

To estimate the impact of inspections on institutional responses, we specify the following two-way fixed effects (TWFE) model at the hospital-month level:

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

where Y_{ht} represents the outcome for hospital h in month t, including total OBGYN admissions, aggregate inpatient revenue, and average revenue per admission. The treatment variable D_{ht} is an indicator equal to 1 if hospital h has been inspected by month t, and 0 otherwise. The coefficient β_1 measures the average effect of inspection on treated hospitals. Because treatment is staggered, this two-way fixed-effects estimand is a weighted average of cohort-specific effects; we therefore complement it below with heterogeneity-robust estimators. We include hospital fixed effects, μ_h, to absorb time-invariant institutional characteristics (e.g., administrative rank and baseline capacity), and year-month fixed effects, γ_t, to account for province-wide temporal shocks. Standard errors are clustered at the hospital level to adjust for serial correlation within hospitals.

To evaluate individual-level infant and maternal health outcomes, we estimate a corresponding specification as follows:

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

[12]Document details are withheld to preserve the anonymity of the study province; the plan is a publicly issued administrative document and is available from the authors on request.

where Y_{iht} denotes the health outcome for patient i admitted to hospital h in month t. Maternal outcomes include indicators for delivery trauma and severe obstetric complications, while infant outcomes capture birth weight and NICU admission. The treatment indicator D_{ht} varies at the hospital-month level. Fixed effects and standard error clustering remain identical to the hospital-level model.

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-month level specification, the event-study specification can be written as:

$$Y_{ht} = \sum_{k=-6,\, k\neq -1}^{10} \beta_k \mathbb{1}\{t - T_h = k\} + \mu_h + \gamma_t + \varepsilon_{ht} \tag{3}$$

where T_h denotes the month in which hospital h is inspected, and $\mathbb{1}\{t - T_h = k\}$ is an indicator equal to one if month t is k periods away from the inspection event. The period prior to inspection ($k = -1$) is omitted and serves as the reference category. The coefficients β_k trace the evolution of outcomes in treated 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 indistinguishable from zero. The individual-level event study follows *mutatis mutandis*: we replace Y_{ht} with the patient-level outcome Y_{iht} and retain the same event-time indicators, fixed effects, and clustering.

Recent econometric advances demonstrate that under staggered treatment timing, canonical TWFE models can yield biased estimates if treatment effects are heterogeneous across cohorts or over time (De Chaisemartin & d'Haultfoeuille, 2020; Goodman-Bacon, 2021). To overcome this negative weighting problem, we complement our canonical event study with the group-time average treatment effect estimator developed by Callaway and Sant'Anna (2021) and the heterogeneity-robust estimator proposed by Sun and Abraham (2021). By estimating cohort-specific dynamic effects before aggregating them, this approach ensures our dynamic estimates remain unbiased by underlying heterogeneous treatment dynamics.

4 Main Results

We present baseline estimates of the inspection's impact on hospital-level resource use and individual clinical outcomes. Our preferred specifications use the clean control group to prevent bias from treatment spillovers. We estimate the policy's effect on aggregate inpatient revenue, decompose it into extensive- and intensive-margin responses, and then assess the welfare implications by examining maternal and infant health outcomes.

4.1 Hospital-Level Resource Reallocation

Table 1 Panel A reports the baseline estimates for aggregate monthly inpatient revenue. Following an inspection, OBGYN revenue in treated hospitals rose significantly by 401,940 RMB, an 18% increase relative to the pre-treatment mean (column 1). Pregnancy-related admissions drive most of the aggregate increase (columns 2–3): delivery care accounts for the largest absolute gain (240,615 RMB), but the largest proportional increases are in prenatal and postnatal care, where revenue rose by 79% in each, while revenue from pregnancy-loss admissions is unchanged (columns 4–7). Panel B shows that this growth operates through the extensive margin: total OBGYN admissions rose by 51 (15%), with the expansion concentrated in discretionary services (prenatal and postnatal admissions increased by 77% and 111%) and no change in pregnancy-loss admissions. Because care for pregnancy loss is dictated by acute biological necessity rather than provider discretion, this null serves as a natural placebo test: the expansions in prenatal and postnatal care are supply-driven rather than reflections of shifting medical need.

The intensive margin, by contrast, is essentially flat. Panel C shows no significant change in average revenue per admission for the full sample or for pregnancy-loss, prenatal, and delivery admissions; the exception is postnatal care, where revenue per admission rose by 680 RMB (16%). Average length of stay (Panel D) fell modestly, by 0.26 days (about 3%), with the reduction concentrated among pregnancy-related and delivery admissions. Post-inspection revenue growth thus came almost entirely from admitting more patients, not from billing more per admission or keeping each patient longer.

Figure 1 plots event-study estimates (Equation (3)) and 95% confidence intervals for aggregate revenue and admission volume, for the full OBGYN department (Panels A–B) and prenatal care (Panels C–D), over a window from six months before to ten months after inspection, with $k = -1$ as the reference period.[13] Pre-treatment coefficients are indistinguishable from zero in every panel, supporting the absence of differential pre-trends, and the canonical TWFE estimates align closely with the heterogeneity-robust estimators of Sun and Abraham (2021) and Callaway and Sant'Anna (2021). Table A4 reports the aggregate effects for every main outcome under all three estimators: they agree in sign throughout, with the heterogeneity-robust aggregates 16 to 44% smaller depending on the outcome, a difference that reflects which pre-inspection baseline each estimator uses rather than negative weighting.[14] The post-inspection increase is persistent rather than transitory and emerges with a three-to-four-month lag that maps directly onto the institutional

[13] While the sample begins in January 2015, reporting from January to April 2015 is incomplete for a number of *control* hospitals, restricting the reliable common pre-treatment window to six months; the treated cohorts are essentially complete over the full window, so the baseline exposure measures in Section 5 are unaffected.

[14] Following Goodman-Bacon (2021), 88% of the estimation weight falls on comparisons between inspected and never-inspected hospitals and 12% on comparisons between the two inspected cohorts; for every hospital-level outcome, the latter comparisons pull the estimate toward zero, so contamination attenuates rather than inflates our estimates.

timeline: after the on-site audit, hospitals submit and execute a formal rectification plan over roughly one quarter, so substantive operational adjustments coincide with the mandatory follow-up re-inspections in the third or fourth month.

Two concurrent policies could confound these estimates: the Zero-Markup Drug Policy (ZMDP), whose second rollout wave overlapped our study period, and the 2016 Two-Child Policy (TCP). Appendix **??** addresses both in detail. In brief, controlling for ZMDP exposure leaves the responses intact (Table A5); the pre-TCP November 2015 cohort reproduces the event-study pattern (Figure A3); gestational timing rules out a TCP explanation, since post-policy conceptions could not generate the second- and third-trimester prenatal admissions that drive the increase; and the response is, if anything, larger among women under 29, predominantly first-time mothers (Figure A5, Panel A).

These dynamics align closely with the income compensation hypothesis. By curtailing illicit gray income without adjusting formal wages, the inspections imposed an acute negative income shock on physicians, who responded by expanding the volume of legitimate, high-discretion services, an extensive-margin response that distinguishes this episode from the intensive-margin adjustments that follow formal fee schedule reforms (Fang et al., 2021). What held constant was the *level* of revenue per admission, not its composition: as Section 5 shows, physicians simultaneously substituted away from drugs and toward billable services within the admission. Whether this expansion reflects welfare-reducing physician-induced demand or the fulfillment of latent, unmet medical needs is a normative question we resolve next with maternal and infant health outcomes.

4.2 Individual-Level Health Outcomes

To determine whether the extensive-margin expansion in obstetric services yields actual health benefits, we estimate the admission-level specification (Equation (2)) using delivery-admission microdata. Table 2 Panel A reports the effects on infant health. After inspection, NICU admissions in treated hospitals decline by 6.89 per 1,000 births, a 13% reduction relative to the pre-treatment mean.[15] The incidence of LBW and VLBW drops by 7.45 (9%) and 2.15 (13%) per 1,000 births, average birth weight rises significantly by 17.96 grams, and the proportion of newborns within one standard deviation of the population mean increases by 13.82 per 1,000 births (following Best et al., 2017), a distributional shift confirming that the aggregate weight gain reflects convergence toward a healthy median rather than an elevated risk of macrosomia.

Panel B evaluates maternal health using delivery diagnoses. Genital tract lacerations fall by 6.71 cases per 1,000 deliveries, a 10.5% reduction relative to a pre-treatment mean of 63.83, and

[15]Considering the potential over-rejection arising from the relatively small number of treated clusters ($N = 25$), we re-estimated the p-values using the Wild Cluster Bootstrap-t procedure (Cameron et al., 2008), implementing Webb weights with 9,999 replications. The wild-bootstrapped p-values, reported in Table 2, remain significant.

hypertensive disorders fall by 8.76 cases (13.6%).[16] Reductions also appear for rarer, more severe events (ARDS, 0.14 cases, and surgical or anesthesia complications, 0.19 cases), though their very low baseline incidence makes the proportional magnitudes imprecise. Together, these estimates indicate a systematic improvement in obstetric safety and quality of care.

Figure 2 plots event-study estimates for NICU admissions, LBW, genital tract lacerations, and hypertensive disorders: pre-treatment coefficients ($k \leq -2$) are indistinguishable from zero, supporting parallel trends, and post-inspection estimates exhibit a sustained downward trajectory consistent with a lasting improvement rather than a transitory disruption of clinical care. To check that administrative burdens did not compromise patient safety, we also estimate Equation (2) on extreme adverse outcomes: maternal and fetal mortality alongside 21 indicators of severe maternal morbidity. Across all of these metrics the point estimates are tightly bounded around zero (Figure A2), with no evidence of operational disruption, defensive medicine, or compromised foundational safety. This reassurance is of a limited kind. These events are rare, twenty-five treated clusters leave little power to detect effects on them, and treated-group maternal mortality is zero throughout the pre-period; the nulls are consistent with safety having been maintained but cannot establish it.

The health results survive the same TCP checks as the hospital-level outcomes (Appendix **??**): the pre-TCP 2015 cohort reproduces the baseline pattern (Figure A4), the infant-health improvement emerges as early as March 2016, too soon for infants conceived under the policy, and consistent gains appear among mothers under 29 (Figure A5, Panel B). Nor do the results hinge on excluding the spillover control group: re-estimating every specification with the spillover hospitals added back as controls leaves all headline effects significant with comparable, slightly attenuated magnitudes (Table A6), the direction expected if the deterrence campaign lightly contaminates nearby hospitals; this more demanding specification, if anything, understates the true effect.

A second check addresses the size gap between the two groups. Because only the province's major hospitals were eligible for inspection, treated hospitals are substantially larger than the clean controls (Table A2), so any province-wide shock whose incidence scales with hospital size would load onto the treatment indicator. We therefore rank all 87 sample hospitals on an index of pre-inspection scale, built from twelve capacity, staffing, volume, and revenue measures recorded in the 2014 yearbook, and retain the top two quartiles; all 25 inspected hospitals fall in that range, so the restriction operates entirely on the control group, which drops from 62 hospitals to 18. The outcomes carrying our mechanism, if anything, strengthen: prenatal revenue and admissions rise to 85% and 86% of their pre-treatment means, against 79% and 77% in the baseline, and birth weight

[16] Table 2 also reports a large coefficient on retained placenta (−8.20 cases, 36% of its pre-treatment mean), which we do not interpret as a treatment effect: the outcome was already declining before the inspection, and estimators that difference against the month immediately preceding inspection place the effect at approximately zero (Table A4). We retain the variable in the table for completeness but exclude it from the maternal health results.

gains 18.73 grams. The aggregate magnitudes shrink, and NICU admission and hypertensive disorders lose conventional significance on a sample half the size (Table A7; Appendix D). The restriction closes the size gap but not the tier boundary, however: eligibility was confined to tertiary hospitals, and just one of the 18 remaining controls is itself tertiary.

The inspections improved maternal and infant health without crowding out care for severe, life-threatening conditions. This resolves the normative ambiguity: rather than reflecting wasteful physician-induced demand, the expansion of discretionary services, particularly prenatal care, appears to satisfy latent unmet medical needs and may generate net welfare gains, which we examine in Section 6.

5 Mechanisms

Although the preceding analysis demonstrates that inspections improved clinical outcomes, two critical issues remain. First, an incentive compatibility puzzle arises: if expanding prenatal care is both revenue-generating and health-enhancing, why was it underprovided prior to the regulatory shock? Second, the mechanisms driving these health gains are unidentified. Improved average outcomes do not necessarily reflect genuine enhancement in care quality. Instead, they may stem from strategic patient selection (i.e., cherry-picking) or defensive medicine practices, as physicians seek to minimize risk and reduce their workload. By systematically turning away high-risk patients, hospitals could artificially inflate aggregate health metrics. Such compositional shifts would represent a distortion of medical access rather than a net welfare gain.

To distinguish between these competing mechanisms, we examine the channels underlying the observed health improvements.

5.1 Providers' Behavioral Responses

If the income compensation hypothesis holds (physicians responding to lost gray income by expanding legitimate services), we should observe two revenue-generating behavioral responses: substitution away from drug prescriptions toward medical services at the intensive margin, and expanded prenatal care for patients with obstetric comorbidities. A third margin bears on quality rather than revenue: whether C-section rates fall once case mix is held fixed. We examine each in turn. Although the baseline estimates show no change in per-admission revenue, that aggregate null may obscure compositional shifts, since providers retain incentives to replace targeted drugs with legitimate, less-scrutinized services. We therefore decompose average revenue per admission into drug, service, diagnostic, and consumable components[17] and estimate Equation (1) for each

[17]Services include nursing, surgical, and anesthesia fees; diagnostics consist of pathology, laboratory, and imaging fees; and consumables include single-use medical supplies for examinations, treatments, and surgeries.

component separately.

Table 3 Panel A shows that inspections significantly reduced average drug revenue per admission by 226.63 RMB (16.6% of the pre-treatment mean) while increasing service revenue by 228.91 RMB (10.1%); diagnostic and consumable revenue remain unchanged. This nearly one-to-one financial offset explains the stable intensive margin in our baseline findings and is precisely the substitution the income compensation hypothesis predicts as regulatory scrutiny raised the expected cost of pharmaceutical kickbacks. Panels B and C confirm a genuine reduction in drug utilization rather than mere repricing, evidenced by the increased prevalence of admissions with zero or very low drug revenue. The substitution also holds at the admission level (drug share -2.8 percentage points, service share +2.4) and survives physician fixed effects; within their own caseloads, physicians cut their drug share by 2.4 points and raise their service share by 2.2, so it reflects changes in how physicians practice rather than reallocation of patients across providers (Table A8). It is also hospital-wide: the service share rises across most departments while the drug share declines most in obstetrics and internal medicine, consistent with an institution-wide shock to drug-based rent extraction (Figure A6).

Turning to the second response, Figure 3 shows where the rise in prenatal admissions is concentrated: almost entirely among medium-risk pregnancies with obstetric comorbidities. Prenatal admissions rise by 149% for women with a prior C-section, 121% for anemia, and 110% for gestational diabetes (GDM) relative to their pre-treatment means; conditions whose prenatal management is dictated by the diagnosis itself (placenta previa, placenta accreta, maternal cardiac disease, chronic renal disease) show little or no change. Admissions for mild and severe preeclampsia also rise (89% and 38%); since these are precisely the conditions inpatient antenatal monitoring is meant to detect and manage, we read them as part of the same expansion, anticipating the reduction in hypertensive disorders in Table 2. This gradient is what income compensation predicts: prenatal care for medium-risk pregnancies is relatively discretionary and appears to have been underprovided before the inspection, while high-risk pregnancies require intensive management regardless of the regulatory environment.

Finally, we ask whether the inspections curbed the overuse of C-sections, a discretionary and revenue-relevant procedure. In the full delivery sample the raw effect is small and insignificant (-0.014), but adjusting for case mix (prior C-section, advanced maternal age, placenta accreta, and placenta previa) reveals a significant decline of 2.6 percentage points, or 6.4% of the pre-treatment mean (Table 3, Panel D). The effect is concentrated among mothers aged 35 and above (-3.6 percentage points), the group with the highest baseline C-section rate (54%) and the greatest scope for discretionary surgery, consistent with physicians paring back lower-indication procedures rather than withholding medically necessary ones. One caveat: the case-mix controls include characteristics whose composition itself responds to the inspection, so the adjusted coefficient

describes behavior within case mix rather than a fully clean causal contrast.

Facing scrutiny of pharmaceutical rents, physicians thus substituted from drugs toward legitimate services, expanded discretionary prenatal care for women with obstetric comorbidities, and pared back low-indication C-sections. These average responses are consistent with income compensation but do not establish it: if lost drug rents drive the reorientation, the response should be strongest where the rents at stake were largest. That is the dose-response prediction we test next.

5.2 Heterogeneity by Hospitals' Exposure to the Inspection

A sharper test of income compensation exploits variation in exposure. Providers differ widely in how intensively they prescribed drugs before the inspection,[18] and hence in how much gray income the campaign placed at risk. If the redirection of effort is driven by the loss of pharmaceutical rents, the providers with the most rent to replace should exhibit the largest substitution toward services and, if the redirected effort is genuinely pro-health, the largest downstream health gains. We test this dose-response logic along three dimensions: baseline drug-prescribing intensity (the drug channel), baseline C-section rates (the delivery channel), and physician seniority (a cross-cutting proxy for the capacity to capture rents in the first place).

5.2.1 Drug Channel: Baseline Prescribing Intensity

Using only pre-inspection admissions, we recover physician (hospital) fixed effects from a case-mix-adjusted regression of the drug share of per-admission revenue and rank providers into quartiles of these fixed effects, so that higher quartiles denote higher baseline drug intensity net of the patients treated.[19] Figure 4 confirms that the substitution scales with exposure. The pattern is clearest at the physician level (Panel A), where the individual attending exercises the most discretion: the highest-intensity physicians raise their service share by 3.6 percentage points ($p < 0.05$) and cut their drug share by 3.3 points, whereas the lowest-intensity physicians show no substitution. The hospital-level split (Panel B) is consistent but noisier, since it averages over physicians and leaves few units per quartile. The health gains follow the same gradient (Figure A7): in the top hospital quartile, birth weight rises by about 32 grams and the incidence of low birth weight and hypertensive disorders falls by roughly 18 and 21 cases per 1,000 deliveries, each the largest effect across the

[18] We define the baseline (pre-inspection) period as January to June 2015, the first six months of the sample and well before the earliest inspection round in November 2015. All baseline exposure measures used in this section are computed over this window.

[19] The case-mix controls comprise patient age, insurance status and type (Urban Employee, Urban Resident, or New Rural Cooperative Medical Insurance), an indicator for a pregnancy-related admission, and indicators for obstetric comorbidities that plausibly affect drug use, including gestational diabetes, hypertensive disorders, asthma, renal disease, and cardiovascular disease. Recovering the fixed effect from this regression isolates a provider's systematic prescribing propensity from the composition of its caseload.

four quartiles; the physician-level split echoes this ranking. Because each quartile is identified from only a fraction of the treated clusters, individual estimates are imprecise; the evidence is in the cross-quartile ordering, where the largest effects consistently line up with the highest baseline prescribing intensity. Sorting providers instead by the antibiotic share of drug revenue, a canonical vehicle for revenue-driven over-prescribing on which physicians within a hospital differ far more (Wilcoxon $p < 0.01$; Figure A10), reproduces and sharpens the gradient, with the physician-level service-share response rising monotonically to 4.9 percentage points in the top quartile (Figures A8 and A9).

5.2.2 Delivery Channel: Baseline C-section Rates

Because gray income is hard to observe directly, we proxy a provider's baseline reliance on financial incentives with its pre-inspection C-section rate, a well-established marker of physician-induced demand under asymmetric information (Gruber & Owings, 1996; Johnson & Rehavi, 2016), and stratify hospitals at a 40% baseline rate.[20] Consistent with the hypothesis, the prenatal-care expansion is concentrated in high-C-section hospitals (Table 4, Panel A): aggregate prenatal revenue rises by 76,182 RMB (86% of the pre-treatment mean) and prenatal admissions by 20.0 per hospital-month (80%), with the revenue gain coming entirely through service revenue rather than drugs; in low-C-section hospitals (Panel B) the effects are uniformly smaller. The same gradient holds when we stratify by the attending physician's own baseline C-section rate (Panels C–D), and it appears at a finer grain across comorbidities: the rise in prenatal admissions for anemia and prior C-section is far larger among high-C-section providers, while structurally high-risk conditions change little in either group; GDM is the exception, with the two groups close (Figure A11). Health gains largely track this behavior (Figure 5). At the hospital level, the improvements concentrate in high-C-section hospitals: birth weight rises by about 26 grams and low birth weight and hypertensive disorders fall by roughly 14% and 24% relative to baseline, against small, insignificant changes in low-C-section hospitals. Lacerations are the exception, falling by a similar 8 to 9% in both groups. The physician-level split is directionally similar for the maternal outcomes but reverses the infant-health ordering (the birth-weight gain is larger among *low*-C-section physicians); given how few deliveries are attributable to any single physician, we rest the infant-health evidence for this channel on the hospital-level split.

[20] 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 revealing substantial variation extending well into the mid-40% range (Li et al., 2017; Zhang et al., 2022).

5.2.3 Cross-Cutting: Physician Seniority

Senior physicians wield greater clinical and administrative authority and are the primary recipients of gray income, so if the shock operates through the loss of such income, its consequences should be most pronounced among their patients. We classify an attending physician as "senior" if their surname matches that of a Chief Physician or Department Head recorded on the same hospitalization, and "junior" otherwise. Because this name-based linkage errs in only one direction (misrecording seniors as juniors), coincidental surname matches can only *attenuate* the senior–junior contrast, and the asymmetry is unchanged when we progressively exclude the four most common Chinese surnames (Appendix E; Table A9). Table 5 bears the prediction out across all three margins. Both groups cut drug revenue and drug share, but only senior physicians significantly redirect the freed-up billing toward services, raising per-admission service revenue by 182 RMB (7.3%) and their service share by roughly 3 percentage points (Panel A); only senior physicians significantly expand prenatal admissions for anemia and prior-C-section patients (Panel B); and the health gains concentrate among their patients, whose deliveries show significant improvements in birth weight (+18.7 g), low birth weight (−9.0 per 1,000), lacerations, and hypertensive disorders, while among junior physicians' patients only lacerations improve significantly (Panel C).

The three heterogeneity cuts tell a consistent story: whether providers are sorted by baseline prescribing intensity, baseline C-section rate, or seniority, both the behavioral reallocation toward legitimate services and the resulting health gains are largest among those with the most illicit income at stake. By tying the health improvements directly to the loss of pharmaceutical rents, this dose-response evidence corroborates income compensation as the force driving the reallocation.

5.3 Channel Attribution: Which Margin Produces the Health Gains?

The inspection moved several margins at once—drug prescribing, prenatal volume, and C-section use—so the health gains cannot be attributed to any one of them by fiat. We do not claim to identify mediation. Instead, we ask whether the gains fall where the prenatal channel alone predicts they should, a joint pattern the competing channels do not reproduce.

The channel carries two testable implications. First, the gains should concentrate among the higher-risk pregnancies whose prenatal admissions expanded most and be muted among structurally low-risk deliveries. We split the delivery sample at advanced maternal age (AMA, $\geq$ 35), a predetermined characteristic that, unlike the anemia or gestational-diabetes diagnoses additional monitoring could itself uncover, treatment cannot manufacture. Table A10 bears the prediction out for infant health: among AMA mothers birth weight rises by 39.8 grams and low birth weight falls by 19.5 per 1,000 (Panel A), roughly two to three times the 14.6-gram and (insignificant) 5.8-per-1,000 changes among younger mothers (Panel B), and largest in precisely the group the prenatal

expansion targeted. Second, the channel should improve outcomes antenatal care can manage yet leave untouched those settled in the delivery room. Read this way the maternal results are diagnostic rather than merely mixed: hypertensive disorders, which inpatient monitoring is meant to detect and control, fall nearly twice as much among AMA mothers (-15.6 versus -7.9 per 1,000), whereas genital tract laceration, an intrapartum event tied to operative delivery and delivery-room technique rather than antenatal management (Friedman et al., 2015), shows no age gradient at all, improving only among younger mothers. The same laceration exception appears in the C-section-rate split of Figure 5. A mechanism running through fewer or better-managed C-sections would predict the opposite ordering for laceration; the antenatal channel predicts exactly the one we observe.

The evidence is a preponderance, gains where the mechanism predicts and nulls where it does not, rather than an identified decomposition.[21] We anchor the channel reading on birth weight, which survives the size-restricted control group and carries almost the entire monetized benefit of the cost-benefit analysis.

5.4 Physician Turnover

The 2015 inspection's stringent penalties for rent-seeking, including license revocation and criminal prosecution, motivate a "selective attrition hypothesis": physicians heavily dependent on informal income may have strategically exited inspected hospitals to evade scrutiny. If these leavers were indeed rent-seeking agents and primary drivers of physician-induced demand, their departure and the reallocation of their patients to less financially incentivized peers could independently drive the observed health improvements.

To test this, we track attending physicians through the surname–title records and classify them as *Exit* (appearing pre-inspection but permanently vanishing), *Entry* (appearing only post-inspection), or *Stay*. Because the records carry no unique physician identifier, the measure tracks surnames rather than individuals: it undercounts turnover where surnames collide and registers promotions or reassignments as exits. We therefore interpret the estimates as evidence of a discrete change in the composition of the attending roster rather than as a count of physicians who left (Appendix E). Panel A of Table 6 reports a significant increase in mobility: monthly exits and entries per hospital

[21] As a quantitative complement to this pattern evidence, we apply the order-invariant decomposition of Gelbach (2016) to the birth-weight and low-birth-weight effects. Prenatal admission volume accounts for roughly 21 percent of the birth-weight gain and 36 percent of the low-birth-weight decline; the drug-share channel, the principal competing margin, accounts for essentially zero (0.4 percent) and 6 percent, respectively. Neither channel meaningfully co-moves with the decline in genital tract laceration, consistent with the intrapartum interpretation in the text and distinct from the pattern observed for antenatally manageable outcomes. The decomposition is an accounting bound, not a mediation analysis: the measured channels are themselves outcomes of the inspection, so the exclusion restriction fails and the remainder reflects this structural limitation together with the noise inherent in 25 treated clusters. The channel shares are imprecisely estimated—the 90-percent confidence interval on the prenatal share for birth weight runs from -13 to 48 percent—as expected at this level of aggregation.

rise by 0.12 and 0.11 within a six-month window around inspection (0.16 and 0.19 within three months), sizable relative to the low baseline turnover rate. The estimates are essentially unchanged when we cumulatively exclude the four most common Chinese surnames, so the result is not an artifact of surname collapse (Table A9).

Was this turnover *selective*? Comparing leavers and stayers within hospitals (Figure A12, with hospital, admission-month, and case-mix fixed effects), departing physicians performed significantly more C-sections at baseline (roughly 0.08 standard deviations, $p < 0.01$) and leaned toward more antibiotic-intensive prescribing, while their overall drug revenue and drug share were indistinguishable from stayers': leavers relied disproportionately on the surgical and antibiotic margins most exposed to anti-corruption scrutiny. Unadjusted correlations point the same way, though individually insignificant given just 25 treated hospitals (Figure A13).

If these departures contribute to the health gains, effects should concentrate in hospitals that experienced exits. And they do. Stratifying on whether a hospital recorded an exit within the six-month window (Table 6, Panels B–C), exit hospitals show a 23.13-gram rise in birth weight, a 10.45 per 1,000 drop in LBW, and 8.5% and 15.5% declines in lacerations and hypertensive disorders, whereas no-exit hospitals show small, insignificant changes throughout.

The evidence is consistent with workforce recomposition as a complementary mechanism alongside the behavioral adaptation of incumbents. Two limitations qualify it: a recorded "no exit" means no *surname* disappeared (a distinction that correlates with department size), and exit status is realized after the inspection, itself an outcome of treatment. The Panel B–C comparison is therefore descriptive: one strand of evidence for a compositional channel, not an identified decomposition of the health gains.

5.5 Ruling out Alternative Channels

Four alternative explanations could still threaten this interpretation: strategic patient selection (cherry-picking healthier patients while diverting complex cases elsewhere), patient socioeconomic status (gains merely tracking, say, a decline in informal "red-packet" payments concentrated among affluent urban patients), hospital information-system capacity, which could mechanically shape how outcomes are recorded, and cross-hospital patient diversion, under which the prenatal expansion reallocates admissions from neighboring hospitals rather than adding care. Appendix F reports additional detail on the cherry-picking, information-system, and patient-diversion checks.

Providers under heightened scrutiny have clear incentives to avoid high-risk patients, both to minimize adverse outcomes during a politically sensitive period and to lighten workload as informal income dries up. Following Leonard et al. (2020), we use ICD-10 codes to identify 27 obstetric comorbidities recorded at admission, characterizing each patient's ex-ante clinical risk

profile,[22] and estimate Equation (2) on the comorbidity composition of admissions. Figure 6 (solid circles) yields no evidence of cherry-picking among delivery admissions: the change in admission probability is insignificant for nearly every severe comorbidity, and the probability of admitting mothers of advanced maternal age or with a prior C-section, both well-documented risk factors, if anything rose, so treated hospitals absorbed a marginally higher-risk population rather than a healthier one. The mirror-image concern, that the prenatal expansion reflects admitting low-risk women for unnecessary stays, also fails: the added prenatal admissions (hollow circles) concentrate among pregnancies carrying a documented comorbidity rather than uncomplicated ones, indicating that the expansion met genuinely unmet clinical needs. Appendix F shows further that modifiable conditions such as anemia rise among prenatal admissions but fall at delivery, the signature of prenatal care that treats rather than merely records risk.

Nor do the gains track socioeconomic status or record-keeping capacity (Figure 7). Splitting patients by rural versus urban hukou,[23] the two groups move in lockstep: the between-group difference is insignificant for every childbirth outcome (p-values from 0.14 to 0.99), and utilization and revenue responses are likewise indistinguishable, so a channel concentrated among higher-SES patients (such as a decline in red-packet payments among the affluent) cannot drive the health gains. Splitting hospitals by baseline information-system quality,[24] the core infant-health effects are indistinguishable across the two groups ($p \geq 0.25$); the codes-based maternal complications (laceration, ARDS) fall by more in better-recorded hospitals, the direction that argues *against* strategic under-coding, which would concentrate apparent improvements where records are weakest.

The last concern is one of accounting. If inspected hospitals drew prenatal patients from their neighbors, the DiD would count the same admissions twice and the expansion would reshuffle care rather than add it, undercutting both the cost-benefit calculation and the reading of the expansion as meeting latent need. Figure A14 tests this at the market level, plotting prenatal admissions in event time for the uninspected hospitals within five kilometers of an inspected one. Their admissions do not fall: a post-inspection indicator, with hospital and calendar-month fixed effects, gives 0.38 (s.e. 2.20) against a pre-inspection mean of roughly 17. The same holds at wider radii, which matter because women travel to reach tertiary hospitals: the coefficient is 7.78 (s.e. 3.69) within ten kilometers and 1.88 (s.e. 4.11) within twenty, positive in every case (Appendix F). The catchment-area total rises, which reshuffling cannot produce. And the magnitudes rule diversion out on their

[22]The identification is based on the 27 Comorbidity Diagnosis Groups outlined in Appendix 2 of Leonard et al. (2020).

[23]We infer a patient's hukou (household-registration) status from the insurance scheme paying for the admission: Urban Resident or Urban Employee Basic Medical Insurance admissions are classified as urban, and New Rural Cooperative Medical Scheme admissions as rural.

[24]We classify hospitals that had established both electronic medical records and a hospital management system by 2014, the year before the inspection, as having a good information system, and all others as having a poor one; the underlying measures come from the Health Statistics Yearbooks.

own: roughly 490 additional prenatal admissions per month across the 25 treated hospitals is about twice the entire baseline prenatal volume of their neighbors. The specification that returns the spillover hospitals to the control group supplies a further sign check: diversion would depress precisely those hospitals and therefore *inflate* the estimated effect, yet the coefficients fall slightly (Table A6). Deterrence spillovers could mask some diversion in the flat nearby series, but the catchment total nets the two forces and rises regardless.

Inspected hospitals thus did not change who they treated, who pays, or how outcomes are recorded, and their prenatal expansion was not drawn from neighboring hospitals: the health improvements reflect genuine changes in the delivery of care.

6 Cost-Benefit Analysis

We monetize the estimated treatment effects in a stylized cost-benefit framework; Appendix G presents the valuation formulas, parameter choices, and full sensitivity analysis. We define the policy cost as the aggregate increase in hospital revenue across the treated hospitals. This is best read as an upper bound on the real resources involved: any portion that replaces income physicians previously captured informally is a transfer rather than a resource cost. Applying our baseline estimate of a 401,940 RMB increase in aggregate monthly revenue per hospital to the 515 post-treatment hospital-months in the treated sample yields a total policy cost of approximately 207 million RMB.[25]

The benefits flow through two channels. For infant health, our estimated 0.215 percentage point decline in the probability of VLBW, applied to the 101,031 post-inspection deliveries in treated hospitals,[26] implies 217 averted cases and short-term medical savings of between 32.58 million and 108.61 million RMB.[27] The estimated 18-gram rise in mean birth weight additionally generates long-term human capital returns (Black et al., 2007; Almond et al., 2018; Bharadwaj et al., 2018): using the estimate that a one-standard-deviation increase in birth weight raises adult annual income by roughly 2.75% (Lambiris et al., 2022), we project the present value of lifetime income gains over a 40-year career and discount the stream back to birth (Appendix G, Equations (4) and (5)). At our benchmark parameters (a discount rate of $r = 3\%$ and real wage growth of $g = 2\%$, the pair implied by the Ramsey rule under log utility and a 1% pure rate of time preference), the per-child gain is 1,918 RMB at the non-private-sector wage, aggregating to a social benefit of 193.73 million

[25]The sample comprises 25 treated hospitals: 14 first-round hospitals with 25 months of post-treatment exposure and 11 second-round hospitals with 15 months (data for five hospitals are unavailable). Administrative inspection costs are excluded; they are absorbed within the existing budgets of the Health Commission's Disciplinary Inspection Team.

[26]Treated hospitals accounted for 163,537 deliveries in the analyzed sample, of which 62,506 occurred before and 101,031 after the inspection.

[27]Drawing on evidence from Chinese NICUs, the medical cost of treating a VLBW infant ranges from 150,000 to 500,000 RMB (Zhu et al., 2020).

RMB; the more conservative private-sector wage implies 1,181 RMB per child and 119.29 million RMB in aggregate.

For maternal health, the estimated 0.67 percentage point reduction in genital tract lacerations translates into 678 averted cases post-inspection, worth 0.64 to 1.29 million RMB in avoided hospitalization and roughly 1.01 million RMB in preserved maternal productivity.[28] This is the least well-identified of the effects we monetize (unlike the infant outcomes, the laceration decline does not track the exposure gradients of Section 5), but it contributes under one percent of the total benefit, so nothing turns on it.

Combining the midpoints of these ranges, total monetized benefits of 229 million RMB against a policy cost of 207 million RMB leave a net social surplus of 22 million RMB, or about 3.26 million USD (roughly the cost of fully funding one year of public health expenditure for 20,185 citizens at 2017 levels).[29] Table 7 probes this figure. Because the valuation depends on r and g only through the ratio $(1+g)/(1+r)$, the sign of the surplus is governed by the *wedge* between them: the surplus is positive whenever the wedge falls below roughly 1.4 percentage points (under the Ramsey rule with log utility, whenever the pure rate of time preference is below 1.4%), and is nearly invariant to the growth rate itself. Two benefit-side parameters (the VLBW treatment cost and the birth-weight–earnings elasticity) can reverse the sign on their own, which is why we treat the magnitude with caution; the cost side only reinforces the result, since recognizing even a modest share of the revenue increase as transferred rent rather than resources consumed raises the surplus. Substituting the heterogeneity-robust estimates of Table A4 brackets the canonical figure (+35.70 million RMB under Sun and Abraham (2021), −14.86 million under Callaway and Sant'Anna (2021)), and our accounting omits averted infant mortality, prevented lifelong disability, and mitigated severe maternal morbidity outright (Baran et al., 2025). The exercise thus supports a qualitative conclusion rather than a point estimate: the enforcement appears to have paid for itself in social terms, with a surplus that is positive at the benchmark, across the discount-rate and growth combinations we regard as defensible, and reinforced by every omitted benefit category.

7 Conclusion

This paper examines the healthcare utilization and welfare consequences of anti-corruption inspections within China's public hospital sector. Exploiting the staggered rollout of the 2015–2016 hospital inspections, we trace a chain running from tighter regulatory oversight of gray income, through the reallocation of provider effort toward legitimate preventive care, to marked improve-

[28] We assume one to two avoided hospitalization days at the sample average of 950 RMB per day and a five-day faster return to work at a 297 RMB average daily wage; obstetric genital tract trauma carries substantial morbidity that can impair work and newborn care (Schmidt & Fenner, 2024).

[29] China's per capita Government Health Expenditure was 1,093.88 RMB in 2017 (Yip et al., 2019). Renminbi are converted to dollars at 6.765 RMB per USD; all figures are nominal RMB (see Appendix G).

ments in maternal and infant health; the two endpoints of this chain are identified off inspection timing, while the intermediate link is inferred from provider behavior rather than measured directly.

The inspections raised aggregate inpatient revenue almost entirely through the extensive margin (admission volume) rather than the intensive margin, with prenatal admissions rising most. This expansion translated into health gains: higher birth weight, less (very) low birth weight, a distributional shift of birth weights toward the clinical optimum, and fewer maternal delivery complications. Mechanism analyses tie these gains to a realignment of provider incentives. As the inspections curtailed pharmaceutical kickbacks, providers substituted away from drug-centric revenue toward legitimate, labor-intensive services, particularly prenatal care; both the behavioral response and, for the infant outcomes and hypertensive disorders, the health gains are largest among providers with the greatest ex-ante reliance on revenue-driven practice: high baseline prescribers, high-C-section providers, and senior physicians. Physician turnover was also elevated, with health gains concentrated in hospitals that experienced post-inspection exits. Our cost-benefit analysis indicates that the health and human-capital gains outweigh the cost of the additional care under our benchmark assumptions, though the sign is not robust to every parameter we vary.

Our central contribution is to show that these health benefits were incidental. The inspections targeted illicit rents, not maternal or infant health; the gains are most plausibly explained by the most attractive legitimate substitute once the drug channel was compromised: a health-improving service: discretionary prenatal care, priced favorably under the fee schedule. This logic carries a clear scope condition: anti-corruption enforcement improves health only where the next-best legitimate source of revenue is itself pro-health. Where the fee schedule instead rewards low-value imaging or unnecessary procedures, the same redirection of rent-seeking effort could leave health unchanged or even worsen it, and our estimates, from a single central Chinese province, an obstetric population, and the 2015–2017 window, should be read against that scope condition.

Two policy implications follow for healthcare governance, especially in developing economies facing structural corruption and scarce healthcare resources. First, our results support the transition from episodic enforcement to institutionalized regulatory inspections: sudden scrutiny neither compromised infant health nor induced strategic risk-selection, but improved hospital performance and generated net social gains. Second, sustainable anti-corruption policy requires formal compensation reform. The rapid expansion of legitimate service supply once gray income was cut off reflects physician responsiveness to financial incentives; the pre-inspection prevalence of corruption was symptomatic of formal wages set too low relative to physicians' training and workload, so enforcement must be paired with decoupling physician income from drug and device sales while raising base salaries and medical service fees to reflect the true value of clinical human capital.

References

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

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.

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.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, *90*(3), 414–427.

Centers for Disease Control and Prevention. (2024). Identifying severe maternal morbidity (smm): Icd codes [https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html (accessed March 8, 2026)].

Chandra, A., & Staiger, D. O. (2020). Identifying sources of inefficiency in healthcare. *The quarterly journal of economics*, *135*(2), 785–843.

Chatterji, P., Li, F., & Markowitz, S. (2026, April). *Exploring trends in severe maternal morbidity in the u.s.* (Working Paper No. 35062). National Bureau of Economic Research. https://doi.org/10.3386/w35062

Chen, Y., & Dillender, M. (2025). *Government monitoring of health care quality: Evidence from the nursing home sector* (tech. rep.). National Bureau of Economic Research.

Chereches, 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.

Corman, H., Dave, D., & Reichman, N. E. (2018). *Effects of prenatal care on birth outcomes: Reconciling a messy literature* (Working Paper No. 24885). National Bureau of Economic Research. https://doi.org/10.3386/w24885

Corman, H., Dave, D., & Reichman, N. E. (2019). The effects of prenatal care on birth outcomes: Reconciling a messy literature. In *Oxford research encyclopedia of economics and finance*.

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.

Cygan-Rehm, K., & Karbownik, K. (2022). The effects of incentivizing early prenatal care on infant health. *Journal of Health Economics*, *83*, 102612.

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.

Einav, L., Finkelstein, A., & Mahoney, N. (2018). Provider incentives and healthcare costs: Evidence from long-term care hospitals. *Econometrica*, *86*(6), 2161–2219.

Eliason, P. J., Grieco, P. L., McDevitt, R. C., & Roberts, J. W. (2018). Strategic patient discharge: The case of long-term care hospitals. *American Economic Review*, *108*(11), 3232–3265.

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

Fang, H., Lei, X., Shi, J., & Yi, X. (2021). *Physician-induced demand: Evidence from china's drug price zero-markup policy* (tech. rep.). National Bureau of Economic Research.

Friedman, A. M., Ananth, C. V., Prendergast, E., D'Alton, M. E., & Wright, J. D. (2015). Evaluation of third-degree and fourth-degree laceration rates as quality indicators. *Obstetrics and Gynecology*, *125*(4), 927–937. https://doi.org/10.1097/AOG.0000000000000720

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.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, *34*(2), 509–543.

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.

Grant, D. (2009). Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of health economics*, *28*(1), 244–250.

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.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

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

Kakolwa, M. A., Woodd, S. L., Aiken, A. M., Manzi, F., Gon, G., Graham, W. J., & Kabanywanyi, A. M. (2021). Overuse of antibiotics in maternity and neonatal wards, a descriptive report from public hospitals in dar es salaam, tanzania. *Antimicrobial Resistance & Infection Control*, *10*(1), 142.

Kose, E., O'Keefe, S., & Rosales-Rueda, M. (2024). Does the delivery of primary health care improve birth outcomes?: Evidence from the rollout of community health centers. *Journal of Human Resources*.

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.

Larkin, I., Ang, D., Steinhart, J., Chao, M., Patterson, M., Sah, S., Wu, T., Schoenbaum, M., Hutchins, D., Brennan, T., et al. (2017). Association between academic medical center pharmaceutical detailing policies and physician prescribing. *Jama*, *317*(17), 1785–1795.

Leder-Luis, J. (2025). Can whistleblowers root out public expenditure fraud? evidence from medicare. *Review of Economics and Statistics*, 1–18.

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.

Li, J., Wu, B., Flory, J., & Jung, J. (2022). Impact of the affordable care act's physician payments sunshine act on branded statin prescribing. *Health Services Research*, *57*(5), 1145–1153.

Lian, Q., Zheng, T., Huo, X., Zhang, J., & Zhang, L. (2022). Prophylactic antibiotic use during labor and delivery in china: A nationwide, multicenter, hospital-based, cross-sectional study. *BMC medicine*, *20*(1), 391.

McGuire, T. G. (2000). Physician agency. *Handbook of health economics*, *1*, 461–536.

Ministry of Human Resources and Social Security of the People's Republic of China. (2017). 2017 statistical bulletin on human resources and social security development [https://www.mohrss.gov.cn/wap/fw/rssj/201805/W020220325394406391270.pdf (accessed March 9, 2026)].

Newham, M., & Valente, M. (2024). The cost of influence: How gifts to physicians shape prescriptions and drug costs. *Journal of Health Economics*, *95*, 102887.

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

Pham-Kanter, G., Alexander, G. C., & Nair, K. (2012). Effect of physician payment disclosure laws on prescribing. *Archives of internal medicine*, *172*(10), 819–821.

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.

Sandall, J., Tribe, R. M., Avery, L., Mola, G., Visser, G. H., Homer, C. S., Gibbons, D., Kelly, N. M., Kennedy, H. P., Kidanto, H., et al. (2018). Short-term and long-term effects of caesarean section on the health of women and children. *The Lancet*, *392*(10155), 1349–1357.

Savva, N., Debo, L., & Shumsky, R. A. (2023). Hospital reimbursement in the presence of cherry picking and upcoding. *Management Science*, *69*(11), 6777–6799.

Schmidt, P. C., & Fenner, D. E. (2024). Repair of episiotomy and obstetrical perineal lacerations (first–fourth). *American Journal of Obstetrics and Gynecology*, *230*(3S), S1005–S1013. https://doi.org/10.1016/j.ajog.2022.07.005

Sentilhes, L., Madar, H., Ducarme, G., Hamel, J.-F., Mattuizzi, A., & Hanf, M. (2019). Outcomes of operative vaginal delivery managed by residents under supervision and attending obstetricians: A prospective cross-sectional study. *American Journal of Obstetrics and Gynecology*, *221*(1), 59–e1.

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

Wang, Y. (2023). Wealth shocks and physician behavior: Evidence from childbirth.

Wang, Y. (2020). China's Corruption Investigations Dataset. https://doi.org/10.7910/DVN/9QZRAD

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

Yin, S., Zhou, Y., Yuan, P., Wei, Y., Chen, L., Guo, X., Li, H., Lu, J., Ge, L., Shi, H., et al. (2023). Hospital variations in caesarean delivery rates: An analysis of national data in china, 2016-2020. *Journal of Global Health*, *13*, 04029.

Yip, W., Fu, H., Chen, A. T., Zhai, T., Jian, W., Xu, R., Pan, J., Hu, M., Zhou, Z., Chen, Q., et al. (2019). 10 years of health-care reform in china: Progress and gaps in universal health coverage. *The Lancet*, *394*(10204), 1192–1204.

Yip, W. C. (1998). Physician response to medicare fee reductions: Changes in the volume of coronary artery bypass graft (cabg) surgeries in the medicare and private sectors. *Journal of health economics*, *17*(6), 675–699.

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.

Zhou, Y., Yu, H., Wang, A., Liu, Y., Li, H., & Liu, J. (2024). Temporal and geographic trends in maternal and paternal ages in china from 2016 to 2020. *Chinese J Reproductive Health*, *35*(4), 301–5.

Zhu, Z., Wang, J., Chen, C., & Zhou, J. (2020). Hospitalization charges for extremely preterm infants: A ten-year analysis in shanghai, china. *Journal of Medical Economics*, *23*(12), 1610–1617.

Tables and Figures

Table 1: Effects of Hospital Inspections on Hospital-Level Outcomes

	Full	Non-pregnancy	Pregnancy	Pregnancy Loss	Prenatal	Delivery	Postnatal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Aggregate Revenue (RMB)							
D	401,940***	64,878	337,063***	3,360	67,797***	240,615***	25,292**
	(108,152)	(42,399)	(89,910)	(6,506)	(25,293)	(70,065)	(11,030)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947	2,947	2,947	2,947
R-squared	0.962	0.954	0.939	0.957	0.729	0.929	0.750
Pretreat Mean T	2,246,162	956,493	1,289,668	207,851	86,057	963,796	31,964
Magnitude	0.179	0.068	0.261	0.016	0.788	0.250	0.791
Panel B: Number of Admissions							
D	50.69***	7.26**	43.43***	0.30	19.74***	15.99*	7.40***
	(12.30)	(2.82)	(11.48)	(0.84)	(5.39)	(8.34)	(2.20)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947	2,947	2,947	2,947
R-squared	0.956	0.959	0.936	0.956	0.744	0.916	0.804
Pretreat Mean T	346.83	98.18	248.64	42.54	25.79	173.63	6.69
Magnitude	0.146	0.074	0.175	0.007	0.766	0.092	1.106
Panel C: Average Revenue (RMB)							
D	-23.38	-15.77	105.86	132.54	208.89	-231.17	679.59***
	(144.29)	(210.43)	(132.81)	(170.31)	(154.27)	(164.24)	(238.04)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947	2,947	2,947	2,947
R-squared	0.938	0.839	0.894	0.637	0.653	0.814	0.486
Pretreat Mean T	6,150.83	8,783.77	4,930.26	4,461.70	3,255.23	5,398.19	4,304.99
Magnitude	-0.004	-0.002	0.021	0.030	0.064	-0.043	0.158
Panel D: Average Length of Stay (Days)							
D	-0.26***	-0.04	-0.27***	-0.08	0.06	-0.27**	-0.04
	(0.08)	(0.20)	(0.09)	(0.17)	(0.17)	(0.13)	(0.19)
Obs. (Hospital-Month)	2,947	2,743	2,947	2,748	2,809	2,796	2,411
R-squared	0.909	0.545	0.885	0.515	0.429	0.765	0.402
Pretreat Mean T	7.71	10.25	6.63	7.31	6.37	6.64	7.05
Magnitude	-0.033	-0.003	-0.041	-0.011	0.009	-0.041	-0.005
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 inspections on hospital outcomes at the hospital-month level. Panel A reports the effects on aggregate medical revenue (RMB), Panel B on the number of admissions, Panel C on the average revenue per admission, and Panel D on the average length of stay. 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 classification of OBGYN admission types is based on ICD-10 codes. In Panel D, average length of stay is defined only for hospital-months with at least one admission of the relevant type, so the number of observations is smaller than in Panels A–C and varies across columns. Panels A–C report separate regressions on total revenue, admission volume, and revenue per admission. Because revenue is the product of the latter two, the three coefficients are not additively related: the mean of the per-admission ratio differs from the ratio of the means even before treatment, and the difference is absorbed by neither marginal regression. Row *D* reports the difference-in-differences (DiD) coefficients on the treatment indicator, which equals one for inspected (treatment) hospitals in the months on or after their inspection and zero otherwise. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); a value of 0.10, for example, denotes a 10% increase. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (87 hospitals, of which 25 are inspected and 62 are clean controls). Data are drawn from provincial administrative OBGYN discharge records and the Health Statistics Yearbooks. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Effects of Hospital Inspections on Maternal and Infant Health Outcomes

	(1)	(2)	(3)	(4)	(5)
Panel A: Infant Health					
	NICU	LBW	VLBW	Birth Weight	Birth Weight Concentration
D (per 1,000; col. 4 in g)	-6.89*	-7.45*	-2.15**	17.96***	13.82***
	(3.80)	(4.18)	(0.97)	(6.67)	(5.00)
Wild bootstrap p-value	[0.075]	[0.074]	[0.044]	[0.022]	[0.003]
Obs. (Delivery Admissions)	332,286	332,286	332,286	332,286	332,286
R-squared	0.039	0.039	0.014	0.034	0.013
Pretreat Mean T	51.68	83.58	16.38	3,229.32	721.63
Magnitude	-0.133	-0.089	-0.131	0.006	0.019
Panel B: Maternal Health					
	Surgical/Anesthesia Complications	Genital Tract Laceration	Retained Placenta	ARDS	Hypertensive Disorders
D (expressed per 1,000)	-0.19*	-6.71***	-8.20**	-0.14**	-8.76**
	(0.10)	(2.41)	(4.03)	(0.06)	(4.37)
Wild bootstrap p-value	[0.054]	[0.019]	[0.065]	[0.006]	[0.051]
Obs. (Delivery Admissions)	332,286	332,286	332,286	332,286	332,286
R-squared	0.002	0.217	0.034	0.000	0.030
Pretreat Mean T	0.02	63.83	22.69	0.13	64.20
Magnitude	—	-0.105	-0.362	—	-0.136
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the estimated effects of hospital inspections on infant and maternal health outcomes, using a sample restricted to delivery admissions. Panel A reports infant health outcomes. The dependent variables in columns (1)–(3) and (5) are binary indicators scaled to represent cases per 1,000 deliveries: NICU (= 1 for an inferred Neonatal Intensive Care Unit admission, identified from neonatal diagnosis codes), LBW (= 1 for low birth weight < 2,500 g), VLBW (= 1 for very low birth weight < 1,500 g), and Birth Weight Concentration (= 1 if birth weight falls within one standard deviation of the sample mean). Column (4) reports continuous birth weight in grams. Panel B reports maternal health outcomes. All dependent variables in Panel B are indicators scaled per 1,000 deliveries, capturing surgical or anesthesia complications, genital tract laceration (any obstetric laceration of the perineum, cervix, or upper vagina), retained placenta, ARDS (Acute Respiratory Distress Syndrome), and hypertensive disorders of pregnancy and eclampsia. Row D reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); it is not reported (—) for surgical or anesthesia complications and for ARDS, whose near-zero pre-treatment incidence (0.02 and 0.13 per 1,000, respectively) renders the proportional change uninformative. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (87 hospitals, of which 25 are inspected). Wild bootstrap p-values, computed with Webb weights and 9,999 replications, are reported in brackets for the coefficient on D to address the small number of treated clusters. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Drug-Service Substitution and C-section Provision

	(1)	(2)	(3)	(4)
Panel A: Components of Medical Revenue per Admission				
	Drug	Service	Diagnosis	Consumable
D	-226.63**	228.91***	-44.16	18.50
	(104.79)	(44.66)	(35.16)	(33.32)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947
R-squared	0.825	0.788	0.961	0.951
Pretreat Mean T	1,363.11	2,260.53	1,586.19	941.00
Magnitude	-0.166	0.101	-0.028	0.020
Panel B: Share of Components of Medical Revenue per Admission				
	Drug	Service	Diagnosis	Consumable
D	-0.03***	0.04***	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.00)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947
R-squared	0.679	0.855	0.841	0.883
Pretreat Mean T	0.21	0.38	0.26	0.15
Magnitude	-0.149	0.102	-0.028	-0.006
Panel C: Quantity Effect of Lower Drug Utilization				
	Drug Rev. = 0	< 10 RMB	< 50 RMB	< 100 RMB
D	1.65	3.90*	7.54*	11.36**
	(1.23)	(2.22)	(4.40)	(4.45)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947
R-squared	0.706	0.757	0.841	0.866
Pretreat Mean T	10.64	15.82	37.22	51.93
Magnitude	0.155	0.247	0.203	0.219
Panel D: C-section Rate				
Sample Restrictions	Full Delivery	Full Delivery	Age<35	Age≥35
D	-0.014	-0.026*	-0.014	-0.036***
	(0.01)	(0.01)	(0.01)	(0.01)
Obs. (Delivery Admissions)	332,286	332,286	297,479	34,807
R-squared	0.040	0.204	0.039	0.051
Pretreat Mean T	0.41	0.41	0.39	0.54
Magnitude	-0.035	-0.064	-0.035	-0.067
Case Mix	No	Yes	No	No
Hospital FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: This table investigates the policy effects on drug-service substitution at the hospital-month level and C-section provision at the admission level. Panel A examines the impact on the distinct components of average revenue per admission. The dependent variables are the average revenue from drugs (col. 1), medical services (col. 2), diagnostics (col. 3), and consumables (col. 4). Panel B examines the impact on the corresponding component shares of medical revenue per admission. Panel C reports the effects on the volume of drug utilization, measured by the prevalence of admissions with zero or very low drug revenue. Columns (1)–(4) report the estimated effects on the number of admissions with drug revenue equal to zero, below 10 RMB, below 50 RMB, and below 100 RMB, respectively. Panel D reports the impact on C-section provision among delivery admissions: columns (1)–(2) report the full delivery sample without and with case-mix adjustment (case mix comprises prior C-section, advanced maternal age, placenta accreta, and placenta previa), and columns (3)–(4) report deliveries by mothers below and at or above advanced maternal age (35 years). Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T). All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Heterogeneity in Prenatal Care Provision by Baseline C-section Rate

	Aggregate Rev.	Admissions	Average Rev.	ALOS	Service Rev.	Drug Rev.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Hospitals with Baseline C-section Rate ≥ 40%						
D	76,182**	20.03***	293	0.27	298***	-2
	(36,562)	(6.84)	(227)	(0.20)	(95)	(110)
Obs. (Hospital-Month)	1,136	1,136	1,136	1,120	1,136	1,136
R-squared	0.688	0.684	0.752	0.462	0.607	0.674
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Pretreat Mean T	88,929	25.12	3,282	6.08	1,055	638
Magnitude	0.857	0.798	0.089	0.044	0.282	-0.002
Panel B: Hospitals with Baseline C-section Rate < 40%						
D	46,679	16.91*	-13	-0.33	119	-85
	(28,110)	(9.83)	(185)	(0.31)	(83)	(78)
Obs. (Hospital-Month)	1,310	1,310	1,310	1,226	1,310	1,310
R-squared	0.840	0.845	0.579	0.435	0.540	0.424
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Pretreat Mean T	80,314	27.12	3,202	6.94	977	605
Magnitude	0.581	0.623	-0.004	-0.047	0.121	-0.141
Panel C: Physicians with Baseline C-section Rate ≥ 40%						
D	8,965	2.62**	367**	0.47*	270***	37
	(5,509)	(1.23)	(173)	(0.26)	(79)	(48)
Obs. (Physician-Month)	3,911	3,911	3,911	2,836	3,911	3,911
R-squared	0.502	0.570	0.349	0.212	0.304	0.326
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Pretreat Mean T	15,638	4.55	2,512	6.33	851	420
Magnitude	0.573	0.575	0.146	0.075	0.317	0.087
Panel D: Physicians with Baseline C-section Rate < 40%						
D	6,176	1.15	223	0.34	144	45
	(4,014)	(0.88)	(300)	(0.34)	(116)	(89)
Obs. (Physician-Month)	3,711	3,711	3,711	2,505	3,711	3,711
R-squared	0.435	0.480	0.255	0.251	0.277	0.150
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Pretreat Mean T	8,559	3.37	2,150	6.67	699	342
Magnitude	0.722	0.342	0.104	0.051	0.206	0.132
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table investigates the heterogeneous effects on the prenatal care provision by hospitals' and physicians' baseline C-section rate. Panels A and B stratify hospitals into two groups by baseline C-section rate (above and below 40%), and Panels C and D stratify physicians into two groups by the same threshold. Columns (1)–(4) report effects on aggregate revenue, admission volume, average revenue, and ALOS for prenatal admissions. Columns (5) and (6) report effects on service and drug revenue for prenatal admissions, respectively. Panels A and B are estimated at the hospital-month level and Panels C and D at the physician-month level. In column (4), ALOS is defined only for units with at least one prenatal admission, so its number of observations is smaller than in the other columns. Row D reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T). All specifications include hospital/physician and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Heterogeneity in Behavioral Response and Health Outcomes by Physician Seniority

Sample Restriction	Senior Attending Physicians				Junior Attending Physicians			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Drug-Service Substitution by Physician Seniority								
Outcomes	Drug Rev.	Service Rev.	Drug Share	Service Share	Drug Rev.	Service Rev.	Drug Share	Service Share
D	-197.68*	182.33***	-0.03***	0.03**	-115.56**	70.62	-0.02***	0.01
	(115.78)	(51.65)	(0.01)	(0.01)	(48.48)	(68.57)	(0.01)	(0.01)
Obs. (Full Admissions)	337,666	337,666	337,634	337,634	228,132	228,132	228,129	228,129
R-squared	0.196	0.206	0.304	0.419	0.116	0.286	0.223	0.490
Pretreat Mean T	1,422.06	2,493.99	0.17	0.41	1,032.40	2,762.72	0.14	0.44
Magnitude	-0.139	0.073	-0.178	0.071	-0.112	0.026	-0.157	0.015
Panel B: Prenatal Care Provision by Physician Seniority								
Outcomes	Anemia	Prior C-section	Age ≥ 35	GDM	Anemia	Prior C-section	Age ≥ 35	GDM
D	0.04*	0.03**	0.01	0.01	0.02	0.02	0.00	0.00
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Obs. (Prenatal Admissions)	33,633	33,633	33,633	33,633	27,838	27,838	27,838	27,838
R-squared	0.126	0.050	0.028	0.046	0.234	0.119	0.026	0.086
Pretreat Mean T	0.13	0.07	0.13	0.13	0.08	0.06	0.13	0.03
Magnitude	0.290	0.427	0.042	0.104	0.229	0.291	0.031	0.005
Panel C: Health Outcomes by Physician Seniority								
Outcomes	Birth Weight	LBW	Genital Tract Laceration	Hypertensive Disorders	Birth Weight	LBW	Genital Tract Laceration	Hypertensive Disorders
D	18.71**	-8.98*	-4.86**	-12.32**	9.81	-0.94	-13.37***	0.27
	(7.63)	(4.61)	(2.11)	(5.30)	(11.29)	(5.58)	(5.04)	(8.90)
Obs. (Delivery Admissions)	188,358	188,358	188,358	188,358	132,323	132,323	132,323	132,323
R-squared	0.038	0.041	0.215	0.030	0.031	0.038	0.215	0.032
Pretreat Mean T	3,225.93	84.94	66.47	65.25	3,235.63	80.78	54.03	61.61
Magnitude	0.006	-0.106	-0.073	-0.189	0.003	-0.012	-0.247	0.004
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports heterogeneous effects on behavioral responses and health outcomes by the seniority of the attending physician. Panel A shows the effects on per-admission drug and service revenue and on the drug and service shares of total revenue among senior and junior physicians. Panel B shows the effects on the admission rate of prenatal admissions with obstetric comorbidities (such as anemia, prior C-section, advanced maternal age, and gestational diabetes mellitus, GDM) among senior and junior physicians. Panel C shows the effects on health outcomes among delivery admissions attended by senior and junior physicians. An attending physician is classified as senior if their surname matches that of a Chief Physician or Department Head recorded on the same hospitalization, and junior otherwise. The discharge records carry no unique physician identifier, so this classification is a proxy and errs in one direction only: a senior attending who coincidentally shares a surname with the resident physician on the same record is recorded as junior, while the converse cannot occur. The junior group is therefore mildly contaminated with senior physicians and the senior group is clean, which attenuates the senior–junior contrast rather than creating it. Table A9 re-estimates Panel A excluding the most common surnames, where such coincidences are concentrated. Because the split requires an identifiable attending physician, the senior and junior subsamples do not sum to the full estimation sample of Table 2: admissions with no linkable attending are excluded from both. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T). All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Physician Turnover and Maternal and Infant Health Outcomes

	(1)	(2)	(3)	(4)
Panel A: Physician Turnover				
	Exits $[-6, +6]$	Entries $[-6, +6]$	Exits $[-3, +3]$	Entries $[-3, +3]$
D	0.12**	0.11*	0.16**	0.19**
	(0.06)	(0.07)	(0.06)	(0.09)
Obs. (Hospital-Month)	1,029	1,029	554	554
R-squared	0.167	0.187	0.225	0.255
Pretreat Mean T	0.09	0.26	0.04	0.16
Magnitude	1.351	0.438	4.108	1.215
Panel B: Health Effects in Hospitals With Physician Exits Within Six Months of Inspection				
	Birth Weight	LBW	Genital Tract Laceration	Hypertensive Disorders
D	23.13***	-10.45**	-6.08*	-10.39*
	(8.04)	(5.01)	(3.40)	(5.39)
Obs. (Delivery Admissions)	211,603	211,603	211,603	211,603
R-squared	0.039	0.044	0.217	0.033
Pretreat Mean T	3,213.15	90.40	71.65	67.07
Magnitude	0.007	-0.116	-0.085	-0.155
Panel C: Health Effects in Hospitals Without Physician Exits Within Six Months of Inspection				
	Birth Weight	LBW	Genital Tract Laceration	Hypertensive Disorders
D	-5.24	3.50	-6.40	-2.34
	(13.25)	(3.53)	(5.06)	(5.61)
Obs. (Delivery Admissions)	109,078	109,078	109,078	109,078
R-squared	0.017	0.012	0.033	0.014
Pretreat Mean T	3,313.30	47.61	13.58	48.12
Magnitude	-0.002	0.074	-0.471	-0.049
Hospital FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: This table reports the impact of hospital inspections on physician turnover and the heterogeneity in health outcomes between hospitals with and without physician exits within six months after the inspection. Panel A presents policy effects on the number of departing (Exits) and newly arriving (Entries) attending physicians, identified by the last and first appearance of a physician's name in the dataset, respectively. Columns (1) and (2) restrict the sample to a six-month window around each hospital's inspection, while columns (3) and (4) use a three-month window. Panels B and C report effects on key health outcomes, splitting the sample into hospitals with at least one physician exit within the six-month window after the inspection (Panel B) and hospitals with no physician exits over the same period (Panel C). In both panels, the dependent variables in columns (1)–(4) are birth weight (in grams), low-birth-weight incidence, laceration incidence, and the incidence of hypertensive disorders of pregnancy and eclampsia, respectively. Because assigning a hospital to Panel B or Panel C requires physician appearances to be tracked over time, the two panels do not sum to the full estimation sample of Table 2: admissions with no linkable attending are excluded from both. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T). All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Sensitivity of the Net Social Surplus

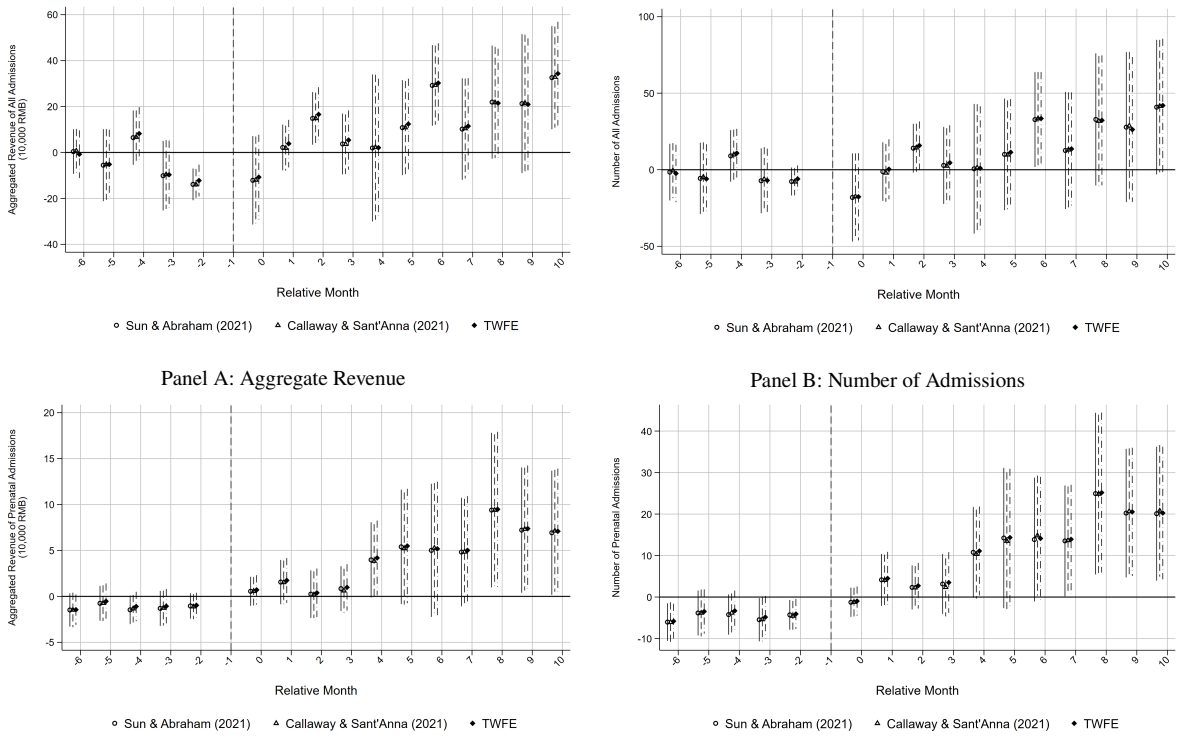
Panel A: Long-run real wage growth g (discount rate $r = 3\%$)

	$g = 1\%$	$g = 2\%$	$g = 3\%$	$g = 4\%$	$g = 5\%$
PV per child, non-private (RMB)	1,338	1,918	2,774	4,049	5,961
PV per child, private (RMB)	824	1,181	1,708	2,493	3,670
Human capital benefit	109.17	156.51	226.42	330.49	486.54
Total benefit	181.74	229.08	298.99	403.06	559.11
Net social surplus	**-25.25**	**22.08**	**91.99**	**196.06**	**352.11**
Memo: wedge $r - g$ (pp)	2.0	1.0	0.0	-1.0	-2.0

Panel B: One-at-a-time variation around the benchmark ($g = 2\%$, $r = 3\%$)

	Net social surplus
Benchmark	22.08
VLBW treatment cost 150,000 RMB	-15.93
VLBW treatment cost 500,000 RMB	60.09
Private-sector wage only	-15.14
Non-private-sector wage only	59.30
Birth-weight elasticity 1.375%	-56.17
Birth-weight elasticity 4.125%	100.33
Labor-market entry at age 22	19.06
Conservative growth ($g = 1\%$)	-25.25
Transfer share of revenue 10%	42.78
Transfer share of revenue 20%	63.48
Robust inputs, Sun and Abraham	35.70
Robust inputs, Callaway and Sant'Anna	-14.86

Notes: All entries are in millions of nominal RMB unless otherwise stated. The benchmark specification is that of Section 6: a discount rate of $r = 3\%$, real wage growth of $g = 2\%$, labor-market entry at age 20, the midpoint of the private and non-private wage benchmarks, a birth-weight–earnings elasticity of 2.75% per standard deviation (Lambiris et al., 2022), and the midpoint of the 150,000–500,000 RMB range for VLBW treatment cost (Zhu et al., 2020). Panel A varies long-run real wage growth, applying it both to the working life and to the years between the wage observation and labor-market entry, as in Equations (4) and (5). Because the valuation depends on g and r only through the ratio $(1+g)/(1+r)$, the memo row reports the equivalent wedge, and a higher discount rate paired with proportionally higher wage growth leaves the columns nearly unchanged: the surplus at $(r, g) = (4\%, 3\%)$ is 21.1 million and at $(5\%, 4\%)$ is 20.2 million, against 22.1 million at our benchmark $(3\%, 2\%)$. The surplus reaches zero at $g \approx 1.58\%$, a wedge of about 1.4 percentage points, and that breakeven wedge is stable across discount rates. Panel B varies one parameter at a time, holding the others at the benchmark. "Transfer share of revenue" is the fraction of the estimated revenue increase treated as rent transferred to providers rather than real resources consumed, so that the policy cost becomes $207 \times (1 - \tau)$ million RMB; at the more conservative $g = 1\%$ the surplus would reach zero at $\tau \approx 12.2\%$. The 401,940 RMB monthly revenue increase underlying the 207 million RMB cost is the canonical two-way fixed effects estimate. The final two rows, labeled "robust inputs", replace it with the corresponding heterogeneity-robust estimate from Table A4, and replace the birth weight, very low birth weight and laceration inputs with the estimates from the same estimator: all four inputs move together, and substituting the revenue estimate alone would cut the cost while leaving the benefits at their canonical values. Data are drawn from provincial administrative OBGYN discharge records.



Panel A: Aggregate Revenue

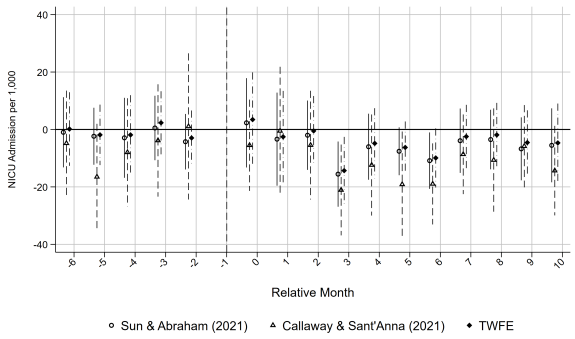
Panel B: Number of Admissions

Panel C: Prenatal Revenue

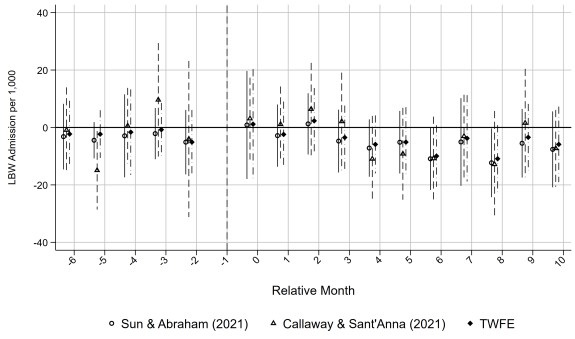
Panel D: Number of Prenatal Admissions

Figure 1: Event Study of Hospital-Level Outcomes

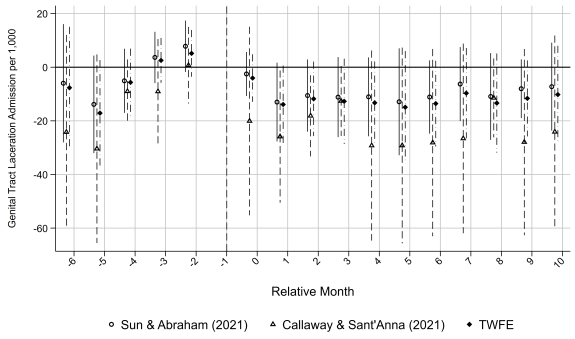
Notes: This figure plots event study estimates illustrating the dynamic impacts of hospital inspections on key hospital outcomes, specifically aggregate revenue and admission volumes. Revenue is measured here in 10,000 RMB, whereas Table 1 reports the same outcome in RMB. Panels A and B present the estimated effects on aggregate revenue and the total number of admissions for all OBGYN patients, respectively. Panels C and D report the corresponding effects on aggregate revenue and admission volumes for prenatal care. To ensure robustness against potential biases in standard difference-in-differences models with staggered timing, the figure 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 ($k = 0$), and vertical bars represent 95% confidence intervals. The event window spans from six months before to ten months after the inspection; observations outside this window are dropped, and $k = -1$ serves as the omitted reference period. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.



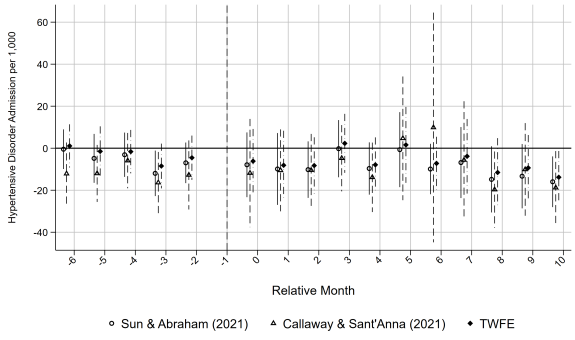
Panel A: Inferred NICU Admission



Panel B: Low Birth Weight



Panel C: Genital Tract Laceration



Panel D: Hypertensive Disorders

Figure 2: Event Study of Maternal and Infant Health Outcomes

Notes: This figure plots event study estimates illustrating the dynamic effects of hospital inspections on key infant and maternal health outcomes. Panels A through D present the estimated effects on the incidence of inferred infant NICU admissions, low-birth-weight cases, genital tract lacerations, and hypertensive disorders of pregnancy and eclampsia, respectively. To ensure robustness against potential biases in standard difference-in-differences models with staggered timing, the figure 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 ($k = 0$), and vertical bars represent 95% confidence intervals. The event window spans from six months before to ten months after the inspection; observations outside this window are dropped, and $k = -1$ serves as the omitted reference period. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

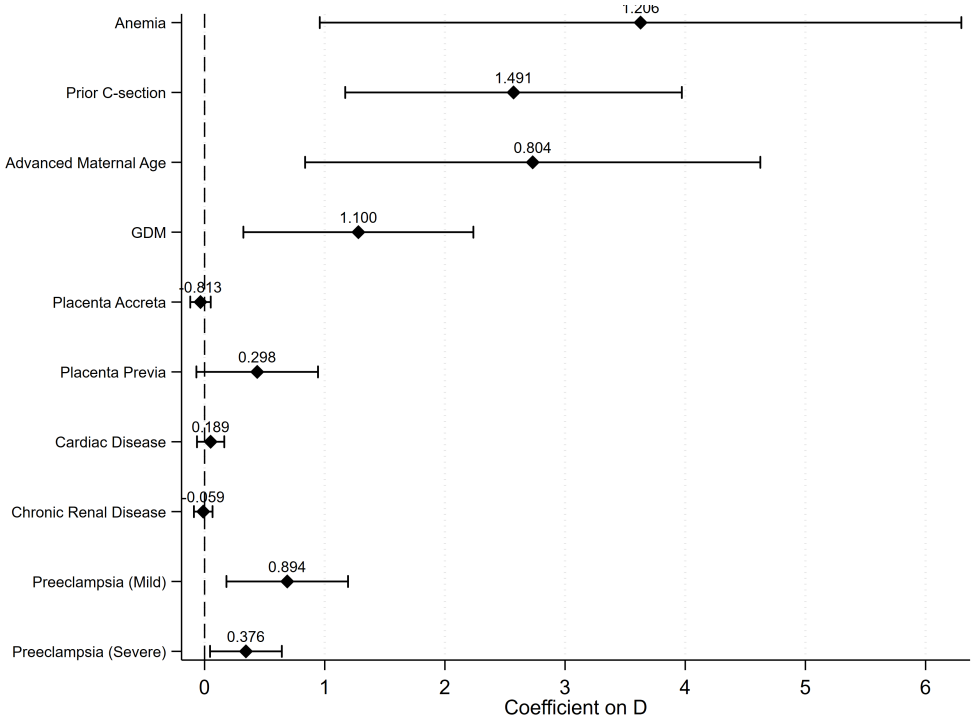
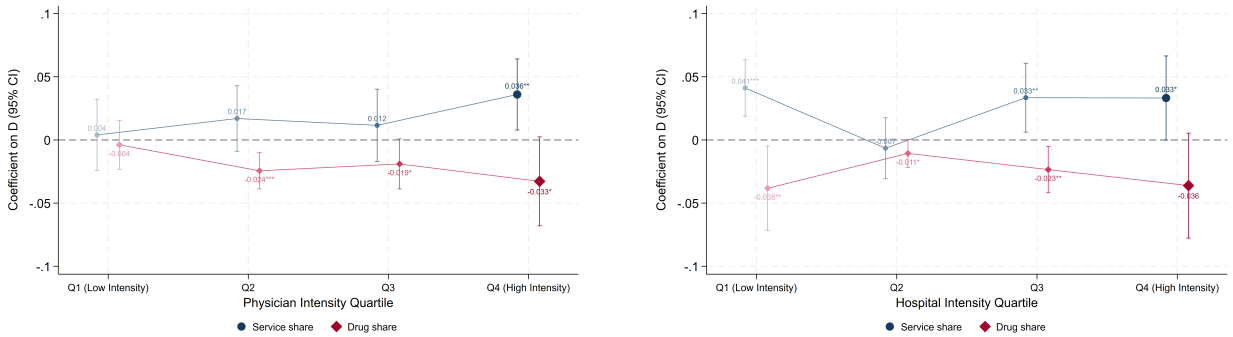


Figure 3: Prenatal Care Provision for Pregnant Women with Obstetric Comorbidities

Notes: This figure plots the impact on the number of prenatal admissions for pregnant women with different obstetric comorbidities at the hospital-month level. The dependent variable in each row is a binary indicator equal to one if the admission carries a diagnosis of the comorbidity listed on the vertical axis, and zero otherwise. The comorbidities follow the clinical risk classification of Leonard et al. (2020) as predictors of severe maternal morbidity. Markers represent point estimates, labeled with the proportional change relative to the pre-treatment mean of the treatment group, and horizontal lines indicate 95% confidence intervals. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

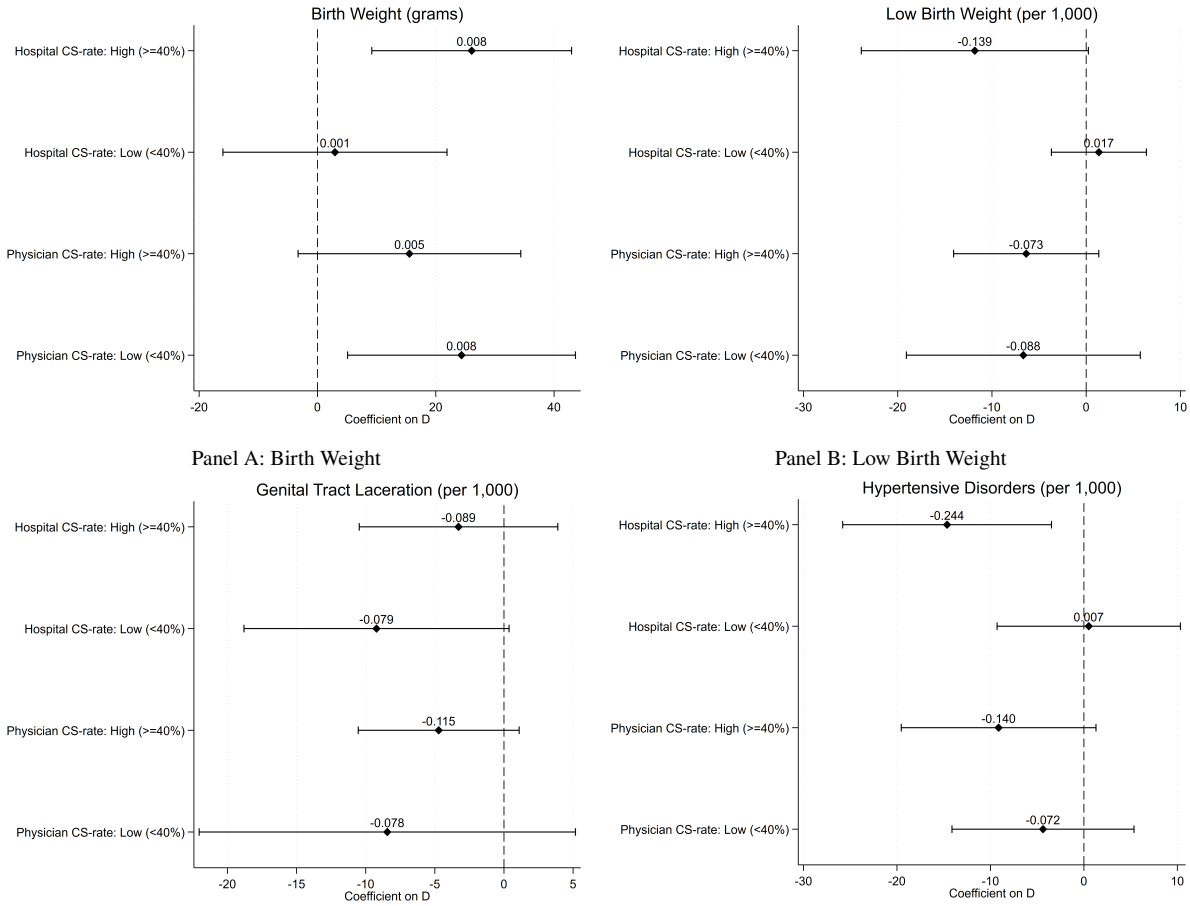


Panel A: By Physician-Level Intensity

Panel B: By Hospital-Level Intensity

Figure 4: Heterogeneity in Drug-Service Substitution by Baseline Drug Prescribing Intensity

Notes: This figure plots heterogeneity in the effects of hospital inspections on drug-service substitution by physicians' and hospitals' baseline drug prescribing intensity. Panel A reports DiD coefficients estimated separately for each quartile of physician prescribing intensity, and Panel B for each quartile of hospital intensity. Exposure is measured by the physician (hospital) fixed effects recovered from a pre-inspection regression of drug revenue shares, so that higher quartiles correspond to physicians (hospitals) with systematically higher baseline drug shares. Navy circles report effects on service revenue shares, and cranberry diamonds report effects on drug revenue shares. Markers represent point estimates, labeled with the coefficient and significance stars, and vertical lines indicate 95% confidence intervals. Specifications in Panel A include attending physician and time fixed effects; specifications in Panel B include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.



Panel A: Birth Weight

Panel B: Low Birth Weight

Panel C: Laceration

Panel D: Hypertensive Disorders

Figure 5: Heterogeneity in Health Outcomes by Baseline C-section Rate

Notes: This figure plots the heterogeneous effects on key maternal and infant health outcomes by hospitals' and physicians' baseline C-section rate. The dependent variables in Panels A-D are birth weight, low-birth-weight cases per 1,000 deliveries, laceration cases per 1,000 deliveries, and hypertensive-disorder cases per 1,000 deliveries. Markers represent point estimates for each sub group, labeled with the proportional change relative to the pre-treatment mean of the treatment group, and horizontal lines indicate 95% confidence intervals. All specifications include hospital/physician and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

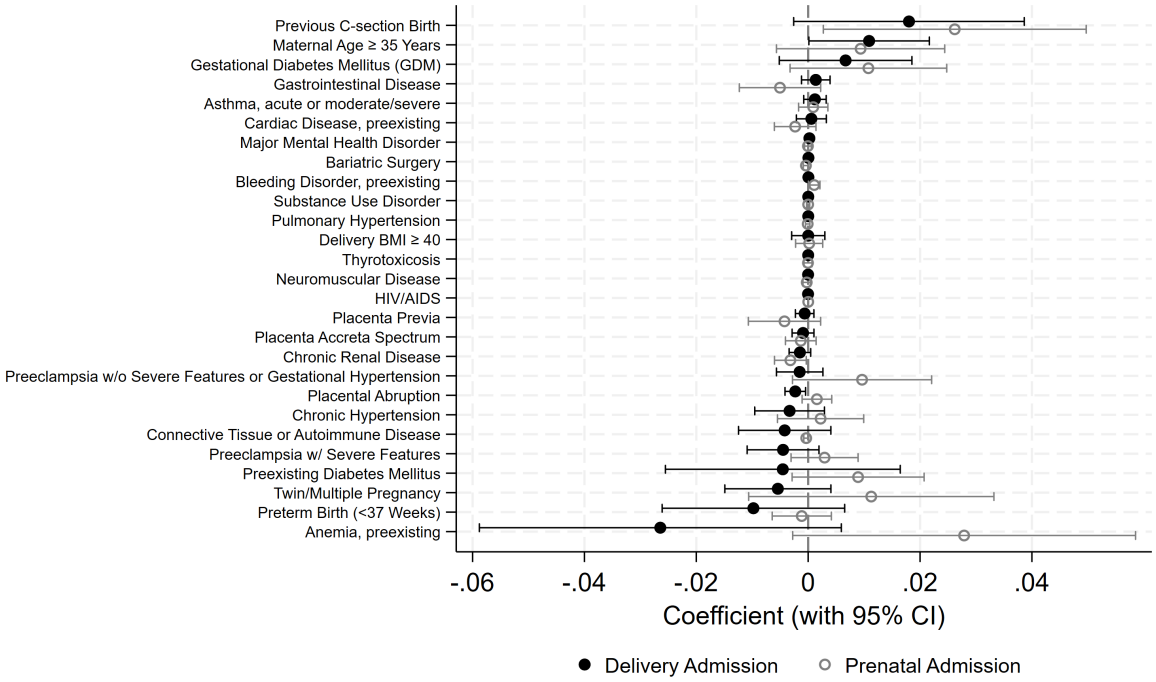


Figure 6: Effects on the Composition of Obstetric Comorbidities

Notes: This figure plots the estimated regression coefficients for the likelihood of 27 specific obstetric comorbidities among prenatal admissions ($N = 63,958$) and delivery admissions ($N = 332,286$). The dependent variable in each regression is a binary indicator equal to one if the patient is diagnosed with the specific comorbidity listed on the y-axis during their admission, 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. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

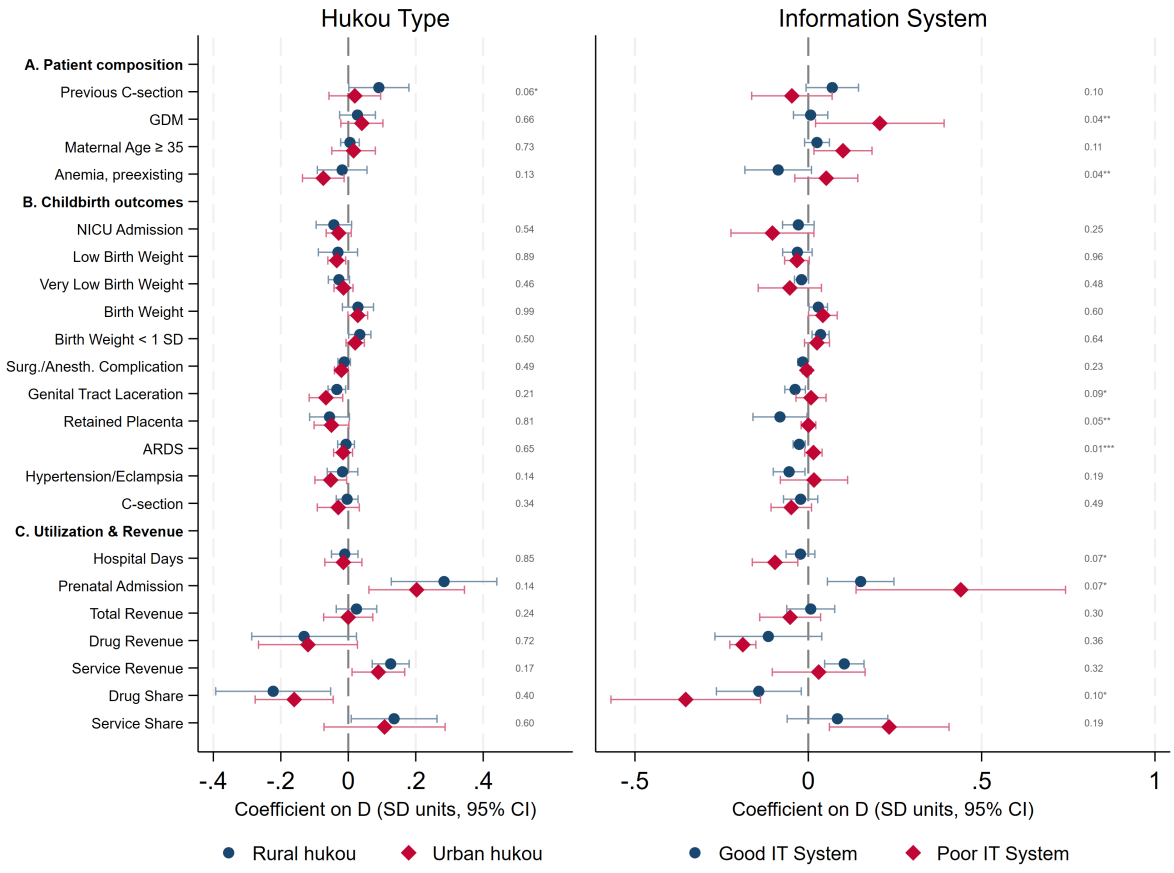


Figure 7: Heterogeneous Effects by Hukou Type and Hospital Information System

Notes: This figure plots heterogeneous effects of hospital inspections by patients' hukou type (left panel) and by hospitals' information system quality (right panel). Within each panel, markers report the DiD coefficients estimated separately for each subgroup: rural ($N = 282,258$) versus urban ($N = 141,542$) hukou patients in the left panel, and hospitals with good ($N = 427,886$) versus poor ($N = 157,753$) information systems in the right panel. Outcomes are grouped into patient composition, childbirth outcomes, and utilization and revenue, with rows aligned across the two panels. The observation counts reported above are the maximum estimation sample for each subgroup, drawn from all OBGYN admissions; the childbirth-outcome rows are estimated on delivery admissions only and therefore use a smaller sample. The two hukou groups do not sum to the full sample because admissions paid by neither an urban scheme nor the NCMS cannot be assigned a hukou type. All outcomes are standardized to z-scores before estimation (patient composition and childbirth outcomes are standardized within the childbirth sample), so coefficients are in standard deviation units. Markers represent point estimates, and horizontal lines indicate 95% confidence intervals. The right margin of each panel reports the p-value of the between-group difference in treatment effects, obtained from a fully interacted pooled model. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records and the Health Statistics Yearbooks. *** p<0.01, ** p<0.05, * p<0.1.

A Appendix Tables and Figures

Table A1: Sample Composition across Sequential Restrictions

	(1) Full Sample	(2) No Small Hosp.	(3) No Hosp. W/O CB	(4) No Outlier Month	(5) No 3rd-Round Hosp.	(6) No Spillover Cont.
Demographic Information						
Age at Adm.	31.36	31.41	31.21	31.21	30.99	31.31
Share Female	0.99	0.99	0.99	0.99	0.99	0.99
Share Married	0.94	0.94	0.94	0.94	0.94	0.94
Share UEBMI	0.15	0.16	0.16	0.16	0.16	0.17
Share URBMI and NCMS	0.53	0.53	0.52	0.52	0.53	0.55
Clinical Information						
Share ED Adm.	0.14	0.14	0.15	0.15	0.15	0.15
Share Delivery Adm.	0.49	0.49	0.59	0.59	0.59	0.57
Length of Stay	6.57	6.59	6.51	6.51	6.41	6.52
Per-adm. Revenue	5,003.24	5,051.04	5,014.25	5,014.99	4,854.65	4,819.54
Per-adm. Drug Revenue	966.65	978.60	886.74	886.79	832.62	840.30
Per-adm. Service Revenue	2,227.90	2,241.31	2,288.13	2,288.43	2,261.14	2,125.86
Per-adm. Diagnosis Revenue	1,135.49	1,147.78	1,151.48	1,151.72	1,112.07	1,183.61
Per-adm. Consumable Revenue	673.21	683.35	687.90	688.05	648.83	669.77
N	929,831	910,572	759,341	758,970	686,586	585,639
% of Sample (col. [1])	100%	97.93%	81.66%	81.62%	73.84%	62.98%
Share Treatment	0.47	0.48	0.53	0.53	0.48	0.56

Notes: This table displays summary statistics across different sample restrictions, with each column adding a further constraint to the raw OBGYN discharge records. Column (1) reports all OBGYN admissions. Column (2) removes small hospitals based on bed count and average monthly admissions, column (3) drops hospitals missing birth weight data for over half the study months, and column (4) excludes outlier months with fewer than 10 admissions. Column (5) removes the seven hospitals inspected during the September 2017 inspection round, while column (6) drops the spillover control group. The final treatment and control groups are built from this restricted sample. "Demographic Information" includes patient age, gender, marital status, and insurance type at admission, while "Clinical Information" comprises the emergency department (ED) admission rate, the share of deliveries among OBGYN admissions, length of stay, and per-admission revenue (drugs, services, diagnostics, and consumables). "N" represents the number of remaining admissions. "% of Sample" is the percentage of the raw data retained, and "Share Treatment" is the percentage of total admissions occurring in inspected hospitals. Data are drawn from provincial administrative OBGYN discharge records.

Table A2: Baseline Summary Statistics of Hospital Characteristics and Clinical Outcomes

	(1) Treatment Group	(2) Spillover Group	(3) Clean Control Group	(4) Diff (1–3)
No. of Admissions	293.98	204.07	106.71	187.27***
	(226.48)	(154.87)	(69.94)	(10.83)
No. of Non-pregnancy Admissions	91.16	17.65	8.66	82.51***
	(77.39)	(16.72)	(9.61)	(3.43)
No. of Pregnancy Admissions	202.81	186.43	98.05	104.76***
	(164.71)	(148.85)	(64.95)	(8.28)
No. of Prenatal Admissions	24.80	20.00	20.86	3.94
	(28.52)	(18.89)	(33.55)	(2.46)
No. of Delivery Admissions	133.26	139.03	54.03	79.23***
	(133.89)	(115.68)	(60.27)	(6.98)
No. of Postnatal Admissions	1.83	1.75	0.73	1.10***
	(2.45)	(2.83)	(1.46)	(0.14)
Ave. Revenue per Admission (RMB)	6,162.19	6,332.45	2,674.46	3,487.72***
	(2,318.16)	(7,780.06)	(954.34)	(117.84)
Ave. Drug Revenue per Admission (RMB)	1,393.62	926.24	450.54	943.08***
	(1,001.23)	(722.06)	(261.77)	(46.74)
Ave. Service Revenue per Admission (RMB)	2,252.06	4,205.37	1,481.64	770.41***
	(603.69)	(6,521.19)	(749.23)	(54.23)
Ave. Diagnosis Revenue per Admission (RMB)	1,575.99	637.25	521.85	1,054.15***
	(655.14)	(423.33)	(310.86)	(34.71)
Ave. Consumable Revenue per Admission (RMB)	940.52	563.59	220.43	720.08***
	(558.76)	(522.60)	(169.37)	(26.64)
ALOS (days)	7.80	6.30	5.55	2.25***
	(1.40)	(1.70)	(1.46)	(0.11)
C-section Rate (%)	0.44	0.37	0.35	0.09***
	(0.14)	(0.11)	(0.18)	(0.01)
LBW Rate (expressed per 1,000)	71.46	23.72	22.42	49.04***
	(60.44)	(22.07)	(38.94)	(3.84)
VLBW Rate (expressed per 1,000)	13.91	2.69	1.35	12.55***
	(18.79)	(5.31)	(4.72)	(0.96)
Birth Weight (gram)	3,261.40	3,334.87	3,345.74	-84.34***
	(131.37)	(76.86)	(147.04)	(11.30)
Maternal Mortality Rate (expressed per 100,000)	0.00	0.00	43.89	-43.89
	(0.00)	(0.00)	(634.56)	(40.15)
N	250	80	524	774

Notes: This table reports the summary statistics of key hospital-level variables in the baseline period, with standard deviations in parentheses for columns (1)–(3) and standard errors in parentheses for column (4). Column (1) restricts the sample to the treatment group. Columns (2) and (3) divide the control group into a "spillover" group (non-inspected hospitals in districts with at least one inspected hospital) and a "clean" group (non-inspected hospitals in districts without any inspected hospitals), respectively. Column (4) reports the difference in means between the treatment group and the clean control group, and significance stars are from a two-sample t-test of the equality of means. "No. of Admissions" represents the monthly volume of OBGYN patients, with subsequent rows detailing admission counts by specific diagnoses. Financial variables are expressed in RMB and represent the average total revenue and its components (drug, service, diagnosis, and consumable revenue) per admission. "ALOS" denotes the average length of stay in days. "C-section Rate" represents the percentage of deliveries performed via C-section. "LBW Rate" and "VLBW Rate" indicate the prevalence of low and very low birth weight cases, respectively, measured per 1,000 deliveries. "Birth Weight" denotes newborn birth weight in grams. "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. Data are drawn from provincial administrative OBGYN discharge records and the Health Statistics Yearbooks. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Balance of Baseline Characteristics across Inspection Cohorts

	(1) First Cohort (Nov. 2015)	(2) Second Cohort (Sept. 2016)	(3) Diff (1–2)	(4) S.E.	(5) Norm. Diff.
Panel A. Hospital structure and capacity (2014)					
Total Beds	971	927	44	(177)	0.102
Health Technical Personnel	1,149	1,316	-167	(201)	-0.332
Licensed Physicians	402	418	-15	(69)	-0.091
Tier 3 Hospital (= 1)	0.93	1.00	-0.07	(0.07)	-0.378
Total Hospital Revenue (million RMB)	572.4	591.4	-18.9	(160.8)	-0.048
Total Inpatient Revenue (million RMB)	385.0	389.0	-4.0	(110.1)	-0.015
Total Outpatient Revenue (million RMB)	146.2	154.8	-8.6	(40.5)	-0.086
Total Patient Number	503,711.9	548,372.2	-44,660.3	(125,031.0)	-0.144
Total Inpatient Number	29,892.1	31,918.6	-2,026.6	(5,236.9)	-0.157
Total Outpatient Number	458,827.9	490,145.0	-31,317.1	(109,476.3)	-0.115
Panel B. OBGYN operations and Health Outcomes (January–October 2015)					
Total Revenue (thousand RMB)	1,520.59	2,506.64	-986.05	(771.75)	-0.534
No. of Admissions	244.60	367.24	-122.65	(99.39)	-0.517
Ave. Revenue per Admission (RMB)	5,925.91	6,492.94	-567.04	(998.48)	-0.234
Drug Share of OBGYN Revenue (%)	0.21	0.22	-0.01	(0.03)	-0.142
Service Share of OBGYN Revenue (%)	0.37	0.39	-0.02	(0.03)	-0.247
C-section Rate (%)	0.42	0.46	-0.04	(0.05)	-0.344
Birth Weight (gram)	3,282.36	3,234.65	47.71	(45.32)	0.431
LBW Rate (expressed per 1,000)	57.01	87.19	-30.18	(22.11)	-0.567
VLBW Rate (expressed per 1,000)	10.22	18.70	-8.48	(6.03)	-0.579
NICU Admission (expressed per 1,000)	39.15	60.33	-21.18	(21.13)	-0.420
Genital Tract Laceration (per 1,000)	60.33	52.48	7.85	(43.09)	0.071
Hypertensive Disorders (per 1,000)	59.99	79.84	-19.85	(22.39)	-0.365
Panel C. Pre-treatment trends (January–October 2015; monthly slope)					
Total OBGYN Revenue	-36,700.8	-17,125.8	-19,575.0	(11,932.6)	—
No. of Admissions	-6.106	-6.898	0.792	(3.643)	—
No. of Prenatal Admissions	-0.118	0.866	-0.983	(0.656)	—
Drug Share of OBGYN Revenue	-0.0009	-0.0019	0.0010	(0.0010)	—
C-section Rate	-0.0031	-0.0059	0.0028	(0.0037)	—
Birth Weight	-2.364	-1.421	-0.943	(2.311)	—
Genital Tract Laceration	0.803	0.981	-0.178	(1.184)	—
No. of Hospitals	14	11	25		

Notes: This table tests whether assignment to the first (November 2015) or second (September 2016) inspection cohort is related to baseline hospital characteristics. The unit of observation is the hospital, and the sample is the 25 inspected hospitals in the main analytical sample. Columns (1) and (2) report cohort means; column (3) reports the difference and column (4) its heteroskedasticity-robust standard error, each from a separate regression of the listed characteristic on a first-cohort indicator. Column (5) reports the normalized difference, defined as the difference in means divided by the square root of the average of the two within-cohort variances (Imbens & Rubin, 2015). Conventional thresholds for this statistic are calibrated for large samples and are uninformative here: with 14 and 11 hospitals its sampling standard deviation under the null of exchangeable cohort labels is 0.40, so individual values near 0.5 lie well within sampling variation. Statistical insignificance in column (4) likewise reflects limited power as much as balance. We therefore benchmark the table against the permutation distribution reported below rather than against a fixed cutoff.

Panel A uses the 2014 wave of the Health Statistics Yearbooks, which precedes both the publication of the inspection roster and the start of the discharge panel. The Tier 3 indicator differs for a single hospital; because the second cohort has no variation in it, its normalized difference is mechanically inflated and should not be read as substantive imbalance. Panel B uses administrative OBGYN discharge records for January–October 2015, the longest window that precedes treatment for both cohorts and therefore the widest one available for a cohort-balance comparison. It is deliberately wider than the January–June 2015 window used to construct the baseline exposure measures in Section 5, which is chosen for a different purpose. Panel C reports the fitted monthly slope of each outcome over that same window; column (3) is the coefficient on the interaction of a linear month trend with the first-cohort indicator, estimated with hospital fixed effects and standard errors clustered at the hospital level, and column (5) does not apply. Because the inspection roster was published in April 2015, this window is partly post-announcement, so a differential trend in Panel C may reflect differential anticipation rather than differing pre-existing dynamics.

A joint permutation test reassigns the cohort label across the 25 hospitals 10,000 times, holding the 14–11 split fixed, and recomputes the balance statistics for the 22 characteristics in Panels A and B. This accounts jointly for the small number of hospitals and for the number of characteristics tested. The largest absolute normalized difference in the data, 0.58, falls at the 20th percentile of its permutation distribution ($p = 0.79$), which has a median of 0.74 and a 95th percentile of 1.21; the average absolute normalized difference, 0.27, falls at the 37th percentile ($p = 0.63$). The imbalance observed across these cohorts is therefore no larger than a random split of the same hospitals would typically produce. Data are drawn from provincial administrative OBGYN discharge records and the Health Statistics Yearbooks.

Table A4: Heterogeneity-Robust Estimates of the Inspection Effect

	(1) TWFE	(2) Callaway and Sant'Anna	(3) Sun and Abraham
Panel A. Hospital-month outcomes			
Aggregate Revenue (RMB)	401,940***	281,361**	274,208**
	(108,152)	(105,423)	(101,228)
No. of Admissions	50.69***	30.60*	28.41*
	(12.30)	(15.46)	(14.69)
Average Revenue per Admission (RMB)	-23.38	-70.56	-46.98
	(144.29)	(124.81)	(116.58)
Prenatal Revenue (RMB)	67,797***	56,858**	56,650**
	(25,293)	(24,239)	(24,122)
No. of Prenatal Admissions	19.74***	16.28**	16.17**
	(5.39)	(6.14)	(6.11)
Panel B. Delivery-admission outcomes (per 1,000; birth weight in g)			
NICU Admission	-6.89*	-5.20	-7.80
	(3.80)	(5.16)	(5.30)
Low Birth Weight	-7.45*	-4.46	-7.93
	(4.18)	(4.21)	(4.87)
Very Low Birth Weight	-2.15**	-1.71	-2.25
	(0.97)	(2.52)	(2.91)
Birth Weight (g)	17.96***	8.26	11.49
	(6.67)	(10.58)	(10.58)
Birth Weight Concentration	13.82***	12.90	18.33*
	(5.00)	(9.33)	(9.84)
Genital Tract Laceration	-6.71***	-7.06	-10.66**
	(2.41)	(8.20)	(4.79)
Retained Placenta	-8.20**	-0.41	-0.81
	(4.03)	(3.34)	(3.40)
Hypertensive Disorders	-8.76**	-7.61	-10.19
	(4.37)	(5.77)	(6.22)

Notes: This table reports the inspection effect for each main outcome under three estimators, on the same estimation sample as Tables 1 and 2. Column (1) reproduces the canonical two-way fixed effects estimate reported there. Column (2) reports the aggregated group-time average treatment effect of Callaway and Sant'Anna (2021), and column (3) the aggregated interaction-weighted estimate of Sun and Abraham (2021), both using the clean control group as the comparison group and averaging over event times $k = 0$ to $k = 10$ with weights proportional to the number of treated observations at each horizon. Standard errors are clustered at the hospital level and reported in parentheses.

The three estimators agree in sign for every outcome, and their event-study coefficients—each measured against $k = -1$—lie close together throughout (Figures 1 and 2). The aggregates in this table nonetheless differ, because the canonical estimator compares the post-inspection period to the entire pre-inspection average while the heterogeneity-robust estimators difference against the month immediately preceding inspection. The difference between the two columns is therefore a statement about which pre-inspection baseline is used, not about negative weighting, which a Goodman–Bacon decomposition of the canonical estimator shows to be immaterial in this setting.

Two entries warrant comment. The mechanism variables—prenatal revenue and prenatal admissions—are the most stable in the table, moving by roughly 13% across estimators. Retained placenta is the least stable: its canonical estimate reflects a decline that was already under way before the inspection, and the robust estimators place the effect at approximately zero, which is why we do not count it among the maternal health results. Because the effect on aggregate revenue grows over the event window rather than shifting once at the inspection, any single aggregate is sensitive to how horizons are weighted; the event studies should be read alongside this table. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Effects on Hospital-Level and Health Outcomes, Controlling for the Zero-Markup Drug Policy

	(1)	(2)	(3)	(4)	(5)
Panel A: Aggregate Revenue (RMB)					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	307,523***	2,154	50,182***	203,227***	21,610**
	(90,158)	(5,918)	(18,808)	(65,182)	(9,940)
D×ZMDP	185,311***	2,366	34,573**	73,381*	7,226
	(65,720)	(4,897)	(16,400)	(39,206)	(9,851)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947	2,947
R-squared	0.963	0.957	0.732	0.929	0.750
Pretreat Mean T	2,246,162	207,851	86,057	963,796	31,964
Magnitude	0.137	0.010	0.583	0.211	0.676
Panel B: Number of Admissions					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	39.15***	0.14	15.15***	15.22*	6.09***
	(10.20)	(1.04)	(4.52)	(8.12)	(2.20)
D×ZMDP	22.64**	0.31	9.02**	1.50	2.57
	(9.49)	(1.02)	(3.65)	(6.30)	(2.49)
Obs. (Hospital-Month)	2,947	2,947	2,947	2,947	2,947
R-squared	0.957	0.956	0.747	0.916	0.805
Pretreat Mean T	346.83	42.54	25.79	173.63	6.69
Magnitude	0.113	0.003	0.587	0.088	0.911
Panel C: Infant Health					
	NICU	LBW	VLBW	Birth Weight	Birth Weight Concentration
D (per 1,000; col. 4 in g)	-4.43	-6.13*	-1.35	16.49**	12.01***
	(3.58)	(3.46)	(1.09)	(6.66)	(4.08)
D×ZMDP	-4.97*	-2.66	-1.61*	2.97	3.66
	(2.55)	(2.86)	(0.85)	(8.14)	(4.42)
Wild bootstrap p-value	[0.241]	[0.081]	[0.241]	[0.031]	[0.002]
Obs. (Delivery Admissions)	332,286	332,286	332,286	332,286	332,286
R-squared	0.039	0.039	0.014	0.034	0.013
Pretreat Mean T	51.68	83.58	16.38	3,229.32	721.63
Magnitude	-0.086	-0.073	-0.082	0.005	0.017
Panel D: Maternal Health					
	Surgical/Anesthesia Complications	Genital Tract Laceration	Retained Placenta	ARDS	Hypertensive Disorders
D (expressed per 1,000)	-0.25*	-3.30	-7.45**	-0.14**	-8.09*
	(0.13)	(4.70)	(3.52)	(0.07)	(4.25)
D×ZMDP	0.12	-6.87	-1.51	0.00	-1.35
	(0.13)	(6.80)	(2.58)	(0.06)	(2.66)
Wild bootstrap p-value	[0.046]	[0.585]	[0.054]	[0.026]	[0.063]
Obs. (Delivery Admissions)	332,286	332,286	332,286	332,286	332,286
R-squared	0.002	0.218	0.034	0.000	0.030
Pretreat Mean T	0.02	63.83	22.69	0.13	64.20
Magnitude	—	-0.052	-0.329	—	-0.126
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: As a robustness check, this table augments the hospital-level and health specifications of Tables 1 and 2 with a control for the Zero-Markup Drug Policy (ZMDP). Panels A (aggregate revenue, RMB) and B (number of admissions) are estimated at the hospital-month level, for all OBGYN patients (Full) and by care type; Panels C (infant health) and D (maternal health) are estimated at the delivery-admission level, with health outcomes expressed per 1,000 deliveries and birth weight in grams. In Panels A and B, columns (1)-(5) report the Full sample and the pregnancy-loss, prenatal, delivery, and postnatal categories. In Panel C, columns (1)-(5) report NICU admission, low birth weight (LBW), very low birth weight (VLBW), birth weight, and birth weight concentration; in Panel D, they report surgical or anesthesia complications, genital tract laceration, retained placenta, ARDS, and hypertensive disorders. 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- or provincial-level hospitals and $\mathbb{1}(\text{PostDec2016})$ is an indicator for months after December 2016. Row D reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); it is not reported (—) for surgical or anesthesia complications and for ARDS, whose near-zero baselines render the proportional change uninformative. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (87 hospitals, of which 25 are inspected). Wild bootstrap p-values, computed with Webb weights and 9,999 replications, are reported in brackets in Panels C and D for the coefficient on D to address the small number of treated clusters. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Effects on Hospital-Level and Health Outcomes, Including the Spillover Control Group

	(1)	(2)	(3)	(4)	(5)
Panel A: Aggregate Revenue (RMB)					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	358,982***	2,127	64,862**	203,813***	25,807**
	(112,202)	(6,532)	(25,134)	(74,401)	(11,004)
Obs. (Hospital-Month)	3,354	3,354	3,354	3,354	3,354
R-squared	0.948	0.947	0.722	0.903	0.714
Pretreat Mean T	2,246,162	207,851	86,057	963,796	31,964
Magnitude	0.160	0.010	0.754	0.211	0.807
Panel B: Number of Admissions					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	45.00***	0.10	18.69***	11.38	7.56***
	(13.18)	(0.85)	(5.30)	(9.24)	(2.20)
Obs. (Hospital-Month)	3,354	3,354	3,354	3,354	3,354
R-squared	0.953	0.962	0.747	0.912	0.761
Pretreat Mean T	346.83	42.54	25.79	173.63	6.69
Magnitude	0.130	0.002	0.725	0.066	1.130
Panel C: Infant Health					
	NICU	LBW	VLBW	Birth Weight	Birth Weight Concentration
D (per 1,000; col. 4 in g)	-7.34*	-8.07*	-2.47**	17.79***	14.04***
	(3.75)	(4.15)	(0.97)	(6.44)	(4.64)
Wild bootstrap p-value	[0.051]	[0.032]	[0.025]	[0.024]	[0.001]
Obs. (Delivery Admissions)	406,431	406,431	406,431	406,431	406,431
R-squared	0.037	0.036	0.013	0.030	0.012
Pretreat Mean T	51.68	83.58	16.38	3,229.32	720.33
Magnitude	-0.142	-0.097	-0.151	0.006	0.019
Panel D: Maternal Health					
	Surgical/Anesthesia Complications	Genital Tract Laceration	Retained Placenta	ARDS	Hypertensive Disorders
D (expressed per 1,000)	-0.14*	-7.20***	-7.34*	-0.14**	-8.48**
	(0.08)	(2.37)	(4.31)	(0.05)	(4.21)
Wild bootstrap p-value	[0.062]	[0.012]	[0.144]	[0.011]	[0.047]
Obs. (Delivery Admissions)	406,431	406,431	406,431	406,431	406,431
R-squared	0.002	0.217	0.061	0.000	0.030
Pretreat Mean T	0.02	63.83	22.69	0.13	64.20
Magnitude	—	-0.113	-0.324	—	-0.132
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: As a robustness check, this table replicates the hospital-level and health specifications of Tables 1 and 2 on an expanded sample that adds the spillover control group. Panels A (aggregate revenue, RMB) and B (number of admissions) are estimated at the hospital-month level, for all OBGYN patients (Full) and by care type; Panels C (infant health) and D (maternal health) are estimated at the delivery-admission level, with health outcomes scaled per 1,000 deliveries and birth weight in grams. Variable definitions follow those tables. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); it is not reported (—) for surgical or anesthesia complications and for ARDS, whose near-zero baselines render the proportional change uninformative. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (25 treated clusters). Wild bootstrap p-values, computed with Webb weights and 9,999 replications, are reported in brackets in Panels C and D for the coefficient on *D* to address the small number of treated clusters. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Effects on Hospital-Level and Health Outcomes, Restricting the Control Group to Comparably Sized Hospitals

	(1)	(2)	(3)	(4)	(5)
Panel A: Aggregate Revenue (RMB)					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	269,237**	-388	73,216**	140,506**	19,136
	(100,532)	(7,252)	(27,611)	(62,548)	(12,217)
Obs. (Hospital-Month)	1,506	1,506	1,506	1,506	1,506
R-squared	0.957	0.950	0.730	0.924	0.760
Pretreat Mean T	2,246,162	207,851	86,057	963,796	31,964
Magnitude	0.120	-0.002	0.851	0.146	0.599
Panel B: Number of Admissions					
	Full	Pregnancy Loss	Prenatal	Delivery	Postnatal
D	29.20**	0.33	22.10***	-2.58	5.09**
	(12.12)	(1.16)	(6.62)	(8.91)	(2.43)
Obs. (Hospital-Month)	1,506	1,506	1,506	1,506	1,506
R-squared	0.951	0.950	0.766	0.918	0.842
Pretreat Mean T	346.83	42.54	25.79	173.63	6.69
Magnitude	0.084	0.008	0.857	-0.015	0.761
Panel C: Infant Health					
	NICU	LBW	VLBW	Birth Weight	Birth Weight Concentration
D (per 1,000; col. 4 in g)	-6.35	-7.85*	-1.86*	18.73**	15.42***
	(3.94)	(4.29)	(1.02)	(7.09)	(5.21)
Wild bootstrap p-value	[0.113]	[0.069]	[0.082]	[0.023]	[0.002]
Obs. (Delivery Admissions)	250,308	250,308	250,308	250,308	250,308
R-squared	0.038	0.038	0.012	0.033	0.014
Pretreat Mean T	51.68	83.58	16.38	3,229.32	721.63
Magnitude	-0.123	-0.094	-0.114	0.006	0.021
Panel D: Maternal Health					
	Surgical/Anesthesia Complications	Genital Tract Laceration	Retained Placenta	ARDS	Hypertensive Disorders
D (expressed per 1,000)	-0.31**	-5.97**	-8.48**	-0.15**	-7.02
	(0.14)	(2.88)	(3.79)	(0.06)	(4.58)
Wild bootstrap p-value	[0.024]	[0.053]	[0.028]	[0.006]	[0.143]
Obs. (Delivery Admissions)	250,308	250,308	250,308	250,308	250,308
R-squared	0.002	0.213	0.034	0.000	0.029
Pretreat Mean T	0.02	63.83	22.69	0.13	64.20
Magnitude	—	-0.094	-0.374	—	-0.109
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: As a robustness check, this table replicates the hospital-level and health specifications of Tables 1 and 2 on a restricted sample that retains only the control hospitals comparable in scale to the inspected ones. We rank the 87 hospitals of the estimation sample on a pre-treatment size index, the mean of a hospital's standardized values of twelve capacity, staffing, volume, and revenue measures recorded in the 2014 Health Statistics Yearbook (the components are listed in Appendix D), and retain the hospitals in the top two quartiles. All 25 inspected hospitals fall in that range, so the restriction operates entirely on the control group, which falls from 62 hospitals to 18, one of which is tertiary. The estimation sample is therefore 43 hospitals, and because the treatment group is unchanged, every pre-treatment mean is identical to its counterpart in Tables 1 and 2. Panels A (aggregate revenue, RMB) and B (number of admissions) are estimated at the hospital-month level, for all OBGYN patients (Full) and by care type; Panels C (infant health) and D (maternal health) are estimated at the delivery-admission level, with health outcomes scaled per 1,000 deliveries and birth weight in grams. Variable definitions follow those tables. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); it is not reported (—) for surgical or anesthesia complications and for ARDS, whose near-zero baselines render the proportional change uninformative. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (43 hospitals, of which 25 are inspected). Wild bootstrap p-values, computed with Webb weights and 9,999 replications, are reported in brackets in Panels C and D for the coefficient on *D* to address the small number of treated clusters. Data are drawn from provincial administrative OBGYN discharge records and the Health Statistics Yearbooks. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Drug-Service Substitution at the Admission and Physician Level

	(1)	(2)	(3)	(4)
Panel A: Components of Medical Revenue (Admission Level)				
	Drug	Service	Diagnosis	Consumable
D	-202.15*	167.71***	-13.33	22.79
	(106.53)	(45.79)	(43.38)	(31.13)
Obs. (Full Admissions)	585,639	585,639	585,639	585,639
R-squared	0.173	0.219	0.371	0.334
Hospital FE	Yes	Yes	Yes	Yes
Pretreat Mean T	1,299.54	2,573.20	1,625.53	978.04
Magnitude	-0.156	0.065	-0.008	0.023
Panel B: Share of Components of Medical Revenue (Admission Level)				
	Drug Share	Service Share	Diagnosis Share	Consumable Share
D	-0.03***	0.02**	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)
Obs. (Full Admissions)	585,602	585,602	585,602	585,602
R-squared	0.267	0.440	0.202	0.461
Hospital FE	Yes	Yes	Yes	Yes
Pretreat Mean T	0.19	0.41	0.25	0.15
Magnitude	-0.147	0.058	0.013	0.006
Panel C: Components of Medical Revenue (Physician Level)				
	Drug	Service	Diagnosis	Consumable
D	-187.03**	166.89***	8.52	16.12
	(85.73)	(47.90)	(49.52)	(38.05)
Obs. (Physician-Month)	15,807	15,807	15,807	15,807
R-squared	0.345	0.473	0.563	0.580
Physician FE	Yes	Yes	Yes	Yes
Pretreat Mean T	1,441.11	2,448.35	1,694.29	994.35
Magnitude	-0.130	0.068	0.005	0.016
Panel D: Share of Components of Medical Revenue (Physician Level)				
	Drug Share	Service Share	Diagnosis Share	Consumable Share
D	-0.02***	0.02**	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.00)
Obs. (Physician-Month)	15,807	15,807	15,807	15,807
R-squared	0.507	0.697	0.499	0.728
Physician FE	Yes	Yes	Yes	Yes
Pretreat Mean T	0.20	0.40	0.26	0.15
Magnitude	-0.124	0.056	0.020	-0.021
Time FE	Yes	Yes	Yes	Yes

Notes: This table investigates the policy effects on drug-service substitution at the admission and physician level. Panels A and C examine the impact on the distinct components of average revenue per admission. The dependent variables are the average revenue from drugs (col. 1), medical services (col. 2), diagnostics (col. 3), and consumables (col. 4). Panels B and D examine the impact on the corresponding component shares of medical revenue per admission. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T). All specifications include hospital/physician and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Robustness to Excluding the Most Common Surnames

	(1)	(2)	(3)	(4)	(5)
Surnames excluded	None	Wang	+ Zhang	+ Li	+ Liu
Panel A. Physician turnover (hospital-month)					
Exits $[-6, +6]$	0.12**	0.12**	0.12**	0.13**	0.14**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Entries $[-6, +6]$	0.11*	0.11*	0.13**	0.11*	0.12*
	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)
Exits $[-3, +3]$	0.16**	0.15**	0.16**	0.15**	0.16**
	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Entries $[-3, +3]$	0.19**	0.19**	0.21**	0.17*	0.19**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Obs. (six-/three-month window)	1,029/554	1,016/547	1,014/546	989/533	984/530
Panel B. Drug–service substitution, senior attending physicians					
Drug Share	-0.029***	-0.021***	-0.024***	-0.024**	-0.025***
	(0.011)	(0.008)	(0.008)	(0.009)	(0.010)
Service Share	0.029**	0.026*	0.025*	0.025*	0.027*
	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)
Obs. (Full Admissions)	337,634	302,552	273,037	250,410	236,303
Panel C. Drug–service substitution, junior attending physicians					
Drug Share	-0.022***	-0.020***	-0.025***	-0.024***	-0.025***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Service Share	0.007	0.003	0.010	0.009	0.010
	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
Obs. (Full Admissions)	228,129	198,475	170,078	151,413	138,293
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: This table re-estimates the two results that depend on identifying physicians from the discharge records on samples that progressively exclude the most common Chinese surnames. Column (1) is the full sample and reproduces Panel A of Table 6 and columns (3), (4), (7), and (8) of Panel A of Table 5. Columns (2) through (5) drop, cumulatively, admissions attended by a physician surnamed Wang; Wang or Zhang; Wang, Zhang, or Li; and Wang, Zhang, Li, or Liu. These are the four most common surnames in China, and because the probability that two physicians coincide is governed by the sum of squared surname shares, they account for the majority of the scope for coincidence.

Panel A reports the effect on the number of departing (Exits) and newly arriving (Entries) attending physicians, using six-month and three-month windows around each hospital's inspection. Panels B and C report the effect on the drug and service shares of revenue per admission, separately for admissions attended by senior and by junior physicians. Coefficients in Panel A are counts of physicians per hospital-month and are reported to two decimal places, as in Table 6; those in Panels B and C are shares and are reported to three decimal places, one more than in Table 5, so that stability across columns can be read directly.

The estimates are stable throughout. In Panel A the six-month exit effect if anything grows as the most ambiguous surnames are removed, which is the direction predicted by the fact that a single surname may stand for more than one physician, since the departures so concealed are concentrated in precisely the surnames being excluded. In Panels B and C the senior–junior asymmetry in the service margin is preserved in every column: the senior estimate is stable at 2.5 to 2.9 percentage points while the junior estimate never exceeds 1.0 and is never significant. The senior service-share estimate is significant at the five percent level in column (1) and at the ten percent level thereafter; its point estimate barely moves, and the loss of precision tracks the roughly thirty percent of admissions that the exclusions remove. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Heterogeneity in Health Outcomes by Advanced Maternal Age

	(1)	(2)	(3)	(4)	(5)
	NICU	LBW	VLBW	Birth Weight	Birth Weight Concentration
Panel A: Infant Health Among Advanced Maternal Age Women					
D (per 1,000; col. 4 in g)	-9.94	-19.54**	-4.81*	39.80**	29.48***
	(7.92)	(7.46)	(2.74)	(16.21)	(8.71)
Wild bootstrap p-value	[0.242]	[0.010]	[0.074]	[0.019]	[0.002]
Obs. (Delivery Admissions)	34,807	34,807	34,807	34,807	34,807
R-squared	0.044	0.039	0.016	0.037	0.015
Pretreat Mean T	63.41	100.62	22.02	3,235.48	682.45
Magnitude	-0.157	-0.194	-0.218	0.012	0.043
Panel B: Infant Health Among Non-Advanced Maternal Age Women					
D (per 1,000; col. 4 in g)	-6.46*	-5.79	-1.81*	14.64**	11.59**
	(3.55)	(3.99)	(0.91)	(6.34)	(5.35)
Wild bootstrap p-value	[0.087]	[0.184]	[0.070]	[0.067]	[0.027]
Obs. (Delivery Admissions)	297,479	297,479	297,479	297,479	297,479
R-squared	0.038	0.039	0.013	0.034	0.013
Pretreat Mean T	49.98	81.11	15.57	3,228.42	727.30
Magnitude	-0.129	-0.071	-0.116	0.005	0.016
	Surgical/Anesthesia Complications	Genital Tract Laceration	Retained Placenta	ARDS	Hypertensive Disorders
Panel C: Maternal Health Among Advanced Maternal Age Women					
D (per 1,000)	0.15	-0.51	-11.68**	-0.48*	-15.64**
	(0.21)	(3.39)	(5.24)	(0.28)	(6.50)
Wild bootstrap p-value	[0.598]	[0.890]	[0.054]	[0.105]	[0.019]
Obs. (Delivery Admissions)	34,807	34,807	34,807	34,807	34,807
R-squared	0.004	0.161	0.026	0.003	0.037
Pretreat Mean T	0.00	38.10	24.93	0.38	97.84
Magnitude	—	-0.013	-0.469	—	-0.160
Panel D: Maternal Health Among Non-Advanced Maternal Age Women					
D (per 1,000)	-0.22**	-6.96***	-7.63*	-0.09	-7.91*
	(0.10)	(2.43)	(3.90)	(0.05)	(4.38)
Wild bootstrap p-value	[0.029]	[0.023]	[0.072]	[0.098]	[0.079]
Obs. (Delivery Admissions)	297,479	297,479	297,479	297,479	297,479
R-squared	0.002	0.231	0.036	0.000	0.027
Pretreat Mean T	0.02	67.56	22.36	0.09	59.34
Magnitude	—	-0.103	-0.341	—	-0.133
Hospital FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated effects of hospital inspections on infant and maternal health outcomes among delivery admissions, split by advanced maternal age (AMA), defined as maternal age ≥ 35 at delivery. Panels A and B report infant health outcomes for AMA and non-AMA mothers, respectively; Panels C and D report maternal health outcomes for the same two groups. In Panels A and B, the dependent variables in columns (1)–(3) and (5) are binary indicators scaled per 1,000 deliveries—NICU (an inferred Neonatal Intensive Care Unit admission, identified from neonatal diagnosis codes), LBW (low birth weight < 2,500 g), VLBW (very low birth weight < 1,500 g), and Birth Weight Concentration (birth weight within one standard deviation of the sample mean)—while column (4) reports continuous birth weight in grams. In Panels C and D, all dependent variables are indicators scaled per 1,000 deliveries, capturing surgical or anesthesia complications, genital tract laceration (any obstetric laceration of the perineum, cervix, or upper vagina), retained placenta, ARDS (Acute Respiratory Distress Syndrome), and hypertensive disorders of pregnancy and eclampsia. Row *D* reports the DiD coefficients. Magnitude denotes the proportional change relative to the pre-treatment mean of the treatment group (Pretreat Mean T); it is not reported (—) for surgical or anesthesia complications and for ARDS, whose pre-treatment incidence is at or near zero in both age groups (see the Pretreat Mean T rows), which renders the proportional change uninformative. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (87 hospitals, of which 25 are inspected). Wild bootstrap p-values, computed with Webb weights and 9,999 replications, are reported in brackets for the coefficient on *D* to address the small number of treated clusters. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

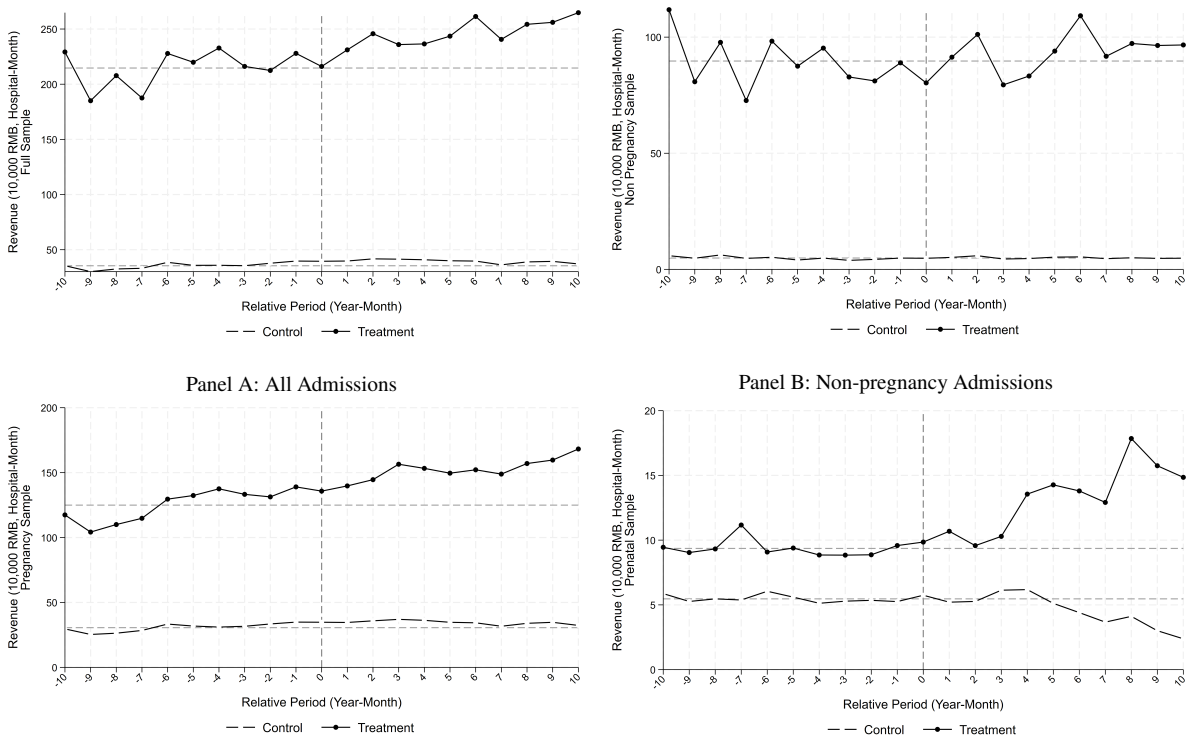
Table A11: Announcement Effects and Pre-Announcement Trends, First Inspection Cohort

	(1) Announcement effect	(2) Pre-period monthly trend
Panel A. Hospital-month outcomes		
Aggregate Revenue (RMB)	-65,914	-26,272***
	(42,003)	(8,165)
No. of Admissions	0.99	-2.68**
	(7.37)	(1.18)
Average Revenue per Admission (RMB)	-139.30	0.28
	(125.69)	(33.84)
Prenatal Revenue (RMB)	13,356	-3,171
	(13,742)	(3,758)
No. of Prenatal Admissions	3.81	-0.08
	(4.59)	(0.84)
Panel B. Delivery-admission outcomes (per 1,000; birth weight in g)		
NICU Admission	10.35***	1.07
	(3.56)	(1.56)
Low Birth Weight	5.89*	0.74
	(3.26)	(1.59)
Very Low Birth Weight	2.64	1.32**
	(1.96)	(0.62)
Birth Weight (g)	-5.28	-3.14
	(10.15)	(3.64)
Birth Weight Concentration	-5.99	-2.86
	(9.69)	(2.58)
Genital Tract Laceration	6.13	0.19
	(4.32)	(0.55)
Retained Placenta	-0.83	0.12
	(3.20)	(0.70)
Hypertensive Disorders	6.19	0.24
	(5.82)	(1.55)
No. of Hospitals (treated / control)	14 / 62	14 / 62

Notes: This table tests whether hospitals began responding when the inspection roster was published in April 2015, seven months before the first on-site inspections. The sample is the 14 first-cohort hospitals and the 62 clean control hospitals; the second cohort is excluded because its tenth pre-inspection month already falls after the first cohort was inspected. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level and reported in parentheses (76 hospitals, of which 14 are inspected).

Column (1) reports the announcement effect, defined as the difference between two blocks of treated-group months relative to controls: April–October 2015, when the roster was public but no on-site inspection had occurred, minus January–March 2015, the only months in the sample that precede publication. Column (2) reports the fitted monthly slope for treated hospitals over the full ten-month pre-inspection window ($k = -10$ to $k = -1$), which distinguishes a smooth differential drift from a level shift at the announcement. A drift loads on column (2) and leaves column (1) near zero; an announcement response does the reverse.

The exercise is deliberately under-powered: it identifies the announcement effect from three pre-announcement months in 14 hospitals, and reporting by a number of control hospitals is incomplete over those same months. An insignificant estimate in column (1) should therefore not be read as evidence that no announcement response occurred. The significant slopes for aggregate revenue and admissions in column (2) are levered by a single seasonal reversal: the extended event study places January 2015 at +357,410 ($t = 3.52$) and February 2015 at $-212{,}866$ ($t = -2.25$) for aggregate revenue, and the Chinese New Year fell in February 2015. Both months lie outside the six-month event window used elsewhere in the paper. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.



Panel A: All Admissions

Panel B: Non-pregnancy Admissions

Panel C: Pregnancy Admissions

Panel D: Prenatal Admissions

Figure A1: Raw Trends in Aggregate Revenue by Admission Type

Notes: This figure illustrates raw trends in aggregate medical revenue at the hospital-month level across different admission types. Panels A–D plot the trends for all OBGYN, non-pregnancy, pregnancy, and prenatal admissions, respectively. The vertical axis measures revenue in 10,000 RMB. Solid lines depict the trajectories for inspected hospitals (the treatment group), while dashed lines represent the clean control group. The horizontal dashed lines indicate the pre-inspection means for the treatment and control groups. The horizontal axis indicates the month relative to the inspection month ($k = 0$). 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 revenue from 10 months prior to 10 months following the inspection (or placebo) date. Data are drawn from provincial administrative OBGYN discharge records.

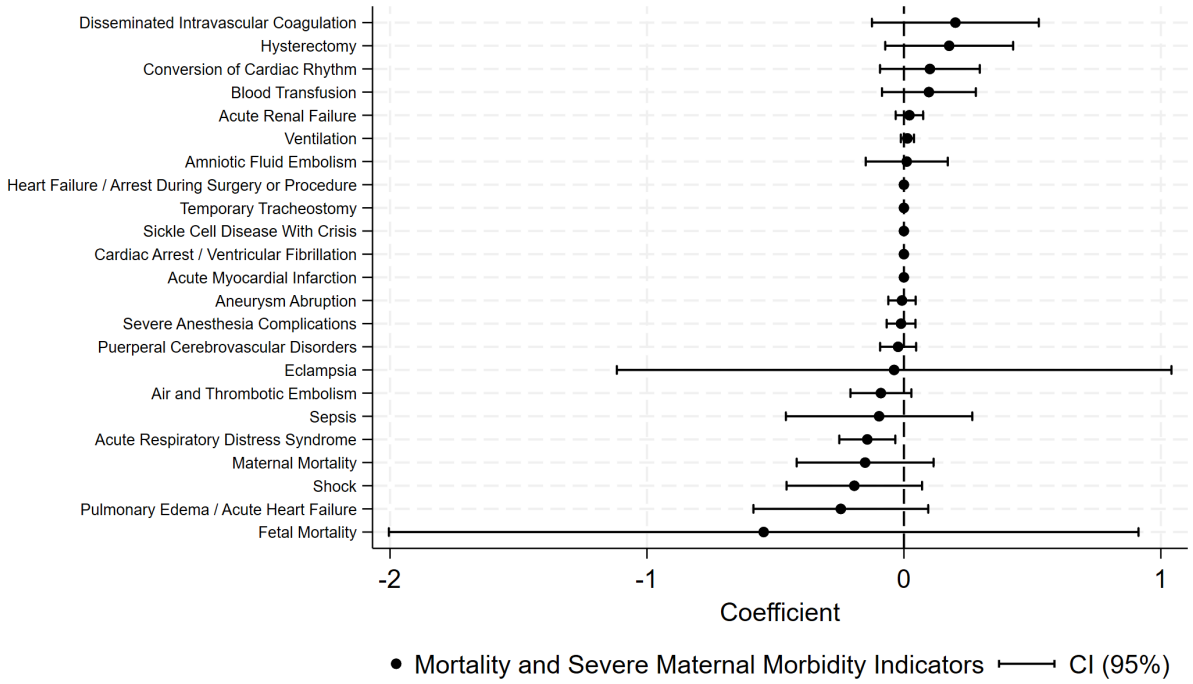
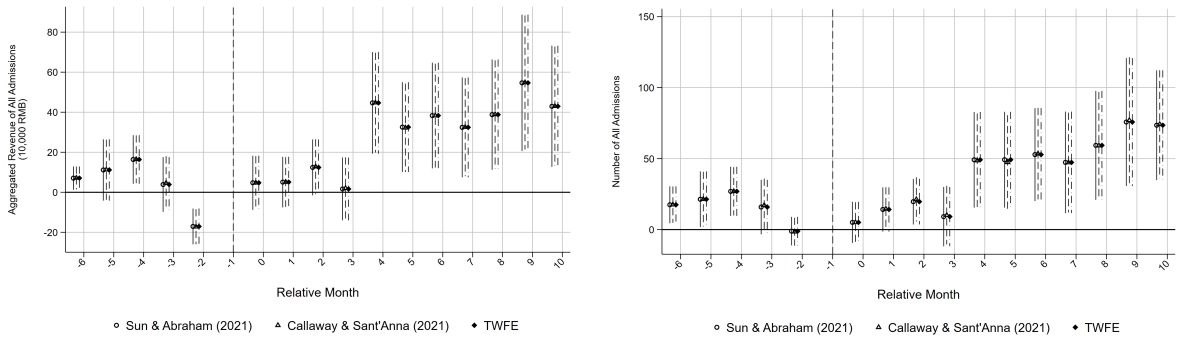


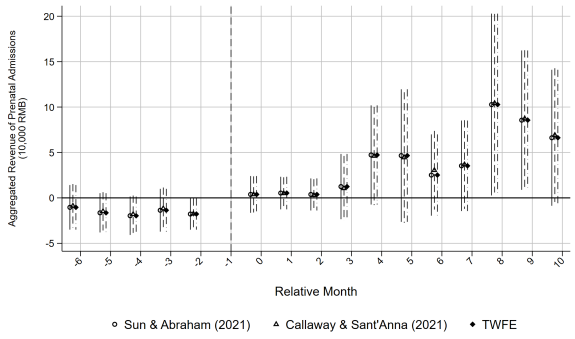
Figure A2: Effects on Maternal Mortality, Fetal Death, and Severe Maternal Morbidity

Notes: This figure plots the estimated regression coefficients capturing the effect of the anti-corruption inspections on the incidence of maternal death, fetal death, and 21 specific indicators of severe maternal morbidity (SMM) (Centers for Disease Control and Prevention, 2024), scaled per 1,000 births. In each separate regression, the dependent variable is a binary indicator that equals one if the patient was diagnosed with the corresponding condition listed on the vertical axis, and zero otherwise. Solid markers denote the point estimates of the treatment effect, and the horizontal error bars represent the 95% confidence intervals. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

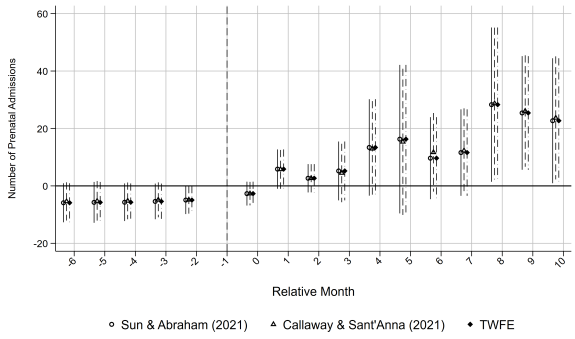


Panel A: Aggregate Revenue

Panel B: Number of Admissions



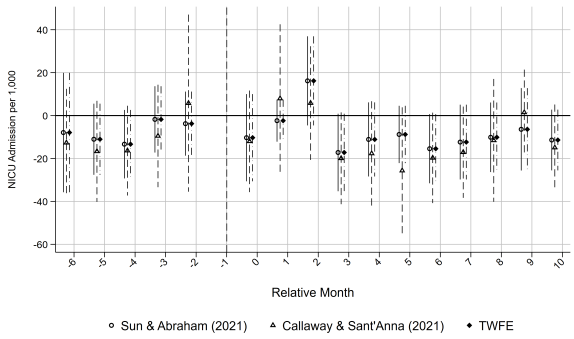
Panel C: Prenatal Revenue



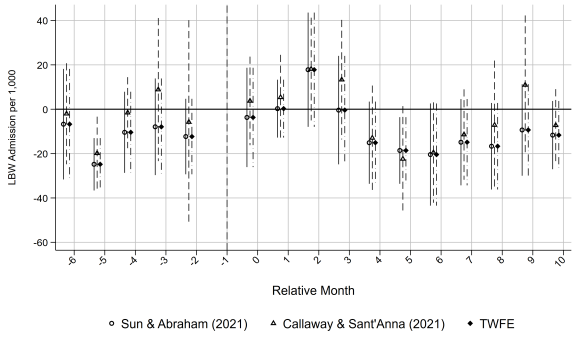
Panel D: Number of Prenatal Admissions

Figure A3: Event Study of Hospital-Level Outcomes, Excluding Hospitals Inspected after the Two-Child Policy

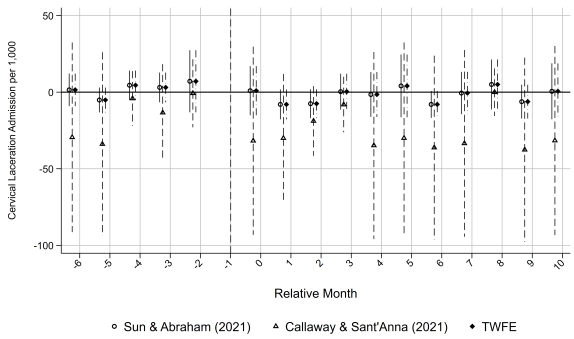
Notes: This figure plots event study estimates illustrating the dynamic impacts of hospital inspections on key hospital outcomes, specifically aggregate revenue (measured in 10,000 RMB) and admission volumes, after excluding treated hospitals inspected after the implementation of two-child policy. Panels A and B present the estimated effects on aggregate revenue and the total number of admissions for all OBGYN patients, respectively. Panels C and D report the corresponding effects on aggregate revenue and admission volumes for prenatal care. To ensure robustness against potential biases in standard difference-in-differences models with staggered timing, the figure 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 ($k = 0$), and vertical bars represent 95% confidence intervals. The event window spans from six months before to ten months after the inspection; observations outside this window are dropped, and $k = -1$ serves as the omitted reference period. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.



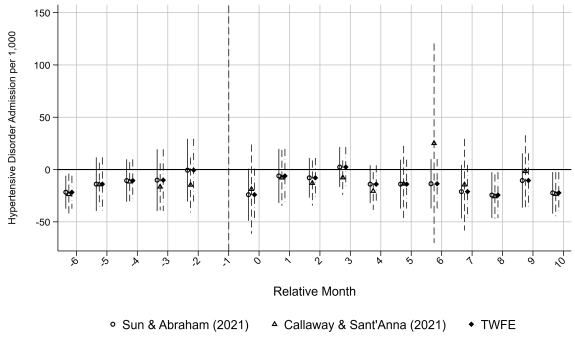
Panel A: Inferred NICU Admission



Panel B: Low Birth Weight



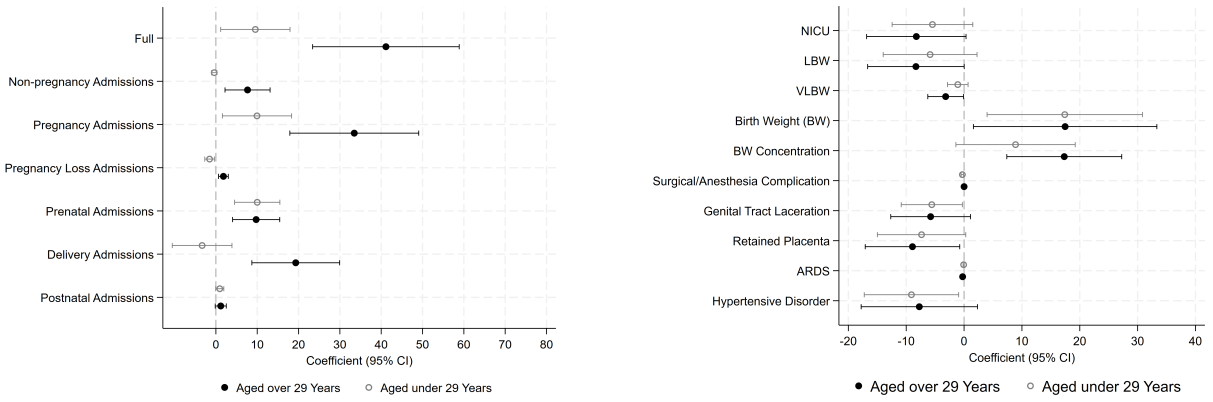
Panel C: Genital Tract Laceration



Panel D: Hypertensive Disorders

Figure A4: Event Study of Maternal and Infant Health Outcomes, Excluding Hospitals Inspected after the Two-Child Policy

Notes: This figure plots event study estimates illustrating the dynamic effects of hospital inspections on key maternal and infant health outcomes after excluding treated hospitals inspected after the implementation of two-child policy. Panels A through D present the estimated effects on the incidence of inferred infant NICU admissions, low-birth-weight cases, genital tract lacerations, and hypertensive disorders of pregnancy and eclampsia, respectively. To ensure robustness against potential biases in standard difference-in-differences models with staggered timing, the figure 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 ($k = 0$), and vertical bars represent 95% confidence intervals. The event window spans from six months before to ten months after the inspection; observations outside this window are dropped, and $k = -1$ serves as the omitted reference period. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

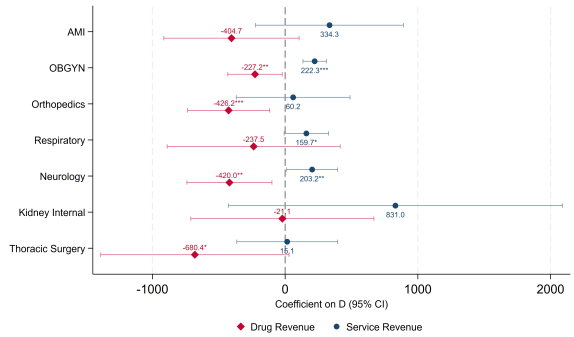


Panel A: Number of Admissions

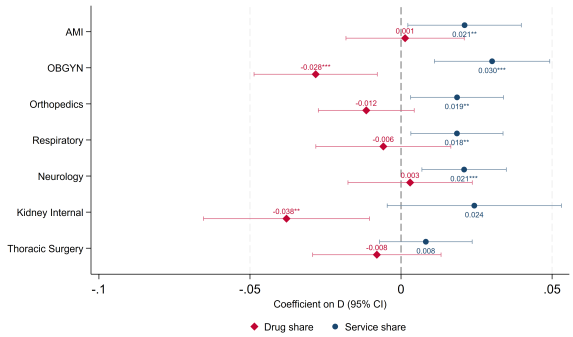
Panel B: Health Outcomes

Figure A5: Heterogeneous Effects of Hospital Inspections by Maternal Age

Notes: This figure illustrates the heterogeneous effects of hospital inspections, with estimates partitioned by maternal age (under 29 vs. 29 and older). Panel A presents the estimated coefficients from Equation (1) regarding admission volumes, while Panel B displays the results from Equation (2) for maternal and infant health outcomes. Filled circles denote point estimates for the older age group, and hollow circles denote point estimates for the younger age group. Horizontal lines represent 95% confidence intervals. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.



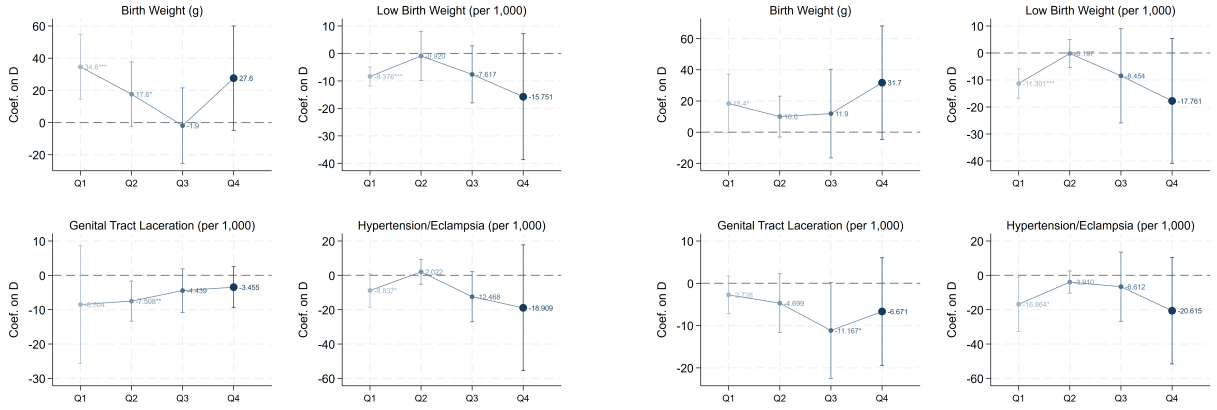
Panel A: Average Drug/Service Revenue per Admission



Panel B: Drug/Service Revenue Shares

Figure A6: Drug-Service Substitution across Hospital Departments

Notes: This figure plots the effects of hospital inspections on drug-service substitution across hospital departments without any sample restrictions. Panel A reports effects on average drug and service revenue per admission, and Panel B reports effects on drug and service revenue shares. Each pair of markers reports the DiD coefficients from separate regressions estimated on the corresponding department subsample, with navy circles for service outcomes and cranberry diamonds for drug outcomes, following the same convention as the quartile figures. Markers represent point estimates, labeled with the coefficient and significance stars, and horizontal lines indicate 95% confidence intervals. All specifications include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

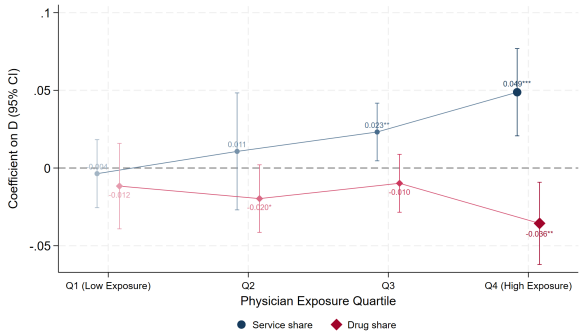


Panel A: By Physician-Level Intensity

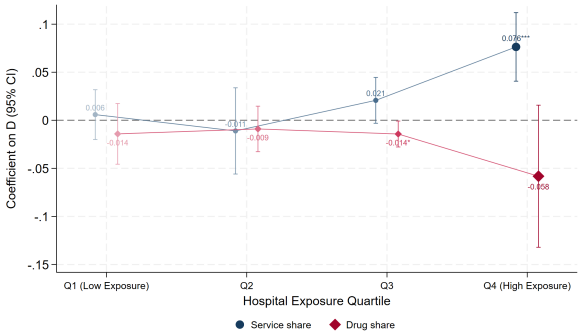
Panel B: By Hospital-Level Intensity

Figure A7: Heterogeneity in Health Outcomes by Baseline Drug Prescribing Intensity

Notes: This figure plots heterogeneity in the effects of hospital inspections on key maternal and infant health outcomes by physicians' and hospitals' baseline drug prescribing intensity. Panel A reports DiD coefficients estimated separately for each quartile of physician intensity, and Panel B for each quartile of hospital intensity. Exposure is measured by the physician (hospital) fixed effects recovered from a pre-inspection regression of drug revenue shares, so that higher quartiles correspond to physicians (hospitals) with systematically higher baseline drug shares. Markers represent point estimates, labeled with the coefficient and significance stars, and vertical lines indicate 95% confidence intervals. Specifications in Panel A include attending physician and time fixed effects; specifications in Panel B include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.



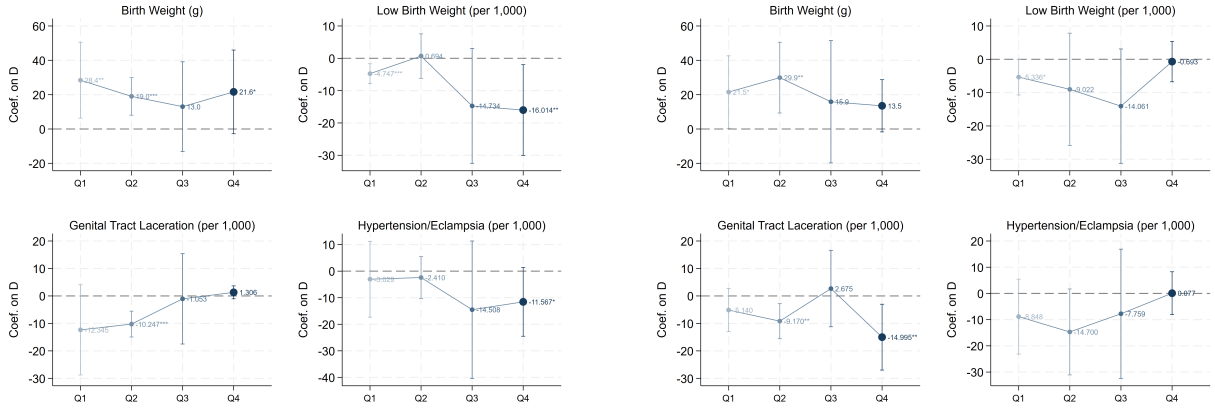
Panel A: By Physician-Level Intensity



Panel B: By Hospital-Level Intensity

Figure A8: Heterogeneity in Drug-Service Substitution by Baseline Antibiotic Prescribing Intensity

Notes: This figure plots heterogeneity in the effects of hospital inspections on drug-service substitution by physicians' and hospitals' baseline antibiotic prescribing intensity. Panel A reports DiD coefficients estimated separately for each quartile of physician prescribing intensity, and Panel B for each quartile of hospital intensity. Exposure is measured by the physician (hospital) fixed effects recovered from a pre-inspection regression of the antibiotic share of drug revenue, so that higher quartiles correspond to physicians (hospitals) with systematically higher baseline antibiotic shares. Navy circles report effects on service revenue shares, and cranberry diamonds report effects on drug revenue shares. Markers represent point estimates, labeled with the coefficient and significance stars, and vertical lines indicate 95% confidence intervals. Specifications in Panel A include attending physician and time fixed effects; specifications in Panel B include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.



Panel A: By Physician-Level Intensity

Panel B: By Hospital-Level Intensity

Figure A9: Heterogeneity in Health Outcomes by Baseline Antibiotic Prescribing Intensity

Notes: This figure plots heterogeneity in the effects of hospital inspections on key maternal and infant health outcomes by physicians' and hospitals' baseline antibiotic prescribing intensity. Panel A reports DiD coefficients estimated separately for each quartile of physician intensity, and Panel B for each quartile of hospital intensity. Exposure is measured by the physician (hospital) fixed effects recovered from a pre-inspection regression of the antibiotic share of drug revenue, so that higher quartiles correspond to physicians (hospitals) with systematically higher baseline antibiotic shares. Markers represent point estimates, labeled with the coefficient and significance stars, and vertical lines indicate 95% confidence intervals. Specifications in Panel A include attending physician and time fixed effects; specifications in Panel B include hospital and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.

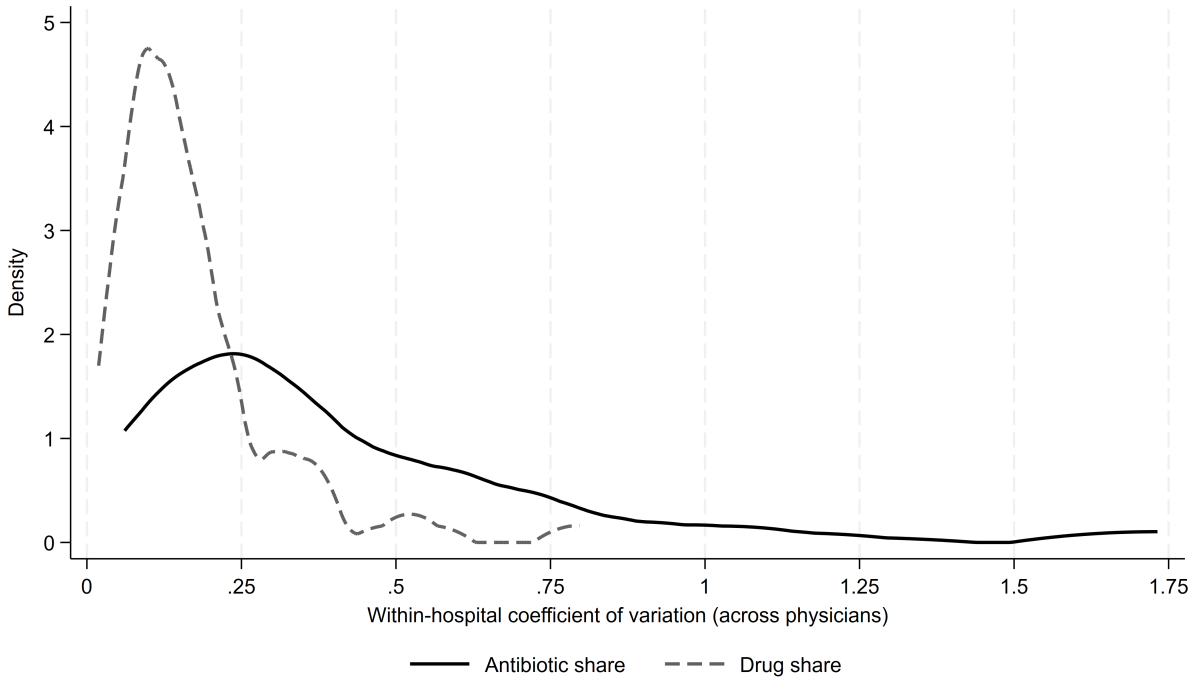
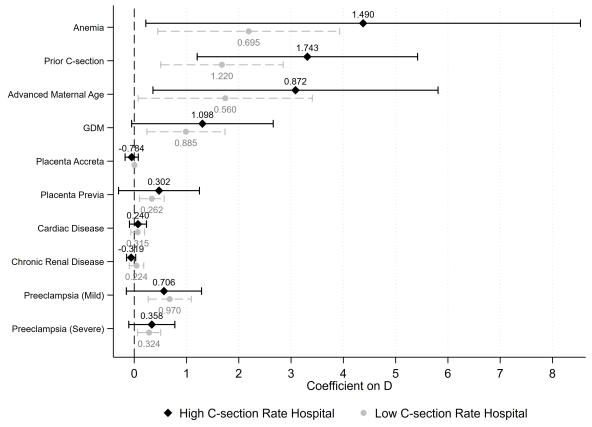
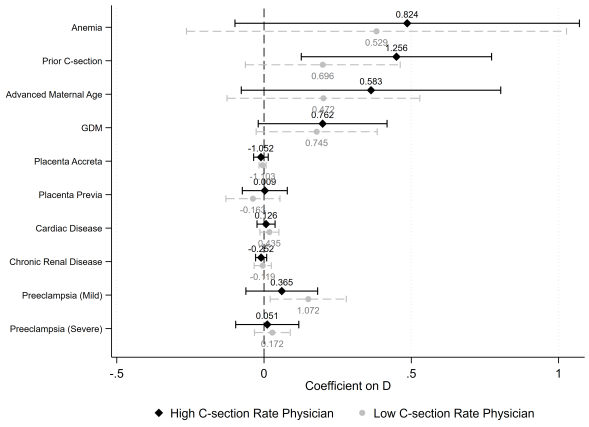


Figure A10: Within-Hospital Variation in Physician Practice Styles: Antibiotic vs. Overall Drug Share

Notes: This figure plots the distribution of within-hospital variation in physician pre-inspection practice styles across antibiotic prescribing (solid black) and overall drug prescribing (dashed dark gray). The unit of observation is a hospital. For each attending physician, we compute the mean antibiotic share of drug revenue and the mean drug share of total revenue across all admissions attributed to that physician. For each hospital, we then compute the coefficient of variation (CV)—the standard deviation divided by the mean—of each practice-style measure across its attending physicians; the CV rescales dispersion by the mean, making variation comparable across measures with different levels. Larger values indicate greater heterogeneity in practice styles among physicians within the same hospital, i.e., dimensions over which individual physicians retain more discretion. Wilcoxon signed-rank tests comparing paired within-hospital CVs reject equality between the antibiotic-share CV and the drug-share CV ($p = 0.00$). Data are drawn from provincial administrative OBGYN discharge records.



Panel A: By Hospital Baseline C-section Rate



Panel B: By Physician Baseline C-section Rate

Figure A11: Heterogeneity in Prenatal Care Provision for Obstetric Comorbidities by Baseline C-section Rate

Notes: This figure plots the heterogeneous effect on the number of prenatal admissions with obstetric comorbidities by hospitals' baseline C-section rate at the hospital-month level (Panel A) and by physicians' baseline C-section rate at the physician-month level (Panel B). The dependent variable in each row is a binary indicator equal to one if the admission carries a diagnosis of the comorbidity listed on the vertical axis, and zero otherwise. The comorbidities follow the clinical risk classification of Leonard et al. (2020) as predictors of severe maternal morbidity. Markers represent point estimates, labeled with the proportional change relative to the pre-treatment mean of the treatment group, and horizontal lines indicate 95% confidence intervals. All specifications include hospital/physician and time fixed effects. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records.

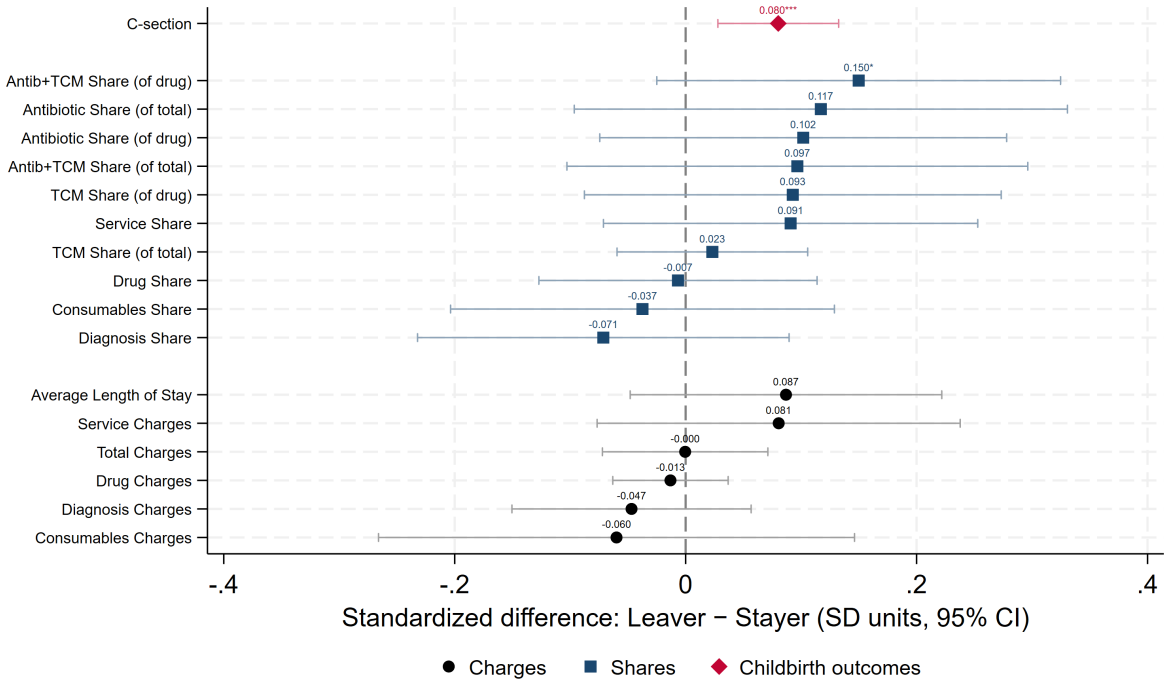
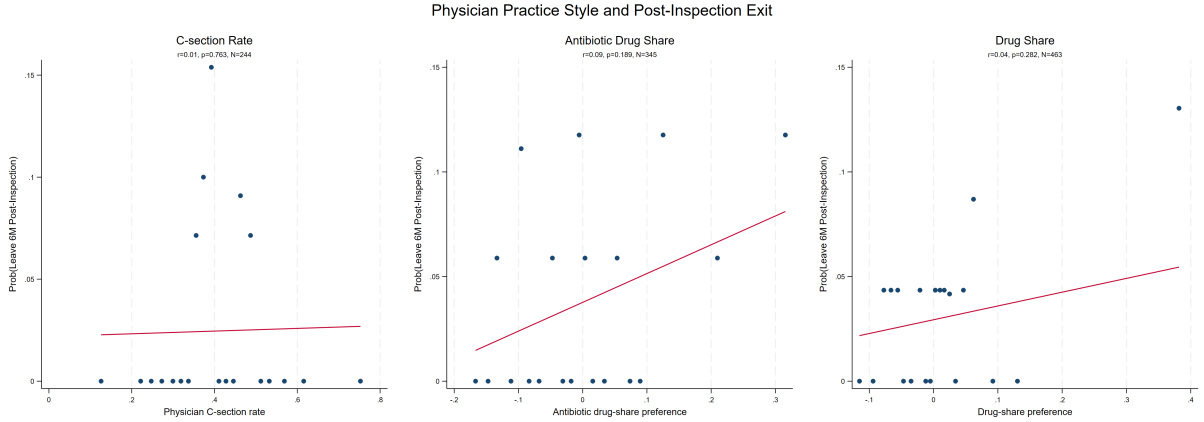


Figure A12: Differences in Baseline Practice Patterns between Leaver and Stayer Physicians

Notes: This figure plots differences in baseline practice patterns between leaver and stayer physicians. Leavers are attending physicians who exit the hospital within the six-month window after the inspection; stayers are those who remain. Each outcome is standardized to a z-score before estimation, so coefficients are in standard deviation units. Black circles report differences in revenue outcomes, navy squares report differences in revenue shares, and the cranberry diamond reports the difference in the C-section rate. Markers represent point estimates, labeled with the coefficient and significance stars, and horizontal lines indicate 95% confidence intervals. All specifications include hospital, admission-month, and case-mix fixed effects, where case mix comprises age group, pregnancy status, prior C-section, multiple pregnancy, chronic hypertension, and GDM. Standard errors are clustered at the hospital level. Data are drawn from provincial administrative OBGYN discharge records. *** p<0.01, ** p<0.05, * p<0.1.



Panel A: Hospital-Level Correlation



Panel B: Physician-Level Correlation

Figure A13: Correlation between Baseline Corruption Margins and Physician Turnover

Notes: This figure plots correlations between baseline corruption-margin proxies and post-inspection physician turnover at the hospital level (Panel A) and the physician level (Panel B). In both panels, the practice-style measures on the x-axes—the baseline C-section rate, the antibiotic drug-share preference, and the drug-share preference—proxy for the scope for rent-seeking, so that higher values indicate hospitals or physicians with more room for corruption-related prescribing and treatment. In Panel A, each dot is a hospital and the line is an OLS fit; untreated hospitals are excluded. The reported β is the slope from regressing the number of physicians leaving within six months post-inspection on the hospital practice-style measure (x-axis); p is the two-sided p-value for $H_0 : \beta = 0$ based on heteroskedasticity-robust standard errors; and N is the number of hospitals. In Panel B, each dot is a bin mean from a 20-bin binned scatter, and the line is an OLS fit on the underlying physician-level data. The outcome is an indicator for leaving within six months post-inspection; r is the Pearson correlation, p is the two-sided robust test of a zero slope, and N is the number of physicians. The fitted lines are plotted without shaded confidence bands; the reported p-values, not the visual fit, are the basis for inference here. Across both panels, physicians and hospitals with higher baseline C-section rates, antibiotic shares, and drug shares tend to exhibit greater post-inspection turnover, though the associations are not significant at conventional levels. Data are drawn from provincial administrative OBGYN discharge records.

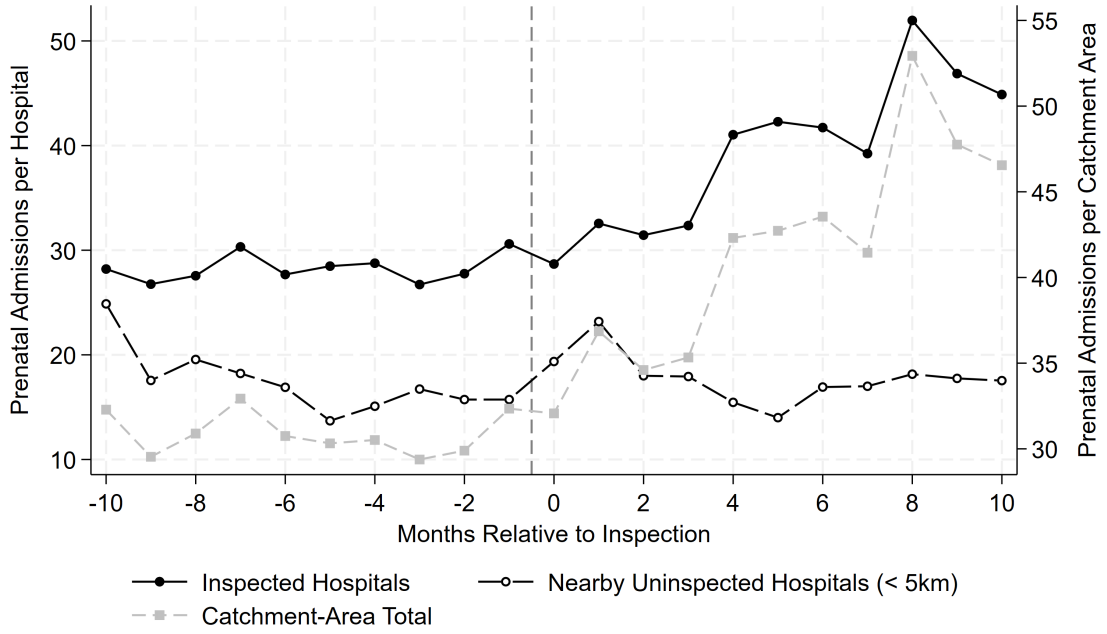


Figure A14: Market-Level Prenatal Admissions around Inspection

Notes: This figure plots monthly prenatal admissions in event time around the inspection for inspected hospitals, for uninspected hospitals located within five kilometers of an inspected hospital, and for the catchment area as a whole. Solid circles report the mean number of prenatal admissions per inspected hospital (left axis). Hollow circles report the mean per nearby uninspected hospital (left axis); each nearby hospital is assigned the inspection month of the inspected hospital within five kilometers of it. Gray squares report the catchment-area total (right axis): prenatal admissions are summed within each catchment area, defined as an inspected hospital together with the uninspected hospitals within five kilometers, and averaged across catchment areas at each event month. The horizontal axis is the month relative to the inspection ($k = 0$), and the dashed vertical line marks the inspection. A regression of nearby hospitals' monthly prenatal admissions on a post-inspection indicator, with hospital and calendar-month fixed effects and standard errors clustered at the hospital level, yields a coefficient of 0.38 (s.e. 2.20) on a pre-inspection mean of roughly 17 admissions per hospital-month. Data are drawn from provincial administrative OBGYN discharge records.

B Additional Institutional Background

This appendix expands on the institutional context summarized in Section 2.

The 2009 reform and its aftermath. The 2009 healthcare reform sought to correct distorted provider incentives chiefly through the Zero-Markup Drug Policy (ZMDP) and the National Essential Drugs List (NEDL) rolled out over the 2010s. By abolishing the pharmaceutical markup, capping drug revenue shares, and standardizing reimbursement lists, policymakers tried to sever the link between provider pay and drug sales. These measures did lower drug prices and weaken the "drug-financed" hospital model, but the underlying structural imbalance persisted: because base salaries stayed low relative to physicians' training and workload, physicians remained in a "high-intensity, low-reward" equilibrium with strong incentives to seek supplementary income through other channels. Pushing rent extraction underground raised the agency costs of corruption and left its net effect on care quality and welfare ambiguous.

The national anti-corruption campaign. In November 2012, at the 18th National Congress of the Chinese Communist Party, the leadership framed official corruption as a fundamental threat to regime legitimacy, prompting the December 2012 promulgation of the "Eight-Point Regulation." By targeting both high-ranking officials and grassroots cadres, the initiative aimed to disrupt the rent-seeking mechanisms entrenched during decades of decentralized economic growth, and enforcement expanded beyond traditional party organs to state-owned enterprises, the military, and public service sectors. Unlike previous episodic crackdowns, this sustained enforcement altered political incentives and bureaucratic norms.

History of the Hospital Inspection program. The program has undergone five rounds. The Ministry of Health centrally administered the first two (2005 and 2010). The third round, launched in 2015, introduced the dual-level oversight structure described in the main text and institutionalized the inspections as a permanent regulatory mechanism, a structure maintained in subsequent rounds (2019–2022 and 2023–2026). Early rounds imposed weak sanctions: in 2005, noncompliant hospitals received only rectification recommendations, and by 2010 repercussions remained confined to internal warnings, with public reprimands or dismissals rare. By contrast, in the 2015 round hospitals with systemic failures faced coordinated cross-agency penalties, such as targeted reimbursement audits, in addition to the individual-level sanctions described in the main text.

The three-stage protocol. Procedurally, the 2015 round standardized a three-stage protocol. First, hospitals conducted internal self-assessments and submitted corrective plans. Second, multidisciplinary inspection teams (government officials, medical experts, and Party disciplinarians) performed on-site evaluations via document reviews, interviews, and random audits. Third, authorities publicly released the results and mandated verified rectification measures. The protocol

emphasized three core areas: strengthening Party Committee oversight over hospital leadership, strictly enforcing the "Nine Prohibitions" against kickbacks and supplier-induced demand, and auditing budgetary and procurement compliance. Following the initial on-site audit, the inspection team issues formal feedback, prompting the hospital to submit and execute a detailed rectification plan; this administrative cycle typically spans one quarter and concludes with mandatory follow-up re-inspections, which matches the three-to-four-month lag with which the treatment effects emerge in Section 4.

C Testing for Announcement Effects

The provincial implementation plan named all 44 hospitals and fixed each hospital's inspection year in April 2015, so the roster was public for all but the first three months of our sample. The pre-inspection months of a treated hospital are therefore months in which it already knew its inspection year, and any differential pre-inspection trend may reflect anticipation rather than a violation of parallel trends in the usual sense. The publication date makes one direct test possible, which this appendix reports.

January through March 2015 are the only months in our sample that precede the roster's publication, and for the first inspection cohort they sit at $k = -10$ through $k = -8$, outside the event window plotted in the main text. Restricting attention to the 14 first-cohort hospitals and the 62 never-inspected control hospitals, we compare these three pre-announcement months with the seven announced-but-not-yet-inspected months from April to October 2015, each measured relative to the control group. The difference between the two blocks is the announcement effect, and column (1) of Table A11 reports it for each of our thirteen outcomes. Eleven are indistinguishable from zero, including all five hospital-level outcomes: aggregate OBGYN revenue (−65,914, s.e. 42,003), total admissions (0.99, s.e. 7.37), prenatal revenue (13,356, s.e. 13,742), prenatal admissions (3.81, s.e. 4.59), and average revenue per Admission(139.30, s.e. 125.69). Two exceptions, NICU and low-birth-weight admissions, carry the opposite sign to its post-inspection effect, and across thirteen tests a t-statistic of this size is close to the largest one would expect under the null.

Column (2) takes the complementary approach and fits a linear monthly trend for treated hospitals across the whole ten-month pre-period, which separates a smooth differential drift from a level response at the announcement. The variables that carry our mechanism are flat: neither prenatal revenue ($t = -0.84$) nor prenatal admissions ($t = -0.09$) exhibits a pre-inspection slope, and among the eight health outcomes only very low birth weight attains conventional significance. Aggregate revenue (−26,272 per month, $t = -3.22$) and total admissions (−2.68 per month, $t = -2.27$) do slope downward. That movement is not smooth, however. In the extended event study the two adjacent months of January and February 2015 swing from +357,410 ($t = 3.52$) to −212,866 ($t = -2.25$), and the Chinese New Year fell in February 2015; much of the fitted slope

is levered by this single seasonal reversal. We therefore read the extended pre-window as unstable for the aggregate series rather than as evidence of a differential trend, and note that the six-month window used in our event studies excludes both months.

Two limitations bound what this exercise establishes. It rests on three pre-announcement months and 14 treated hospitals, so it is under-powered, and an announcement effect that is statistically insignificant here is not the same as an announcement effect of zero. Reporting by a number of control hospitals is also incomplete over precisely these months, so they contribute unevenly to the comparison. We treat the exercise accordingly, as a bounded check rather than a decisive test: it finds no evidence that treated hospitals began to respond when the roster was published, and it establishes that the outcomes underpinning our mechanism carry no pre-inspection drift.

D Robustness of the Main Estimates

This appendix reports three checks on the main estimates summarized in Section 4: two concurrent policies that could confound them, and the comparability of the control group.

The Zero-Markup Drug Policy. The implementation of the ZMDP partially overlapped with our study period and is known to alter physician incentives, encouraging the substitution of prescribed drugs with medical services (Fang et al., 2021).[30] To separate the two policies, we augment Equations (1) and (2) with a ZMDP exposure indicator,

$$\text{ZMDP}_{ht} = \mathbb{1}(\text{municipal or provincial hospital}) \times \mathbb{1}(t > \text{December 2016}),$$

which marks the hospital-months covered by the second wave of the rollout, together with its interaction with D_{ht}; the reported D_{ht} coefficient is therefore the inspection effect for hospital-months outside that wave. Table A5 shows that after controlling for ZMDP rollout, the extensive-margin responses (aggregate revenue and admission volumes) remain robust, and every maternal and infant health effect retains its health-improving sign, with the core improvements (birth weight, low birth weight, and hypertensive disorders) remaining significant.

The Two-Child Policy. The 2016 TCP might increase maternal healthcare utilization and change the composition of women giving birth.[31] We conduct two robustness checks. First, we restrict the treatment group to hospitals inspected in November 2015, prior to TCP implementation. Figures A3 and A4 present the event studies: aggregate revenue and admission volumes again rise persistently, with the increase emerging around the fourth month after inspection, consistent

[30] In the studied province, the ZMDP was initially piloted in rural county-level hospitals in 2012 and fully implemented in these hospitals by July 2015. The policy was subsequently expanded to municipal and provincial hospitals beginning in December 2016, achieving a province-wide elimination of drug markups across all hospital tiers by July 2017.

[31] Announced nationwide in January 2016, China's Universal TCP formally ended the decades-long one-child policy by allowing all couples to legally have up to two children.

with the rectification and re-inspection cycle, and the health results closely mirror the baseline findings. Gestational timing makes a TCP explanation implausible over this window. The TCP took effect two months after the first-round inspections and can raise obstetric volume only through new conceptions, so a pregnancy begun in January 2016 would be at most eight to ten weeks along when revenue and admissions rise in March 2016, far too early to generate the second- and third-trimester inpatient prenatal admissions that account for the increase; such a pregnancy would not reach delivery until late 2016, beyond the end of our event window. The infant-health improvement likewise emerges as early as March 2016, only three months after the announcement, too soon to be a TCP effect, since these infants would have been conceived months before the policy took effect. Second, to account for the "legalization" of second pregnancies, we stratify the sample into mothers aged $\geq$29 and $<$29 years, on the assumption that women under 29 are predominantly experiencing their first pregnancy and therefore less likely to be affected by the TCP.[32] Figure A5 shows that both age groups experienced an increase in prenatal care provision, with the effect if anything larger among the younger cohort, the opposite of what a TCP-driven story would predict, and consistent health gains among mothers under 29.

Control-group comparability. Because inspection eligibility was confined to the province's major hospitals, treated hospitals sit at the top of the size distribution while the clean controls, defined by district, do not. Parallel pre-inspection trends are the identifying assumption, but level gaps of this magnitude leave room for a province-wide shock whose incidence scales with hospital size to load onto the treatment indicator: the second wave of the ZMDP rollout, shifts in referral patterns, and any differential incidence of the TCP at larger facilities all have this character. To restrict the comparison to hospitals of comparable scale, we rank the 87 hospitals of the estimation sample on an index of pre-inspection size, defined as the mean of a hospital's standardized values of twelve measures recorded in the 2014 Health Statistics Yearbook: staffed beds, total bed-days available, total bed-days occupied, discharges, total visits, outpatient visits, total staff, licensed physicians, registered nurses, and medical, outpatient, and inpatient revenue. All twelve predate every inspection, so the restriction cannot be a function of treatment; because three of them are revenue measures, it also narrows the baseline gap in the level of the dependent variable itself. Retaining the top two quartiles of the index leaves all 25 inspected hospitals but only 18 of the 62 clean controls, so the exercise imposes common support rather than matching, and it does so at the cost of roughly half the sample: 1,506 hospital-months in place of 2,947, and 250,308 delivery admissions in place of 332,286.

Table A7 reports the result, which separates the finding into a part the restriction leaves intact and a part it does not. The outcomes carrying our mechanism are, if anything, stronger among

[32] According to Zhou et al. (2024), the average age at first childbirth in China from 2016 to 2020 was 28.8 ± 5.1 years.

comparable hospitals: prenatal revenue rises by 73,216 RMB and prenatal admissions by 22.10, or 85 and 86 percent of their pre-treatment means against 79 and 77 percent in the baseline, and the infant-health gains that follow them survive, with birth weight up 18.73 grams and the birth-weight concentration measure up 15.42 per 1,000. The aggregate magnitudes shrink: aggregate revenue falls from 401,940 to 269,237 RMB and total admissions from 50.69 to 29.20; the delivery-admission coefficient turns slightly negative and insignificant, though delivery revenue remains significant at a reduced 140,506 RMB; postnatal revenue loses significance; and NICU admission and hypertensive disorders weaken to wild bootstrap p-values of 0.113 and 0.143. Discarding half the sample widens every standard error, so power accounts for part of this, but the point estimates on the aggregates move as well. We therefore read the exercise as locating the result rather than overturning it: what survives a comparison confined to the province's larger hospitals is the compositional shift toward prenatal care and the infant-health improvement that follows it, which is the mechanism the paper advances, while aggregate revenue and delivery volume, the components most exposed to a size-scaling confounder, are estimated less robustly. One limitation deserves stating plainly. The restriction closes the size gap but not the tier boundary: eligibility was confined to tertiary hospitals and just one of the 18 remaining controls is itself tertiary, so the comparison still runs across the margin the program selected on. Tier is also the dimension along which the confounders named above actually travel, the ZMDP exposure indicator being defined as municipal or provincial status interacted with the months after December 2016.

E Measuring Physicians through the Surname–Title Linkage

The discharge records carry no unique physician identifier: providers can be tracked only through the surname and professional title of the attending and supervising physicians recorded on each hospitalization. This appendix collects the measurement properties of the two variables built on this linkage (the seniority classification and the turnover measure used in Section 5) and the robustness checks that address them.

Seniority. An attending physician is classified as "senior" if their surname matches that of a Chief Physician or Department Head recorded on the same hospitalization, and "junior" otherwise. Because the linkage is made on names, it can err in only one direction: a senior attending who happens to share a surname with the resident physician on the same record is recorded as junior, while the converse cannot occur, since an attending who is the resident necessarily shares that surname. The junior group is therefore mildly contaminated with senior physicians while the senior group is clean, and the effect is to *attenuate* any senior–junior contrast: coincidence cannot manufacture the asymmetry documented in Table 5, only dilute it. The rate at which it occurs is the probability that two physicians in the same department share a surname, which does not vary systematically with department size and is concentrated in the commonest surnames.

Table A9 accordingly re-estimates the seniority results on samples that progressively exclude the four most common Chinese surnames, removing most of the scope for coincidence; the senior–junior asymmetry in the service margin is unchanged in every case.

Turnover. The turnover measure tracks surnames rather than individuals, and it is conservative in one direction and permissive in the other. Two attendings who share a surname within the same department are indistinguishable, so a departure registers only when the last physician of that surname stops appearing; measured exits and entries are therefore undercounts, and the undercount grows with the number of attendings in a department. In the opposite direction, a physician who ceases to be listed as the attending of record (through promotion, reassignment, or prolonged leave) registers as an exit while remaining on staff. Short rotations do not, however, generate exits: we require a name to be absent for the remainder of the three-year panel rather than merely for a few months, so an exit recorded around the first-round inspection means the name does not reappear for roughly two years. We accordingly interpret the turnover estimates as evidence that inspections coincided with a discrete change in the composition of the attending roster, and do not interpret the level of the turnover rate, or its comparison across hospitals of different size, as a count of physicians who left.

Because what we observe are surname appearances, the concern is sharpest for the commonest surnames, where a single recorded name is most likely to conceal more than one physician. Panel A of Table A9 therefore reproduces the turnover estimates on samples that cumulatively exclude the four most common Chinese surnames (Wang, Zhang, Li, and Liu, together roughly a quarter of the population). The estimates are unchanged: across the four exclusions, exits within the six-month window rise by 0.12 to 0.14 and entries by 0.11 to 0.13; the three-month estimates are likewise stable at 0.15 to 0.16 and 0.17 to 0.21, and every coefficient remains significant at conventional levels. If anything the proportional response grows as the most ambiguous surnames are removed: the six-month exit effect rises from 1.35 to 2.09 times its pre-treatment mean, which is the direction surname collapse predicts, since the departures it conceals are concentrated in exactly the names being dropped. The exercise speaks to this source of error rather than to the residual cases in which a physician stops attending without leaving the hospital. Relatedly, the exit/no-exit stratification in Table 6 inherits these properties: a hospital recorded as having no exit is one in which no surname disappeared, which is not the same as one in which no physician left, and the probability that a departure is masked rises with department size, so the split is partly a split on department size and the two groups differ substantially at baseline.

F Additional Detail on Alternative Explanations

Interpreting the comorbidity coefficients. The reference class matters for the prenatal-admission results in Figure 6 and differs from the one used in the mechanism analysis: relative to an uncom-

plicated pregnancy, the added prenatal admissions are higher-risk, which is what argues against revenue-driven over-provision, while relative to the structurally high-risk conditions of Figure 3 they are of intermediate risk, which is what makes them discretionary enough for physicians to expand. Both statements describe the same coefficients. Comparing the delivery and prenatal samples is itself revealing. For *modifiable* conditions, the coefficients diverge across the two stages: the prevalence of anemia, for instance, rises among prenatal admissions but falls at delivery, consistent with earlier detection and management lowering its prevalence by the time of birth. For *non-modifiable* demographic factors such as advanced maternal age and prior C-section, by contrast, the increase appears in both samples alike, reflecting a genuine shift in the patient population rather than differential management. This divergence is the signature of prenatal care that treats rather than merely records risk.

The information-system channel. Our prior was that hospitals with weaker information systems, which were harder for regulators to monitor beforehand and thus had more scope for undetected rent-seeking, would respond more strongly to the inspection. The data do not bear this out: the effects on the core infant-health outcomes (NICU admission, low and very low birth weight, and birth weight) are indistinguishable across hospitals with good and poor baseline information systems ($p \geq 0.25$), so the infant-health improvement is broad-based rather than concentrated in previously less-monitored hospitals. The maternal complications are the exception: the reductions in genital tract laceration and ARDS are larger in hospitals with better information systems ($p = 0.09$ and 0.01). These are precisely the outcomes identified from diagnosis codes rather than from a directly measured quantity such as birth weight, so better-recorded settings plausibly detect a given decline more reliably. The direction is also reassuring: had the inspections induced strategic under-coding of complications, the apparent improvements would have concentrated in the hospitals whose records were weakest, not the strongest.[33]

Cross-hospital patient diversion: the radius, and a sign check. The market-level test in Section 5 defines a catchment area by a five-kilometer radius. Because women travel further to reach tertiary hospitals, we re-estimate the nearby-hospital regression at ten and twenty kilometers, in each case regressing the monthly prenatal admissions of the uninspected hospitals within the radius on a post-inspection indicator, with hospital and calendar-month fixed effects and standard errors clustered at the hospital level. The coefficient is 0.38 (s.e. 2.20) within five kilometers, on a pre-inspection mean of 17.2 admissions per hospital-month across 14 hospitals; 7.78 (s.e. 3.69) within ten, on a mean of 16.7 across 17 hospitals; and 1.88 (s.e. 4.11) within twenty, on a mean of 19.6 across 26 hospitals. No radius produces a decline. The point estimates are positive but

[33] We measure information-system capacity using the annual Health Statistics Yearbooks, which record each hospital's information-technology construction: whether it has electronic medical records (EMR), a management system such as a Hospital Information System (HIS) or Hospital Resource Planning (HRP), a Picture Archiving and Communication System (PACS), and a Laboratory Information System (LIS).

imprecise and not monotone in the radius, and the ten-kilometer figure rests on the three hospitals that ring adds to the five-kilometer one, so we read the exercise as ruling out a fall rather than as establishing a positive spillover. One qualification travels with the wider radii. The magnitude argument of the main text, that the roughly 490 additional monthly prenatal admissions across the treated hospitals exceed what their neighbors could supply, is decisive at five kilometers, where the neighbors' entire baseline volume is about 240 admissions per month, and weaker at twenty, where it is about 510; over the wider catchment it is the direct estimate rather than the accounting that carries the conclusion.

The specification that returns the spillover hospitals to the control group offers a check on the sign. Diversion would depress prenatal volume at exactly those hospitals, lowering the control-group counterfactual and therefore *inflating* the estimated inspection effect. Table A6 shows the opposite: relative to Table 1, aggregate revenue falls from 401,940 to 358,982 RMB, total admissions from 50.69 to 45.00, and prenatal admissions from 19.74 to 18.69. The neighbors thus move with the inspected hospitals rather than against them, which is the signature of deterrence rather than of diversion. Because the spillover group is defined by administrative district and the catchment areas by distance, the two exercises partition the neighboring hospitals differently and reach the same answer. Neither, however, is immune to the two forces offsetting: a deterrence spillover that raises nearby volume can conceal a diversion that lowers it, in the flat event-time series and in the attenuated coefficient alike. The catchment-area total plotted in Figure A14 is the statistic that nets them, and it rises.

G Cost-Benefit Valuation: Formulas, Parameters, and Sensitivity

This appendix details the human-capital valuation underlying Section 6 and the sensitivity of the net surplus to its parameters.

Valuing the birth-weight gain. Prior literature establishes that a one-standard-deviation increase in birth weight raises adult annual income by roughly 2.75% (Lambiris et al., 2022). To monetize our estimated 18-gram increase in mean birth weight, we project the present value (PV) of lifetime income gains using a standard formulation:

$$PV = [p \times w_0] \times \sum_{s=1}^{N} \frac{(1+g)^{s-1}}{(1+r)^s} \tag{4}$$

where $r = 3\%$ is the discount rate, g is the annual growth rate of real wages, $N = 40$ denotes working years, w_0 is the baseline annual wage, and p scales the income premium by our estimated

birth weight gain relative to its sample standard deviation.[34] The summation index s runs over working years, so that $s = 1$ is the first year of labor-market participation, which we take to begin at age 20, and $N = 40$ carries the stream through to retirement at 60.

Equation (4) values the earnings stream as of labor-market entry, roughly two decades after the births that generate it, whereas the policy cost is borne during our 2015–2017 observation window. Differencing the two requires a common valuation date, so we discount the stream back to birth. Because g is the growth rate of economy-wide real wages rather than the slope of an individual age-earnings profile, it also applies to the intervening years: a child born during our window enters the labor market facing a wage level of $w_0(1+g)^T$ rather than w_0. Both adjustments are captured by a single factor,

$$PV_{\text{birth}} = PV \times \left(\frac{1+g}{1+r}\right)^T, \tag{5}$$

where $T = 19$ is the gap between the year in which wages are observed and labor-market entry at age 20. The factor falls below one only when real wage growth runs behind the discount rate, and equals exactly one when $g = r$. What governs the valuation is thus the *wedge* between the two rates, not their levels.

Benchmark parameters. We set $g = 2\%$. Paired with $r = 3\%$, this is the combination implied by the standard Ramsey rule $r = \delta + \eta g$ under a pure rate of time preference of $\delta = 1\%$ and log utility ($\eta = 1$), so the two parameters are chosen jointly rather than one being tuned against the other. As a growth projection it is deliberately undemanding: real wages in urban China grew at roughly 8 to 10 percent a year between 2000 and 2017, several times the rate we carry forward over these children's working lives. What the calculation turns on is the resulting one-percentage-point wedge rather than the growth rate itself. Applying it, the non-private-sector wage yields a per-child lifetime income gain of 1,918 RMB valued at birth, aggregating to a social benefit of 193.73 million RMB; the more conservative private-sector benchmark implies a per-child gain of 1,181 RMB and an aggregate benefit of 119.29 million RMB.[35]

Sensitivity. Table 7 reports how the net surplus responds to the parameters it rests on. Panel A varies long-run real wage growth. The surplus reaches zero at $g \approx 1.58\%$ (equivalently, at a wedge of about 1.4 percentage points between the discount rate and wage growth), and is positive at every

[34] $p = (\Delta_{BW}/\sigma_{BW}) \times 0.0275$, where Δ_{BW} indicates the increase in birth weight, and σ_{BW} indicates the sample standard deviation of birth weight among delivery admissions, which is equal to 515 grams.

[35] Baseline annual wages (w_0) in 2017 were 74,318 RMB in the urban non-private sector and 45,761 RMB in the private sector (Ministry of Human Resources and Social Security of the People's Republic of China, 2017). All monetary figures are nominal RMB. The wage and public-health-expenditure benchmarks are 2017 values and the revenue estimates are averages over 2015–2017, a window over which consumer prices rose by under four percent in total, an order of magnitude smaller than the parameter uncertainty documented in Table 7. Renminbi are converted to dollars at 6.765 RMB per USD.

higher rate, reaching 91.99 million RMB at $g = 3\%$. Because the valuation depends on r and g only through the ratio $(1 + g)/(1 + r)$, this threshold is a statement about the wedge rather than about either rate alone: at $r = 4\%$ the surplus turns positive at $g \approx 2.6\%$, at $r = 5\%$ at $g \approx 3.6\%$, and the breakeven wedge holds at roughly 1.4 percentage points throughout. Under the Ramsey rule with log utility the wedge is simply the pure rate of time preference, so the sign of the surplus is governed by δ alone and is very nearly independent of the growth rate assumed: at $\delta = 1\%$ the surplus lies between 20 and 23 million RMB for every g between 1% and 4%, and it stays positive for any δ below 1.4%. The exercise therefore does not ask the reader to forecast Chinese wage growth; it asks only for a position on how heavily to weigh the welfare of a generation not yet born.

Panel B varies the remaining parameters one at a time. Two of them (the VLBW treatment cost and the birth-weight–earnings elasticity) are strong enough to reverse the sign on their own, which is why we treat the magnitude with caution. The cost side, by contrast, only reinforces the result: because the calculation charges the entire revenue increase to society, recognizing even a modest share of it as transferred rent rather than resources consumed raises the surplus, and at the more conservative $g = 1\%$ a transfer share above 12.2% would be enough to make the surplus positive on its own, a threshold comfortably inside the range implied by the kickback literature (Yip et al., 2010; Fu et al., 2023).

Estimator choice. A third source of sensitivity is the estimator itself. Substituting the heterogeneity-robust estimates of Table A4 for the canonical ones (which requires replacing the birth weight, very low birth weight, and laceration inputs alongside the revenue input, since all four move together) raises the surplus to 35.70 million RMB under Sun and Abraham (2021) and lowers it to -14.86 million under Callaway and Sant'Anna (2021). The two robust estimators bracket the canonical figure rather than displacing it, and the spread between them comes from the benefit side rather than the cost side: the revenue estimate that sets the cost is significant under all three estimators and varies by less than three percent between the two robust ones, whereas birth weight and very low birth weight, which together carry almost the entire monetized benefit, are individually indistinguishable from zero under both. We therefore read the canonical figure as the central estimate and the robust ones as indicating the range the data can support.

Omitted benefit categories. Pushing in the opposite direction, our accounting omits several benefit categories outright. Unlike more comprehensive assessments (Baran et al., 2025), we do not value averted infant mortality, prevented lifelong childhood disabilities, or mitigated severe maternal morbidity; in high-income settings the associated mortality-risk reduction alone can exceed 1.3 million USD per VLBW case, which would dominate everything we are able to measure. We exclude these dimensions because context-specific parameters for the Value of a Statistical Life and Quality-Adjusted Life Years are scarce in our setting, not because they are small.