Information Provision and Streamlined Medical Service: Evidence from a Mobile Appointment App

Changcheng Song* Nan Yang† Junjian Yi‡ Ye Yuan§

September 20, 2018

Abstract

We examine the launch of a mobile outpatient appointment app in China to study the effect of information provision and a streamlined appointment process on hospital operations and the alignment of healthcare supply and demand. Using a longitudinal dataset on hospital operations and a difference-in-differences model, we document that the app increases completed hospital consultations by 9.5%, through boosting registrations by 4.8% and reducing appointment cancellations by 3.4%. The app improves queuing efficiency in overcrowded hospitals and draws demand for underutilized ones. Supported by additional evidence from a subset of patients' electronic medical records, we also find that the app directs patients to the hospital and department more suitable to their medical conditions and to less busy days, resulting in a better match between patient demand and hospital service.

Keywords: Healthcare Information Technology; Information Provision; Hospital Operations; Patient Sorting; Patient Choices

JEL Codes: I11, I12, I18

*Department of Economics, National University of Singapore; email: ecsscc@nus.edu.sg.

†Business School, National University of Singapore; email: bizyn@nus.edu.sg.

‡Department of Economics, National University of Singapore; email: junjian@nus.edu.sg.

§Department of Economics, National University of Singapore; email: ye.yuan@nus.edu.sg

1 Introduction

Before dawn on an ordinary day, the queue for consultation registration outside Beijing Union Hospital—one of the nation's flagship research hospitals—has stretched for hundreds of meters. While this particular hospital may be among the worst cases in the country, the media has long reported that "queues at Chinese hospitals are legendary" (Economist, 2017). Though much of the overcrowding can be attributed to a shortage of medical resources, the mismatch between healthcare supply and demand must not be overlooked. In China, the hospital-bed utilization rate has reached 102% in large research hospitals, often relegating some inpatients to makeshift beds, but is only around 60% in small primary-care facilities (National Bureau of Statistics, 2015). Elsewhere in the world, suboptimal allocation of limited medical resources has been documented in both developing and developed countries. In the US, a shortage of general practitioners and oversupply of specialists contributes to both underprovision and excess spending (Japsen, 2016; Jauhar, 2014). Despite wide recognition of the supply-demand mismatch in the Chinese healthcare system, changes have been scarce and late in coming (Eggleston et al., 2008). Institutional reforms, such as introducing a general-practitioner-based referral system and meaningful differential pricing between large hospitals and small primary-care providers, are necessary to correct the mismatch. However, these reforms are progressive and expensive.

In this paper, we study the impact of a mobile outpatient appointment application (app)—a lightweight information technology (IT) innovation—on hospital operations and the alignment of healthcare supply and demand in China. The app was designed by a private Chinese healthcare IT company, which we refer to as the Company hereafter, and launched by the Company's client hospitals. Prior to the app's launch, these hospitals, like most Chinese hospitals, primarily accepted only walk-in registration for outpatients, who would then be placed in a queue for consultation in the order of registration. Launching the app only changes hospitals' operations by informing their patients of physicians' availability and allowing advance online booking for hourly consultation slots.

Our objective in analyzing the app is twofold. First, to provide managerial insights to healthcare providers, we evaluate the app's impact on hospital operations. Like many other health IT products, the app requires a sizeable set-up cost: An adopting hospital must pay around 2 million CNY (0.3 million USD) to the Company for the app and to train its staff to accommodate the procedural changes. Meanwhile, there has been widespread skepticism about the app's cost effectiveness, with prominent hospitals already operating close to their full capacity doubtful that the app could help. The Company, in turn, was concerned that the app's easy appointment cancellation feature might waste scarce slot resources. In the absence of convincing evidence, hospitals were often incentivized by public-relational and political pressure to adopt the app.[1] The use of similar top-down incentives for health IT product adoption (Schilling, 2011) testifies to the importance of our study.

Second, and more importantly, we examine whether the app improves the match between supply and demand across hospitals and time. Before the app launch, hospitals' overreliance on walk-in registrations made it difficult for patients to learn about hospital crowdedness and physician availability without physically visiting the hospital. Using a simple analytical model, we show that this uncertainty could lead to patients' suboptimal hospital choices. The model also predicts that by eliminating the uncertainty, the app could improve match efficiency by directing patients to hospitals more suitable to their medical conditions, which we set out to test empirically. Jensen (2007) demonstrates that in the context of the Indian fisheries sector, mobile technology is capable of mitigating information asymmetry and achieving the "law of one price" across different fishery markets. We investigate the role of information provision through mobile technology in a market in which the price mechanism is largely absent due to government regulation.

1. See an administrative guideline published by the National Health and Family Planning Commission of the People's Republic of China for an example (in Chinese): http://www.gov.cn/gongbao/content/2014/content_2600086.htm.

We gained access to the Company's longitudinal dataset of hospital-department-day outpatient records from 2013 to 2016 for 22 urban hospitals in a developed coastal province of China. All of the sampled hospitals use the Company's Hospital Information System (HIS), which ensures a uniform data standard. By the end of the observational window, 9 of the 22 hospitals had launched the app. The app was launched at different times, and the sequentiality of app adoption allows a staggered difference-in-differences analysis. A substantial time gap exists between a hospital's decision to adopt and the actual app launch date, which is at least partially outside the adopting hospital's control. This suggests that it is impractical to coordinate launch timing with pre-launch trends in hospital operations, which is the key identification assumption of our analysis. We further validate this assumption in three statistical tests: using data from the emergency department, using placebo launch timing, and directly testing pre-launch trends. To supplement the hospital-department level analysis, we also examine two separate datasets. From over 130,000 patients' electronic medical records (EMR), we examine how the app affects hospital choices made by patients of various medical conditions. Using the app-user data, we study the impact of the app's subsequent launches on its first batch of users.

Our difference-in-differences analysis first establishes that the app improves hospital operations by increasing completed outpatient consultations by 9.5%. This is achieved by boosting registration counts by 4.8% and reducing cancellations—which are defined as a registered patient's failure to see a doctor on the registration day—by 3.4%. The cancellation rate averaged 12% in our sampled hospitals prior to the app launch, which suggests that nearly one out of every eight walk-in patients left after registration and before consulting a physician. Batt and Terwiesch (2015) document that patients waiting in hospital lines are sensitive to visual queue lengths and often abandon queues. Given the typically crowded condition of Chinese hospitals, such a high cancellation rate is most likely caused by unexpectedly long queues. By informing patients of physicians' availability and allowing advance booking, the app effectively reduces cancellation waste. The extent of reduction implies tens of millions per year in increased hospital revenue and savings on patient opportunity costs.

Second, the app serves as a "soft" triage mechanism that facilitates better matching between patients and hospitals:

1. In large tier-three hospitals, we find that the app induces the substitution of walk-in registrations for online bookings, and increases total consultations by reducing cancellations. In tier-two hospitals, we find that a net increase in registrations explains the increase in consultations. This is a welcome outcome, as in China, tier-three hospitals are often overcrowded and tier-two hospitals are underutilized (Yip and Hsiao, 2008, 2014).

2. We find that the app increases consultations mostly in departments that handle more severe medical conditions in tier-three hospitals and those that handle less severe conditions in tier-two hospitals, which suggests that patients are better able to sort into the "right" hospital according to their medical needs. We find consistent results in the EMR data, where we directly observe the patient's visit history and the severity of their medical conditions through recorded diagnoses.

3. We find that in post-adoption hospitals, patients avoid busier weekdays and increase their consultations during less busy weekends. App-user data show that although subsequent app launch reduces the average scheduled waiting period for users across all days of the week, the reduction on weekends is the largest.

Overall, our findings demonstrate that the app leads to a better match between patient demand and hospital service by directing patients to the hospital and department more suitable to their medical conditions and to less busy days.

Our work is related to three streams of literature. First and foremost, our paper addresses healthcare resources' allocative efficiency: how to align the supply and demand of healthcare services. Previous studies have documented extensive supply-side factors in the lack of allocative efficiency, such as an oversupply of specialists (Baicker and Chandra, 2004), distortion in physician incentives (Clemens and Gottlieb, 2014), and expensive but narrowly targeted medicine and technologies (Chandra and Skinner, 2012). Our work contributes by documenting the importance of demand-side factors in allocative efficiency: When better informed and given more choices, patients choose more suitable healthcare providers and less busy days, implying more efficient utilization of healthcare resources.

Second, our work studies non-price factors that influence healthcare demand. Previous research finds that the patient's choice of hospital is responsive to information, such as that found on social networks (Moscone et al., 2012), hospital report cards (Dranove and Sfekas, 2008), and government websites (Varkevisser et al., 2012). Studies by Gaynor et al. (2016) and Cooper et al. (2011) demonstrate that patients "shop" for hospitals when provided with more information and choices. In this paper, we study the effect of providing appointment availability information and scheduling flexibility. Our findings of improved hospital operations and demand-supply alignment show that non-price factors, such as length of waiting, ease of hospital access, and information technologies, are clearly important.

Finally, our analysis of the app adds to a growing literature of digital healthcare, such as the adoption of electronic medical records (Agarwal et al., 2012; Agha, 2014; Dranove et al., 2014; McCullough et al., 2016; Miller and Tucker, 2011); handheld mobile technologies for physicians (Prgomet et al., 2009); mobile-phone-based interventions for patient adherence to medical treatment (Chow et al., 2015; Lester et al., 2010; Ramachandran et al., 2013); the impact of health IT on clinical routines (Goh et al., 2011); online health communities (Goh et al., 2016); online physician ratings (Gao et al., 2015); telemedicine consultations (Serrano and Karahanna, 2016); and IT's facilitating role in hospital quality disclosure (Angst et al., 2014). Our work differs in its focus on documenting information provision as a way to improve patient sorting in healthcare markets.

2 Conceptual Framework

Before proceeding to the empirical analysis, we introduce a conceptual framework that guides our analysis. This starts with discussions on the mismatch between demand and supply in Chinese hospitals. We then describe the app's design, functionality, and launch process, followed by a simple model that illustrates how the app could improve sorting between patients and hospitals.

2.1 Why Are Large (Tier-three) Hospitals Overcrowded in China?

In our study, we use China's official three-tier system to differentiate large research hospitals from their smaller and primary-care-focused counterparts. This system was established between the 1950s and 1960s, and its structure and criteria have not radically changed since then (National Bureau of Statistics, 2015). The three tiers can generally be regarded as communal- (tier one), township- (two), and city-level (three) hospitals. The classification is slightly different in rural and urban areas, and depends on weighted scores of hospital characteristics. Size measures, such as the number of inpatient beds, are the predominant classifiers (Ministry of Health, 1989). Tier-three hospitals, or what we sometimes loosely refer to as "large hospitals," provide comprehensive care, such as acute care and specialist services. They also play a dominant role in medical education and research. Lower-tier hospitals mainly provide primary and preventive care, perform basic surgical procedures, and serve local communities. We discuss the Chinese healthcare market in more detail in online Appendix A.

Although tier-three and lower-tier hospitals provide care of vastly different quality, the outpatient consultation fee is regulated and capped at a very low rate across all hospitals.[2] Prices for prescription drugs and basic diagnostic tests are also heavily regulated and have limited variation between hospitals in different tiers (Yip and Hsiao, 2008). Moreover, there is no general-practitioner-based referral system in China or any other formal protocol for patient

2. In our sampled hospitals, this is generally less than 15 CNY (just above 2 USD).

referrals (Xu et al., 2010). Consequently, the nation's public health insurance system typically does not require a valid referral from lower-tier hospitals or general practitioners to reimburse costs incurred in higher-tier hospitals. The absence of price differentials and a referral system makes tier-three hospitals the preferred option for many patients: They skip lower-tier hospitals and flood into high-tier ones (Eggleston et al., 2008), which causes further disparities between hospital tiers. This vicious cycle quickly escalated after market-oriented healthcare reforms in the 1980s and 1990s. Tier-three hospitals have grown rapidly in size and captured the lion's share of skilled physicians, patients, and revenue. In contrast, lower-tier facilities are increasingly understaffed and underfunded, which creates a downward spiral for the quality and reputation of lower-tier facilities. In 2014, the average tier-three hospital in China employed 26 times more physicians and nurses, treated 27 times more patients, and generated 60 times more revenue than their tier-one counterparts. The bed-occupancy rate in tier-three hospitals averages 101.8%, in contrast to only 60.1% in tier-one hospitals (National Bureau of Statistics, 2015). We provide descriptive statistics that contrast hospital characteristics in different tiers in Appendix Table A1.

As much as one could argue that the overwhelming preference for tier-three hospitals is rational, the pre-mobile era of hospital appointment systems almost certainly created inefficiency. Before having access to an online appointment option, most patients simply walked into a hospital's outpatient department, registered at the reception, and were placed by the registration system in a queue for consultation. Not knowing the queuing time until after registration, many patients—especially those visiting busy tier-three hospitals—were deterred by unexpectedly long queues and abandoned their registrations (Blumenthal and Hsiao, 2015; Yip and Hsiao, 2008). We depict this traditional (offline) registration process in the left panel of Figure 1.

2.2 The Mobile Outpatient Appointment App

App Designer The mobile outpatient-appointment app was designed by a publicly listed healthcare IT company (the Company) in China. By the end of 2016, the Company had launched the app in more than 300 client hospitals in China and had close to 10 million app users. The Company typically charges a hospital around 2 million CNY (0.3 million USD) for the app and its future maintenance, but does not otherwise interfere with the client hospital's operations.

In addition to the app, one of the Company's main products is a hospital information system (HIS), which is a one-stop IT solution for hospitals. To ensure a uniform data standard, we only sample hospitals that use the Company's HIS.

Main Functions of the App The app has two main functions: displaying physicians' availability and allowing advance booking. The app is a mini-program add-on on WeChat, which is a massive social-network application: Daily active users account for more than 88% of all Internet users in China (Clover, 2016). The app works on all mobile platforms and does not require a separate app-store download or installation. To access the app for appointments in an adopting hospital, a user must subscribe to the hospital's WeChat public page, where the link to the app appears as a functional button. Clicking through the button leads to the app's landing pages, as shown in Figure 2, which are standard across adopting hospitals. First, the user is presented with the left-hand panel, which provides visual guidance to help match the user's illness or symptoms to a hospital department. Clicking on leads to a selection screen for departmental choice, as shown in the middle panel. After choosing a particular department, the user can browse on-duty physicians' available hourly consultation slots for the next two weeks on the right-hand panel and book the desired slot. After paying the registration fee via the app upon booking, the user can arrive at the designated time, skip the queue, and see the physician with a minimum on-site wait. The user can also cancel the appointment using the app for a full registration fee refund by midnight the day before the appointment date. In the right-hand panel of Figure 1, we depict the consultation process using the app. It is important to note that since the app is tied to a hospital's public page, it does not rank hospitals or provide crowdedness information from other hospitals for comparison.

Launch of the App Upon signing a sales contract, the Company dispatches IT technicians to integrate the client hospital's existing outpatient appointment system with the app. Meanwhile, the client hospital must train its staff to receive patients who go through the app channel for consultations. How soon the pre-launch preparation is completed varies greatly, and depends on the IT system's complexity and compatibility, as well as constraints on the Company's and hospital's manpower. Typically, the preparation time ranges from four to six months and is hard for hospitals to precisely predict. Therefore, it is highly unlikely that an adopting hospital can synchronize the app launch time with anticipated demand shocks or supply changes, such as hiring additional staff. Moreover, we conduct formal tests in the months before the launch time to detect whether there exist other hospital interventions or unobserved pre-trends in adopting hospitals that drive the estimated effects of app adoption. We find no such contaminations. In Section 4.2, we present further evidence from robustness analyses to show that the app's launch time is arguably exogenous and has no confounding factors.

On or around the app launch date, the adopting hospital often holds a press conference or uses other marketing techniques, such as website banners and on-premises advertising posters, to announce the launch. These marketing efforts are mostly app-centric. Therefore, though our estimated effect may pick up these marketing efforts, they should be interpreted as bundled with app adoption.

App-adopting hospitals typically allocate only a small fraction (ranging from 8-18% for tier-three hospitals in our sample) of registration slots online and retain the majority for offline patients. Based on conversations with the Company and some sampled hospitals, we do not believe that hospitals change work schedules or increase their workforce upon app launch. Indeed, we find no post-adoption increase in total registrations for hospitals that, before app launch, were likely to operate at close to capacity.

2.3 How Can the App Help?

The app is designed to streamline the queuing process and reduce the uncertainty of queuing time. The mechanism by which it affects patients' choice of hospitals on different tiers is not straightforward. To illustrate this point, we develop a stylized model. This model has a unit mass of patients and two hospitals: a tier-three (h) and a lower-tier (l). Each patient must decide which hospital to visit for one trip. They have a uniformly distributed idiosyncratic utility $V_h \sim U[\underline{v}, \overline{v}]$ for visiting the tier-three hospital for consultation. This value could be interpreted as the match value of one's medical condition to hospital—i.e., how efficaciously a patient will be treated by this hospital. Patients must queue in the hospital before consultation, and hence are affected by a visit-specific disutility from queuing, Q. Before the app launch, we assume that Q follows a uniform distribution: $Q \sim [0, \overline{q}]$. For simplicity, we fix the utility that all patients receive from visiting the lower-tier hospital at v_l, normalize the queuing disutility at lower-tier hospitals to 0, and assume that patients are risk-neutral. We analyze the case in which an average patient prefers to visit the tier-three hospital when facing the average queuing disutility, or $(\overline{v} + \underline{v} - \overline{q})/2 \geq v_l > 0$. We use capital letters V_h and Q to denote random variables and reserve lower case for their realizations and other fixed values.

Figure 3 shows the effect of the app on the patient's hospital choice and cancellation. Panels (A) and (B) present the probability density function of V_h and patient's hospital choice before and after the app adoption, respectively. Before the app's launch, patients observe their own realized v_h before deciding which hospital to visit. However, they do not learn about the realization of Q until they visit the tier-three hospital, and hence base their visit decisions on the average of queuing disutility $\overline{q}/2$. Panel (A) plots v_h on the x-axis. Those with $v_h - \overline{q}/2 \leq v_l$, or Area (I), choose to visit the lower-tier hospital. The rest choose to visit the tier-three hospital. Suppose that a patient with realized utility v_h visits the tier-three hospital and discovers the queuing disutility to be q. She will only remain in the queue to receive the net utility $v_h - q$ if $v_h - q \geq 0$. Otherwise, she will cancel her registration. Areas (II) and (III) correspond to the proportions of patients who stay through the queue and those who cancel their registrations, respectively.

After the app's launch, patients who secure consultation slots via the app experience minimum queuing time at the tier-three hospital. However, some patients may find their ideal consultation slots taken before they make their appointments. For algebraic simplicity, we assume that the app changes Q's distribution to a binomial distribution: With probability p, patients cannot secure their ideal slot ($q = \overline{q}$), and with probability $1 - p$, patients can ($q = 0$). A crucial difference from the pre-launch case is that patients observe the realization of Q before visiting the tier-three hospital. Panel (B) of Figure 3 plots the case after the app's launch, with v_h on the x-axis. Area (I*) corresponds to the proportion of patients who choose to visit the lower-tier hospital, which comprises all those with low match values with the tier-three hospital ($v_h \leq v_l$) and those who prefer the tier-three hospital to a certain extent ($v_l \leq v_h \leq v_l + \overline{q}$), but face $\overline{q}$ as queuing disutility. Area (II*) corresponds to patients who visit the tier-three hospital regardless of queuing disutility.

Under our simplified parametrization and compared with before the app launch, the lower-tier hospital receives more patients if $p > 1/2$, a lower-bound condition on the proportion of patients who are not able to secure their ideal slots. If $p < 1/2 + 1/(2\overline{q}^2) * (\overline{q}/2 - v_l)^2$, an upper-bound condition on the proportion of patients who are not able to secure their ideal slots, the tier-three hospitals also complete more consultations than before the app launch. This gain is achieved by eliminating cancellations. Since some values of p satisfy both lower- and upper-bound conditions, both the lower-tier and tier-three hospitals may benefit from the app launch.

The model is obviously highly stylized. For instance, it assumes a fixed number of patients and ignores a possible market expansion effect: More patients may be drawn to visit hospitals under information clarity. Nonetheless, the mechanism highlighted by the model produces three key empirical predictions for the app's effect. First, it may increase visits to lower-tier hospitals. Second, it may increase consultations in tier-three hospitals by reducing cancellations. And third, "marginal" patients who are more likely to switch from tier-three to lower-tier hospitals are those with less serious medical conditions (lower v_h) or who face higher queuing disutility (higher $\overline{q}$). We set out to test these predictions in our empirical analyses.

3 Data

We collect three datasets for the empirical analyses: hospital operations data, electronic medical record (EMR) data, and app-user data. Hospital operations data, our primary data source, consist of daily hospital-department-level outpatient records for 22 urban hospitals in a developed coastal province of China from January 2013 to September 2016. To study individual-level choices, we collect two supplementary datasets from a subset of hospitals and the sample period. EMR data consist of outpatient consultation records from four out of the 22 hospitals. App-user data consist of app-users' mobile registration records in all adopting hospitals after the app is adopted.

3.1 Hospital Operations Data

Our primary dataset consists of daily hospital-department-level outpatient records for 22 urban hospitals in a developed coastal province of China. The Company constructed the data by aggregating patient records in the hospitals' HIS. This baseline sample covers the period from January 1, 2013, to September 20, 2016. The panel is unbalanced, because four hospitals installed their HIS progressively in the course of 2013, and their observations are not available until HIS is up and running. Nonetheless, from September 3, 2013, onward, all 22 hospitals are observed until the end of the sample period. The first app launch was in August 2014. By the end of the sample period, nine hospitals had adopted the app.

We examine three categories of outcome variables to measure hospital operations: (1) total registrations for outpatient consultation (green ovals in Figure 1); (2) total consultations, defined as consultation sessions that actually occurred (blue rectangles in Figure 1); and (3) registration cancellation rate, defined as the ratio between registration-consultation difference and total registrations (red diamonds in Figure 1). A cancellation can be an "informed" one, in which a registered patient informs the hospital by canceling through the app or revisiting the registration counter for a registration fee refund, or an "uninformed" one, in which a registered patient simply forfeits his/her registration by abandoning the queue and leaving the hospital. In our empirical analyses, we do not distinguish between these two types of cancellations.

The three outcome variables are further classified as either "online" (right panel of Figure 1) or "offline" (left panel of Figure 1). Online registrations are those made in advance using the app, and offline registrations are made on-site at the hospitals. Online/offline consultations and cancellation rate refer to those following online/offline registrations. Since online outcome variables are definitionally unavailable before the app launch, we use the three total outcome variables and the three offline ones as dependent variables in the analyses.

We present some descriptive statistics in Table 1. The sample covers nearly 30,000 hospital-days and the app is functioning on 17.1% of these days. Table 1's Panel A shows that on average, app-adopting hospitals receive more daily registrations and consultations and experience a higher cancellation rate than non-adopters. The high cancellation rate is particularly worth noting: adopters receive a total of 2,228 daily registrations, of which 1,846 eventually consult a physician, leaving 14.6% of registrations unfulfilled. In Table 1's Panel B, we conduct a simple before-after app launch comparison. Adopters' daily consultations increased from 1,704 to 2,026, and daily registrations increased from 2,153 to 2,323. Even with the ease of cancelling registrations on the app, the online/offline combined cancellation rate decreased by

5.4 percentage points for adopters, but barely changed for non-adopters.

3.2 Supplementary Individual Data

Electronic Medical Records (EMRs) We procure more than 130,000 individual patients' electronic medical records (EMRs) from four hospitals in the same district of the provincial capital. These four hospitals are among the 22 included in the hospital operations data. The EMR data's sample period, from January 2015 to September 2016, is also covered by the hospital operations data. Three of the four hospitals, two tier-two and one tier-three, launched the app before the sample period for the EMR data. The fourth one, a tier-three hospital, adopted the app during the sample period. EMR data are longitudinal. Each record includes a patient identifier, some patient characteristics, the department name, and the diagnosis.

App-user Data We use app-user data to study whether app adoption shortens the time between mobile booking and actual consultation. App-user data contain more than 1.7 million mobile booking records from August 1, 2014—the first launch date—to September 30, 2016. For each registration record, available variables include the patient identifier and basic demographics, the registered hospital department, and the timestamps when a mobile booking is made and when the consultation begins. In the following analyses, we define the time difference between the mobile booking and the actual consultation as the on-app scheduled waiting period.[3] In our sample, the on-app scheduled waiting period averages 44.6 hours, which implies that app users secure a consultation slot roughly two days in advance.

These three datasets complement each other to offer a holistic view of the app's impact. Our main analysis uses hospital operations data, which contain essential information on the app's performances and provide the most comprehensive coverage in terms of time and geography. EMR data contain detailed information on patients' medical conditions and allows us to track

3. See Figure 1 for how the on-app scheduled waiting period differs from the queuing time in traditional walk-in visits.

individuals' hospital choices over time. We use this dataset to show that the app directs patients to hospitals that are suitable for their medical conditions. However, only one app launch occurs during the EMR data's sample period, which restricts our analysis to an "event study." App user data are generated when users book consultation slots on mobile phones. Therefore, it only captures booking activities at adopting hospitals after the app was launched. In other words, the patient's behavior off the app or before the app was not observed. Bearing this limitation in mind, we use this dataset to study the effect of subsequent app launch on the scheduled waiting period of the first batch of app users.

4 Hospital Operations

4.1 Benchmark Results

We estimate the app's impact on hospital operations using hospital operations data and the following difference-in-differences model:

$$y_{it} = \alpha + \beta App_{it} + \lambda_i + \lambda_t + \epsilon_{it}, \tag{1}$$

where y_{it} is one of the six outcome variables for hospital i on day t: total registrations, total consultations, overall cancellation rate, offline registrations, offline consultations, and offline cancellation rate. Binary variable App_{it} is one if hospital i adopted the app on or before day t and zero otherwise. The coefficient of interest, β, measures the effect of the app on the outcomes. The hospital's time-invariant characteristics are controlled by hospital fixed effect λ_i. Dummy variables λ_t include both month fixed effect and year fixed effect, used to control for time-varying but hospital-invariant factors. In all analyses, the number of registrations and consultations are in logarithm, such that we can interpret the estimated effect as percentage changes. Robust standard errors are clustered at the hospital level.

Table 2's Panel A presents the estimates for Equation (1). We compute p-values using the wild bootstrap procedure proposed by MacKinnon and Webb (2018) to account for a small number of clusters. The simple model seems to fit the data well, scoring R^2 around 0.8 for consultations and registrations and around 0.66 for cancellations. The app increases an adopting hospital's total daily consultations by 9.5% (Column 1). This is achieved by both increasing registrations by 4.8% (Column 2) and reducing the cancellation rate by 3.4 percentage points (Column 3). Daily offline registrations shrink by 4.4% after app launch (Column 5), suggesting a straightforward substitution of registrations from offline to online. This reduction is offset by a drop of 3.6 percentage points in the offline cancellation rate (Column 6), leaving the number of consultations via the offline channel almost unchanged after the app launch (Column 4).

We attribute the reduction in cancellations to two possible factors. First, patients with higher costs of queuing—hence a higher chance of abandoning queues—are more likely to switch the registration from offline to online. Since these patients can now arrive at their precise appointment times, they cancel less often. Second, with online patients arriving at their appointed times, queues become visibly shorter in hospitals' waiting rooms, which encourages patients who register offline not to cancel their appointments.

Without directly observing health outcomes, we refrain from speculative normative assessments of the changes in registrations and consultations. However, it is difficult to believe that the reduction in cancellations is not efficiency-improving. Some conservative back-of-the-envelope calculations help to put the 3.4-percentage-points reduction into perspective. The nine adopting hospitals in our sample combined had, on average, 15,741 daily registrations before the app launch. A 3.4-percentage-points reduction in cancellations means that after the app launch and on each day, around 535 registered patients who would otherwise fail to consult a physician can now go through the queue. On average, each outpatient consultation generated around 250 CNY (36 USD) in revenue for adopting hospitals before the app launch. Hence, the reduced cancellation implies an increased revenue of 133,750 CNY (19,110 USD) per day, or 48.8 million CNY (7 million USD) per year for the nine adopting hospitals. Suppose that each of these patients would have wasted two hours each for their cancelled hospital consultations. Monetizing the wasted hours using the average hourly wage in the province's urban cities, 35 CNY (5 USD), the app thus reduces losses in opportunity costs by 37,450 CNY (5,350 USD) per day, or 13.7 million CNY (2 million USD) per year. In total, even ignoring the probable health benefits, the app produces efficiency gains of 62.5 million CNY (9 million USD) annually through boosting revenue for all nine adopting hospitals and reducing wasted time for patients.

4.2 Identification and Robustness Checks

Model (1)'s causal identification requires that pre-launch trends in the dependent variables be parallel—that is, after controlling for hospital and time fixed effects, adopting and non-adopting hospitals experience no systematically different trends in registration, cancellation, or consultations. In Section 2.2, we argue that the partially exogenous pre-launch preparation time makes it difficult for hospitals to anticipatively coordinate their adoption choices with a special time trend. In addition, we conduct formal statistical tests to support the parallel trend assumption.

Using "Placebo" Launch Times In the first test, we construct pseudo launch times for adopting hospitals to determine whether they experienced pre-adoption trends that were different from those of non-adopters. To this end, we restrict the sample to the first 19 months of the data (January 2013 to July 2014), at which time no hospital had adopted the app. Then, we further create App_{it}^{T} as a "placebo dummy" that switches to one after a pseudo launch time T if hospital i is one of the nine adopting hospitals. We set time T to be each of the 13 months between April 2013 and April 2014 to ensure that there are at least 3 months of observations on either side of T. Figure 4 depicts the β estimates from 13 regressions and their 90% confidence intervals, obtained by estimating Equation (1) using App_{it}^{T} in place of App_{it}. We report regression results in Appendix Table A2. If, supposedly, the number of registrations were already rising faster or the number of cancellations were falling faster in adopting hospitals prior to app launch, some placebo dummies would capture this pre-trend and have statistically significant coefficients. However, all placebo estimates are close to zero and statistically insignificant, and significantly different from baseline estimates (dashed horizontal line). This leads us to reject differences in pre-adoption trends between adopting and non-adopting hospitals.

Using Lead and Lag Launch Times In the second test, we construct pseudo launch times using the leads and lags of the actual launch time. Figure 5 plots estimates using 11 launch times and their 90% confidence intervals. Five of these launch times are five, four, ..., and one quarter(s) earlier than the actual one (App_{it}^{-5} to App_{it}^{-1}). Five of these are five, four, ..., and one quarter(s) later than the actual one (App_{it}^{1} to App_{it}^{5}). The estimate in the middle uses the actual launch time. Corresponding regression results are reported in Appendix Table A3. Similar to the first test, we find no significant effect in lead quarters, which further substantiates that the launch timing is unrelated to pre-existing trends.

This exercise also rules out an alternative explanation for the estimated effects in Table 2. One may be concerned that, in addition to launching the app, adopting hospitals may implement other measures to improve the offline queuing procedure, such as improving the waiting area environment and hiring more experienced staff to work at the registration desk. However, the app's exact launch time is usually months away and not precisely predictable when a hospital decides to adopt. In this case, it is unlikely that hospitals synchronize improvement measures to occur in the same month as the app launch. It is more plausible that many of the improvements will take place shortly after the decision to launch the app but before the actual launch, in which case the lead dummies would have turned out significantly positive.

The estimated coefficients of the lag dummies also have economic implications: They capture the dynamic effect of the app after its launch. There exists a persistent rise in total consultations and registrations, and a persistent decline in cancellations after app launch for adopting hospitals. This is consistent with a continuous diffusion of the awareness of the app and patients' evolving taste for app-provided convenience, though coefficients of lag dummies should be treated with some caution, as post-adoption data become more sparse for later lags.

Using Emergency Room Visits The benchmark analysis excludes Emergency Room (ER) data. As visits to the ER are by nature unplanned and unexpected, the app does not handle registration for ER admissions. This feature allows us to use ER visits as a falsification test for the exogeneity of app-launch timing. If adopting hospitals coordinate app launch with anticipated demand shocks, such as an epidemic or patient population spike, estimating Equation (1) using ER visits and registrations would show statistically significant positive effects. Also, since the average cancellation rate for ER visits amounts to 10.9% before the app launch, if adopting hospitals implement measures other than the app to improve queuing, such as additional staffing, we expect that such measures would also impact ER visits. In contrast, Table 2's Panel B shows negligibly small and statistically insignificant effects.

Robustness Checks We take advantage of the large sample size and its ensuing large statistical power and reestimate Equation (1) in a few variations as robustness checks. First, noting that adopting hospitals are systematically different from non-adopters, in addition to the tests on the sufficient parallel pre-launch trends assumption, we drop non-adopting hospitals from the sample and use later adopters as the control group. Second, to parametrically account for possible differences in pre-launch trends, we include city-specific time trends and the interaction between the linear time trend and pre-launch hospital-level average cancellation rates. This specification accounts for the possible selection by which earlier adopters are those whose pre-launch cancellation rates were rising faster. Finally, we also add linear and quadratic time trends to Equation (1) and test robustness in: (1) a subsample that drops the specialty hospitals (dermatology, neurology, and dentistry); (2) a subsample that drops weekend observations; (3) a subsample that is restricted to hospitals in the provincial capital; and (4) a subsample that drops observations before September 3, 2013, to get a balanced panel. We plot the estimated effect of app adoption (β and its 90% confidence intervals) in Appendix Figure A1 and report regression results in Appendix Table A4. All estimated effects are similar to the baseline.

5 Patient Sorting

The increase in consultations and the reduction in cancellations suggest that patients have responded to the additional information and flexibility provided by the app. The stylized model in Section 2.3 produces three predictions on how the app facilitates patient sorting across hospitals: First, the app may increase visits to lower-tiered hospitals. Second, it may increase consultations in tier-three hospitals by reducing cancellations. And third, "marginal" patients who are more likely to switch from tier-three to lower-tier hospitals are those with less serious medical conditions or those who face higher queuing disutility. We test these predictions in this section. Beyond these model predictions, we also explore how the app improves patient sorting across time, that is, from busier days to less busy days. In a nutshell, we consider three channels through which patients' sorting across hospitals may be improved: sorting across hospitals for average patients, sorting across hospitals based on the medical condition's severity, and sorting across time.

5.1 Patient Sorting across Hospitals

We first examine post-adoption changes in the hospital choice of the average patient. Of the nine app-adopting hospitals, four are tier-three and five are tier-two. All but one of the tier-two hospitals are comparatively small, receiving only one-third of the outpatients received by the tier-three hospitals before the app's first launch. These small tier-two hospitals mainly provide primary healthcare. Only one tier-two hospital is comparable in size to the tier-three hospitals, and has a strong focus on maternity and child healthcare. All of our results in this paper are qualitatively unchanged if we exclude this large tier-two hospital. We use the following model to estimate the app's heterogeneous impact on hospitals of different tiers:

$$y_{it} = \alpha + \beta_{TierTwo} TierTwo_i \times App_{it} + \beta_{TierThree} TierThree_i \times App_{it} + \lambda_i + \lambda_t + \epsilon_{it}, \quad (2)$$

where $TierTwo_i$ and $TierThree_i$ are binary indicators for whether hospital i is a tier-two or tier-three hospital, respectively.

We report estimation results in Table 3, and plot estimates for $\beta_{TierTwo}$ and $\beta_{TierThree}$ and their 90% confidence intervals in Appendix Figure A2. Tier-two and tier-three hospitals experience 9.4% and 9.5% increase in outpatient consultations, respectively. Both estimates are statistically significant at the 1% level and similar in magnitude to the baseline result. However, the result reveals important differences in how the app works for different tiers. Tier-three hospitals may have already operated close to capacity before the app launch, limiting further increase in total registrations. Instead, as in the baseline results, registrations shift from offline to online. The streamlined queuing leads to a 6.4-percentage-points reduction in overall cancellation rate, which explains the increase in consultations. For tier-two hospitals, we find no evidence that pre-launch offline registrations shift to online; the 9.4% increase in outpatient consultations is mostly driven by an increase in online registrations. These results are consistent with the predictions of our model that the app increases visits to lower-tiered hospitals and reduces cancellations in tier-three hospitals. These results also demonstrate the app's effectiveness for both hospitals that are experiencing long queues and those that are underutilized.

We further examine individuals' choice patterns in the supplementary app-user data. Since the data, by construction, only include app users in incumbent adopters and only record patient choices via the app, we cannot assess the effect of the first app launch on the general population. Instead, our primary objective is to examine whether subsequent app launches in the same region shorten a patient's "in-app" scheduled waiting period, defined as the number of hours that elapse between the moment one makes an appointment and the actual consultation (see the right panel of Figure 1). This scheduled waiting period may be either voluntary, if one sets the appointment time to suit his/her own schedule, or involuntary, if one's most desired appointment slot has been taken, or a combination of both. However, if we find evidence that subsequent launches reduce the scheduled waiting period, the reduction is most likely for the involuntary portion. Essentially, with the supplementary app-user data, we estimate the spillover effect of new app launches on the in-app scheduled waiting period of the first batch of app users.

We estimate the following model:

$$w_{ijt} = \alpha + \beta_{New} New_{ijt} + \lambda_i + \lambda_j + \lambda_t + \epsilon_{ijt}, \tag{3}$$

where w_{ijt} is the in-app scheduled waiting period for patient i in city j with a consultation appointment on day t. New_{ijt} denotes the number of subsequent app launches in city j until day t since patient i's first online registration. λ_i and λ_j are the individual and city fixed effects, respectively. λ_t includes both month fixed effect and year fixed effect.

Table 4's Column (1) presents the results of Equation (3). On average, each additional app launch reduces the in-app scheduled waiting period by 2.6 hours. This reduction in scheduled waiting period implies that app-users could either secure a consultation slot sooner or make their registration later. Both scenarios suggest more flexibility for existing app users. Furthermore, we investigate the heterogeneous effects of app launches by different hospital tiers by estimating the following model:

$$w_{ijt} = \alpha + \beta_{TierTwo} New_{ijt}^{TierTwo} + \beta_{TierThree} New_{ijt}^{TierThree} + \lambda_i + \lambda_j + \lambda_t + \epsilon_{ijt}, \tag{4}$$

where $New_{ijt}^{TierTwo}$ and $New_{ijt}^{TierThree}$ measure the number of subsequent apps launched by tier-two and tier-three hospitals, respectively. Table 4's Column (2) reports our statistically significant findings: The in-app scheduled waiting period is reduced by only 0.7 hours after an additional adoption by a tier-three hospital, but by 5 hours after an additional adoption by a tier-two hospital. This is intuitive: Tier-three hospitals' in-app slots are often claimed quickly, which limits the app's pressure-releasing effect on already adopting hospitals.

5.2 Patient Sorting and Medical Condition Severity

With evidence pointing to changes in hospital choice following app launch, a natural question is whether patients have been induced to make the "right" choice. Our analytical model predicts that the app improves patient sorting: Marginal patients with less severe medical conditions are more likely to go to tier-two hospitals after the app launch, and those with more severe conditions to tier-three hospitals. In this section, we present empirical evidence at two levels of aggregation: first at the hospital-department-day level using our primary dataset, and then at the individual level using the EMR data.

Hospital-level Evidence Using our primary dataset, we map the ambulatory-care-sensitive conditions listed by Shigeoka (2014) to the departments of endocrinology, cardiology, pulmonology, urology, cardiothoracic surgery, and orthopedics and refer to them as "more severe" departments. We then categorize ophthalmology, otolaryngology, dermatology, dentistry, health promotion, rehabilitation, nutrition, and Chinese medicine as "less severe" departments. We estimate Equation (2) separately for the two categories, excluding uncategorized departments from the analysis.

Table 5 reports the estimated effect of the app on patients' department choices across different hospitals. We also plot the coefficient estimates in Appendix Figure A3. For tier-two hospitals, whose designated role is front-line primary care providers, the less severe departments experience a 14.2% increase in total registrations and a 15.6% increase in total consultations, whereas for the more severe departments, registrations drop by 5% (statistically insignificant), and consultations remain almost unchanged. For tier-three hospitals, there is a large and statistically significant increase in total consultations in the more severe departments (18.9%), and a much smaller increase in the less severe departments (4.5%, marginally significant). Overall cancellation rate also takes the larger dip (4.9%) in the more severe departments, implying a much shortened queue, which promotes an increase of 14.6% in offline consultations. Across the board, the app appears to serve as a self-triage mechanism and directs patients to consult the "right" hospital based on their medical conditions.

Individual-level Evidence To supplement the hospital-level analysis, where we indirectly infer patients' medical conditions from the departments they visit, we also look for evidence from the individual-level EMR data, where we directly observe patients' diagnoses. Before the EMR data's sample period, three hospitals have already adopted the app. The fourth one, a tier-three hospital, adopted the app during the sample period. To examine how the app launch affects patients' hospital choices, we restrict the sample to patients who have hospital visits on both sides of the launch date with the same diagnosis and estimate the following two models:

$$Tier3Adopter_i = \alpha + \beta_1 PreTier2_i + \beta_2 Severe_i + \beta_3 PreTier2_i \times Severe_i + \gamma X_i + \lambda_i^{Pre} + \lambda_i^{New} + \epsilon_i, \tag{5}$$

and

$$\begin{aligned} Tier2_i = \alpha + \beta_1 PreTier3Adopter_i + \beta_2 NonSevere_i + \beta_3 PreTier3Adopter_i \times NonSevere_i \\ + \gamma X_i + \lambda_i^{Pre} + \lambda_i^{New} + \epsilon_i. \end{aligned} \tag{6}$$

Equation (5) examines whether patients switch to the tier-three adopter from a tier-two hospital. The subscript i indexes patients and $Tier3Adopter_i$, $PreTier2_i$, and $Severe_i$ are dummy indicators. $Tier3Adopter_i$ is valued 1 if patient i's first visit after the app launch occurs in the newly adopting tier-three hospital, and 0 otherwise; $PreTier2_i$ is valued 1 if patient i's previous visit (before the app launch by definition) occurs in a tier-two hospital; $Severe_i$ is valued 1 if patient i has a severe medical condition, which is classified as such if the patient's diagnosis is related to cancer, hypertension, diabetes with complications, angina pectoris or myocardial infarction, atherosclerosis or other diseases of the arteries, chronic obstructive pulmonary disease or pneumonia, hepatic failure, acute kidney failure or chronic kidney disease, or traumatic brain injury (Shigeoka, 2014).

Equation (6) examines whether patients switch from the tier-three adopter to a tier-two hospital. Dummy indicators $Tier2_i$ and $PreTier3Adopter_i$ are defined analogous to their counterparts in Equation (5). Binary variable $NonSevere_i$ is valued 1 if patient i has a non-severe medical condition, which is classified as such if the patient's diagnosis is related to ophthalmology, dermatology, dentistry, health promotion, rehabilitation, nutrition, or Chinese medicine.

In both Equations (5) and (6), we include in X_i the set of individual characteristics: gender, age, and ID-registry location dummy. We also include a set of monthly and yearly time dummies to control for seasonal factors such as epidemic outbreaks at the time of pre- and post-launch visits: λ_i^{Pre} denotes month and year dummies for pre-launch visits, and λ_i^{New} for post-launch visits.

Table 6, panel A reports estimation results for Equation (5). Our preferred specification is model (3), which includes all controls. The estimates reveal that, of the patients who have visited tier-two hospitals before the app launch, those with severe medical conditions are 4.2% more likely than those with non-severe medical conditions to visit the tier-three hospital after the app launch. Table 6, panel B reports estimation results for Equation (6). Model (3), with all controls, shows of patients who visited the tier-three adopter before the app launch, those with non-severe medical conditions are 6.4% more likely than those with severe medical conditions to visit a tier-two hospital after the app launch. We obtain qualitatively similar results across the board when we separately examine 17 subcategories of diagnoses (e.g., cancer as "severe," rehabilitation as "non-severe"). Through this event study on more than 130,000 patients, we find consistent evidence that the app directs patients to hospitals that are more suitable for their medical conditions.

5.3 Patient Sorting across Time

In this section, we examine whether the app induces patients to adjust the timing of hospital visits to avoid busy hours. This is another potential channel through which information provision may improve the allocative efficiency of healthcare resources. In China, although some outpatient departments, such as radiology, may be closed on Sundays, most departments are generally open seven days a week. In many Chinese hospitals, including those in our sample, weekdays usually see more consultations—and consequently more congestion—than weekends. We report the descriptive statistics by day of the week in Appendix Table A7. Before the first app launch, average daily consultations on weekdays are 40.5% higher than on weekends (1,630 vs. 1,160). The average cancellation rate is also higher on weekdays than on weekends (12.9% vs. 11.8%). Monday is the busiest day of the week, with an average of 1,865 consultations and a 13.4% cancellation rate.

To examine whether the app encourages patients to shift consultations to less busy weekends, we estimate the following model separately for Monday, Tuesday, ..., Friday, and Weekends:

$$y_{it} = \alpha + \beta_{DOW} App_{it} \times DOW_t + \beta_{Other} App_{it} \times Other_t + \lambda_i + \lambda_t + \epsilon_{it}, \tag{7}$$

where binary dummy DOW_t indicates the tth day of the week. Binary $Other_t$ is DOW_t's complement dummy. For instance, when we estimate Equation (7) for weekends, $DOW_t/Other_t$ is one/zero when day t is either Saturday or Sunday and zero/one otherwise. λ_i is hospital fixed effect, and λ_t includes both month fixed effect and year fixed effect. Note that we run six regressions, one for each weekday, instead of including six weekday dummies in one regression. This is to allow the fixed effects to be flexibly different across regressions. As such, the estimate for β_{DOW} should be interpreted as the app's impact on a given day of the week relative to this day's pre-app average, rather than relative to other days of the week.

We plot the six estimates for β_{DOW} and their 90% confidence intervals in Figure 6. Detailed regression results are included in Appendix Table A8 and Table A9. Both total registrations and total consultations increase disproportionally more on weekends than weekdays, suggesting that demand is directed to the hospitals' less busy hours. Overall cancellations and offline cancellations drop across all days of the week. Offline registrations drop across weekdays, but remain almost unchanged on weekends. Due to the reduction in offline weekend cancellations, offline weekend consultations increase by nearly 5%. The levelled offline weekend registrations could be because the offline-to-online substitution does not occur on less busy weekends, or because the app—perhaps by displaying abundant weekend time slots—attracts more patients on weekends through offline channels, which offsets the offline-to-online substitution. The latter would suggest that the app's impact spills over to the offline channel.

We examine individuals' choice patterns in the supplementary app-user data. In this analysis we estimate separately for Monday, Tuesday, ..., Friday, and Weekends:

$$w_{ijt} = \alpha + \beta_{DOW} New_{ijt} \times DOW_t + \beta_{Other} New_{ijt} \times Other_t + \lambda_i + \lambda_j + \lambda_t + \epsilon_{ijt}, \tag{8}$$

where w_{ijt} is the in-app scheduled waiting period for patient i in city j with a consultation appointment on day t. New_{ijt} denotes the number of new apps adopted in city j on day t since patient i's first online registration. Binary dummies DOW_t and $Other_t$ are defined as in Equation (7). λ_i and λ_j are individual and city fixed effect, respectively. λ_t includes both month fixed effect and year fixed effect.

Figure 7 Panel (A) plots estimated scheduled waiting-period reduction by day of the week and their 90% confidence intervals. Detailed regression results are included in Appendix Table A10. New app launches lead to a larger reduction in scheduled waiting period for consultations on weekends, which is consistent with the proposition that the app incentivizes patients to avoid busy days. Since app users generally prefer weekend consultations, a subsequent app launch by another hospital attracts away like-minded app users and relieves weekend slots by a larger margin than weekday slots.

Interacting hospital tiers with day-of-week provides further evidence that avoiding overcrowding is an important factor in the patients' hospital choice. Figure 8, Panel (B) plots coefficient estimates for the interaction terms of hospital tier and day-of-week. Detailed regression results are included in Appendix Tables A11 to A13. Tier-three hospitals see a statistically significant increase in consultations only on the weekends, while tier-two hospitals experience consistent increases across the entire week. This is consistent with our analytical model's prediction that patients who face high queuing disutility (visiting on busy days) are more likely to go to tier-two hospitals using the app. A 4.8% weekend increase in offline registrations in tier-two hospitals similarly implies that the app's effect may spill over to offline patients. In further examing the "more severe" departments in tier-three hospitals (Appendix Figure A4), we find a similar result: The increase in registrations is only significant on the weekends. This analysis explains a somewhat puzzling result in Section 5.2, in which the app increases registrations at these departments, even though they were already operating close to capacity prior to the app launch.

Using the supplementary app-user data, Figure 7 presents consistent evidence of the reduction in scheduled waiting period following subsequent app launches. Detailed regression results are included in Appendix Table A14. The reduction is large and stable across the week when the new app is adopted by a tier-two hospital, but only statistically significant on the weekend when the new app's adoption is by a tier-three hospital.

6 Conclusion

Our study shows that an inexpensive, lightweight, and non-intrusive IT innovation simultaneously improves hospital operations and the alignment of healthcare supply and demand. In

2015, Chinese hospitals performed more than three billion outpatient consultations.[4] Holding supplied resources fixed, even a tiny percentage increase in total consultations could imply a radical improvement in healthcare access and efficiency; our estimates of the app's effect in increasing total consultations is 9.5%. Therefore, we believe that the app could be categorized as a "Type I" intervention in Chandra and Skinner's (2012) typology—namely, a cost-effective "home run" innovation with the potential to benefit a large population.

Given the severe rationing problem and extremely outdated appointment booking process in Chinese hospitals, one might argue that the app simply picks up low-hanging fruit. We would argue, however, that there is a lot of low-hanging fruit in the healthcare systems of developing—and even developed—countries. These are places in which simple economic mechanisms and technological interventions can go a long way toward improving the quality of life, and therefore warrant more research.

We close by discussing three possible extensions. First, the caveat of aggregate data and the limited scope of our individual EMR data imply that we cannot definitively assert that the net increase in registrations in tier-two hospitals is because some patients switch from crowded tier-three hospitals and other non-adopting hospitals, or due to newly generated demand for medical services. This is, however, an important question regarding the app's scalability: What if all hospitals in the medical system adopt such an app? In the extreme case, in which the net increase is purely due to patients who switch from other non-adopting hospitals, the app's effect on tier-two hospitals will diminish to zero when all hospitals adopt this app; this implies that the tier-two hospitals' app launches are a business-stealing social waste. Comprehensive individual choice data are necessary to assess the extent of business-stealing.

Second, in the partially privatized Chinese healthcare market—and especially if the app can empower tier-two hospitals to gain market share from competitors—one would expect that market structure plays a role in the app's adoption and diffusion. However, our analysis, except

4. Source: (in Chinese) National Health and Family Planning Commission of the People's Republic of China http://www.nhfpc.gov.cn/guihuaxxs/s10748/201607/da7575d64fa04670b5f375c87b6229b0.shtml.

for examing hospitals' pre-trends, ignores the app's two-sided nature and focuses primarily on the demand side. An extension in this direction would require data with universal coverage of multiple local healthcare markets.

Lastly, an IT-based medical intervention runs the risk of creating new social inequalities based on technological savvy. In terms of this particular app, using aggregate patient demographic data to examine responses from different age groups, our results seem to suggest otherwise. For those age 60 and above, the app increases their consultations by over 8%, largely from increased offline registrations (Appendix Table A15). It appears that even though elderly patients are not as active in using the app, they benefit from the shortened queues in adopting hospitals. However, the aggregate data are too sparse to answer this question definitively.

References

Agarwal, Ritu, Abhay Mishra, Corey Angst, and Catherine Anderson, "Digitizing Healthcase: The Ability and Motivation of Physician Practices and Their Adoption of Electronic Health Record Systems," *Information Systems Research*, 2012, *23*, 738–760.

Agha, Leila, "The Effects of Health Information Technology on the Costs and Quality of Medical Care," *Journal of Health Economics*, 2014, *34*, 19–30.

Angst, Corey, Ritu Agarwal, Guodong Gordon Gao, Jiban Khuntia, and Jeffrey S McCullough, "Information Technology and Voluntary Quality Disclosure by Hospitals," *Decision Support Systems*, 2014, *57*, 367–375.

Baicker, Katherine and Amitabh Chandra, "The Productivity of Physician Specialization: Evidence from the Medicare Program," *American Economic Review*, 2004, *94* (2), 357–361.

Batt, Robert J and Christian Terwiesch, "Waiting Patiently: An Empirical Study of Queue Abandonment in An Emergency Department," *Management Science*, 2015, *61* (1), 39–59.

Blumenthal, David and William Hsiao, "Lessons from the East – China's Rapidly Evolving Health Care System," *New England Journal of Medicine*, 2015, *372* (14), 1281–1285.

Chandra, Amitabh and Jonathan Skinner, "Technology Growth and Expenditure Growth in Health Care," *Journal of Economic Literature*, 2012, *50* (3), 645–680.

Chow, Clara K, Julie Redfern, Graham S Hillis, Jay Thakkar, Karla Santo, Maree L Hackett, Stephen Jan, Nicholas Graves, Laura de Keizer, Tony Barry et al., "Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients with Coronary Heart Disease: A Randomized Clinical Trial," *Journal of the American Medical Association*, 2015, *314* (12), 1255–1263.

Clemens, Jeffrey and Joshua D Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Clover, Charles, "Tencents WeChat Dominates Chinas Lucrative Messaging App Market," *Financial Times*, March 2016. Available at https://www.ft.com/content/aadd256e-d0ef-11e5-831d-09f7778e7377.

Cooper, Zack, Stephen Gibbons, Simon Jones, and Alistair McGuire, "Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms," *Economic Journal*, 2011, *121* (554), F228–F260.

Dranove, David and Andrew Sfekas, "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards," *Journal of Health Economics*, 2008, *27* (5), 1201–1207.

__ , Chris Forman, Avi Goldfarb, and Shane Greenstein, "The Trillion Dollar Conundrum: Complementarities and Health Information Technology," *American Economic Journal: Economic Policy*, 2014, *6* (4), 239–270.

Economist, "China Needs Many More Primary-care Doctors," *Economist*, April 2017. Available at http://www.economist.com/news/china/21721948-memories-barefoot-ones-put-some-people-seeing-them-china-needs-many-more-primary-care.

Eggleston, Karen, Li Ling, Meng Qingyue, Magnus Lindelow, and Adam Wagstaff, "Health Service Delivery in China: A Literature Review," *Health Economics*, 2008, *17* (2), 149–165.

Gao, Guodong Gordon, Brad N Greenwood, Ritu Agarwal, and Jeffrey S McCullough, "Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perception of Quality," *MIS Quarterly*, 2015, *39* (3), 565–589.

Gaynor, Martin, Carol Propper, and Stephan Seiler, "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service," *American Economic Review*, 2016, *106* (11), 3521–57.

Goh, Jie Mein, Guodong Gao, and Ritu Agarwal, "Evolving Work Routines: Adaptive Routinization of Information Technology in Healthcare," *Information Systems Research*, 2011, *22* (3), 565–585.

Goh, Jie-Mein, Guodong (Gordon) Gao, and Ritu Agarwal, "The Creation of Social Value: Can an Online Health Community Reduce Rural-Urban Health Disparities?," *MIS Quarterly*, 2016, *40* (1), 247–263.

Japsen, Bruce, "U.S. Faces Doctor Shortage Despite Specialist Growth," *Forbes*, December 2016. Available at https://www.forbes.com/sites/brucejapsen/2016/12/30/u-s-still-faces-doctor-shortage-despite-specialist-growth/#2c267aa11aa7.

Jauhar, Sandeep, "One Patient, Too Many Doctors: The Terrible Expense of Overspecialization," *Time*, August 2014. Available at http://time.com/3138561/specialist-doctors-high-cost/.

Jensen, Robert, "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *Quarterly Journal of Economics*, 2007, *122* (3), 879–924.

Lester, Richard T, Paul Ritvo, Edward J Mills, Antony Kariri, Sarah Karanja, Michael H Chung, William Jack, James Habyarimana, Mohsen Sadatsafavi, Mehdi Najafzadeh et al., "Effects of a Mobile Phone Short Message Service on Antiretroviral Treatment Adherence in Kenya (Weltel Kenya1): A Randomised Trial," *Lancet*, 2010, *376* (9755), 1838–1845.

MacKinnon, James G and Matthew D Webb, "The wild bootstrap for few (treated) clusters," *The Econometrics Journal*, 2018, *21* (2), 114–135.

McCullough, Jeffrey S, Stephen T Parente, and Robert Town, "Health Information Technology and Patient Outcomes: The Role of Information and Labor Coordination," *RAND Journal of Economics*, 2016, *47* (1), 207–236.

Miller, Amalia and Catherine Tucker, "Can Health Care Information Technology Save Babies?," *Journal of Political Economy*, 2011, *119* (2), 289–324.

Ministry of Health, *Public Hospital Classification Standard*, Beijing,China, 1989.

Moscone, Francesco, Elisa Tosetti, and Giorgio Vittadini, "Social interaction in Patients Hospital Choice: Evidence from Italy," *Journal of the Royal Statistical Society*, 2012, *175* (2), 453–472.

National Bureau of Statistics, *China Health Statistical Yearbook 2015 (Chinese Edition)*, China Statistics Press, 2015.

Prgomet, Mirela, Andrew Georgiou, and Johanna I Westbrook, "The Impact of Mobile Handheld Technology on Hospital Physicians' Work Practices and Patient Care: A Systematic Review," *Journal of the American Medical Informatics Association*, 2009, *16* (6), 792–801.

Ramachandran, Ambady, Chamukuttan Snehalatha, Jagannathan Ram, Sundaram Selvam, Mary Simon, Arun Nanditha, Ananth Samith Shetty, Ian F Godsland, Nish Chaturvedi, Azeem Majeed et al., "Effectiveness of Mobile Phone Messaging in Prevention of Type 2 Diabetes by Lifestyle Modification in Men in India: A Prospective, Parallel-Group, Randomised Controlled Trial," *Lancet Diabetes & Endocrinology*, 2013, *1* (3), 191–198.

Schilling, Brian, "The Federal Government Has Put Billions into Promoting Electronic Health Record Use: How Is It Going?," *Quality Matters*, 2011, *June/July.*

Serrano, Christina I and Elena Karahanna, "The Compensatory Interaction Between User Capabiities and Technology Capabilities in Influencing Task Performance: An Empirical Assessment in Telemedicine Consultations.," *MIS Quarterly*, 2016, *40* (3), 597–621.

Shigeoka, Hitoshi, "The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection," *American Economic Review*, 2014, *104* (7), 2152–2184.

Varkevisser, Marco, Stephanie a van der Geest, and Frederik T Schut, "Do Patients Choose Hospitals with High Quality Ratings? Empirical Evidence from the Market for Angioplasty in the Netherlands," *Journal of Health Economics*, 2012, *31* (2), 371–378.

Xu, Jing, Wenxin Wang, Yongbin Li, Juan Zhang, Milena Pavlova, Hua Liu, Ping Yin, and Zuxun Lu, "Analysis of Factors Influencing the Outpatient Workload at Chinese Health Centers," *BMC Health Services Research*, 2010, *10* (1), 151.

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

_ **and William Hsiao**, "Harnessing the Privatisation of China's Fragmented Health-Care Delivery," *Lancet*, 2014, *384* (9945), 805–818.

TABLE 1
Descriptive Statistics Of Hospital Operations Data

	Mean	Mean	Mean	Difference
Panel A: Adopter-Non-adopter Comparison for All Hospitals				
Variable	All	Non-adopter	Adopter	Adopter – Non-adopter
All Hospitals (22 Hospitals)				
Daily Consultations	1,634	1,498	1,846	348
Daily Registrations	1,887	1,668	2,228	560
Daily Cancellation Rate	0.120	0.103	0.146	0.043
Observations	29,731	18,132	11,599	
% of Days with App Launched	0.171		0.439	
Tier-three Hospitals (8 Hospitals)				
Daily Consultations	2,766	2,719	2,813	94
Daily Registrations	3,196	2,986	3,413	427
Daily Cancellation Rate	0.134	0.091	0.178	0.088
Observations	10,976	5,579	5,397	
Tier-two Hospitals (14 Hospitals)				
Daily Consultations	971	955	1,003	48
Daily Registrations	1,120	1,082	1,196	114
Daily Cancellation Rate	0.112	0.109	0.118	0.009
Observations	18,755	12,553	6,202	
Panel B: Before-After Comparison for Adopters				
Variable	Adopter	Before	After	After – Before
All Adopters (9 Hospitals)				
Daily Consultations	1,846	1,704	2,026	322
Daily Registrations	2,228	2,153	2,323	170
Daily Cancellation Rate	0.146	0.170	0.116	-0.054
Observations	11,599	6,509	5,090	
Tier-three Adopters (4 Hospitals)				
Daily Consultations	2,813	2,582	3,142	559
Daily Registrations	3,413	3,311	3,559	248
Daily Cancellation Rate	0.178	0.219	0.120	-0.099
Observations	5,397	3,167	2,230	
Tier-two Adopters (5 Hospitals)				
Daily Consultations	1,003	872	1,157	285
Daily Registrations	1,196	1,056	1,360	304
Daily Cancellation Rate	0.118	0.123	0.112	-0.011
Observations	6,202	3,342	2,860	

Notes: All descriptive statistics are tabulated from hospital-level daily operations data, which contain 22 hospitals. Emergency room admissions are excluded. Panel A compares non-adopters and adopters, while Panel B compares the before and after for adopters. Statistics for tier-three and tier-two hospitals are listed separately. See Section 3 for details on variable definitions.

TABLE 2
Benchmark Results

	(1)	(2)	(3)	(4)	(5)	(6)
		Total			Offline	
Dependent Variable	Consultations	Registrations	Cancellation Rate	Consultations	Registrations	Cancellation Rate
Panel A: Non-Emergency Departments						
App	0.095***	0.048*	-0.034***	0.005	-0.044**	-0.036***
	(0.028)	(0.025)	(0.012)	(0.026)	(0.022)	(0.012)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.801	0.817	0.663	0.797	0.813	0.663
Hospitals	22	22	22	22	22	22
Bootstrap p-value[a]	0.010	0.081	0.020	0.455	0.076	0.005
Month & Year FE	YES	YES	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES	YES	YES

		Total	
Dependent Variable	Consultations	Registrations	Cancellation Rate
Panel B: Emergency Room			
App	0.008	0.006	-0.001
	(0.068)	(0.057)	(0.016)
Observations	19,732	19,732	19,732
R-squared	0.894	0.903	0.519
Hospitals	15	15	15
Month & Year FE	YES	YES	YES
Hospital FE	YES	YES	YES

Notes: Panel A is for all departments except for the emergency room (ER); Panel B is for the ER. In Panel B, we exclude seven hospitals that either do not have an ER or have fewer than 10 daily ER admissions for more than half of the sample period. Numbers of registrations and consultations are in logarithm. The dependent variable in Column (1) is the log total daily consultations, and in Column (2) is the log total appointments registered through the offline channel, i.e., excluding registrations made through the app. Column (3) is the total appointment cancellation rate calculated as $\frac{\text{total registrations} - \text{total consultations}}{\text{total registrations}}$. Dependent variables in Column (4) to (6) are offline counterparts. *App* is the dummy that switches to one after the app is launched in treatment hospitals. See detailed definitions in Section 3. Robust standard errors are clustered at the hospital level: $***p < 0.01, **p < 0.05, *p < 0.1$

[a]: Bootstrap p-value is computed using the wild bootstrap procedure proposed by MacKinnon and Webb (2018) to deal with few clusters.

TABLE 3
Heterogeneous Effect by Hospital Tier

	(1)	(2)	(3)	(4)	(5)	(6)
		Total			Offline	
Dependent Variable	Consultations	Registrations	Cancellation Rate	Consultations	Registrations	Cancellation Rate
$App \times TierTwo$	0.094**	0.085**	-0.007	0.007	-0.012	-0.015
	(0.040)	(0.041)	(0.011)	(0.027)	(0.031)	(0.012)
$App \times TierThree$	0.095***	0.008	-0.064***	0.003	-0.080***	-0.060***
	(0.035)	(0.015)	(0.021)	(0.046)	(0.024)	(0.022)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.801	0.817	0.676	0.797	0.813	0.671
Hospitals	22	22	22	22	22	22
Month FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES	YES	YES

Notes: Dependent variables are log total consultations, log total registrations, and total cancellation rate in Columns (1) to (3), and log offline consultations, log offline registrations, and offline cancellation rate in Columns (4) to (6). *App* is the dummy that switches to one after the app is launched in treatment hospitals. $TierThree$ and $TierTwo$ are dummies for tier-three and tier-two hospitals, respectively. Robust standard errors clustered at the hospital level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE 4
Reduction in Scheduled Waiting Period by Hospital Tier

Dependent Variable	(1) Scheduled waiting period (hrs)	(2) Scheduled waiting period (hrs)
New	-2.654***	
	(0.128)	
$New^{TierTwo}$		-5.067***
		(0.200)
$New^{TierThree}$		-0.705***
		(0.192)
Observations	1,705,283	1,705,283
R-squared	0.008	0.008
Number of Individuals	278,909	278,909
Year and Month FE	YES	YES
City FE	YES	YES
Individual FE	YES	YES

Notes: Dependent variables are in-app scheduled waiting period measured in hours. Independent variable "New" measures the number of new apps adopted in the same city since the individual's first appearance in the sample. $New^{TierTwo}$ measures the number of new apps adopted by tier-two hospitals, and $New^{TierThree}$ measures the number of new apps adopted by tier-three hospitals. Robust standard errors clustered at the individual level are in parentheses: $*** \, p < 0.01, ** \; p < 0.05, * \, p < 0.1$

TABLE 5
Heterogeneous Effect by Hospital Tier and by Type of Department

	(1)	(2)	(3)	(4)	(5)	(6)
		Total			Offline	
Dependent Variable	Consultations	Registrations	Cancellation Rate	Consultations	Registrations	Cancellation Rate
Panel A: Departments for Less Severe Conditions						
$App \times TierTwo$	0.156***	0.142***	-0.011	0.069**	0.046**	-0.019*
	(0.040)	(0.033)	(0.010)	(0.033)	(0.023)	(0.010)
$App \times TierThree$	0.045*	0.028	-0.016**	-0.043	-0.059	-0.016**
	(0.026)	(0.024)	(0.008)	(0.041)	(0.038)	(0.008)
Observations	28,085	28,085	28,085	28,085	28,085	28,085
R-squared	0.846	0.848	0.532	0.843	0.845	0.526
Panel B: Departments for More Severe Conditions						
$App \times TierTwo$	0.009	-0.050	-0.043*	-0.008	-0.069	-0.045*
	(0.100)	(0.077)	(0.025)	(0.094)	(0.073)	(0.026)
$App \times TierThree$	0.189***	0.124**	-0.050***	0.146**	0.082	-0.049***
	(0.069)	(0.061)	(0.018)	(0.071)	(0.064)	(0.017)
Observations	25,509	25,509	25,509	25,509	25,509	25,509
R-squared	0.865	0.870	0.693	0.866	0.870	0.689

Notes: Panel A presents results for less severe departments: health promotion, rehabilitation, nutrition, Chinese medicine, ophthalmology, otolaryngology, dermatology, and dentistry. Panel B presents results for more severe departments: endocrinology, cardiology, pulmonology, urology, cardiothoracic surgery, and orthopedics. *App* is the dummy that switches to one after the app is launched in treatment hospitals. $TierThree$ and $TierTwo$ are dummies for tier-three and tier-two hospitals, respectively. All regressions include hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital level are in parentheses: $***\, p < 0.01, **\ p < 0.05, *\, p < 0.1$

TABLE 6
Patient Sorting across Hospitals After New App Adopter

	(1)	(2)	(3)
Panel A: Switching from Tier 2 to Tier 3			
Dependent Variable	Tier 3 Adopter	Tier 3 Adopter	Tier 3 Adopter
$Pre.Severe$	0.137***	0.113***	0.112***
	(0.005)	(0.005)	(0.005)
$Pre.Tier2$	-0.084***	-0.088***	-0.090***
	(0.002)	(0.002)	(0.002)
$Pre.Tier2 \times Severe$	0.047***	0.044***	0.042***
	(0.006)	(0.006)	(0.006)
Observations	132,469	132,469	132,469
R-squared	0.052	0.058	0.059
Demographic Controls	No	Yes	Yes
Month and Year FE	No	No	Yes
Panel B: Switching from Tier 3 to Tier 2			
Dependent Variable	Tier 2	Tier 2	Tier 2
$Pre.NonSevere$	0.110***	0.114***	0.117***
	(0.003)	(0.003)	(0.003)
$Pre.Tier3Adopter$	-0.516***	-0.488***	-0.484***
	(0.003)	(0.004)	(0.004)
$Pre.Tier3Adopter \times NonSevere$	0.066***	0.065***	0.064***
	(0.016)	(0.016)	(0.017)
Observations	132,469	132,469	132,469
R-squared	0.035	0.083	0.091
Demographic Controls	No	Yes	Yes
Month and Year FE	No	No	Yes

Notes: In Panel A, the dependent variable is the indicator for visiting the new tier-three adopter hospital after the app launch. $Pre.Severe$ is the indicator for a severe condition in the previous visit; $Pre.Tier2$ is the indicator for visiting a tier-two hospital in the previous visit. Severe conditions are defined by diagnosis name: cancer, hypertension, diabetes with complications, angina pectoris or myocardial infarction, atherosclerosis or other diseases of the arteries, chronic obstructive pulmonary disease or pneumonia, hepatic failure, acute kidney failure or chronic kidney disease, or traumatic brain injury, mapped from the ambulatory-care-sensitive conditions listed by Shigeoka (2014). In Panel B, the dependent variable is the indicator for visiting the tier-two hospital after the app launch. $Pre.NonSevere$ is the indicator for a nonsevere condition in the previous visit; $Pre.Tier3Adopter$ is the indicator for visiting the tier-three adopter hospital in the previous visit. Non-severe conditions include diagnoses related to ophthalmology, dermatology, dentistry, health promotion, rehabilitation, nutrition, kidney disease, and Chinese medicine. Robust standard errors clustered at the individual level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

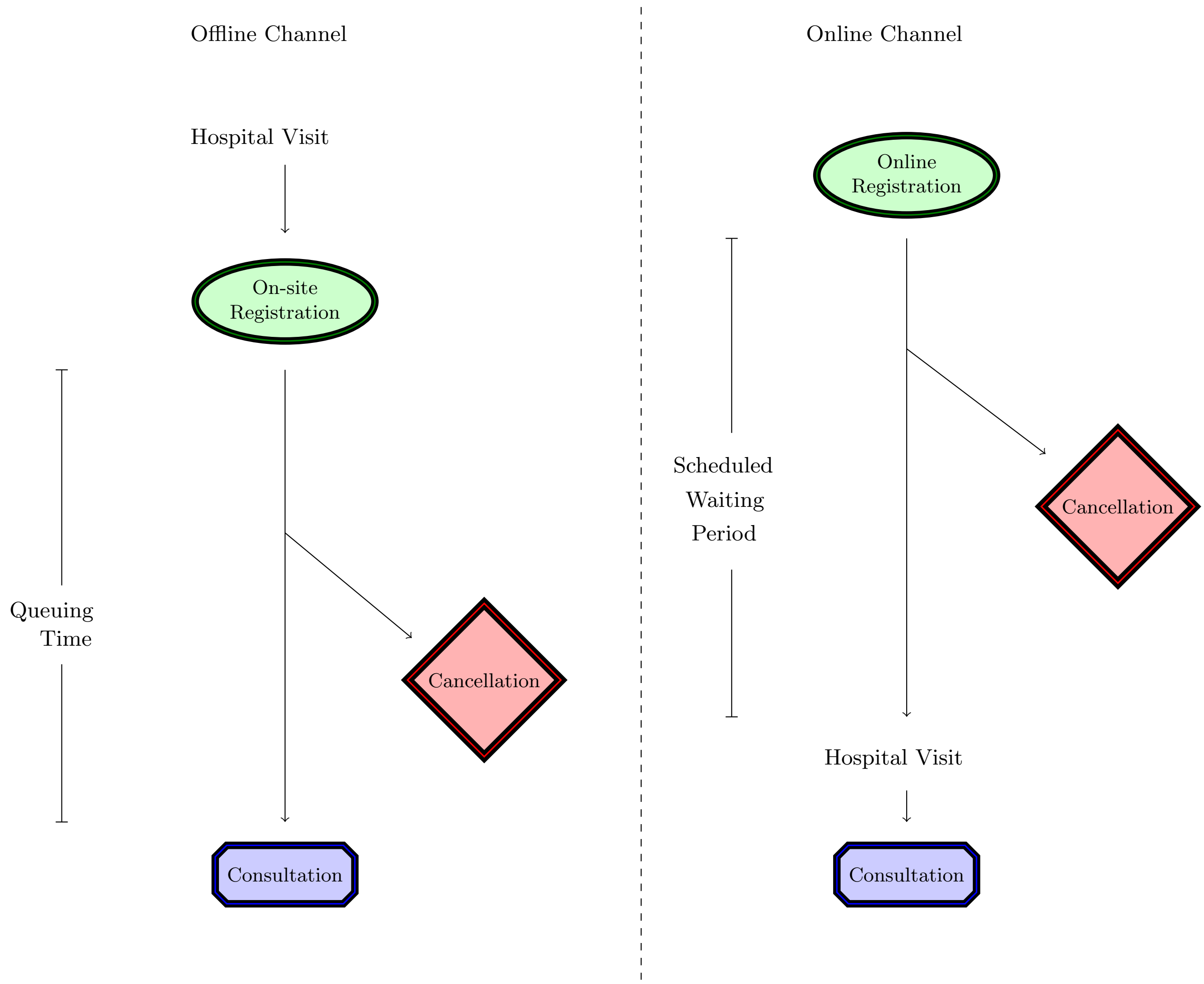


FIGURE 1
Consultation Processes Online and Offline

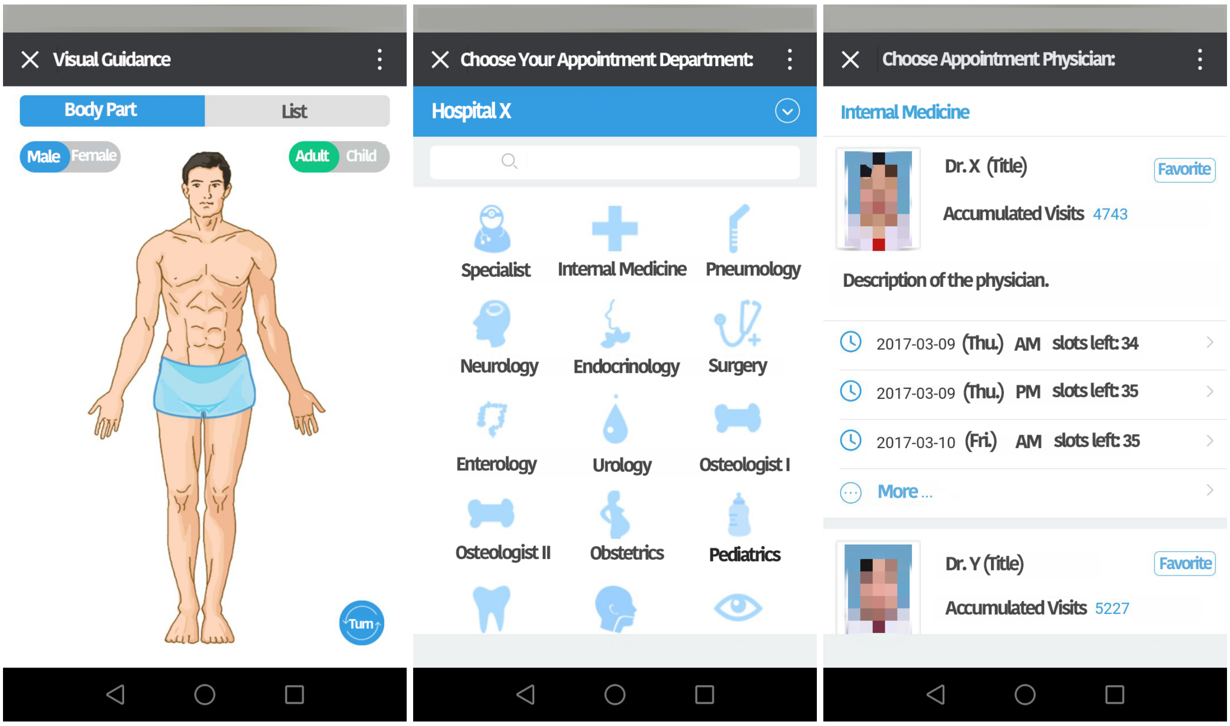


FIGURE 2
Screenshots of the App

Notes: This figure shows the screenshots of the app's basic functionality. The left panel offers visual guidance to help match the patient's symptom to a hospital department. The middle panel guides the patient to the appropriate department. The right panel allows patients to browse on-duty physicians' available hourly slots and make appointments for the next two weeks. English translation is provided by the authors.

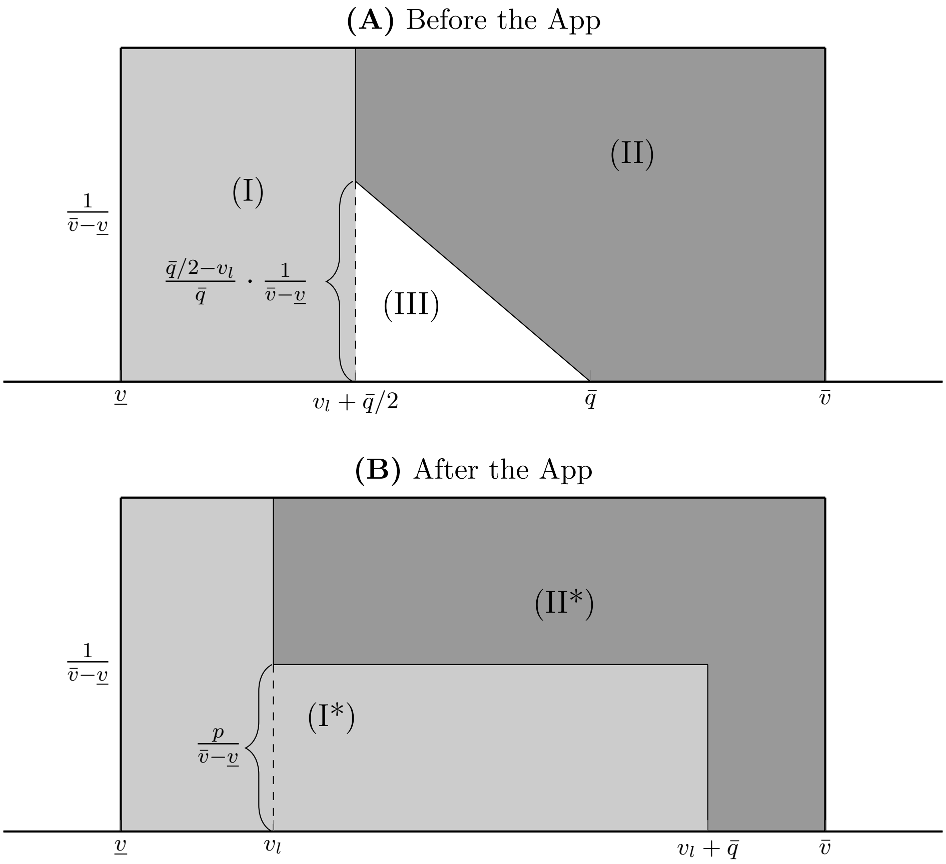
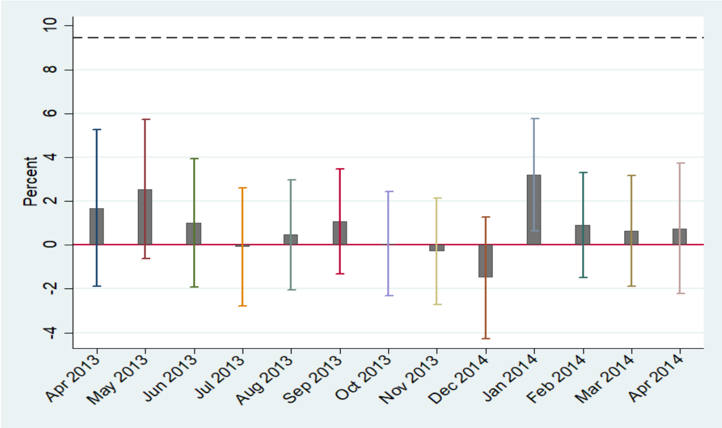


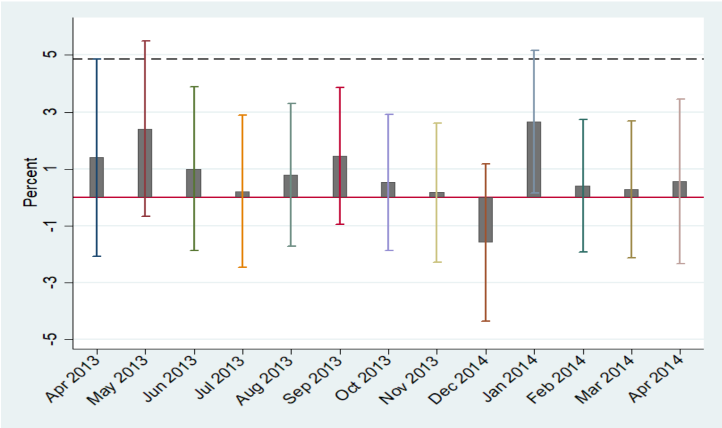
FIGURE 3
Effect of the App on Hospital Choice and Cancellation

Notes: The figure shows the effect of the app on patient's hospital choice and cancellation. Panels (A) and (B) present the probability density function of V_h and patient's hospital choice before and after the app adoption, respectively.
In Panel (A), before the app's launch, utility from visiting a tier-three hospital is uniformly distributed: $V_h \sim U[\underline{v}, \overline{v}]$; and utility from visiting a tier-two hospital is fixed at v_l. Disutility from queuing in a tier-three hospital is uniformly distributed: $Q \sim [0, \overline{q}]$; and that in a tier-two hospital is zero. We plot v_h on the $x-$axis. Those with $v_h - \overline{q}/2 \leq v_l$, or Area (I), choose to visit lower-tier hospitals. The rest choose to visit tier-three hospitals. A patient who visits a tier-three hospital discovers the queuing disutility to be q. She will only remain in the queue to receive the net utility $v_h - q$ if $v_h - q \geq 0$: Areas (II). Otherwise, she will cancel her registration: Area (III). In particular, for a patient with $v_h = v_l + \overline{q}/2$, she will cancel if $v_h - q < 0$, i.e.,$v_l + \overline{q}/2 - q < 0$. The probability of this patient cancelling her registration is $\Pr(q > v_l + \overline{q}/2) = (\overline{q}/2 - v_l)/\overline{q}$.
In Panel (B), after the app's launch, we assume that the app changes Q's distribution to be binomial: With probability p, patients cannot secure their ideal slot ($q = \overline{q}$), and with probability $1 - p$, patients can ($q = 0$). Area (I*) corresponds to patients who visit tier-two hospitals: Those with low valuations of tier-three hospitals ($v_h \leq v_l$) and those who prefer tier-three hospitals to a certain extend ($v_l \leq v_h \leq v_l + \overline{q}$) but face $\overline{q}$ as queuing disutility. Area (II*) corresponds to patients who visit tier-three hospitals. In particular, for any patient with $v_l \leq v_h \leq v_l + \overline{q}$, there is a probability p that she cannot secure ideal slot ($q = \overline{q}$) and chooses to visit the tier-two hospital.
If $1/2 < p < 1/2 + 1/(2\overline{q}^2) * (\overline{q}/2 - v_l)^2$, Area (I*)>Area(I) and Area(II*)>Area(II), i.e., both lower-tier and tier-three hospitals benefit from the app launch.

(A) Total Consultations



(B) Total Registrations



(C) Total Cancellation Rate

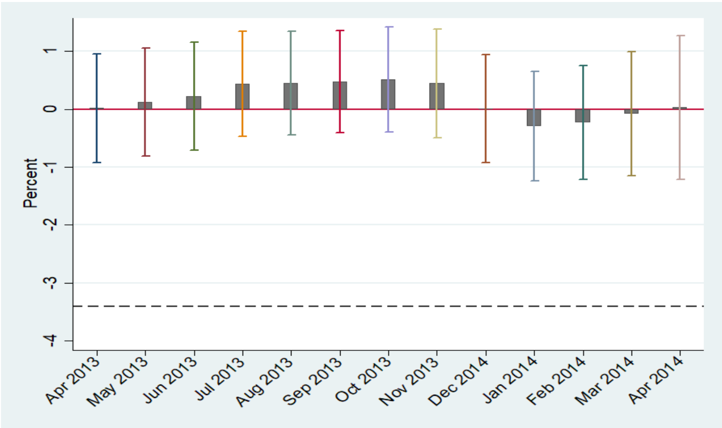
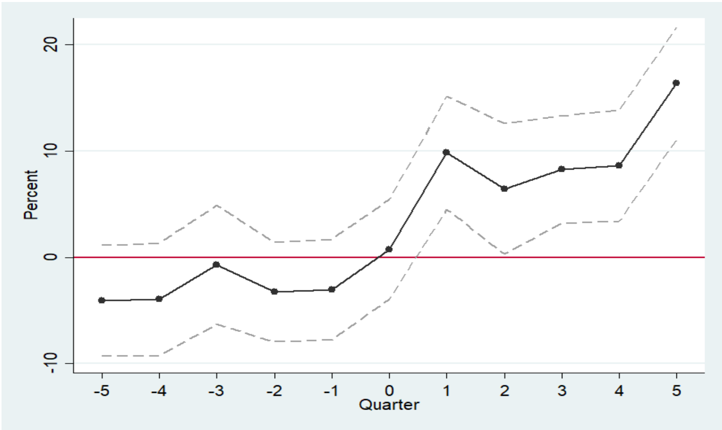


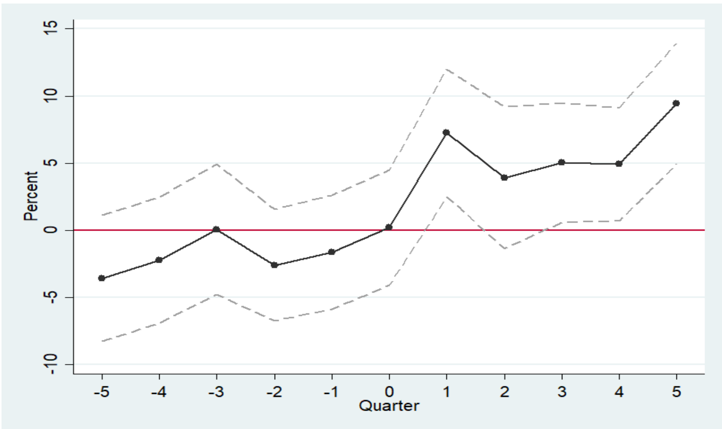
FIGURE 4
Testing Parallel Trends before App Adoption

Notes: The figure tests whether there exist differential pre-existing trends between adopting and non-adopting hospitals in the pre-app period from January 2013 to July 2014. The regression specification is $y_{it} = \alpha + \beta App_{it}^{T} + \lambda_i + \lambda_t + \epsilon_{it}$, where $T \in \{Apr2013, May2013, ..., Apr2014\}$, and App_{it}^{T} is a placebo dummy that switches to one for adopting hospitals after month T. Each shaded bar represents the estimated β from a separate regression, coupled with the the 90% confidence interval. Detailed regression results are reported in Appendix Table A2. Shaded dashed lines represent the estimated effect of actual app adoption in the baseline sample, obtained from Table 2.

(A) Total Consultations



(B) Total Registrations



(C) Total Cancellation Rate

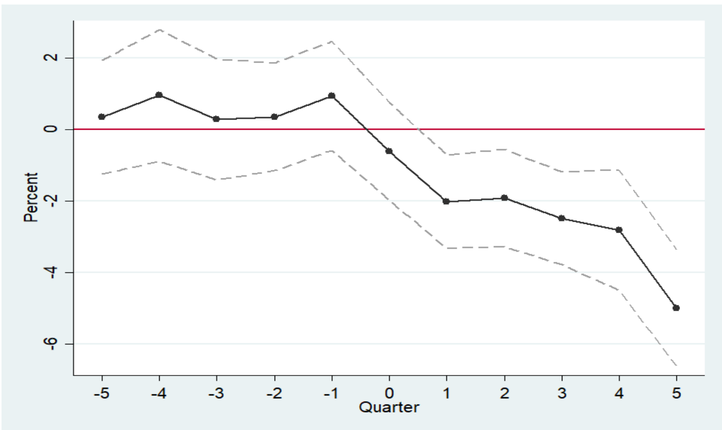
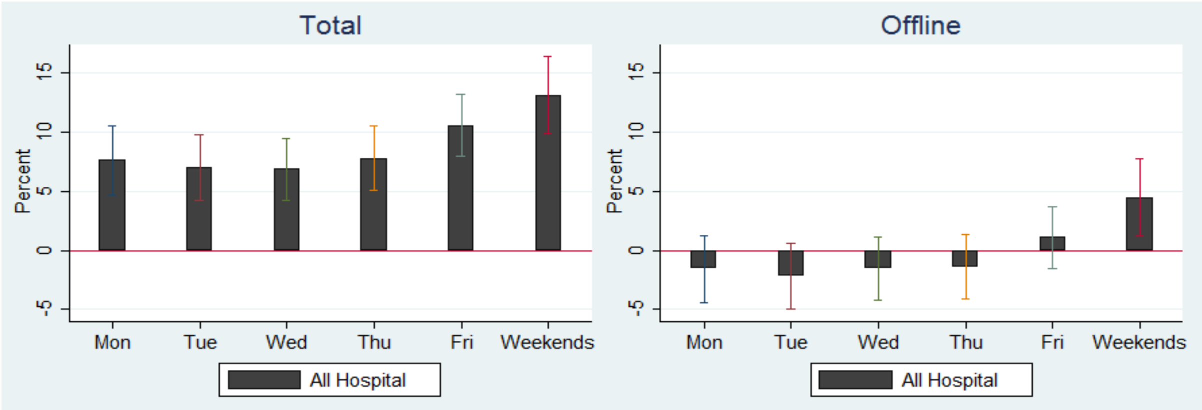


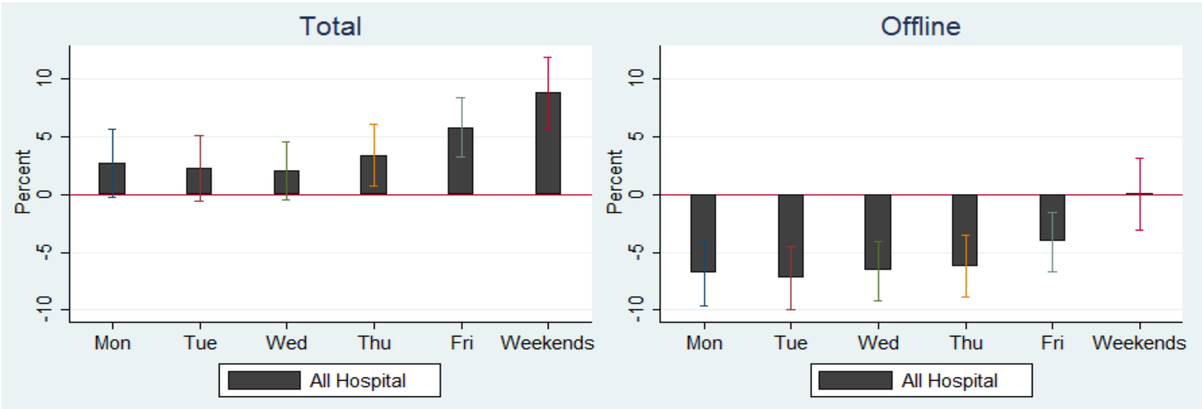
FIGURE 5
Dynamic Response before and after App Adoption

Notes: The figure estimates the dynamic effect of the app on the hospital's total consultations, total registrations, and total cancellation rate. The regression specification is $y_{it} = \alpha + \sum_Q \beta^Q App_{it}^Q + \lambda_i + \lambda_t + \epsilon_{it}$, where $Q \in \{-5, -4, ..., 6\}$. App^{-5} to App^{5} are lag and lead quarter dummies defined for each treatment hospital. App^{0} represents the actual app launch, and App^{6} the 6th quarter and later. All quarter dummies are zero for control hospitals. Solid dots are estimated β^Qs from this regression, coupled with the 90% confidence interval. All regressions include hospital fixed effect, year fixed effect, and month fixed effect. Detailed regression results are reported in Appendix Table A3.

(A) Consultations



(B) Registrations



(C) Cancellation Rate

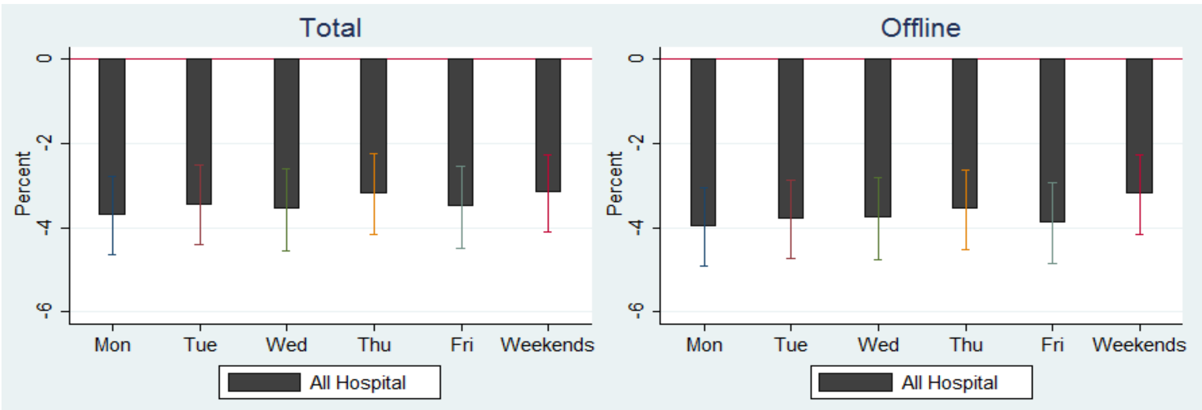
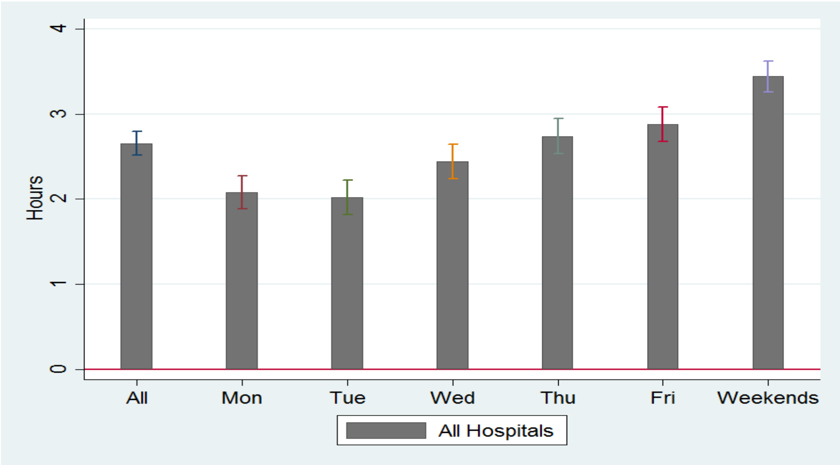


FIGURE 6
Heterogeneous Effect by Days of the Week

Notes: The figure estimates the heterogeneous effect of the app across different days of the week. We run $y_{it} = \alpha + \beta_{DOW} App_{it} \times DOW_t + \beta_{Other} App_{it} \times Other_t + \lambda_i + \lambda_t + \epsilon_{it}$ on hospital-level data, where DOW_t is the dummy for a given day of the week, and $Other_t$ is the dummy for other days of the week. Shaded bars are estimated β_{DOW}, coupled with the 90% confidence interval. Detailed regression results are reported in Appendix Tables A8 (total) and A9 (offline).

(A) By Days of the Week



(B) By Days of the Week and Hospital Tiers

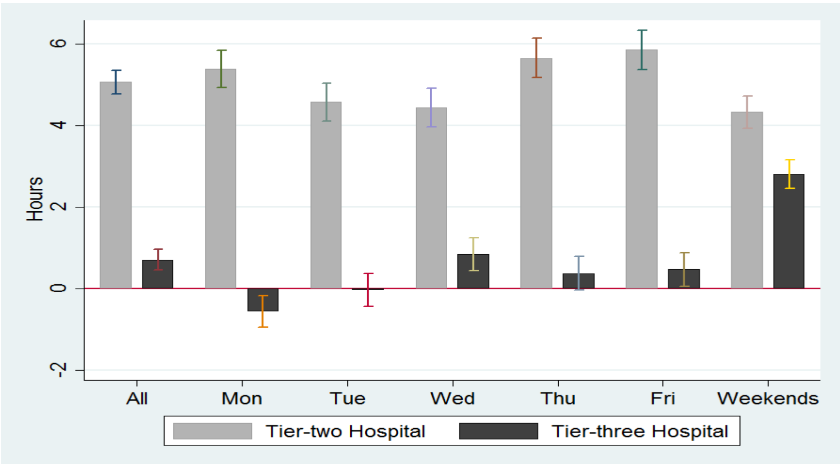
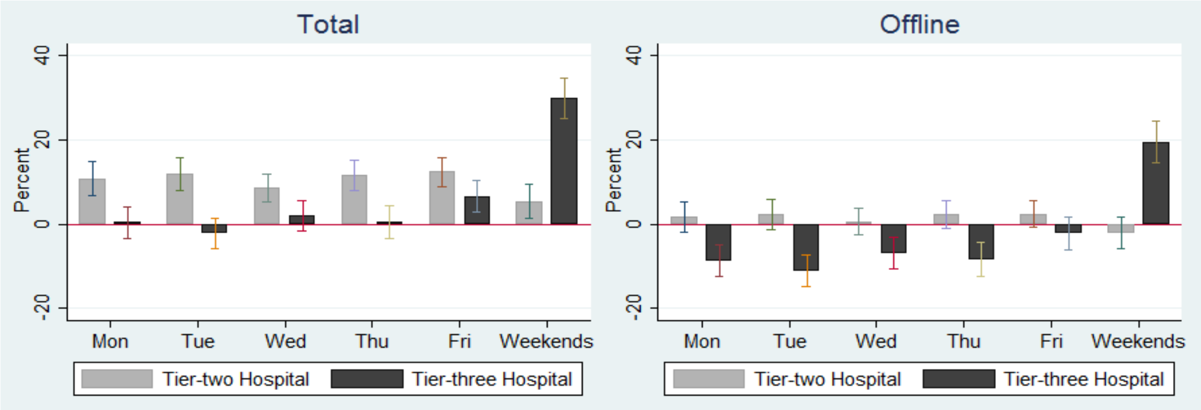


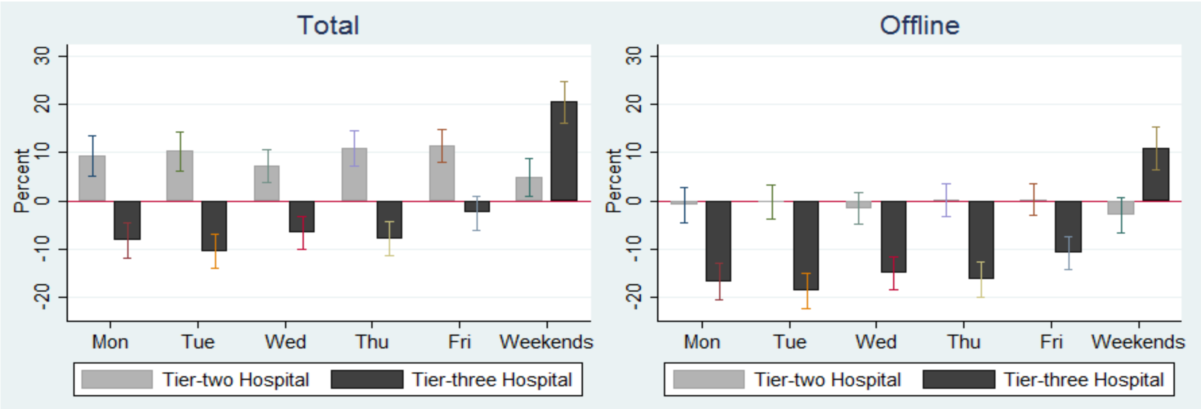
FIGURE 7
Reduction in Scheduled Waiting Period by Days of the Week and Hospital Tier

Notes: This figure estimates the spillover effect of new app adoption on reducing scheduled waiting period across different days of the week. Data contain the longitudinal appointment-booking records of the first batch of app users, which allows us to study the spillover effect of subsequent app-adopters on the incumbent app adopter's first batch of app users. Panel (A) esitmates the average reduction across different days of the week; Panel (B) estimates the average reduction across both days of the week and hospital tiers. In Panel (A), we estimate $w_{ijt} = \alpha + \beta_{DOW} New_{ijt} \times DOW_t + \beta_{Other} New_{ijt} \times Other_t + \lambda_i + \lambda_j + \lambda_t + \epsilon_{ijt}$; in Panel (B), we run $w_{ijt} = \alpha + \beta_1 DOW_t + \beta_2 New_{ijt}^{TierTwo} \times DOW_t + \beta_3 New_{ijt}^{TierThree} \times DOW_t + \beta_4 New_{ijt}^{TierTwo} \times Other_t + \beta_5 New_{ijt}^{TierThree} \times Other_t + \lambda_i + \lambda_j + \lambda_t + \epsilon_{it}$, where w_{ijt} denotes the scheduled waiting period on the mobile waiting list, and New_{ijt} denotes the number of new apps adopted in city j at time t since patient i's first online appointment; $New_{ijt}^{TierTwo}$ and $New_{ijt}^{TierThree}$ are the number adopted by tier-two and tier-three hospitals, respectively. DOW_t is the dummy for a given day of the week, and $Other_t$ is the dummy for other days of the week. Shaded bars are estimated β_{DOW}, coupled with the 90% confidence interval. Detailed regression results are reported in Appendix Tables A10 and A14.

(A) Consultation



(B) Registration



(C) Cancellation Rate

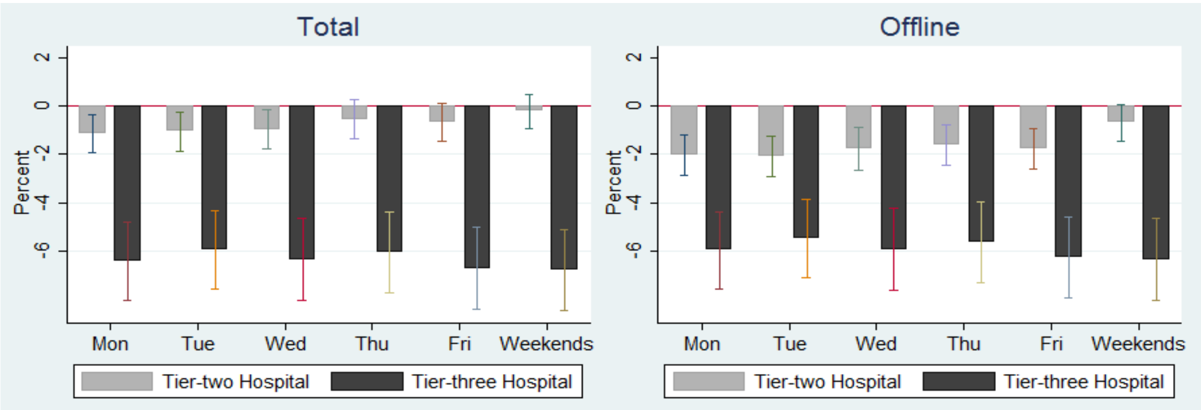


FIGURE 8
Heterogeneous Effect by Days of the Week and Hospital Tier

Notes: This figure estimates the heterogeneous effect of the app across hospitals of different tiers and across different days of the week. We run $y_{it} = \alpha + \beta_1 DOW_t \times TierTwo_i + \beta_2 DOW_t \times TierThree_i + \beta_3 DOW_t \times TierTwo_i \times App_{it} + \beta_4 DOW_t \times TierThree_i \times App_{it} + \beta_5 Other_t \times TierTwo_i \times App_{it} + \beta_6 Other_i \times TierThree_i \times App_{it} + \lambda_i + \lambda_t + \epsilon_{it}$ on hospital-level data, where DOW_t is the day of the week dummy, and $TierThree_i$ is the dummy for tier-three hospitals. Lighter shaded bars are estimated β_3, and darker shaded bars are estimated β_4, both coupled with the 90% confidence interval. Detailed regression results are reported in Appendix Table A11 (consultations), Table A12 (registrations), and Table A13 (cancellation rate).

Information Provision and Streamlined Medical Service:

Evidence from a Mobile Appointment App

Online Appendix

September 5, 2018

Contents

Appendix A The Chinese Healthcare System

China developed its primary-care system after the establishment of the People's Republic of China in 1949. One of the major achievements was an innovative three-tier healthcare system in public hospitals.[1] The first tier generally consists of community health stations and clinics that have fewer than 100 beds and are tasked with providing primary care, preventive care, and rehabilitation services at the community level. The second tier is generally represented by township hospitals in mid-size cities. They are equipped with 100 to 500 beds, and are responsible for more comprehensive health services and medical training for health-workers in tier-one facilities. The third tier contains general hospitals at the city, provincial, or national level with a bed capacity exceeding 500. They provide the most sophisticated acute care and specialist services. They also play a dominant role in medical education and research, and serve as medical hubs for multiple regions.

Backed by government funding, the three-tier system was successful in improving the population's health and life expectancy across the country. Between 1952 and 1982, life expectancy increased from 45 years to 68 years, the infant mortality rate fell from 200 to 34 per 1,000 live births, and longstanding scourges such as schistosomiasis were largely eradicated (Blumenthal and Hsiao, 2005). In 1984, the "China Model" was highly praised by the World Bank and the World Health Organization as an effective model for other developing countries (World Bank, 1984).

Following the economic reforms initiated in December 1978, however, the market-oriented healthcare reforms of the 1980s and 1990s moved the Chinese healthcare system onto a different track (Blumenthal and Hsiao, 2015; Yip and Hsiao, 2008). The reforms gave more autonomy to hospitals and dramatically cut government financing. Government subsidies fell to a mere 10% of a hospital's total revenues by the early 1990s and stayed low ever since (Yip and Hsiao,

1. The three-tier classification of hospitals is based on weighted scores that measure the number of beds, level of service provision, medical technology, medical equipment, and quality of management and medical care. In practice, the three tiers are further subdivided into 3 subsidiary levels (Ministry of Health, 1989).

2008).

Even through the three-tier structure remains, the disparity between tier-three and tier-one hospitals has rapidly widened. Tier-three hospitals have grown quickly in size and captured the lion's share of skilled physicians, patient flow and revenue. For example, in 2014 the average tier-three hospitals in China employed 26 times more physicians and nurses, treated 27 times more patients, and received 60 times more revenue than their tier-one counterparts (National Bureau of Statistics, 2015). (For additional comparisons between Chinese hospitals in different tiers, see Appendix Table A1.) In contrast, lower-tier facilities are increasingly understaffed and underfunded.[2] This has created a downward spiral in the quality and reputation of the lower-tier facilities, and motivates patients to flock to tier-three hospitals regardless of the severity of their illness.

Tier-three hospitals are increasingly overcrowded, and lower-tier hospitals are increasingly underutilized. By 2014, the bed occupancy rate in tier-three hospitals is overwhelmingly 101.8%, in contrast to only 60.1% in tier-one hospitals (National Bureau of Statistics, 2015). In US hospitals, the most commonly targeted bed occupancy rate is 85% (Green, 2006). The problems of waiting and rationing quickly escalated in tier-three hospitals — which has deterred patients with acute conditions from receiving timely care, and drives many patients to forfeit treatment without being seen.

Facing these challenges, the Chinese government launched a nationwide systemic health reform in 2009, pledging to provide more affordable and equitable access to health care for all citizens by 2020. The reform marked a departure from the market-oriented strategy used since 1978, and reinstated the government's role in the financing of health care and provision of public goods (Chen, 2009; Eggleston et al., 2008).[3] Although the 2009 healthcare reform has

2. China has a large shortage of general practitioners (GP): in 2014, there were only 0.13 GPs per 1,000 population. GPs only account for 5.6% of all physicians across China (National Bureau of Statistics, 2015). In contrast, in the UK there are 0.8 GPs per 1,000 population, which account for 28.7% of all physicians (OECD, 2016).

3. The 2009 health reform has five objectives. First, expand public health insurance to gradually cover more than 90% of the Chinese population, including improved coverage for urban residents, the new rural cooperative Medicare scheme for rural residents, and the improved Medicaid scheme for the poor. Second, establish a

made great progress in expanding insurance coverage, much work remains to improve healthcare delivery.

Three long-lasting problems stand out, which also mark the main differences between the Chinese healthcare system and most western healthcare systems: (1) there is no effective referral system in outpatient settings that directs traffic to primary-care or acute-care hospitals. (2) The price differential in registrations fee is too narrow to serve as a screening device to enforce more appropriate use of different levels of healthcare services. (3) The public health insurance system generally does not cover outpatient consultations and contributes little to establish a price differential. As a result, all patients are inclined to visit large hospitals regardless of the severity of illness, which causes overcrowding in large hospitals and underutilization in smaller ones.

Despite lackluster development on the supply side, the demand for health and health services is booming in China, driven by the growing middle class and an aging population. The total number of annual hospital consultations has tripled from 2005 to 2014 (from 51.8 million to 153.7 million), and average hospital revenue has increased almost five times during the same period (from 55.7 million to 273.4 million) (National Bureau of Statistics, 2015). A report by the McKinsey Global Institute predicts that the healthcare spending in China will reach 1 trillion USD by 2020, up from 350 billion in 2012 (Le Deu et al., 2012). Facing this ever-growing demand, improving the healthcare delivery has become the top priority for both the government and the society as a whole to ensure the most effective development of China's healthcare system.

nationwide drug system with dedicated high reimbursement rates for a list of essential drugs to provide an affordable drug supply. Third, provide more public financing and infrastructure support to grassroots health facilities and county hospitals to expand health service network in rural areas and reduce the workload for urban hospitals. Fourth, promote basic public health services. Fifth, launch the pilot reform in public hospitals. See Chen (2009) for more detail.

Appendix B Tables

TABLE A1
Comparison of Hospital Performance by Hospital Tier

Hospital Tier	(1) Tier 3	(2) Tier 2	(3) Tier 1
Number of Hospitals	1954	6850	7009
Patients per Hospital	715478	167458	26363
Physicians per Hospital	341	93	16
Nurses per Hospital	524	129	17
Bed Utilization (percent)	101.8	87.9	60.1
Revenue per Hospital (million CNY)	663	101	11
Expenditure per Hospital (million CNY)	631	97	10
Outpatient Registrations Fee (USD)	2.6	1.3	1

Source: China Health Statistical Yearbook 2015.

TABLE A2
Testing the Parallel Trends Between Adopting and Non-Adopting Hospitals Before the Treatment

	(1) Apr 2013	(2) May 2013	(3) Jun 2013	(4) Jul 2013	(5) Aug 2013	(6) Sep 2013	(7) Oct 2013	(8) Nov 2013	(9) Dec 2013	(10) Jan 2014	(11) Feb 2014	(12) Mar 2014	(13) Apr 2014
Panel A: Total Consultations													
App	0.017	0.025	0.010	-0.001	0.005	0.011	0.001	-0.003	-0.015	0.032**	0.009	0.006	0.007
	(0.022)	(0.019)	(0.018)	(0.016)	(0.015)	(0.015)	(0.014)	(0.015)	(0.017)	(0.016)	(0.014)	(0.015)	(0.018)
Observations	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549
R-squared	0.789	0.719	0.687	0.714	0.658	0.855	0.659	0.788	0.665	0.778	0.712	0.841	0.689
Panel B: Total Registrations													
App	0.014	0.024	0.010	0.002	0.008	0.015	0.005	0.002	-0.016	0.027*	0.004	0.003	0.005
	(0.021)	(0.019)	(0.017)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.017)	(0.015)	(0.014)	(0.015)	(0.018)
Observations	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549
R-squared	0.811	0.745	0.854	0.749	0.789	0.774	0.811	0.803	0.889	0.814	0.789	0.745	0.800
Panel C: Total Cancellation Rates													
App	0.000	0.001	0.002	0.004	0.004	0.005	0.005	0.004	0.000	-0.003	-0.002	-0.001	0.000
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)
Observations	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549	11,549
R-squared	0.679	0.653	0.670	0.680	0.598	0.650	0.672	0.681	0.785	0.632	0.602	0.599	0.680

Notes: The table tests whether there exist differential pre-existing trends between adopting and non-adopting hospitals in the pre-app period from January 2013 to July 2014. This table reports the regression results for Figure 4. The regression specification is $y_{it} = \alpha + \beta App_{it}^T + \lambda_i + \lambda_t + \epsilon_{it}$, where $T \in \{Apr2013, May2013, ..., Apr2014\}$, and App_{it}^T is a placebo dummy that switches to one for adopting hospitals after month T. All regressions include hospital fixed effects, month fixed effects and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A3
Comparing the Dynamic Response Between Adopting and Non-Adopting Hospitals Before and After the Treatment

VARIABLES	(1) Total Consultations	(2) Total Registrations	(3) Total Cancellation Rates
App^{-5}	-0.036	-0.041	0.003
	(0.029)	(0.032)	(0.010)
App^{-4}	-0.022	-0.040	0.009
	(0.028)	(0.032)	(0.011)
App^{-3}	0.001	-0.007	0.003
	(0.029)	(0.034)	(0.010)
App^{-2}	-0.026	-0.033	0.003
	(0.025)	(0.028)	(0.009)
App^{-1}	-0.016	-0.030	0.009
	(0.026)	(0.029)	(0.009)
App^{1}	0.002	0.008	-0.006
	(0.026)	(0.028)	(0.008)
App^{2}	0.073**	0.098***	-0.020**
	(0.029)	(0.032)	(0.008)
App^{3}	0.039	0.064*	-0.019**
	(0.032)	(0.037)	(0.008)
App^{4}	0.050*	0.082***	-0.025***
	(0.027)	(0.031)	(0.008)
App^{5}	0.049*	0.086***	-0.028***
	(0.026)	(0.032)	(0.010)
App^{6}	0.095***	0.163***	-0.050***
	(0.027)	(0.032)	(0.010)
Observations	28,832	28,832	28,832
R-squared	0.814	0.800	0.734

Notes: The table estimates the dynamic effect of the app on the hospital's total consultations, total registrations, and total cancellation rates. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Figure 5. The regression specification is $y_{it} = \alpha + \sum_Q \beta^Q App_{it}^Q + \lambda_i + \lambda_t + \epsilon_{it},$ where $Q \in \{-5, -4, ..., 6\}$ where App^Q is the dummy indicating the Qth quarter from the actual app launch. App^1 is the dummy for the first three months after the app launch, and App^6 is the dummy for all subsequent months after the 6th quarter following the actual app launch. All quarterly lags and leads remain zero for control hospitals. All regressions include hospital fixed effects, month fixed effects and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $*** \ p < 0.01, ** \ p < 0.05, * \ p < 0.1$

TABLE A4
Robustness Analysis

VARIABLES	(1) Total Consultations	(2) Total Registrations	(3) Total Cancellation Rate	(4) Offline Consultations	(5) Offline Registrations	(6) Offline Cancellation Rate
Robustness 1: Subsample of adopter hospitals						
App	0.093***	0.050**	-0.034***	0.026	-0.020	-0.037***
	(0.021)	(0.020)	(0.008)	(0.020)	(0.019)	(0.008)
Observations	11,630	11,630	11,630	11,630	11,630	11,630
R-squared	0.887	0.904	0.754	0.887	0.905	0.759
Robustness 2: Adding city-specific time trend						
App	0.094***	0.048***	-0.034***	0.005	-0.044***	-0.036***
	(0.014)	(0.014)	(0.006)	(0.014)	(0.013)	(0.006)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.801	0.817	0.664	0.797	0.813	0.663
Robustness 3: Adding interaction between pre-app cancellation rates and time trend						
App	0.083***	0.055***	-0.021***	-0.005	-0.035***	-0.023***
	(0.014)	(0.014)	(0.005)	(0.014)	(0.014)	(0.005)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.801	0.817	0.724	0.797	0.813	0.727
Robustness 4: Adding linear and quadratic time trend						
App	0.096***	0.050***	-0.035***	0.008	-0.042***	-0.037***
	(0.014)	(0.013)	(0.006)	(0.014)	(0.013)	(0.006)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.801	0.817	0.664	0.797	0.814	0.664
Robustness 5: Subsample of non-specialty hospitals						
App	0.107***	0.055***	-0.038***	0.015	-0.040***	-0.041***
	(0.015)	(0.014)	(0.006)	(0.014)	(0.014)	(0.006)
Observations	26,970	26,970	26,970	26,970	26,970	26,970
R-squared	0.826	0.841	0.664	0.822	0.838	0.664
Robustness 6: Subsample of weekdays observations						
App	0.099***	0.054***	-0.034***	0.008	-0.041***	-0.037***
	(0.014)	(0.014)	(0.006)	(0.014)	(0.014)	(0.006)
Observations	21,232	21,232	21,232	21,232	21,232	21,232
R-squared	0.862	0.875	0.671	0.860	0.873	0.673
Robustness 7: Subsample of provincial capital hospitals						
App	0.082***	0.040***	-0.032***	0.012	-0.028*	-0.030***
	(0.016)	(0.015)	(0.008)	(0.017)	(0.016)	(0.008)
Observations	20,252	20,252	20,252	20,252	20,252	20,252
R-squared	0.805	0.812	0.369	0.800	0.807	0.375
Robustness 8: Subsample of balanced panel						
App	0.087***	0.054***	-0.024***	0.002	-0.034**	-0.027***
	(0.014)	(0.013)	(0.005)	(0.014)	(0.013)	(0.005)
Observations	25,126	25,126	25,126	25,126	25,126	25,126
R-squared	0.812	0.826	0.681	0.808	0.822	0.678

Notes: The table checks the robustness by re-estimating Equation 1 with modified specifications. All regressions include hospital fixed effects, month fixed effects and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A5
Heterogeneous Effect of Patient Sorting From Tier 2 to Tier 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: Whether visiting the tier-three adopter hospital										
Pre. Severe	0.112***	0.098***	0.094***	0.085***	0.098***	0.093***	0.097***	0.105***	0.087***	0.049***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)	(0.010)
Pre. Tier 2	-0.090***	-0.080***	-0.081***	-0.082***	-0.076***	-0.086***	-0.081***	-0.077***	-0.082***	-0.075***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Pre. Tier 2 and Severe	0.042***	0.036***	0.037***	0.046***	0.024***	0.043***	0.036***	0.028***	0.038***	0.039***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.006)	(0.007)	(0.013)
Observations	132,469	132,469	132,469	132,469	132,469	132,469	132,469	132,469	132,469	132,469
R-squared	0.059	0.049	0.049	0.047	0.042	0.051	0.049	0.052	0.048	0.036
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis Category	Baseline	1	2	3	4	5	6	7	8	9

Notes: This table reestimate Equation 5 by defining the severe conditions at nine sub-catogories. The first column replicate the baseline results in Panel A Column (3) in Table 6. The diagnosis catogories of severe conditions are: (1) cancer, (2) hypertension, (3) diabetes with complications, (4) angina pectoris and myocardial infarction, (5) atherosclerosis and other diseases of arteries, (6) chronic obstructive pulmonary disease and pneumonia, (7) hepatic failure, (8) acute kidney failure and chronic kidney disease, and (9) traumatic brain injury. See detailed regression specifications in Section 5.2 of the main text. Demographic controls include gender, age, and ID-registry location dummy. Time dummies include monthly and yearly time dummies for both pre- and post-launch visits. Robust standard errors are in parentheses: $***\, p < 0.01, **\, p < 0.05, *\, p < 0.1$

TABLE A6
Heterogeneous Effect of Patient Sorting From Tier 3 to Tier 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable: Whether visiting tier-two hospitals									
Pre. NonSevere	0.120***	0.032***	0.097***	0.084***	0.034***	0.062***	0.082***	0.073***	0.097***
	(0.003)	(0.006)	(0.007)	(0.007)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)
Pre. Tier3 Adopter	-0.489***	-0.482***	-0.478***	-0.481***	-0.484***	-0.493***	-0.480***	-0.480***	-0.479***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Pre. Tier3 Adopter and NonSevere	0.061***	0.103***	0.054	0.093**	0.106***	0.133***	0.092**	0.091**	0.045
	(0.016)	(0.031)	(0.041)	(0.038)	(0.029)	(0.024)	(0.041)	(0.041)	(0.036)
Observations	132,469	132,469	132,469	132,469	132,469	132,469	132,469	132,469	132,469
R-squared	0.090	0.081	0.082	0.082	0.081	0.082	0.081	0.081	0.082
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis Category	Baseline	1	2	3	4	5	6	7	8

Notes: This table reestimate Equation 6 by defining the nonsevere conditions at eight sub-catogories. The first column replicate the baseline results in Panel B Column (3) in Table 6. The diagnosis catogories of nonsevere conditions are: (1) ophthalmology, (2) otolaryngology, (3) dermatology, (4) dentistry, (5) health promotion, (6) rehabilitation, (7) nutrition, and (8) Chinese medicine. Demographic controls include gender, age, and ID-registry location dummy. Time dummies include monthly and yearly time dummies for both pre- and post-launch visits. See detailed regression specifications in Section 5.2 of the main text. Robust standard errors are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A7
Descriptive Statistics By Days of the Week Before App Launch

VARIABLES	**Daily Consultations**		**Daily Registrations**		**Daily Cancellation Rate**	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Monday	1,865	1,600	2,200	1,838	0.134	0.0904
Tuesday	1,632	1,439	1,902	1,625	0.128	0.0843
Wednesday	1,558	1,326	1,830	1,520	0.130	0.0891
Thursday	1,538	1,294	1,797	1,487	0.126	0.0840
Friday	1,547	1,310	1,803	1,499	0.125	0.0800
Saturday	1,320	1,070	1,549	1,258	0.120	0.0887
Sunday	1,002	699.4	1,174	877.2	0.116	0.0908
Weekdays	1,630	1,405	1,909	1,608	0.129	0.0857
Weekends	1,160	915.9	1,360	1,099	0.118	0.0898

Notes: This table shows hospital summary statistics broken down by days of the week for the year 2013.

TABLE A8
Heterogeneous Effect by Days of the Week (Part 1: Total)

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
Panel A: Log Total Consultations						
DOW	0.229***	0.127***	0.070***	0.039***	0.046***	-0.306***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)	(0.012)
App × *DOW*	0.076***	0.070***	0.069***	0.078***	0.105***	0.131***
	(0.018)	(0.017)	(0.016)	(0.017)	(0.016)	(0.020)
App × *Other*	0.098***	0.099***	0.099***	0.097***	0.093***	0.079***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.811	0.804	0.802	0.801	0.801	0.831
Panel B: Log Total Registrations						
DOW	0.234***	0.134***	0.074***	0.041***	0.051***	-0.321***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)	(0.011)
App × *DOW*	0.027	0.023	0.021	0.034**	0.058***	0.088***
	(0.018)	(0.017)	(0.015)	(0.016)	(0.016)	(0.019)
App × *Other*	0.052***	0.053***	0.053***	0.051***	0.047***	0.032**
	(0.014)	(0.013)	(0.014)	(0.013)	(0.014)	(0.015)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.827	0.820	0.818	0.817	0.817	0.848
Panel C: Total Appointment Cancellation Rate						
DOW	0.005***	0.006***	0.004***	0.002***	0.005***	-0.013***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
App × *DOW*	-0.037***	-0.035***	-0.036***	-0.032***	-0.035***	-0.032***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
App × *Other*	-0.034***	-0.034***	-0.034***	-0.034***	-0.034***	-0.035***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.664	0.664	0.664	0.663	0.664	0.669

Notes: This table reports the estimated effects of the app on hospital's total consultations (Panel A), total registrations (Panel B), and total cancellation rate (Panel C), by days of the week. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Figure 6. Dependent variables are in logarithmic terms. Independent variable *DOW* is dummy for a given day of the week, and *Other* is dummy for other days of the week. All regressions include the hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A9
Heterogeneous Effect by Days of the Week (Part 2: Offline)

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
Panel D: Log Offline Consultations						
DOW	0.229***	0.127***	0.070***	0.039***	0.046***	-0.306***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)	(0.012)
App × *DOW*	-0.016	-0.022	-0.015	-0.014	0.011	0.045**
	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)	(0.020)
App × *Other*	0.008	0.010	0.008	0.008	0.004	-0.012
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.807	0.800	0.798	0.797	0.797	0.828
Panel E: Log Offline Registrations						
DOW	0.234***	0.134***	0.074***	0.041***	0.051***	-0.321***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)	(0.011)
App × *DOW*	-0.068***	-0.073***	-0.066***	-0.062***	-0.041***	0.001
	(0.017)	(0.016)	(0.015)	(0.016)	(0.015)	(0.019)
App × *Other*	-0.040***	-0.039***	-0.040***	-0.041***	-0.044***	-0.062***
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.015)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.824	0.817	0.814	0.814	0.814	0.845
Panel F: Offline Appointment Cancellation Rate						
DOW	0.005***	0.006***	0.004***	0.002***	0.005***	-0.013***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
App × *DOW*	-0.040***	-0.038***	-0.038***	-0.036***	-0.039***	-0.032***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
App × *Other*	-0.036***	-0.036***	-0.036***	-0.037***	-0.036***	-0.038***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.664	0.664	0.663	0.663	0.664	0.669

Notes: This table reports the estimated effects of the app on hospital's offline consultations (Panel D), offline registrations (Panel E), and offline cancellation rate (Panel F), by days of the week. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Figure 6. Dependent variables are in logarithmic terms. Independent variable *DOW* is dummy for a given day of the week, and *Other* is dummy for other days of the week. All regressions include hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A10
Reduction in Waiting Time by Days of the Week

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
DOW	1.096***	-0.092	0.962***	1.368***	0.592***	-3.300***
	(0.139)	(0.151)	(0.157)	(0.145)	(0.142)	(0.122)
New × *DOW*	-2.081***	-2.022***	-2.441***	-2.737***	-2.882***	-3.440***
	(0.153)	(0.167)	(0.165)	(0.164)	(0.157)	(0.140)
New × *Other*	-2.771***	-2.771***	-2.693***	-2.637***	-2.613***	-2.404***
	(0.129)	(0.128)	(0.128)	(0.128)	(0.129)	(0.133)
Observations	1,705,283	1,705,283	1,705,283	1,705,283	1,705,283	1,705,283
R-squared	0.008	0.008	0.008	0.008	0.008	0.009
Number of Individuals	278,909	278,909	278,909	278,909	278,909	278,909
Year and Month FE	YES	YES	YES	YES	YES	YES
City FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimated effects of subsequent app launches on the waiting time of existing app users by days of the week. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Figure 7. Dependent variables are waiting time on mobile waiting list measured in hours. Independent variable "New" measures the number of new apps adopted in the same city since the individual's first appearance in the sample. *DOW* is the day of the week dummy for Monday to Friday and weekends, respectively, from Columns (1) to (6). *Other* is the dummy for days of the week other than DOW. Robust standard errors clustered at the individual level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A11
Heterogeneous Effect by Days of the Week and by Hospital Tier
(Part 1: Total and Offline Consultations)

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
Panel A: Log Total Consultations						
$DOW \times TierTwo$	0.162***	0.057***	0.009***	-0.013***	0.010**	-0.136***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
$DOW \times TierThree$	0.350***	0.252***	0.179***	0.134***	0.111***	-0.616***
	(0.009)	(0.012)	(0.009)	(0.010)	(0.009)	(0.019)
$DOW \times TierTwo \times App$	0.108***	0.117***	0.085***	0.116***	0.123***	0.053**
	(0.025)	(0.024)	(0.020)	(0.022)	(0.020)	(0.025)
$DOW \times TierThree \times App$	0.004	-0.023	0.020	0.005	0.066***	0.297***
	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)	(0.029)
$Other \times TierTwo \times App$	0.092***	0.090***	0.095***	0.090***	0.089***	0.109***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)
$Other \times TierThree \times App$	0.110***	0.115***	0.108***	0.110***	0.100***	0.014
	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)	(0.019)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.813	0.805	0.803	0.802	0.802	0.846
Panel B: Log Offline Consultations						
$DOW \times TierTwo$	0.162***	0.057***	0.009***	-0.013***	0.010**	-0.136***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
$DOW \times TierThree$	0.350***	0.253***	0.179***	0.134***	0.111***	-0.616***
	(0.009)	(0.012)	(0.009)	(0.010)	(0.009)	(0.019)
$DOW \times TierTwo \times App$	0.016	0.023	0.006	0.022	0.024	-0.021
	(0.022)	(0.021)	(0.019)	(0.020)	(0.019)	(0.023)
$DOW \times TierThree \times App$	-0.087***	-0.110***	-0.069***	-0.084***	-0.022	0.194***
	(0.023)	(0.023)	(0.023)	(0.025)	(0.023)	(0.030)
$Other \times TierTwo \times App$	0.006	0.005	0.007	0.005	0.005	0.018
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.019)
$Other \times TierThree \times App$	0.017	0.021	0.015	0.017	0.007	-0.075***
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.809	0.801	0.799	0.798	0.798	0.844

Notes: This table reports the estimated effects of the app by days of the week and by hospital tiers. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Panel (A) of Figure 8. Dependent variables are log total consultations in Panel (A) and log offline consultations in Panel (B). *DOW* is the dummy for a given day of the week, i.e. Monday to Friday and weekends, respectively, from Columns (1) to (6). *Other* is the dummy for days of the week other than *DOW*. *App* is the dummy that switches to one after the app is launched in treatment hospitals. $TierThree$ and $TierTwo$ are the dummies for tier-three and tier-two hospitals, respectively. All regressions include the hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A12
Heterogeneous Effect by Days of the Week and by Hospital Tier
(Part 2: Total and Offline Registrations)

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
Panel A: Log Total Registrations						
DOW × *TierTwo*	0.172***	0.068***	0.016***	-0.009**	0.015***	-0.158***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
DOW × *TierThree*	0.347***	0.253***	0.179***	0.132***	0.117***	-0.617***
	(0.009)	(0.011)	(0.009)	(0.010)	(0.009)	(0.018)
DOW × *TierTwo* × *App*	0.094***	0.104***	0.072***	0.108***	0.114***	0.049**
	(0.025)	(0.025)	(0.021)	(0.022)	(0.021)	(0.024)
DOW × *TierThree* × *App*	-0.081***	-0.105***	-0.066***	-0.078***	-0.025	0.206***
	(0.022)	(0.021)	(0.020)	(0.021)	(0.021)	(0.026)
Other × *TierTwo* × *App*	0.083***	0.081***	0.087***	0.081***	0.080***	0.098***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)
Other × *TierThree* × *App*	0.023	0.027	0.021	0.023	0.014	-0.072***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.828	0.821	0.819	0.818	0.818	0.861
Panel B: Log Offline Registrations						
DOW × *TierTwo*	0.172***	0.068***	0.016***	-0.009**	0.015***	-0.158***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
DOW × *TierThree*	0.347***	0.253***	0.179***	0.132***	0.117***	-0.617***
	(0.009)	(0.011)	(0.009)	(0.010)	(0.009)	(0.018)
DOW × *TierTwo* × *App*	-0.009	-0.003	-0.016	0.003	0.003	-0.030
	(0.023)	(0.022)	(0.020)	(0.021)	(0.020)	(0.022)
DOW × *TierThree* × *App*	-0.167***	-0.187***	-0.151***	-0.162***	-0.108***	0.108***
	(0.023)	(0.022)	(0.021)	(0.023)	(0.021)	(0.027)
Other × *TierTwo* × *App*	-0.012	-0.013	-0.011	-0.014	-0.014	-0.005
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.020)
Other × *TierThree* × *App*	-0.065***	-0.062***	-0.068***	-0.066***	-0.075***	-0.156***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.825	0.818	0.815	0.814	0.814	0.859

Notes: This table reports the estimated effects of the app by days of the week and by hospital tiers. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Panel (B) of Figure 8. Dependent variables are log total registrations in Panel (A) and log offline registrations in Panel (B). *DOW* is the dummy for a given day of the week, i.e. Monday to Friday and weekends, respectively, from Columns (1) to (6). *Other* is the dummy for days of the week other than *DOW*. *App* is the dummy that switches to one after the app is launched in treatment hospitals. *TierThree* and *TierTwo* are the dummies for tier-three and tier-two hospitals, respectively. All regressions include the hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A13
Heterogeneous Effect by Days of the Week and by Hospital Tier
(Part 3: Total and Offline Cancellation Rate)

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
Panel A: Log Total Cancellation Rate						
$DOW \times TierTwo$	0.009***	0.010***	0.006***	0.004***	0.004***	-0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$DOW \times TierThree$	-0.001	0.000	0.000	-0.002**	0.005***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$DOW \times TierTwo \times App$	-0.011**	-0.011**	-0.010**	-0.005	-0.007	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
$DOW \times TierThree \times App$	-0.064***	-0.059***	-0.064***	-0.061***	-0.067***	-0.068***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
$Other \times TierTwo \times App$	-0.006	-0.006	-0.006	-0.007	-0.007	-0.009*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
$Other \times TierThree \times App$	-0.064***	-0.065***	-0.064***	-0.065***	-0.064***	-0.063***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.677	0.677	0.676	0.676	0.676	0.684
Panel B: Log Offline Cancellation Rate						
$DOW \times TierTwo$	0.009***	0.010***	0.006***	0.004***	0.004***	-0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$DOW \times TierThree$	-0.001	0.000	0.000	-0.002**	0.005***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$DOW \times TierTwo \times App$	-0.020***	-0.021***	-0.018***	-0.016***	-0.018***	-0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$DOW \times TierThree \times App$	-0.060***	-0.055***	-0.059***	-0.057***	-0.063***	-0.063***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
$Other \times TierTwo \times App$	-0.014***	-0.014***	-0.015***	-0.015***	-0.015***	-0.019***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$Other \times TierThree \times App$	-0.060***	-0.061***	-0.060***	-0.061***	-0.060***	-0.059***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.672	0.672	0.671	0.671	0.671	0.679

Notes: This table reports the estimated effects of the app by days of the week and by hospital tiers. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Panel (C) of Figure 8. Dependent variables are log total cancellation rate in Panel (A) and log offline cancellation rate in Panel (B). *DOW* is the dummy for a given day of the week, i.e. Monday to Friday and weekends, respectively, from Columns (1) to (6). *Other* is the dummy for days of the week other than *DOW*. *App* is the dummy that switches to one after the app is launched in treatment hospitals. *TierThree* and *TierTwo* are the dummies for tier-three and tier-two hospitals, respectively. All regressions include hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A14
Reduction in Waiting Time by Hospital Tier and by Days of the Week

	(1) Monday	(2) Tuesday	(3) Wednesday	(4) Thursday	(5) Friday	(6) Weekends
DOW	1.092***	-0.101	0.958***	1.363***	0.601***	-3.310***
	(0.139)	(0.151)	(0.157)	(0.145)	(0.142)	(0.122)
$New^{TierTwo} \times DOW$	-5.392***	-4.575***	-4.445***	-5.654***	-5.858***	-4.329***
	(0.267)	(0.274)	(0.301)	(0.280)	(0.279)	(0.228)
$New^{TierThree} \times DOW$	0.571**	0.044	-0.839***	-0.371	-0.464*	-2.813***
	(0.251)	(0.275)	(0.283)	(0.275)	(0.269)	(0.230)
$New^{TierTwo} \times Other$	-4.974***	-5.155***	-5.170***	-4.959***	-4.937***	-5.194***
	(0.202)	(0.202)	(0.201)	(0.201)	(0.202)	(0.209)
$New^{TierThree} \times Other$	-0.994***	-0.846***	-0.688***	-0.762***	-0.740***	-0.143
	(0.195)	(0.193)	(0.193)	(0.194)	(0.194)	(0.202)
Observations	1,705,283	1,705,283	1,705,283	1,705,283	1,705,283	1,705,283
R-squared	0.008	0.008	0.008	0.008	0.008	0.009
Number of Individuals	278,909	278,909	278,909	278,909	278,909	278,909
Year and Month FE	YES	YES	YES	YES	YES	YES
City FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimated effects of subsequent app launches on the waiting time of existing app users by days of the week and by hospital tiers. The estimated coefficients and the corresponding 90% confidence intervals are plotted in Figure 7. Dependent variables are waiting time on mobile waiting list measured in hours. $New^{TierTwo}$ measures the number of new apps adopted by tier-two hospitals, and $New^{TierThree}$ measures the number of new apps adopted by tier-three hospitals. *DOW* is the dummy for a given day of the week, i.e. Monday to Friday and weekends, respectively, from Columns (1) to (6). *Other* is the dummy for days of the week other than *DOW*. Robust standard errors clustered at the individual level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

TABLE A15
Heterogeneous Effect by Age Groups

	(1)	(2)	(3)	(4)	(5)	(6)
		Total			Offline	
VARIABLES	Consultations	Registrations	Cancellation Rate	Consultations	Registrations	Cancellation Rate
Panel A: Age below 20						
App	0.158***	0.163***	0.004*	0.028	0.029	0.000
	(0.021)	(0.021)	(0.002)	(0.021)	(0.020)	(0.002)
Observations	29,728	29,728	29,728	29,728	29,728	29,728
R-squared	0.846	0.855	0.649	0.846	0.854	0.650
Panel B: Age between 20 to 40						
App	0.105***	0.071***	-0.024***	-0.007	-0.043***	-0.027***
	(0.014)	(0.014)	(0.004)	(0.014)	(0.014)	(0.004)
Observations	29,731	29,731	29,731	29,731	29,731	29,731
R-squared	0.827	0.837	0.739	0.824	0.835	0.737
Panel C: Age between 40 to 60						
App	0.057***	0.016	-0.032***	0.024**	-0.018*	-0.033***
	(0.013)	(0.011)	(0.005)	(0.012)	(0.011)	(0.005)
Observations	29,727	29,727	29,727	29,727	29,727	29,727
R-squared	0.770	0.780	0.716	0.775	0.784	0.715
Panel D: Age above 60						
App	0.083***	0.041***	-0.031***	0.056***	0.017	-0.032***
	(0.015)	(0.013)	(0.006)	(0.014)	(0.012)	(0.006)
Observations	29,717	29,717	29,717	29,717	29,717	29,717
R-squared	0.783	0.790	0.627	0.784	0.792	0.626

Notes: Dependent variables are logarithmic of total consultations, total registrations, and total cancellation rates respectively in Columns (1) to (3), and logarithmic of offline consultations, offline registrations and offline cancellation rates in Columns (4) to (6). Panel A is for patients aged below 20, Panel B for age between 20 and 40, Panel C for age between 40 and 60; and Panel D for age above 60. All regressions include hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital-month level are in parentheses: $***\ p < 0.01, **\ p < 0.05, *\ p < 0.1$

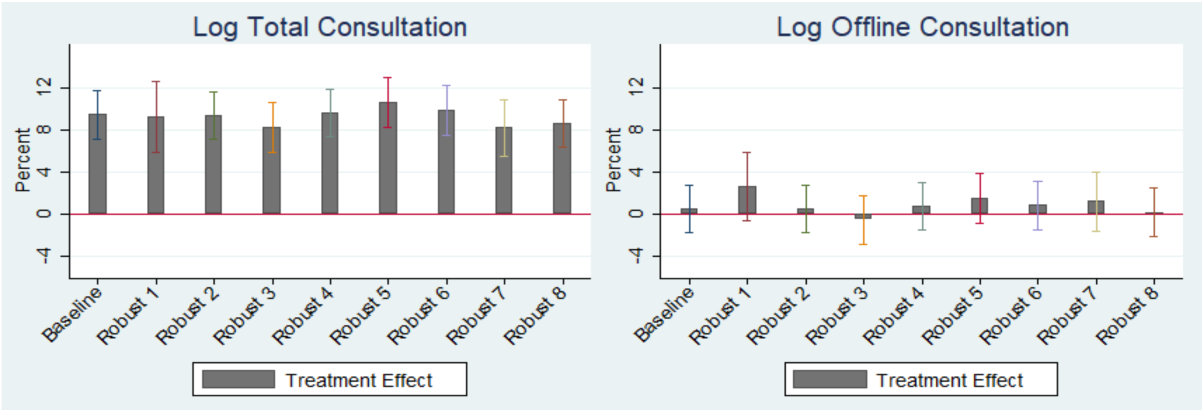
TABLE A16
Effect of App in Non-categorized Departments by Hospital Tier

	(1)	(2)	(3)	(4)	(5)	(6)
		Total			Offline	
Dependent Variable	Consultations	Registrations	Cancellation Rate	Consultations	Registrations	Cancellation Rate
$App \times TierTwo$	0.091	0.080	-0.009	0.020	0.001	-0.017
	(0.059)	(0.060)	(0.015)	(0.065)	(0.073)	(0.018)
$App \times TierThree$	0.071**	-0.046	-0.073**	-0.066*	-0.167***	-0.061**
	(0.033)	(0.033)	(0.029)	(0.038)	(0.026)	(0.029)
Observations	27,355	27,355	27,355	27,355	27,355	27,355
R-squared	0.849	0.855	0.637	0.841	0.847	0.635

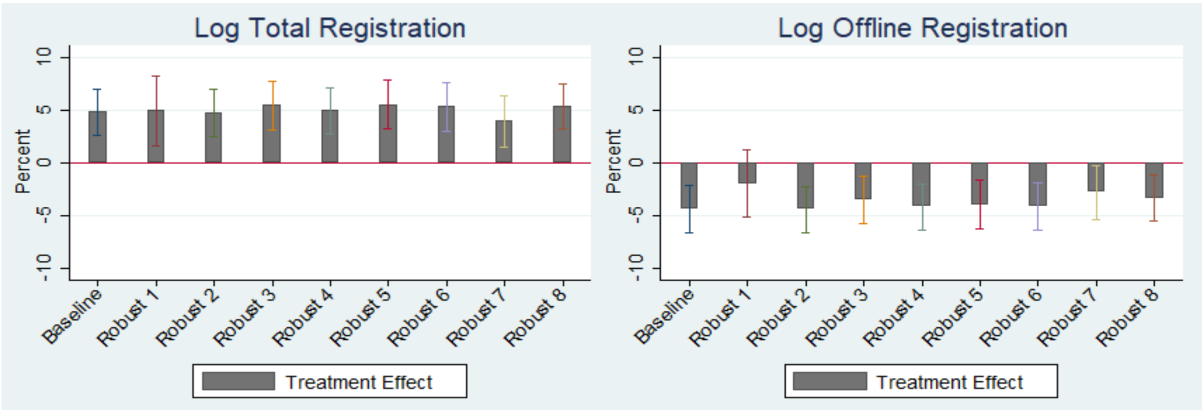
Notes: This table shows the effect of the appointment app on non-categorized departments, that is, all departments excluding the more severe and less severe departments defined in Table 5. *App* is the dummy that switches to one after the app is launched in treatment hospitals. $TierThree$ and $TierTwo$ are dummies for tier-three and tier-two hospitals, respectively. All regressions include hospital fixed effects, month fixed effects, and year fixed effects. Robust standard errors clustered at the hospital level are in parentheses: $*** \, p < 0.01, ** \, p < 0.05, * \, p < 0.1$

Appendix C Figures

(A) Consultation



(B) Registration



(C) Cancellation Rate

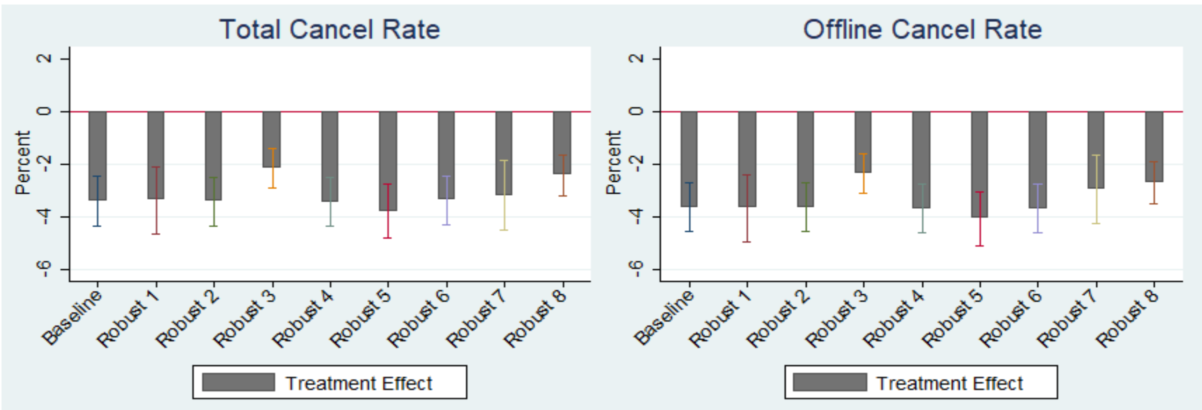


FIGURE A1 Robustness Analysis

Notes: The figure checks the robustness of our baseline results by re-estimating Equation 1 with modified specifications. Each bar represents the estimated β, coupled with the 90% confidence interval. Robust 1 uses subsample with only adopter hospitals; Robust 2 adds city-specific time trend; Robust 3 adds interaction term between linear time trend and pre-app average cancellation rates; Robust 4 adds linear and quadratic time trend; Robust 5 uses subsample that drops three specialty hospitals (dermatology, neurology, and dentistry); Robust 6 uses subsample that drops weekend observations; Robust 7 uses subsample that restricts to hospitals in the provincial capital; Robust 8 uses subsample that drops observations before September 3, 2013 to get a balanced panel. We report the detailed regression results in Appendix Table A4.

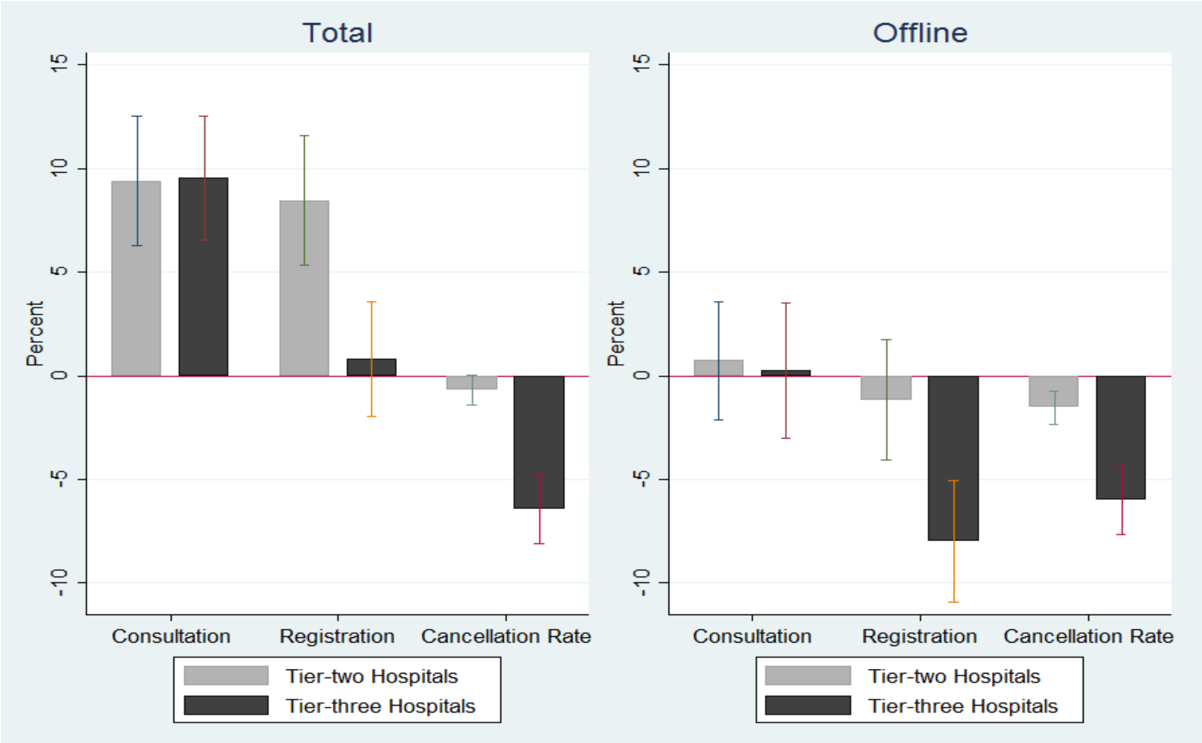


FIGURE A2
Heterogeneous Effect by Hospital Tier

Notes: This figure estimates the heterogeneous effect of the app across hospitals of different size. We run $y_{it} = \alpha + \beta_{TierTow} App_{it} \times TierTwo_i + \beta_{TierThree} App_{it} \times TierThree_i + \lambda_i + \lambda_t + \epsilon_{it}$ on hospital-level data, where $TierThree_i$ and $TierTwo_i$ are dummies for tier-three and tier-two hospitals, respectively. Lighter shaded bars are estimated $\beta_{TierTow}$, while darker shaded bars are estimated $\beta_{TierThree}$, both coupled with 90% confidence intervals. We report the detailed regression results in Table 3.

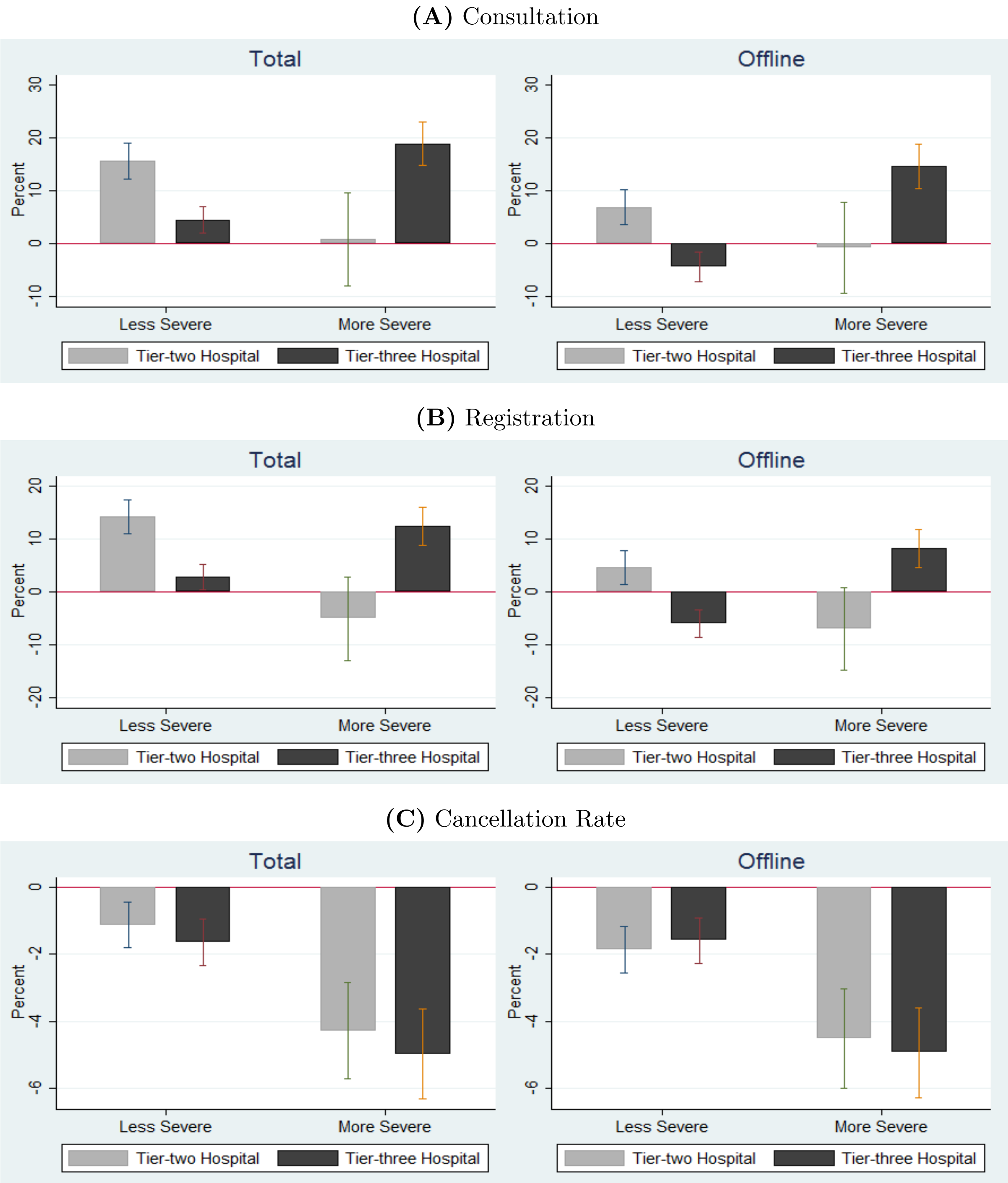


FIGURE A3
Heterogeneous Effect by Type of Department and by Hospital Tier

Notes: This figure estimates the heterogeneous effect of the app across different types of departments and across hospitals of different tiers. We run $y_{it} = \alpha + \beta_{TierTwo} App_{it} \times TierTwo_i + \beta_{TierThree} App_{it} \times TierThree_i + \lambda_i + \lambda_t + \epsilon_{it}$ on hospital-level data from two types of departments: less serious, and more serious. $TierThree_i$ and $TierTwo_i$ are dummies for tier-three and tier-two hospitals, respectively. Light shaded bars are estimated $\beta_{TierTwo}$ and dark shaded bars are estimated $\beta_{TierThree}$, both coupled with 90% confidence intervals. We report the detailed regression results in Table 5. Department categories are defined in Section 5.2.

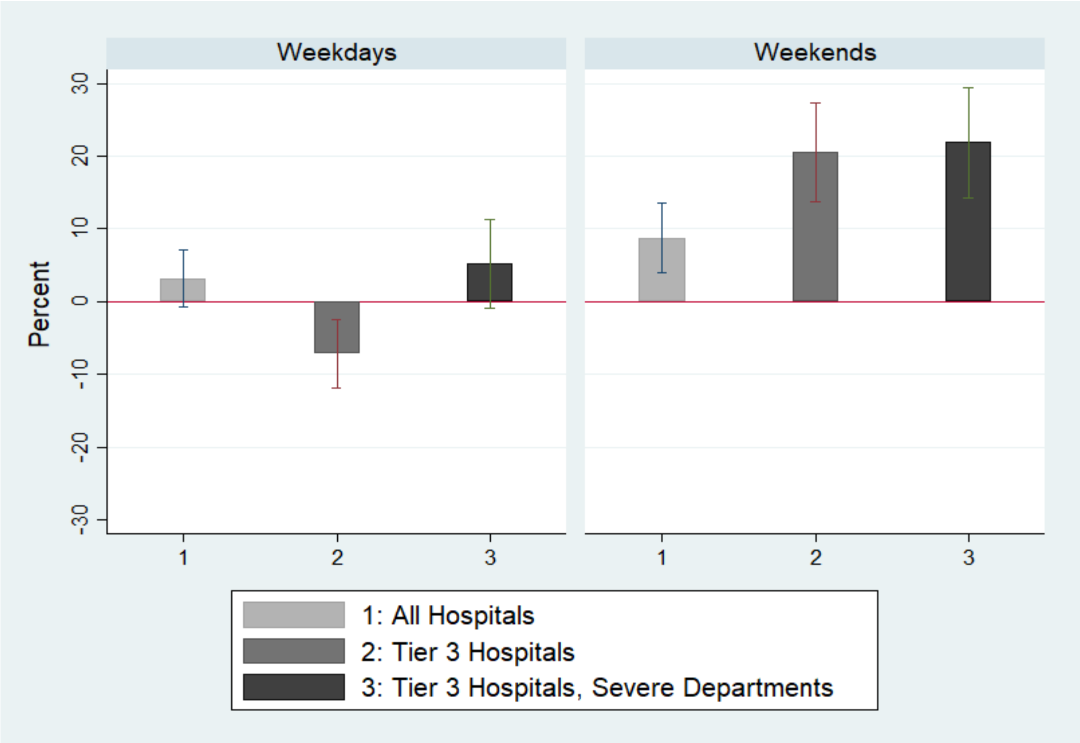


FIGURE A4
Patients with Severe Conditions Sort to Weekend Slots

Notes: This figure presents evidence that the rise of total consultation in tier-three hospitals is largely due to that patients with severe conditions switching from the over-utilized weekday slots to the under-utilized weekend slots.

References

Blumenthal, D and W Hsiao, "Privatization And Its Discontents – The Evolving Chinese Health Care System," *New England Journal of Medicine*, 2005, *353* (11), 1165–1170.

Blumenthal, David and William Hsiao, "Lessons from the East – China's Rapidly Evolving Health Care System," *New England Journal of Medicine*, 2015, *372* (14), 1281–1285.

Chen, Zhu, "Launch of the Health-Care Reform Plan in China," *Lancet*, 2009, *373* (9672), 1322–1324.

Deu, Franck Le, Rajesh Parekh, Fangning Zhang, and Gaobo Zhou, "Healthcare in China: Entering Uncharted Waters," *Shanghai: McKinsey & Company*, 2012.

Eggleston, Karen, Li Ling, Meng Qingyue, Magnus Lindelow, and Adam Wagstaff, "Health Service Delivery in China: A Literature Review," *Health Economics*, 2008, *17* (2), 149–165.

Green, Linda, "Queueing Analysis in Healthcare," in "Patient Flow: Reducing Delay in Healthcare Delivery," Springer, 2006, pp. 281–307.

Ministry of Health, *Public Hospital Classification Standard*, Beijing,China, 1989.

National Bureau of Statistics, *China Health Statistical Yearbook 2015 (Chinese Edition)*, China Statistics Press, 2015.

OECD, *OECD Health Statistics 2016* 2016.

World Bank, "China: The Health Sector.," 1984, *Report No. 4464 – CHA.*

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