Hospital Competition and Quality: Evidence from the Entry of High-Speed Train in South Korea*

Hyesung Yoo† Maria Ana Vitorino‡ Song Yao§

November 10, 2020

Preliminary and Incomplete. Please do not cite without authors' permission

Abstract

This paper leverages the entry of a high-speed train (HST) system in South Korea as a natural experiment to establish the causal effect of competition among hospitals on health care quality and consumer welfare. Using a difference-in-differences estimator, we examine the effects of competition on hospitals depending on their proximity to train stations, specifically how increased competition impacts health outcomes as measured by 30-day mortality rates. Our results suggest that increased competition intensity leads to better quality of clinical care. To evaluate the overall impact of the HST on patients welfare, we estimate a structural model of hospital choice, allowing for a flexible formation of patients' consideration sets. We find that patients living near a HST station experience an improvement in welfare arising from the reduction in travel time as well as improvements in hospital quality. Patients living further away from HST stations also experience an improvement in welfare although they do not gain from the reduced travel time due to the improvement in the quality of treated hospitals. We also find that the HST can have a beneficial impact on patient helath by facilitating patients' sorting to better hospitals, even while holding quality of clinical care constant.

*We are grateful for the helpful comments provided by Mark Bergen, Thomas Holmes, Amil Petrin, Joel Waldfogel, Linli Xu, and participants of the Applied Microeconomic Workshop, Marketing Brownbag Seminar at the University of Minnesota, INFORMS Marketing Science Conference 2018, and seminar participants at Sorbonne University and the University of Chicago. The opinions expressed in this article are the authors own and do not reflect the view of the National Health Insurance Services (NHIS-2018-2-139) and Health Insurance Review Services.

†PhD Candidate, Olin Business School, Washington University in St. Louis; hyesung.yoo@wustl.edu

‡Associate Professor of Marketing, INSEAD; maria-ana.vitorino@insead.edu

§Associate Professor of Marketing, Olin Business School, Washington University in St. Louis; songyao@wustl.edu.

1 Introduction

It is important to understand how competition affects service quality in the health care industry. However, empirical evidence on this topic is mixed. Policies to improve the efficiency and the quality of health care have been introduced in several countries, but their effectiveness remains ambiguous. Difficulty in assessing the impact of competition is partly due to the fact that competition in health care markets is geographically based, as pointed out by Propper et al. (2008) and Gaynor et al. (2013a).

Many existing studies rely on cross-sectional and over time variation in hospital market structure to identify the impact of competition on service quality of hospitals. However, the market structure may be endogenous because the quality of incumbent hospitals and potential entrants may affect their strategic entry and exit decisions, hence the market structure. Other studies exploit changes in health-related policies, which are exogenous shocks that spur competition. Yet the analysis are often complicated by the fact that when policys are *health*-related, they may affect the incentives of the agents involved in ways unanticipated by researchers. If such incentive changes are not accounted for in the analysis, the conclusions may be biased.

In this article we exploit the entry of high-speed train (HST henceforth) system in South Korea to examine the effects of competition on the quality of health care. As of April 2004, Korea Train eXpress (KTX) started operating in South Korea, connecting most major cities by high-speed rail. An important aspect of the South Korean healthcare industry is that patients have the full freedom to go to any hospital of their choice and prices are fixed. The introduction of the HST represents an exogenous shock to the healthcare market in that it greatly reduced patients' travel time, and enabled patients to consider hospitals that were previously unreachable due to long travel distances, thereby increasing substitutability between hospitals. According to news reports, the proportion of rural patients choosing the top four largest hospitals in Seoul increased from 41.2% in 2002 to 48.5% in 2007 as a result of the HST.[1] In addition, when Kim et al. (2008) randomly surveyed HST passengers arriving in Seoul and asked them: "Have you used the HST to seek treatment in hospitals located in Seoul at least once?" 36% (out of 561 passengers) responded "Yes". The news reports and the survey provide some evidence that patients indeed use the HST for medical purposes. Clearly the reduction in travel time facilitates access to better hospitals, implying that hospitals that previously competed locally are now competing with those located further away.

[1]Source: http://news20.busan.com/controller/newsController.jsp?newsId=20110804000124 (in Korean), accessed on July 10, 2018.

We rely on the fact that the HST does not extend to all regions, thereby increasing competition only for hospitals that are located sufficiently close to HST stations. In the current context, there are treated hospitals - hospitals that are located close to a HST station, as well as treated patients - patients that live close to a HST station (more discussion on this subject in the next section). Although our primary interest is to study the impact of competition on hospital quality, we distinguish patients in the treated group from those in the control group so as to provide descriptive evidence on patients' responses to the entry of the HST, as well as to investigate differential changes in patients' welfare and health outcomes based on where they live.

Our analysis relies on the health insurance claims dataset from the National Health Insurance Services (NHIS). We first proceed by providing descriptive evidence to show that patients in the treated group traveled further distances to visit a hospital after the entry of the HST, whereas we do not see such a pattern for patients in the control group, suggesting that patients responded differently to the entry of the HST depending on the proximity from their home to the HST station. Using a difference-in-differences estimator, we then examine the impact of increased hospital competition on the hospital clinical quality, as measured by 30-day risk-adjusted mortality rates following admissions for a surgery. Specifically, we look at all surgeries that were conducted during this period where mortality rate can be used as a measure of quality of clinical care. We find that increased competition improves the clinical quality: hospitals affected by the entry of the HST experience a decrease in adjusted mortality rates.

We then estimate a structural model of hospital choice and use the model estimates to quantify the impact of the entry of the HST on patient welfare. We find that patients living in treated regions experience an improvement in welfare due to both reduction in travel costs as well as enhanced clinical quality. Although patients living in control regions do not benefit from reduced travel costs (becauase there is no HST station near their homes), they also experience an increase in welfare because many of them choose to go to hospitals that are affected by HST. We further use the model estimates to measure the effect of patients' sorting to better hospitals (due to lower travel costs) on their health outcomes (survival from the surgery). This is implemented by comparing the number of death in the post-HST period to a counterfactual scenario when the HST is removed. From this analysis we find that a substantial number of lives can be saved annually with the HST as a result of patients sorting to better hospitals. Our research contributes to the literature on hospital competition and quality in health care. The most influential study of health care markets with fixed prices is Kessler and McClellan (2000), who examine the impact of market concentration on both costs and mortality rates for US Medicare Acute Myocardial Infarction (AMI) patients. They find that in the 1980s competition led to higher costs but lower mortality rates, but find that after 1990, competition resulted in both lower costs and lower mortality rates, and conclude that competition is unambiguously welfare improving post-1990.[2] Exploiting the 2006 English pro-competitive policy shift, Gaynor et al. (2013a) study the impact of the competition on quality (as measured by death rates from heart attack) as well as other measures of quality such as hospital productivity and expenditures (hospital operating expenditures and expernditures per admission) using a difference-in-differences research design. They find that increased competition improves the quality of clinical care without increasing expenditures.[3] Leveraging the same reform, Gaynor et al. (2016) find that patients became more responsive to clinical quality post-reform, and that hospitals responded to changes in demand by improving quality.

Some other papers that study the relationship between competition and healthcare quality, however, find opposite results. Using Medicare data for AMI and pneumonia patients, Gowrisankaran and Town (2003) also estimate the impact of hospital market structure on mortality rates for Medicare patients and find that mortality rate is worse for patients treated in hospitals with more intense competition. This is in contrast to the classical theoretical literature that increased competition under fixed prices results in improved quality. Gowrisankaran and Town (2003) provide a possible explanation: If the profit margin on Medicare patients is sufficiently low, then greater competition for these Medicare patients can cause the hospital to focus on more profitable HMO patients and give up on investing in Medicare patients. In fact, Brekke et al. (2011) show theoretically that under fixed prices, increasing competition through either lower transportation costs (increased substitutability) or a higher number of hospitals may have ambiguous effects on quality if profit margins are low or negative, or if hospitals deviate from profit-maximizing behavior. Several papers find further empirical evidence that support the results of Gowrisankaran and Town (2003) (Propper et al. (2004), Propper et al. (2008), Lewis and Pflum (2017), Colla et al. (2016)). Leveraging the 1991 health reform in the UK National Health Service, Propper et al. (2004) find that the relationship between competition and AMI mortality rates are negative. Propper et al. (2008) investigate the changes further and find that increased competition reduces waiting times, suggesting that hospitals facing more competition reduce services that affect mortality rates (that are unobserved) in order

[2]Other papers such as Shen (2003) finds mixed effects, and Shortell and Hughes (1988) find no effects of competition on quality using Medicare patient data.

[3]For measures of hospital productivity, they use simple measure of labor productivity- the number of admissions per clinical staff.

to increase other activities which are better observed by the health-care buyers. Findings of Lewis and Pflum (2017) also suggest that in response to competition, hospitals divert the resources away from investing in clinical quality, which is imperfectly observed, in order to increase investment in amenities that are better observed by the patient.[4]

Our research advances the existing literature in health economics by studying the effects of competition with fixed prices following an exogenous shock. Because the shock (entry of HST) that increases competition is orthogonal to hospital market structure or any other aspect of healthcare, our setting provides a unique and novel natural-experiment that helps answer our research question. Furthermore, because the HST only reaches certain regions of the country, not only can we do a pre-post analysis, but we are also able to explore the variations in the degree of treatment for more convincing insights. To the best of our knowlege, Gaynor et al. (2013a) and Propper et al. (2008) are the only papers that employ difference-in-differences approach to study this question.

Our research is also the first in which the competition is driven by a shock that reduces travel costs. In their theoretical model, Brekke et al. (2011) measure intensified competition in two ways; more hospitals in the market, and lower transportation costs (increased substitutability between hospitals). The importance of tradeoff between quality and travel timethat patients face is highlighted in Tay (2003). This tradeoff between quality and travel time is what gives market power to hospitals. The entry of the HST reduces the travel time faced by the patients, thereby alleviating this tradeoff. As long as hospital quality remains unchanged, the introduction of the HST therefore should be unambiguously welfare improving. We show that this is indeed the case by decomposing the changes in patient welfare resulting from changes in travel time and changes in mortality rates.

Our research is also closely related to the literature on constrained choice sets. Ho (2006), Dafny et al. (2013) and Gaynor et al. (2016) also analyze the effects of removing choice constraints within the health care context. Our setting is more similar to that of Gaynor et al. (2016) in which the patients' choice sets are unobserved. To exploit the unique feature of our setting in which the entry of the HST increased the number of hospitals in a patient's consideration set, we adopt a modeling approach used in the geography/transportation literature. Specifically, we explicitly model the formation of consideration sets for the post-HST period when individuals have limited time resource, and evaluate the welfare effects of the removal of the HST in a counterfactual scenario.

[4]In contrast to Medicare patients, however, both Gowrisankaran and Town (2003) and Lewis and Pflum (2017) find that competition improves clinical quality for HMO patients. Lewis and Pflum (2017) explain that because HMOs can better evaluate the clinical quality of hospitals than individual patients, hospitals have higher incentives to improve clinical quality levels when competing for inclusion in HMO provider networks.

Finally, our research adds to the fast growing literature on the economic impacts of transportation infrastructure (Banerjee et al. (2012); Qin (2016); Donaldson (2018); Heuermann and Schmieder (2018); Qin et al. (2018)). While these papers mainly study the impact on economic activities that are directly affected by the new transportation system, such as inter-regional trade, per capita GDP, co-opetition between transportation modes, and housing/commute decisions, our paper looks at the unexpected externality caused by the HST.

The rest of this paper is structured as follows. In the next section we describe the relevant aspects of the industry; Section 3 describes the data; In Section 4 we describe our estimation strategy and present the results. Section 5 outlines the structural demand model, and Section 6 presents the estimation results. Section 7 analyzes the welfare effects of the entry of the HST and Section 8 concludes.

2 Industry Details

2.1 Health Care Industry

National Health Insurance (NHI) program in South Korea is a compulsory single-payer public insurance system which covers the entire resident population. The social insurance system of South Korea was established in 1977, and initially covered only 8.79% of the population, but expanded to approximately 97% of the population by 1989. It operated as a multi-insurance fund system with more than 370 insurers until July 2000, when the funds were integrated to form a single-payer system. It is managed by a single insurer, the National Health Insurance Corporation (NHIC), and is supervised by the Ministry of Health, Welfare and Family Affairs (MIHWFA). The Health Insurance Review and Assessment Service (HIRA), also supervised by MIHWFA, reviews the cost and healthcare benefits and evaluates the appropriateness of health care services provided by hospitals. The system is funded by compulsory contributions from the entire resident population and government subsidies. The amount paid as NHIC contributions by an individual depends on his income and wealth; the elderly and disabled pay less.

As opposed to public-sector dominant healthcare financing, healthcare delivery in South Korea is predominantly provided by the private sector: approximately 90% of hospitals are private institutions. Since the launch of the NHI program, private providers are not allowed to opt out from the program. This is to ensure that private health-care providers respond to changes in demand which the public health insurance has brought about.

The NHIC negotiates the level of medical service fees annually with provider associations. The fee schedule includes fees for all medical services and materials including drugs, as well as remuneration of providers for the services they provide. Patients are responsible for any co-payments applicable to the medical services they received, and the NHIC reimburses healthcare providers the share of medical costs not borne directly by the patient on the basis of the fee schedule. Fee regulation has been the subject of recurrent complaints by providers in South Korea, who claim that they are not adequately compensated for their services as a result of historically low levels of NHI fees, which did not keep pace with inflation until the mid-1990s.

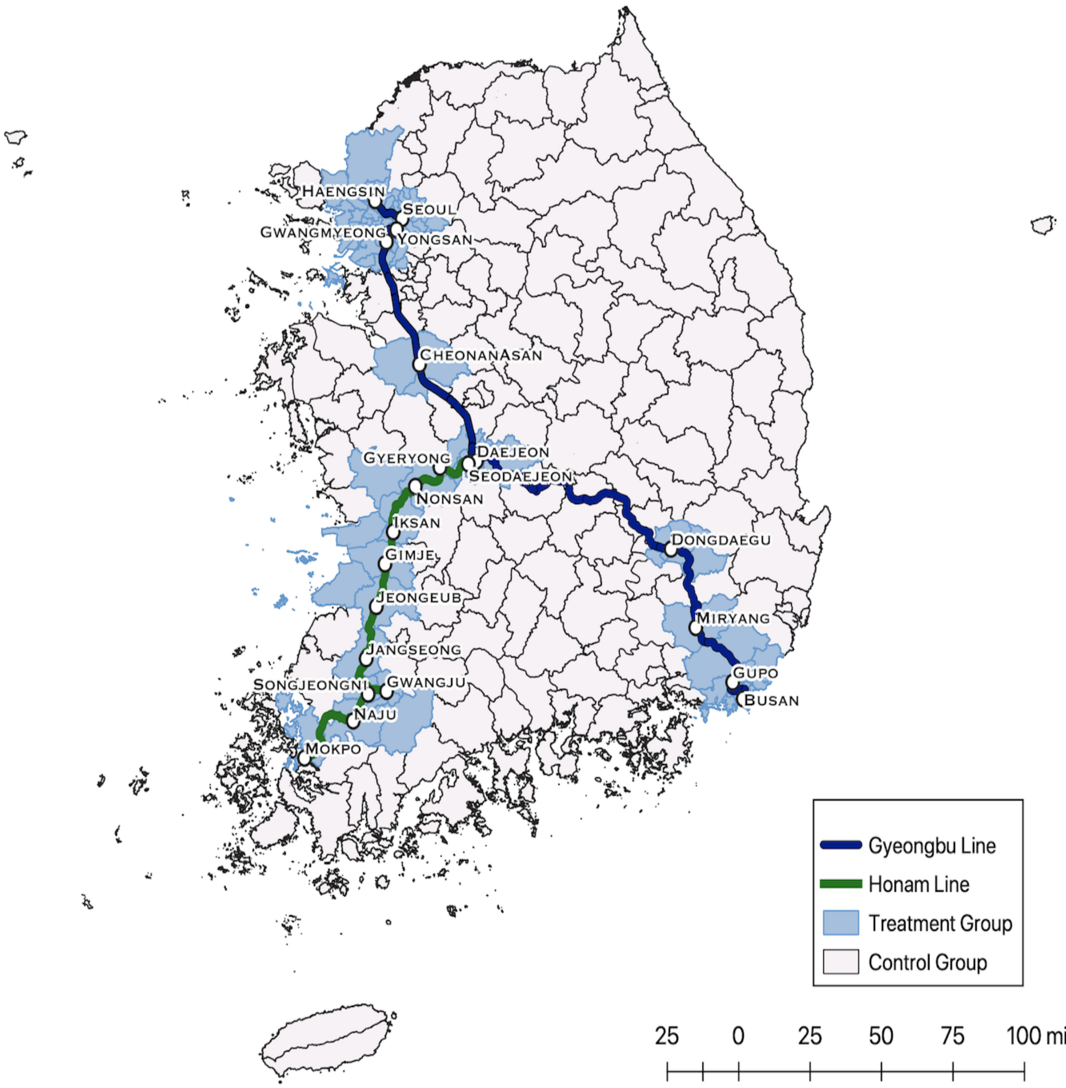
In fact, according to a report published by Health Insurance Review Assessment (HIRA) in 2006, the fixed fee schedule covers on average only 73.9% of the costs incurred by providers.[5] Differential margins for different medical services lead physicians to provide more of those services with higher margins. Specialties of which services are paid relatively generously attract a greater number of applicants for the residency training. Popular specialties include psychiatry, ophthalmology and dermatology, whereas radiology, thoracic surgery and anesthesiology are the least popular. Moreover, to mitigate the effects of negative margins, physicians encourage patients to receive uninsured medical services, for which hospitals have the full freedom to set their own price.

Healthcare delivery system in South Korea is classified into three tiers: primary, secondary, and tertiary care. Although NHI service flow is designed to progress from primary to secondary to tertiary care, patients have the complete freedom to choose a healthcare provider at any level within this system with some financial incentives. To achieve an efficient distribution of limited healthcare resources, outpatient insurance coverage largely depends on the tier of the hospital. Patients must be referred by primary or secondary care hospitals to receive outpatient treatment in tertiary hospitals, in which case 40% of their bills are covered by insurance (Otherwise, they can expect to pay 100% of the bill). The insurance coverage is identical at all levels of hospitals for inpatient care, where patients pay 20% of medical expenses. In our analysis we only focus on inpatient surgical treatments.

2.2 Entry of High-Speed Train

South Korea's HST system, Korea Train eXpress (KTX), commenced commercial operations on April 1st 2004, substantially altering patterns of long-distance travel. Construction of the HST

[5]Source: http://www.medicaltimes.com/News/39629, accessed on November 30, 2018



Notes: This map displays first-stage HST lines. Shaded areas represent regions that are treated - districts whose centroids are located within 15 miles of the HST station.

Figure 1: Treated Regions

system occurred in two stages.[6] The first-stage construction involved building Gyeongbu HST Line connecting Seoul to Daegu and electrifying the existing Gyeongbu Line connecting Daegu-Busan, as well as electrifying the existing Honam Line connecting Daejeon-Mokpo.[7] The second-stage HST system, which involved the construction of the new Gyeongbu HST line connecting Daegu to Busan replacing the existing electrified tracks, went into service in November of 2010. In this paper we only focus on the first-stage HST system. Although the launch of the second-stage HST system enabled the HST to reach full speed through Daegu-Busan corridor, this shock was much smaller in magnitude compared to the shock generated by the first-stage HST system. Figure 1 displays two HST lines of the first-stage HST system, Gyeongbu Line (blue) connecting Seoul-Busan and Honam line (green) connecting Seoul-Mokpo. We define a "treatmented area" as an area located within 10 miles of the HST station. Shaded areas in Figure 1 represent treated areas whose centroids are within 10 miles of a HST station.[8]

At the time of the lauch in 2004, the HST operated 128 times per day (94 times on Gyeongbu Line, and 34 times on the Honam Line), and the daily frequency increased to 163 in the following years. HST fares were fixed and kept low, at approximately 55% of the corresponding air fares for the same routes, to encourage the use of the HST.[9] The HST system has reduced the travel time from Seoul to Busan from more than 5 hours by car to 2 hours 40 minutes by train.

3 Data

We rely on a number of data sources at the patient, hospital and city-county-district level.[10]

Our patient data comes from the National Health Insurance Services (NHIS) which is a health

[6]Note that here we are referring to the construction of Gyeongbu HST system. The construction of additional HST systems were completed only after 2015. Additional electrified (existing) lines were added by the end of 2010.

[7]Newly constructed links included 51.6 mi of viaducts and 47.0 mi of tunnels. Electrification of the existing rail comprised of 82.5 mi across Daegu to Busan, 12.9 mi across Daejeon, and 164.3 mi from Daejeon to Mokpo and Gwangju. First stage Gyeongbu HST stations include Seoul Station, Gwangmyeong, Cheonan-Asan, Daejon, Dongdaegu stations, and the electrified Gyeongbu line connecting Dongdaegu and Busan includes Miryang, Gupo and Busan stations. Honam line includes Yongsan station, Seodaejeon, Dungyae, Nonsan, Iksan, Gimje, Jeongeub, Jangseong, Songjeongni, Gwangju, Naju, and Mokpo stations. There exists a depot for HST along the Gyeongui Line at Haengsin station. Thus some HST services continue beyond Seoul and Yongsan station and terminate at Haengsin station. For detailed information on HST services see Cho and Chung (2008).

[8]The robustness of our 10-mile definition of treatment is discussed in Appendix **??**

[9]In addition to low regular prices, various discounts (60% off the regular passes and 20% off the reserved tickets) were available to attract as many passengers as possible.

[10]South Korea is made up of 17 first-tier administrative divisions (province level). These are further subdivided into cities (si), counties (gun), districts (gu), towns (eup), townships (myeon), neighborshoods (dong) and villages (ri). Once a country attains a population of at least 150,000, it becomes a city. Cities with a population of over 500,000 are subdivided into districts. Districts are then further divided into neighborhoods (dong). Cities with a population of less than 500,000 are directly diveded into neighborhoods (dong).

	Pre-HST				Post-HST			
	Mean	SD	Min	Max	Mean	SD	Min	Max
				Control Hospitals (N = 55)				
Total admissions	156	138.1	13	598	193.1	188.0	13	789
Hospital beds	567.3	296.3	99	1256	567.3	296.3	99	1256
Mortality rates	0.050	0.041	0	0.280	0.052	0.050	0	0.308
				Treated Hospitals (N = 112)				
Total admissions	243.9	256.4	17	1501	290	313	13	1943
Hospital beds	703.2	453.9	121	2993	703.2	453.9	121	2993
Mortality rates	0.043	0	0.176	0.040	0.040	0.029	0	0.176

Notes: hospital-treatment in this table is defined as being located within 15-miles of train stations

Table 1: Summary Statistics: Hospital Characteristics

insurance claims dataset collected by the single insurer system NHI (NHIS-2018-2-139). Our data are of a nationally representative random sample, which accounts for approximately 2% of the entire South Korean population for years 2003 to 2007. The data contain patient-level information on medical procedures received at the hospitals. Detailed information on patient demographics, diagnosis, patients' location at the district level[11] and hospital choice are observed, as well as the date of admission, number of inpatient treatment days, and the month/year of the patient's death.

The identity of the hospitals in the NHIS dataset are anonymized and hospital location is observable only at the provincial level. To get a more precise location of the hospitals, which is essential for our analysis, we combine the NHIS dataset with that obtained from HIRA (Health Insurance Review Assessment) which, in addition to the hospital characteristics in the NHIS dataset, also provides hospital location at the district level[12]. The identity of the hospitals in the HIRA dataset is also anonymized, but we are able to match this dataset to NHIS dataset using hospitals characteristics.

Our sample selection process is as follows: To study the causal impact of increased competition on the quality of clinical care, we define January 2003 to March 2004 as the pre-HST time period and January 2006 to March 2007 as the post-HST time period. The data are collapsed into pre- and post-HST period.

We focus on patients who underwent any kind of surgery. Specifically, we consider all surgeries

[11]More precisely, patients' locations are at the city-county-district level because some counties are not populated enough to qualify for a city and hence are not sub-divided into districts.

[12]For the same reasons as discussed in footnote 11, hospitals' locations are at the city-county-district level

	Control Patients				Treated Patients			
	Pre-HST Period		Post-HST Period		Pre-HST Period		Post-HST Period	
(Fractions of)	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	0.490	0.450	0.486	0.500	0.497	0.500	0.483	0.500
Ages 0-19 Years	0.150	0.357	0.147	0.354	0.149	0.356	0.155	0.362
Ages 20-39 Years	0.188	0.391	0.167	0.373	0.213	0.410	0.177	0.382
Ages 40-59 Years	0.290	0.454	0.290	0.454	0.300	0.458	0.303	0.460
Ages 60-79 Years	0.328	0.470	0.344	0.475	0.299	0.458	0.320	0.466
Ages 80 Years +	0.044	0.204	0.052	0.221	0.039	0.194	0.044	0.206
Income Group 0-1	0.084	0.277	0.077	0.267	0.068	0.251	0.067	0.250
Income Group 2-4	0.198	0.399	0.188	0.391	0.186	0.389	0.179	0.383
Income Group 5-7	0.296	0.456	0.286	0.452	0.295	0.456	0.289	0.453
Income Group 8-10	0.422	0.494	0.449	0.497	0.451	0.498	0.465	0.499
Comorbidity	0.873	0.333	0.847	0.360	0.854	0.353	0.844	0.363
Nobs	18,639		22,431		17,252		20,664	

Notes: patient-treatment in this table is defined as living within 10-miles of train stations

Table 2: Summary Statistics: Patient Characteristics

that were conducted during this period that resulted in at least one death. Since our data is a 2% sample of the entire population, 1 death in the data can be inferred as 50 deaths in the entire population. Ideally we want to look at patients suffering from one specific illness, or who underwent one specific type of surgery in order to minimize the contamination of hospital quality (mortality rates) with patient selection.[13] Constraining our analysis to a single type of surgery, however, leaves us with too few observations (too few patients as well as too few hospitals). Limiting our attention to only one "category" of surgery (e.g. cardiovascular surgery) also leaves us with too few observations. To attenuate the contamination of hospital quality from pooling patients across multiple types of surgeries, we control for the riskiness of each type of surgery as well as the patients' diagnosed disease when obtaining the adjusted mortality rates. Details of this procedure are provided in the Appendix 8.

The key feature of our setting is that the entry of HST enables patients to exercise choice among alternatives with different travel distances. To take advantage of this feature, we drop the following patients who were less likely to exercise choice based on hospital location: First, patients who arrived at the hospital via ambulance because the emergency ambulance usually takes patients to a nearby hospital. Second, patients who arrived at the hospital via intrahopital transfer as it is the physician who makes the choice of the hospital in this case. Next, we drop patients living on islands (Jeju

[13] Gowrisankaran and Town (2003)look at pneumonia patients, Kessler and McClellan (2000), Propper et al. (2004) look at acute myocardial infarction (AMI) patients, and Gaynor et al. (2016) look at patients receiving coronary artery bypass grafting (CABG) surgery.

Table 3: Descriptive Evidence of Changes in Travel Distance

	Control Patients			Treated Patients		
	Pre-HST	Post-HST		Pre-HST	Post-HST	
Distance Traveled	Mean	Mean	%Δ	Mean	Mean	Δ
	(st.dev)	(st.dev)	(t-stat)	(st.dev)	(st.dev)	Δ(t-stat)
Panel A: Patient Treatment: within 10 miles of train station						
			A1. Excluding ambulance and transfer patients			
	27.338	27.825	1.78%	13.713	14.781	7.79%
	(43.183)	(43.295)	1.135	(36.365)	(37.602)	2.797
Nobs	18,639	22,431	0.2562	17,252	20,664	0.0052
			A2. Excluding patients living in Seoul			
	29.601	29.704	0.35%	21.957	24.458	11.4%
	44.344	44.313	0.224	48.372	50.821	3.192
Nobs	16,867	20,520	0.8225	7,587	8,557	0.0014
			A3. Ambulance and transfer patients			
	22.112	22.363	1.14%	11.096	9.3890	-15.4%
	(36.525)	(35.887)	0.114	(34.202)	(25.805)	1.041
Nobs	387	853	0.9095	478	901	0.2981
Panel B: Patient Treatment : within 15 miles of train station						
			B1. Excluding ambulance and transfer patients			
	33.071	33.472	1.21%	13.334	14.128	5.95%
	(46.240)	(46.425)	0.747	(34.746)	(35.567)	2.485
Nobs	13,556	16,580	0.4550	22,335	26,515	0.0130
			B2. Excluding patients living in Seoul			
	33.071	33.472	1.21%	19.963	21.113	5.76%
	(46.240)	(46.425)	0.7471	(44.109)	(45.381)	1.9593
Nobs	13,556	16,580	0.4550	10,898	12,497	0.0501
			B3. Ambulance and transfer patients			
	28.190	25.529	-9.44%	10.652	10.079	-5.38%
	(39.746)	(37.536)	0.953	(32.310)	(26.379)	0.396
Nobs	265	638	0.3408	600	1,116	0.6921

and Ulleng Islands, as well as Shin-ahn and Ong-jin Gun) because we are unable to calculate the travel time to hospitals by car and ferry for these patients, a necessary component for estimating our demand model and performing counterfactuals.

We drop outpatient admissions, where the patient stayed at the hospital less than 24 hours to ensure that the patients in our sample are sick enough. Following Tay (2003) and Ho (2006), we exclude hospitals with fewer than 10 admissions per period, and we only keep hospitals that appear in both pre- and post-HST periods to facilitate the comparison of hospital quality. Our final sample consists of 167 hospitals and 78,986 patients.

In our setting, there are "Treated Hospitals" and "Treated Patients". "Treated Hospitals" are hospitals that are located within 15 miles of the HST staion, and "Treated Patients" are patients who live within 15 miles of the HST station. In our main analysis, we define Treated Hospitals as hospitals that are located within 15 miles of the HST station unless otherwise specified. For Treated Patients, we show descriptive statistics using both 10 mile and 15 mile definitions of treatment. In the Appendix, we show that our results hold consistently even if we change the definition of treament as being located within 5 miles, 10 miles, 20 miles. Table 1 and Table 2 provides summary statistics of hospital characteristics and patient characteristics, respectively.[14]

In Table 3 we present descriptive evidence on changes in patients' travel patterns following the entry of HST. Panel A.1 reports the average travel distances (in miles) before-and after the introduction of HST, defining patients living within 10 miles of the HST station as treated. While there are no changes in travel distances for patient living in control regions, patients living in treated regions traveled significantly further distances after the entry of the HST (approximately 8 percent increase post HST). These differences become more salient when we only focus on patients living in non-Seoul areas (Panel A.2): while there is no difference in travel distance for patients living in control regions, the average travel distance increased by 11 percent for patients living in treated regions.[15]

As mentioned earlier, our final sample excludes patients who transferred from other hospital and who arrived at a hospital via ambulance because these patients are less likely (if any) to exercise choice. If the increase in travel distance for patients living in treated regions is a consequence of the entry of the HST, we should not see changes in travel distance for patients who arrived at hospitals via transfer or ambulance because these patients did not take the HST. Table 3 Panel A.3 reports the mean travel distances for patients who arrived at hospitals via transfer or ambulance. As expected, we do not see significant changes in travel distance for patients living in treated regions.

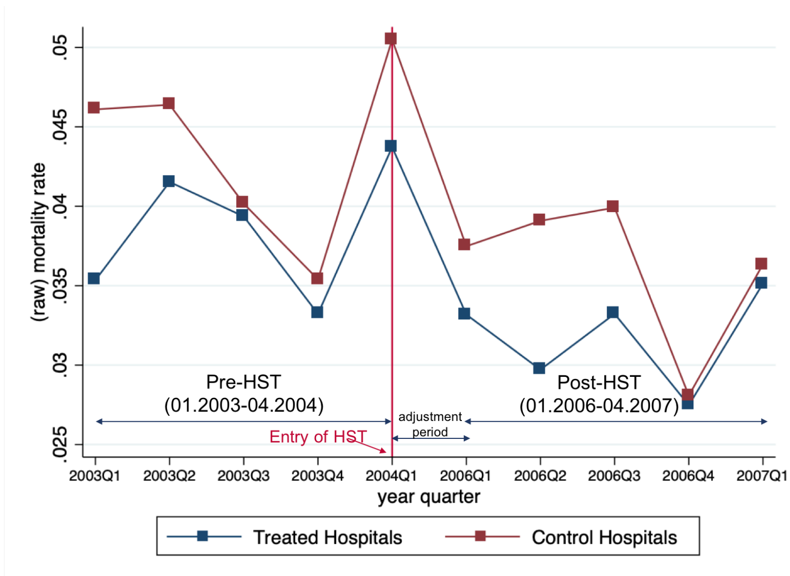
Panel B reports the average travel distances by period and region, defining patients living within 15 miles of the HST station as treated. The patterns reported in this table are consistent with those in Panel A.

4 Difference-in-Differences Estimation and Results

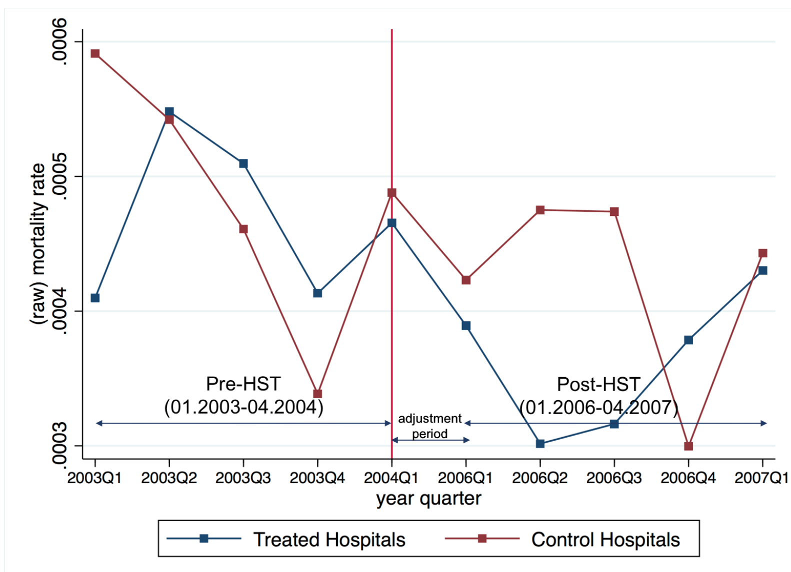
In this section we study the impact of hospital competition on the quality of clinical care using a difference-in-differences approach. Specifically, we compare pre- and post-HST mortality rates of

[14] In our data we observe up to two diagnosis per patient, main diagnosis and sub-diagnosis.

[15] More precisely, we exclude Seoul and the surrounding metro area.



(a) Hospital-level mortality rates



(b) Patient-level mortality rates

Figure 2: Trend of hospital-level mortality rates

hospitals that are located near the HST stations, using hospitals located further away from the HST stations as the control group. Figure 2 plots the hospital-level raw mortality rates by quarter for pre- and post-HST periods.[16] From this figure, we can see that mortality rates at the treated and control hospitalsfollow a similar trend.

We first briefly describe our estimation strategy, and then proceed to describe the issue concerning the use of raw mortality rates as a measure of hospital clinical quality, followed by a description of how to resolve this problem. We then report our estimation results.

4.1 Difference-in-Differences

We analyze the impact of hospital competition on the quality of clinical care using difference-in-differences estimator. Specifically, we estimate the equation as below:

$$Y_{jt} = \beta_0 + \beta_1 Post_t + \beta_2 Treated_j \cdot Post_t + \mu_j + \varepsilon_{jt} \tag{1}$$

where Y_{jt} denotes the quality of clinical care at hospital j in period t, $Post_t$ is a dummy variable which equals 1 if post-HST period, $Treated_j$ is a dummy variable which takes value 1 if hospital j is located in a treated region, $Treated_j \cdot Post_t$ is an interaction term of $Treated_j$ and $Post_t$. We control for hospital-specific characteristics with hospital fixed effects, μ_j. Coefficient β_2 captures the impact of increased competition on Y_{jt} and is of primary interest.

4.2 Adjusted Mortality Rate

Using raw mortality rates as a measure of quality is problematic due to patient selection bias: severely ill patients may choose high quality hospitals. The existing literature address this selection bias by obtaining adjusted mortality rates (Gowrisankaran and Town (1999), Gowrisankaran and Town (2003), Kessler and McClellan (2000), Geweke et al. (2004), Tay (2003)). Specifically, Gowrisankaran and Town (1999) propose controlling for patients' severity of illness with an instrumental variables (IV) framework using geographic location data, i.e. distance from each patient to *all* hospitals. Although the distance to the *chosen* hospital will be correlated with the patient's severity of illness, and hence cannot be a valid instrument, where a patient chooses to live relative to *all* hospitals is uncorrelated to patient's severity of illness. This assumption is commonly used in

[16] In the following analysis, we collapse all the pre-HST and post-HST quarters into a single pre-HST and post-HST period, respectively. This is becuase if we calculate hospital-level mortality rates at the quarter level, some hospitals are left with too few admissions per period.

empirical models of hospital choice, e.g. Kessler and McClellan (2000), Gowrisankaran and Town (1999), Capps et al. (2003), Gaynor and Vogt (2003), Ho (2009), Beckert et al. (2012)

In our setting, the HST facilitates long-distance travel for severely ill patients, and hence the degree of patient selection may be aggravated as a result of the entry of the HST. To allow for this change in the degree of patient selection resulting from the reduction in travel time, we use different sets of instruments in pre- and post-HST periods. We follow Gowrisankaran and Town (1999) but use travel *time* rather than travel distance from each patient to all hospitals as instruments for hospital choice. This is to account for the changes in travel time for patients living sufficiently close to the HST station in post-HST era (because even with HST, the actual distance to the hospitals does not change - what changes in the post-HST period is the travel time).

Specifically, we obtain an adjusted mortality rate by estimating a linear probability model where we regress an indicator for whether a patient dies approximately 30 days following the admission (conditional on choosing hospital j) on a set of hospital/time period dummies and patient's observed characteristics.[17] The mortality of patient i in period t is given as

$$\mu_{it} = \boldsymbol{\psi}' \boldsymbol{c}_i + \boldsymbol{\gamma}' \boldsymbol{h}_i + s_{it} + \eta_{it} \tag{2}$$

where μ_{it} is a dummy variable that denotes the death of patient i within 30 days of the admission, $\boldsymbol{c}_i$ is a vector of dummy variables $(c_{i1pre}, ..., c_{iJpre}, c_{i1post}, ...c_{iJpost})$ where c_{ijt} equals 1 if patient i chooses hospital j in period t, $\boldsymbol{h}_i$ is a vector of patient characteristics that can affect mortality, s_{it} is unobserved (by the researcher) severity of illness, and η_{it} is an i.i.d. normal error term. The parameter vectors to estimate are $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$. With the linear probability model, the elements of estimated fixed effects $\hat{\boldsymbol{\psi}}$ are interpreted as the incremental probability of death from choosing a particular hospital conditional on observed health status, and is used as our measure of quality of care. The coefficient vector $\boldsymbol{\gamma}$ captures the impact of patients' observed health status on the probability of death. We will refer to the estimated measure of quality of care, $\hat{\boldsymbol{\psi}}$ as the adjusted mortality rate. Note that we are slighltly abusing the terminology as $\hat{\boldsymbol{\psi}}$ is not adjusted mortality probabilities per se, but is the hospital's impact on patients' mortality conditional on observed characteristics. Nevertheless we use this terminology for the simplicity. Because hospital choice is likely to be correlated with patients' unobserved severity of illness, estimating equation (2) using OLS will lead to biased estimates. For instance, if sicker patients are more likely to choose a certain

[17] The reason for why we use a linear probability model is because it is difficult to use non-linear models in the presence of endogenous variables. Detailed explanation on this is discussed in Gowrisankaran and Town (1999).

hospital j, s_{it} and c_{ijt} will be positively correlated, and hence $\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}$ will be overestimated.

To address the endogeneity of hospital choice, we use two sets of instrumental variables for hospital choice dummy variables ($\boldsymbol{c}_i$) : (i) the travel time to each hospital, and (ii) a set of dummy variables indicating whether a hospital is located withim 10 miles of the patient's home. As mentioned before, this is to account for the changes in travel time for patients living sufficiently close to the HST station in the post-HST era, and is based on the assumption that where a patient chooses to live relative to *all* hospitals is uncorrelated to her severity of illness. We define travel time for patient i to hospital j in period t as

$$\text{traveltime}_{ijt} = \begin{cases} \min(\text{cartime}_{ij}, \text{traintime}_{ij}) & \text{if } t = \text{post-HST} \\ \text{cartime}_{ij} & \text{if } t = \text{pre-HST} \end{cases} \tag{3}$$

where cartime_{ij} denotes the drive time from patient i's location to hospital j by car, and traintime_{ij} is the travel time from patient i's location to hospital j by HST.[18] Tests of validity of our IV strategy and further details on estimating adjusted mortality rates are provided in Appendix 8.

Having obtained the adjusted mortality rates using the instrumental variable approach outlined above, we use this measure of clinical quality to look the impact of hospital competition on the quality of clinical care using a difference-in-differences estimator.

4.3 Estimation Results

As a starting point to this analysis, we first estimate equation (1) using hospital-level raw mortality rates as a dependent variable. Hospital-level raw mortality rates, however, do not correctly reflect the true quality of clinical care due to differences in patients' health status across hospitals (referred to as hospital's "case-mix") i.e., hospitals with a larger number of sicker patients are more likely to have higher mortality rates. It is therefore essential to take into account differences in patient case-mix across hospitals, especially since we are using patients undergoing various types of different surgeries. Specifically, we include hospital-level case-mix as control variables (details on hospital-level case mix and estimation results provided in Appendix). Table 4 Panel A reports the results, defining hospitals located within 15 miles of HST as treated. We implement a simple difference regression (pre vs post) in Column 1 to analyze the changes in hospital quality after the entry of

[18] Note that traintime_{ij} is obtained by summing the following three components: (i) drive time from i's location to i's nearest HST station h, (ii) travel time from station h to station k, which is the closest HST station to hospital j and (iii) drive time from station k to hospital j. We obtain driving time by car by using *georoute* routine developed by Weber and Péclat (2017) which calculates the driving time between two points under normal traffic conditions.

the HST. The coefficient on Post dummy variable is negative ($\beta_1 = -0.003$, interpreted as decrease in mortality rates by 0.3 percentage points) but not significant. Column 2 reports the difference-in-differences estimates. Since hospitals located near the HST station are the ones that are most affected by the entry of the HST and hence are exposed to increased competition, the estimated diff-in-diff coefficient captures the impact of increased hospital competition. We can see that the diff-in-diff coefficient is negative and significant. Increased competition due to the entry of HST decreased mortality rates by 1.4 percentage points.

In order to control for the case-mix at the patient-level, we estimate equation (2) using OLS, and use estimated $\hat{\psi}$ as a measure of clinical quality to estimate equation (1). Note that although this measure of quality controls for observed health status at the individual patient-level, it does not control for unobserved (to the researcher) severity of illness which may be correlated with patients' hospital choice, and hence may be biased. The results are reported Table 4, Panel A Column 3. The diff-in-diff coefficient is negative and significant, and has a similar magnitude as the one obtained using raw-mortality rates, i.e. increased competition due to the entry of HST decreased adjusted mortality rates by 1.4 percentage points. As already mentioned, however, simply controlling for observed patient case-mix is not sufficient to correctly measure the quality of clinical care. Patients' unobserved (to the researcher) severity of illness, which may be correlated with hospital choice, may contaminate the quality of clinical care. We further control for patients' unobserved severity of illness by instrumenting hospital choice dummy variables for each period with travel time to each hospital, and use thus (using IV) obtained adjusted mortality rates as the dependent variable to estimate equation (1). The results are reported in Table 4, Panel A Column 4. After controlling for unobserved severity of illness, we see that the (absolute) magnitude of diff-in-diff coefficient has become larger. The diff-in-diff coefficient is -0.082 and significant, suggesting that increased hospital competition leads to an improvement in the clinical quality by approximately 8 percentage points. Since we obtained the quality of clinical care using a linear probability model (by estimating estimated equation 2 using OLS), estimated $\hat{\psi}$ does not necessarily lie within [0,1] interval. To facilitate interpretation, we rescale the IV-estimated quality of clinical care so that estimates lie within the same bounds as our raw hospital-level mortality rates. Specifically, let $[raw_{min}, raw_{max}]$ be the range of the raw hospital-level mortality rates, and let $[IV_{min}, IV_{max}]$ be the ragen of IV-estimated mortality rates. For each $\hat{\psi}_j$ obtained via IV estimation, we calculate a rescaled IV-estimated mortality rate $\hat{\phi}_j$ as

Table 4: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

	raw mortality	raw mortality	adjusted mortality OLS	adjusted mortality IV	adjusted mortality IV (rescaled)
	(1)	(2)	(3)	(4)	(5)
		Panel A: Excluding ambulance and transfer patients			
Post	-0.003	0.007	0.004	0.025	0.004
	(0.002)	(0.005)	(0.007)	(0.030)	(0.004)
Treated×Post		**-0.014****	**-0.014***	**-0.082****	**-0.012****
		(0.006)	(0.007)	(0.038)	(0.006)
R-squared	0.832	0.840	0.601	0.618	0.618
		Panel B: Including ambulance and transfer patients			
Post	-0.0043	-0.0014	0.003	0.029	
	(0.0027)	(0.0034)	0.007	0.033	
Treated×Post		**-0.0082***	**-0.014****	**-0.084****	
		(0.0046)	0.007	0.040	
R-squared	0.813	0.816	0.5967	0.6152	
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		55	55	55	55
Treated Hospitals		112	112	112	112
Observations	334	334	334	334	334

Notes: Models are estimated using OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. Treatment is defined as being located within 15-miles of the HST station. When estimating the model with raw-mortality rates, we controlled for hospital level case-mix. Full set of regression results of hospital-level case-mix for raw-mortality rates are presented in Appendix.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

$$\hat{\phi}_j = (\hat{\psi}_j - IV_{min}) \times \left(raw_{max} / \max_j(\hat{\phi}_j) \right)$$

This transformation ensures that $\hat{\psi}_j$ lies within the range $[raw_{min}, raw_{max}]$. Table 4, Panel A Column 5 reports the diff-in-diff estimation results using $\hat{\phi}_j$ as a measure of hospital quality. After rescaling, the magnitude of the diff-in-diff coefficient is similar to those in Columns 2 and 3.

As aforementioned, our final sample excludes patients who transferred from other hospital and who arrived at a hospital via ambulance. Including these patients in our sample should not change our results because the quality of clinical care should be independent from how patiens arrived at a hospital. We include these patients in our sample and estimate equation (1) and report the results in Table 4, Panel B. Our results hold consistently, and the DID estimates are similar to those in Panel A.

4.4 Discussion of the Results

The results in the previous subsection suggest that increased competition leads to an improvement in hospital clinical quality. To evaluate the impact on patient welfare, we next estimate a demand model of hospital choice and use the model estimates to perform welfare analysis and counterfactuals.

5 The Model of Hospital Choice

To evaluate the impact of the HST on patient welfare we need to look at hospital choice that patients would have made had the HST not been launched. To do this, we estimate a hospital choice model, and conduct a reverse counterfactual analysis by switching off the impact of the HST. The entry of the HST reduces travel time and thereby increases number of hospitals in a choice set for patients living close to a HST station. To capture the changes in patients' choice set in our model, we extend the basic conditional logit model by imposing travel time constraints on patients, following literature in geography and transportation. We assume that travel time to each hospital determines whether that hospital is included in patients' choice set or not. If a hospital is located too far away from a patients' location, a patient with travel-time constraints will exclude it from his choice set. This translates to a decrease in the size of the choice set for patients living close to a HST station once the HST is removed.

5.1 Utility and Demand

Each patient i chooses from $J_i \subseteq J$ hospitals in his choice set, indexed $j = 1, ...J_i$ where J is the total number of hospitals in our data. The indirect utility of patient i from choosing hospital j, $j = 1, ..., J$ is defined as

$$u_{ij} = \sum_{l=1}^{L} X_{j,k} \mathbf{Y}_i^{'} \beta_{\cdot,l}^{xy} + Z_j \mathbf{Y}_i^{'} \alpha^z + f(D_{ij}) + \mathbf{X}_j^{'} \beta^x + \alpha Z_j + \varepsilon_{ij} \tag{4}$$

where $\mathbf{X}_j$ is a L vector of hospital characteristics; $\mathbf{Y}_i$ is a K vector of patient-specific demographics; D_{ij} is the travel time from patient i's home to hospital j; Z_j denotes the quality of clinical care at hospital j; ε_{ij} is an idiosyncratic taste shock that is distributed i.i.d. type I extreme value. β^{xy}, α^z and β^x are $K \times L$, $K \times 1$, and $L \times 1$ matrices of coefficients, respectively. Following previous literature on hospital choice, we assume that all patients are admitted to some hospital, and hence there is no outside option in our model.

We define the function $f(\cdot)$ as

$$f(D_{ij}) = \beta_{d_1} D_{ij} + \beta_{d_2} D_{ij}^2 + D_{ij} \mathbf{Y}_i^{'} \beta_{d_3} \tag{5}$$

We estimate equation (4) using logit maximum likelihood approach. One might be concerned about the endogeneity of quality of clinical care in the utility function. Previous literature has found that treating a larger number of cases is associated with better outcomes. Hospitals with higher unobserved quality will attract larger volume of patients, and this will in turn lead to higher quality of clinical care.[19] However, since our measure of quality is controlled for patient case-mix, this issue does not arise (Gaynor et al. (2013b)).

5.2 Choice Set Formation

The entry of the HST has enlarged patients' consideration sets by reducing the travel cost. Hospitals that would not previously have been considered by the patient may now be considered. We model this change consideration sets by imposing a travel-time constraint on patients. We assume that time is a limited resource that constrains the choice options from being evaluated. This assumption is consistent with theoretical and empirical literature in geography and regional science

[19] For more literature on volume-quality relationship, see Birkmeyer et al. (2002); Silber et al. (2010); and Halm et al. (2002).

where a relationship between the available time budget and individuals' destination choice has been established. Our modeling approach follows the Approximate Nested Choice-Set Destination Choice (ANCS-DC) model developed by Thill and Horowitz (1997) which explicitly models the formation of choice sets when individuals have limited time resources.

Each patient has a travel-time threshold T_i which confines his choice set. We let T_i to be a random variable with cumulative distribution $P_T(t;\theta)$, where parameterization by θ allows $P_T(t;\theta)$ to depend on observable patient characteristics. Then, the unconditional probability of patient i choosing hospital j is given as

$$Pr(y_{ij}=1)=\int_{t=0}^{\infty} Pr(y_{ij}=1|J_i)dP_T(t;\theta) \tag{6}$$

where J_{it} is a choice set of individual i who has a travel-time threshold t. Hospitals are discrete and mutually exclusive alternatives. Hence, if the hospitals are sorted according to their travel time from patient's location in ascending order, equation (13) can be simplified to a summation over all the nested sets of hospitals defined by incremental travel-time thresholds, given as

$$Pr(y_{ij}=1)=\sum_{r=1}^{J} Pr(y_{ij}=1|J_i r)p_T(r;\theta) \tag{7}$$

where $p_T(r;\theta)$ is the probability that travel time threshold is between travel times to destinations r and $r+1$, i.e.,

$$p_T(r;\theta)=P_T(t_{r+1};\theta)-P_T(t_r;\theta). \tag{8}$$

The attractive feature of this modeling approach is that it enables us to avoid considering all subset combinations of hospitals which would result in 2^{J-1} choice sets for each patient. The number of possible choice sets is substantially reduced by exploiting the non-random ordering of hospitals based on their travel time from patients' location and travel-time constraints. Therefore, all hospitals that are located closer than any hospital that satisfies the inclusion criterion set by the travel-time threshold are also included in the choice set, and all hospitals that are located further than any hospital that does not satisfy the inclusion criterion are excluded.

Nevertheless, the computational complexity still remains due to large number of hospitals in our data. To further reduce the computational burden, we reduce the support of p_T by restricting the entire series of travel-time thresholds to take only a few discrete values.

Specifically, let $T_{r'}$ denote the travel-time threshold with $r^{'}=1,...,R_T$, where R_T is the number of possible travel-time thresholds after the number of discrete thresholds has been approximated to a few manageable points. We denote the probability that patient i's threshold is $T_{r'}$ as $\pi_{i,r'}$. Let $\pi_{i,r'}$ be a function of concomitant (demographic) variables, defined as

$$\pi_{i,r'} = \frac{\exp(\gamma_r + \mathbf{Y}_i' \phi_{r'})}{\sum_l^{R_T} \exp(\gamma_l + \mathbf{Y}_i' \phi_{r'})} \tag{9}$$

where $\mathbf{Y}_i$ is a $K \times 1$ vector of patient demographics (Gupta and Chintagunta (1994)). Then the probability that hospital j is chosen is

$$Pr(y_{ij} = 1) = \sum_{r'=1}^{R_T} Pr(y_{ij} = 1 | J_{ir'}) \pi_{i,r'} \tag{10}$$

where $J_{ir'}$ is the set of all hospitals h such that $D_{ih} \leq T_{r'}$. The model is estimated by maximizing the following log likelihood function:

$$LL = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \log \left(\sum_{r'=1}^{R_T} Pr(y_{ij} = 1 | J_{ir'}) \pi_{i,r'} \right). \tag{11}$$

6 Demand Estimation Results

We estimate the conditional logit model of hospital choice under travel-time constraint (ANCS-DC). The covariates that enter the utility function are as follows: "TravelTime" refers to travel time (in minutes) between the patient and a hospital in the choice set, and is defined in units of 100 minutes. Age1 is a dummy variable that equals 1 if a patient is between 25 and 50 years of age and 0 otherwise; Age2 is a dummy variable that equals 1 if a patient is between 50 and 75 years of age and 0 otherwise; Age3 is a dummy variable that equals 1 if a patient is above 75 years of age and 0 otherwise; LowIncome is a dummy variable that equals 1 if a patient falls into the lowest income group (total 10 groups); HighSeverityMainSick is a dummy variable that equals 1 if a patient is diagnosed with a disease of mortality rate greater than 0.2 and 0 otherwise; SeverityMainSick is a dummy variable that equals 1 if a patient is diagnosed with a disease of mortality rate within the range [0.1, 0.2) and 0 otherwise; HighSeveritySubSick is a dummy variable that equals 1 if a patient is diagnosed with a comorbidity of mortality rate greater than 0.2 and 0 otherwise; SeveritySubSick is a dummy variable that equals 1 if a patient is diagnosed with a comorbidity of mortality rate within the range [0.1, 0.2) and 0 otherwise; HighSeveritySurgery is a dummy variable that equals 1 if a patient is undergoing a surgery of mortality rate greater than 0.2 and 0 otherwise; SeveritySubSurgery is a dummy variable that equals 1 if a patient is undergoind a surgery with a mortality rate within the range [0.1, 0.2) and 0 otherwise; Disabled is a dummy variable that equals 1 if disabled with kidney and other dysfunction and 0 otherwise; HospitalBed is number of beds in a hospital, and is defined

Table 5: Demand Model Estimates

	(1) Multinomial Logit		(2) ANCS-DC	
	Coefficient	Standard error	Coefficient	Standard error
TravelTime	-4.9798***	0.0242	-3.6825***	0.0312
Mortality	-0.6850***	0.1258	-0.6643***	0.0601
Mortality^2	-5.7107***	0.1878	-5.5142***	0.1707
Mortality×Female	0.1241	0.0939	0.1258**	0.0633
Mortality×Age[25-50)	0.0393	0.1520	0.0856	0.0762
Mortality×Age[50-75)	-0.0087	0.1406	0.0858	0.0704
Mortality×Age[75+)	0.3312**	0.1626	0.3264***	0.1023
Mortality×LowIncome	0.6514***	0.1752	0.6565***	0.1084
Mortality×HighSeverityMainSick	0.6026*	0.3108	0.6262***	0.2045
Mortality×SeverityMainSick	-0.3047*	0.1660	-0.3007***	0.0838
Mortality×HighSeveritySubSick	-0.3024	0.2866	-0.3103	0.1944
Mortality×SeveritySubSick	-0.7323***	0.1731	-0.7429***	0.1232
Mortality×HighSeveritySurgery	-0.3195**	0.1345	-0.2745***	0.0898
Mortality×SeveritySurgery	-0.6331***	0.1411	-0.6161***	0.1122
Mortality×Disabled	0.6148*	0.3573	0.6091***	0.2288
HospitalBed	0.0995***	0.0019	0.1014***	0.0019
HospitalBed×Female	0.0006	0.0014	0.0008	0.0014
HospitalBed×Age[25-50)	-0.0015	0.0023	0.0004	0.0023
HospitalBed×Age[50-75)	0.0145***	0.0021	0.0121***	0.0021
HospitalBed×Age[75+)	-0.0066***	0.0025	-0.0185***	0.0030
HospitalBed×LowIncome	-0.0195***	0.0029	-0.0180***	0.0029
HospitalBed×HighSeverityMainSick	-0.0204***	0.0051	-0.0216***	0.0051
HospitalBed×SeverityMainSick	0.0266***	0.0022	0.0258***	0.0022
HospitalBed×HighSeveritySubSick	-0.0204***	0.0046	-0.0190***	0.0050
HospitalBed×SeveritySubSick	0.0250***	0.0023	0.0250***	0.0023
HospitalBed×HighSeveritySurgery	0.0125***	0.0019	0.0136***	0.0021
HospitalBed×SeveritySurgery	0.0148***	0.0021	0.0162***	0.0021
HospitalBed×Disabled	0.0044	0.0051	0.0021	0.0053
Log-Likelihood	-162,533.49		-1.60,575.7	

Notes: *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

in units of 100 beds.

The estimation results are reported in Column 2 of Table 5. The results are, for the most part, intuitive. Travel time to the hospital plays an important role in patients' decisions when choosing a hospital. The coefficients suggest that patients are less likely to go to hospitals that are located further away from their home.

Our estimates suggest that patients dislike hospitals with poor clinical quality (as measured by adjusted mortality rates) and hospital quality enters patients' utility nonlinearly. We find that patients with more severe comorbidities and patients who are undergoing a more risky surgery are more sensitive to the quality of clinical care. We do not find differences in sensitivity to mortality rates between patients of different genders, ages.

Patients generally prefer larger hospitals (as measured by the number of hospital beds). Lower income patients are less likely to choose larger hospitals. Sicker patients are generally also likely to choose larger hospitals.

Table 7presents the estimates of the parameters of travel-time threshold probabilites. We discretize travel-time threshold into 9 points: 30, 60, 90, 120, 180, 240, 300, 360, and 420 minutes.[20] To reflect decreasing marginal disutility of time spent traveling, threshold points are 30 minutes apart (instead of 60) below 120 minutes of travel time. Several of our estimates show bimodality over time constraints which makes the interpretation complicated. Patients living in metro areas are likely to have choice set to be within 30 minutes or 360 minutes. Our estimates suggest that low income patients are more likely to be time constrained in their choice. This may be due to the monetary cost of traveling long distances. For example, low income patients may not have a car, which is not uncommon given the public transportation infrastructure in South Korea. Older patients are less likely to be time constrained within 30 minutes, but are also less likely to be constrainted within 360 minutes. Since older patients are more likely to be sicker (and have more time if they have retired) they may be less time constrained than younger people, and are willing to travel longer distances. At the same time, since they are older, they may experience difficulty traveling too much, resulting in a bimodal distribution. Coefficients on disease, comorbidity are ambiguous.

We also estimate the hospital choice model using conventional multinomial logit model (without travel-time constraints). The estimates of the parameters are reported in Column 1 of Table 5. In most respects, the signs and magnitude of the estimates are very similar to those obtained using the ANCS-DC model. We prefer to use the ANCS-DC model because the general theory of choice behavior postulates that individuals follow a two-stage decision process in which the alternatives are reduced to a smaller set (consideration set). The construction of these choice sets depend on

[20] Travel time threshold of 420 minutes includes all the hospitals in our data.

factors such as the individual's awareness, feasibility, saliency or accessibility of the alternatives, and mis-specifying the considerations sets may lead to inconsistent parameter estimates. In our setting, we are not able to use an ad-hoc rule such as "15 miles within a patients' home" to define a choice set because a substantial number of patients travel very long distances (even prior to the entry of the HST) to seek better health care services. The ANCS-DC model that we employ is flexible in this manner because it allows the travel time thresholds to be probabilistic, and also to depend on patients' demographic chracteristics. We also use the likelihood ratio test to test whether modelling of the choice set incorporated in the formulation of the ANCS-DC model enhances the representation of the observed hospital choice over the conventional multinomial logit model. The χ^2 statistic for this test is $-2 \times (-162,533 + 160,575) = 3,916$ with 89 degrees of freedom, leading to significance at the 0.01 level. This establishes the relevance of travel-time constraints in modelling the hospital choice problem.

7 Counterfactual Analysis

Using the estimates from the demand model we evaluate the impact of the entry of the HST on patient welfare. We decompose changes in patient welfare arising from (i) the reduced travel time and (ii) changes in hospital quality. We implement this using the following steps: Using pre-HST travel times and pre-HST clinical quality as a baseline, we first calculate changes in patient welfare arising from reduced travel time, assuming hospital quality did not change. Next, using the same baseline, we calculate changes in welfare arising from improved hospital quality, assuming that travel time did not change. Finally, we calculate changes in welfare arising from both, reduced travel time and changes in clinical quality.

We then evaluate the impact of the entry of the HST on patients' health outcomes through its effect on patients' sorting to better hospitals. In other words, we are only interested in quantifying the impact of patients' sorting to better hospitals on their health outcomes (ignoring hospitals' response to greater competition). We compare the number of deaths in post-HST period to a counterfactual scenario where the train is removed while keeping hospital quality constant.

7.1 Changes in Patient Welfare

We compute the changes in patient welfare from the advent of the HST: changes in travel time and changes in hospital quality. Using the parameter estimates from the demand model, we simulate a post-HST scenario where the HST is removed and travel time remains the same as that of the pre-HST level. Recall from our demand model that when travel-time becomes longer (i.e. if the

Table 7: Estimates of the Time Constraint Parameters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	30 min	60 min	90 min	120 min	180 min	240 min	300 min	360 min	360+ min
Intercept	0.6369***	-1.8963***	-73.2109***	-76.4531***	-59.2107***	-57.3386***	-44.5481***	-1.0408***	
	(0.0429)	(0.1960)	(13.5117)	(14.4226)	(10.1155)	(11.3740)	(11.3885)	(0.0930)	
Metro	13.0493***	-57.1276***	-17.0177***	4.0103***	0.0710	10.4360***	-106.4734***	11.5887***	12.1525***
	(0.6013)	(10.3937)	(3.5591)	(1.3847)	(1.0312)	(1.1504)	(22.2776)	(0.5728)	(0.6371)
LowIncome	4.6395***	5.9514***	-2.4540**	5.7028***	0.4013	8.7282***	34.4375***	4.4003***	5.0452***
	(0.4925)	(0.3976)	(1.0001)	(1.9156)	(1.0807)	(1.4648)	(6.1478)	(0.4886)	(0.6189)
Female	-9.4019***	-50.6345***	-23.0915***	-6.8108***	-2.4730**	-25.3802***	-43.1934***	-9.6039***	-9.0977***
	(0.5090)	(8.3970)	(4.4153)	(1.3689)	(1.0027)	(3.8074)	(7.0902)	(0.5097)	(0.5062)
Age[25-50)	-2.2862***	-19.8424***	-8.7353***	12.7399***	-1.5121	-14.9623***	-7.7816***	-0.0645	-2.9835***
	(0.0978)	(3.8921)	(1.4663)	(3.4379)	(1.3587)	(2.0792)	(1.5691)	(0.0848)	(0.1559)
Age[50-75)	-10.1412***	-8.3876***	-9.4569***	-25.5520***	1.4453	6.7198*	0.4536	-7.2567***	-11.1217***
	(0.3708)	(0.4112)	(1.6510)	(4.0144)	(1.1601)	(3.5826)	(2.2918)	(0.4097)	(0.8617)
Age[75+]	-5.6053***	-3.6618***	-10.1214***	-4.5019***	-14.3094***	-19.5667***	51.1738***	-18.0609***	-4.2723***
	(0.2709)	(0.6129)	(1.8590)	(1.0331)	(2.7953)	(2.8174)	(13.5017)	(2.9500)	(0.2721)
MainSickRisk	-14.4068***	-24.4119***	-2.8452***	-5.3036***	-4.1232***	21.7233***	27.6011***	-9.2244***	-16.2573***
	(0.4837)	(2.0860)	(1.0809)	(1.4082)	(1.3202)	(4.6032)	(8.8364)	(0.6646)	(0.9178)
SubSickRisk	-15.4288***	-22.6418***	4.3966***	-1.1159	-4.8191***	4.2365***	138.5419***	-13.3901***	-17.3975***
	(0.4781)	(1.7303)	(1.0936)	(1.0002)	(1.2984)	(1.2838)	(30.3550)	(0.8009)	(0.6010)
SurgeryRisk	23.4182***	24.4731***	-4.6975***	1.5432	-5.7751***	28.6760***	-287.2755***	22.3997***	24.5485***
	(0.8438)	(1.0127)	(1.2966)	(1.1338)	(1.2748)	(5.4696)	(60.4609)	(1.1730)	(0.7425)

Metro area corresponds to 7 metropolitan cities consisting of Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, and Ulsan. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

travel time is that of the pre-HST level), constraints imposed on patients' travel-time will force them to remove further-located hospitals (which are included in the choice set if the travel time is that of the post-HST level) from the consideration set.

The expected patient surplus (in utils) for patient i with post-HST travel time can be expressed as

$$E(Surplus_i^{train}) = \sum_{r=1}^{R_T} E(Surplus_{i|r}^{train}) \cdot \pi_{ir} = \sum_{r=1}^{R_T} E\left(\max_{j \in J_{i|r}^{train}} (\bar{U}_{ij} + \varepsilon_{ij})\right) \cdot \pi_{ir} \tag{12}$$

while in with pre-HST travel time, it is expressed as

$$E(Surplus_i^{no\ train}) = \sum_{r=1}^{R_T} E(Surplus_{i|r}^{no\ train}) \cdot \pi_{ir} = \sum_{r=1}^{R_T} E\left(\max_{j \in J_{i|r}^{no\ train}} (\bar{U}_{ij} + \varepsilon_{ij})\right) \cdot \pi_{ir}. \tag{13}$$

For patients living in treated regions, the choice set $J_{i|r}^{train}$ can differs from $J_{i|r}^{no\ train}$ because changes in travel time changes the composition of hospitals in a choice set. Assuming that ε_{ij} is distributed i.i.d extreme value, the above expression can be rewritten as a logit-inclusive value

$$E(Surplus_i^{train}) = \sum_{r=1}^{R_T} \ln\left(\sum_{j \in J_{i|r}^{train}} \exp(\bar{U}_{ij})\right) \pi_{ir} \tag{14}$$

and

$$E(Surplus_i^{no\,train}) = \sum_{r=1}^{R_T} \ln\left(\sum_{j \in J_{i|r}^{\text{no train}}} \exp(\bar{U}_{ij})\right) \pi_{ir} \tag{15}$$

The average change in surplus per patient is given as

$$E(\Delta Surplus_i) = \frac{1}{N} \sum_{i=1}^{N} E(Surplus_i^{train}) - E(Surplus_i^{no\ train}) \tag{16}$$

where N is the number of patients in post-HST period.

We first calculate the quantity in equation (16) assuming the quality of clinical care did not change. This allow us to evaluate the changes in welfare from the reduction in travel time only. To obtain these quantities, we used pre-HST travel time as well as pre-HST quality of clinical care as our baseline. The results are reported in Table 9, panel A. Assuming the quality of clinical care did not change, patients living in treated regions experience an average increase of 0.2815 units in expected utility. This increase in welfare arises from reduction in travel time, and the resulting

Table 9: Counterfactual Analysis

Panel A. Changes in Patient Welfare				
		Change in travel time No change in quality	No change in travel time Chage in quality	Change in travel time Chage in quality
Treated Patients	ΔUtility	0.4058	0.0170	0.2973
	Dollar Value	$1,269	$76.82	$1,336
Control Patients	ΔUtility	0	0.0197	0.0197
	Dollar Value	0	$88.51	$88.51

Panel B. Impact of Sorting on Patient Survival (number of lives saved)		
	No change in quality	Chage in quality
Treated Patients	0.2510	12
Control Patients	0	4
Total	0.2510	16

Notes: Panel A reports the changes in patient welfare in terms of expected utility (unit in expected utility) and dollar value under various counterfactual scenarios. Panel B reports the number of patients that would survive as a result of sorting to better hospitals.

ability of patients to sort to better hospitals. There is no change in welfare for patients living in control regions as they do not benefit from the entry of HST. Since there is no price coefficient in the demand model due to the absence of price mechanism in this market, we cannot direcly convert the welfare change from utils into a dollar value. Therefore, following Gaynor et al. (2016), we first translate the gains in terms of the preferece over distance, and then convert the welfare estimates into a dollar value using additional data from other sources.[21] Comparing the gains in utils to the preference over distance, we find that the welfare effect of the reduction in travel distance for the treated patients corresponds to 7.6 minutes reduction in travel time.[22] Applying a $167 value per minute reduction in travel time (Gaynor et al. (2016); Gowrisankaran et al. (2015)), the reduction in travel time yields a welfare effect of approximately $2,071 ($167 \times 7.6 = 1,269$) per patient.[23]

Next, we calculate the changes the welfare arising from changes in quality of clinical care, holding the changes in travel time constant. Patients living in treated regions experience an average increase

[21] Gowrisankaran et al. (2015) estimate that a one minute reduction in travel time to hospitals increases patient surplus by $167.

[22] $0.2815/(-3.6825) = -0.0764$, where -3.6825 is the coefficient on travel time. Travel time in the regression is defined in units of 100 minutes.

[23] Due to the travel time constraint in our model, the number of hospitals that a patient considers changes when travel time to each hospital changes. Increased number of hospitals will affect the welfare gains because the term in parentheses in equations 14 and 15 is simply the denominator of the logit choice probability (which is simply the outcomes of the mathematical form of the extreme value distribution, and has no economic meaning (Train (2009)). Therefore, we also calculate changes in welfare arising from reduced travel time (holding hospital quality constant) while holding the number of hospitals in the consideration set constant. This method yields an increase of XXX units in expected utility, which translates to a reduction in 11 minutes of travel time, and amounts to approximately XXX per patient.

of 0.0170 units in expected utility. Patients living in control regions experience an average increase of 0.0197 units in expected utility. The increase in expected utility for patients living in control regions arises from the fact that they face higher clinical quality although they do not benefit from the new transportation system. Applying the same back of the envelope calculation as before to monetize the gains in utils, the improvement in clinical quality yields a welfare effect of approximately \$76.82 per patient for patients living in treated regions, and \$88.51 per patient for patients living in control regions.[24]

Finally, we calculate the changes the welfare arising from both, changes in travel time and changes in quality of clinical care. Patients living in treated regions experience an average increase of 0.2973 units in expected utility. Patients living in control regions experience an average increase of 0.1189 units in expected utility (identical to the case when quality of clinical care changes, holding the changes in travel time constant). This yields a welfare effect of approximately \$1,336 per patient for patients living in treated regions, and \$88.51 per patient for patients living in control regions.[25]

7.2 The Impact of Patients' Sorting on Survival

The HST has enabled patients to choose hospitals that were previously difficult to consider due to long travel distances. Therefore the HST has not only improved the quality of clinical care through increased competition among hospitals, but has also increased the size of the choice set for the patients which in turn has resulted in patients' sorting to better hospitals. One way to directly measure the benefits generated by the HST through its impact on patient sorting is to calculate how many patients would have died in the post-HST period if the HST were to be removed, i.e. post-HST period patients are faced with the pre-HST level travel time to the hospitals.

To implement this, we closely follow Gaynor et al. (2016) and calculate the expected differences in mortality across all patients:

$$E(\Delta Mortality) = \sum_i \left[E(Mortality_i)^{\text{train}} - E(Mortality_i)^{\text{no train}}\right] \tag{17}$$

[24]The welfare effect of the improvement in clinical care for the treated patients corresponds to approximately 0.46 minutes reduction in travel time, $0.0170/(-3.6825) = -0.0046$. Multiplying this by the value per minute reduction in time, we get $0.46 \times 167 = 76.82$. Similarly, the welfare effect of the improvement in clinical care for the control patients corresponds to approximately 0.53 minutes reduction in travel time. Multiplying this by the value per minute reduction in time, we get $0.53 \times 167 = 88.51$.

[25]The welfare effect of the improvementfor the treated patients corresponds to approximately 8 minutes reduction in travel time, $0.2973/(-3.6825) = -0.0807$. Multiplying this by the value per minute reduction in time, we get $8 \times 167 = 1,336$.

where

$$E(Mortality_i)^{\text{train}} = \sum_j Pr_{ij}^{\text{train}} \cdot Prob(Mortality_i|choice = j, Health_i) \tag{18}$$

and

$$E(Mortality_i)^{\text{no train}} = \sum_j Pr_{ij}^{\text{no train}} \cdot Prob(Mortality_i|choice = j, Health_i). \tag{19}$$

Equations (18) and (19) denote the mortality probability with post-HST travel time and pre-HST travel time, respectively. $Mortality_i$ is an indicator variable which takes value 1 if the patient dies and 0 otherwise. Term $Prob(Mortality_i|choice = j, Health_i)$ that appears in both equations is the predicted probability of death conditional on choice of hospital and patient's health status. Since we estimated adjusted mortality rates and the coefficients on patient case-mix using linear probability model, the predicted mortality probability may not necessarily lie within the (0,1) interval. Therefore, we obtain the predicted probability of death using the Linear Discriminant Model (LDM) method which transforms the coefficients from the linear probability model into maximum likelihood estimates of the parameters of a linear discriminant model. [26] The LDM implies a logistic regression model for the dependence of the outcome on the predictors. This method ensures that the predicted probabilities lie within the (0,1) interval.

The results are reported in Table 9, panel B. Our estimates from this counterfactual analysis suggest that 0.25 more lives can be saved from patients' sorting. Since our data is a 2 percent random sample of the entire population, this translates to approximately 12 lives over the five quarters, which is equivalent to 10 lives on an annual basis.[27]

Next, we calculate how many more lives are saved due to patient sorting when the quality of clinical care also responds to the entry of HST. Our estimates suggest that 12 lives (480 lives on an annual basis) of patients living in treated regions and 4 lives (160 lives) of patients living in control regions can be saved.[28]

8 Conclusion

This paper exploits the entry of HST in South Korea, which reduced patients' travel costs, increasing substitutability among hospitals and thereby increasing hospital competition. This exogenous shock allows us to look at the impact of reduced travel time on patient behavior as well as to study the

[26] https://statisticalhorizons.com/better-predicted-probabilities

[27] $0.25 \times 50 \times (4/5) = 10$

[28] $12 \times 50 \times (4/5) = 480$ and $4 \times 50 \times (4/5) = 160$

causal impact of competition on hospital quality. Taking advantage of the differential effects of the entry of the HST on hospitals located in different regions of the country, we use a difference-in-differences estimator to examine the impact of competition on health outcomes measured by 30-day mortality rates following admissions for cardiovascular or neurological surgeries. On the methodological side, we utilize the heterogeneous effects of the entry of the HST on patients living in different areas of the country to obtain a reliable measure of hospital-level quality of clinical care.

We find that the entry of the HST improves patient mobility, and that intensified hospital competition leads to an improvement in clinical quality. To evaluate the overall impact of HST on patient welfare, we estimate a structural model of hospital choice, allowing for a flexible formation of patients' consideration set. We find that patients living near a HST station experience an improvement in welfare arising from reduction in travel time as well as improvement in hospital quality. Patients living further away from HST stations also experience an improvement in welfare because while they do not benefit from the reduced travel time, they benefit from the improvement in the quality of treated hospitals. We also find that HST has led to a substantial improvement on the probability of patient survival through its effect on patient sorting, even while holding hospital quality constant.

Overall, our paper suggests that increased hospital competition can lead to beneficial health outcomes and that an improvement in transportation infrastructure can have a beneficial impact on patient helath by facilitating patients' sorting to better hospitals through lower travel costs.

References

BANERJEE, A., E. DUFLO, AND N. QIAN (2012): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *NBER Working Paper No. 17897*.

BECKERT, W., M. CHRISTENSEN, AND K. COLLYER (2012): "Choice of NHS-funded Hospital Services in England," *The Economic Journal*, 122, 400–417.

BIRKMEYER, J. D., A. E. SIEWERS, E. V. FINLAYSON, T. A. STUKEL, F. L. LUCAS, I. BATISTA, H. G. WELCH, AND D. E. WENNBERG (2002): "Hospital Volume and Surgical Mortality in the United States," *The New England Journal of Medicine*, 346, 1128–37.

BREKKE, K. R., L. SICILIANI, AND O. R. STRAUME (2011): "Hospital Competition and Quality with Regulated Prices," *The Scandinavian Journal of Economics*, 113, 444–469.

CAPPS, C., D. DRANOVE, AND M. SATTERTHWAITE (2003): "Competition and Market Power in Option Demand Markets," *The RAND Journal of Economics*, 34, 737–763.

CHO, N.-G. AND J.-K. CHUNG (2008): "High speed rail construction of Korea and its impact," *Korea Research Institute for Human Settlement*, 12.

COLLA, C., J. BYNUM, A. AUSTIN, AND J. SKINNER (2016): "Hospital Competition, Quality, and Expenditures in the U.S. Medicare Population," *NBER Working Paper No. 22826*.

DAFNY, L., K. HO, AND M. VARELA (2013): "Let Them Have Choice: Gains from Shifting Away from Employer-Sponsored Health Insurance and toward an Individual Exchange," *American Economic Journal: Economic Policy*, 5, 32–58.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

GAYNOR, M., R. MORENO-SERRA, AND C. PROPPER (2013a): "Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service," *American Economic Journal: Economic Policy*, 5, 134–66.

GAYNOR, M., C. PROPPER, AND S. SEILER (2013b): "Free to Choose? Reform and Demand Response in the English National Health Service," *Working Paper*.

——— (2016): "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service," *American Economic Review*, 106, 3521–57.

GAYNOR, M. AND W. B. VOGT (2003): "Competition among Hospitals," *The RAND Journal of Economics*, 34, 764–785.

GEWEKE, J., G. GOWRISANKARAN, AND R. J. TOWN (2004): "Bayesian Inference for Hospital Quality in a Selection Model," *Econometrica*, 71, 1215–1238.

GOWRISANKARAN, G., A. NEVO, AND R. TOWN (2015): "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry," *American Economic Review*, 105, 172–203.

GOWRISANKARAN, G. AND R. J. TOWN (1999): "Estimating the quality of care in hospitals using instrumental variables," *Journal of Health Economics*, 18, 747–767.

——— (2003): "Competition, Payers, and Hospital Quality," *Health Services Research*, 38, 1403–1422.

Gupta, S. and P. K. Chintagunta (1994): "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models," *Journal of Marketing Research*, 31, 128–136.

Halm, E., C. Lee, and M. Chassin (2002): "Is volume related to outcome in health care? A systematic review and methodologic critique of the literature," *Annals of Internal Medicine*, 137, 511–20.

Heuermann, D. F. and J. F. Schmieder (2018): "The Effect of Infrastructure on Worker Mobility: Evidence from High-Speed Rail Expansion in Germany," *NBER Working Paper No. 24507*.

Ho, K. (2006): "The welfare effects of restricted hospital choice in the US medical care market," *Journal of Applied Econometrics*, 21, 1039–1079.

——— (2009): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 99, 393–430.

Kessler, D. P. and M. B. McClellan (2000): "Is Hospital Competition Socially Wasteful?" *The Quarterly Journal of Economics*, 115, 577–615.

Kim, J., J. Lee, W. Yoo, and S. Park (2008): "KTX ui GungangYunghyangPyungka," *Korea Institute for Health and Social Affairs.*

Lewis, M. S. and K. E. Pflum (2017): "Competition and Quality Choice in Hospital Markets," Working Paper.

Propper, C., S. Burgess, and D. Gossage (2008): "Competition and Quality: Evidence from the NHS Internal Market 1991-9," *The Economic Journal*, 118, 138–170.

Propper, C., S. Burgess, and K. Green (2004): "Does competition between hospitals improve the quality of care?: Hospital death rates and the NHS internal market," *Journal of Public Economics*, 88, 1247–1272.

Qin, M., G. John, and M. A. Vitorino (2018): "Planes, Trains and Co-Opetition: Evidence from China," *Working Paper*.

Qin, Y. (2016): ""No county left behind?" The distributional impact of high-speed rail upgrades in China," *Journal of Economic Geography*, 17, 489–520.

Shen, Y. C. (2003): "The effect of financial pressure on the quality of care in hospitals," *Journal of Health Economics*, 22, 243–269.

Shortell, S. M. and E. F. Hughes (1988): "The Effects of Regulation, Competition, and Ownership on Mortality Rates Among Hospital Inpatients," *New England Journal of Medicine*, 318, 1100–07.

Silber, J. H., P. R. Rosenbaum, J. Tanguy, R. N. Ross, L. J. Bressler, O. Even-Shoshan, S. A. Lorch, and K. G. Volpp (2010): "The Hospital Compare Mortality Model and the Volume-Outcome Relationship," *Health Services Research*, 45, 1148–6773.

Tay, A. (2003): "Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Differentiation," *The RAND Journal of Economics*, 34, 786–814.

Thill, J.-C. and J. L. Horowitz (1997): "Modelling Non-Work Destination Choices with Choice Sets Defined by Travel-Time Constraints," *Recent Developments in Spatial Analysis*, 186–208.

Train, K. (2009): *Discrete Choice Methods with Simulation*, Cambridge University Press, 2 ed.

Weber, S. and M. Péclat (2017): "A simple command to calculate traveldistance and travel time," *The STATA Journal*, 17, 962–971.

Appendix A: Hospital-Level Case-Mix

Hospital-level raw mortality rates do not correctly reflect the true quality of clinical care due to differences in patients' health status across hospitals (referred to as hospital's "case-mix") i.e., hospitals with a larger number of sicker patients are more likely to have higher mortality rates. It is therefore essential to take into account differences in patient case-mix across hospitals, especially since we are using patients undergoing various types of different surgeries. Specifically, we include the following hospital-level case-mix as control variables: Above70 (fraction of patients older than 70 years of age), SurgeryRisk (average deathrate of all surgeries conducted in each hospital, where deathrate of a surgery is calculate as the death rate of each surgery over our entire sample), DiseaseRisk (average deathrate of patients' diagnosed disease, where deathrate of a disease is calculate as the death rate of each disease over our entire sample), ComorbidityRisk (average deathrate of patients' diagnosed comorbidity, where deathrate of a comorbidity is calculate as the death rate of each comorbidity over our entire sample. If a patient does not have a comorbidity, this variable equals 0), DisabledFrac (fraction of patients with a kidney and other dysfunction), DisabilitySeverity (severity of disability, 1 mild, 2 severe), LowIncomeFrac (fraction of patients with low income).[29]

Table A.1 reports the Diff-in-Diff estimates of the impact of competition on raw mortality rates controlling for hospital-level case mix. As expected, hospitals with more riskier diseases, riskier comorbidities, and more severe disabilities have higher mortality rates. Income and Age do not seem to affect hospital level mortality rates. After controlling for DisabilitySeverity, the coefficient on Disabled becomes negative. In Table A.2 we check the robustness of our results using only a subset of the control variables. Our results remain unchanged.

[29] In the data there are various categories of disabilities, such as intelectual disorder, mental disorder, hearing disability, etc. Since some of these disabilities are not likely to affect the mortality of a patient, we only consider Kidney Dysfunction and "Other Dysfuction". Other dysfunction includes (but does not distinguish between) speech disability, austistic disorder, cardiac dysfunction, respiratory dysfunction, liver dysfunction, facial disfigurement, intestinal fistular/urinary fistular. Although speech disability and autism may be unrelated to deathrate, we are not able to distinguish these disabilities from more critical ones such as cardiac and liver dysfunction.

Table A.1: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Raw Mortality Rates	(1)	(2)	(3)	(4)	(4)
Post	-0.002	0.001	0.004	0.007	0.008
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)
Treated×Post		**-0.008***	**-0.011****	**-0.014****	**-0.014****
		(0.005)	(0.005)	(0.006)	(0.006)
Above70	0.001	0.004	0.007	0.006	-0.000
	(0.048)	(0.047)	(0.047)	(0.046)	(0.047)
SurgeryRisk	-0.039	-0.026	-0.024	-0.027	-0.029
	(0.112)	(0.110)	(0.108)	(0.108)	(0.108)
DiseaseRisk	0.697***	0.703***	0.702***	0.725***	0.721***
	(0.234)	(0.229)	(0.228)	(0.224)	(0.226)
ComorbidityRisk	1.113***	1.122***	1.126***	1.136***	1.128***
	(0.392)	(0.388)	(0.379)	(0.373)	(0.376)
DisabledFrac	-1.939***	-2.187***	-2.280***	-2.274***	-2.106***
	(0.707)	(0.766)	(0.718)	(0.710)	(0.691)
DisabilitySeverity	2.310***	2.562***	2.660***	2.654***	2.471***
	(0.820)	(0.879)	(0.835)	(0.825)	(0.801)
LowIncomeFrac	-0.065	-0.071	-0.069	-0.069	-0.057
	(0.141)	(0.138)	(0.140)	(0.138)	(0.137)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.8321	0.8351	0.8372	0.8398	0.8386

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table A.2: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

height	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Raw Mortality Rates	(1)	(2)	(3)	(4)	(4)
Post	-0.003	0.000	0.004	0.006	0.008
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Treated×Post		**-0.008***	**-0.011****	**-0.014****	**-0.014****
		(0.005)	(0.005)	(0.006)	(0.006)
DiseaseRisk	0.662**	0.683***	0.687***	0.705***	0.693***
	(0.261)	(0.253)	(0.245)	(0.239)	(0.244)
ComorbidityRisk	1.0751***	1.092***	1.097***	1.106***	1.100***
	(0.349)	(0.346)	(0.334)	(0.330)	(0.333)
DisabledFrac	-2.019***	-2.259***	-2.341***	-2.342***	-2.177***
	(0.751)	(0.804)	(0.757)	(0.746)	(0.727)
DisabilitySeverity	2.393***	2.636***	2.723***	2.725***	2.545***
	(0.872)	(0.925)	(0.880)	(0.867)	(0.844)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.8315	0.8344	0.8366	0.8391	0.8382

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Appendix B: Adjusted Mortality Rates

We want to use 30-day mortality following a surgery as our measure of hospital quality as it is the most commonly used outcome-based measure. However, we do not observe the exact date of the surgery in our data. To complicate matters further, we only observe the year and month of patients' death instead of the exact date. Therefore our (proxy) measure of 30-day mortality rate is obtained as follows: We construct a dummy variable M whose element μ_i takes value 1 if (i) patient i who was admitted to hospital in month mm_i day dd_i and year $yyyy_i$ dies either in month mm_i and year $yyyy_i$ or in month $mm_i + 1$ and year $yyyy_i$ for $mm_i = 1, ...11$ and (ii) length of hospital-stay does not exceed 30 days. If patient was admitted to hospital in $mm_i = 12$ and year $yyyy_i$, μ_i takes value 1 if patient dies in month mm_i and year $yyyy_i$ or in Janurary of year $yyyy_i + 1$.

We then use this mortality dummy variable M to obtain the case-mix adjusted mortality rate by estimating the following linear probability model pooled across both pre- and post-HST periods:

$$M = C\boldsymbol{\psi} + H\boldsymbol{\gamma} + (S + \eta) \tag{20}$$

where M is a vector of dummy variable whose elements are switched on if a patient died. C is a matrix of hospital-time period dummy variables, and H and S are patients' observed and unobserved health status, respectively.[30] The estimated hospital fixed effects parameter, $\boldsymbol{\psi}$, is the case-mix adjusted mortality rate that will be used in our difference-in-differences estimation as well as in our structural model of hospital choice. The corresponding expression for an individual observation is as follows:

$$\mu_{it} = \boldsymbol{\psi}'\boldsymbol{c}_i + \boldsymbol{\gamma}'\boldsymbol{h}_i + s_{it} + \eta_{it}$$

Following Gaynor et al. (2013b) hospital dummies are stacked in a block-diagonal matrix where each block represents each period. Along with patients' observed case-mix, the data are arranged as

$$X = \begin{bmatrix} C_{pre} & & H_{pre} \\ & C_{post} & H_{post} \end{bmatrix}$$

where all elements in the matrix other than C_t and H_t are equal to zero. C_t is given by

[30] In patient characteristics matrix H, we include female dummy, age, income group, riskiness of the patient's surgery, riskiness of the patient's disease, riskiness of the patient's comorbidity and a disability dummy variable.

Table A.3: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

	undadjusted raw mortality	unadjusted raw mortality	adjusted mortality OLS	adjusted mortality IV	adjusted mortality IV (rescaled)
	(1)	(2)	(3)	(4)	(5)
Post	-0.00078	0.0020	0.004	0.025	0.004
	(0.0031)	(0.0077)	(0.007)	(0.030)	(0.004)
Treated×Post		-0.0042	**-0.014***	**-0.082****	**-0.012****
		(0.0082)	(0.007)	(0.038)	(0.006)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		55	55	55	55
Treated Hospitals		112	112	112	112
Observations	334	334	334	334	334
R-squared	0.6885	0.6892	0.6011	0.6181	0.6181

Notes: This table compares the DID estimates using raw mortality rates which do not control for hospital-level case mix. Models are estimated using OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. Treatment is defined as being located within 15-miles of the HST station.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table A.4: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care (unadjusted raw mortality)

unadjusted raw mortality	Pre-Post (1)	5-mile Treatment (2)	10-mile Treatment (3)	15-mile Treatment (4)	20-mile Treatment (4)
Post	-0.00078	-0.00044	0.001	0.0020	0.0048
	(0.0031)	(0.0043)	(0.006)	(0.0077)	(0.0096)
Treated×Post		-0.00091	-0.0031	-0.0042	-0.0075
		(0.0061)	(0.0069)	(0.0082)	(0.010)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		55	55	55	55
Treated Hospitals		112	112	112	112
Observations	334	334	334	334	334
R-squared	0.6885	0.6885	0.6889	0.6892	0.6905

Notes: This table compares the DID estimates using raw mortality rates which do not control for hospital-level case mix. Models are estimated using OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. Treatment is defined as being located within 15-miles of the HST station.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

$$C_t = \begin{bmatrix} c_{11}^t & \dots & c_{1J-1}^t \\ \vdots & \ddots & \vdots \\ c_{n_t 1}^t & \dots & c_{n_t J-1}^t \end{bmatrix}$$

where n_t is the number of patients in period t, and the elements c_{ij}^t takes value one if patient i chooses hospital j among J alternatives in period t, and zero otherwise.

Allowing the hospital fixed effects to vary for each period, we need to instrument $(2 \cdot J - 1)$ hospital choice dummies for each period, requiring us of at least as many number of instruments. We use travel time to each of the J hospitals and additional J set of a dummy variables which equals 1 if a given hospital is the closest one to the patient, which gives us a total of $2 \cdot J$ instruments for each period. Specifically, we define travel time for patient i to hospital j in period t as

$$\text{TravelTime}_{ijt} = \begin{cases} \min(\text{cartime}_{ij}, \text{traintime}_{ij}) & \text{if } i \text{ lives in treated region in } t = post \\ \text{cartime}_{ij} & \text{otherwise} \end{cases}$$

The matrix of instrumental variables is constructed as

$$Z = \begin{bmatrix} Z_{pre} & & H_{pre} \\ & Z_{post} & H_{post} \end{bmatrix}$$

where Z_t is a matrix of $2 \cdot J$ instruments which is given by

$$Z_t = \begin{bmatrix} z_{11}^t & \dots & z_{1K}^t \\ \vdots & \ddots & \vdots \\ z_{n_t 1}^t & \dots & z_{n_t K}^t \end{bmatrix}$$

and $K = 2 \cdot J$ denotes the number of instruments.

Formal specification tests for the validity of our instruments are provided in Table B.1. Our overidentifying restrictions are valid as we fail to reject the null of the Sargan-Hansen overidentification test. We reject the null hypothesis of the Hausman Endogeneity test which means that our OLS and IV estimates are statistically different. We also perform the Wald-Test of Weak Instruments and reject the hypothesis that our instruments are weak. These tests provide support for the validity of our IV specification.

In Table B.2 we report the estimates of the effect of patients' observed case-mix on patient mortality from OLS and IV methods. Figure 3 shows the scatterplot of the point estimates of hospital fixed effects (quality of clinical care) obtained using IV against raw (case-mix unadjusted) mortality rates. Consistent with previous research that use the same methodology, the IV point estimates have a much wider spread than the raw mortality rates.

Sargan-Hansen	χ^2	274.0181
Overidentification Test	P-value	0.9936
Hausman	χ^2	3,068
Endogeneity Test	P-value	0.0001
Wald-Test of	χ^2	31,610
Weak Instruments	P-value	0.0001

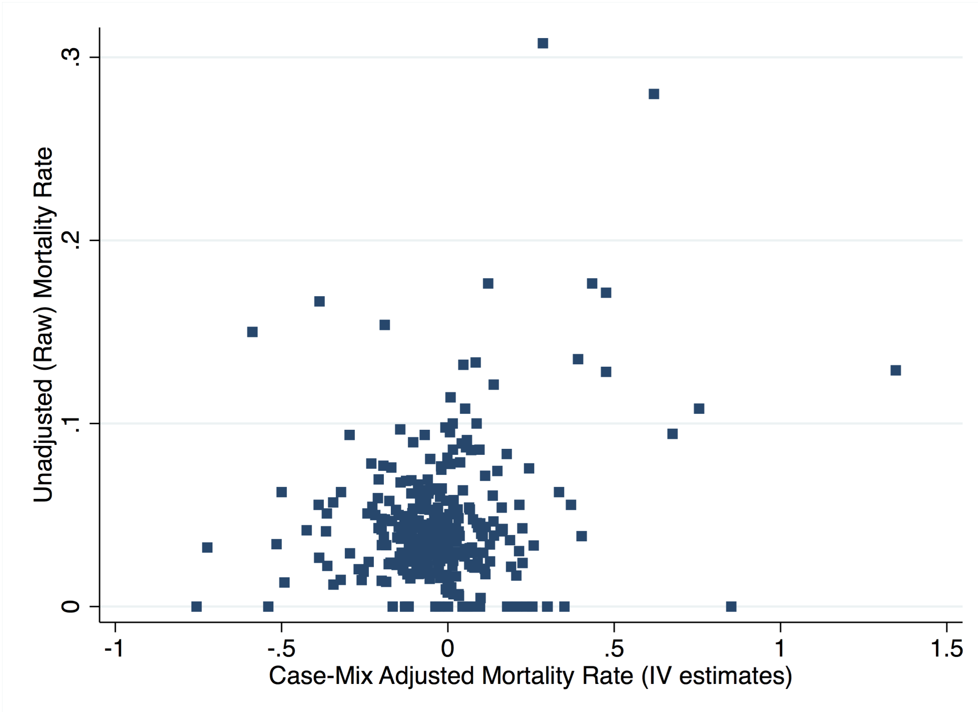
Table B.1: Tests for Validity of Instruments

Table B.2: Estimates of the effect of patient characteristics on mortality from OLS and IV methods (standard errors in parentheses)

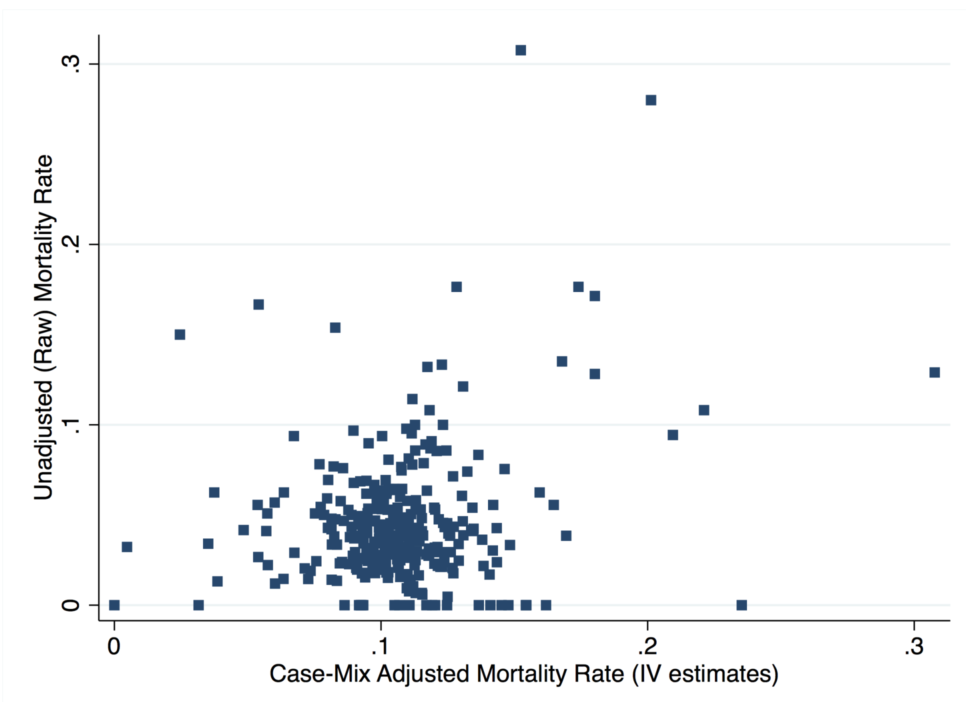
	OLS Coefficients	IV Coefficients
Female	0.003**	0.002
	(0.001)	(0.003)
MediumIncome	0.008***	0.009**
	(0.002)	(0.004)
HighIncome	0.007***	0.006
	(0.002)	(0.004)
Age[20-40)	-0.002***	-0.003
	(0.002)	(0.004)
Age[40-60)	-0.011	-0.010**
	(0.002)	(0.004)
Age[60-80)	0.0064***	0.008*
	(0.002)	(0.005)
Age[80+)	0.079***	0.072***
	(0.005)	(0.010)
MainsickRisk	0.518***	0.512***
	(0.010)	(0.020)
SubsickRisk	0.523***	0.537***
	(0.010)	(0.017)
SurgeryRisk	0.383***	0.382***
	(0.005)	(0.012)
Disabled	-0.022***	-0.032***
	(0.004)	(0.008)
DisabilitySevere	0.003***	0.006
	(0.012)	(0.016)

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants.

*** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.



(a) Case-mix adjusted IV estimates



(b) Case-mix adjusted IV estimates (rescaled)

Figure 3: Relationship between adjusted and raw mortality rates

Appendix C: Alternative Definitions of Treatment

Table C.1: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

OLS Estimated	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Adjusted Mortality Rates	(1)	(2)	(3)	(4)	(5)
Post	-0.005**	-0.002	0.001	0.004	0.007
	0.003	0.004	0.005	0.007	0.008
Treated×Post		**-0.001****	**-0.010***	**-0.014***	**-0.016***
		0.005	0.006	0.007	0.008
	(p: 0.045)	(p:0.047)	(p:0.076)	(p:0.052)	(p:0.056)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.5868	0.5947	0.5961	0.6011	0.6040

Notes: This table shows diff-in-diff estimates for case-mix adjusted mortality rates obtained via OLS for various definitions of hospital treatment (hospitals located within 5 miles, 10 miles, 15 miles, and 20 miles). For various definitions of treatment, we can consistenly see that the diff-in-diff coefficient is negative and significant at the 10 percent level. Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table C.2: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care

IV Estimated Adjusted Mortality Rates	Pre-Post (1)	5-mile Treatment (2)	10-mile Treatment (3)	15-mile Treatment (4)	20-mile Treatment (4)
Post	-.0296	-0.025	0.016	0.025	0.026
	(0.019)	(0.026)	(0.027)	(0.030)	(0.036)
Treated×Post		**-0.014**	**-0.078****	**-0.082****	**-0.075***
		(0.037)	(0.038)	(0.038)	(0.043)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.6086	0.6088	0.6181	0.6181	0.6156

Notes: This table shows diff-in-diff estimates for case-mix adjusted mortality rates obtained via IV method for various definitions of hospital treatment (hospitals located within 5 miles, 10 miles, 15 miles, and 20 miles). For various definitions of treatment (except for 5 mile treatment), we can consistenly see that the diff-in-diff coefficient is negative and significant at either 5 percent or 10 percent level. Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table C.3: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care [Rescaled]

IV Estimated	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Adjusted Mortality Rates	(1)	(2)	(3)	(4)	(4)
Post	-0.004	-0.004	0.002	0.004	0.004
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Treated×Post		**-0.002**	**-0.011****	**-0.012****	**-0.011***
		(0.005)	(0.006)	(0.006)	(0.006)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.6086	0.6088	0.6181	0.6181	0.6156

Notes: This table shows diff-in-diff estimates for case-mix adjusted mortality rates obtained via IV method (rescaled for interpretation purpose) for various definitions of hospital treatment (hospitals located within 5 miles, 10 miles, 15 miles, and 20 miles). For various definitions of treatment (except for 5 mile treatment), we can consistenly see that the diff-in-diff coefficient is negative and significant at either 5 percent or 10 percent level. Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Appendix D

	Treated Hospitals			Control Hospitals		
	Pre-HST	Post-HST	t-stat	Pre-HST	Post-HST	t-stat
			Patient Treatment: 10 miles			
Control Patients	0.288	0.292	t:0.7130	0.073	0.076	t:0.8973
	(0.453)	(0.455)		(0.260)	(0.266)	
N	7,202	8,065	p:0.4758	7,865	9,552	p:0.3696
Treated Patients	0.132	0.151	**t:3.3867**	0.335	0.376	t:1.2550
	(0.338)	(0.358)		(0.473)	(0.485)	
N	7,202	8,065	p:0.0007	385	492	p:0.2098
			Patient Treatment: 15 miles			
Control Patients	0.478	0.476	t:0.2291	0.067	0.072	t:1.1415
	(0.500)	(0.500)		(0.251)	(0.258)	
N	6,075	7,496	p:0.8188	7,487	9,084	p:0.2537
Treated Patients	0.110	0.120	**t:2.0532**	0.254	0.263	t:0.3915
	(0.313)	(0.324)		(0.436)	(0.440)	
N	10,185	11,617	p:0.0401	713	880	p:0.2537

Notes: This table shows the changes in proportion of patients (excluding Seoul and surrounding area) who traveled more than 50 miles to arrive at the hospitals. There is a significant increase in proportion of treated patients traveling more than 50 miles to arrive at treated hospitas. Standard deviation in parentheses.

Table C.1: Proportion of Patients who Traveled to arrive at Hospitals

Table C.2: Changes in Hospital-Level (Raw) Mortality Rates

Control Hospitals				Treated Hospitals			
Hospital Treatment: 10 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
0.051	0.052	0.001	0.1342	0.041	0.039	-0.002	-0.4939
(0.039)	(0.048)			(0.031)	(0.027)		
[69]	[69]			[98]	[98]		
Hospital Treatment: 15 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
0.050	0.052	0.002	0.2319	0.043	0.040	-0.002	-0.5331
(0.041)	(0.050)			(0.031)	(0.029)		
[55]	[55]			[112]	[112]		

Notes: This table shows the mean changes in raw mortality rates at the hospital level. Using raw mortality rates, we do not see any changes pre and post HST for both, control and treated hospitals. Standard deviation in parentheses. Number of hospitals in brackets.

Table C.3: Changes in Hospital-Level (IV adjusted) Mortality Rates

Control Hospitals				Treated Hospitals			
Hospital Treatment: 10 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
-0.030	-0.014	0.016	0.4523	0.002	-0.059	**-0.062**	-2.346
(0.192)	(0.229)	t: 0.4523		(0.203)	(0.164)	t:-2.346	
[69]	[69]			[98]	[98]		
Hospital Treatment: 15 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
-0.029	-0.004	0.025	0.5870	-0.002	-0.059	**-0.056**	-2.345
(0.212)	(0.237)			(0.192)	(0.168)		
[55]	[55]			[112]	[112]		

Notes: This table shows the mean changes in case-mix adjusted mortality rates at the hospital level (obtained using IV method). Adjusted mortality rates decrease post-HST for treated hospitals whereas there is no difference for the control hospitals. Standard deviation in parentheses. Number of hospitals in brackets.

Table C.4: Changes in Patient Level Mortality Rates by Destination

Control Hospitals				Treated Hospitals			
Hospital Treatment: 10 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
0.044	0.036	**-0.008**	-3.341	0.039	0.032	**-0.007**	-4.13
(0.205)	(0.187)			(0.193)	(0.176)		
[13,648]	[16,385]			[22,243]	[26,710]		
Hospital Treatment: 15 mile radius							
Pre-Train	Post-Train	Δ in Means	t-stat	Pre-Train	Post-Train	Δ in Means	t-stat
0.047	0.037	**-0.009**	-3.231	0.039	0.032	**-0.006**	-4.293
(0.211)	(0.189)			(0.193)	(0.177)		
[8,577]	[10,620]			[27,314]	[32,475]		

Notes: This table shows the mean changes in patient level (raw) mortality rates by destination (whether patients went to control or treated hospitals). Standard deviation in parentheses. Number of patients in brackets.

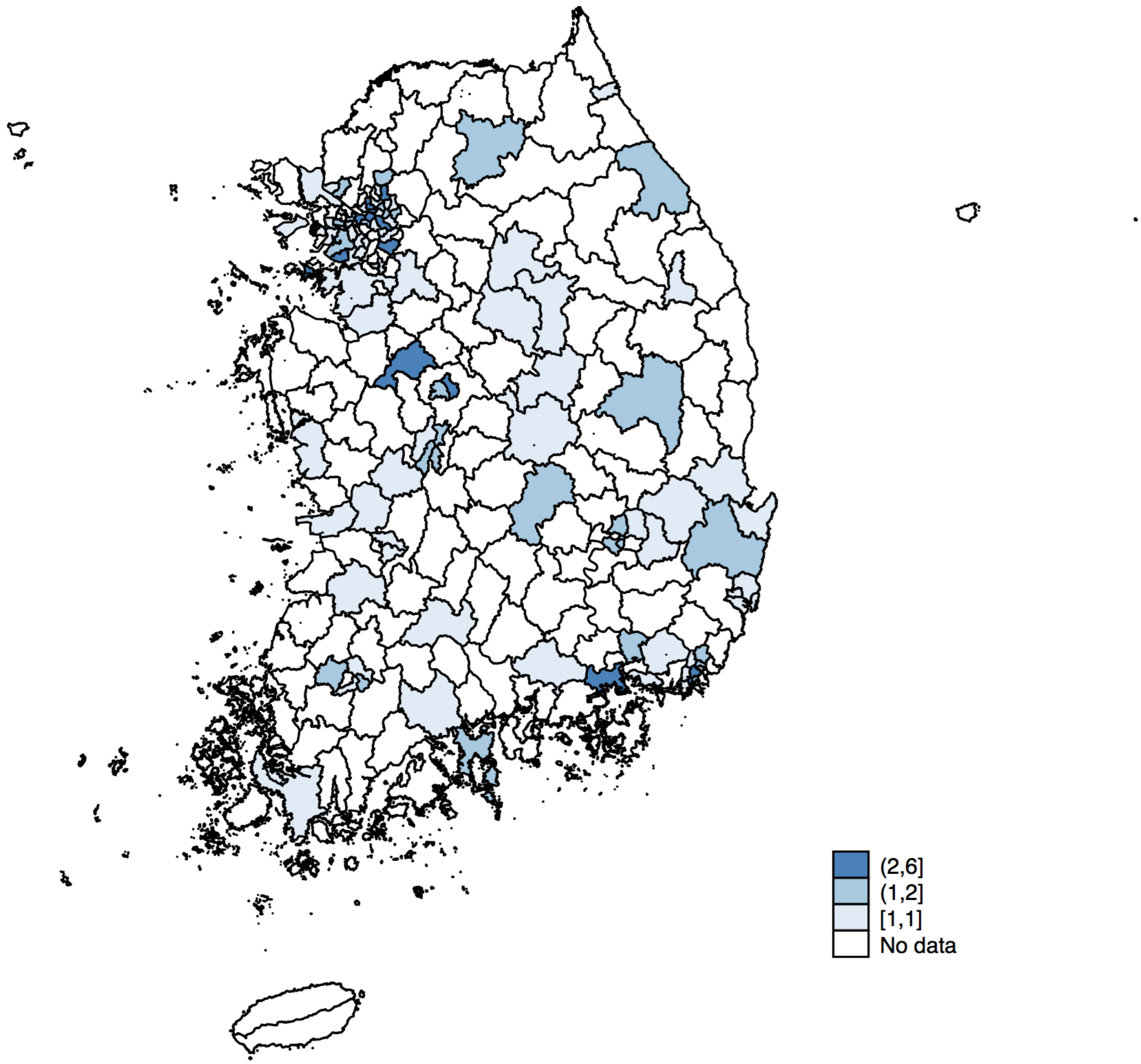
Table C.5: Patients's expected mortality rate at the hospital of his choice (Treated Hospitals)

	Pre-HST	Post-HST	Δ	t-stat	t-stat of diff in Δ
Patients who took the train	0.037	0.02	-0.017	1.7362*	1.191
to arrive at the hospital	(0.188)	(0.149)			
Number of patients	792	970			
Patients who did not take the train	0.039	0.032	-0.007	3.8864***	
to arrive at the hospital	(0.193)	(0.176)			
Number of patients	21,451	25,740			

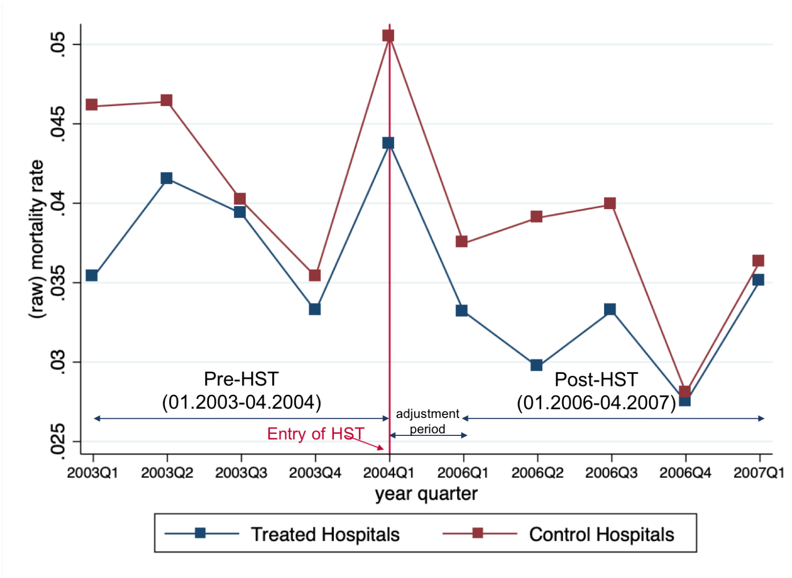
Notes: We define "patients who took the train" as patients who traveled more than 50 miles. Difference between the changes in means is not statistically significant ($t = -1.1914$). Standard deviation in parentheses.

Appendix E

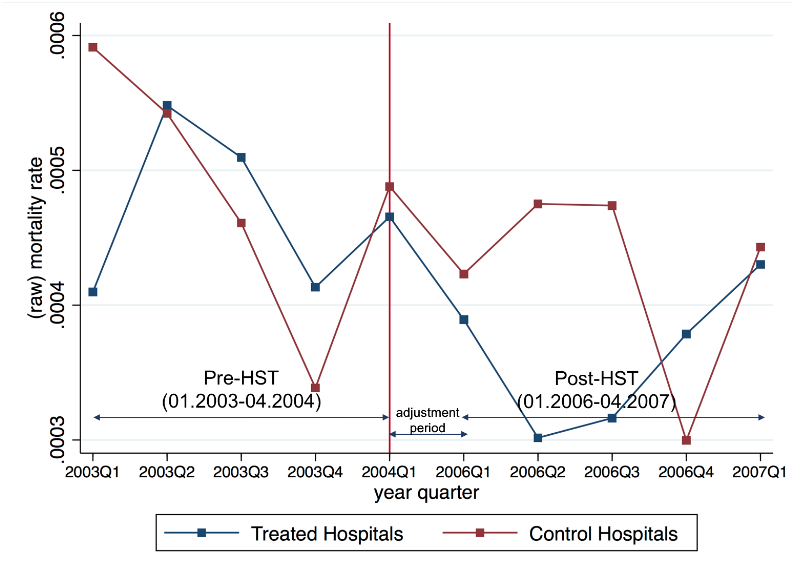
Figure 5: Distribution of hospitals



Notes: This figure displays number of hospitals by si-gun-gu in our final sample



(a) Hospital-level mortality rates



(b) weighted mortality rates

Figure 4: Trend of hospital-level mortality rates

Appendix F : Emergency admissions

Table C.6: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care (including emergency admissions)

	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Raw Mortality Rates	(1)	(2)	(3)	(4)	(4)
Post	-0.0007	-0.00067	0.0028	0.0029	0.0058
	(0.0032)	(0.0045)	(0.0061)	(0.0075)	(0.0095)
Treated×Post		-0.000064	-0.0060	-0.0054	-0.0088
		(0.0062)	(0.0070)	(0.0081)	(0.0099)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.6680	0.6680	0.6697	0.6693	0.6709

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table C.7: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care (including emergency admissions)

	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
OLS	(1)	(2)	(3)	(4)	(4)
Post	-0.007**	-0.003	0.0003	0.003	0.005
	0.003	0.004	0.006	0.007	0.008
Treated×Post		-0.009*	-0.012*	-0.014*	-0.016*
		0.005	0.006	0.007	0.009
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.5823	0.5888	0.5935	0.5967	0.5988

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table C.8: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care (including emergency admissions)

	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
IV	(1)	(2)	(3)	(4)	(4)
Post	-0.027	-0.025	0.026	0.029	0.027
	0.019	0.026	0.029	0.033	0.039
Treated×Post		-0.006	-0.091**	-0.084**	-0.072
		0.038	0.039	0.040	0.045
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.6052	0.6052	0.6181	0.6152	0.6115

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

Table C.9: Diff-in-Diff Estimates of the Impact of Competition on Quality of Clinical Care (including emergency admissions)

	Pre-Post	5-mile Treatment	10-mile Treatment	15-mile Treatment	20-mile Treatment
Raw Mortality Rates	(1)	(2)	(3)	(4)	(4)
Post	-0.00427	-0.00140	0.00283	0.00513	0.00587
	(0.00271)	(0.00345)	(0.00436)	(0.00536)	(0.00623)
Treated×Post		**-0.00819***	**-0.0128****	**-0.0144****	**-0.0136****
		(0.00464)	(0.00500)	(0.00574)	(0.00659)
Above70	0.0130	0.0138	0.0200	0.0165	0.0105
	(0.0536)	(0.0531)	(0.0516)	(0.0515)	(0.0526)
SurgeryRisk	0.0459	0.0558	0.0558	0.0505	0.0495
	(0.122)	(0.118)	(0.117)	(0.117)	(0.118)
DiseaseRisk	0.481*	0.492*	0.495*	0.519*	0.510*
	(0.280)	(0.275)	(0.271)	(0.267)	(0.270)
ComorbidityRisk	1.184***	1.198***	1.191***	1.204***	1.198***
	(0.415)	(0.411)	(0.402)	(0.400)	(0.404)
DisabledFrac	-2.047***	-2.285***	-2.459***	-2.399***	-2.208***
	(0.724)	(0.789)	(0.738)	(0.732)	(0.704)
DisabilitySeverity	2.504***	2.746***	2.929***	2.871***	2.663***
	(0.877)	(0.941)	(0.891)	(0.884)	(0.851)
LowIncomeFrac	-0.00920	-0.0160	-0.0185	-0.0193	-0.00678
	(0.138)	(0.134)	(0.134)	(0.134)	(0.132)
Hospital FE	YES	YES	YES	YES	YES
Number Hospitals	167	167	167	167	167
Control Hospitals		105	69	55	43
Treated Hospitals		62	98	112	124
Observations	334	334	334	334	334
R-squared	0.813	0.816	0.820	0.822	0.820

Notes: Models are estimated by OLS with standard errors (in parentheses under coefficients) robust to arbitrary heteroskedasticity. All regressions include constants. *** Significant at the 1 percent level; ** Significant at the 5 percent level;* Significant at the 10 percent level.

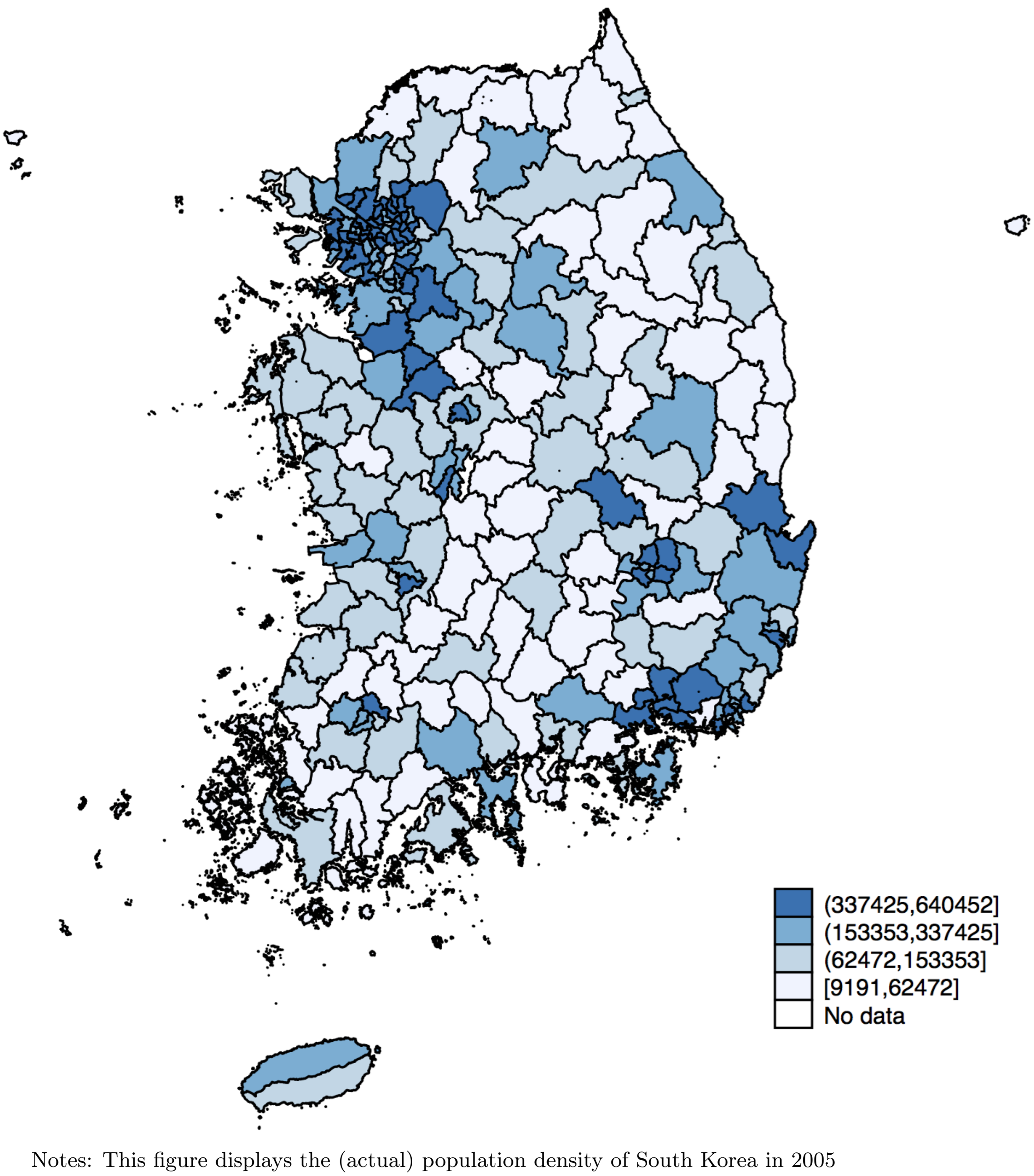
Appendix G

Table C.10: Descriptive Evidence of Changes in Travel Distance (Patients that appear in both periods)

	Control Patients			Treated Patients		
	Pre-HST	Post-HST		Pre-HST	Post-HST	
Distance Traveled	Mean	Mean	%Δ	Mean	Mean	Δ
	(st.dev)	(st.dev)	(t-stat)	(st.dev)	(st.dev)	Δ(t-stat)
Patients that appear in both periods (Patient Treatment: 15 miles)						
			Panel A. Distance Traveled (miles)			
	30.965	33.288		13.636	15.876	
	(42.155)	(45.306)	t:1.413	(35.687)	(38.381)	**t:2.045**
Nobs	1,333	1,530	p:0.1577	2,158	2,461	p:0.0409
			Panel B. Traveled to arrive at treated hospitals			
	0.483	0.485		0.066	0.084	
	(0.500)	(0.500)	t:0.092	(0.249)	(0.277)	**t:2.186**
Nobs	605	742	p:0.9264	2,081	2,354	p:0.0289
			Panel C. Traveled to arrive at control hospitals			
	0.074	0.086		0.260	0.318	
	(0.262)	(0.281)	t:0.866	(0.441)	(0.468)	t:0.850
Nobs	728	788	p:0.3865	77	107	p:0.3967

Notes: This table shows summmary statistics for patients who appear in both, pre- and post-HST periods. Patient treatment is defined as living within 15 miles of the HST station. Hospital treatment is defined as being located within 15 mile of the HST station. Panel A shows changes in travel distance (miles) in each period for treated and control patients. Panel B shows changes in proportion of patients that arrived at the treated hospitals via traveling. Panel C shows the changes in proportion of patients that arrived at the control hospitals via traveling.

Figure 6: Population Density



Notes: This figure displays the (actual) population density of South Korea in 2005