The Effects of Road Access on Health Outcomes in Rural India

Yeshwant Chillakuru

Department of Economics, The George Washington University

2115 G St. NW Suite 340, Washington, DC 20052

5 April 2017

Abstract

This is my abstract.

Introduction

Improving transportation infrastructure is vital to India's long-term economic development. Paved roads are needed to facilitate trade between communities and improve access to markets, schools, and health facilities, especially in rural India, where most of the country's population lives. In 2000, India began the rural road expansion program known as Pradhan Mantri Gram Sadak Yojama (PGSY), which aims to connect 170,000 villages with paved all-weather roads by 2020. As of 2010, 72,000 habitations have been connected by paved roads (World Bank, 2010). However, few studies have examined the efficacy of road improvement on health outcomes in India, and what literature exists is often plagued by weak methodology. The Ministry of Rural Development in India assessed PMGSY in 2004 and claimed an overwhelmingly positive impact of roads on health outcomes, stating improved travel times, higher levels of patient satisfaction, and increased hospital visits in several districts. However, the assessment was limited by a small sample size, no control groups, and failure to control for confounding variables, such as increased disease communication due to improved travel conditions causing more hospital visits.

The effect of road expansion on health outcomes has recently been researched, but the results, thus far, offer limited understanding of roads' impact. Bell and van Dillen (2014) examined health outcomes after road expansion under PMGSY in upland Orissa, India, with a cross-sectinoal analysis of 2010 survey data. The theoretical framework underlying their analysis stated that roads would not alter the disease environment (e.g. sanitation, strains of local bacteria/viruses) of the villages in the short-term but would improve treatment times and outcomes as patients travel to health centers and as ambulances reach patients more easily. They found that roads had no effect on morbidity rates, mortality rates, or treatment times. Bell and van Dillen's findings refute the positive impact touted by the Ministry of Rural Development, providing evidence that Indian rural health services are so inadequate that improved access may not improve health outcomes of villagers. However, their analysis was limited to the Orissa uplands and their data lacked a more robust, longitudinal analysis.

The effect of road access can lead to unexpected negative health outcomes. Contrary to intuition, Djemai (2011) has shown that HIV morbidity increases with road access in six sub-Saharan African countries. Djemai proposes a framework of two opposing forces:

1. Increased road access allows individuals to access markets to purchase condoms and receive HIV-related education, thus reducing the risk of HIV infection.
2. Increased road access allows for more communicability through both high-risk mobile groups and individuals having more sexual partners, thus increasing risk of HIV infection.

Djemai found that both effects independently occur and are significant, but the latter overpowers the former. By applying Djemai's framework with the work done by Bell and van Dillen, rural roads expansion can effect health outcomes in India through the following mechanisms:

1. By improving access to treatment facilities, roads will improve overall health outcomes. However, if rural health care facilities lack effective treatments or contain non-sterile pathogenic environments, health outcomes may worsen.
2. By accelerating economic development and thus personal hygiene and sanitation, roads will improve overall health outcomes.
3. By increasing the likelihood of communication from high-risk mobility groups, roads will worsen health outcomes related to communicable disease, but will have no effect for non-communicable disease.

This paper uses panel data from the India Human Development Survey (IHDS) to demonstrate the relationship between road expansion and health outcomes, as defined by incidence of a disease in the past 60 days, specifically the following 13 diseases: cataracts, tuberculosis, high blood pressure, heart disease, diabetes, leprosy, cancer, asthma, polio, paralysis, epilepsy, mental illness, STD/AIDS, and other major morbidity diseases. Using a linear regression model revealed that road expansion, whether by decreasing distance to nearest paved road or building a paved road in the village, significant had no impact on the incidence of any disease in rural India.

Data

IHDS is a nationally representative survey of 41,554 households in 1,503 villages and 971 urban neighborhoods in India. IHDS-I was conducted from 2004-2005, and IHDS-II was conducted from 2011-2012. The data of interest for this study was divided into several levels. Health outcomes, such as incidence of diabetes or polio, and individual-level controls, such as education and alcohol consumption, were stored in the individual-level dataset. Household-level controls, such as income and access to electricity, were stored household-level dataset. Road data, such as distance to road, and village-level controls were stored in the village-level dataset. The datasets were all encoded to be mapped together. Each village had an I.D. number. Each household had an I.D. number that was mapped to the village I.D. number. Each individual had an I.D. number that was then mapped to the household and village I.D. numbers. Therefore, data was aggregated to the village level, which had the road data, and as a result, an individual village is the unit of observation. When aggregating data from the specific to the broad (i.e. individual/household to village), the mean was used if data was an interval or ratio variable. If the data was a categorical variable, dummy variables were created for each category and the means of the dummy variables were aggregated to the village level. After data cleaning, the 2005 data (IHDS-I) contained 1457 observations (i.e. unique villages), and the 2012 dataset (IHDS-II) contained 1345 observations. 1314 villages were the same in both 2005 and 2012.

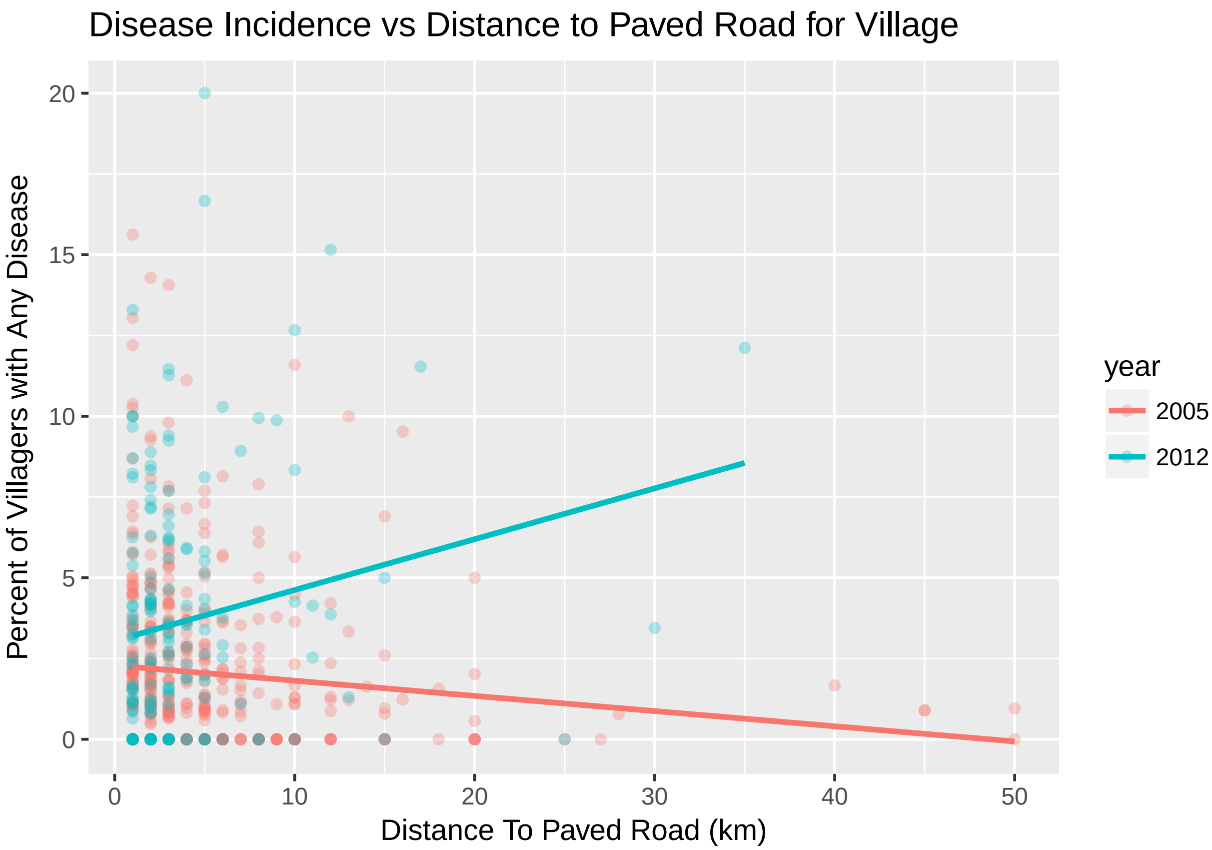
From the survey data, the impact of PGMSY on rural India can be seen. In 2005, 66.71% of villages in the IHDS survey had paved roads, but in 2012, 86.9%had paved roads. Additionally, for villages without paved roads, the closest paved road was 4.93 km away on average in 2005, but only 3.86 km away on average in 2012. Disease morbidity increased from 2005 to 2012 for all diseases in the dataset, except STD/AIDS, which remained the same.

Table 1: Average percent of villagers in each village with disease

Year	mbCataract	mbTuberculosis	mbHighBP	mbHeartDisease	mbDiabetes	mbLeprosy
2005	0.5097277	0.2750774	1.140681	0.3784203	0.5758473	0.0507374
2012	0.9399871	0.2861963	2.503536	0.6506209	1.1971111	0.0581898

Year	mbCancer	mbAsthma	mbPolio	mbParalysis	mbEpilepsy	mbMentalIllness	mbSTDorAIDS
2005	0.0595485	0.6151725	0.1160445	0.1542915	0.1164206	0.1289937	0.0549214
2012	0.0574084	1.0072603	0.1011937	0.3793866	0.2306951	0.3284036	0.0341595

A plausible explanation for this may be improved diagnosis from medical campaigns. A simple analysis examining “Percentage of Village with Any Disease vs Distance to Road” reveals a negative relationship between distanceToPavedRoad and mbAnyDisease in 2005 but a positive relationship in 2012. The 2005 correlation opposes the prevailing notion that roads access would reduce disease incidence, but may be attributed to increased diagnosis.



Model

Two OLS regressions models with time and village-fixed effects will be used to analyze the impact of road expansion on incidence of disease:

$$diseaseIncidence_{it} = \beta RoadPaved_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \delta_3 Medical_{it} + \gamma_{it} + \varepsilon_{it} \tag{2}$$

$$diseaseIncidence_{it} = \beta ln(DistanceToPavedRoad)_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \delta_3 Medical_{it} + \gamma_{it} + \varepsilon_{it} \tag{2}$$

$diseaseIncidence$ is the percent of villagers in village i during year t (t is 2005 or 2012) that fall into one of three disease categories:

- Sick with at least one of the 14 disease categories collected in the IHDS ($diseaseIncidence = mbAnyDisease$)
- Sick with at least one of three communicable diseases, STD/AIDS, Polio, or Tuberculosis ($diseaseIncidence = mbComDisease$)
- Sick with at least one of the remaining non-communicable diseases ($diseaseIncidence = mbNonComDisease$)

$Household_{it}$ represents village-level controls of household characteristics, such as income and electricity access, aggregated to the corresponding village of the household. $Village_{it}$ represents controls for characteristics of the village, such as number of primary health centers in village or the number of immunization campaigns for polio conducted in the village in the past year. $Medical_{it}$ represents village-level controls for where individuals in that village go for treatment, such as same village or different village, and for who individuals seek treatment, such private doctor or public doctor. γ_{it} represents village fixed effects, and ε_{it} is the error term.

The only difference between Models 1 and 2 is the primary indepedent variable of interest: $RoadPaved_{it}$ vs $ln(DistanceToPavedRoad)_{it}$. Equation (1) allows us to understand the impact of how introducing a paved road affects $diseaseIncidence_{it}$, while Equation (2) allows us to understand how becoming closer to a road for villages with no paved road can improve $diseaseIncidedence_{it}$. $roadPaved_{it}$ indicates if there is or is not a paved road in the village. $ln(DistanceToPavedRoad)_{it}$ indicates the natural log distance to the nearest paved for villages without a paved road. Model 2 is only applied to villages without a paved road to separate the effects of road being present or not versus improvement in distance to the road.

Simultaneous causality may present a problem in the current model. Individuals may move to areas where roads are built to improve economic opportunities. This may increase risk of communicability, similar to Djemai's study, where high-risk mobile groups increased HIV risk in areas closer to roads. To isolate the impact of increased communicability on disease incidence, I performed seperate regressions for communicable vs non-communicable diseaes. For example, if road access is bringing high-risk groups, who partake in risky sexual behavior, STD/HIV incidence would increase, while non-communicable diseases, like mental illness would not be effected.

Another issue with simultaneous causality is that villages with higher incomes and larger populations could be more likely to receive roads. Therefore, road placement may not be random. The PMGSY program is run at the national, and not state or district, level, and is less prone to influence of local village-level politics.

The time and village-fixed effects allowed to account for several potentials areas of omitted variable bias. Differences between villages in healthcare technology and access and sanitation infrastructure may arise non-randomly as villages near urban areas may benefit from spillover effects and earlier development. However, this is adjusted for by village-fixed effects. Similarly, healthcare technology and sanitation infrastructure may improve from 2005 to 2012 in villages, but again this are accounted for in time-fixed effects.

Results

Road Paved vs Road Unpaved

Distance to Nearest Road

Robustness Checks

- TO-DO:

- (maybe do it in the RESULTS section) Additionally, to control for road-induced migration, the data can be subsetted for people who have lived in the current village for over 7 years. This means that they would partake in both the 2005 and 2012 surveys.

Table 3:

	Dependent variable:		
	mbAnyDisease		
	(1)	(2)	(3)
roadPaved	1.486***	0.442*	0.073
	(0.254)	(0.257)	(0.176)
income		0.00000**	−0.00000***
		(0.00000)	(0.00000)
illiterate		3.108**	1.036
		(1.319)	(0.890)
smokeTobacco		0.826***	−0.105
		(0.153)	(0.107)
ownToilet		0.014**	0.006
		(0.006)	(0.004)
electricity		0.006	
		(0.005)	
ImmuniCamps			0.011
			(0.017)
PctPipedWater			−0.004
			(0.003)
primaryHealthCntr			−0.502**
			(0.242)
Treat.SameVil			0.494***
			(0.024)
Treat.OthrVil			0.542***
			(0.028)
Treat.OthrTwn			0.521***
			(0.023)
Treat.DstrctTwn			0.463***
			(0.030)
seenDoctor		−0.847	−0.780
		(1.287)	(0.869)
Observations	2,800	2,742	2,590
R^2	0.025	0.145	0.635
Adjusted R^2	−1.081	−0.862	0.147
F Statistic	34.265*** (df = 1; 1311)	30.454*** (df = 7; 1259)	148.154*** (df = 13; 1109)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4:

	Dependent variable:		
	mbAnyDisease		
	(1)	(2)	(3)
lnDistToPvdRd	−0.661	0.0003	0.249
	(0.469)	(0.463)	(0.328)
income		−0.00001	−0.00000
		(0.00001)	(0.00001)
illiterate		1.160	2.194
		(3.892)	(2.547)
smokeTobacco		2.307^{***}	0.381
		(0.495)	(0.365)
ownToilet		0.018	−0.003
		(0.017)	(0.012)
electricity		0.0002	
		(0.013)	
ImmuniCamps			−0.003
			(0.068)
PctPipedWater			0.022^{*}
			(0.012)
primaryHealthCntr			−0.971
			(1.387)
Treat.SameVil			0.530^{***}
			(0.082)
Treat.OthrVil			0.602^{***}
			(0.075)
Treat.OthrTwn			0.484^{***}
			(0.080)
Treat.DstrctTwn			0.377^{***}
			(0.091)
seenDoctor		0.456	−1.412
		(3.313)	(2.156)
Observations	658	642	602
R^2	0.016	0.290	0.752
Adjusted R^2	−4.174	−3.026	−0.583
F Statistic	1.985 (df = 1; 125)	6.601^{***} (df = 7; 113)	21.970^{***} (df = 13; 94)
Note:			*p<0.1; **p<0.05; ***p<0.01