The Effects of Road Access on Health Outcomes in Rural India

Yeshwant Chillakuru

March 8, 2017

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 ongoing rural road expansion program known as Pradhan Mantri Gram Sadak Yojama (PMGSY), which aims to connect 170,000 habitations villages with paved all-weather roads by 2020 with an expected total cost of US$54.6 billion. As of 2010, 72,000 habitations have been connected by paved roads (World Bank, 2010). However, literature examining the efficacy of road improvement on health outcomes in India is sparse, and what literature exists often is 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, patient satisfaction, and increased hospital visits in several districts. However, the assessment was limited by small sample size and no controls.

The effect of road expansion can lead to unexpected health outcomes. Contrary to intuition, Djemai has shown that HIV morbidity increases as road access increases (2011). Djemai proposes a framework of opposing forces with a negative health effect due to increased risky sexual behavior by high HIV-risk mobile groups overpowering a positive health effect of increased access to HIV education due to roads (2011). This framework can be extended to rural India. Roads may have a positive health effect by increasing access to health centers and proper treatment for diseases like polio, STDs, or tuberculosis. However, roads may also increase likelihood of communication due to increased travel by high risk morbidity groups or exposure to pathogens in non-sterile hospital settings. Additionally, if rural hospitals lack effective treatments or sterile environments, increased road access may not affect health outcomes or may lead to worse outcomes. Morbidity rates, which percent of population that has become ill in a fixed time period, is a common measure of assessing health outcomes in the developing world. Using the data from the India Human Development Survey, this study assesses the impact of increased access to paved roads on the incidence of major morbidity in rural India by examining morbidity rates for individual diseases and groups of communicable and non-communicable diseases.

Literature Review

The effect of road expansion on health outcomes has recently been studied, but the results so far offer a limited understanding. Bell and van Dillen examined health outcomes after road expansion under PMGSY in upland Orissa, India, using 2010 survey data (2014). The theoretical framework underlying their analysis stated that roads would not alter the disease environment of the villages in the short-term but would improve treatment times and outcomes as patients travel to health centers and ambulances reach patients more easily. They performed a cross-sectional analysis using both univariate and multivariate Tobit models of the 1,292 villagers across 30 villages, 456 of whom lived in one of 11 villages with a paved road. Bell and van Dillen found that roads had no effect on morbidity rates or treatment times. An additional analysis on mortality rates concluded the same null effect of roads. While their analysis was limited to the region of Orissa and their dataset did not allow for a more robust, longitudinal analysis, Bell and van Dillen offered a theoretical framework to approach the lack of improvement of health outcomes. As expected, roads did not change the disease environment in villages, but they failed to improve treatment times and mortality rates, despite survey responses by villages stating they were more likely to use and visit healthcare facilities for treatment. Therefore, rural health services in India may simply be inadequate (2014).

While limited literature exists on road access in rural India, Djemai has performed a robust analysis of roads in Africa. Djemai examined the effect of paved road access on risk of HIV infection in Cameroon, Ethiopia, Ghana, Kenya, Malawi, and Zimbabwe (2011). Djemai found an inverse relationship between distance to paved road and likelihood of HIV infection. This unexpected result can be explained with Djemai's framework of analysis, which suggested two opposing forces:

1. Increased road access allows individuals to access markets to purchase condoms and also learn about HIV risks, thus reducing the risk of HIV infection. Djemai demonstrated this effect to be true.
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 also showed this to be true.

Djemai's overall result showed that the latter effect to overpower the former. This theoretical framework can be transferred to the analysis of the effect of roads in rural India, as explained earlier. Djemai's analysis controlled for a wide variety of factors and also implemented an instrumental variables approach using terrain ruggedness and slope to improve robustness (2011). While the IHDS data lacks latitude and longitude data, preventing a similar IV approach, the IHDS data allows me to both control for individual and community level characteristics and investigate potential causal mechanisms behind the relationship. Djemai's robust analysis provides an ideal framework, upon which this study can build.

Data Description

The data used was collected from the India Human Development Survey (IHDS: http://ihds.info), 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 1501 observations (i.e. unique villages), and the 2012 dataset (IHDS-II) contained 1345 observations. The same indexing was used in both 2005 and 2012, and 1343 villages appeared in both IHDS-I and IHDS-II. Summary statistics are provided below:

2005

```
##  distanceToPavedRoad   roadPaved
##  Min.   : 0.0000     Min.   :0.000
##  1st Qu.: 0.0000     1st Qu.:1.000
```

```
##  Median : 0.0000     Median :1.000
##  Mean   : 0.5131     Mean   :0.869
##  3rd Qu.: 0.0000     3rd Qu.:1.000
##  Max.   :35.0000     Max.   :1.000
##  NA's   :43          NA's   :2
##    mbCataract      mbTuberculosis       mbHighBP       mbHeartDisease
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.00000   Median :0.000000   Median :0.00000   Median :0.000000
##  Mean   :0.00860   Mean   :0.004073   Mean   :0.01215   Mean   :0.004438
##  3rd Qu.:0.01053   3rd Qu.:0.000000   3rd Qu.:0.01504   3rd Qu.:0.006211
##  Max.   :0.18182   Max.   :0.119048   Max.   :1.00000   Max.   :0.111111
##    mbDiabetes          mbLeprosy            mbCancer
##  Min.   :0.000000   Min.   :0.0000000   Min.   :0.0000000
##  1st Qu.:0.000000   1st Qu.:0.0000000   1st Qu.:0.0000000
##  Median :0.000000   Median :0.0000000   Median :0.0000000
##  Mean   :0.006537   Mean   :0.0006073   Mean   :0.0007687
##  3rd Qu.:0.007194   3rd Qu.:0.0000000   3rd Qu.:0.0000000
##  Max.   :0.250000   Max.   :0.0769231   Max.   :0.0952381
##     mbAsthma           mbPolio         mbParalysis
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.000000   Median :0.00000   Median :0.000000
##  Mean   :0.006618   Mean   :0.00129   Mean   :0.001715
##  3rd Qu.:0.010101   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :0.131579   Max.   :0.04286   Max.   :0.075949
##    mbEpilepsy       mbMentalIllness     mbSTDorAIDS
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.0000000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000000
##  Median :0.000000   Median :0.000000   Median :0.0000000
##  Mean   :0.001282   Mean   :0.001613   Mean   :0.0007416
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.0000000
```

```
## Max.   :0.100000   Max.   :0.222222   Max.   :0.1052632
## mbOtherLongTerm   mbDaysIncapacitated
## Min.   :0.00000   Min.   : 0.0000
## 1st Qu.:0.00000   1st Qu.: 0.2857
## Median :0.01333   Median : 1.6923
## Mean   :0.02273   Mean   : 3.5212
## 3rd Qu.:0.03488   3rd Qu.: 4.7473
## Max.   :0.20000   Max.   :61.0870
```

2012

```
## distanceToPavedRoad   roadPaved
## Min.   :-3.000      Min.   :0.0000
## 1st Qu.: 0.000      1st Qu.:0.0000
## Median : 0.000      Median :1.0000
## Mean   : 1.559      Mean   :0.6476
## 3rd Qu.: 2.000      3rd Qu.:1.0000
## Max.   :50.000      Max.   :1.0000
```

```
##   mbCataract      mbTuberculosis       mbHighBP       mbHeartDisease
## Min.   :0.00000   Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
## 1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
## Median :0.00000   Median :0.000000   Median :0.02985   Median :0.00000
## Mean   :0.02209   Mean   :0.006872   Mean   :0.05048   Mean   :0.01332
## 3rd Qu.:0.03247   3rd Qu.:0.009950   3rd Qu.:0.07292   3rd Qu.:0.01942
## Max.   :0.35484   Max.   :0.173913   Max.   :0.58065   Max.   :0.23077
##   mbDiabetes        mbLeprosy           mbCancer           mbAsthma
## Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Min.   :0.00000
## 1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.00000
## Median :0.00000   Median :0.000000   Median :0.000000   Median :0.01020
## Mean   :0.02412   Mean   :0.001184   Mean   :0.001256   Mean   :0.02042
## 3rd Qu.:0.03030   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.03200
## Max.   :0.51613   Max.   :0.125000   Max.   :0.125000   Max.   :0.50000
```

```
##     mbPolio          mbParalysis         mbEpilepsy
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.000000   Median :0.000000   Median :0.000000
##  Mean   :0.002048   Mean   :0.007779   Mean   :0.004689
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :0.084507   Max.   :0.406780   Max.   :0.250000
##  mbMentalIllness     mbSTDorAIDS         mbOtherLongTerm
##  Min.   :0.000000   Min.   :0.0000000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.0000000   1st Qu.:0.02247
##  Median :0.000000   Median :0.0000000   Median :0.06250
##  Mean   :0.006675   Mean   :0.0007053   Mean   :0.08091
##  3rd Qu.:0.000000   3rd Qu.:0.0000000   3rd Qu.:0.11538
##  Max.   :0.285714   Max.   :0.1111111   Max.   :0.64000
##  mbDaysIncapacitated
##  Min.   : 0.0000
##  1st Qu.: 0.9375
##  Median : 2.9118
##  Mean   : 4.4673
##  3rd Qu.: 6.3542
##  Max.   :75.0000
```

Distance to Paved Road (`distanceToPavedRoad`) is in kilometers and morbidity (`mb...`) is a probability value from 0 to 1, determined by taking the mean of disease prevalence (1 = had the disease in the past year and 0 = did not have the disease in the past year) for all individuals surveyed in a village. Only 43 out of 1345 were missing from the `distanceToPavedRoad` in 2012, and no observations were missing from the health data.

Description of Variables

Most of this information will be placed in appendix

The 76 variables in the dataset are listed below:

[1] "id" "state"

[3] "district" "village"

[5] "year" "roadPaved"

[7] "distanceToPavedRoad" "roadMonsoonUsability"

[9] "yearsWithPavedRoad" "ImmunizationCampaignsNumber" [11] "distanceToNearestTown" "distanceToNearestDistrictHQ" [13] "area" "populationCategories"

[15] "householdNumbersOf" "mbCataract"

[17] "mbTuberculosis" "mbHighBP"

[19] "mbHeartDisease" "mbDiabetes"

[21] "mbLeprosy" "mbCancer"

[23] "mbAsthma" "mbPolio"

[25] "mbParalysis" "mbEpilepsy"

[27] "mbMentalIllness" "mbSTDorAIDS"

[29] "mbOtherLongTerm" "mbDaysIncapacitated"

[31] "mbTreatmentRecieved" "mbTreatmentWho1"

[33] "mbTreatmentWhere1" "mbTreatmentWho2"

[35] "mbTreatmentWhere2" "mbDaysHospitalized"

[37] "mbCostOfTreatment" "mbCostIncludesMedication"

[39] "mbCostOfMeds" "mbTravelExpenses"

[41] "smokeTobacco" "chewTobacco"

[43] "drinkAlcohol" "attendedSchool"

[45] "gradeCompleted" "readingScore"

[47] "mathScore" "writingScore"

[49] "urban" "income"

[51] "caste" "yearsLivingHere"

[53] "ownToilet" "electricity"

[55] "electricityHours" "drinkingWaterSource"

[57] "healthSubCenter" "primaryHealthCenter"

[59] "communityHealthCenter" "govtMaternityCenter"

[61] "pvtClinicTrainedDoc" "pvtClinicUntrainedDoc"

[63] "pvtHospital" "pvtPharm"

[65] "otherGovtMedFacility" "distHealthSubCenter"

[67] "distPrimaryHealthCenter" "distCommunityHealthCenter"

[69] "distDistrictHospital" "distGovtMaternityCenter"

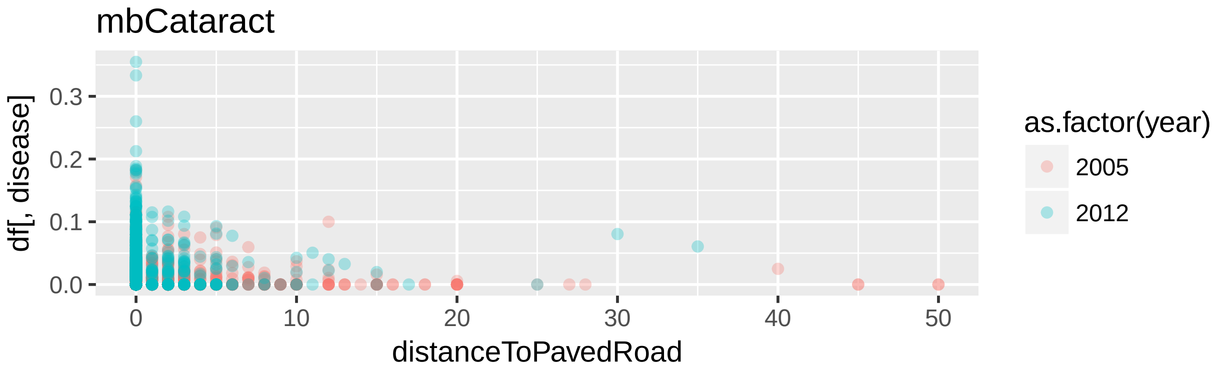
[71] "distPvtClinicTrainedDoc" "distPvtClinicUntrainedDoc"

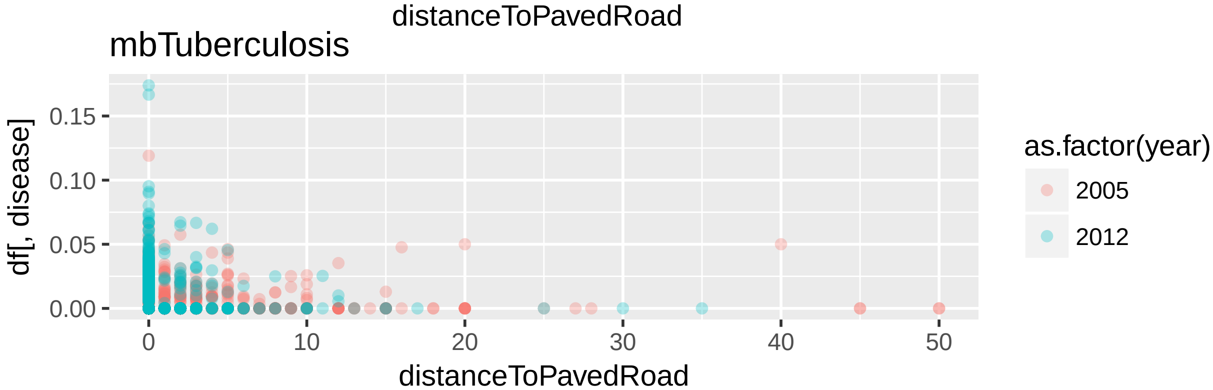
[73] "distPvtHospital" "distPvtPharm"

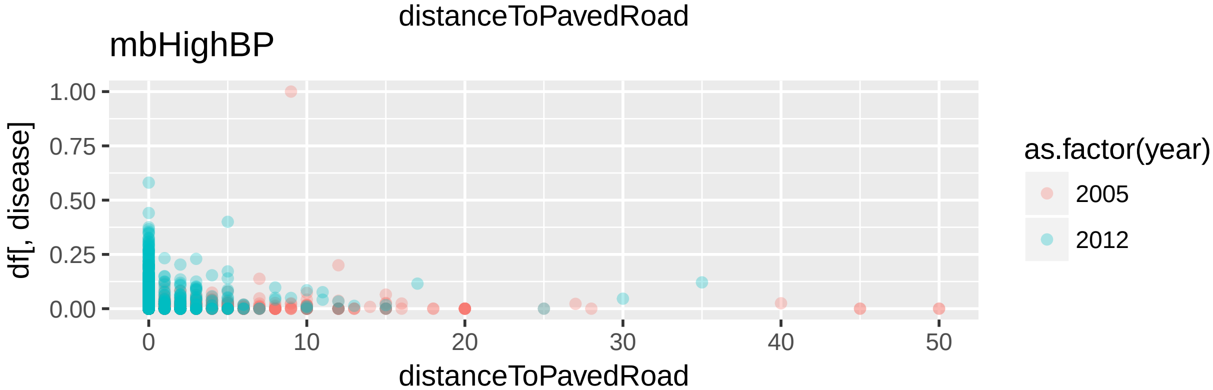
[75] "distOtherGovtMedFacility"

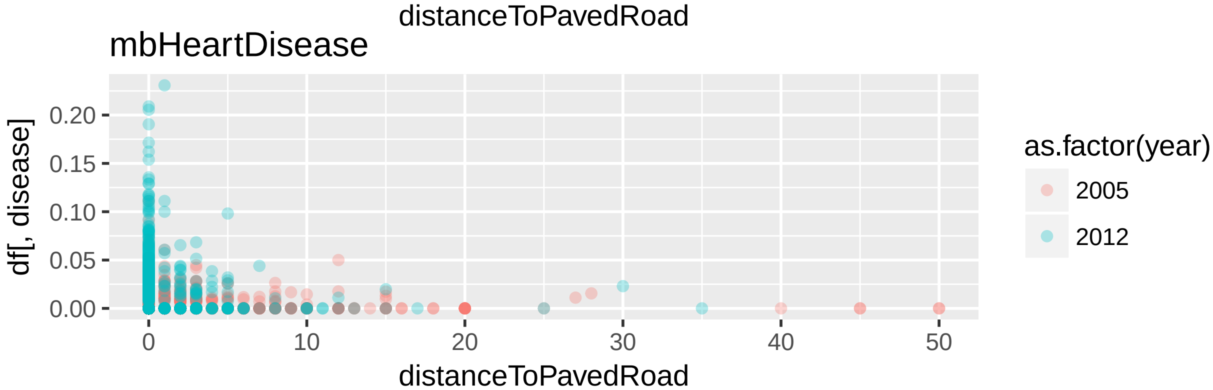
Most variables names have been renamed to be self-explanatory. Important notes regarding variables are: * Two additionally variable needs to be constructed: * `mbProbCommDisease` which is the likelihood of having a communicable disease (STD/AIDS, polio, tuberculosis, leprosy) * `mbProbNonCommDisease which is the likelihood of having non-communicable disease (other mb... that do not fall into mbProbCommDisease) *id= unique to each village each year.`state`and`district`are id values for state and district village is located in. *`cost...`in beginning of name indicates the cost in Rupees *`dist...`in beginning of variable name indicates distance in kilometers *`mb...`in beginning of variable name indicates major morbidity disease *`pvt...`and`govt...`indicates private and public sector facility (e.g.`pvtPharm`indicates if the village has a private-sector pharmacy and distPvtPharm indicates the distance to the nearest private-sector pharmacy if the village did not have one in the village). * Variables 41 (`smokeTobacco`) to 48 (`writingScore`) are Individual-level controls * Variables 49 (`caste`) to 66 (`drinkingWaterSource`) are Household-level controls * Caste will be turned into dummy variables for each caste type * Variables 67 (`healthSubCenter`) to 76 (`distOtherGovtMedFacility`) are Village-level controls * Dummy variables for`mbTreatmentWhere1`,mbTr and caste have yet to be created, but will be soon. *`area`,‘populationCategories, and `householdNumbersOf` are missing in the 2012 Dataset. Upon inquiry with IHDS, I discovered they are still filling in these values from the 2011 Indian Census data. If this data is not complete within the next 2 weeks, these variables will be dropped from the dataset. * R code for data cleaning and analysis can be accessed on Github: https://github.com/yesh747/RuralRoadsIndia

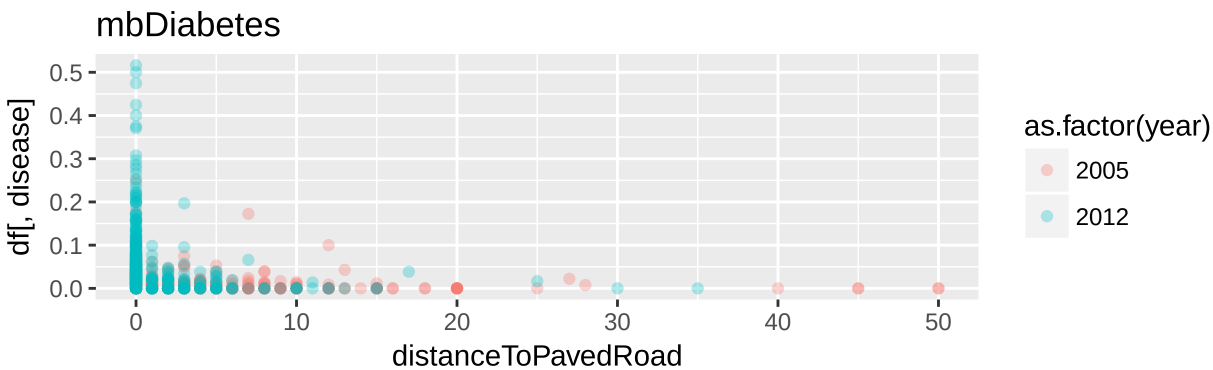
Graphs: Disease incidence vs Road Distance in 2005 and 2012

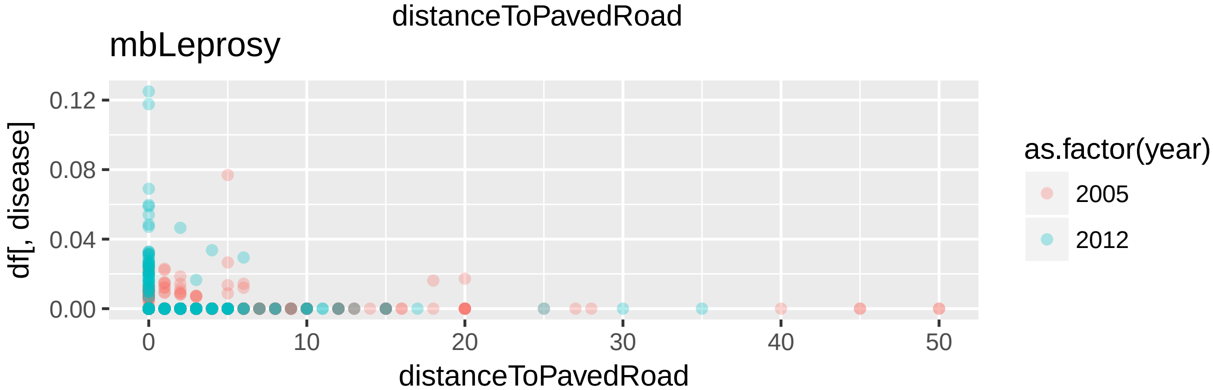


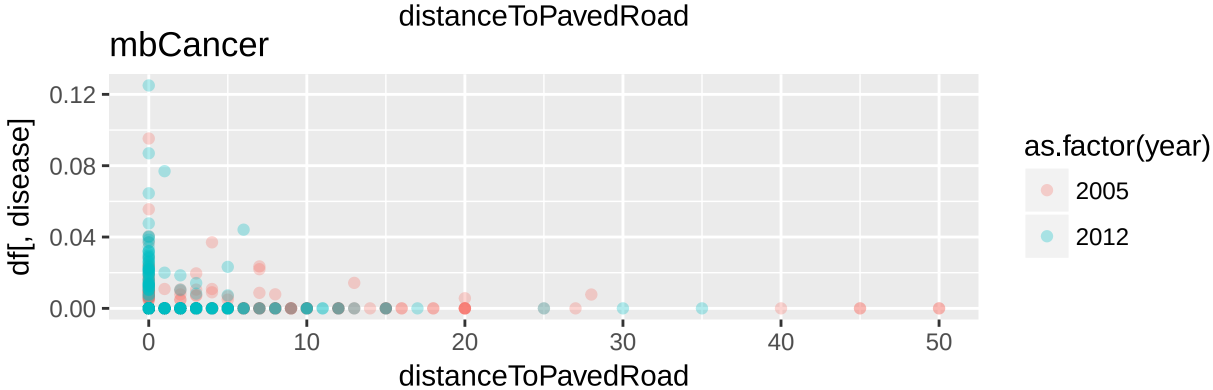


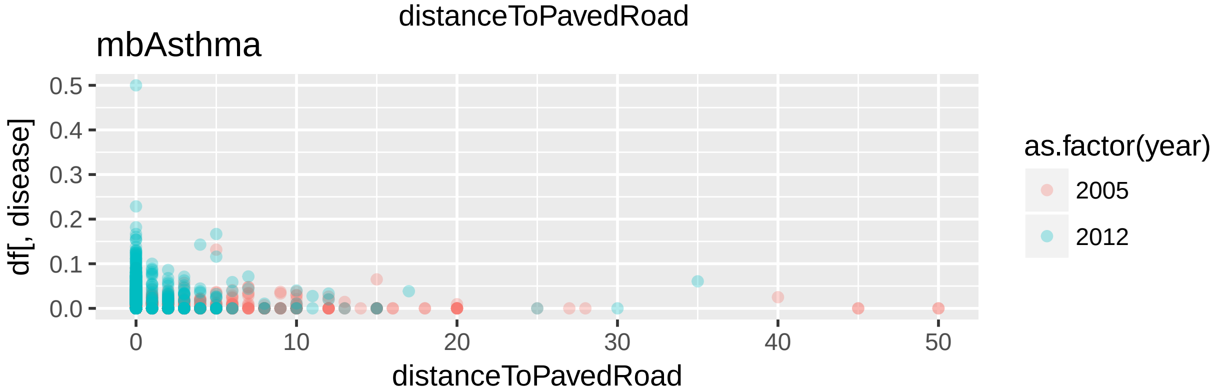


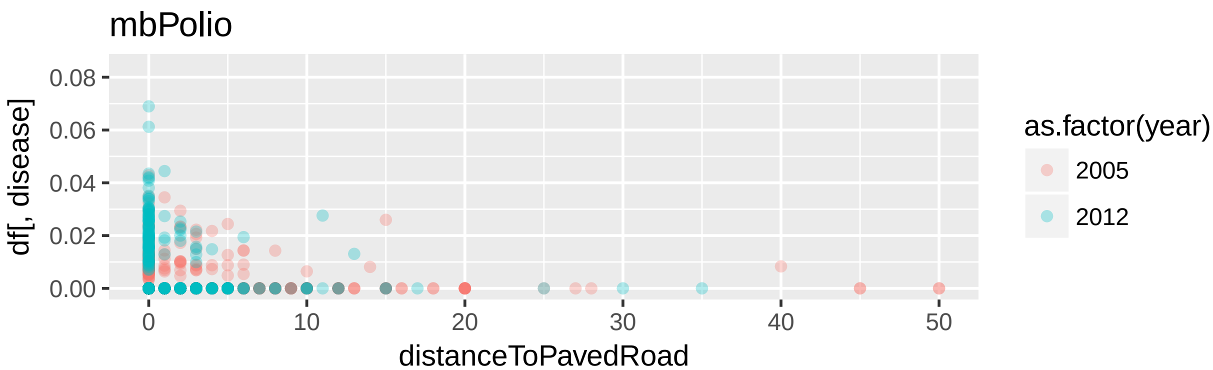


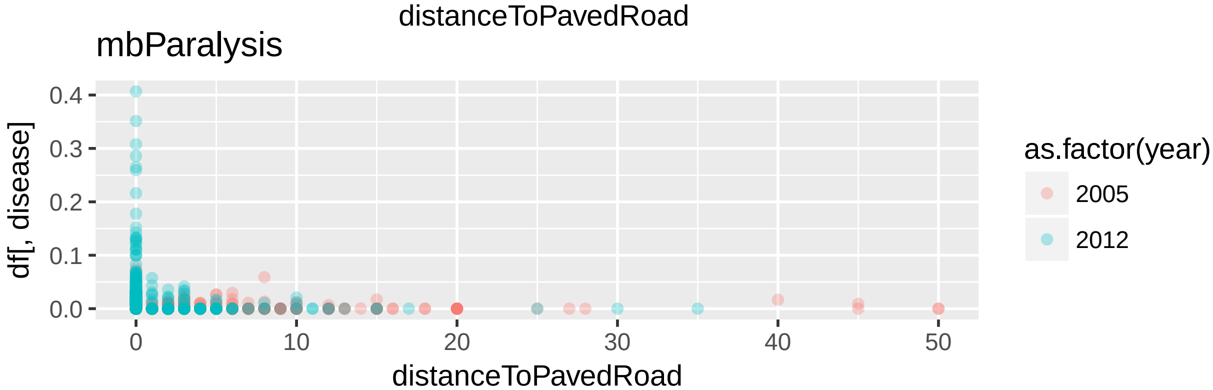

mbDiabetes
df[, disease]
distanceToPavedRoad
as.factor(year)
2005
2012

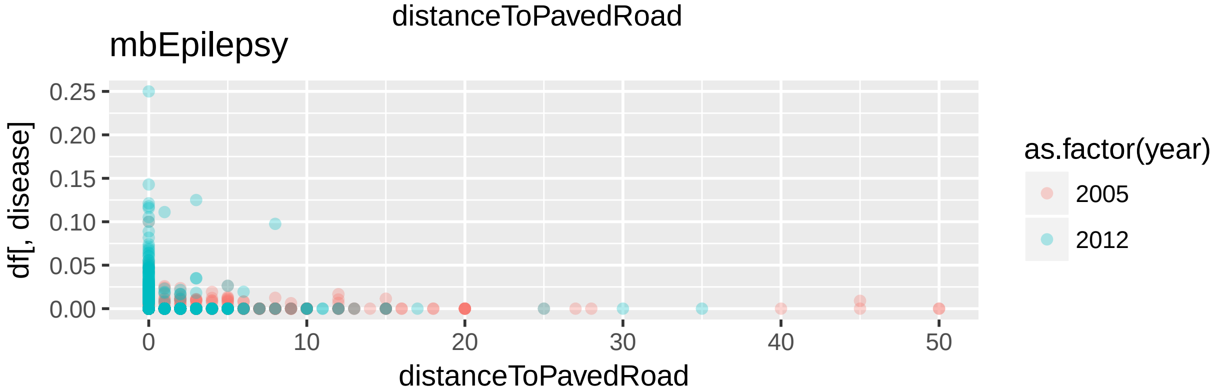

mbLeprosy
df[, disease]
distanceToPavedRoad
as.factor(year)
2005
2012

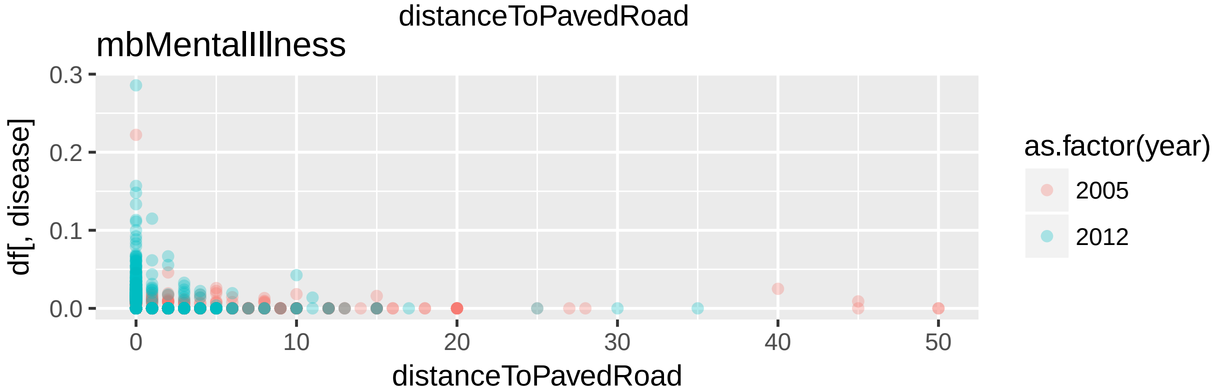

mbCancer
df[, disease]
distanceToPavedRoad
as.factor(year)
2005
2012

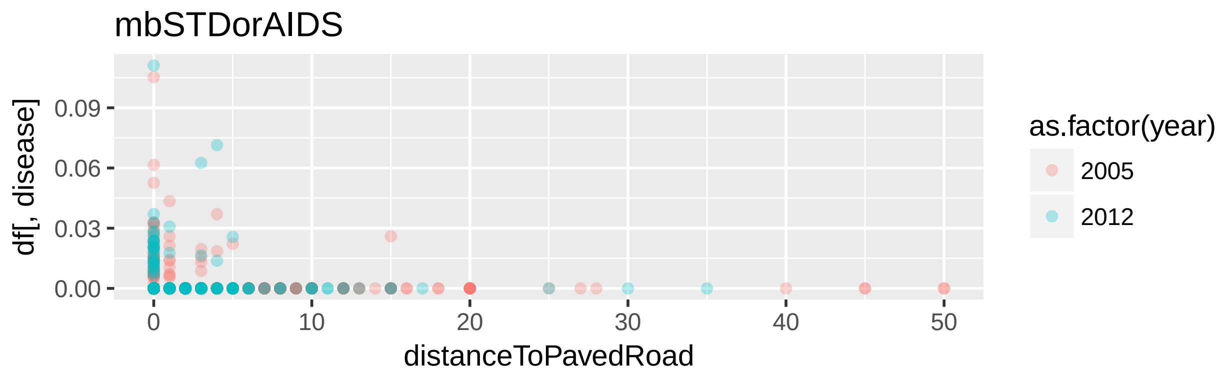

mbAsthma
df[, disease]
distanceToPavedRoad
as.factor(year)
2005
2012


mbPolio
df[, disease]
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
distanceToPavedRoad
as.factor(year)
2005
2012


mbParalysis
df[, disease]
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
distanceToPavedRoad
as.factor(year)
2005
2012


mbEpilepsy
df[, disease]
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
distanceToPavedRoad
as.factor(year)
2005
2012


mbMentalIllness
df[, disease]
0.3
0.2
0.1
0.0
0
10
20
30
40
50
distanceToPavedRoad
as.factor(year)
2005
2012



The above graphs of all the diseases in the dataset were constructed as part of exploratory data analysis. An inverse relationship between road distance and disease prevalence appears to exist. However, this could be due to a number of factors that have yet to be accounted for. One can also see that there is an overall decrease in the distance to a paved road as we move from 2005 to 2012, an effect confirming the presence of PMGSY's road expansion program in the sample of surveyed villages.

Model Description

The following OLS linear regression model will be used:

$$diseaseIncidence_{it} = \beta distroad_{it} + \delta_1 Ind_{it} + \delta_2 household_{it} + \delta_3 Village_{it} + \varepsilon_{it}$$

A panel regression will be performed where i is the village and t is the year (i.e. 2005 or 2012). diseaseIncidenceit is the disease of interest. This regression will be conducted for all the separate diseases in the dataset (shown in the above graphs) and for the constructed variables of `mbProbCommDisease` and `mbProbNonCommDisease.` Indit are the individual-level controls, such as the percent of individuals smoking tobacco in the village. householdit are the household-level controls, such as the mean income of the household. Villageit are the village-level controls, such as the number of primary Health Centers in the village. Potentially, the distRoad variable will be transformed to be log(1 + distRoad). This transformation was done by Djemai (2011), to account for non-linear relationship of distance to a paved road on access to medical facilities via travelling (e.g. individuals will be indifferent to travel if the distance to paved road changes from 50km to 20km, but if the distance to paved road changes from 10km to 5km, then individuals will be more likely to travel).

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. Performing separate regressions on communicable vs non-communicable diseases would allow us to isolate the impact of increased communicability on disease incidence. 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. 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.

```
Another issue with simultaneous causality is that villages with higher incomes and larger populations c
```

Expected Results

Djemai found distance to paved road in Africa to be inversely correlated with communicable diseases, such as STD/AIDS, polio, tuberculosis, or leprosy, and to be positively correlated with non-communicable diseases (2011). However, Bell and Van Dillen found that roads in rural India had no effect on morbidity or mortality, potentially due to the insufficient healthcare services provided in rural India (2014). I quickly conducted univariate and multivariate fixed effects panel regressions of road distance on various disease types.

Univariate Regression Results

```
##
## Results
## ==============================================================================================
##                                                       Dependent variable:
##                                     ----------------------------------------------------------
##                                                         df[, disease]
##                                       mbCataract   mbTuberculosis    mbHighBP    mbHeartDisease
##                                          (1)            (2)            (3)            (4)
## ----------------------------------------------------------------------------------------------
## distanceToPavedRoad               -0.00050948*** -0.00009805   -0.00109148*** -0.00042748***
##                                    (0.00015822)   (0.00006944)   (0.00030783)   (0.00010657)
##
```

```
## Constant                        0.01546407***  0.00555846***  0.03151324***  0.00891772***
##                                 (0.00056803)   (0.00024930)   (0.00110514)   (0.00038261)
##
## ---------------------------------------------------------------------------------------------
## Observations                       2,759          2,759          2,759          2,759
## R2                                0.00374682     0.00072270     0.00453942     0.00580188
## Adjusted R2                       0.00338546     0.00036025     0.00417835     0.00544128
## Residual Std. Error (df = 2757)   0.02837892     0.01245484     0.05521282     0.01911541
## F Statistic (df = 1; 2757)      10.36882000***   1.99393200   12.57225000*** 16.08914000***
## =============================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
##
## Results
## ============================================================================================
##                                                 Dependent variable:
##                                 ------------------------------------------------------------
##                                                   df[, disease]
##                                  mbCataract   mbTuberculosis   mbHighBP    mbHeartDisease
##                                     (1)            (2)            (3)            (4)
## --------------------------------------------------------------------------------------------
## distanceToPavedRoad             -0.00090360***  -0.00001352    -0.00002892   -0.00045120***
##                                 (0.00020912)   (0.00003165)  (0.00003213)   (0.00013442)
##
## Constant                        0.01565483***  0.00089395***  0.00102470*** 0.01360503***
##                                 (0.00075077)   (0.00011362)  (0.00011535)   (0.00048257)
##
## --------------------------------------------------------------------------------------------
## Observations                       2,759          2,759          2,759          2,759
## R2                                0.00672630     0.00006620    0.00029376     0.00407037
## Adjusted R2                       0.00636603    -0.00029649   -0.00006885     0.00370913
## Residual Std. Error (df = 2757)   0.03750867     0.00567636    0.00576284     0.02410913
## F Statistic (df = 1; 2757)      18.66999000***   0.18253050    0.81013510    11.26787000***
```

```
## ===========================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
##
## Results
## =================================================================================================================
##                                                              Dependent variable:
##                                    ------------------------------------------------------------------------------
##                                                                 df[, disease]
##                                     mbCataract   mbTuberculosis   mbHighBP    mbHeartDisease  mbDiabete
##                                        (1)            (2)            (3)           (4)            (5)
## -----------------------------------------------------------------------------------------------------------------
## distanceToPavedRoad                 -0.00003888  -0.00026659**  -0.00013776** -0.00020439***  -0.0000097
##                                    (0.00003189)   (0.00011201)  (0.00006459)   (0.00007484)  (0.0000279
##
## Constant                           0.00164619*** 0.00492930***  0.00308653*** 0.00422511***  0.00070084
##                                    (0.00011450)   (0.00040213)  (0.00023189)   (0.00026869)  (0.0001001
##
## -----------------------------------------------------------------------------------------------------------------
## Observations                           2,759         2,759          2,759         2,759          2,759
## R2                                  0.00053869     0.00205036     0.00164708     0.00269773     0.0000440
## Adjusted R2                         0.00017617     0.00168839     0.00128496     0.00233600    -0.000318
## Residual Std. Error (df = 2757)     0.00572024     0.02009058     0.01158527     0.01342384     0.0050046
## F Statistic (df = 1; 2757)          1.48597600     5.66445800**   4.54848200**   7.45776300***  0.1215360
## =================================================================================================================
## Note:                                                                        *p<0.1; **p<0.05; ***p<0
```

Multivariate FE Panel Regression Results

```
##
## Results
## ===================================================================================
##                                               Dependent variable:
```

```
##                              ------------------------------------------------------
##                                                  df[, disease]
##                                mbCataract  mbTuberculosis   mbHighBP   mbHeartDisease
##                                   (1)            (2)           (3)           (4)
## ------------------------------------------------------------------------------------
## distanceToPavedRoad           0.00031279    0.00002307   -0.00006356    0.00009104
##                              (0.00024488)  (0.00011818)  (0.00040175)  (0.00016229)
##
## electricity                   0.00272689   -0.00032542    0.00215146   -0.00015827
##                              (0.00405657)  (0.00195779)  (0.00665527)  (0.00268851)
##
## income                       -0.00000000    0.00000001*  -0.00000003  -0.00000002**
##                              (0.00000002)  (0.00000001)  (0.00000003)  (0.00000001)
##
## distanceToNearestTown        -0.00011809   -0.00003171    0.00008676    0.00001285
##                              (0.00008397)  (0.00004053)  (0.00013776)  (0.00005565)
##
## writingScore                 -0.00156720   -0.00088576   -0.00313837  0.00344226***
##                              (0.00184285)  (0.00088940)  (0.00302340)  (0.00122135)
##
## smokeTobacco                  0.00028957   -0.00033446    0.00323832   -0.00083128
##                              (0.00122441)  (0.00059093)  (0.00200878)  (0.00081148)
##
## ------------------------------------------------------------------------------------
## Observations                     2,545         2,545         2,545         2,545
## R2                            0.00431901    0.00372068    0.00358756    0.01209612
## Adjusted R2                  -1.29439500   -1.29577400   -1.29608100   -1.27647400
## F Statistic (df = 6; 1104)  0.79814450    0.68716100    0.66248690   2.25293800**
## ====================================================================================
## Note:                                                  *p<0.1; **p<0.05; ***p<0.01
##
## Results
```

```
## ===================================================================================
##                                               Dependent variable:
##                               -----------------------------------------------------
##                                                  df[, disease]
##                               mbCataract   mbTuberculosis  mbHighBP  mbHeartDisease
##                                  (1)            (2)           (3)          (4)
## -----------------------------------------------------------------------------------
## distanceToPavedRoad            0.00016598     0.00000413   -0.00010639*  0.00002662
##                               (0.00025331)   (0.00005435)  (0.00005804)  (0.00022577)
##
## electricity                    0.00419994     0.00100382   -0.00001485   -0.00341122
##                               (0.00419619)   (0.00090038)  (0.00096147)  (0.00374008)
##
## income                        -0.00000003**   0.00000000   -0.00000000    0.00000000
##                               (0.00000002)   (0.00000000)  (0.00000000)  (0.00000001)
##
## distanceToNearestTown          0.00004901    -0.00001146   -0.00000425   -0.00010199
##                               (0.00008686)   (0.00001864)  (0.00001990)  (0.00007742)
##
## writingScore                  -0.00056834    -0.00068571*   0.00018537   -0.00139174
##                               (0.00190627)   (0.00040903)  (0.00043679)  (0.00169907)
##
## smokeTobacco                   0.00188069    -0.00013979   -0.00036371    0.00081339
##                               (0.00126655)   (0.00027176)  (0.00029020)  (0.00112888)
##
## -----------------------------------------------------------------------------------
## Observations                     2,545          2,545         2,545         2,545
## R2                             0.00581909     0.00506681    0.00526636    0.00347171
## Adjusted R2                   -1.29093900    -1.29267200   -1.29221200   -1.29634800
## F Statistic (df = 6; 1104)     1.07698000     0.93704000    0.97414000    0.64102050
## ===================================================================================
## Note:                                                  *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Results
## ==========================================================================================================
##                                                        Dependent variable:
##                              -----------------------------------------------------------------------------
##                                                           df[, disease]
##                                mbCataract   mbTuberculosis   mbHighBP    mbHeartDisease   mbDiabetes
##                                   (1)            (2)            (3)           (4)             (5)
## ----------------------------------------------------------------------------------------------------------
## distanceToPavedRoad          -0.00002286    -0.00004061    0.00005063    0.00006011    -0.00014060***
##                              (0.00006292)   (0.00015591)  (0.00008458)  (0.00013302)   (0.00004808)
##
## electricity                   0.00033340     0.00290024   -0.00012059    0.00021733    -0.00101339
##                              (0.00104230)   (0.00258269)  (0.00140106)  (0.00220365)   (0.00079644)
##
## income                      -0.00000001***  -0.00000001    0.00000000   0.00000002**    0.00000000
##                              (0.00000000)   (0.00000001)  (0.00000001)  (0.00000001)   (0.00000000)
##
## distanceToNearestTown         0.00001331    -0.00000268   -0.00000847   -0.00003382     0.00001154
##                              (0.00002158)   (0.00005346)  (0.00002900)  (0.00004561)   (0.00001649)
##
## writingScore                  0.00046128    -0.00175237   -0.00083289    0.00126897     0.00007638
##                              (0.00047351)   (0.00117328)  (0.00063648)  (0.00100109)   (0.00036181)
##
## smokeTobacco                  0.00021774     0.00074875   -0.00031719  -0.00155046**   -0.00001544
##                              (0.00031460)   (0.00077954)  (0.00042289)  (0.00066513)   (0.00024039)
##
## ----------------------------------------------------------------------------------------------------------
## Observations                    2,545          2,545          2,545         2,545           2,545
## R2                            0.01048724     0.00407693     0.00443986    0.00883507     0.00907738
## Adjusted R2                  -1.28018200    -1.29495300    -1.29411700   -1.28398900    -1.28343000
## F Statistic (df = 6; 1104)   1.95010400*     0.75322640     0.82057830    1.64014400     1.68553700
## ==========================================================================================================
```

```
## Note:                                                        *p<0.1; **p<0.05; ***p<0.01
```

From the univariate results, only the non-communicable diseases appeared to have a statistically significant negative relationship between disease probability and road type. For example, a one kilometer increase in road distance results in a 0.05% decrease in the percentage of individuals in a village that have cataracts. However, when controlling for electricity access, income, distance to nearest down, writing ability, and tobacco smoking habits, this significance is lost. To understand these results better, a more robust analysis must be conducted controlling for more factors, such as differences in medical facilities, which is possible given the IHDS data. Djemai offers a robust example of which factors to control for in his regression model (2011). Additionally, as explained before, transforming road distance to log(1 + roadDistance) may account for the non-linear relationship between road distance in the next regression analysis. If the relationship between road distance and health outcomes remains insignificant, then potentially healthcare centers are such low quality that increasing access confers no health advantage. I can investigate this by examining the type of medical facilities sick people visit, how long they stay there, and how long they remain too sick to function. This secondary analysis can offer a deeper insight into whether roads are increasing access to health facilities and whether these health facilities are improving patient outcomes.

References

Bell, Clive and Susanne van Dillen (2014). "How Does India's Rural Road Program Affect the Grassroots? Findings from a Survey in Upland Orissa." Land Economics 90(2): 372-394. https://www.researchgate.net/publication/254456306_How_Does_India's_Rural_Roads_Program_Affect_the_Grassroots_Findings_from_a_Survey_in_Upland_Orissa.

Djemai, Elodie (2011). "HIV and Access to Road." Working Paper. http://www.ipc.umich.edu/pdf/Djema-HIV-Access-to-Road.pdf.

Ministry of Rural Development, Government of India (2004). Impact Assessment of Pradhan Mantri Gram Sadak Yojana: http://pmgsy.nic.in/pmg122.asp.

World Bank (2010). Project Appraisal Document. Report No. 570810-IN: http://documents.worldbank.org/curated/en/879581468041646928/pdf/570810PAD0P1241e0only1910BOX353794B.pdf.