Referee Response:

The Effect of Road Access on Disease Incidence in Rural India

Yeshwant Chillakuru

Department of Economics, The George Washington University

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

5 April 2017

Requests from Professor Sinclair

- Professor requested to use the smallest sample for all analysis as certain controls lowered the sample size during regression analysis. This was addressed by dropping all observations with missing values for all regressor terms. This dropped the sample size from 2800 observations to 2545 obersvations (9.11% drop in sample size).

- Professor commented why so many obersvations were missing for Equations (2) regressions (599 observations used in Equation (1) regressions compared to 2585 used in Equation (2) regressions). This is because Equation (2) only examines how change in distance to paved affects health outcomes, and so the analysis includes only villages with a paved road (599 villages). This is stated Section III (Model) of the paper, and this approach for analysis was agreed upon with the Professor for ease of interpreability and for isolating the two effects (distance to road and if a road is present or not).

- Professor commented to examine the R^2 values between the two models to see which would be a better model. An analysis of Adjusted R^2 was included in the Robustness Checks section of the paper. Adjusted R^2 is more appropriate for analysis because it penalizes additional of more variables into the regression, thus is a more rigorous measure of fit. Analysis of it was included in the Robustness Checks section because I discuss the extremely weak Adjusted R^2's as a weakness in the model.

- Professor suggested to examine which control variables cause a large drop in the coefficient when looking at the impact of road on any disease. This analysis is now included in the results section

and the Professor's suggestion was expanded to look at a similar drop in non-communicable disease. The "smokeTobacco" variable was the primary cause the drop, with the "income" variable playing an additional minor role in some circumstances.

- Professor suggested to include base model (univariate without panel effects) regressions. These are now included in the analysis in the Results Section.

Requests from Jessica Goldenring (referee)

- Minor grammatical, spelling, and formatting changes were made.

Intro Section

- Jessica suggested to define "rural" and explain why/how villages were chosen. This explanation is now given. "Rural" as used in our analysis is determined by the guidelines used by IHDS when constructing their dataset.
- Clarifications were made regarding the usage of "robust" when describing data and the upland Orissa region.
- As requested by referee, brief definitions of communicable and non-communicable diseases were included.
- Referee suggested use of acronym STI instead of STD. However, IHDS data encodes data as STD, and I decided to use STD acronym in order to maintain consistency with parent dataset and the IHDS documentation.
- Road map now includes the conclusion, as suggested by the referee.
- As suggested by referee, discussion of my analysis on STD/AIDS is included in the introduction.

Data Section

- Jessica suggested to add coefficients to the slop of the lines in the graph. These are now included in a footnote.
- As suggested by referee, relationship between second to last paragraph and abstract was clarified.
- A discussion was included in the Robustness Section of why not all observations of villages were used in the analysis.
- Jessica was confused to the meaning of the "$diseaseIncidence_{it} = mbAnyDisease_{it}$" notation. This notation was replaced with "$diseaseIncidence_{it}$ is set to $mbAnyDisease_{it}$" as a result.

- As suggested, an explanation of the log transformation of distance to paved road was included in a footnote in the Model section.
- As suggested by the referee, summary statistics were added for the control variables in Table 3 and

Results

- Referee suggested rewording "percentage point increase" increase to "point increase in percent" stating that I was not performing a logit/probit analysis. However, "percentage point increase" and "point increase in percent" are referring to the same thing. "Percent increase" would be the appropriate terminology for a logit/probit model: https://www.reed.edu/academic_support/pdfs/qskills/percent.pdf.

Robustness Checks Section

- Non-random road placement was addressed here, as suggested by referee.
- As suggested by referee, omitted variable vias was further discussed here.
- As suggested by referee, internal validity, with respect to survey biases and self-reporting health-outcomes, was discussed.
- As suggested by referee, external validity, regarding generalization of results to India and broader, is discussed.

Conclusion

- As suggested by referee, discussion about null findings and their reltinoaship to the Bell and van Dillen (2014) and Djemai (2011) papers was moved to the Robustness checks.