Journal of Rural & Industrial Development
9 (2) 2021, 27-37
http://publishingindia.com/jrid/

The Causal Effect of Road Connectivity on Healthcare in Previously Unconnected Villages in India: Fuzzy Regression Discontinuity Estimation of the Impact of PMGSY

T. Lakshmanasamy*

Abstract

The Pradhan Mantri Gram Sadak Yojana (PMGSY) programme of India provides new road connectivity for unconnected habitations with a population of at least 500. This paper evaluates the causal effect of the new road connectivity on the healthcare benefits accrued to the previously unconnected village households, specifically the pregnancy care and contraceptive use, awareness and hygiene, and social interaction. The nonparametric fuzzy regression discontinuity design (FRDD) is applied to the data derived from the 2007-2008 District Level Household Survey (DLHS-3) and the Socioeconomic High-Resolution Rural-Urban Geographic Platform for India (SHRUG) data. The FRDD estimates show that in the treatment villages, more women seek antenatal care, have delivery conducted in hospitals and use modern contraceptive methods. In the villages newly connected with roads, awareness of government health care programmes like prevention of sex selection and female foeticide, treat water and take health insurance coverage has increased. The provision of all-weather roads to unconnected villages is also more likely to impact social interaction and more participation in women's self-help groups, and village assembly takes important decisions on preventive healthcare.

Keywords: PMGSY, Villages, All-Weather Road, Connectivity, Healthcare, Fuzzy Regression Discontinuity Design

Introduction

Public programmes are designed to reach certain goals and larger beneficiaries. The accruing benefits of a policy or programme intervention are long-term and can have substantial spillover effects. Some of the programmes might appear potentially promising before implementation yet they may fail to generate expected impacts or benefits. Whether policy interventions or public programmes generate intended effects are generally obtained by evaluation studies. Programme evaluation helps policymakers to promote accountability in the allocation of resources across public programmes and provides an understanding of what works, what does not, and how measured benefits are attributable to a particular project or policy intervention. However, the effects of the new policy or programme can be ascertained only after the implementation of the programme and any amount of prior data will not capture the possible outcomes of the programme. Moreover, the realised outcome of a programme, however obtained, cannot be attributed to the programme alone as the obtained outcome may be due to other influencing factors and there may also be potential selection bias. This makes the identification of the causal effect of the programme impossible.

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is one such programme aimed at creating larger benefits through road connectivity to villages to a set of basic health and economic facilities in India. The PMGSY was launched on December 25, 2000. The PMGSY seeks to build a new all-weather road to reach the nearest link road for hitherto unconnected rural habitats with a population of at least 500. Under this programme, priority is given to larger villages and the possibility that the new roads benefit other rural habitats. Between 2000 and 2014, 38710 roads were sanctioned of which 102,828 roads, covering 391,991 kms benefitting 172,052 habitations

* ICSSR Senior Fellow and Formerly Professor and Head, Department of Econometrics, University of Madras, Chennai, Tamil Nadu, India. Email: tlsamy@yahoo.co.in

have been completed. Under this scheme, sometimes it is also possible that a village with a fewer population might get a new road before a village with a larger population.

The benefits of a road programme are not limited to direct benefits like increased transportation, faster transit, etc. Road connectivity importantly reduces the costs of travel cost including time, benefits higher household income, allows information and awareness, improve the access to health care services, and facilitates social interactions within and outside the village. People living in unconnected villages will often find it difficult to access various services outside the village. In fact, low usage of preventive healthcare is a major challenge in India. Access to roads can increase preventive healthcare usage through various channels like reduction in transportation cost and travel time. All these factors together can additionally increase the usage of preventive healthcare. An assessment of the impact of such a wider programme is not feasible with a simple focus on the road connectivity benefit alone or a cost-benefit analysis. With confounding accrual of various benefits, identification of the actual impact of the programme needs wider evaluation methods. Fortunately, the impact or programme evaluation methods help in estimating the intervention effects of a larger programme.

Impact evaluation assesses the mechanisms by which beneficiaries are responding to the intervention. Unlike several evaluation approaches such as comparing outcomes before and after the implementation of the programme or comparing the outcome of beneficiary or target group of the programme with non-beneficiary or control groups, the impact evaluation assesses the benefits of the programme within the group of targeted beneficiaries. To know the outcome before the implementation of the programme is impossible. Comparison with a control group is problematic because of individual heterogeneity and ineligibility to the benefits of the programme. This makes impact evaluation different from other evaluation approaches. The impact evaluation approaches construct a potential outcome without the programme and then compares it with the actual outcome of the programme. In the absence of data on outcome prior to the implementation of the programme, the potential outcome is measured by a counterfactual outcome, the outcome for participants had they not been exposed to the programme. The difference in the counterfactual outcome and the actual outcome is identified as the actual benefits of the programme. Thus, impact evaluation isolates the causal effect of the programme from other factors and potential selection bias, identifying the accrued benefits are indeed due to the programme intervention and not to other factors.

Successful impact evaluation hinges on finding a good comparison group. The comparison group should be just like the targeted group. As it is impossible to identify such a comparable group because of individual heterogeneity and eligibility to the programme, the impact evaluation identifies the eligible beneficiaries of the programme themself as the comparison group. That is the impact evaluation compares how the same beneficiary would have fared with and without the programme intervention. The main challenge of impact evaluation is to find a good counterfactual, the situation a participating subject would have experienced had the beneficiary not been exposed to the programme. The problem is that while the programme's impact can be assessed only by computing actual and counterfactual outcomes, the counterfactual is not observed. Therefore, the challenge is to create a convincing and reasonable comparison group for beneficiaries in light of this missing data. Fortunately, a beneficiary's outcome in the absence of the intervention, if constructed, would be its counterfactual. But one cannot do so because at a given point in time the individual cannot be in both the treatment and control groups at the same time. Without information on the counterfactual, the next best alternative is to compare outcomes of the treated group with those of the comparison group that has not been treated. In doing so, one attempts to pick a comparison group that is very similar to the treated group, such that those who received the treatment would have had outcomes similar to those in the comparison group in the absence of treatment.

This paper analyses the impact of the massive road-building programme undertaken by the Government of India, the Pradhan Mantri Gram Sadak Yojana (PMGSY) on benefits accrued to households, specifically the preventive healthcare benefits. The main objective of this paper is to estimate the causal effect of new road connectivity in Indian villages on preventive healthcare - pregnancy care and contraceptive use, awareness and hygiene, and social interaction - among people in previously unconnected villages in India. This paper uses secondary data from two sources. Information about preventive healthcare is derived from the District Level Household Survey 2007-2008 (DLHS-3) and the PMGSY programme placement data is derived from the Socioeconomic High-Resolution Rural-Urban Geographic Platform for India (SHRUG) of the Development Data Lab. A nonparametric econometric method, the Fuzzy Regression Discontinuity Design (FRDD) is used to estimate the causal effect of the road-building programme on preventive healthcare in the Indian villages.

Review of Literature

Goel and Gupta (2017) estimate the causal impact of Delhi Metro in reducing nitrogen dioxide (NO2), carbon monoxide (CO) and particulate matter (PM2.5), the three localised transportation source pollutants present at dangerously high levels in Delhi. The RDD estimates show that there is a significant reduction in two transportation source pollutants, carbon monoxide and nitrogen dioxide.

Aggarwal (2018) analyses the causal impact of the PMGSY programme on rural households. The impacts considered are the type and variety of goods consumed, adoption of agricultural technology, investments in child human capital and adult occupation. A nonparametric method, the Difference-in-Difference (DID) technique is used to estimate the difference between treatment and control groups over time. The study finds that the residents in districts with greater road connectivity pay a reduced price for urban goods. Also, road connectivity improved the consumption variety, especially goods that are not locally produced. Further, fertilizer and hybrid seeds use have improved in the villages with paved roads. In these districts, the school enrollment of younger children has also increased.

Lehne et al. (2018) study the causal effect of politically driven corruption on road construction using the nonparametric RDD estimation method. The study finds that in close elections political interference in contracting road construction is associated with higher costs and a greater likelihood that roads go missing.

Prakash et al. (2019) analyse the causal impact on the economic performance of elected politicians with criminal accusations. The study finds that in constituencies that elected these politicians the annual road length construction under the PMGSY is 24-percentage points lower than in constituencies with a non-accused politician elected. This implies huge aggregate economic costs as well as individual costs due to lower quality persons being elected to assemblies. Ultimately, electing a criminally accused person to the legislature results in foregoing access to public services.

Asher and Novosad (2020) examine the causal effect of the PMGSY programme on local economic development by estimating the treatment effect of new roads using the FRDD framework. The study finds that the effect of the PMGSY programme on village economies are smaller than those anticipated or the existing evidence on the benefits of roads. The study reports no major changes in agricultural outcomes, income or assets, but a slight expansion of employment in firms in the villages. The paper notes that rural growth in India is constrained more by the poor state of transportation infrastructure and suggests that the main economic benefit of rural transportation infrastructure is the connection of rural workers to new employment opportunities rather than facilitating the growth of village farms and firms.

Adukia et al. (2020) study the causal impact of roads on educational outcomes in villages that received new roads by 2015 under the PMGSY programme using panel fixed effects regression that exploits the timing of road construction. The study estimates the impact of new roads on schooling using the RDD framework. Their estimated results show that road construction has significantly increased middle school enrollment in the villages that are connected by new roads. Further, children stay in schools longer and perform better in examinations.

Data and Methodology

This paper uses two secondary datasets. The data on roads constructed under the Pradhan Mantri Gram Sadak Yojana (PMGSY) programmes are obtained from the Socioeconomic High-Resolution Rural-Urban Geographic Platform for India (SHRUG) from the Development Data Lab. It contains programme placement data of PMGSY from 2000 to 2014. The data on preventive healthcare is derived from the District Level Household Survey 2007-2008 (DLHS-3) data published by the International Institute of Population Studies, Mumbai. The period considered spans from December 2007 to December 2008. For the analysis, the DLHS-3 data are matched to the SHRUG data identifying the key variables. The variable used in the analysis is spread across three different surveys of the DLHS-3, survey of village, survey of ever-married women and survey of households. The SHRUG data is matched with the DLHS-3 data at the village level, owing to the fact that the treatment is provided at the village level. The resulting data set has 2172 observations each representing a village. This data set is then matched to the ever-married women survey and household survey separately. The matched women survey data set has 50,132 observations. The matched household data set has 55,382 observations.

For the application of the fuzzy regression discontinuity design, two variables, a treatment variable and a running variable need to be identified. The treatment variable used is the sanction year of the earliest completed new road. The treatment variable is to be binary i.e. 1 representing assignment to treatment and 0 representing not assigned to treatment. In this paper, the running variable is the village population based on the 2001 census. All the outcome variables are to be binary. From the DLHS-3, the outcome variables considered for the analysis are pregnancy and antenatal care, contraceptive use and social interactions in villages with responses as 1 for 'yes', and 0 for 'no'. Empirically, the impact of new roads on the outcome variables are estimated by the nonparametric fuzzy regression discontinuity design (FRDD) commands in the software R programme (Calonico et al., 2015; 2016; Thoemmes et al., 2016; Dimmery, 2017).

Fuzzy Regression Discontinuity Design (RDD) Method

Regression Discontinuity designs (RDD) has been an effective policy evaluation strategy that provides unbiased estimates of policy intervention. The RDD is a popular quasi-experimental design used when random assignment is not feasible, but the precise knowledge of the rules determining the eligibility into treatment is known. Given the exogenously fixed cut-off that determines eligibility for treatment, the comparison of outcomes closely around the cut-off i.e. above and below the eligibility criterion filters the factors unrelated to the programme. The RD design was first introduced by Donald Thistlethwaite and Donald Campbell (1960) as a way of estimating treatment effects in a nonexperimental setting where treatment is determined by whether an observed assignment variable exceeds a known cut-off point. Jinyong Hahn, Petra Todd and van der Klaauw (2001) formalised the RD design estimation of the causal effect of interventions. The RD designs require seemingly mild assumptions compared to those needed for other non-experimental approaches. The RD design is not just another evaluation strategy as the causal inferences from RD designs are potentially more credible than those from typical natural experiment strategies. One need not assume that the RD design isolates treatment variation that is as good as randomised, instead, such randomised variation is a consequence of agents' inability to precisely control the assignment variable near the known cut-off (Trochim, 1984; Lee, 2008; Khandker, 2010; Lee & Lemiux, 2010; Jocob et al., 2012; Cattaneo et al., 2020). In addition to providing a highly credible and transparent way of estimating programme effects, the RD designs can be used in a wide variety of contexts covering a large number of important economic questions.

In any intervention programme that desires to change the outcomes for certain targeted groups, the eligibility for inclusion or exclusion to the programme is normally conditioned by some threshold or cut-off value of an observed variable. In the impact evaluation literature, this value is known as the assignment or forcing or running variable and is denoted as x. An individual is assigned to a treatment i.e. benefit of a programme if the value of x is greater than or lesser than or equal to a known cut-off value 'c' of x, depending on the intention of the intervention or programme. Therefore, the outcome variable is a discontinuous function of the assignment variable. This discontinuous 'jump' or 'sharp discontinuity' at the cut-off in the linear relationship between the outcome and assignment variables is attributed to the causal effect of the intervention. A simple way of estimating the treatment effect τ is by a simple dummy in linear regression:

$$y = \alpha + d\tau + \beta x + \varepsilon \tag{1}$$

where y is the outcome of interest, x is the assignment variable, and d is a binary treatment indicator such that $d \in \{0.1\}$, $d = 1$ if $x > c$ and $d = 0$ if $x < c$.

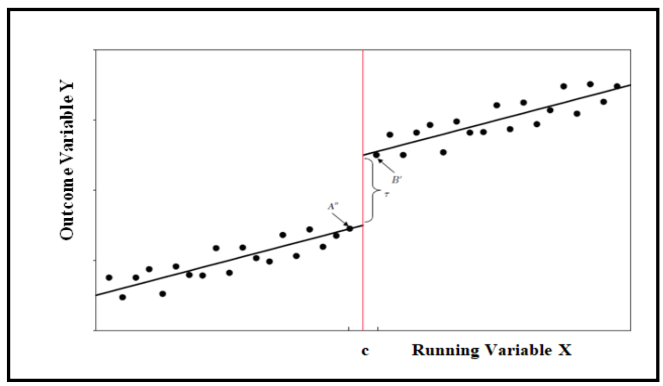

Fig. 1: Regression Discontinuity Design

Fig. 1 represents a regression discontinuity design. The outcome is measured on the Y-axis and the assignment variable is measured on the X-axis, where the vertical line that cuts the graph is the cut-off or eligibility threshold. All units to the right of this cut-off represent the treatment group i.e. units that get treated and all units to the left of this cut-off represent the control group i.e. units that do not get treated. Regression estimation with observations close (both above and below) to the cut-off produces the outcome with treatment and the counterfactual outcome of not receiving the treatment , and the difference being the causal effect of the intervention programme τ.

A key assumption of the RD design is that all other factors that determine the outcome should be smooth with respect to the eligibility variable i.e. the baseline characteristics determined prior to the realisation of the assignment variable should have the same distribution just above and just below the cut-off. Moreover, using only those observations very close to the cut-off value of the assignment variable is costly in terms of fewer observations and ignores data away from the cut-off. Following the pioneering formalisation of the RD design by Hahn et al. (2001), a 'potential outcomes' framework solves the issues of smoothness and continuity assumptions with respect to the running variable.

Consider an individual who is eligible for treatment under a programme, but he may be observed in or out of the treatment. Hence, for this individual, there exists a pair of 'potential' outcomes: yi(1) for what would occur if the unit were exposed to the treatment and yi(0) if not exposed. Therefore, the causal effect of the treatment is simply the difference yi(1) - yi(0), the latter operating as the valid counterfactual. However, a fundamental problem of this causal inference is that yi(0) and yi(1) pair cannot be observed simultaneously. To circumvent this problem, instead of estimating unit level effects, estimate the treatment effects to each sub-group i.e. eligible and treated and eligible but not treated, separately and then take the average effects of the treatment i.e. averages of yi(1) - yi(0). The underlying relationships between average outcomes and x can be represented as E[yi(1)|x] and E[yi(0)|x]. Hence, all individuals observed above the cut-off eligibility are exposed to the treatment and all individuals observed below it are untreated. Then, an estimate of the causal effect with observables is given by the outcomes of the treatment difference:

$$TE(= B' - A'') = \lim_{\varepsilon \downarrow 0} E[\, y(1) \mid x\,] = c + \varepsilon - \lim_{\varepsilon \uparrow 0} E[\, y(0) \mid x\,] = c + \varepsilon \tag{2}$$

Therefore, is the 'average treatment effect' at the cut-off c is:

$$ATE = E[\, y(1) \mid x\,] - E[\, y(0) \mid x\,] = E[y(1) - y(0) \mid x = c] \tag{3}$$

However, this form of 'sharp' RD design requires deterministic assignment i.e. the probability of treatment jumps from 0 to 1 at the precise cut-off value, with no cross-overs. If individuals can manipulate the assignment variable, then factors other than the threshold rule may influence the probability of programme participation. In the absence of the strict assignment rule, some ineligible will be there in the treatment group and some eligible would have been omitted from the benefits of the intervention, allowing cross-overs at the cut-off value. Therefore, the eligibility criterion acts as a nudge. With the probabilistic assignment, the regression discontinuity takes a 'fuzzy' RD design, allowing for a small jump at the threshold value (Trochim, 1984). Fig. 2 represents the sharp and fuzzy regression discontinuity designs.

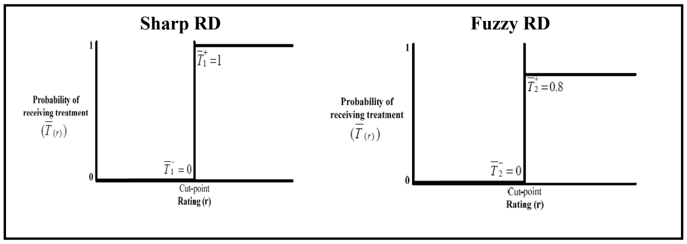

Fig. 2: Sharp and Fuzzy RD Designs

With imprecise control over the assignment rule, the variation in the treatment in a neighbourhood of the threshold is as good as randomisation (Lee & Lemieux, 2010). Even if some are likely to have values of x near the cut-off, every individual will have approximately the same probability of having an x that is just around the cut-off. With randomisation, the RD design takes the simple dummy variable formulation:

$$\begin{aligned} y &= \tau d + \gamma_1 z + u \\ d &= 1[x > c] \\ x &= \gamma_2 z + v \end{aligned} \tag{4}$$

where z is a vector of predetermined characteristics that affect the outcome and/or assignment variable. Therefore,

$$\lim_{\varepsilon \downarrow 0} Pr[\, d = 1 \mid x = c + \varepsilon\,] \neq \lim_{\varepsilon \uparrow 0} Pr[\, d = 1 \mid x = c + \varepsilon\,] \tag{5}$$

and the causal effect is estimated as the outcomes of the treatment difference.

However, the estimate is no longer the average treatment effect as the probability of treatment jumps by less than one at the threshold. The treatment effect is recovered from the discontinuity jump in the relation between d and x, as the ratio of the jump in the relationship between y and x at c to the fraction induced to take up the treatment at the threshold. Thus, the treatment effect under fuzzy RDD can be specified as the ratio of the RD gaps in y and d:

$$\tau^F = \frac{\lim_{\varepsilon \downarrow 0} E[y|x = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[y|x = c + \varepsilon]}{\lim_{\varepsilon \downarrow 0} E[d|x = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[d|x = c + \varepsilon]} \tag{6}$$

Estimation Strategy

The RD design can be conveniently described by a two equation system. The regression lines to the left and right of the cut-off point (x-c):

$$y = \alpha_l + f_l(x - c) + \varepsilon \tag{7}$$

$$y = \alpha_r + f_r(x - c) + \varepsilon \tag{8}$$

where l and r represent the left and right sides of the cut-off value c, and f represents the functional form. The treatment effect then is the difference between the intercepts on the two sides of the cut-off point, $\tau = (\alpha_r - \alpha_l)$. A pooled RD design linear regression can be specified as:

$$y = \alpha_l + d\tau + \{\beta_l(x - c) + d[\beta_r(x - c) - \beta_l(x - c)]\} + \varepsilon \quad (9)$$

Allowing interaction between d and x, the regression equation is specified as:

$$y = \alpha_l + d\tau + \beta_l(x - c) + (\beta_r - \beta_l)d(x - c) + \varepsilon \quad (10)$$

While the parametric estimation of the RDD regression equation requires a priori distributional assumptions, nonparametric estimation such as kernel estimation allows empirical distribution. The nonparametric RDD regression can be specified with cut-offs $c - h \leq x \leq c + h$. However, nonparametric estimation requires the identification of optimal bandwidth, the width of the bins that balances the precision and bias in the regression estimates. The nonparametric econometrics literature offers various methods for choosing optimal bandwidth, like rule-of-thump and cross-validation criteria. Often the nonparametric RDD results are also depicted graphically. An advantage of the RDD graph is its visual show of discontinuity and transparent identification of the treatment effect. Hahn et al. (2001), Porter (2003), Imbens and Lemieux (2008) and Porter and Yu (2015) show that the nonparametric local linear or polynomial regression, based on cross-validation bandwidth selection, provides consistent estimates of the treatment effect. Under monotonicity, the estimated treatment effect is the local average treatment effect.

Given treatment t, the fuzzy RD design can be specified as:

$$\begin{aligned} y &= \theta t + \gamma_1 z + u_1 \\ t &= \mu d + \theta z + u_2 \\ d &= 1[x > c] \\ x &= \gamma_2 z + v \end{aligned} \quad (11)$$

where the parameter θ is the effect of treatment on the outcome. The probability of treatment can be specified as:

$$\Pr(d = 1 \mid x = \tilde{x}) = \delta + \gamma t + g(x - c) \quad (12)$$

where $t = 1\ [x \geq c]$ indicates whether the assignment variable exceeds the eligibility threshold c. Since $d = \Pr(d = 1|x = x_0)$, the fuzzy RD design can be described in terms of the two equation system:

$$y = \alpha + \tau d + f(x - c) + \varepsilon \quad (13)$$

$$d = \theta + \gamma t + g(X - c) + v \quad (14)$$

Substituting the treatment determining equation into the outcome equation yields the reduced form equation:

$$y = \alpha_r + \tau_r t + f_r(X - c) + \varepsilon_r \quad (15)$$

where $\tau_r = \tau\gamma$ is interpreted as the 'intent-to-treat' effect. Since the model is exactly identified, 2SLS estimates are numerically identical to the ratio of the reduced form coefficients τ_r/γ of the local linear or local polynomial regressions when the same bandwidth is used in both the treatment and outcome regressions (Imbens & Lemieux, 2008).

Empirically, the two estimating equations are specified as:

$$t_{ivd} = \beta_0 + \beta_1 x_{vd} + \beta_2 d_i + \varepsilon_{ivd} \quad (16)$$

$$y_{ivd} = \gamma_0 + \gamma_1 x_{vd} + \gamma_2 t_{ivd} + \gamma_3 d_i + \varepsilon_{ivd} \quad (17)$$

where y_{ivd} indicates the outcome variable, t_{ivd} is the treatment indicator whether the village has got a new road, di is a dummy to indicate whether the individual is above or below the cut-off, y_2 is the treatment effect at cut-off, x_{vd} is the running variable i.e. population of the village. The local average treatment effect τ_F is then estimated as the ratio of the gaps in treatment outcomes to the gaps in the eligibility i.e. y and d, as implied by equation (4).

Empirical Results

The DLHS-3 2007-08 survey results presented in Table 1 reveal that more than 75 percent of women in India received antenatal check-up, out of which 55 percent of women received antenatal care check-up from a government health facility and 36 percent from a private healthcare facility and around 10 percent from community-based services like non-government hospital/trust hospital or clinic, own home, parent's home and others. Government health institutions account for about 47 percent of child delivery while 52 percent of child deliveries are at home. Childbirths in health institutions for women under age 35 years is higher than for women aged 35 years and above.

Table 1: Antenatal Checkup and Place of Delivery in India

Age Group	*Total*	*Government*	*Private*	*Community*	*No. of Women*
Antenatal Checkup					
15-19	76.6	54.3	35.7	11.8	14,006
20-24	78.9	54.7	37.4	9.0	73,455
25-29	76.7	54.5	37.3	8.7	72,061
30-34	71.4	53.4	35.3	10.3	35,246
35+	60.5	55.6	29.6	12.3	20,280
Place of Delivery and Assistance					
	Institutional Delivery	Home Delivery	Assisted Home Delivery	Safe Delivery	No. of Women
15-19	47.1	52.1	5.4	52.5	14,006
20-24	50.4	48.9	5.6	56.0	73,455
25-29	48.7	50.6	5.3	54.0	72,061
30-34	43.7	55.7	5.2	48.9	35,246
35+	33.1	66.1	5.1	38.2	20,280

Source: DLHS-3 2007-08.

The descriptive statistics of the outcome variables are presented in Table 2. The outcome variables - pregnancy care and contraceptive use, awareness and hygiene, and social interaction - are binary i.e. 1 means the presence of the outcome and 0 means absence of the outcome. The descriptive statistics show a higher proportion of female sterilisation, delivery at home, awareness of personal hygiene, AIDS and DOTS, and participation in a self-help group and inter-village assembly but low use of pills as contraception, awareness of health scheme, and participation in welfare committee among the people in the villages of India.

Table 2: Descriptive Statistics of Outcome Variables

Variable	*Yes (1)*	*No (0)*	*Mean*	*Std. Dev*	*Obs.*
Pregnancy Care and Contraceptive Use					
Male sterilisation	1275	2038	0.385	0.487	3313
Female sterilisation	1665	1648	0.503	0.500	3313
Daily pills	1235	2047	0.382	0.486	3313
Weekly pills	289	3023	0.087	0.282	3313
Condom	807	2505	0.244	0.429	3313
Delivery at hospital	5291	14056	0.273	0.446	193447
Delivery at home	14056	5291	0.726	0.446	19347
Awareness and Hygiene					
Awareness about AIDS	22547	33835	0.407	0.491	55382
Awareness about DOTS	29104	26278	0.525	0.499	55382
Awareness about sex selection	20383	34999	0.368	0.482	55382
Awareness about personal hygiene	41260	14122	0.745	0.436	55382
Use treated water	13422	41930	0.247	0.465	55382
Joined a health scheme	1229	53524	0.113	0.859	55382
Village Level Social Interaction					
Youth club	462	1710	0.213	0.409	2172
Women's help group	781	1391	0.360	0.471	2172
Self-help group	1005	1167	0.463	0.499	2172
Village welfare committee	291	1181	0.134	0.341	2172
Improvement in clinic	290	1882	0.134	0.340	2172
Inter-village assembly decision making	825	724	0.533	0.499	1549

Table 3 presents the treatment effect estimates of the outcome variables for different bandwidth choices. The RDD plots are presented in Fig. 3 showing the discontinuity and the treatment effects visually. The coefficient estimates are the local average treatment effect (LATE) for the compliers, those individuals who receive treatment when assigned to the treatment group when they satisfy the cut-off rule and do not receive the treatment when assigned to the control group. The Fuzzy RDD estimates on contraceptive use in the treatment villages show that women-centric contraception is the dominant form and reliance is more on condoms and daily pills as a means of contraceptive. Women in treatment villages are 91 percent more likely to use condoms and 65 percent more likely to use daily pills. The likelihood of female sterilisation decreases by 96 percent in these villages. These results are encouraging as women are more aware of healthy childbearing practices and family planning.

However, in the treatment villages, childbirth in hospitals is likely to decrease significantly and at the same time, childbirth at home is likely to increase significantly. This implies that connecting a village to the all-weather road via a link road improves the mobility of people in and out of the village. This makes the provision of antenatal care by villages nurses and ASHAs more effective.

In the villages newly connected by roads, the likelihood that a household is aware of the directly observed treatment short course (DOTS) on Tuberculosis increases by 88 percent and that of prevention of sex selection increases by 91 percent. Households in the treatment villages are well aware of personal hygiene whose likelihood increases by 52 percent. The connectivity improves the information and awareness on preventive techniques and these influences the behaviour of people in remote towards preventive healthcare.

In the treatment villages with a population of 500 and above, the provision of a new road increases the likelihood of participation in the inter-village assembly by 94 percent and in the village welfare committee by 44 percent, and the likelihood that the treated village has women self-help group increases by 88 percent. However, the construction of a new road in these villages decreases the likelihood of the presence of a youth club by 79 percent and a self-help group participation by 66 percent. However, none of the estimates at the village level is statistically significant. Overall, the availability of a new road has little relevance for social interaction in the villages of India.

Table 3: Fuzzy Regression Discontinuity Estimates of Treatment Effects

Variable	*LATE*	*Bias-Corrected*	*Robust*	*BW=0.5*	*BW=2*
Pregnancy Care and Contraceptive Use					
Male sterilisation	0.020 (0.317)	-0.130 (0.317)	-0.130 (0.363)	0.321 (0.503)	0.417* (0.243)
Female sterilisation	-0.959* (0.360)	-0.884* (0.360)	-0.884* (0.421)	-0.845** (0.823)	-0.644* (0.240)
Daily pills	0.658* (0.244)	0.704* (0.244)	0.704** (0.297)	0.865** (0.497)	0.475* (0.184)
Weekly pills	0.166 (0.110)	0.197*** 9.110)	0.197 (0.125)	0.279 (0.171)	0.133*** (0.072)
Condom	0.915* (0.246)	0.915* (0.246)	0.934* (0.303)	1.026* (0.538)	0.811*** (0.188)
Delivery at home	-0.145* (0.101)	-0.167* (0.101)	-0.167** (0.155)	-0.075 (0.101)	-0.292** (0.127)
Delivery at hospital	0.705* (0.421)	0.731* (0.421)	0.731** (0.493)	0.201 (0.308)	0.799*** (0.417)
Awareness and Hygiene					
Awareness about AIDS	0.003 (0.166)	-0.045 (0.166)	-0.045 (0.180)	-0.123 (0.243)	0.231 (0.220)
Awareness about DOTS on Tuberculosis	0.877* (0.251)	0.886* (0.251)	0.886* (0.299)	0.670* (0.211)	0.907* (0.314)
Awareness about sex selection	0.912* (0.225)	0.943* (0.225)	0.943* (0.266)	0.906* (0.219)	1.035* (0.325)
Awareness about personal hygiene	0.523** (0.217)	0.468** (0.217)	0.468*** (0.248)	0.284 (0.196)	0.820* (0.274)
Village Level Social Interaction					
Youth club	-0.794 (1.325)	-0.726 (1.325)	-0.726 (1.614)	-0.198 (0.877)	-0.569 (0.637)
Women's help group	0.885 (1.377)	0.683 (1.377)	0.183 (1.639)	0.475 (1.715)	0.358 (0.681)
Self-help group	-0.663 (2.191)	-0.212 (2.191)	-0.212 (2.573)	-0.804 (2.150)	-0.600 (0.755)
Village welfare committee	0.442 (1.771)	0.987 (1.771)	0.987 (2.121)	0.499 (0.803)	0.107 (0.733)
Inter-village assembly	0.937 (0.741)	0.843 (0.741)	0.843 (0.907)	0.009 (0.617)	0.929 (0.656)

Note: Standard errors in parentheses. Village cut-off: 500. *, **, *** significant at 1, 5, 10 percent levels.

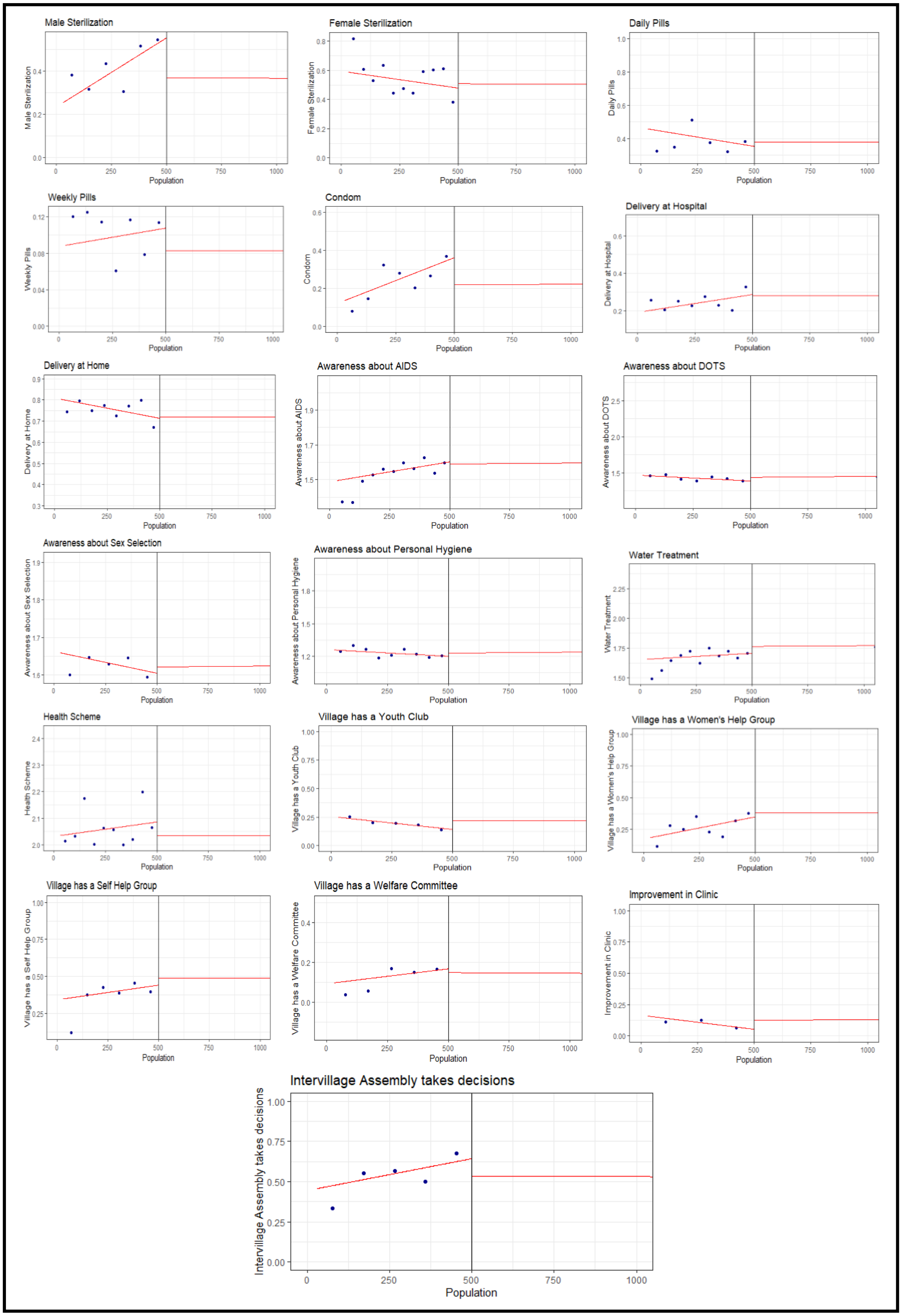

Fig. 3: Fuzzy Regression Discontinuity Plots

Conclusion

The massive Pradhan Mantri Gram Sadak Yojana (PMGSY) programme in India provides new road connectivity for unconnected habitations with a population of at least 500. The provision of new roads under the PMGSY scheme makes it possible to connect villages that were unconnected earlier to the nearest all-weather road. Road connectivity confers numerous benefits to rural masses in villages. The provision of new roads lowers travel costs and increases the opportunities for gainful employment and income of people. Provision of roads to unconnected places increases communication, awareness and use of healthcare facilities and social interaction in villages. This paper evaluates the causal effect of the roads on benefits accrued to households, specifically the preventive healthcare use - pregnancy care and contraceptive use, awareness and hygiene, and social interaction - in previously unconnected villages. The analysis is based on preventive healthcare data from the 2007-2008 District Level Household Survey (DLHS-3) and information about the programme placement from the Socioeconomic High-Resolution Rural-Urban Geographic Platform for India (SHRUG) data. Since the treatment i.e. a new road is provided at the village level, the DLHS-3 data was matched to the SHRUG data at the village level. An impact evaluation method is used to evaluate the causal effect of the new roads on preventive healthcare and other socioeconomic and demographic benefits. A nonparametric econometric method, the Fuzzy Regression Discontinuity Design (FRDD) is used to estimate the causal effect of the road-building programme on preventive healthcare in the Indian villages.

The FRDD estimates show that connecting villages with all-weather roads increases the usage of preventive healthcare. Expect for the village level variables, almost every other variable turns out to be statistically significant. In the treatment villages, larger women go for antenatal care, deliver the child in hospitals and use modern contraceptives. Households in treated villages are likely to be more aware of the various government health care programmes such as prevention of sex selection and female foeticide, treat water and take health insurance coverage. In the treatment villages, social interactions and women's self-help groups are likely to be higher. It is also more likely that in these villages the village assembly takes important decisions regarding preventive healthcare. Except for women's help group and decision making of village assembly, all other variables show increased participation and usage. Overall, the provision of new roads has a significant impact on preventive healthcare usage but an insignificant effect on social interactions in villages in India.

References

Adukia, A., Asher, S., & Novosad, O. (2020). Educational investment responses to economic opportunity: Evidence from Indian road construction. *American Economic Journal: Applied Economics, 12*(1), 348-376.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics, 133*, 375-395.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American Economic Review, 110*(3), 797-823.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Rdrobust: An R package for robust nonparametric inference in regression-discontinuity designs. *R Journal, 7*, 38-51.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2016). Rdrobust: Robust data-driven statistical inference in regression-discontinuity designs. *The Comprehensive R Archive Network /CRAN.*

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2020). *A practical introduction to regression discontinuity designs: Foundations and extensions.* Cambridge, MA: Cambridge University Press.

Dimmery, D. (2017). Rdd: R package for regression discontinuity design. *The Comprehensive R Archive Network /CRAN.*

Goel, D., & Gupta, S. (2017). The effects of metro expansions on air pollution in Delhi. *World Bank Economic Review, 31*(1), 271-294.

Hahn, J., Todd, P., & Klaauw, W. V. D. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-635.

Jacob, R., Zhu, P., Somers, M.-A., & Bloom, H. (2012). *A practical guide to regression discontinuity.* New York: Manpower Demonstration Research Corporation.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. World Bank.

Lee, D., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281-355.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics, 142*(2), 675-697.

Lehne, J., Shapiro, J. N., & Eynde, O. V. (2018) Building connections: Political corruption and road construction in India. *Journal of Development Economics, 131*(1), 62-78.

Porter, J. (2003). *Estimation in the regression discontinuity model* (Unpublished paper, University of Wisconsin-Madison).

Porter, J., & Yu, P. (2015). Regression discontinuity designs with unknown discontinuity points: Testing and estimation. *Journal of Econometrics, 189*(1), 132-147.

Prakash, N., Rockmore, M., & Uppal, Y. (2019). Do criminally accused politicians affect economic outcomes? Evidence from India. *Journal of Development Economics, 141*.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*(6), 309-317.

Thoemmes, F., Liao, W., & Jin, Z. (2016). The analysis of the regression-discontinuity design in R. *Journal of Educational and Behavioral Statistics, 42*(3), 342-360.

Trochim, W. M. K. (1984). *Research design for programmeme evaluation: The regression-discontinuity approach.* Beverly Hills: Sage Publications.