University of Heidelberg

Department of Economics

Discussion Paper Series | No. 602

On the Way to Good Health? Rural Roads and Morbidity in Upland Orissa

Clive Bell and Susanne van Dillen

September 2015

On the Way to Good Health? Rural Roads and Morbidity in Upland Orissa*

Clive Bell[a,b] and Susanne van Dillen[c]

[a] University of Heidelberg, INF 330, D-69120 Heidelberg, Germany

[b] Chr. Michelsen Institute, P.O. Box 6033, N-5892 Bergen, Norway

[c] Schulstrasse 20, D-69221, Dossenheim, Germany

email: clive.bell@urz.uni-heidelberg.de ; susanne@vandillen.de

First version: January, 2015. This version: September, 2015.

*We are indebted to Magnus Hatlebakk, Stefan Klonner and Abhiroop Mukhopadhyay for valuable discussions and written comments, and also to participants in seminars at CMI and Heidelberg for their constructive suggestions. The responsibility for all remaining errors of analysis and opinion is ours alone. The survey rounds in 2009 and 2013 were financed by the World Bank and the Research Council of Norway, respectively. This financial support is gratefully acknowledged, but the opinions expressed herein do not necessarily represent those of either institution.

Abstract

This paper investigates the effects of India's rural roads program (PMGSY) on morbidity, using data on 279 households drawn from 30 villages in a region of upland Orissa. The households were surveyed in 2010 and 2013, yielding an unbalanced panel of 1580 individuals, 1076 of whom were present in both years. Ten villages had received a direct all-weather road connection by 2013. Treating the village as a unit within the whole road network, the provision of a connection, whether direct or in the neighbourhood, is estimated to have reduced an inhabitant's probability of falling sick by an estimated 3.6 percentage points, and the expected duration of incapacitating illness by 0.46 days, for each km. of unpaved track so replaced.

Keywords: Rural roads, morbidity, India

JEL Classification: H54, I10, I15, R41

1 Introduction

The interplay between the extension of transportation networks and economic activity is a topic that has received much attention of late. The contributions feature various combinations of roads and railroads, urbanisation, local growth and trade, in historical and contemporary settings alike (see, for example, Atack et al., 2010; Baum-Snow et al., 2013; Donaldson, 2013; Duranton et al., 2014; Jedwab et al., 2013). There is little, however, on how improved transportation affects morbidity. This lack is rather troubling in the light of the revived emphasis on infrastructure in promoting development in poor countries. One important component of such investment programmes is the large-scale provision of all-weather rural roads, which would serve correspondingly large populations.

Timely and competent treatment can, of course, make the difference between life and death; but in the normal run of things, not only are patients spared some period of pain and suffering, but they are also able to resume earlier the ordinary business of life, be it working or attending school. A failure to get such treatment can also have longer term consequences, such as the permanent neurological damage inflicted by a severe bout of malaria. Thus, even if the provision of all-weather rural roads does little to promote 'normal' preventive care, such as the immunisation of children, it enables treatment that can have preventive effects in the future as well as reducing the duration and severity of patients' current ailments.

A prime example of a rural roads programme is India's *Pradhan Mantri Gram Sadak Yojana* (hereinafter PMGSY), which is one of the largest public sector construction programs in Asia, claiming annually about 0.2% of India's GDP. Launched in 2000, it aimed to provide all of India's habitations with populations exceeding 500 persons (250 in hilly and desert areas) with an all-weather road connection by 2015. At that time, about 170,000 habitations were eligible; at this time of writing, about 60% have their road. By the end of 2010, accumulated expenditures had amounted to about US$14.6 billion, and it was estimated that a further US$40 billion would be required to complete the program by 2020 (World Bank, 2010).

Its architects envisaged that once so connected to the main road network, villagers would enjoy benefits in three spheres: first, the commercial one, as lower transport costs yield better terms of trade in goods and services; second, improved school attendance, by pupils and teachers alike; and third, timely access to medical treatment, especially in the event of accidents and acute illness.

The three spheres of benefits listed above look tidy enough, but there is an interplay between the first and the third that influences the ultimate outcome where morbidity and medical treatment are concerned. A good road is a two-way street in more senses than one; for the commerce that it promotes goes with more frequent personal contacts of various kinds, and towns and markets pool both people and various infectious diseases. The villagers who return from such trips may well bring back pathogens freshly acquired in the course of doing business there. Yet in contrast to this incidental effect of trading, broadly construed, the decision to seek treatment for an ailment involves more than logistics; for the commercial sphere now comes into play in another way. The network of health facilities takes the form mainly of government hospitals and primary health centres (PHCs), which are supposed to provide free treatment. In fact, fees are often demanded, especially in hospitals, and staff, especially doctors, are often absent from PHCs (Chaudhuri *et al.*, 2006; Muralidharan *et al.*, 2010). There are also private practices, not a few of them run by doctors who hold positions in public facilities and so have access to potential clients. The better commercial terms of trade ensuing from the provision of an all-weather road should enhance villagers' capacity – and willingness – to pay for treatment, with some choice among facilities forming part of the available options.[1]

The point that the extension of the all-weather road network to village communities brings with it new hazards to health as well as new opportunities to get and stay well is hardly new. While there are numerous references in the literature to the fact that the provision of all-weather roads improves the rural population's access to health facilities, as well as the specific findings that trip-time and -cost hinder the uptake of formal health care (see, e.g., Wong *et al.*, 1987; Gertler and van der Gaag, 1990; Mwabu *et al.*, 1993), there appears to be nothing on whether the provision of such roads actually affects morbidity. This paper, in contrast, is concerned precisely with analysing that *outcome*, as measured by the duration of individuals' incapacitating bouts of illness, if any.[2]

This paper investigates PMGSY's effects on morbidity in upland Orissa, a remote and backward region widely known – if not actually infamous – for its poverty and low scores on other social indicators. The emphasis is on episodes of acute illness, especially those of the infectious kind, which account for the lion's share of the overall

[1]Klemik et al. (2009) investigate the effects of the quality of trip connections on the choice of facility in rural Tanzania. There appear to be no such studies for PMGSY.

[2]Banerjee and Sachdeva (2015) find that PMGSY has promoted preventive health care, but as they remark in their conclusions, the data were lacking to say anything about outcomes.

burden of morbidity. Here, climate and season play their part. The monsoon rains can make *kutcha* (dirt) tracks impassable, and they usher in malaria, other fevers and water-borne diseases. As the land dries out during winter and the blazing hot summer that follows, so these diseases retreat somewhat and the tracks become passable once more.

A salient feature of the analysis is the treatment of the village as a unit within the road network as a whole. For our purposes, it is the separate stretches on track, rural road and highway comprising the entire trip to the nearest hospital and PHC, respectively, that matter, not simply whether the village has an all-weather road or the total lengths of these journeys. Such precision in defining the trip variables is essential if the effects of the programme, if there are any, are to be estimated with any precision.

The data relate to 1580 individuals belonging to a panel of 279 households drawn from a spatially stratified sample of 30 villages. The investigation covers the *kharif* (monsoon) seasons of 2009 and 2013; 1076 individuals were present in both. Morbidity in *kharif* is much higher than in *rabi* (winter); moreover, data for *rabi* 2009 were not collected. Six of the 30 villages received a direct connection between 2004 and 2009, another four between 2009 and 2013. Five other villages experienced an indirect connection, in the form of a new road in the neighbourhood, whose completion reduced the stretch on a *kutcha* track.

In summary, the estimated (individual) random effects model yields the following reductions in morbidity when one km. of *kutcha* track is replaced by one km. of PMGSY road. First, the probability that an individual will fall sick at all is lower by 3.6 percentage points per km.; second, an individual's expected number of days of sickness is lower by 0.46 per km. These estimated effects on the two measures of morbidity should be viewed in the light of an average stretch so replaced of just over 3 km. in the recent, mature stage of PMGSY. The said average was about twice as long in the early phase.

The paper is structured as follows. The region, the surveys and the general nature of the data are briefly described in Section 2. Section 3 addresses the question of what influenced the pattern of PMGSY connections, as actually implemented up to the end of 2013. Section 4 gives a summary account of the main illnesses and their incidence, followed by definitions and brief discussions of the regressors that are employed in the analysis. Section 5 analyses the association between the trip variables to health facilities and the levels of individuals' morbidity in the *kharif* seasons of 2009 and 2013. In particular, estimates of PMGSY's impact are presented and discussed in Sections

5.2 and 5.4. The chief conclusions are drawn together in Section 6.

2 The Surveys and the Data

The study area comprises five, contiguous administrative blocks, four of them in Bolangir District and one in Kalahandi District. This remote region's chief topographical features are small river basins and often densely forested hills. Numerous tribal groups and Hindu communities make up its population. Its general poverty and proneness to drought have made it a byword in India's political discourse.

The original survey was carried out in 2001-02 with the object of investigating the variability of households' incomes and how they cope with drought (see van Dillen [2008] for an extensive analysis). To ensure a fairly even spatial distribution, six clusters of villages were defined in each block and one village was drawn at random from each cluster. Eight households were then selected from each village, randomly drawn from different parts of the village settlement area so as to include, with high probability, at least one household from each of its various communities. It should be noted that at the outset of the survey, PMGSY was in its infancy and neither the investigating team nor the villagers had any notion that it would be implemented in the region in the near future.

These households were re-surveyed in the early months of 2010 in connection with an evaluation of the socio-economic effects of PMGSY, which had then been under way in the area for a few years. Particular attention was paid to trade in goods, education and health (Bell and van Dillen, 2014). The period covered was the *kharif* season (July 1 to December 31) of 2009. The households' individual members and their salient characteristics in 2004-05, a complete list of which was available from the previous survey round, provided the basis for these newly canvassed data, especially in the domains of education and health. Of the original 240 households, 236 could be traced and re-surveyed, though births, deaths, marriage and individual migration had combined to alter the character of many families in the interim.

The next re-survey was conducted in two waves between October 2013 and March 2014, and covered the calendar year 2013.[3] Of the 1291 individuals comprising the sample in 2009, 1076 were also present in the 2013 round, with death, exogamous marriage and above all outmigration claiming the other 215. Arrivals in the form of

[3]The 2009 round was financed by the World Bank, the 2013 round by the Research Council of Norway.

births, marriage and return migration yielded 289 newly present individuals in 2013, for a total sample of 1365 in that round.

3 The Placement of PMGSY Connections

The pattern in which PMGSY connections extend the whole network of all-weather roads is influenced by diverse factors – formal eligibility requirements involving settlements' population size and remoteness, political pressures, administrative efficiency and engineering considerations. What ultimately matters for present purposes, however, is whether the pattern of connections was random in relation to villages' disease environments.

With this consideration in mind, the fact that all villages are located in a network calls for a remark before any examination of how direct connections were assigned among the survey villages. Villages that did not enjoy a direct connection to the main all-weather road network would still have benefited from the extension of that network through PMGSY if a new road happened to pass through their general neighbourhood. For a journey that hitherto began with a long stretch on a *kutcha* track could change into one with just a couple of kilometres of track to a nearby, newly constructed stretch of PMGSY road. Five of the survey villages benefited in this way between 2003 and 2009, two of them with populations under 250. The existence of such indirect connections strengthens the random element in the allocation of improvements in all-weather connections among villages, and mitigates, in part at least, any potential concern arising from the discussion that follows.

Under the rules, some villages were 'ineligible' for a direct connection, the lower limit on population in this region being 250 persons (with reference to the 2001 census). Nine of the 30 sample villages were thus disqualified, and none had received a direct connection at the time of the last survey. Another village happened to sit astride a PWD district road, which leaves 20 villages as candidates at the start of the program in 2000.

To the eye, the 10 villages that had received a PMGSY connection by 2013 form no obvious spatial pattern on the map.[4] A second, related possibility is that certain blocks were favoured administratively or politically. Table 1 sets out the position by block in 2009 and 2013. Since there are rather few villages, sampling fluctuation alone could be

[4] A map of the area, with the sample villages placed within it, is found in van Dillen (2008).

responsible for the outcome in which both the two eligible villages in Block 2 had hit the jackpot by 2009, but none of the four in Block 3. For this outcome must be seen against a background in which scores of villages in each of the five blocks were vying for connections. In fact, Fisher's exact test of the null that there was no association between blocks and connections in 2009 yields a borderline result ($p = 0.07$), though that for 2013 survives at conventional levels ($p = 0.12$).

Extending this approach further, it is desirable to check whether there are any differences between other observable, plausibly influential characteristics of 'connected' and 'unconnected' villages. An immediate possibility is that the more populous eligible villages had better chances: the more beneficiaries – and voters – the better. The politics of caste and the growing maoist insurgency may also influence such allocative decisions, which suggests that the proportion of a village's inhabitants belonging to the politically powerful group legally classified as 'other backward castes' (OBC), as opposed to scheduled tribes (ST) and scheduled castes (SC), may come into play. Getting a connection to the electricity grid involves similar political and administrative processes, so it is arguable that those villages that enjoyed such connections in 2001 would have been well-placed in the queue to get a PMGSY connection thereafter. A closely related factor is a village's distance from the *Panchayat's* HQ, which is the lowest level of government in India's federal system.

Summary statistics of these various characteristics in 2001 for the sample villages, classified by their PMGSY status in 2013, are given in Table 2. For the set of connections established up to 2009, the null is rejected at the 5% level ($F(1, 18) = 5.566$) when population is the sole criterion for separating the two groups. It is just rejected at that level ($F(2, 17) = 3.444$) when the village's caste composition is added; but the successive introduction of an existing electricity connection and the distance to the *Panchayat's* HQ yields a clear rejection of the null once more. The new connections provided after 2009 caused four villages to switch categories: the null is then rejected only at the 18% level when population is the sole criterion (see Hotelling's T^2 statistic, $F(1, 18) = 1.882$ in Table 2). Yet when the village caste indicator is added, the null is rejected at the 1% level. Adding the other two variables confirms this rejection, albeit at the 5% level. We conclude that in both periods, political factors influenced whether a village was favoured with a PMGSY road, even when the statutory lower limit of 250 persons is imposed.

Yet settlements dominated by OBCs might have much land under irrigation, and the local HQ, established long ago, might lie close to lowland now become swampy. Both would tend to be plagued by malaria and water-borne diseases. Village characteristics designed to address, and control for, such possibilities will be discussed and defined in Section 4.1.

To complete this account, and in keeping with the importance of viewing villages as units in a network, we define the trip to a medical facility for treatment as the vector of the lengths and quality of the various stretches of paths, cart tracks and roads that constitute it. The definition employed here involves two alternative trips, one to the nearest hospital, the other to the nearest PHC. This definition places the village within the route network, and so is unaffected by villagers' actual choices of facility. Thus, the trip to the nearest hospital is defined by the following elements:

- $h1\_d0$ denotes the stretch of *kutcha* track, in km.;
- $h1\_d1$ denotes the stretch of PMGSY or all-weather village road, in km.;
- $h1\_d2$ denotes the stretch of district (PWD) road, in km.;
- $h1\_d2c3$ denotes a stretch of PWD road in poor condition, in km.;
- $h1\_d3$ denotes the stretch of highway, in km.;
- $h1\_d2c3$ denotes a stretch of highway in poor condition, in km..

Likewise, the trip to the nearest PHC is defined by $h2\_d0, \ldots$. For brevity, these descriptions of the village's position will be termed the network regressors. It should be remarked that a direct PMGSY connection almost invariably involves replacing the stretch hf_d0 with a stretch hf_d1 ($f = 1, 2$) of almost the same length. As noted above, there are also indirect connections, which involve a partial reduction in hf_d0 in exchange for a positive value of hf_d1. Both kinds of improvements may induce households to choose a different health facility, and hence a different trip after the end of the stretch hf_d1, but this matter will not be pursued here.

4 Illnesses and Covariates

The chief ailments afflicting this population are infectious diseases of the acute kind, especially malaria, various other fevers and water-borne diseases, and then especially during the *kharif* season. Some individuals do, of course, suffer from chronic conditions, such as anaemia, tuberculosis, rheumatism and alcoholism; but relatively few respondents reported that family members did so (see below). How those who drink view their relationship with the bottle, and how they speak of it to others, is known, moreover, to be frequently coloured by denial and social convention. This certainly holds in upland Orissa, where a hangover is also unlikely to deter work in the fields or on a construction site. In what follows, we shall confine our attention to bouts of acute illness, including those caused by chronic conditions.

As formulated in the questionnaire, a bout of acute illness is defined to be one so severe as to have prevented the individual from working or attending school. As with all self-reported conditions, there is no stringently common standard here. The very poor may well be under a strong compulsion to work however miserable they feel, and children may be treated lightly, since official enforcement of attendance is unlikely to be strict. In the event, all but a handful of the individuals reportedly suffering such a bout of illness were also treated for it, if only by a traditional healer. It is therefore quite possible that investigators and respondents alike conflated incapacitating illness with the decision to have it treated; so that those who took ill but received no treatment were not deemed to be unfit for work or school, whatever their actual condition. With these reservations in mind, the measure of an individual's morbidity is defined to be the total number of days of acute sickness he or she suffered in the season in question, a measure that also comfortably accommodates those who were too young to go to school or too old to work, but still received treatment, as well as those who were reportedly not sick at all in that season.

As a preliminary, we begin with a brief account of the morbidity immediately preceding those few cases of death and of that associated with chronic diseases. There were just four deaths among the 1291 individuals comprising the sample in *kharif* 2009, two each due to malaria and tuberculosis or pneumonia. Fifty-two individuals suffered from chronic ailments, 27 of them from anaemia, tuberculosis or both, and the rest from conditions such as lameness, polio and eye problems. Of these 52, 18 reported at least one bout of acute illness in that season, a proportion much like that among all those reportedly free of chronic conditions; and the incidence of the associated acute diseases was also similar in the two groups. The said 52 individuals are included in the following analysis of acute illness, but the morbidity associated with their chronic conditions only if it satisfied the definition given above. The four individuals who died are not included. There were 10 deaths in *kharif* 2013, five of them attributed to old age, the rest to malaria, fevers and childbirth. Seventy-eight individuals suffered from chronic ailments, frequently rheumatism among the old.

The distributions of the chosen measure of morbidity in the *kharif* seasons of 2009 and 2013 are set out in Table 3, with villages classified by the period in which they received a direct PMGSY connection, if at all. In what follows, the subset of 20 villages that never received a direct connection will be denoted by S_0, that comprising the six that got one early (2004-2009) by S_1, and that comprising the four that got one late (2009-2013) by S_2.

The overall incidence of morbidity in the whole sample was much lower in 2013: only 26% of all individuals suffered any days of sickness, as opposed to 44% in 2009. Among those who did fall sick, the average duration of illness was the same, at 12.97 and 13.0 days in 2009 and 2013, respectively, albeit with very different s.d.'s of 6.8 and 12.7 days, respectively. The null hypothesis that the two overall distributions are drawn from the same population is decisively rejected, by both Pearson's chi-square and the Kruskal-Wallis tests. It is clear that this striking fall in morbidity cannot be laid largely at the door of the provision of just four PMGSY connections in the interim. There was much heavier rainfall in the 'good' monsoon of 2009, a fact that will be taken up below.

Turning to the differences in the distributions for villages with and without an all-weather road within a season, nothing striking meets the eye in 2009. There is a hint that morbidity was lower in the group S_2, the fraction of individuals reportedly free of sickness being several percentage points lower than in either of the other groups; but the null of independence is in no danger of rejection at conventional levels using either test. Inspection of the picture for 2013 reveals no obvious differences. The null is roughly a coin-toss using Kruskal-Wallis; that it is rejected at the 7% level using Pearson's chi-square can be viewed as yet another instance of a well-known tendency when the sample is large.

Despite the apparent absence of differences in the overall distributions within periods, there are notable differences in the changes in means over time. Although all three means were lower in 2013, the falls from 2009 were rather different: for S_0, S_1 and S_2, they were 2.313, 3.412 and 0.893 days, respectively. Regressing the difference in village mean levels on dummy variables for S_1 and S_2 yields the finding that the fall in S_1 villages was greater than that in the reference subset S_0 at the 4% level of significance with robust s.e.s, but only at the 23% level with conventional OLS s.e.s. The coefficient on the dummy for S_2 is not all significant with either kind. These differences will be pursued further in Section 5.4, which deals with village fixed-effect specifications.

Many individuals suffered more than one bout of acute illness, and some more than one ailment. In *kharif* 2009, 210 of the 567 individuals who fell sick had two such bouts and another 66 had three bouts. Of these 276 individuals who had at least two bouts, 89 suffered from a single ailment, principally malaria, followed by viral fevers. Thus, 177 individuals suffered from at least two different ailments in that season. A partial picture of the incidence of the various acute ailments in the two seasons is presented in Table 4, which gives an individual's main ailment, as defined by the level of morbidity it brought about. To this it should be added that the incidence of second and third ailments, if any, was quite similar, again with malaria and viral fevers the chief ones. Where the accuracy of these reported illnesses is concerned, nearly all individuals received medical treatment; even so, the household's respondent may not have been quite clear about the diagnosis at the time of interview. The categories 'viral fever' and 'influenza and colds', in particular, may well be elastically interchangeable.

Comparing 2009 and 2013 in aggregate, the lower incidence of overall morbidity in the latter is seen to arise from a sharp fall in the number of cases of malaria, viral fever and respiratory diseases. This fall is substantially offset by a rise in the number of cases in the residual category 'other', which includes accidents and toothache, especially in the group S_2. The null hypothesis that the incidence of the various ailments was the same in the two years is correspondingly decisively rejected. The null that, within each season, there was the same incidence of the said illnesses in all three categories of villages is just rejected at the 5% level for 2009, with a heavy incidence of malaria in group S_1; the rejection is comprehensive for 2013.

4.1 The regressors

The regressors are ordered into groups of characteristics at the individual, household and village levels, respectively. The associated summary statistics for both *kharif* seasons are presented in Table 5. These differ in some measure, due principally to the movements of individuals into and out of the sample, the splitting of households and changes in the rural road network over the period in question.

The trip variables defined in Section 3 are supplemented by an array of covariates as controls. Each member of a household is distinguished not only by his or her age and sex, but also by a particular blood- or marital relationship to the head of household. That relationship may determine how an individual is treated in the allocation of both the tasks to be performed and the family's consumption, especially where food and medical care are concerned, independently of the individual's age and sex. Let this relationship be denoted by r_{kl}, where k denotes a generational-cum-blood or marriage connection and $l\,(=1,2)$ the individual's sex. Thus,

- r_{11} denotes the head of household (almost invariably male) and r_{12} his spouse;
- r_{21} denotes a son, r_{22} a daughter;
- r_{31} denotes a grandson, r_{32} a granddaughter;
- r_{42} denotes the mother;
- r_{72} denotes a daughter-in-law;
- r_8 is a catch-all for all others, including (infrequently) fathers and brothers.

All of the above are dummy variables, with r_{11} as the reference case. Each category's proportion in the whole sample is the mean value reported in Table 5.

Turning to age, there are five groups: infants and toddlers, 0-4 years; school-age children, 5-15 years; young adults, 16-25 years; prime-age adults, 26-45 years; and old adults, over 45 years. These seem fine enough to capture age-specific morbidity, and they are not perfectly collinear with the four-year gap between the surveys. They are assigned dummy variables, with prime-age adults forming the reference group.

The household's productive endowments and its demographic structure generate both income and claims on the common pot. Particularly important are its owned holding (in acres) and the numbers of adult males and females of working age (15 to 65 years). The same arguably holds for the head of household's educational attainment (in years) and sex (male denoted by 0). The former normally influences the household's overall productivity, and it may well influence nutrition, hygiene and the choice of medical treatment – if any. The introduction of the head's sex allows for the possibility that the head is female, and then usually a widow, again with the possibility that this influences morbidity through the channels just described. Caste may also have an influence on outcomes, even controlling for all other factors; for beliefs about the sources of illness and how to treat them are arguably bound up with the family's wider view of the natural order, which may well be influenced by caste. Taking STs, whose beliefs have certain animistic elements, as the reference group, separate dummy variables are defined for SCs and OBCs.

The household's immediate environment is the village. Since infectious diseases account for the lion's share of morbidity, the village's total population is potentially in play, as is the proportion of its total area under forest and irrigation, respectively. Other physical features may also exert an influence. A village settlement lies not only at a specific altitude (in meters), but also in a specific topographical environment. Three categories are distinguished: dry basins (geo1), riverine basins (geo2), and hilly, forested areas (geo3). All three are dummy variables, with geo1 as the reference category. The administrative block in which the village is located also has particular characteristics, environmental as well as infrastructural, that may influence morbidity.

5 Morbidity

We now investigate the association between individuals' morbidity, the network regressors and village characteristics in the *kharif* seasons of 2009 and 2013.[5] It should be emphasised at the outset that PMGSY involves the substitution of an all-weather road for a stretch of track, so that it is not enough to examine the coefficients on the corresponding network regressors in isolation from one another. An exact formulation of the programme effect follows below. The associated estimates are set out and discussed in Section 5.2.

Let y_{ijt} denote the level of morbidity experienced by individual i in village j in period t. The form to be estimated is

$$y_{ijt} = \alpha \cdot \mathbf{x}_{ijt} + \beta \cdot \mathbf{z}_{jt} + u_i + v_j + w_t + \epsilon_{ijt}\,, \tag{1}$$

where $\mathbf{x}_{ijt}$ is a vector of i's characteristics at time t, $\mathbf{z}_{jt}$ is the corresponding vector for the village j in which i resides at t, the terms u_i and v_j represent unobservable, time-invariant heterogeneity among individuals and villages, respectively, w_t represents a time-varying common shock, and ϵ_{ijt} is a white noise term. It is assumed that w_t and ϵ_{ijt} are serially uncorrelated. Recalling the seasons recently surveyed, let $t = 1, 2$ denote 2009 and 2013, respectively.

The central policy question is, what is the effect on morbidity of replacing the whole of a village's *kutcha* track with a PMGSY road? As it turns out, the first two components of the trip to the nearest hospital or PHC are usually common to both facilities, that is, $(h1\_d0, h1\_d1) = (h2\_d0, h2\_d1)$. That being so, the effect of providing the

[5] A detailed analysis of the how morbidity is related to kinship, caste and other household variables is the subject of a separate paper. Discussion of these relationships here is very cursory.

road on (individual) morbidity is

$$\delta \cdot d0 \equiv (\beta_{11} - \beta_{10} + \beta_{21} - \beta_{20}) \cdot d0, \tag{2}$$

where $d0$ is the length of track replaced, β_{11} is the coefficient of $h1\_d1$, the other β's are analogously defined and δ is the net effect per km. so replaced. It should be noted that this expression also covers indirect connections, for which only part of the *kutcha* track is replaced.

5.1 The two periods combined

Three points should be recalled at the outset. First, morbidity was much higher in 2009 than 2013. Second, four villages changed the first stage of their connection to the main road network, from a *kutcha* track to a PMGSY road. Third, only 68% of all the individuals comprising the sample were present in both periods, which could pose an obstacle fully to exploiting their unobserved heterogeneity. Of these 1076 individuals, 483 were sick in 2009, but only 276 in 2013; 460 were sick in neither period, and 143 in both. Thus, 56% of them experienced the same state in both periods.

The lower level of overall morbidity in 2013 calls for a shift term ($s13\_02$) to supplement the constant in (1). It seems plausible that the coefficients on elements of the trip to the health facilities are also influenced by monsoonal conditions. For if the associated pathogenic environment is generally less hazardous, the efficacy of a good transport connection in suppressing morbidity should be lower. Indeed, in the limit, in which there are no pathogens at all, neither the road network nor the health facilities will have any effect on such morbidity. Since the key part of the connection is the first stretch of the whole trip, be it a *kutcha* track or an all-weather road, we therefore introduced interaction terms between $s13\_02$ and the elements involving $\_d0$ and $\_d1$, respectively. Thus, the effects of PMGSY substituting the one for the other would depend on variations in the monsoon. As it turned out, these additional regressors proved to be statistically quite insignificant and were therefore dropped.

For reasons of efficiency, with just four villages gaining a direct connection between 2009 and 2013, we choose the random-effects estimator, with unobserved individual heterogeneity.

As a first, exploratory step, we employ a linear probability model (LPM), with the measure of morbidity transformed into the discrete variable $\{0, 1\}$ (not-sick/sick) as regressand. Although the said transformation discards information, it may well reduce measurement errors relative to morbidity defined in days, for which some degree of digital preference is apparent in respondents' answers. The LPM usually provides good estimates of the partial effects of changes in the regressors near the center of their distribution (Wooldridge, 2002: 455).

Starting with the LPM (see Table 6), the component for unobserved individual heterogeneity accounts for just 4.02% of the residual variance. The only trip variable that is significant at the 5% level is that involving a stretch of highway in bad condition to the nearest PHC. The season dummy is highly significant: *cet. par.*, the estimated probability of falling sick in 2013 was 0.185 lower than in 2009. The only village characteristic that is significant at the 5% level is pop01: an extra 100 persons yields an estimated increase of 0.04 in the probability that any resident will fall sick. The effect of an individual's relationship to the head of household shows a definite pattern. Relative to that person and controlling for age, spouses are more likely to fall sick, others less so, and the associated coefficients for sons, daughters, grandsons, mothers and others are significant at the 10% level. More females in the household is also associated with a statistically significant reduction in individual morbidity.

Turning to the tobit specification, the component for unobserved individual heterogeneity accounts for just 4.4% of the total residual variance. The LR test cannot reject the alternative of simply pooling the individual observations ($p = 0.201$), so that such pooling of the data would also be defensible. The slope parameter associated with the seasonal shift term $s13\_02$ has a point estimate of 8.26 days fewer than its counterpart in 2009, and it is statistically highly significant.

The network regressors $h1\_d0$ and $h1\_d1$ have coefficients of opposite sign, but neither is significant even at the 10% level. The stretch to the nearest PHC, $h2\_d0$, is significant at the 1% level, but $h2\_d1$, also of opposite sign, is not at all significant.[6] As in the LPM model, a trip to the nearest PHC along a stretch of highway in poor condition, $h2\_d3c3$, is significant, now at the 1% level.

A discussion of these findings is in order. For the great majority of villages, the nearest hospital is farther off than the nearest PHC; but it is also almost certain to be staffed the whole time, which cannot be said of the PHC. Those who are gravely sick, or simply able to afford it, are therefore more likely to seek treatment in the hospital. Faced with a long stretch on a *kutcha* track in order to reach the main road network

[6] The two trips often have a stretch of *kutcha* track in common, and the correlation coefficient between $h1\_d0$ and $h2\_d0$ is correspondingly quite high, at 0.71.

and thereafter a hospital, those who are not so very ill may well choose to visit the nearest PHC instead. A similar stretch on a PMGSY road, however, is easily covered under all conditions. In this reduced-form specification, the findings may indicate that individuals in graver condition were taken to hospitals, which are usually more distant, whereas those in somewhat better shape were taken to PHCs. The negative, and statistically highly significant, coefficient on the PHC-highway stretch in poor condition is consistent with this interpretation.

Among the village characteristics, pop01 and forest are significant at the 5% level, altitude at the 10% level, all three with the expected signs. Living in Block 2 is statistically less hazardous to health. The findings concerning the relationship and household variables are qualitatively the same as those yielded by the LPM, but most are rather more precisely estimated, notably those for spouses and the number of females.

To probe more deeply, consider a Heckman-type 'hurdle model', in which the first-tier equation is to be interpreted as modelling the probability of falling sick at all. The identifying restrictions involve the village's total population, the proportions of its area under forest and irrigation, and the environmental variables, all of which should matter at the first stage, but none of which is arguably likely to affect the number of days of sickness, conditional on falling sick. (Recall that just over one half of all individuals who did so had but one bout of illness.)

Beginning with the two-step specification, the seasonal shift term enters very strongly in the first stage, where much of the action is concentrated (see Table 7). The trip variables $h2\_d0$ and $h2\_d3c3$ are significant at the 5% and 1% levels, respectively, pop01 and forest at the 1% and 5% levels, respectively. In the second stage, the trip variables $h1\_d0$ and $h2\_d0$ are significant at the 10% and 5% levels, respectively, with opposite signs. The inverse Mills ratio is not at all significant.

Various attempts to obtain convergence with maximum likelihood specifications were unsuccessful.

5.2 PMGSY's effects

The next step is to answer the central policy question, as set out formally in eq.(2), namely, how large is δ and is it statistically significant? The values of δ and their associated standard errors for all periods and specifications are reported in Table 8. In the single-equation specifications, the null hypothesis of a zero net effect is rejected at the 10% level and 5% levels in the LPM and tobit models, respectively. In the hurdle model, the null is clearly rejected at the hurdle stage, but not at the second (duration) stage.

As for the estimated size of PMGSY's effects, the LPM estimate of δ implies a reduction in the probability of falling sick of 0.0357 per km. The tobit estimate is 1.322 days; adjusting the latter by the fraction of non-zero observations yields $1.322 \times (919/2656) = 0.457$ days per km. The average length of *kutcha* track facing the inhabitants of the 10 qualified villages still awaiting a PMGSY connection in 2013 was 3.1 km. Providing them with such a road would, for each one of them, reduce the probability of falling sick by 0.111 (LPM) and the expected number of days of sickness by 1.42 (tobit). The two-step hurdle model yields a virtually identical estimate of the former when allowance is made for the proportion of non-zero observations at the hurdle stage. The estimated reduction in the expected duration of sickness is somewhat more modest, at 0.92 days.

It might well be asked whether the same findings would emerge from a more parsimonious specification of the trips to the health facilities. A natural alternative is to replace the whole set of network regressors with a simple indicator variable for the existence of a direct all-weather connection. This does indeed yield point estimates for both the LPM and tobit models that are similar to those reported above, but with a decisive drawback: the associated standard errors are so large that the null of no effect on morbidity is in no danger of rejection, even at the 35 percent level. Using the indicator variable as a short-cut would therefore lead to a serious error in drawing inferences about the programme's effects. This finding has potentially important implications for investigations of the effects of extending the all-weather road network in other semi-arid regions.

5.3 Heterogeneity across periods

When combining the two periods in the preceding subsections, a dummy was used to capture the large difference in their prevailing levels of morbidity in the aggregate. The other notable difference between the two periods, though evidently much more modest in scale, was that another four villages got a PMGSY connection. Now, it is possible that the program's effects varied over time in a more complicated way: in particular, six villages had already gained a direct connection between 2004 and 2009, and another five had benefited indirectly. To pursue this possibility, we give a summary account of the results yielded by separate regressions for each period, confining the discussion to the network regressors, which are of central interest (see Table 9).[7]

Beginning with the LPM estimates, each additional km. along a PMGSY road to the nearest PHC was associated with a 0.0615 lower probability of falling sick in 2009 and 0.0475 in 2013, which are significant at the 5% and 1% levels, respectively. The corresponding estimates from the tobit specification are 2.017 and 2.404 days, which are significant at the 5 and 10% levels, respectively. Only the estimates for the first tier of the hurdle model are significant, and then only for 2013. It is noteworthy that all of the above are larger, in absolute magnitude, than those obtained from combining the two periods.

5.4 Village fixed effects

A salient feature of the specifications employed thus far is that unobserved village fixed effects have been ignored: implicitly, the corresponding term for unobserved heterogeneity, v_j in eq.(1), has been combined with ϵ_{ijt}. Although a whole battery of village characteristics has been employed – to good statistical effect – in the above sub-sections, there still remains a potential omitted variable problem. It is therefore desirable to investigate whether a fixed-effects estimator produces different results from those obtained above.

Two preliminary remarks are called for. First, there is no fixed-effects estimator for tobit or probit, so that the existence of a mass point of observations at zero emerges as a countervailing drawback when running OLS regressions on either measure of morbidity with village fixed effects. Giving up specifications that respect the fact that the measures have natural restricted ranges comes at a cost. Second, the specific data at hand involve just four new connections between 2009 and 2013, so that any findings for that period relate to those recent connections, but not to earlier ones.

Recalling the classification of villages by the period of their direct all-weather connection, if any, let $d0_{jt}$ denote the distance along a track from j to an all-weather road at time t $(= 1, 2)$. Then, ignoring indirect connections, the stretches of *kutcha* track to the all-weather road network are

$$d0_{j1} = d0_{j2} > 0 \; \forall j \in S_0; \; d0_{j1} = d0_{j2} = 0 \; \forall j \in S_1; \; d0_{j1} > 0, \; d0_{j2} = 0 \; \forall j \in S_2.$$

[7] A full set of results is available from the authors upon request.

Forming first-differences using eq. (1) and ignoring any changes in individuals' observable characteristics, we have

$$\Delta y_{ijt} = \Delta w_t + \Delta\epsilon_{ijt},\ \forall j \in S_0, S_1;\ \Delta y_{ijt} = \beta_1 d0_{j1} + \Delta w_t + \Delta\epsilon_{ijt},\ \forall j \in S_2.$$

This procedure has the advantage of ridding the data of both individual and village fixed effects. By assumption, $E\Delta\epsilon_{ijt} = 0$, so that this difference-in-difference scheme yields estimates of the common shock relative to 2009, Δw_t, and the response per km. of a new PMGSY connection, β_1, from the observations in S_0 and S_1, and S_2, respectively. This is just the comparison of changes in mean levels of morbidity set out and briefly discussed in Section 4.

As for the other covariates, there were some changes in some individuals' characteristics between 2009 and 2013, but these proved to be rather minor. There were more extensive changes in the family variables involving the head of household and the numbers of adult males and females, so these have been included as controls. It turns out that, for the four villages comprising S_2, the differences in the trip variables involving $d0$ and $d1$ are perfectly correlated; for the first stage of the trip from the village to either health facility was the same, and the new connection resulted in $hf\_d1\,(f = 1, 2)$ in 2013 being the same as $hf\_d0\,(f = 1, 2)$ in 2009. Hence, only one of the four differences in each could be employed as a regressor. Table 10 is correspondingly parsimonious.

After forming first differences, the regressand for the LPM takes one of the values $\{-1, 0, 1\}$. The common shock relative to 2009 is very precisely estimated, but the effects of replacing the *kutcha* track with an all-weather road, $\Delta h1\_d1_t\,(= -\Delta h1\_d0_t)$, not at all so. The regressand for the duration measure (in days) ranges from about -100 to $+100$, with a heavy mass point at zero (460 out of 1076 observations). The coefficient on the 'program variable' $\Delta h1\_d1_t$ is likewise insignificant at the 10% level.

An alternative procedure is to use the whole sample and estimate in levels employing the OLS within-estimator, whereby the problem of multicollinearity among the trip variables arises once more. The results for the LPM are qualitatively the same (see Table 10), despite the presence of unobserved individual heterogeneity and different restrictions on the set of values taken by the respective regressands. When the duration measure is employed, however, the 'program variable' $h1\_d1_t$ is significant at the 10% level when the controls are introduced. This finding of a *positive* association between morbidity and new connections between 2009 and 2013 stands in contrast to the results obtained for all connections in Section 5.2 . The fact that the four villages in S_2 had rather low levels of morbidity in 2009 may be at work. More generally, OLS has evident shortcomings as a description of the pattern in these data.

It is not, perhaps, very surprising that these findings are rather ‘fuzzy’ in comparison with those in the preceding sub-sections. The sample of villages is rather small, with only four ‘treated’ during the second sub-period. One set of specification problems is exchanged for another, with attendant problems in drawing inferences. There is also the attenuation that creeps in when differences are used in any part of the calculations, including those for the within-estimator, since errors in level-variables are then magnified (Angrist and Pischke, 2009).

6 Conclusions

The setting is one in which the population is poor and ill-educated, and the overall burden of disease arises chiefly from the infectious kinds, especially in the monsoon season. The provision of all-weather roads should, therefore, reduce morbidity and mortality, provided the sick receive moderately competent treatment upon arrival at a PHC or hospital. Defining morbidity as relating to bouts of illness so acute, even if stemming from a chronic condition, that the sufferer cannot work or go to school, the estimates of PMGSY’s net effects in this region of upland Orissa can be summarised by two measures. First, the probability that an individual will fall sick at all is lower by 0.036 per km of road. Second, an individual’s expected number of days of sickness is lower by 0.46 per km. Recent direct connections have been about 3 km. long on average, which implies that the inhabitants of such villages have enjoyed an 11 percentage point reduction in the probability of falling sick and each of them, on average, 1.4 fewer days of sickness. It must be conceded that these are panel estimates with individual random effects, in which some residual village fixed effects may be at work. The associated fixed-effects estimates are much less precise than, and somewhat at odds with, those just set out. It has been argued that the fixed-effects estimates are subject to substantial reservations. Even so, they call for a certain caution in accepting results based on the assumed absence of unobserved village heterogeneity, despite the deployment of an extensive battery of village-level characteristics.

The mechanisms underlying these outcomes can be only partly uncovered using the survey data. The results from the hurdle models indicate that most of the effect is attributable to a reduction in the probability of falling sick rather than the duration of the bout conditional on falling sick, which is open to the interpretation that all-weather roads have a preventive effect, though exactly how is unclear.

To the extent that the above estimates of PMGSY's quantitative effects on morbidity are regarded as reliable, it is natural to ask whether they might hold more widely. Upland Orissa is but one part of a vast semi-arid tract in India's interior, whose inhabitants share not only its climate, topography and the same sort of economic activity, but also the chief diseases that flourish in such conditions. The health system to which they must resort for treatment and prevention is also broadly much the same. It is tempting, perhaps, to assume that rural road-building would yield similar quantitative effects in semi-arid tracts elsewhere, in sub-Saharan Africa, for example; but that would involve a big leap of faith.

To close, it should be emphasised that the PMGSY roads appearing in the study were all quite new, so that their longer term effects on morbidity may deviate from the estimates obtained here. More frequent visits for preventive care, such as immunisation, is one hope; better nutrition from higher incomes is another; and over the long haul, there is the prospect that girls will reach motherhood better educated. Then again, getting to town easily, especially with more money in one's pocket, holds clear dangers of contracting various diseases; and an extra bottle or two of liquor can be bought locally and consumed at home. Investigating the full effects of providing poor rural communities with all-weather roads will require a succession of refined surveys and much patience.

References

[1] Angrist, J.D., and Pischke, J.-S. (2009), *Mostly Harmless Econometrics*, Princeton, NJ.: Princeton University Press.

[2] Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010), 'Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60', *Social Science History*, 34: 171-197.

[3] Banerjee, R., and Sachdeva, A. (2015), 'Pathways to Preventive Health, Evidence from India's Rural Road Program', USC Dornsife Institute for New Economic Thinking, Working Paper No. 15-19.

[4] Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2013), 'Roads, Railroads and Decentralization of Chinese Cities', mimeo, Brown University.

[5] Bell, C., and van Dillen, S. (2014), 'How Does India's Rural Roads Program Affect the Grassroots? Findings from a Survey in Upland Orissa', *Land Economics*, 90(2): 372-394.

[6] Chaudhury, N., Hammer J., Kremer, M., Muralidharan, K., and Rogers, F.H. (2006), 'Missing in Action: Teacher and Health Worker Absence in Developing Countries', *Journal of Economic Perspectives*, 2(1): 91-116.

[7] Donaldson, D. (2013), 'Railroads of the Raj: Estimating the Effect of Transportation Infrastructure', *American Economic Review*, (forthcoming).

[8] Duranton, G., Morrow, P. M., and Turner, M. A. (2014), 'Roads and Trade: Evidence from the U.S.', *Review of Economic Studies*, 81(2): 684-724.

[9] Gertler,P., and van der Gaag, J. (1990), *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries*, Baltimore, MD: Johns Hopkins University Press.

[10] Klemik, H., Leonard, K.L., and Masatu, M.C. (2009), 'Defining Access to Health Care: Evidence on the Importance of Quality and Distance in Rural Tanzania', *American Journal of Agricultural Economics*, 91(2): 347-358.

[11] Jedwab, R., Kerby, E., and Moradi, A. (2013), 'Out of Africa: Colonial Railroads, White Settlement and Path Dependence in Kenya', mimeo, George Washington University.

[12] Muralidharana, K., Chaudhury, N., Hammer, J., Kremer, M., and Rogers, F.H. (2010), 'Is there a Doctor in the House? Medical Worker Absence in India', mimeo, University of California, San Diego.

[13] Mwabu, G., Ainsworth, M., and Nyamete, A. (1993). 'Quality of Medical Care and Choice of Medical Treatment in Kenya: An Empirical Analysis', *Journal of Human Resources* 28(4): 838-862.

[14] van Dillen, S. (2008), 'Income and its Variability in a Drought-prone Region: Seasonality, Location and Household Characteristics', *European Journal of Development Research*, 20(4): 579-596. DOI: 10.1080/09578810802464912.

[15] Wong, E., Popkin, B., Guilkey, D., and Akin, J. (1987), 'Accessibility, Quality of Care and Prenatal Care Use in the Philippines', *Social Science and Medicine*, 24(11): 927-944.

[16] Wooldridge, M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

[17] World Bank (2010), 'Project Appraisal Document', Report No. 57081-IN, Washington, DC.

Table 1: Eligible village connections by block, *kharif* 2009 and 2013

		2009[a]			2013[b]	
Connection	PMGSY	*kutcha*	total	PMGSY	*kutcha*	total
Block 1	2	1	3	3	0	3
Block 2	2	0	2	2	0	2
Block 3	0	4	4	1	3	4
Block 4	1	5	6	3	3	6
Block 5	1	4	5	1	4	5
Total	6	14	20	10	10	20

[a] Fisher's exact test: $p = 0.07$. [b] Fisher's exact test: $p = 0.12$.

Table 2: Eligible villages' characteristics in 2001, by connection in 2013

PMGSY connection in 2013 ($n = 10$)

Characteristic	mean	s.d.	min.	max.
Population	667.0	227.6	407	1060
OBC (percent)	59.4	21.9	22.1	80.8
Electricity	0.7	0.843	0	1
Panchayat HQ (km.)	2.3	1.23	0	4.0

kutcha connection in 2013 ($n = 10$)

Characteristic	mean	s.d.	min.	max.
Population	529.0	222.9	274	854
OBC (percent)	30.8	26.5	0.7	76.0
Electricity	0.6	0.516	0	1
Panchayat HQ (km.)	3.65	1.63	1.5	7.0

Hotelling's T^2. Population: $F(1, 18) = 1.882, p = 0.187$.

Population, OBC: $F(2, 17) = 6.480, p = 0.008$.

Population, OBC, Electricity: $F(3, 16) = 4.094, p = 0.0247$.

Population, OBC, Electricity, Panchayat HQ: $F(4, 15) = 3.471, p = 0.0338$.

Table 3: Individuals' days of sickness, classified by village's date of connection

kharif 2009 ($n = 1291$)

No. of days	0	1-5	6-10	11-15	16+	total
Never	483	55	104	127	88	857
	(56.4)	(6.4)	(12.1)	(14.8)	(10.2)	(100)
2004-2009	133	25	31	38	29	256
	(52.0)	(9.8)	(12.1)	(14.8)	(11.3)	(100)
2010-2013	108	9	24	29	8	178
	(60.7)	(5.1)	(13.5)	(16.3)	(4.5)	(100)
Total[a]	724	89	159	194	125	1291
	(56.1)	(6.9)	(12.3)	(15.0)	(9.7)	(100)

$\chi^2(8) = 12.092, p = 0.147$. Kruskal-Wallis: $\chi^2(2) = 3.868$ with ties, $p = 0.145$.

kharif 2013 ($n = 1365$)

No. of days	0	1-5	6-10	11-15	16+	total
Never	656	46	107	40	46	895
	(73.3)	(5.1)	(12.0)	(4.5)	(5.1)	(100)
2004-2009	212	14	35	5	14	280
	(75.7)	(5.0)	(12.5)	(1.8)	(5.0)	(100)
2010-2013	145	10	11	14	10	190
	(76.3)	(5.3)	(5.8)	(7.4)	(5.3)	(100)
Total[a]	1013	70	153	59	70	1365
	(74.2)	(5.1)	(11.2)	(4.3)	(5.1)	(100)

$\chi^2(8) = 14.463, p = 0.070$. Kruskal-Wallis: $\chi^2(2) = 1.022$ with ties, $p = 0.600$.

[a] $\chi^2(4) = 136.06, p = 0.000$. Kruskal-Wallis: $\chi^2(1) = 110.37$ with ties, $p = 0.000$.

Row percentages in parentheses.

Table 4: Individuals' chief acute diseases, *kharif* 2009 and 2013

kharif 2009

Disease	viral fever	gastro-enter.	'flu/cold	malaria	OBGYN	other	total
Never	125	13	44	145	16	31	374
	(33.4)	(3.5)	(11.8)	(38.8)	(4.3)	(8.3)	(100)
2004-2009	27	8	14	65	4	5	123
	(22.0)	(6.5)	(11.4)	(52.8)	(3.3)	(4.1)	(100)
2010-2013	26	2	8	25	6	3	70
	(37.1)	(2.9)	(11.4)	(35.7)	(8.6)	(4.3)	(100)
Total[a]	178	23	66	235	26	39	567
	(31.4)	(4.1)	(11.6)	(41.4)	(4.6)	(6.9)	(100)

Row percentages in parentheses. $\chi^2(10) = 18.329,\ p = 0.050$.

kharif 2013

Disease	viral fever	gastro-enter.	'flu/cold	malaria	OBGYN	other	total
Never	58	13	32	89	8	39	239
	(24.3)	(5.4)	(13.4)	(37.2)	(3.3)	(16.3)	(100)
2004-2009	13	8	8	26	3	10	68
	(19.1)	(11.8)	(11.8)	(38.2)	(4.4)	(14.7)	(100)
2010-2013	7	8	4	7	1	18	45
	(15.6)	(17.8)	(8.9)	(15.6)	(2.2)	(40.0)	(100)
Total[a]	78	29	44	122	12	67	352
	(22.2)	(8.2)	(12.5)	(34.7)	(3.4)	(19.0)	(100)

$\chi^2(10) = 28.39,\ p = 0.002$. [a] $\chi^2(10) = 44.62,\ p = 0.000$ Fisher's exact test, $p = 0.000$.

Viral fevers mostly respiratory, gastric ailments overwhelming infectious, in nature.

Table 5: Regressors, summary statistics

	2009 ($n = 1291$)				2013 ($n = 1365$)			
variable	mean	s.d.	min.	max.	mean	s.d.	min.	max.
r_{12}	.1642	.3706	0	1	.1897	.3922	0	1
r_{21}	.2463	.4310	0	1	.2381	.4261	0	1
r_{22}	.1565	.3634	0	1	.1758	.3808	0	1
r_{31}	.0627	.2426	0	1	.0432	.2034	0	1
r_{32}	.0480	.2139	0	1	.0366	.1879	0	1
r_{42}	.0356	.1854	0	1	.0315	.1747	0	1
r_{72}	.0565	.2311	0	1	.0469	.2115	0	1
r_8	.0232	.1507	0	1	.0212	.1443	0	1
age0-4	.0976	.2969	0	1	.0777	.2677	0	1
age5-15	.2417	.4283	0	1	.2410	.4279	0	1
age16-25	.1882	.3910	0	1	.1941	.3957	0	1
age46+	.2200	.4143	0	1	.2505	.4335	0	1
ownhold	3.137	3.826	0	30.0	2.210	2.249	0	11.83
hhedu	3.421	3.311	0	14	3.575	3.371	0	18
hhsex	.0418	.2003	0	1	.0381	.1915	0	1
males	1.945	1.066	0	5	1.766	1.007	0	5
females	1.820	.828	0	4	1.733	.889	0	5
children	2.125	1.365	0	6	1.888	1.429	0	7
SC	.2115	.4085	0	1	.1802	.3845	0	1
OBC	.3656	.4818	0	1	.3626	.4809	0	1
pop01	445.0	259.7	61	1060	442.8	259.2	61	1060
forest	7.26	12.30	0	60.23	7.57	13.06	0	60.23
elevation	225.7	34.3	164	303	225.4	34.1	164	303
geo2	.1464	.3536	0	1	.1458	.3530	0	1
geo3	.2974	.4573	0	1	.3048	.4605	0	1
block2	.1875	.3904	0	1	.1993	.3996	0	1
block3	.2014	.4012	0	1	.1846	.3881	0	1
block4	.1998	.4000	0	1	.1985	.3990	0	1
block5	.2130	.4096	0	1	.2212	.4152	0	1
$h1\_d0$	2.335	2.022	0	8	2.090	2.169	0	8
$h1\_d1$	2.306	3.884	0	13	2.682	4.007	0	13
$h1\_d2$	5.439	9.406	0	33	5.321	9.287	0	33
$h1\_d2c3$	.697	2.781	0	17	.610	2.588	0	17
$h1\_d3$	8.264	8.587	0	28	8.067	8.470	0	28
$h1\_d3c3$	1.649	4.635	0	21	1.664	4.733	0	21
$h2\_d0$	2.359	1.810	0	8	2.087	1.955	0	8
$h2\_d1$	1.957	3.532	0	13	2.322	3.679	0	13
$h2\_d2$	2.404	3.608	0	12	2.381	3.622	0	12
$h2\_d2c3$	.757	2.392	0	12	.690	2.282	0	12
$h2\_d3$	1.302	2.288	0	8	1.304	2.290	0	8
$h2\_d3c3$	.513	1.752	0	8	.488	1.688	0	8

Table 6: Morbidity 2009 and 2013: panel estimates

	LPM [a]		Tobit [b]	
Variable	coeff.	s.e.	coeff.	s.e.
r_{12}	.0512	.0345	3.916***	1.330
r_{21}	−.0706*	.0418	−3.728*	1.965
r_{22}	−.0895*	.0515	−4.701**	2.209
r_{31}	−.1021*	.0590	−5.080*	2.937
r_{32}	−.1088	.0701	−4.998	3.201
r_{42}	−.0891*	.0493	−4.172*	2.538
r_{72}	−.0556	.0547	−.099	2.483
r_8	−.1454**	.0620	−5.319	3.390
hhedu	.0007	.0031	.053	.146
hhsex	.1105*	.0660	3.599	2.239
males	−.0148	.0099	−.705	.511
females	−.0230**	.0111	−1.686***	.585
pop01/100	.0340**	.0166	1.269**	.543
forest	.0033	.0023	.1513**	.0714
irrigation	.0001	.0035	−.0304	.1135
alt/100	−.0707	.0706	−4.704*	2.665
geo2	.0007	.0569	2.343	2.374
geo3	.0011	.0613	.827	1.787
block2	−.1196	.0742	−6.123**	2.788
block3	.0003	.0631	−2.621	2.173
block4	−.0209	.0834	−-.787	3.360
block5	−.0383	.0567	−1.594	1.951
$h1\_d0$	−.0035	.0128	−.615	.590
$h1\_d1$	.0071	.0120	.529	.427
$h1\_d2$	.0001	.0038	.116	.125
$h1\_d2c3$	−.0022	.0126	−.414	.591
$h1\_d3$	.0025	.0030	.134	.128
$h1\_d3c3$	.0046	.0034	.219	.166
$h2\_d0$	.0314	.0219	1.868***	.688
$h2\_d1$	−.0149	.0113	−.598	.487
$h2\_d2$	−.0037	.0064	−.326	.240
$h2\_d2c3$	−.0057	.0152	.340	.728
$h2\_d3$	.0027	.0127	.499	.529
$h2\_d3c3$	−.0352**	.0174	−2.031***	.740
$s13\_2$	−.1849***	.0254	−8.261***	.892
cons	.5286***	.1564	4.160	6.573
σ	.4466		17.360	.599

30 groups (villages). $n = 2656$, 919 uncensored obs.
Other controls: age0-4, age5-15, age16-25, age46+, ownhold, children, SC, OBC.
[a] Discrete variable $\{0, 1\}$. R^2 (overall) = 0.0823. Robust s.e.'s, clustering for villages. Individual random effects, $\rho = 0.0403$.
[b] Wald $\chi^2(43) = 203.62$, log-likelihood = −4825.0. Individual random effects, $\rho = 0.0439$.

Table 7: Morbidity: hurdle model, two-step estimates, *kharif* 2009 and 2013 combined

	first tier		second tier	
Variable	coeff.	s.e.	coeff.	s.e.
$h1\_d0$	$-.0126$	.0353	$-.679$	.378
$h1\_d1$	.0215	.0262	.154	.262
$h1\_d2$	.0026	.0066	.071	.062
$h1\_d2c3$	$-.0183$	.0174	.090	.145
$h1\_d3$	.0085	.0078	$-.044$	.065
$h1\_d3c3$	.0143	.0096	.047	.111
$h2\_d0$	.0960	.0409	.940	.420
$h2\_d1$	−-.0401	.0297	$-.191$	.258
$h2\_d2$	$-.0127$	.0135	$-.208$	.143
$h2\_d3$	.0151	.0310	.152	.329
$h2\_d3c3$	$-.1123$	.0396	$-.152$	.389
$s13\_2$	$-.5313$	.0545	$-.690$	1.822
pop01/100	.0996	.0308		
forest	.0096	.0043		
irrigation	.0002	.0064		
alt/100	$-.2083$	.1600		
geo2	.0320	.1292		
geo3	.0132	.1076		
cons	$-.1076$	.3960	13.658	3.304
σ	9.251			

30 groups. $n = 2656$; 919 uncensored obs. $\rho = 0.213$. Wald: $\chi^2(36) = 76.36$.
Controls: $r_{12}, \ldots, r_8$, age0-4, age5-15, age16-25, age46+, ownhold, hhedu, hhsex, males, females, children, SC, OBC, block2, ..., block5.

Table 8: The effect of PMGSY on morbidity

Model	discrete [a]		duration [b]	
	δ	s.e.	δ	s.e.
LPM	$-.0357^{*}$	.0214	–	–
Tobit	–	–	-1.322^{**}	.669
Heckman two-step	$-.1020^{**}$	.0404	$-.298$	.231
Heckman ML [c]	–	–	–	–

The null hypothesis is $H_0 : \delta \equiv \beta_{11} - \beta_{10} + \beta_{21} - \beta_{20} = 0$.
[a] Probability of falling sick. [b] In days. [c] Convergence not obtained.

Table 9: The effect of PMGSY on morbidity, cross-section estimates

Model	discrete [a]		duration [b]	
	δ	s.e.	δ	s.e.
2009				
LPM	$-.0615^{**}$	.0235	–	–
Tobit	–	–	-2.017^{**}	.845
Heckman two-step	$-.1680^{***}$	.0637	$-.271$	.223
Heckman ML	$-.0502$	.0434	$-.199$	.138
2013				
LPM	$-.0475^{***}$	.0139	–	–
Tobit	–	–	-2.404^{*}	1.435
Heckman two-step	$-.1330^{*}$	.0709	$-.248$	.507
Heckman ML	$-.1209^{***}$	.0440	$-.373$	.280

The null hypothesis is $H_0 : \delta \equiv \beta_{11} - \beta_{10} + \beta_{21} - \beta_{20} = 0$.
[a] Probability of falling sick. [b] In days.

Table 10: The effect of PMGSY on morbidity: village fixed effects

Model	first-diff. [a]				level [b]			
	discrete		duration		discrete [c]		duration [d]	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
$h1\_d1_t$	0.0321	0.0398	0.869	0.557	0.0264	0.0285	0.876^{*}	0.398
$s13\_2$	–	–	–	–	-0.1864^{***}	0.0269	-2.525^{***}	0.476
const.	-0.2054^{***}	0.0252	-2.700^{***}	0.526	0.4683^{***}	0.0798	5.664^{***}	1.463

Standard errors clustered on villages.
[a] $n = 1076$ individuals. Regressand: $\Delta y_t \equiv y_{2013} - y_{2009}$.
Regressor: $\Delta h1\_d1_t \equiv h1\_d1_t - h1\_d1_{t-1}$. Controls: S_1, hhedu, hhsex, males, females.
LPM: $F(7, 29) = 1.58$. Duration: $F(7, 29) = 2.31^{*}$.
[b] $n = 2656$ observations, 30 groups.
[c] R^2: within = 0.0473, between = 0.0019. $F(8, 29) = 15.45$.
[d] R^2: within = 0.0298, between = 0.0006. $F(8, 29) = 25.00$.