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: April, 2018.

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 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. By 2013, ten villages had received a direct all-weather road connection since the inception of PMGSY in 2004. Treating the village as a unit within the whole network of roads and medical facilities, the provision of a connection, whether direct or in the neighbourhood, is estimated (random effects) to have reduced an inhabitant's probability of falling sick by 4.3 percentage points, and the expected duration of incapacitating illness by 0.54 days, for each km of unpaved track so replaced. The fixed-effects estimates are qualitatively the same, but less precise. A simple indicator variable for the presence or absence of such a connection yields qualitatively similar estimates, but with very large standard errors, which confirms the importance of employing fine measures of the network regressors.

Keywords: Rural roads, morbidity, India, PMGSY

JEL Classification: H54, I10, I15, R41

1 Introduction

The relationship between economic activity and the expansion of transportation networks 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, 2010; Duranton et al., 2014; Jedwab et al., 2013). Few, however, investigate how improved transportation affects morbidity and mortality. Zimran (2017) concludes that the development of the transportation network in the rural regions of the antebellum United States had a marked adverse effect on average height. Burgess and Donaldson (2017) find that mortality became less responsive to rainfall as the railways spread out in colonial India. These valuable historical investigations notwithstanding, the lack of contemporary evidence is rather troubling in the light of the revived emphasis on infrastructure to promote development in poor countries. For their health systems and provision of social protection, for all their failings, are undoubtedly much superior to those of a century and more ago.

One important component of such investment programmes is the provision of all-weather rural roads. A prime example 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. About 170,000 habitations were eligible; about 75% now 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 programme 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.

There is an interplay between the first and the third spheres 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, and since towns and markets pool both people and various communicable diseases, villagers become more exposed. Yet a decision to seek treatment for an ailment, however contracted, involves more than logistics; for the commercial sphere is now in 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 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.[1]

The point that extending 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. 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). Yet 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*,

---

[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.

as measured by the duration of individuals' incapacitating bouts of illness, if any.[2] Its central question is thus related to Zimran's (2017) finding that expanding access to markets stunted the rural population of the antebellum United States.

This paper investigates PMGSY's effects on morbidity in upland Orissa, a remote, backward, semi-arid region widely known – if not infamous – for its poverty and low scores on other social indicators. The emphasis is on episodes of acute illness, whereby communicable diseases account for the lion's share of the overall 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 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, distinguishing the separate stretches in km on track, rural road and highway comprising the entire trip to the nearest hospital and PHC, respectively. The two trips so described are effectively exogenous, whatever be villagers' actual choices. Such precision in defining the trip variables turns out to be essential if the effects of the programme are to be estimated with any precision. This approach has much in common with that employed by Donaldson and Hornbeck (2016), who estimate the aggregate impact of railroads on the U.S. agricultural sector in 1890.[3]

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* seasons (July 1 to December 31) of 2009 and 2013; 1076 individuals were present in both. Morbidity in *kharif* was much higher than in *rabi* (January 1 to June 30); moreover, data for *rabi* 2009 were not collected. Six of the 30 villages received a direct connection between 2004 and 2009, another four between 2010 and 2013. Five

---

[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.

[3]Their measure is least-cost, arrived at 'by constructing a network database of railroads and waterways and calculating lowest-cost county-to-county freight routes.'

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 random-effects models yield the following reductions in morbidity when one km. of *kutcha* track was replaced by one km. of PMGSY road. First, the probability that an individual fell sick at all was lowered by 4.3 percentage points; second, an individual's expected number of days of sickness was reduced by 0.54. These estimated effects 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 average in the early phase was somewhat over twice as long.

The paper is structured as follows. The region, the surveys and the general nature of the data are described in Section 2. Section 3 gives a summary account of the main illnesses and their incidence in relation to the villages' connections to the main road network. Section 4 addresses the question of what influenced the pattern of PMGSY connections, as actually implemented up to the end of 2013. There follow the definitions and brief discussions of the regressors 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. Alternative estimates of PMGSY's impact are presented and discussed in Sections 5.2 and 5.3. The chief conclusions and some remarks are drawn together in Section 6.

2   The Surveys and the Data

The study area comprises five, contiguous administrative blocks. One of the four in Bolangir District includes Titlagarh, a Sub-Division HQ. These four are separated from Kesinga Block in Kalahandi District by the river Tel. 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 households' vulnerability to drought. 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. At that time, 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 a socio-economic evaluation of PMGSY. The focus was on trade in goods, education and health; income data were not collected. The period covered was the *kharif* season of 2009. 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. 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 births, marriage and return migration yielded 289 newly present individuals in 2013, making 2656 observations in all. The total number of observations varied over villages, ranging from 60 to 118, but with an original sample of eight households per village, independently of village populations. The departures of daughters and arrivals of daughters-in-law, usually at a young age, are fixed by custom. Much of the migration was of the revolving kind.

3   Illnesses and Connections

The chief ailments are communicable diseases, especially malaria, various other fevers and water-borne diseases, and then especially during the *kharif* season. Some individuals suffer from chronic conditions, such as anaemia, tuberculosis, rheumatism and alcoholism; but relatively few respondents reported that family members did so.

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 common standard here. The very poor may be under a strong compulsion to work however miserable they feel; and children may get off lightly, since official enforcement of attendance is unlikely to be strict. In the event, all but a handful of all individuals reportedly suffering a bout of illness were also treated for it, if only by a traditional healer. It is therefore 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. Since virtually all those reportedly sick were treated, regardless of age, this definition also covers those who were too young to go to school or too old to work.

The distributions of the chosen measure of morbidity in the *kharif* seasons of 2009 and 2013 are set out in Table 1, with villages classified by the period in which they received a direct PMGSY connection, if at all. 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 13.0 days, 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 Pearson's chi-square and the

Kruskal-Wallis test. It is clear that this striking fall in morbidity cannot be attributed to the provision of just the four PMGSY connections in the interim. There was much heavier rainfall in the 'good' monsoon of 2009.

Turning to the differences in the distributions for villages with and without an all-weather road within a season, there is a hint that morbidity in 2009 was lower in the four villages that got a connection afterwards; but the null of independence cannot be rejected 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. Its rejection at the 7% level using Pearson's chi-square can be viewed as yet another instance of a well-known tendency of that test when the sample is large.

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 2, which gives an individual's main ailment, as defined by the resulting level of morbidity. The incidence of second and third ailments, if any, was quite similar, again with malaria and viral fevers the chief ones. Although nearly all individuals received medical treatment, 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 the years in aggregate, the lower incidence of morbidity in 2013 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 four villages that got a late connection. The null hypothesis that the incidence of the various

ailments was the same in both years is clearly rejected. The null that, in 2009, the incidence of the said illnesses was the same in all three categories of villages is just rejected at the 5% level, with a heavy incidence of malaria in the villages that got an early connection. The rejection is comprehensive for 2013. A difference-in-difference test involves a comparison of the four villages that got a late connection with the rest. The tests do not permit negative values; but it is seen that all of the counts for 2009 exceed those for 2013, with the exception of the category 'other'. Discarding the latter, Fisher's exact test yields a borderline rejection of the null ($p = 0.075$).

4 The Placement of PMGSY Connections

The pattern and timing of PMGSY connections are 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 is whether the changes in the network of all-weather roads were random over villages.

Villages that did not receive a direct connection would still have benefited from PMGSY if such a road happened to have an alignment that reduced their first stretches on a track. Five of the survey villages benefited in this way between 2004 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.[4] It mitigates, in part at least, any potential concern arising from the discussion that follows.

Some villages were ineligible for a direct connection: the lower limit on population in this region was 250 persons. 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 Public Works Department (PWD) district road, which leaves

---

[4]Indirect extensions play an important role in the identification strategy in Atack et al. (2010).

20 villages as candidates at the start of the program in 2000-01.

The 10 villages that had received a direct connection by 2013 form no obvious spatial pattern on the map. A related possibility is that certain blocks were favoured administratively or politically. Table 3 sets out the position in 2009 and 2013 (Titlagarh is in Block 1). Since there are rather few villages, sampling fluctuation could be 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. 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 ($p = 0.12$).

Villages have other observable characteristics that plausibly influence their chances of getting a connection. The more populous eligible villages have one obvious advantage: the more beneficiaries – and voters – the better. The politics of caste and the growing maoist insurgency may also have been at work. The proportion of a village's inhabitants belonging to the 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 4. For the set of six 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. For the 10 connections in 2013, the null is not rejected when population is the sole criterion (see $F(1, 18) = 1.882$ in Table 4). Yet when the village caste indicator is added, the null is rejected at the 1% level. Adding the other two variables confirms this rejection. We conclude that in both periods, political factors influenced whether a village was favoured with a PMGSY road, even imposing the statutory lower limit of 250 persons.

This conclusion raises the question of whether the provision of PMGSY roads to particular villages was accompanied by other public expenditures that especially favoured them. The latter could affect morbidity indirectly by raising incomes. In that event, the estimates presented below would overstate PMGSY's effects. This possibility cannot be excluded; but the data to investigate it are lacking.

4.1 The network and other regressors

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 all-weather roads that constitute the shortest route, which is almost surely the fastest for any given combination of transport modes. There are two trips, one to the nearest hospital, the other to the nearest PHC. These trips place the village within the route network, independently of villagers' actual choices of facility. The trip to the nearest hospital comprises the following elements:

- $h1\_d0$ denotes the stretch of *kutcha* track, in km;

- $h1\_d1$ the stretch of PMGSY road, in km;

- $h1\_d2$ the stretch of district (PWD) road, in km;

- $h1\_d20$ the stretch of PWD road in poor condition, in km;

- $h1\_d3$ the stretch of highway, in km;

- $h1\_d30$ the stretch of highway in poor condition, in km.

Likewise, the trip to the nearest PHC is defined by $h2\_d0, \ldots$. These descriptions of the village's position are 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.

The covariates are ordered in 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 somewhat, 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.

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 (the reference case, and almost invariably male). 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. There are spouses, sons, daughters, grandsons, granddaughters, aged parents, and other relatives. Each category's proportion in the whole sample is the mean value reported in Table 5.

There are five age 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 as the reference group.

13

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. Female heads of household are almost invariably widows, who may have other priorities. Caste, too, may influence outcomes; 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 vary with 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 communicable diseases account for the lion's share of morbidity, the village's total population is potentially in play, as are the proportions of its total area under forest and irrigation. 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 environmental and infrastructural characteristics 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 that PMGSY involves the substitution of an all-weather road for a stretch

---

[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.

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.

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

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

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, assumed to be serially uncorrelated with w_t. Let $t = 1, 2$ denote 2009 and 2013, respectively. In view of the much lower level of overall morbidity in 2013, a year fixed effect, denoted by t_{13}, is needed in $\mathbf{z}_{jt}$.

The central policy question is, what is the effect on morbidity of replacing part or all of a village's *kutcha* track with a PMGSY road? Before the advent of PMGSY, village j was connected to the main road network by a track of length d_j km. As the programme proceeds, this stretch may be replaced, either wholly with a direct connection, or partly, as connections are completed in j's general neighbourhood. In keeping with the data, let the effect be one of simple substitution, so that the trip to the main network at time t comprises $d0_{jt}$ km of track and $d1_{jt} = d_j - d0_{jt}$ km of PMGSY road, respectively.

Classifying all 30 villages by the existence or otherwise of any change in the date of their all-weather connections and the period in which the change occurred, we have the following scheme of 'treatments'.

Type 0: 16 villages experienced no changes at any time: $d0_{jt} = d_j$, $d1_{jt} = 0$ and, with the exception of the element t_{13}, $\mathbf{z}_{jt-1} = \mathbf{z}_{jt}$ $\forall t$.

15

Type 1: 6 villages received a direct connection between 2004 and 2009, ranging in length from 4.5 to 13 km. Another 5 villages received an indirect one, ranging from 2 to 10 km. To allow for the possibility that these changes in connections had not only a short-run effect on morbidity, as observed in 2009, but also an additional one over the longer run ending in 2013, $\mathbf{z}_{j2}$ also includes the lagged value $d1_{j1}$.

Type 2: 4 additional villages had received a direct connection between 2010 and 2013; for one village, this completed its earlier indirect connection. These connections were quite short, ranging from 1 to 3 km.

These subsets of villages will be denoted by S_0, S_1 and S_2, respectively. With village fixed effects, it is the variation associated with S_2 and S_1 that yields the possibility of estimating the programme's lagged and short-run effects, respectively.

The first two components of the trip to the nearest hospital or PHC are, with three exceptions, common to both facilities, that is, the contemporaneous values of $(h1\_d0, h1\_d1)$ and $(h2\_d0, h2\_d1)$ are equal. Hence, the effect of replacing $d0$ km of track on (individual) morbidity is

$$\delta \cdot d0 \equiv (\beta_{11} - \beta_{10} + \beta_{21} - \beta_{20} + \beta_{11,-1}) \cdot d0, \tag{2}$$

where β_{11} is the coefficient of $h1\_d1$, the other β's are analogously defined, $\beta_{11,-1}$ is the coefficient of the lagged value of $h1\_d1$ in 2013, and δ is the net effect per km so replaced. The village is the natural choice of group variable, with corresponding clustered standard errors when the estimator permits.

It seems plausible that in addition to the year fixed effect t_{13}, 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

16

is the first stretch of the whole trip, be it a *kutcha* track or an all-weather road, we therefore introduced interaction terms between t_{13} 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. These additional regressors proved to be statistically quite insignificant and were dropped.

5.1 Panel estimates: random effects

For reasons of efficiency, with just four new direct connection between 2009 and 2013, and no changes in the indirect connections, we begin with the random-effects estimator. As a first step, we employ a linear probability model (LPM), with the measure of morbidity transformed into the discrete variable $\{0, 1\}$ (not-sick/sick). Although this discards information, it may reduce measurement errors relative to morbidity defined in days, for which some digital preference is apparent in respondents' answers. The LPM usually provides good estimates of the partial effects of changes in the regressors near the centre of their distribution (Wooldridge, 2002: 455).

The only trip variables that are significant at the 5% level are the *kutcha* stretch to the nearest hospital and that involving a stretch of highway in poor condition to the nearest PHC, $h2\_d30$, with positive and negative signs, respectively (see Table 6). The time dummy t_{13} is highly significant: the estimated probability of falling sick in 2013 was 0.168 lower than in 2009. The only village characteristic that is significant at the 5% level is its population, pop01: an extra 100 persons yields an estimated increase of 0.04 in the probability that any resident fell sick.

Morbidity measured in days is analysed with a tobit model. The component for unobserved village heterogeneity accounts for just 5.0% of the total residual variance. The LR test cannot reject the alternative of simply pooling the individual observations $(p = 0.18)$, so that such pooling would also be defensible. The slope parameter associated with the year fixed effect has a point estimate of 6.99 days fewer than its

counterpart in 2009, and it is statistically highly significant.

None of the coefficients of the contemporaneous network regressors hf_d0 and hf_d1 is significant, even at the 10% level, though the coefficient of the lagged value of $h1\_d1$ is negative and significant at that level.[6] As in the LPM model, a trip to the nearest PHC on a stretch of highway in poor condition, $h2\_d30$, is significant at the 5% 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. 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 track in order to reach the main road network and thereafter a hospital, those who are not so very ill may choose to visit the nearest PHC instead. A similar stretch on a PMGSY road, however, is easily covered under all conditions. The findings from the reduced-form specification eq.(1) may indicate that individuals in graver condition were taken to hospitals, 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.

To probe a little more deeply, consider a Heckman-type 'two-tier 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 (at the second tier), conditional on falling sick. (Recall that just over one half of all individuals who did so had only one bout of illness.)

The term t_{13} enters very strongly in the first stage, where much of the action is concentrated (see Table 7). The coefficients of the trip variables $h1\_d0$ and $h2\_d30$ are significant at the 1% level, those of $h2\_d0$ and $h2\_d1$ at the 10% level. The identifying

---

[6]The correlation coefficient between $h1\_d1$ and $h2\_d1$ is high, at 0.87.

variable pop01 is highly significant. In the second stage, the lagged value of the trip variable $h1\_d1$ is negative and borderline significant at the 1% level, but no other trip variable is significant at conventional levels. The inverse Mills ratio is not at all significant.

5.2 PMGSY's effects

The the central policy question, as set out formally in eq.(2), is: 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 5% level and 10% levels in the LPM and tobit models, respectively. In the two-tier model, the null is clearly rejected at the selection stage, but is in no danger at the second (duration) stage.

The LPM estimate of δ implies a reduction in the probability of falling sick of 0.0427 per km. The tobit estimate of δ is -1.577 days. The associated marginal effect is -0.536 days per km, which is virtually identical to that yielded by adjusting δ by the fraction of non-zero observations ($-1.577 \times (919/2656) = -0.546$). The average length of 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 0.132 (LPM) and the expected number of days of sickness by 1.66 (tobit). The two-tier model yields a virtually identical estimate of the former when allowance is made for the proportion of non-zero observations at the selection stage. The estimated reduction in the expected duration of sickness is close to zero.

It might well be asked whether the same findings would emerge from a more parsimonious specification of the trips to the health facilities. A simple alternative is to replace the whole set of network regressors with an indicator variable for the existence of a direct all-weather connection. This yields estimates for both the LPM and to-

bit models that are qualitatively somewhat 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 serious errors 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.

In view of the sampling design, it is also pertinent to ask whether the above findings, which are based on equally weighted observations, are robust to this assumption. As noted in Section 2, a village's population had no effect on its total number of observations; but all villages had the same probability of being drawn, so that the inhabitants of small villages had correspondingly better chances of being sampled. This does not settle things, however. For it was established in Section 4 that, even among qualified villages, more populous ones had better chances of getting a direct connection; and among the unqualified ones, only two of nine benefited from an indirect connection. Thus, the inhabitants of bigger villages had a better chance of being 'treated'. On balance, the use of equal weights seems quite defensible. Taking a conservative position, one could ignore differences in the chances of getting a connection and weight observations by population size. The results are presented in Table 8. It is seen that the findings are fairly robust to the change in weights, with both estimated effects and standard errors being a bit larger.

5.3   Village fixed effects

Thus far unobserved village fixed effects have been ignored: 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 specifications, there remains a potential omitted variable problem. It is

therefore desirable to investigate whether a fixed-effects estimator produces different results.

Two preliminary remarks are called for. First, there is no fixed-effects estimator for tobit or probit. 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. Second, since $d1_{jt} + d0_{jt} = d_j$, one of the changes in the said trip components must be dropped.

Starting with the first-difference form, ignoring any changes in individuals' observable characteristics, and recalling the above classification of villages, eq.(1) yields $(t = 2)$

$$\Delta y_{ijt} = \Delta w_t + \Delta \epsilon_{ijt}, \ \forall j \in S_0 \, ; \ \Delta y_{ijt} = \beta_{11,-1} \cdot d1_{j1} + \Delta w_t + \Delta \epsilon_{ijt}, \ \forall j \in S_1 \, ;$$

and

$$\Delta y_{ijt} = \beta_{11} \cdot d1_{j2} + \Delta w_t + \Delta \epsilon_{ijt}, \ \forall j \in S_2.$$

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; the additional, lagged response, $\beta_{11,-1}$, per km of an earlier PMGSY connection; and the response per km of a new connection, β_{11}, from the observations in S_0 and S_1, and S_2, respectively.

Only 68% of all the individuals comprising the sample were present in both periods. Of these 1076 individuals, 483 were sick in 2009, but only 276 in 2013; 460 were sick in neither period, and 143 in both. There were some exogenous changes in some individuals' characteristics between 2009 and 2013, notably in age-groups and whether the household head was female, so these have been included as controls. There were also fairly extensive changes in the family relationship variables. These arose overwhelmingly from the splitting of households, which is arguably an endogenous event.

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, contemporaneous and lagged, of replacing the *kutcha* track with an all-weather road, $\Delta h1\_d1_t \, (= -\Delta h1\_d0_t)$, $t = 1, 2$, 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 lagged 'programme variable' $\Delta h1\_d1_{t-1}$ is negative, but not quite significant at the 10% level.

An alternative is to use the whole, unbalanced sample of 2656 observations and analyse levels of morbidity 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 9), 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, the lagged 'programme variable' $h1\_d1_{t-1}$ is significant at the 10% level. The contemporaneous value has a positive sign, but is significant only at the 25% level. With the population weights, the lagged term is not quite significant at the 10% level. If, however, the villages whose populations are fewer than 250 are discarded and the remaining observations are weighted inversely with respect to their populations to reflect the smaller chances of inhabitants of larger villages of being sampled, then $h1\_d1_{t-1}$ is significant at the 5% level.

It is not, perhaps, surprising that these findings, although broadly consistent with those in Section 5.2, are less precise. 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: OLS has evident shortcomings as a description of the duration data. 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).

5.4 Discussion

The placement of PMGSY roads is not random, neither spatially, nor in timing. Indirect connections might well be so, but with only five out of 30 villages so 'treated', all early ones, this identification strategy is unpromising. The 30 villages were, however, drawn randomly before PMGSY started, so it is arguable that the above findings reflect PMGSY's effects, as actually implemented, in the whole region between 2004 and 2013. Whether they also hold for later stages is a somewhat open question.

The underlying mechanisms can be only partly uncovered using the survey data. The possibility that the more grievously ill were more likely to be taken to a hospital if their village was well-connected was discussed above. This suggests using McFadden's random utility model to investigate whether the provision of a PMGSY road would induce some switching from PHCs or traditional healers to hospitals. In fact, no such conclusion can be drawn from this model if the first part of the journey to the main road network is common to both facilities, as is almost invariably the case. For the odds-ratio is a function of, *inter alia*, the vector of differences between the trip components $(h1\_d0, h1\_d1)$ and $(h2\_d0, h2\_d1)$, respectively, and both differences are necessarily zero, both before the PMGSY road is built and afterwards.

The results from the two-tier model indicate that PMGSY's effects are attributable more to a reduction in the probability of falling sick than in the duration of the bout conditional on falling sick. This is open to the interpretation that all-weather roads have a preventive effect. As noted in Section 3, there were no reported visits for preventive treatment. Even without such visits, preventive effects can occur in other ways, though perhaps with a lag. One possibility is improved nutrition as incomes rise. More directly, Khandker et al. (2009), found substantial, across-the-board increases in consumption following the provision of all-weather roads in Bangladesh. Another possibility is that better and more timely treatment of present ailments reduces future morbidity. For example, the high fever that usually accompanies a bout of malaria can

lead to neurological damage. This particular hazard is especially acute for children.

6 Conclusions

The region's population is poor and ill-educated, and the overall burden of disease arises chiefly from the communicable kind, especially in the monsoon season. In such an environment, the provision of all-weather roads will have ambiguous effects. Increased trafficking, in various senses of the term, will promote the propagation of disease; but better access should reduce mortality and the duration of morbidity, provided the sick receive moderately competent treatment upon arrival at a local clinic or hospital. Higher incomes should also lead to improved nutrition, as well as the willingness to meet the costs of treatment when fees are demanded.

The empirical findings presented above indicate that the net effect in the short- to medium run is a reduction in morbidity. The random-effects estimates are, first, that the probability that an individual fell sick at all was lowered by 0.043 for each km of track replaced. Second, an individual's expected number of days of sickness was reduced by 0.54 per km. Recent direct connections have been about 3 km long on average, which implies that the inhabitants of such villages have enjoyed an 13 percentage point reduction in the probability of falling sick and each of them, on average, 1.6 fewer days of sickness. The associated fixed-effects estimates are much less precise, but qualitatively consistent. The latter estimates are subject to substantial reservations. Yet they call for caution in accepting results based on the assumed absence of any correlation between the regressors and unobserved village heterogeneity, despite the deployment of an extensive battery of village-level characteristics as controls.

Insofar as these estimates of PMGSY's effects on morbidity are reliable, it is natural to ask whether they might hold more widely. Upland Orissa is part of a vast semi-arid tract in India's interior, whose populations 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 prevention and treatment is also broadly much the same. Pending further research, our findings provide a basis.

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] Burgess, R., and Donaldson, D. (2017). 'Railroads and the Demise of Famine in Colonial India', mimeo, LSE.

[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. (2010), 'Railroads of the Raj: Estimating the Effect of Transportation Infrastructure', NBER Working Paper 16487.

[8] Donaldson, D. and Hornbeck, R. (2016). 'Railroads and American Economic Growth: A 'Market Access' Approach', *Quarterly Journal of Economics*, 131(2): 799-858.

[9] 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.

[10] 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.

[11] Khandker, S.R., Bakht, Z., and Koolwal, G. B. (2009), 'The Poverty Impact of Rural Roads: Evidence from Bangladesh', *Economic Development and Cultural Change*, 57: 685-722.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

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

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

[19] Zimran, A. (2017), 'Transportation and Health in a Developing Country: The United States, 1820-1847', mimeo, Vanderbilt University.

[19] Zimran, A. (2017), 'Transportation and Health in a Developing Country: The United States, 1820-1847', mimeo, Vanderbilt University.

Table 1: Individuals' days of sickness, classified by village's date of connection

(Remark for reviewers: this table to go on-line as a supplement)

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 2: Individuals' chief acute diseases, by connection: *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	8	29	89	8	47	239
	(24.3)	(3.3)	(12.1)	(37.2)	(3.3)	(19.7)	(100)
2004-2009	13	5	7	26	3	14	68
	(19.1)	(7.4)	(10.3)	(38.2)	(4.4)	(20.6)	(100)
2010-2013	7	5	4	7	1	21	45
	(15.6)	(11.1)	(8.8)	(15.6)	(2.2)	(46.7)	(100)
Total[a]	78	18	40	122	12	82	352
	(22.2)	(5.1)	(11.4)	(34.7)	(3.4)	(23.3)	(100)

$\chi^2(10) = 25.33$, $p = 0.005$. [a] $\chi^2(10) = 54.97$, $p = 0.000$. Fisher's exact test, $p = 0.000$.

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

30

Table 3: Eligible village connections by block, *kharif* 2009 and 2013

(Remark for reviewers: this table to go on-line as a supplement)

Connection	2009 [a]			2013 [b]		
	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 4: Eligible villages' characteristics in 2001, by connection in 2013

PMGSY connection in 2013 ($n = 10$)				
Characteristic	mean	s.d.	min.	max.
Population	667	228	407	1060
OBC (percent)	59.4	21.9	22.1	80.8
Electricity connection	0.7	0.84	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	223	274	854
OBC (percent)	30.8	26.5	0.7	76.0
Electricity connection	0.6	0.52	0	1
Panchayat HQ (km.)	3.7	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 5: Regressors, summary statistics

(Remark for reviewers: this table to go on-line as a supplement)

Variable	2009 ($n = 1291$)				2013 ($n = 1365$)			
	mean	s.d.	min.	max.	mean	s.d.	min.	max.
spouse	.1642	.3706	0	1	.1890	.3916	0	1
son	.2463	.4310	0	1	.2381	.4261	0	1
daughter	.1565	.3634	0	1	.1758	.3808	0	1
grandson	.0627	.2426	0	1	.0432	.2034	0	1
granddtr.	.0480	.2139	0	1	.0366	.1879	0	1
mother	.0356	.1854	0	1	.0315	.1747	0	1
dtr.-in-law	.0565	.2311	0	1	.0469	.2115	0	1
other rel.	.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	1.944	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
S(ched.)C	.2115	.4085	0	1	.1802	.3845	0	1
OBC	.3656	.4818	0	1	.3626	.4809	0	1
population	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
hosp-track	2.430	1.868	0	8	2.203	2.001	0	8
hosp-PMGSY	2.338	3.732	0	13	2.671	3.841	0	13
hosp-PWD	5.695	9.339	0	33	5.600	9.217	0	33
hosp-PWD0	.697	2.781	0	17	.610	2.588	0	17
hosp-HWY	7.689	8.527	0	26	7.460	8.456	0	26
hosp-HWY0	1.649	4.635	0	21	1.664	4.733	0	21
PHC-track	2.486	1.780	0	8	2.245	1.920	0	8
PHC-PMGSY	1.795	3.322	0	13	2.112	3.487	0	13
PHC-PWD	2.662	3.649	0	12	2.660	3.670	0	12
PHC-PWD0	.757	2.392	0	12	.690	2.282	0	12
PHC-HWY	1.045	1.933	0	8	1.047	1.934	0	8
PHC-HWY0	.513	1.752	0	8	.488	1.688	0	6

Table 6: Morbidity 2009: random effects estimates

Variable	LPM [a] coeff.	s.e.	Tobit [b] coeff.	s.e.
pop01/100	.0404***	.0140	1.523***	.581
forest	.0026	.0021	.1099	.0833
irrigation	.0013	.0032	−.0097	.1276
altitude/100	−.0271	.0663	−.756	2.904
geo2	−.0190	.0594	1.770	2.544
geo3	.0252	.0489	−.960	1.791
block2	−.1163	.0825	−4.801	3.543
block3	.0068	.0495	−2.869	2.362
block4	.0007	.0880	−.098	3.863
block5	−.0072	.0519	.351	2.185
hosp-track	.0661**	.0296	1.963	1.294
hosp-PMGSY	.0100	.0102	.600	.518
hosp-PWD	.0008	.0033	.130	.124
hosp-PWD0	−.0051	.0122	−.432	.634
hosp-HWY	.0022	.0026	.137	.108
hosp-HWY0	.0029	.0054	.121	.213
PHC-track	−.0405	.0327	−1.065	1.312
PHC-PMGSY	−.0228*	.0134	−.897	.638
PHC-PWD	−.0059	.0058	−.423	.275
PHC-PWD0	−.0032	.0145	.224	.778
PHC-HWY	−.0083	.0069	−.238	.358
PHC-HWY0	−.0336***	.0118	−1.555***	.654
lag-hosp-PMGSY	−.0042	.0039	−.382*	.224
t_{13}	−.1676***	.0272	−6.986***	1.023
constant	.4019**	.1564	−6.102	7.211
σ	.4564		17.829***	.479

30 groups (villages). $n = 2656$, 919 uncensored obs.

Other controls: family variables, age groups hhedu, ownhold, SC, OBC.

[a] Linear probability model, discrete variable $\{0,1\}$. R^2 (overall) = 0.0816. Robust s.e.'s, clustering for villages.

[b] Wald $\chi^2(44) = 182.88$, log-likelihood $= -4833.5$, $\rho = 0.050$.

Table 7: Morbidity: two-tier model, *kharif* 2009 and 2013

(Remark for reviewers: this table to go on-line as a supplement)

	first tier		second tier	
Variable	coeff.	s.e.	coeff.	s.e.
hosp-track	.1994***	.0705	−.619	.791
hosp-PMGSY	.0305	.0284	.204	.303
hosp-PWD	.0034	.0061	.097	.063
hosp-PWD0	−.0213	.0174	.139	.150
hosp-HWY	.0073	.0057	.012	.057
hosp-HWY0	.0084	.0115	−.019	.113
PHC-track	−.1257*	.0723	.897	.800
PHC-PMGSY	−.0655*	.0345	−.025	.303
PHC-PWD	−.0183	.0137	−.193	.157
PHC-HWY	−.0201	.0192	.129	.229
PHC-HWY0	−.0999***	.0320	−.058	.330
lag-hosp-PMGSY	−.0111	.0137	−.443**	.173
t_{13}	−.4810***	.0618	−.185	1.525
pop01/100	.1161***	.0296		
forest	.0070	.0045		
irrigation	.0030	.0063		
altitude/100	−.0006	.0016		
geo2	−.0325	.1284		
geo3	−.0585	.0963		
constant	−.2732	.3914	9.403***	3.267
σ	9.546			
Mills lambda	3.477	3.705		

30 groups. $n = 2656$; 919 uncensored obs. $\rho = 0.364$. Wald: $\chi^2(37) = 76.07$.
Controls: family variables, age groups, ownhold, hhedu, hhsex, SC, OBC, block2,..., block5.

Table 8: The effect of PMGSY on morbidity

Model	discrete[a]		duration[b]	
	δ	s.e.	δ	s.e.
LPM	−.0427**	.0211	—	—
Tobit	—	—	−1.577*	.852
Heckman two-step	−.1199***	.0459	.106	.313
Population wts.[c]				
LPM	−.0483*	.0257	—	—
Tobit	—	—	−1.926*	1.137

The null hypothesis is no programme effect. $H_0 : \delta \equiv \beta_{11} - \beta_{10} + \beta_{21} - \beta_{20} + \beta_{11,-1} = 0$.
[a] Probability of falling sick. [b] In days. [c] No random-effects option, pooled estimator.

Table 9: 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.
$h1\_d1_t$	.0319	.0357	.705	.535	.0181	.0366	.740	.616
$h1\_d1_{t-1}$	$-.0017$	.0033	$-.151$	.091	$-.0028$	.0035	$-.161^*$	.086
t_{13}	—	—	—	—	$-.184^{***}$	.026	-2.23^{***}	.42
const.	$-.226^{***}$	.029	-2.63^{***}	.54	$.422^{***}$	.098	4.10^{**}	1.73
Population wts.								
$h1\_d1_t$	.0334	.0372	0.804	0.541	.0195	.0386	.810	.631
$h1\_d1_{t-1}$	.0006	.0039	-0.130	0.097	$-.0021$	.0037	$-.153$	.093
t_{13}	—	—	—	—	$-.195^{***}$	.034	-2.36^{***}	.45
const.	-0.240^{***}	0.038	-2.94^{***}	0.48	$.437^{***}$	.153	3.17	2.72

Standard errors clustered on villages.

[a] $n = 1076$ individuals. Regressand: $\Delta y_t \equiv y_{2013} - y_{2009}$.
Regressors: $\Delta h1\_d1_t \equiv h1\_d1_t - h1\_d1_{t-1}$, $t = 1, 2$. Controls: family relations, age groups, hhsex.
LPM: $F(7, 29) = 1.40$. Duration: $F(7, 29) = 0.78$.
[b] $n = 2656$ observations, 30 groups.
[c] R^2: within $= 0.0672$, between $= 0.0001$. $F(18, 29) = 28.05$.
[d] R^2: within $= 0.0498$, between $= 0.0017$. $F(18, 29) = 22.82$.