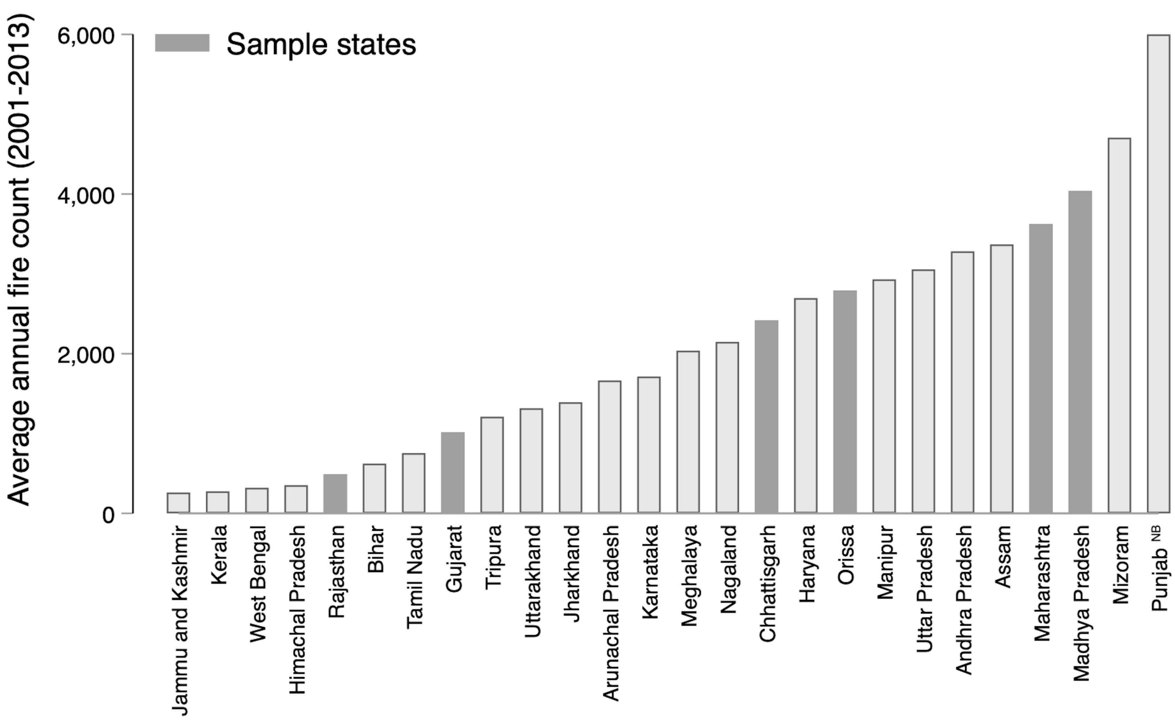
Rural Roads, Farm Labor Exits, and Crop Fires

Teevrat Garg, Maulik Jagnani, and Hemant K. Pullabhotla

Online Appendices

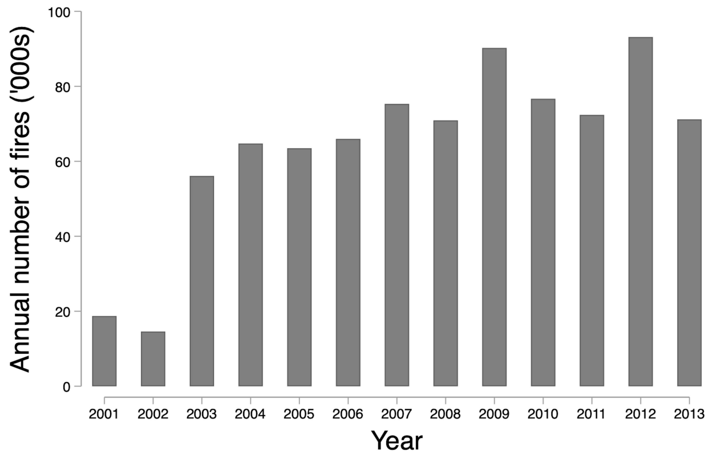
A Appendix Figures and Tables

Figure A.1: Annual average fire counts per state

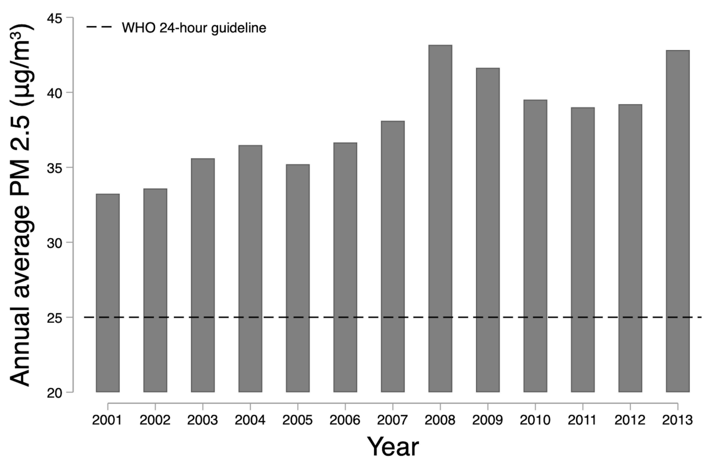


Notes: Figure shows the annual average fire counts for 2001 - 2013 for each state. States in the analysis sample are highlighted in darker shaded bars. Note that the value for Punjab is truncated at 6000 for ease of visual representation. The average fire counts per year for Punjab is $\approx$ 15,000 fires.

Figure A.2: Annual number of fires and average PM 2.5 across India



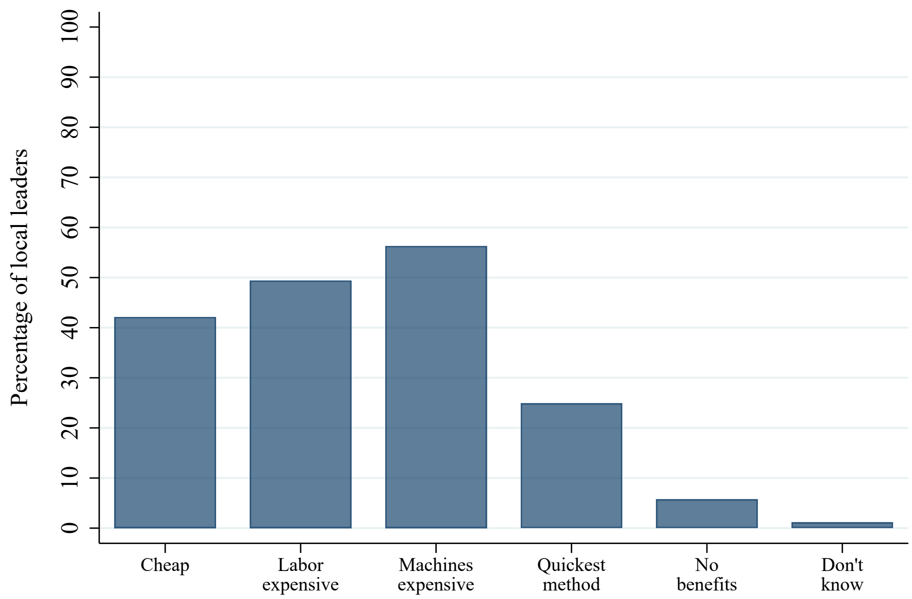
(a) Fire counts



(b) PM 2.5

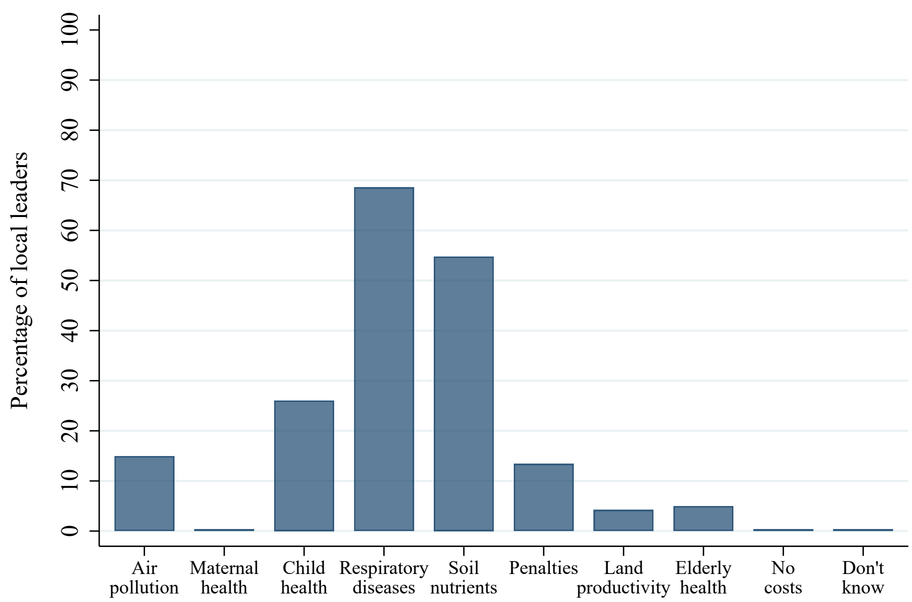
Notes: Panel (a) shows the number of fire pixels detected each year over India from MODIS TERRA and AQUA satellite data from 2001 to 2013. MODIS AQUA became operational only after July 2002. Therefore, the total number of fires detected for 2001 and 2002 are lower compared to later years when both TERRA and AQUA were in operation. Panel (b) shows the average annual PM 2.5 across India using data from Van Donkelaar et al. (2016). The horizontal dashed line shows the WHO 24-hour guideline of 25 $\mu g/m^3$ for PM 2.5.

Figure A.3: Benefits and costs of crop fires



Notes: Survey data from 261 panchayat mukhiyas (community leaders) across Punjab, India, December 22 - January 23.

(a) Benefits of crop fires

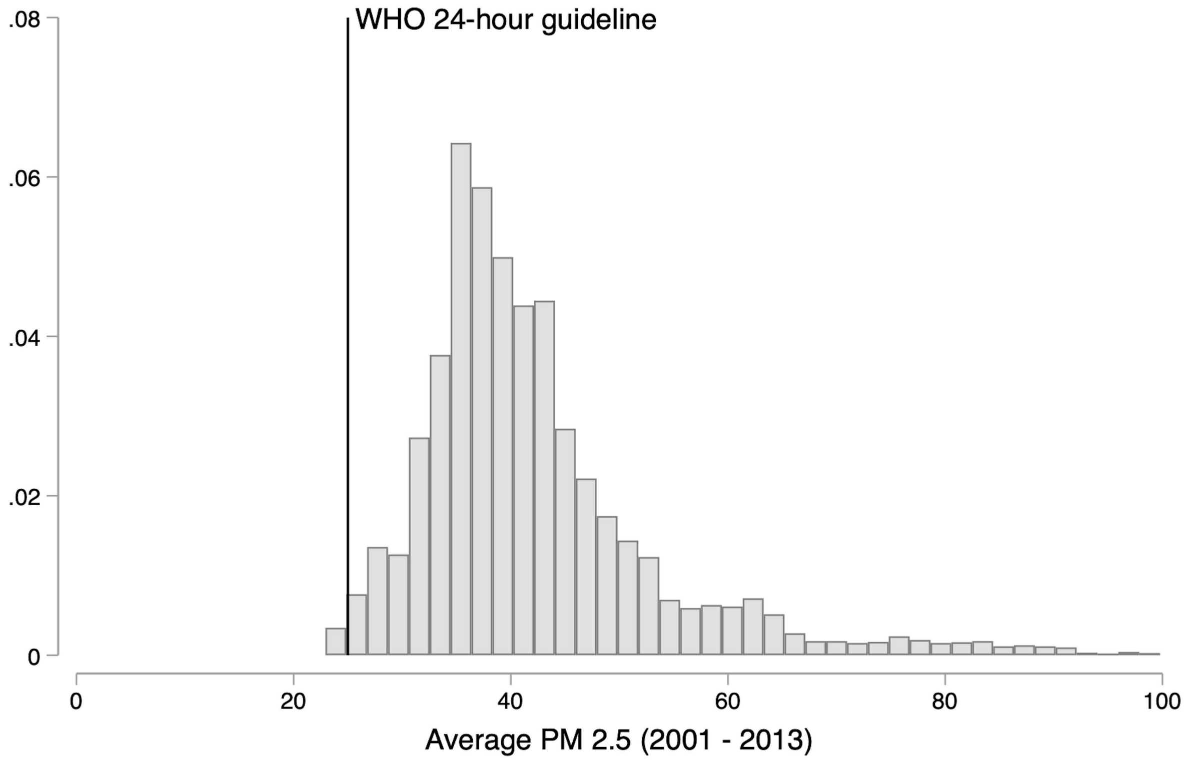


Notes: Survey data from 261 panchayat mukhiyas (community leaders) across Punjab, India, December 22 - January 23.

(b) Costs of crop fires

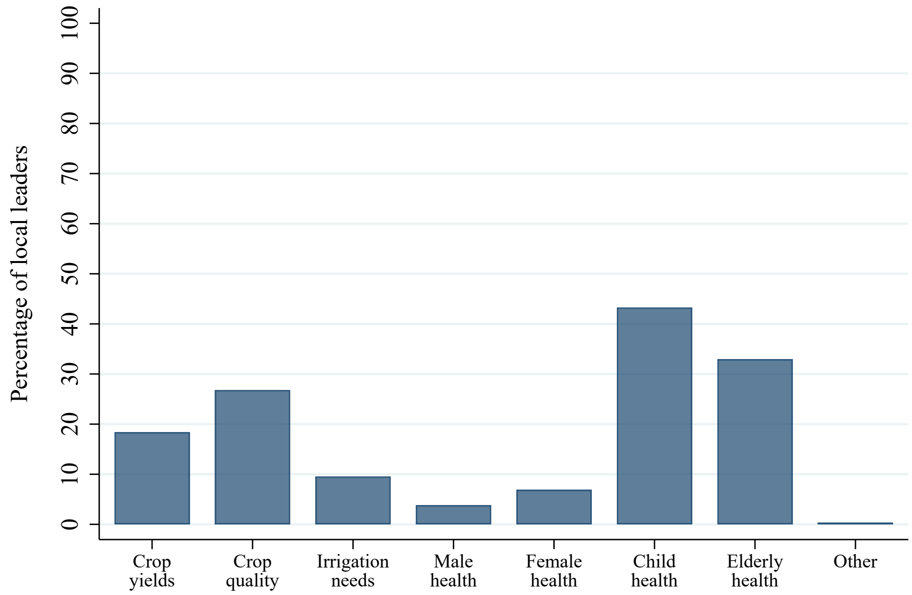
Source: Jagnani and Mahadevan (2023)

Figure A.4: Distribution of annual average PM2.5 at the village level



Notes: Figure plots the distribution of annual average PM 2.5 (μg/m^3) between 2001 and 2013 in PMGSY villages. The vertical line indicates the WHO 24-hour guideline of 25 μg/m^3.

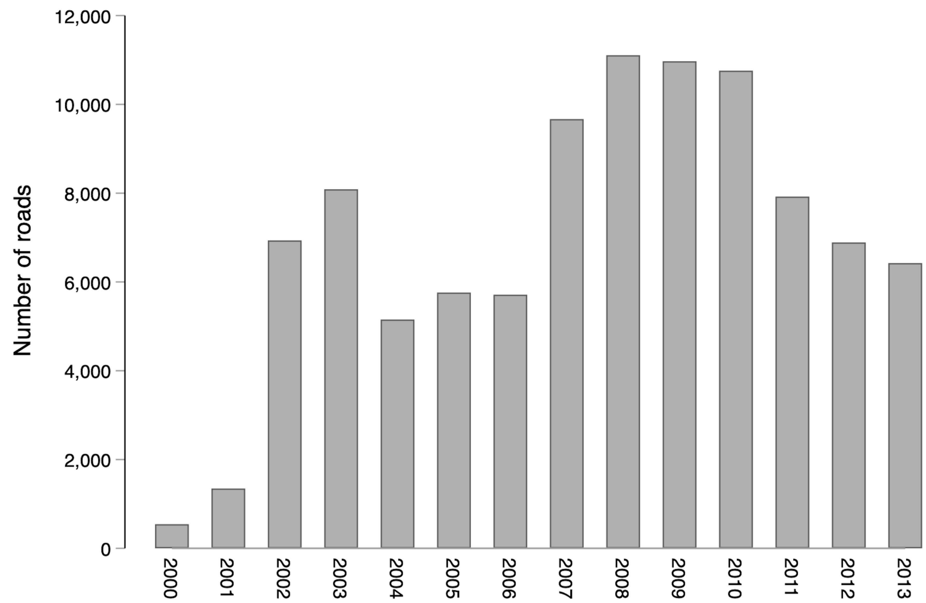
Figure A.5: Personal costs of crop fires



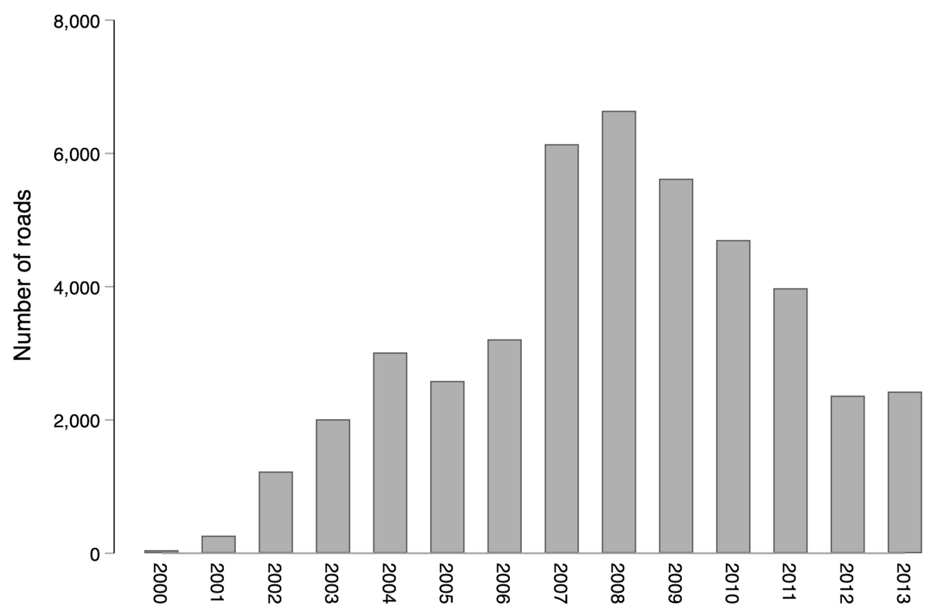
Notes: Survey data from 261 panchayat mukhiyas (community leaders) across Punjab, India, December 22 - January 23.

Source: Jagnani and Mahadevan (2023)

Figure A.6: Roads constructed under the PMGSY between 2000 and 2013



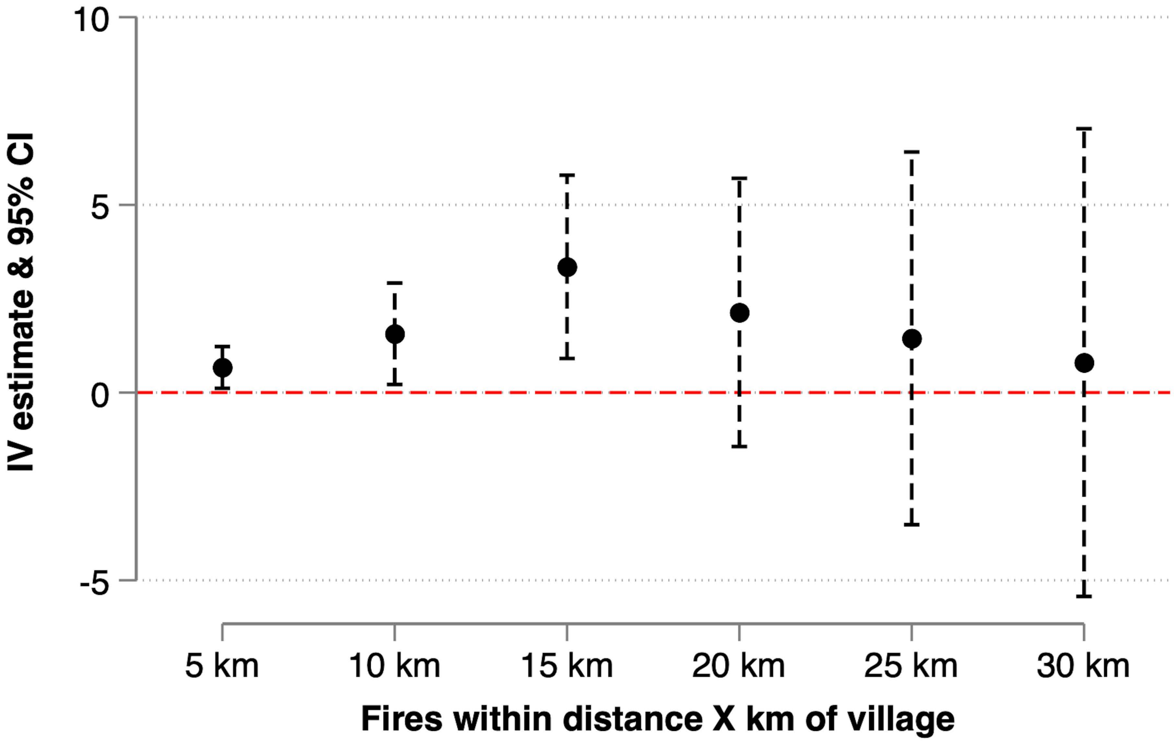
(a) All States



(b) Sample States

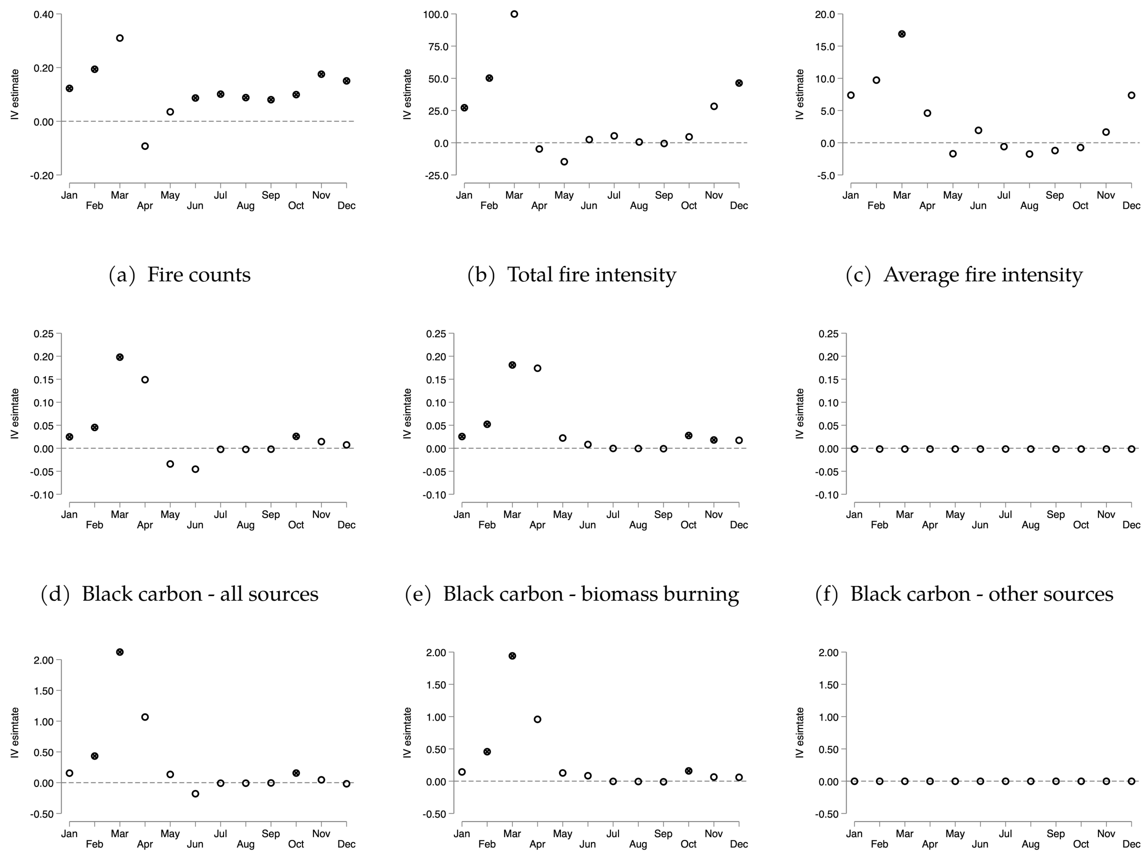
Notes: Panel (a) shows the total number of roads constructed per year under the PMGSY for all Indian states and Panel (b) shows total number of roads constructed per year in sample states. The road construction data is drawn from Asher et al. (2021) and Asher and Novosad (2020).

Figure A.7: Impact of rural roads on fires (count) with varying buffer radius around villages



Notes: This figure shows regression discontinuity IV estimates of receiving a new road on the annual count of fires within different buffer radii around villages. Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. Each estimate is from a separate regression. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. 95% confidence intervals are clustered at the PMGSY village level.

Figure A.8: Impact of rural roads on agricultural fires (count) and particulate emissions ($ng/m^2/s$) by month and point sources



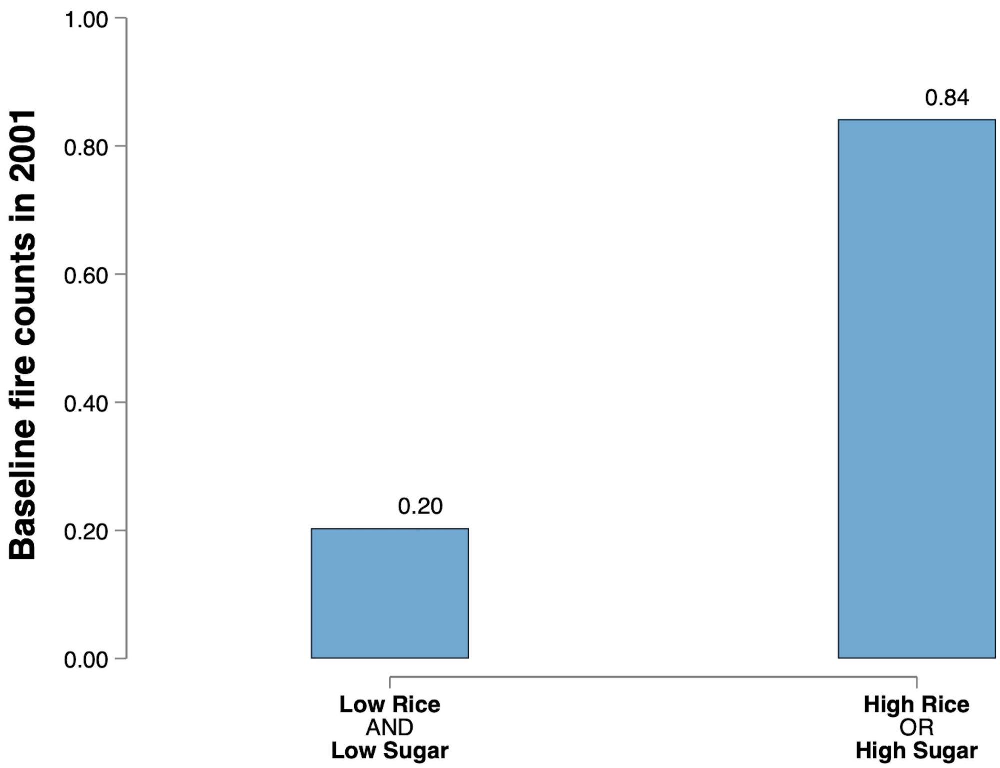
(a) Fire counts (b) Total fire intensity (c) Average fire intensity

(d) Black carbon - all sources (e) Black carbon - biomass burning (f) Black carbon - other sources

(g) Organic carbon - all sources (h) Organic carbon - biomass burning (i) Organic carbon - other sources

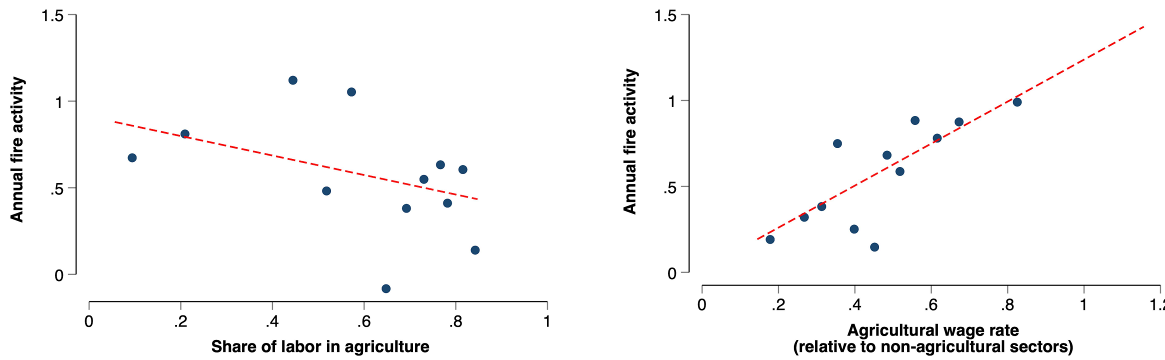
Notes: This figure shows the month-by-month effects of rural roads on fire counts, total fire intensity, average fire intensity, and black carbon (BC) and organic carbon (OC) emissions in the winter harvest and post-harvest months and the rest of the year from 2002-2013. Winter harvest and post-harvest period comprises the months from October - March. Fire activity is measured in counts and emissions are measured in nano-gram per square meter per second ($ng/m^2/s$). Fire intensity is captured by the brightness temperature (in Kelvin, K) of a fire pixel. The brightness temperature is a measure of the intensity of infrared radiation captured by Band 21 (thermal infrared channel) of the MODIS instrument. Panels (d) and (g) show impact on emissions from all sources, panels (e) and (h) are emissions from biomass sources only, and panels (f) and (i) are emissions from other (non-biomass sources) sources. All regressions control for district-threshold FE, year FE, and baseline controls. Standard errors in parentheses are clustered at village level. Estimates shown with a $\otimes$ are significant at 90% level.

Figure A.9: Fires in baseline (2001) by fire-suited cropping patterns



Notes: This bar graph shows the average fire counts at baseline (2001) for villages in i) districts with high rice or high sugarcane production (fire-suited cropping patterns) and ii) districts with low rice and low sugarcane production. High (low) rice districts are defined as districts with rice acreage share above (below) the sample median at baseline. High (low) sugarcane districts are defined as districts with sugarcane acreage share above (below) the sample median at baseline.

Figure A.10: Correlation between farm labor share and crop fires and relative farm wage rate and crop fires at baseline



(a) Farm labor share and crop fires (count)

(b) Relative farm wage rate and crop fires (count)

Notes: Both panels show a binned scatter plot and linear fit line, conditional on state fixed effects. Figure (a) shows the correlation between annual crop fires (count) and share of wage labor in agriculture in 2001. Figure (b) shows the correlation between annual crop fires (count) and relative agricultural wage rate in 2001. Rural labor shares and agricultural and non-agricultural daily labor wage rates are district-level estimates based on data from the 1999 - 2000 NSSO survey data (Round 55). Annual crop fires is the baseline annual count of fires at the village level in 2001 within the sample of villages used in the regression discontinuity estimates.

Table A.1: Impact of rural roads on annual fire intensity (in Kelvin, K)

	Annual fires intensity	
	(1) Total	(2) Average
Road built	3124.420**	1768.106*
	(1396.319)	(912.488)
N	133,788	133,788
Control group mean	5,412.68	4,276.38

Notes: This table shows the effects of rural roads on intensity of agricultural fires. The sample consists of the panel of villages from 2002 - 2013. Fire intensity is captured by the brightness temperature (in Kelvin, K) of a fire pixel. The brightness temperature is a measure of the intensity of infrared radiation captured by Band 21 (thermal infrared channel) of the MODIS instrument. It is the primary input used to detect thermal anomalies that signify the presence of a fire (Giglio et al., 2021). The higher the temperature, more intense is the thermal radiation from a fire. Column (1) shows the effect on annual total fire intensity of all fire pixels detected within a 10 km radius around sample villages. Column (2) shows the impact on annual average fire intensity within 10 km around sample villages. All regressions include district-threshold fixed effects, year fixed effects, and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table A.2: Impact of rural roads on annual agricultural fires (count) and PM2.5 ($\mu g/m^3$) by agricultural mechanization at baseline

	High mechanization index		Low mechanization index	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	1.007*	0.272	1.937	0.345
	(0.576)	(0.251)	(1.404)	(0.262)
N	57,552	57,552	59,388	59,388
Control group mean	2.17	47.52	3.31	39.21

Notes: This table shows the effects of rural roads on village-level annual fires and PM 2.5 ($\mu g/m^3$). The sample consists of the panel of PMGSY villages from 2002 - 2013. High (low) mechanization index sample consists of districts which had above (below) sample median agricultural mechanization index at baseline (2001). Mechanization index is measured using district-level data on mechanized input usage from the Agricultural Input Census (2001). The Agricultural Input Census provides district-level counts of the mechanized equipment used for various agricultural operations. We standardize the individual equipment counts into z-scores and construct the mechanization index as the sum across all z-scores. All regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table A.3: Impact of rural roads on agricultural fires (count) and PM2.5 ($\mu g/m^3$) by baseline agricultural mechanization in high rice acreage districts

	High mechanization index		Low mechanization index	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	2.915*	0.839	5.267	1.166***
	(1.622)	(0.570)	(3.373)	(0.408)
N	29,700	29,700	28,452	28,452
Control group mean	2.63	39.83	4.00	34.52

Notes: This table shows the effects of rural roads on village-level annual fire activity and PM 2.5 ($\mu g/m^3$). The sample consists of the panel of PMGSY villages from 2002 - 2013 within districts with high (above sample median) share of cropped area under rice. High (low) mechanization index sample consists of districts which had above (below) sample median level of agricultural mechanization index at baseline (2001). Mechanization index is measured using district-level data on mechanized input usage from the Agricultural Input Census (2001). All regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table A.4: Impact of rural roads on agricultural fires (count) and PM2.5 ($\mu g/m^3$) by baseline mechanization specific to rice harvesting in high rice acreage districts

	High mechanization index		Low mechanization index	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	2.845	1.730	3.562*	0.753**
	(3.396)	(1.413)	(2.075)	(0.294)
N	14,928	14,928	43,224	43,224
Control group mean	2.98	44.00	3.42	34.91

Notes: This table shows the effects of rural roads on village-level annual fire activity and PM 2.5 ($\mu g/m^3$). The sample consists of the panel of PMGSY villages from 2002 - 2013 within districts with high (above sample median) share of cropped area under rice. High (low) mechanization index sample consists of districts which had above (below) sample median level of agricultural mechanization index at baseline (2001). Mechanization index is measured using district-level data on mechanized input usage from the Agricultural Input Census (2001). We standardize the equipment counts for rice harvesting equipment – self-propelled combine harvesters and tractor drawn combines – into z-scores and construct the mechanization index as the sum across all z-scores. All regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

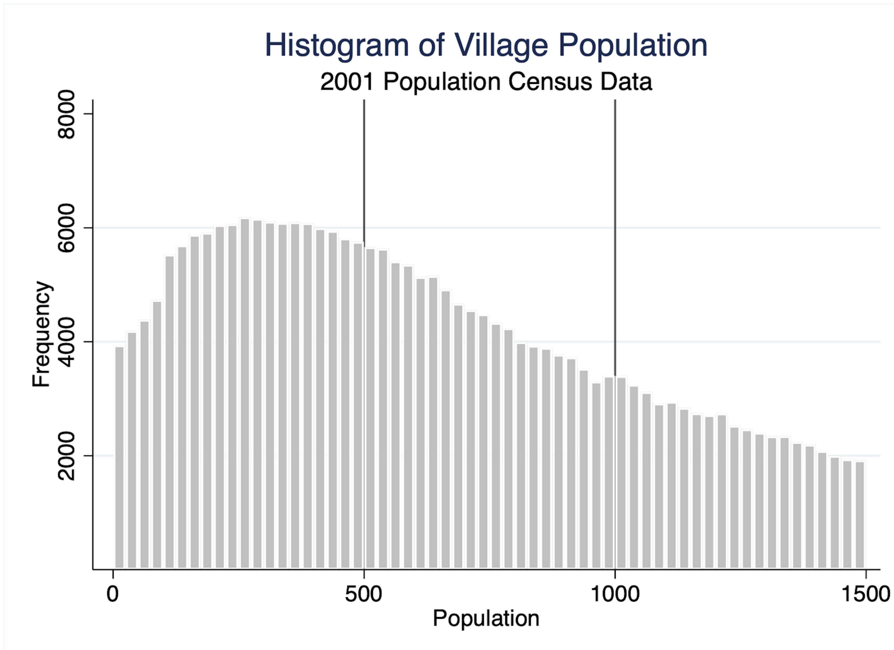
Table A.5: Impact of rural roads on agricultural fires (count) activity by size of the expected opportunity cost and returns to education effects

	Opportunity cost effect		Returns to education effect	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
Road built	4.537***	-0.186	1.773**	1.686
	(1.629)	(0.646)	(0.871)	(1.055)
N	65,172	68,616	67,620	64,596
Control group mean	3.20	2.35	2.87	2.53

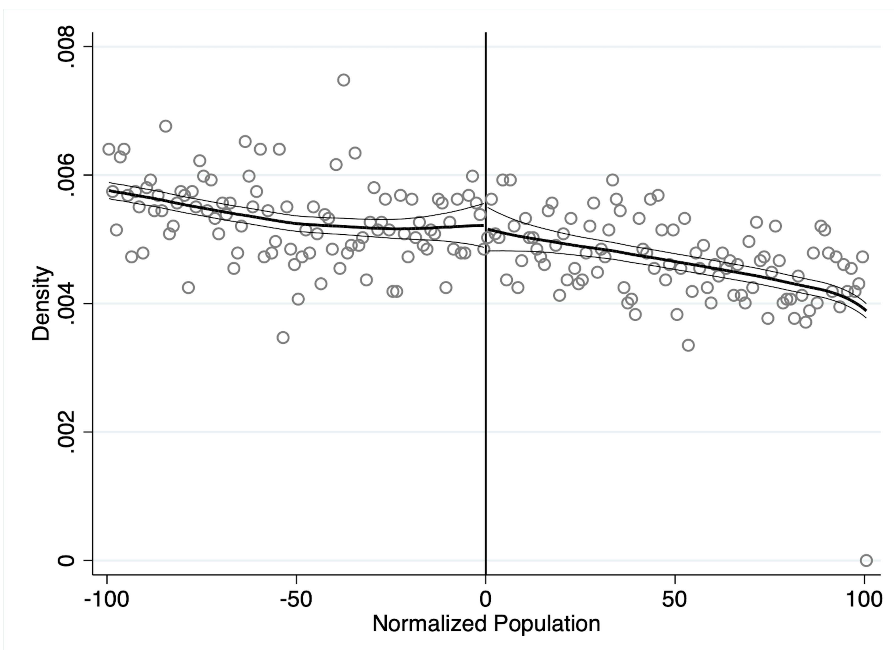
Notes: This table shows the effect of rural roads on village-level annual fire activity. The sample consists of the panel of villages from 2002 - 2013. Proxies for the opportunity cost and returns to education effects are estimated using pre-program survey data from the 1999 - 2000 NSSO survey data (Round 55). The size of the opportunity cost effect is proxied by the district-level mean low-skill urban wage minus the mean low-skill rural wage. The size of the returns to education effect is proxied by the difference between the urban and rural Mincerian returns to one additional year of education. The Mincerian returns are estimated by regressing log wage on years of education, age, square of age, and the log of household land owned, separately for individuals in urban and rural locations in each district. For each measure, the sample is split into high/low based on whether the proxy is above/below the sample median. Regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

B Replication of Asher and Novosad (2020)

Figure B.1: Distribution of running variable



(a) Histograph of Village Population



(b) McCary Test

Notes: Panel (a) and (b) replicate results from Asher and Novosad (2020) showing the distribution of village population around the population thresholds. Panel (a) is a histogram of village population as recorded in the 2001 Population Census. The vertical lines show the program eligibility thresholds at 500 and 1,000. Panel (b) uses the normalized village population (reported population minus the threshold, either 500 or 1,000). It plots a non-parametric regression to each half of the distribution following McCrary (2008), testing for a discontinuity at zero. The point estimate for the discontinuity is -0.010, with a standard error of 0.048.

Table B.1: Impact of rural roads on indices of major outcomes

	(1) Transportation	(2) Ag. occupation index	(3) Firms	(4) Agriculture	(5) Consumption
Road built	0.432**	-0.370**	0.237	0.050	0.016
	(0.190)	(0.162)	(0.159)	(0.127)	(0.138)
N	11,151	11,151	10,405	11,151	11,151
Control group mean	-0.02	-0.00	0.01	0.00	-0.00

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on effect of a new road on indices of the major outcomes in each of the five families of outcomes: transportation, occupation, firms, agriculture, and welfare. Each index is generated following Anderson (2008). The transportation index consists of availability of public buses, private buses, vans, taxis and auto-rickshaws at the village level. The agricultural occupation index is a combination of the share of workers in agriculture and the inverse of the share of workers in manual labor. The firms index is the log of employment in all non-farm firms in the village. The agriculture production index is a combination of remotely-sensed Normalized Difference Vegetation Index (NDVI), share of households owning farm equipment, share of households owing irrigation equipment, share of households owning land, log total cultivated acres and presence of non-cereal or pulse crops among the primary three crops in the village. The consumption index consists of log per capita consumption, primary component asset index, log of night light luminosity and share of households with primary earner having a monthly earning of more than INR 5,000. All regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.2: Impact of rural roads on transportation

	(1) Govt. bus	(2) Pvt. bus	(3) Taxi	(4) Van	(5) Autorickshaw
Road built	0.130**	0.131*	0.012	-0.015	0.068
	(0.056)	(0.076)	(0.049)	(0.054)	(0.044)
N	11,151	11,151	11,151	11,151	11,151
Control group mean	0.12	0.20	0.07	0.14	0.06
R^2	0.30	0.09	0.09	0.43	0.26

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on availability of transportation services at the village level. Columns (1) - (5) estimate the impact of new roads on five categories of transport services available at the village level from the Census in 2011. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.3: Impact of rural roads occupation and income source

	Occupation				Income source	
	(1) Agriculture	(2) Manual labor	(3) Unemployed	(4) Unclassifiable	(5) Agriculture	(6) Manual labor
Road built	-0.100**	0.086**	0.011	-0.008	-0.045	0.003
	(0.044)	(0.044)	(0.025)	(0.010)	(0.045)	(0.044)
N	11,151	11,151	11,151	11,151	11,151	11,151
Control group mean	0.48	0.45	0.43	0.02	0.41	0.51
R^2	0.29	0.26	0.30	0.17	0.31	0.28

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on occupational choice among workers and households' primary income source based on data from the 2011 - 2012 Socioeconomic and Caste Census (SECC). Column (1) shows the effect on share of workers in agriculture, while column (2) shows the effect on share of workers in non-agriculture manual labor. These two categories make up 92% of the workers, on average, in the sample villages. To examine any potential changes in the workforce, columns (3) and (4) estimate the impact on the share of working age adults (18-60) who are unemployed (including those who are studying, only involved in household work, etc.) and the share of working age adults whose work force participation is unclear in the SECC data, respectively. The results in columns (3) and (4) suggest that roads do not affect the share of adults in these categories, alleviating any concerns that the change in share of workers in agriculture or non-agricultural wage work is biased due to changes in the workforce. Additionally, results in Table B.8 show that rural roads had no effect on permanent out-migration. The absence of any impact on migration suggests that the observed sectoral reallocation of labor is the result of changes in occupational choice rather than compositional effects due to selective migration in the village population. Column (5) shows the impact on share of households in the village that report their main income source as cultivation. Column (6) shows the impact on share of households whose primary income source is manual labor. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.4: Impact of rural roads on employment in village non-farm firms

Panel A: Log employment growth

	(1) Total	(2) Livestock	(3) Manufacturing	(4) Education	(5) Retail	(6) Forestry
Road built	0.241	0.229	0.243	0.129	0.318**	-0.107
	(0.162)	(0.190)	(0.196)	(0.145)	(0.157)	(0.110)
N	10,405	10,405	10,405	10,405	10,405	10,405
Control group mean	2.95	0.69	0.91	1.50	1.23	0.17
R^2	0.30	0.42	0.24	0.18	0.23	0.35

Panel B: Level employment growth

	(1) Total	(2) Livestock	(3) Manufacturing	(4) Education	(5) Retail	(6) Forestry
Road built	2.801	-1.918	2.432	0.277	1.811	2.473
	(7.773)	(3.399)	(3.910)	(0.993)	(1.582)	(4.144)
N	10,405	10,405	10,405	10,405	10,405	10,405
Control group mean	32.31	6.85	5.88	5.11	4.55	2.88
R^2	0.30	0.46	0.18	0.13	0.16	0.36

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on employment in village level non-farm firms based on data from the 2013 Economic Census. Column (1) shows the impact on total non-farm employment. Columns (2) - (6) show results for the five largest sectors in the sample which constitute 79% of the non-farm firm employment (livestock, manufacturing, education, retail, and forestry). Panel A shows the impact on log employment in all non-farm firms and Panel B present same estimates using levels. The sample is limited to villages where total employment reported in the firm-level data is less than the total inhabitants in the village. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.5: Impact of rural roads on agricultural yields

	NDVI			EVI		
	(1)	(2)	(3)	(4)	(5)	(6)
	Max - June	Cumulative	Max	Max - June	Cumulative	Max
Road built	0.028	0.002	0.015	0.043	0.000	0.024
	(0.027)	(0.013)	(0.014)	(0.034)	(0.016)	(0.019)
N	11,054	11,053	11,054	11,054	11,053	11,054
Control group mean	8.24	10.52	8.81	7.96	10.17	8.48
R^2	0.70	0.88	0.80	0.71	0.85	0.69

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on village-level measures of agricultural activity using three different remote-sensing NDVI-based proxies for agricultural yields. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.6: Impact of rural roads on agricultural inputs

	(1)	(2)	(3)	(4)	(5)
	Mech.	Irri.	Own ag. land	Non-cereal/pulse	Cult. land (log)
Road built	-0.008	-0.009	0.005	0.002	-0.062
	(0.012)	(0.029)	(0.040)	(0.075)	(0.117)
N	11,150	11,151	11,151	8,001	10,885
Control group mean	0.04	0.14	0.57	0.40	5.04
R^2	0.24	0.40	0.29	0.45	0.46

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on the impact of roads on agricultural inputs based on data from the 2011-2012 Socioeconomic and Caste Census (Columns 1-3) and the 2011 Population Census (Columns 4-5). The outcome variables in Columns (1) and (2) are the share of households owning mechanized farm equipment and the share of households owning irrigation equipment, respectively. Column (3) shows the impact on the share of households owning agricultural land. The outcome in Column (4) is a dummy variable taking the value 1 if a non-cereal and non-pulse crop is listed as one of its three major crops at the village-level census data. Column (5) examines the impact on log total cultivated land in the village (villages that have no cultivable land are excluded from the sample). Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.7: Impact of rural roads on consumption and asset ownership

Panel A: Consumption indicators and asset index

	(1) Consumption per capita (log)	(2) Night lights (log)	(3) Share of HH earning $\geq 5k$	(4) Asset index
Road built	0.013	0.037	-0.008	0.120
	(0.039)	(0.169)	(0.032)	(0.135)
N	11,151	10,828	11,151	11,151
Control group mean	9.56	1.58	0.15	-0.01
R^2	0.42	0.66	0.25	0.52

Panel B: Individual asset ownership

	(1) Solid house	(2) Refrigrator	(3) Any vehicle	(4) Phone
Road built	0.037	0.004	-0.007	0.020
	(0.030)	(0.013)	(0.024)	(0.041)
N	11,151	11,151	11,151	11,151
Control group mean	0.22	0.04	0.14	0.44
R^2	0.66	0.26	0.38	0.48

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on indicators of consumption and asset ownership. Panel A, Column (1) shows the impact on imputed log consumption per capita (see Asher and Novosad (2020), Data Appendix for details). Column (2) estimates the effect on log of mean total night light luminosity in 2011-13. Column (3) is the share of households whose highest earning member earns more than INR 5000 per month based on data from the 2011-12 Socioeconomic and Caste Census (SECC). Column (4) is village-level average of the primary component of indicator variables for all household assets included in the SECC. Panel B shows the impact on the village-level share of households owning major assets in the SECC. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table B.8: Impact of rural roads on village-level population, age distribution, and gender ratios

Panel A: Population growth (2001 - 2011)

	(1) Log	(2) Level
Road built	-0.023	-9.833
	(0.030)	(20.783)
N	11,151	11,151
Control group mean	6.43	654.59
R^2	0.79	0.84

Panel B: Age group share

	(1) 11 - 20	(2) 21 - 30	(3) 31 - 40	(4) 41 - 50	(5) 51 - 60
Road built	-0.004	-0.004	0.002	-0.002	0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)
N	11,151	11,151	11,151	11,151	11,151
Control group mean	0.24	0.19	0.15	0.11	0.07
R^2	0.22	0.19	0.25	0.37	0.40

Panel C: Male share by age group

	(1) 11 - 20	(2) 21 - 30	(3) 31 - 40	(4) 41 - 50	(5) 51 - 60
Road built	-0.013	0.004	0.004	-0.007	0.016
	(0.009)	(0.008)	(0.008)	(0.010)	(0.013)
N	11,151	11,151	11,151	11,151	11,151
Control group mean	0.52	0.52	0.51	0.52	0.51
R^2	0.12	0.19	0.10	0.07	0.06

Notes: This table replicates results from Asher and Novosad (2020) showing regression discontinuity treatment estimates of the effect of new village roads on population growth and demographic structure of the village population. Panel A shows the effect on village population from the Population Census 2011. Column (1) shows the outcome in log, and (2) shows the level effect. Panel B shows the effect on share of population in 10 year age groups, starting from 11 to 20 in Column (1) to 51 to 60 in Column (5). The outcome variables in Panel C are the share of men in the population within each 10-year age group. Outcomes in Panels B and C are based on 2011-12 Socioeconomic and Caste Census (SECC). Panel A shows that new roads do not affect population growth, suggesting no permanent migration impacts (short-term migrants and daily commuters are accounted for in the Census and SECC listings). The effect on population could be attenuated towards zero if there had been changes in fertility or mortality that offset any net migration. However, the null results in Panels B and C on age distribution and sex ratios by age cohort show changes in fertility and mortality were unlikely to be a contributing factor. Together, the results in this table suggest that roads did not lead to any substantial out-migration. Absent any impact on migration, the effects observed on labor reallocation (in Table B.3) can be seen as a result of occupational choice rather than the result of selective migration. Regressions include district-threshold fixed effects and baseline control variables. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

C Rural Economic and Demographic Survey (REDS)

We use village- and household-level surveys from the 1999 and 2006 rounds of the Rural Economic and Demographic Survey (REDS), administered by the National Council of Applied Economic Research (NCAER), to estimate the effect of rural roads on agricultural wages, local stock of combine harvesters, and use of hired mechanized agricultural equipment. REDS is a nationally representative survey of rural households in India spanning 221 villages across 100 districts in 17 major states. It includes a village survey that collects information on the prevailing wage rates for agricultural labor as well as stock of agricultural machinery at the village-level . The household questionnaire provides detailed information on the use of agricultural inputs, including the use and cost of hired mechanized equipment.

Because the data includes few villages in states that followed population thresholds to determined rural road construction under PMGSY, we cannot use a regression discontinuity design. Instead, we estimate simple difference-in-difference specifications. 'Treat' is an indicator variable that takes the value 1 if a village receives a rural road between 1999 and 2006, 0 otherwise. 'Post' is an indicator variable that takes the value 1 if the year is 2006, 0 otherwise. 'TreatXPost' captures the effect of the construction of a rural road between 1999 and 2006. Since political patronage likely played a role in placement of roads as discussed in Section I.B, the results discussed below should only be considered suggestive.

Table C.1 estimates the effect of rural roads on agricultural wages. We find access to rural roads increases agricultural wages for both men and women by 1.2% and 2%, respectively. This result is consistent with the key takeaway from Asher and Novosad (2020) that rural roads induce movement of workers out of agriculture.

Table C.2 estimates the effect of rural roads on the village-level stock of combine harvesters. Using a village-panel regression with village and year fixed effects, we fail to find an economically or statistically significant impact on the village-level stock agricultural machinery across a variety of agricultural implements either on the extensive (Panel A) or the intensive margin (Panel B). In particular, we see no effect on the presence or number of combine harvesters. Next, using a household-panel regression with household and year fixed effects, we fail to detect an impact on the use of hired mechanized agricultural equipment at the household level (Table C.3).

Table C.1: Impact of roads on agricultural wage rates - REDS village panel

	Wage rate			Log wage		
	(1) Male	(2) Female	(3) Average	(4) Male	(5) Female	(6) Average
Treat X Post	0.722*	0.918***	0.856**	0.012*	0.020**	0.016**
	(0.420)	(0.349)	(0.352)	(0.007)	(0.008)	(0.007)
N	442	442	442	442	442	442
Control group mean	62.99	47.65	55.25	62.99	47.65	55.25
R^2	0.76	0.78	0.77	0.75	0.76	0.76

Notes: Table reports regression estimates showing the impact of receiving a PMGSY road on agricultural wage rates (in 2006 prices), in both levels and log wages. Sample used is village level panel from REDS 1999 - 2006. The coefficient "Treat X Post" takes the value 1 in the post (2006) year for villages that receive a road by 2006. Regressions include village and year fixed effects. Standard errors in parentheses clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table C.2: Impact of roads on village-level stock of agricultural machinery - REDS village panel

Panel A: Present (Yes = 1)

	Combines	Threshers	Tractors	Power tillers
Treat X Post	0.001	0.005	0.016*	0.021
	(0.009)	(0.015)	(0.010)	(0.014)
Sample mean	0.08	0.61	0.88	0.24
N	442	442	442	442
R^2	0.58	0.77	0.80	0.59

Panel A: Present (Yes = 1)

	Combines	Threshers	Tractors	Power tillers
Treat X Post	0.007	0.017	0.033	0.029
	(0.023)	(0.045)	(0.026)	(0.030)
Sample mean	0.10	1.40	2.48	0.42
N	442	442	442	442
R^2	0.54	0.76	0.90	0.54

Notes: Table reports regression estimates showing the impact of receiving a PMGSY road in agricultural machinery stock at the village level. Sample used is village level panel from REDS 1999 - 2006. The coefficient "Treat X Post" takes the value 1 in the post (2006) year for villages that receive a road by 2006. Regressions include village and year fixed effects. Standard errors in parentheses clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table C.3: Impact of roads on household use of mechanized agricultural equipment - REDS household panel

	(1) Used? (Yes=1)	(2) Log cost	(3) Log cost per acre	(4) Share of total cost
Treat X Post	0.002	-0.048	-0.046	-0.004
	(0.010)	(0.081)	(0.070)	(0.004)
Sample mean	0.67	4.77	3.91	0.15
N	6090	5548	5548	5474
R^2	0.57	0.60	0.62	0.69

Notes: Table reports regression estimates showing the impact of receiving a PMGSY road on household use of tractors, harvester, threshers or other mechanized equipment. Sample used is household panel from REDS 1999 and 2006. The coefficient "Treat X Post" takes the value 1 in the post (2006) year for households in villages that receive a road by 2006. Regressions include household and year fixed effects. Columns (1), (2) and (4) also control for cropped area. Standard errors in parentheses clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

D Harvest and Planting Dates

In this section, we evaluate the effect of rural roads on harvest and planting dates. We procure satellite-based measures of harvest (end-)dates and planting dates for the kharif season, aggregated up to the village level from 250 m pixel data. The planting and harvest dates are estimated using Enhanced Vegetation Index (EVI) data from MODIS and were validated using ground data.[1] Harvest (planting) date is measured as the median pixel value of the harvest (planting) dates within a 10 km buffer around the village. Unfortunately, these data are not available for states that followed population thresholds to determine rural road construction under PMGSY. Therefore, we are not able to use a regression discontinuity design. Instead, we estimate the following event study specification:

$$\Delta Date_{v,d,y} = \sum_{\tau, \tau \neq -1} \delta_\tau D_{t^0+\tau} + \lambda_v + \mu_{d,y} + \alpha_y X_v + \varepsilon_{v,d,y} \tag{1}$$

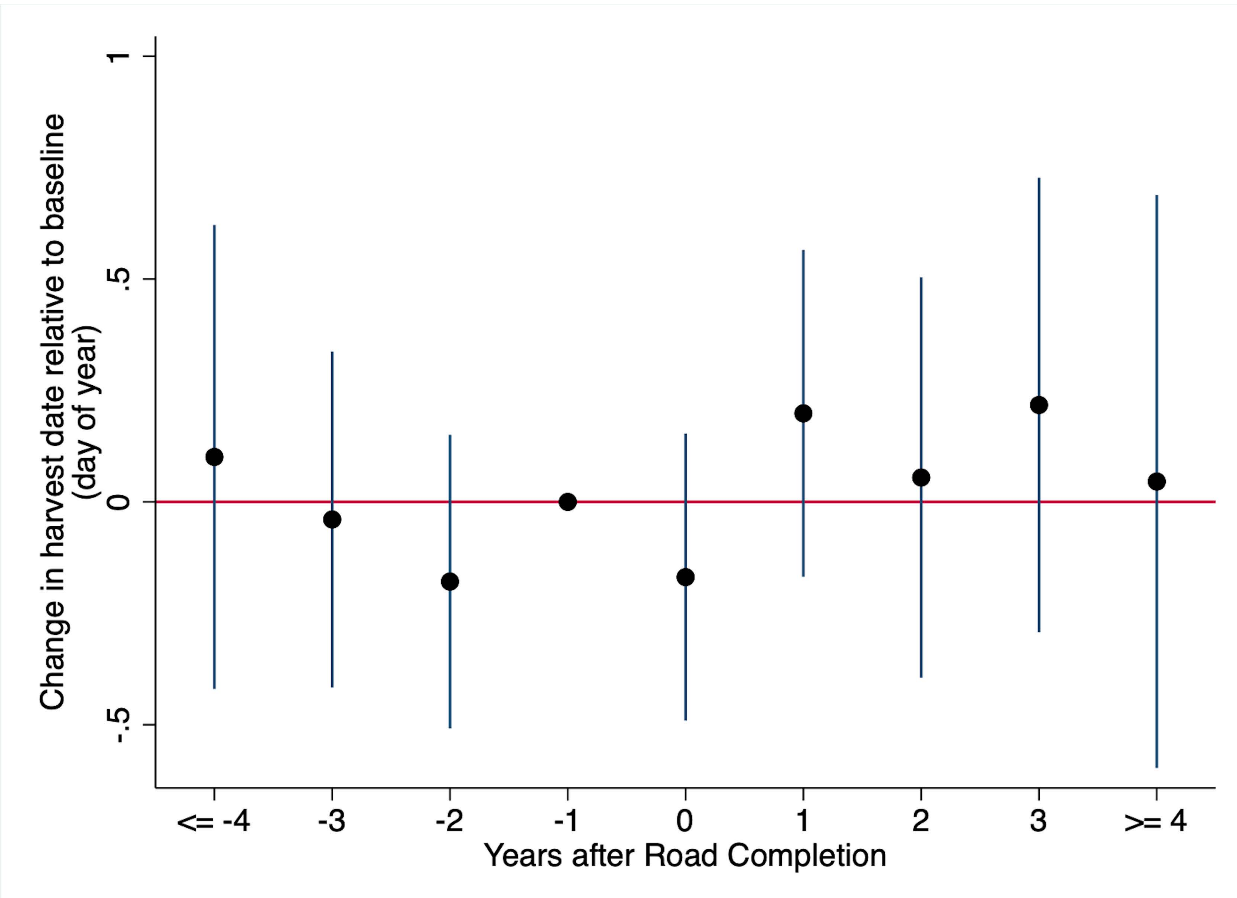
where $\Delta Date_{v,d,y}$ is the change in harvest or sowing date for village v, located in district d in year y from the baseline (2002). $D_{t^0+\tau}$ are event time indicator variables that capture the average treatment effect, where τ indicates the year relative to when a village receives access to a rural road, with the year prior to treatment being the excluded category. λ_v are village fixed effects and $\mu_{d,y}$ are district-by-year fixed effects. Village fixed effects control for time-invariant unobservables at the village level (e.g., soil type). District-by-year fixed effects control for time-varying district-specific confounders. For instance, the National Rural Employment Guarantee Scheme was rolled-out in a staggered manner across India between 2006 and 2008. Lastly, we include an interaction of baseline village characteristics X_v with year fixed effects α_y. Standard errors are clustered at the village level.

The identifying assumption here is that there exist no village-specific time-varying confounders that are correlated with both access to rural roads as well as local agriculture. E.g., if rural roads are placed in villages where agricultural activities are changing, our estimates would be biased. While lack of pre-trends would bolster our confidence in said assumption, if change in local agriculture and rural road construction were to occur simultaneously, our estimates will still be biased.

Figures D.2 and D.1 present our results. First, we don't find evidence for any pre-trends. Second, and more importantly, we fail to find evidence that access to rural roads affects either harvest or planting dates. The point estimates are small and statistically insignificant.

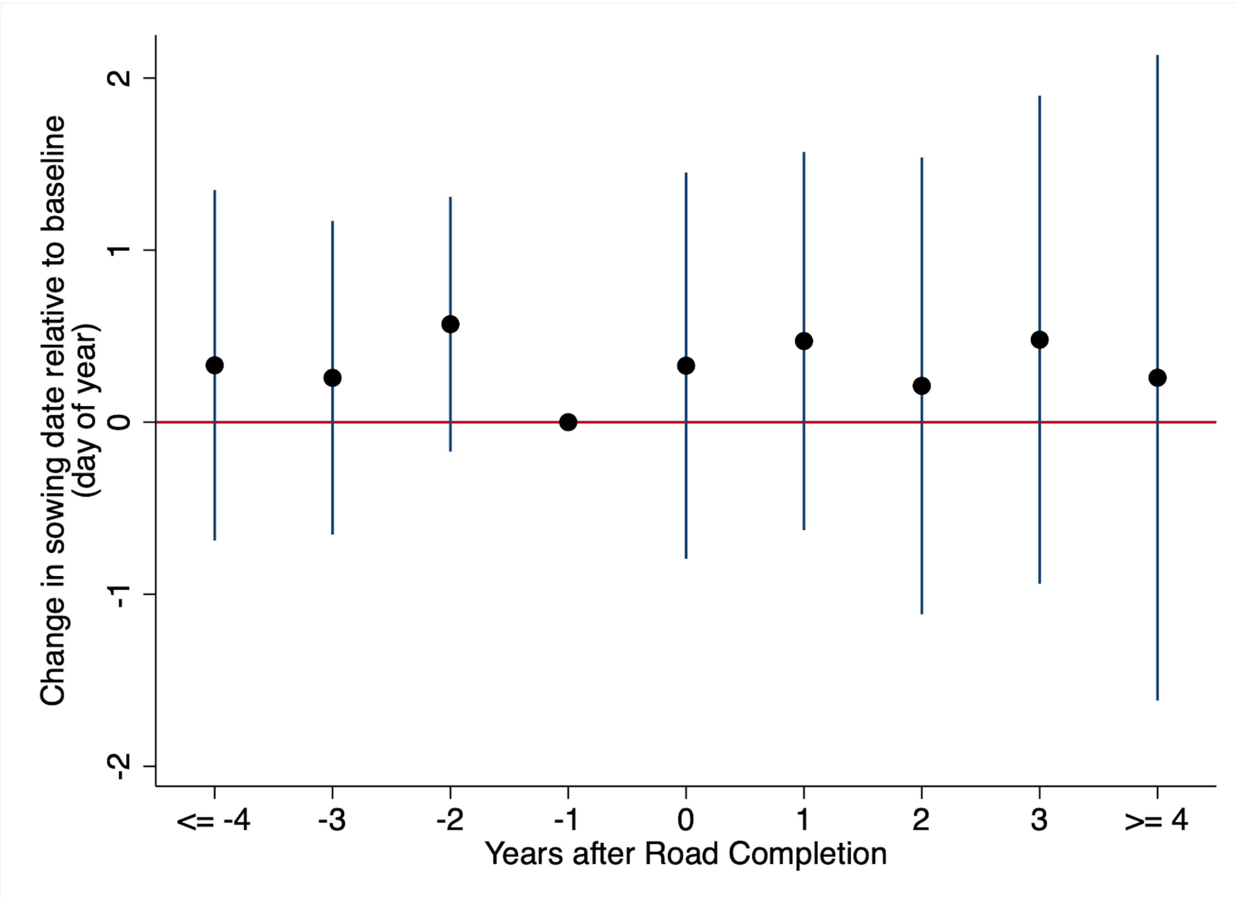
[1] We are grateful to Meha Jain, School for Environment and Sustainability, University of Michigan, for sharing these data.

Figure D.1: Event study estimates: Effect of access to rural roads on monsoon (kharif) harvest date



Notes: The graph shows event study estimates for the effect of new roads on monsoon harvest dates. Sample is limited to the states of Punjab, Haryana, Uttar Pradesh, and Bihar for which satellite-based sowing date measures are available. The sample period is 2003-2013. The dependent variable is the change in harvest completion day from the baseline (2002). The horizontal axis shows the event year relative to the year of road completion. Each point shows the coefficient and confidence interval on each event-time fixed effect relative to the omitted category which is the year before road completion ($t = -1$). All regressions include village FE, district × year fixed effects and the interactions of year fixed effects with baseline village characteristics and harvesting date/week in 2002. Standard errors in parentheses are clustered at village level.

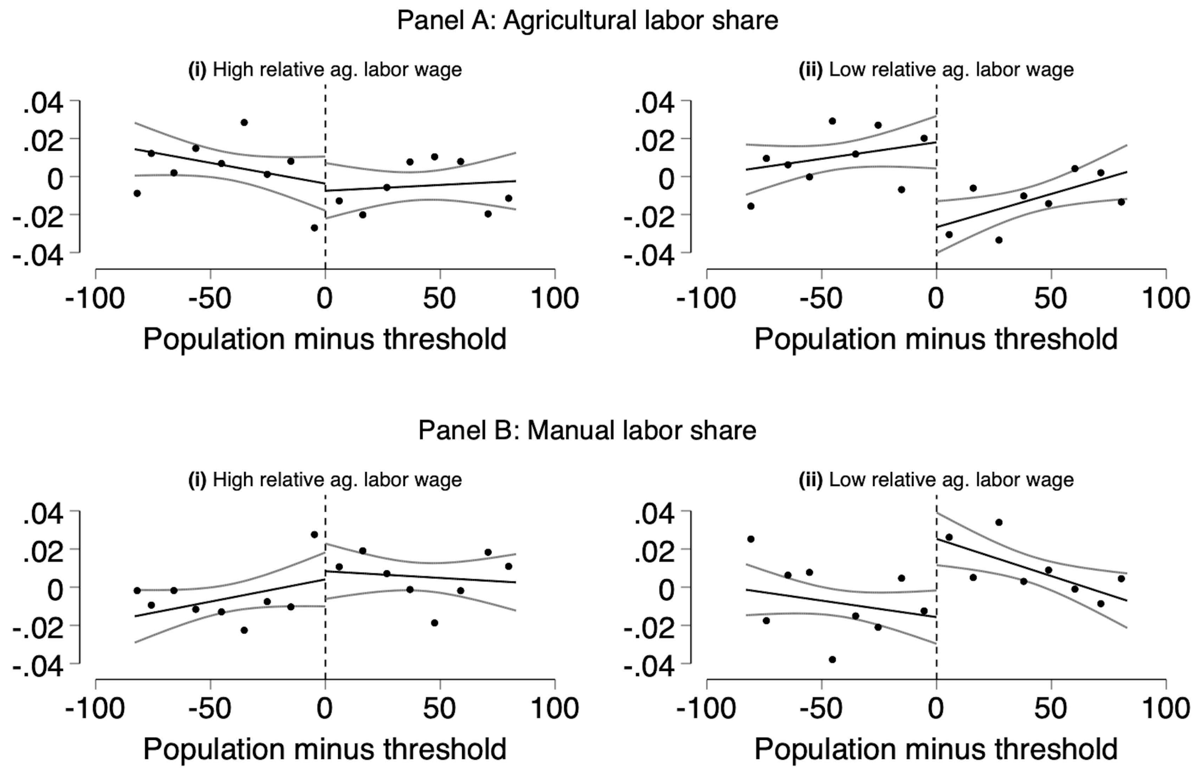
Figure D.2: Event study estimates: Effect of access to rural roads on monsoon (kharif) sowing date



Notes: The graph shows event study estimates for the effect of new roads on monsoon sowing dates. Sample is limited to the states of Punjab, Haryana, Uttar Pradesh, and Bihar for which satellite-based sowing date measures are available. The sample period is 2003-2013. The dependent variable is the change in day of sowing from the baseline (2002). The horizontal axis shows the event year relative to the year of road completion. Each point shows the coefficient and confidence interval on each event-time fixed effect relative to the omitted category which is the year before road completion ($t = -1$). All regressions include village FE, district $\times$ year fixed effects and the interactions of year fixed effects with baseline village characteristics and sowing date/week in 2002. Standard errors in parentheses are clustered at village level.

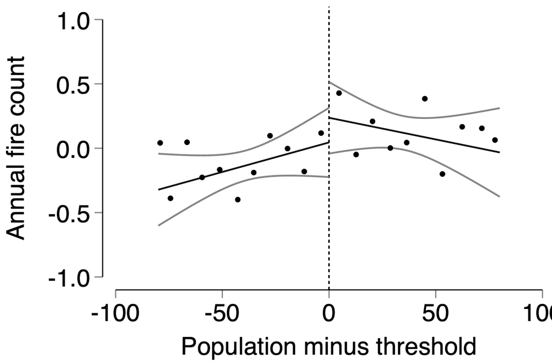
E Heterogeneity by Relative Agricultural Wage at Baseline

Figure E.1: Heterogeneity by relative agricultural labor wages at baseline: Impact of rural roads on share of village labor in agriculture and non-agricultural sectors

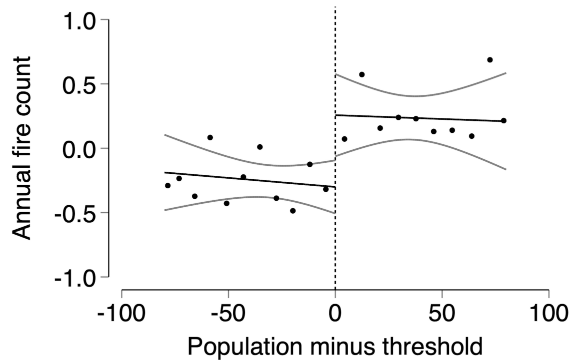


Notes: All panels show regression discontinuity estimates by plotting the residualized values of village-level share of manual labor in agriculture (Panel A) and non-agricultural sectors (Panel B), after controlling for all variables in the main specification other than population, as a function of the normalized 2001 village population relative to the threshold. Each point represents the mean of all villages in a given population bin. Figure (i) of each panel plots the regression discontinuity relationship for districts that had high (above sample median) agricultural labor wages relative to non-agricultural labor wage rates in rural areas at baseline. Figure (ii) of each panel plots the regression discontinuity relationship for districts with below median relative agricultural wage rates. The outcome variables are based on the Socioeconomic and Caste Census 2011-2012 (see (Asher and Novosad, 2020) for details). Baseline rural agricultural and non-agricultural daily labor wage rates are based on the 1999 - 2000 NSSO survey data (Round 55). Estimates in all panels control for district-threshold fixed effects, year fixed effects, and baseline village characteristics in 2001. Population is centered around the state-specific threshold used for road eligibility - either 500 or 1000, depending on the state. Standard errors clustered at the village level.

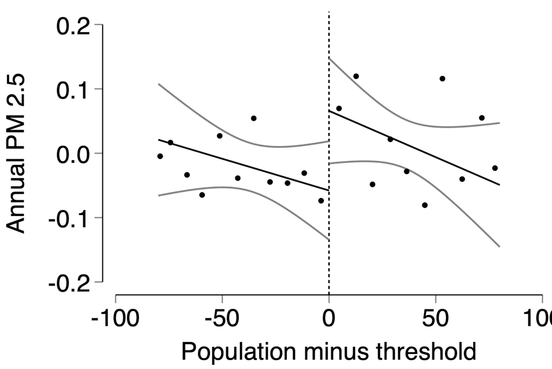
Figure E.2: Heterogeneity in impact of rural roads on annual agricultural fire activity by high versus low relative agricultural wage rates at baseline



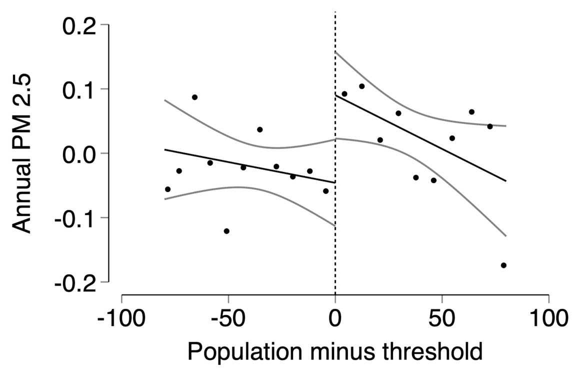
(a) Fires: High relative ag. labor wage



(b) Fires: Low relative ag. labor wage



(c) PM2.5: High relative ag. labor wage



(d) PM2.5: Low relative ag. labor wage

Notes: All panels show regression discontinuity estimates by plotting the residualized values of outcomes (after controlling for all variables in the main specification other than population) as a function of the normalized 2001 village population relative to the threshold. Each point represents the mean of all villages in a given population bin. Panels (a) and (b) show results for the annual number of fires between 2002 - 2013 , while (c) and (d) show the same annual average PM 2.5 ($\mu g/m^3$). Panels (a) and (c) plot the regression discontinuity relationship for districts which had high (above sample median) agricultural labor wages relative to non-agricultural labor wage rates in rural areas at baseline. Panels (b) and (d) plot the regression discontinuity relationship for districts with below median relative agricultural wage rates. Rural agricultural and non-agricultural daily labor wage rates are based on the 1999 - 2000 NSSO survey data (Round 55). Estimates in both panels control for district-threshold fixed effects, year fixed effects, and baseline village characteristics in 2001. Population is centered around the state-specific threshold used for road eligibility - either 500 or 1000, depending on the state. Standard errors are clustered at the village level.

Table E.1: Impact of rural roads on share of village labor in agriculture and non-agricultural sectors by relative agricultural labor wage rates

	Share of labor in agriculture		Share of non-agricultural manual labor	
	(1) High rel. ag. wage	(2) Low rel. ag. wage	(3) High rel. ag. wage	(4) Low rel. ag. wage
Road built	-0.029	-0.245***	0.030	0.210**
	(0.046)	(0.093)	(0.046)	(0.092)
N	5,402	5,483	5,402	5,483
Control group mean	0.49	0.46	0.45	0.46

Notes: This table shows the effect of rural roads on share of manual labor at village-level in agriculture and non-agricultural sectors. The outcome variables are based on the Socioeconomic and Caste Census 2011-2012 (see (Asher and Novosad, 2020) for details). "High rel. ag labor wage" sample consists of districts which had high (above sample median) agricultural labor wages relative to non-agricultural labor wage rates, while "Low rel. ag labor wage" sample are districts with below median relative agricultural wage rates. Wage rates are based on the 1999 - 2000 NSSO survey data (Round 55). Regressions include district-threshold fixed effects, year and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

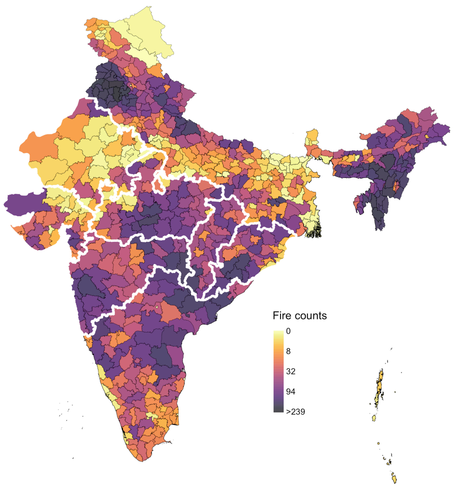
Table E.2: Impact of rural roads on annual agricultural fire activity and PM2.5 by high versus low relative agricultural wage rates at baseline

	High rel. ag. wage		Low rel. ag. wage	
	(1) Fires	(2) PM 2.5	(3) Fires	(4) PM 2.5
Road built	0.778	0.378*	3.195**	0.617*
	(0.721)	(0.220)	(1.439)	(0.367)
N	62,880	62,880	67,740	67,740
Control group mean	2.68	45.38	2.88	40.09

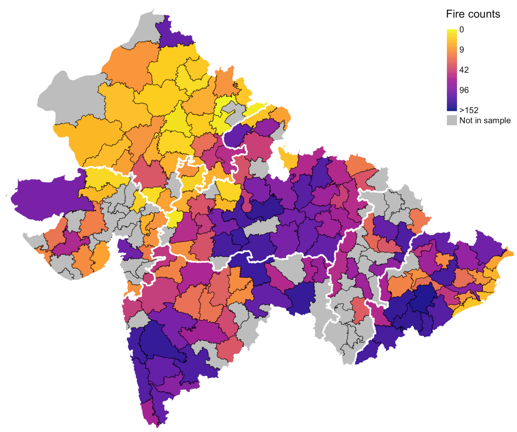
Notes: This table shows the effect of rural roads on village-level annual fire activity and PM 2.5 ($\mu g/m^3$). The sample consists of the panel of villages for the 5 year period from 2002 - 2013. "High rel. ag labor wage" sample consists of districts which had high (above sample median) agricultural labor wages relative to non-agricultural labor wage rates, while "Low rel. ag labor wage" sample are districts with below median relative agricultural wage rates. Wage rates are based on the 1999 - 2000 NSSO survey data (Round 55). Regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

F Heterogeneity by Crops Grown

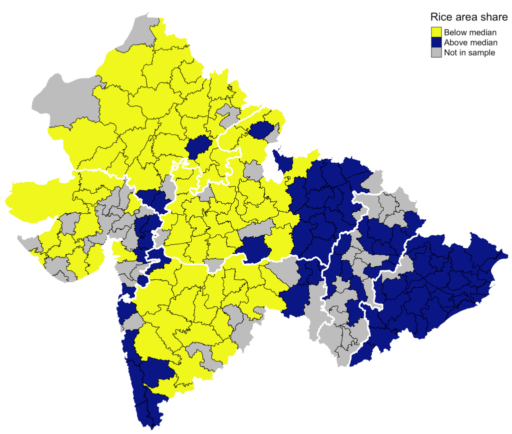
Figure F.1: Spatial distribution of average annual fire activity and baseline rice and sugarcane acreage shares



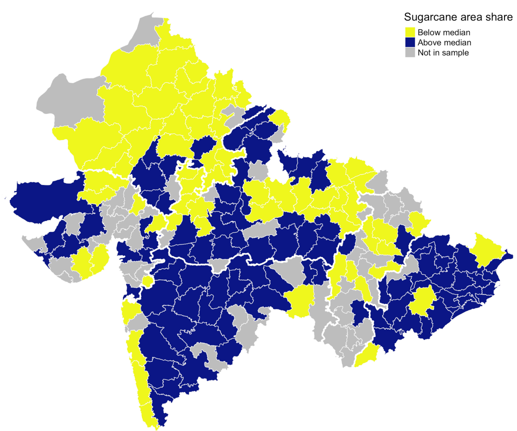
(a) Average annual fire counts across India



(b) Average annual fire counts in sample districts



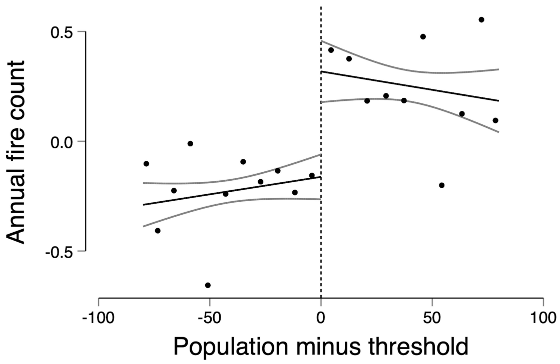
(c) Rice area share



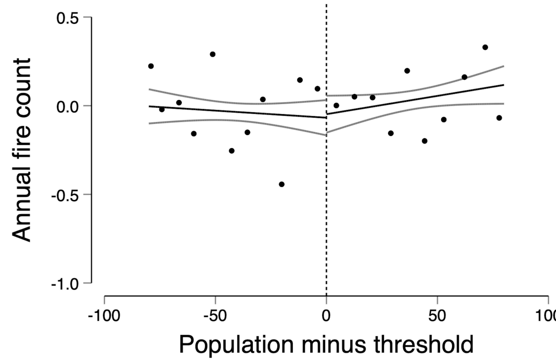
(d) Sugarcane area share

Notes: Panels (a) and (b) show the mean annual number of fire pixels detected in each district from MODIS satellite data for the period 2001 to 2013 for whole of India and within the sample districts, respectively. Panels (c) and (d) show districts with above/below sample median share of cropland under rice and sugarcane, respectively, at baseline (2001).

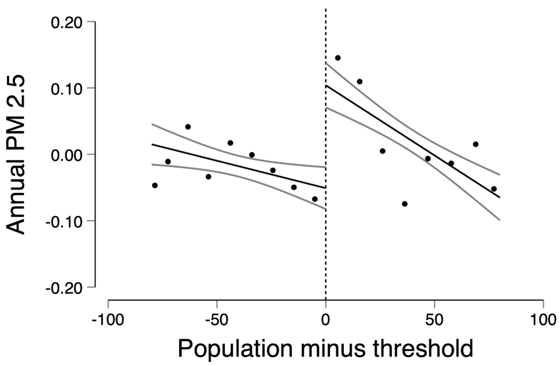
Figure F.2: Heterogeneity in impact of rural roads on annual agricultural fire activity and PM2.5 by high rice *or* high sugarcane districts versus low rice *and* low sugar districts



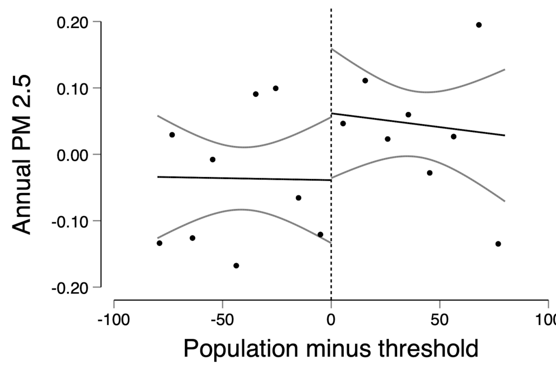
(a) Fires: High rice *or* high sugar



(b) Fires: Low rice *and* low sugar



(c) PM 2.5: High rice *or* high sugar



(d) PM 2.5: Low rice *and* low sugar

Notes: All panels show regression discontinuity estimates by plotting the residualized values of outcomes as a function of the normalized 2001 village population relative to the threshold (after controlling for fixed effects and all baseline variables in the main specification other than population). Each point represents the mean of all villages in a given population bin. Panels (a) and (b) plot the regression discontinuity relationship for annual fire counts, while panels (c) and (d) plot the regression discontinuity relationship for annual average PM 2.5 ($\mu g/m^3$). The sample used is the panel of villages for the period from 2002 - 2013. Panels (a) and (c) plot the regression discontinuity relationship for districts with high (above sample median) rice *or* sugarcane acreage share at baseline. Panels (b) and (d) plot the regression discontinuity relationship for districts with low (below sample median) rice *and* sugarcane acreage share at baseline. Estimates in all panels control for district-threshold fixed effects, year fixed effects, and baseline village characteristics in 2001. Population is centered around the state-specific threshold used for road eligibility - either 500 or 1000, depending on the state. Standard errors are clustered at the village level.

Table F.1: Impact of rural roads on annual agricultural fire activity and PM2.5 by high rice *or* high sugarcane districts versus low rice *and* low sugar districts

	High rice or high sugar		Low rice and low sugar	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	3.193**	1.048***	0.151	0.187
	(1.330)	(0.310)	(0.393)	(0.279)
N	92,772	92,772	31,920	31,920
Control group mean	3.30	38.86	1.43	52.93

Notes: This table shows the effect of rural roads on village-level annual fire activity and annual average PM 2.5 ($\mu g/m^3$). The sample consists of the panel of villages from 2002 - 2013. "High rice or high sugar" sample consists of districts with high (above sample median) rice *or* sugarcane acreage share at baseline (2001). "Low rice and low sugar" sample consists of districts with below median acreage share of rice and below median sugarcane acreage share at baseline. Regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table F.2: Impact of rural roads on annual agricultural fire activity and PM2.5 in high rice versus high sugarcane districts

	High rice		High sugar	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	3.560*	1.392***	3.245**	0.705**
	(2.022)	(0.453)	(1.315)	(0.305)
N	62,400	62,400	65,712	65,712
Control group mean	3.46	37.05	3.09	38.54

Notes: This table shows the effect of rural roads on village-level annual fire activity and annual average PM 2.5 ($\mu g/m^3$) from 2002-2013. The sample for columns (1) and (2) consists of villages in districts with high (above sample median) rice acreage share at baseline (2001). The sample in columns (3) and (4) consists of villages in districts with high (above sample median) sugarcane acreage share at baseline (2001). Regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table F.3: Impact of rural roads on annual agricultural fire activity and PM2.5 in high rice and low sugarcane versus high sugarcane and low rice districts

	High rice & low sugar		High sugar & low rice	
	(1)	(2)	(3)	(4)
	Fires	PM 2.5	Fires	PM 2.5
Road built	3.169	2.223**	2.641*	0.479
	(3.951)	(1.133)	(1.495)	(0.423)
N	27,060	27,060	30,372	30,372
Control group mean	3.80	39.61	2.96	42.76

Notes: This table shows the effect of rural roads on village-level annual fire activity and annual average PM 2.5 ($\mu g/m^3$) from 2002-2013. The sample for columns (1) and (2) consists of villages in districts with high (above sample median) rice acreage share but low (below median) sugarcane share at baseline (2001). The sample in columns (3) and (4) consists of villages in districts with high (above sample median) sugarcane acreage but low rice share at baseline (2001). Regressions include district-threshold fixed effects, year fixed effects and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

G Fire Activity in Cropland versus Non-Cropland Areas

We use remotely-sensed land cover classification data from 1995 to categorize fires into those occurring on cropland versus fires on non-cropland land cover classes around the PMGSY sample villages (Roy et al., 2016). The land cover data assigns 100 m X 100 m grid cells to 19 land use categories based on the International Geosphere-Biosphere Programme (IGBP) classification scheme. These categories include cropland, various type of forests, built-up land, and other categories. We overlay the land cover data on 1-km grid cells around each fire pixel (this reflects the resolution of the MODIS data we use to measure fires). Fire pixels that have most of their 1-km grid cell falling in the cropland land class are categorized as cropland fires. We find the increase in fire activity in response to rural roads is concentrated in cropland areas, with small and statistically insignificant effects in non-cropland areas (Table G.1).

Measurement error. It is important to note that classifying remotely-sensed fires by type of vegetation using land cover data as we do here is prone to measurement error. The first source of error arises due to the coarse spatial resolution of the fire data. The MODIS fire data we use only tells us the coordinates at the center of a 1 km grid cell that contains a fire – but cannot place the exact location of the fire within that grid cell. As a result, a fire occurring on cropland at the edge of a 1 X 1 km grid cell, with the majority of that grid cell consisting of non-cropland land classes, would be mislabelled as a non-cropland fire. Second, the accuracy of land cover data itself can be poor, particularly in regions with poor ground-level validation data and in fragmented landscapes such as those at forest edges (Zubkova et al., 2019). Finally, the land cover data we use reflects average land usage over the year in 1995. It will not capture seasonal changes in land use or the changes in land use that occurred in the years leading to the start of the PMGSY program in 2002. Studies in the remote sensing literature have noted that relying on land cover data to classify fires by type of vegetation (crops, forests, or other) is not validated and can result in errors of omission and commission (Roy and Kumar, 2017). Instead, a more accurate approach would rely on extensive labeled field-level data to serve as a training and validation data set (Hall, Argueta and Giglio, 2021). However, such ground truth data for agricultural fires in India, especially in the time period for our analysis, is unavailable. Given these limitations, we focus our primary analysis on all types of vegetative fires that are detectable from satellite data around the sample villages.

Table G.1: IV estimates of impact of rural roads on fires (count) in cropland areas versus non-cropland areas

	Annual fire count	
	(1)	(2)
	Cropland	Non-cropland
Road built	1.061*	0.248
	(0.558)	(0.387)
N	133,788	133,788
Control group mean	1.71	1.27

Notes: This table shows the IV estimates for the effects of rural roads on the annual count of fires in cropland (column (1)) versus non-cropland (column (2)) land cover categories. Classification of cropland versus non-cropland areas is based on pre-program (1995) land cover classification data from Roy et al. (2016). The sample used is the panel of villages for the period from 2002 - 2013. All regressions include district-threshold fixed effects, year fixed effects, and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

H Infant Mortality

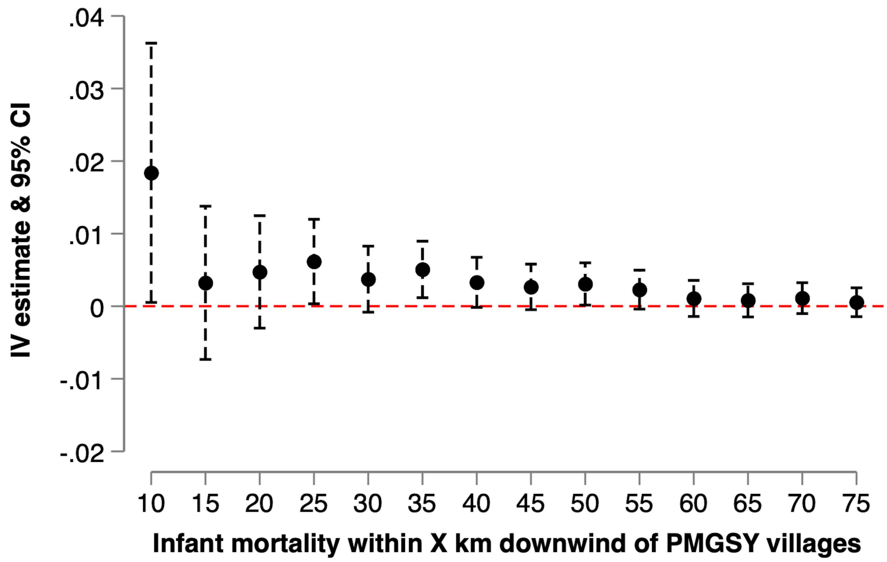
Does variability or predictability of annual wind direction play an important role for our infant mortality estimates? The year-on-year variation in annual wind direction is limited: the correlation between wind direction (in degrees) for PMGSY villages in year T and Year T-1 is extremely high (Table H.4, below). Consequently, the downwind or non-downwind categorization of NFHS villages in year T and Year T-1 is highly correlated. That is, most NFHS clusters that were categorized as downwind of PMGSY village based on wind direction in 2002, were also categorized as downwind based on wind direction between 2003 and 2013. Therefore, it is important to evaluate the role of both variability and predictability in annual wind direction for our infant mortality estimates.

First, we evaluate the role of variability in annual wind direction by i) assigning annual wind direction at baseline (2001) to all years (Table H.5, below) and ii) dropping NFHS clusters that switch categorization from downwind to non-downwind in any year between 2002 and 2013 (Table H.6, below). Our results are insensitive to both these robustness tests suggesting that inter-annual variability plays little to no role for our infant mortality estimates. Second, insofar as predictability is concerned, it is plausible that that households in downwind NHFS villages account for pollution-carrying winds from PMGSY villages, and engage in avoidance behavior (e.g., sorting). Therefore, avoidance behavior may be baked into our infant mortality estimates. However, we find no evidence for significant impacts on total population in villages downwind of the PMGSY village, either in logs or levels (Table H.7, below). The limitation of population growth as an outcome is that any impacts on net migration could be offset by changes to fertility and mortality. But such offsetting effects would cause changes in village demographics, which we can estimate in the comprehensive census data: downwind villages observed no changes to the age distribution or gender ratios for age cohorts between 11 and 60.[2] Furthermore, the NFHS asked mothers if they have always resided in the current location, and if not, how many years ago did they move to the current location. Therefore, as an indirect test of whether migration patterns are responsible for our infant mortality results, we restrict the NFHS sample to births to mothers who have resided in the current location since the start of PMGSY program in 2001. This restriction excludes 14.7% of our NFHS sample. However, the point estimates remain qualitatively similar (Table H.8, below). Taken together, these four pieces of evidence suggest that air pollution from fires in PMGSY villages that received roads did not lead to out-migration in downwind villages.

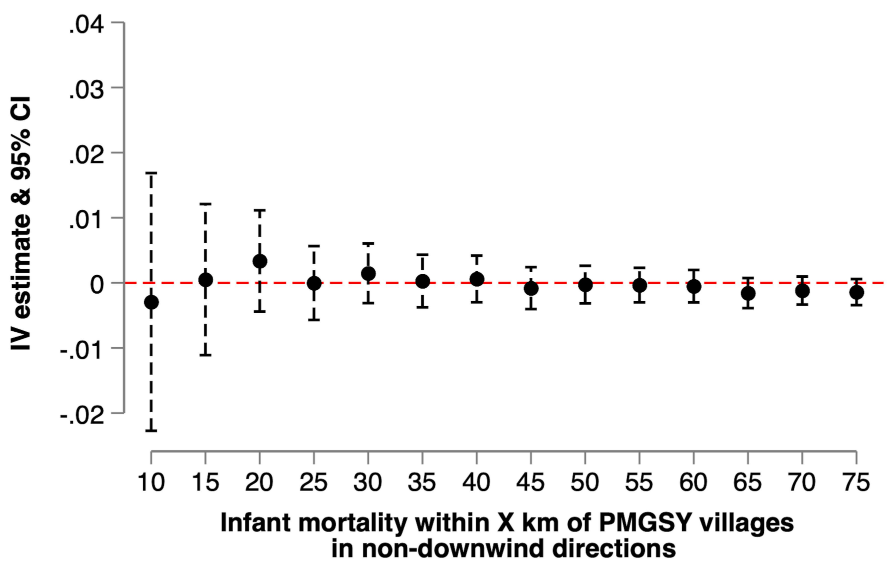
Of course, downwind households may engage in other types of avoidance behavior (e.g., staying indoors). Although our seasonal estimates suggest that because increases in particulate pollution occur for four to six months in a year, staying indoors might not be a feasible avoidance strategy.

[2] Note that these population and demographic data is not available in the NFHS data. Therefore, we use the 2011 Village Population Census and the 2011-12 Socioeconomic and Caste Census (SECC) to identify and examine changes in population or demographics in downwind villages.

Figure H.1: Impact of rural roads on infant mortality in NFHS villages located at varying distances in downwind versus non-downwind direction of PMGSY villages



(a) Downwind direction



(b) Other directions

Notes: This figure shows regression discontinuity IV estimates of receiving a new road on likelihood of infant mortality (death within 12 months of birth). Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. Each estimate is from a separate regression. The sample for each regression consists of all births recorded during 2002 - 2013 in the NFHS-IV survey that were located within the specified distance from the PMGSY villages in the downwind direction for panel (a) and in non-downwind directions for panel (b). The outcome variable "infant mortality" takes the value 1 if the child born died within the first year of birth, 0 otherwise. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. 95% confidence intervals are clustered at the PMGSY village level.

Table H.1: Differences between PMGSY villages that are matched vs. not matched with NFHS-IV villages within 50 km, and balance and falsification tests for the matched sample

Variable	(1) Full sample	(2) NFHS-IV matched sample	(3) No NFHS-IV match sample	(4) Difference of means	(5) p-value on difference	(6) RD estimate (matched sample)	(7) p-value on RD estimate
Primary school	0.959	0.958	0.969	0.011	0.07	-0.029	0.42
Medical center	0.166	0.164	0.180	0.015	0.09	-0.090	0.20
Electrified	0.430	0.435	0.367	-0.067	0.67	0.018	0.84
Distance from nearest town (km)	26.490	26.514	26.190	-0.324	0.78	-3.146	0.43
Land irrigated (share)	0.281	0.282	0.266	-0.016	0.81	-0.011	0.83
Ln land area	5.151	5.155	5.108	-0.047	0.03	-0.099	0.38
Literate (share)	0.457	0.459	0.436	-0.023	0.64	-0.002	0.95
Scheduled caste (share)	0.143	0.144	0.127	-0.017	0.16	-0.006	0.86
Land ownership (share)	0.733	0.734	0.719	-0.014	0.20	0.013	0.74
Subsistence ag (share)	0.435	0.433	0.459	0.026	0.13	0.043	0.35
HH income > INR 250 (share)	0.754	0.756	0.738	-0.018	0.90	-0.003	0.96
Annual fires (count)	0.685	0.668	0.892	0.224	0.27	0.195	0.51
Annual PM2.5 ($\mu g/m^3$)	35.886	35.943	35.190	-0.752	0.16	0.179	0.67
Down-wind IMR		0.067				0.002	0.63
Other directions IMR		0.072				-0.002	0.72
Down-wind PM 2.5		36.038				0.431	0.36
Other directions PM 2.5		39.254				0.901	0.15
N	11151	10310	841				

Notes: This table presents mean values for village characteristics, measured in the baseline period. The first eight variables are from the 2001 Population Census, the next three (below the first line) are from the 2002 BPL Census, while the final six variables are our outcome variables measured at baseline (2001). Columns 1-3 show the unconditional means for all villages, villages matched to NFHS-IV survey locations within 50 km, and villages with no NFHS-IV matches, respectively. Column 4 shows the difference of means across matched versus unmatched samples, and Column 5 shows the p-value for the difference of means between (2) and (3). Column 6 shows the regression discontinuity estimate for matched PMGSY villages, following the main estimating equation, of the effect of being above the treatment threshold on the baseline variable (with the outcome variable omitted from the set of controls), and Column 7 is the p-value for this estimate, using heteroskedasticity robust standard errors.

Table H.2: First stage and IV estimates of impact of rural roads on agricultural fires (count) and PM2.5 ($\mu g/m^3$) for PMGSY villages matched to NFHS-IV villages

	Road built	Annual fire activity	Annual average PM 2.5
	(1)	(2)	(3)
	1^{st} stage	IV	
Above threshold pop.	0.224***		
	(0.018)		
Road built		2.182***	0.724***
		(0.740)	(0.226)
Control group mean	0.260	2.895	42.841
Observations	129,379	129,379	129,379

Notes: This table shows the first stage estimates (probability of receiving a rural road) as well as the effects of rural roads on count of agricultural fires and PM 2.5. The sample consists of the panel of PMGSY villages matched to NFHS-IV villages from 2002 - 2013. "Above threshold pop." is an indicator for the village population being above the treatment threshold. Column (1) shows the first stage, with the dependent variable ("Road built") taking the value 1 if the village received a new road during 2002-2013, 0 otherwise. Columns (2) and (3) present the IV estimates of the treatment effects of new roads on annual fire counts and annual average PM 2.5 ($\mu g/m^3$), respectively. All regressions include district-threshold fixed effects, year fixed effects, and baseline control variables. Standard errors in parentheses are clustered at village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.3: Impact of rural roads on PM2.5 and infant mortality in NFHS villages located downwind versus all other directions from PMGSY with standard errors clustered at district level

	Downwind		Other directions	
	(1)	(2)	(3)	(4)
	PM 2.5	Infant mortality	PM 2.5	Infant mortality
Road built	1.0072***	0.0031**	-0.1551	-0.0005
	(0.3833)	(0.0016)	(0.2533)	(0.0015)
N	129,405	9,648,574	104,052	8,191,400

Notes: This table shows the effects of rural roads on PM 2.5 ($\mu g/m^3$) and infant mortality in NFHS-IV clusters – within 50 km of PMGSY villages – located in downwind (column 1) versus non-downwind (column 2) directions. Downwind NFHS-IV sample locations lie within 45-degrees on either side along the wind direction vector for each PMGSY village. Locations outside of this comprise the other directions sample. Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. The sample consists of all births recorded during 2002 - 2013 in the NFHS-IV sample. The outcome variables for columns (1) and (3) are the mean annual PM 2.5 averaged over all NFHS-IV locations in downwind and non-downwind directions, respectively. The outcome variable "infant mortality" takes the value 1 if the child born died within the first year of birth, 0 otherwise. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. Standard errors in parentheses are clustered at the district level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.4: Temporal correlation in wind direction at PMGSY villages and temporal correlation in likelihood that a NFHS village is downwind of a PMGSY village

	PMGSY villages	NFHS-IV clusters
	(1)	(2)
	Wind direction in $year\_t$	Downwind in $year\_t$
Wind direction in $year\_t-1$	0.6444***	
	(0.0051)	
Downwind in $year\_t-1$		0.7194***
		(0.0104)
N	137,184	69,701
R^2	0.401	0.512

Notes: This table shows the temporal correlation in wind direction at PMGSY villages and the temporal correlation in likelihood that a NFHS village is downwind of a PMGSY village from 2002-2013. In column (1) annual wind direction at a PMGSY village in year t is regressed on the lagged wind direction in year $t-1$. In column (2), an indicator taking the value 1 if an NFHS village is located downwind of a PMGSY village within 50 km of its location is regressed on the lagged value of the indicator. Standard errors in parentheses are clustered at the PMGSY village level in column (1) and at the NFHS village level in column (2). Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.5: Impact of rural roads on PM2.5 and infant mortality in NFHS villages located downwind versus all other directions from PMGSY villages based on baseline (2001) wind direction

	Downwind		Other directions	
	(1)	(2)	(3)	(4)
	PM 2.5	Infant mortality	PM 2.5	Infant mortality
Road built	1.0904***	0.0028*	-0.1696	-0.0002
	(0.2086)	(0.0015)	(0.1370)	(0.0016)
N	130,392	9,688,043	125,425	8,151,931
Control group mean	42.8941	0.0567	43.1935	0.0591

Notes: This table shows the effects of rural roads on PM 2.5 ($\mu g/m^3$) and infant mortality in NFHS-IV clusters – within 50 km of PMGSY villages – located in downwind (column 1) versus non-downwind (column 2) directions. Downwind is defined based on 2001 wind direction: NFHS-IV sample locations that lie within 45-degrees on either side along the wind direction vector for each PMGSY village. Locations outside of this comprise the other directions sample. Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. The sample consists of all births recorded during 2002 - 2013 in the NFHS-IV sample. The outcome variables for columns (1) and (3) are the mean annual PM 2.5 averaged over all NFHS-IV locations in downwind and non-downwind directions, respectively. The outcome variable "infant mortality" takes the value 1 if the child born died within the first year of birth, and 0 otherwise. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. Standard errors in parentheses are clustered at the PMGSY village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.6: Impact of rural roads on PM2.5 and infant mortality in NFHS villages located downwind versus all other directions from PMGSY - excluding NFHS villages that switched downwind or non-downwind categorization in any year between 2002 and 2013

	Downwind		Other directions	
	(1)	(2)	(3)	(4)
	PM 2.5	Infant mortality	PM 2.5	Infant mortality
Road built	0.9731***	0.0031**	-0.2206	-0.0003
	(0.1958)	(0.0015)	(0.1397)	(0.0015)
N	129,247	9,614,240	103,034	8,005,549
Control group mean	42.8951	0.0569	43.5783	0.0586

Notes: This table shows the effects of rural roads on PM 2.5 ($\mu g/m^3$) and infant mortality in NFHS-IV clusters – within 50 km of PMGSY villages – located in downwind (column 1) versus non-downwind (column 2) directions. Downwind NFHS-IV sample locations lie within 45-degrees on either side along the wind direction vector for each PMGSY village. Locations outside of this comprise the other directions sample. NFHS-IV villages that switched categorization from downwind to non-downwind in any year between 2002 and 2013 are excluded from the sample (1.23% of the main estimation sample). Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. The sample consists of all births recorded during 2002 - 2013 in the NFHS-IV sample. The outcome variables for columns (1) and (3) are the mean annual PM 2.5 averaged over all NFHS-IV locations in downwind and non-downwind directions, respectively. The outcome variable "infant mortality" takes the value 1 if the child born died within the first year of birth, 0 otherwise. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. Standard errors in parentheses are clustered at the PMGSY village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.7: Impact of rural roads on downwind village-level population, age distribution, and gender ratios

Panel A: Population growth (2001 - 2011)

	(1) Log	(2) Level
Road built	0.0165	3.5292
	(0.0151)	(7.6293)
N	11,147	11,147
Control group mean	6.54	1,094.48

Panel B: Age group share

	(1) 11 - 20	(2) 21 - 30	(3) 31 - 40	(4) 41 - 50	(5) 51 - 60
Road built	0.0013	-0.0018*	-0.0012	-0.0012	-0.0006
	(0.0013)	(0.0010)	(0.0010)	(0.0011)	(0.0010)
N	11,147	11,147	11,147	11,147	11,147
Control group mean	0.24	0.19	0.15	0.12	0.07

Panel C: Male share by age group

	(1) 11 - 20	(2) 21 - 30	(3) 31 - 40	(4) 41 - 50	(5) 51 - 60
Road built	0.0019	0.0002	-0.0016	-0.0014	-0.0017
	(0.0013)	(0.0014)	(0.0012)	(0.0014)	(0.0016)
N	11,147	11,147	11,147	11,147	11,147
Control group mean	0.52	0.52	0.51	0.52	0.52

Notes: This table shows the effects of rural roads on the population growth and demographic structure for villages downwind of PMGSY villages. Downwind outcomes are based on census locations (excluding the PMGSY sample villages) that lie within 45-degrees on either side along the wind direction vector for each PMGSY village. Outcome variables are the mean values averaged across all census villages downwind of each PMGSY village. Panel A shows the effect on average downwind village population from the Population Census 2011. Column (1) shows the outcome in log, and (2) shows the level effect. Column (1) includes control for baseline (2001) downwind log of population and (2) includes 2001 downind population in levels as control. Panel B shows the effect on average share of population in 10 year age groups in downwind villages, starting from 11 to 20 in Column (1) to 51 to 60 in Column (5). The outcome variables in Panel C are the average share of men in the population in downwind villages within each 10-year age group. Outcomes in Panels B and C are based on 2011-12 Socioeconomic and Caste Census (SECC). All regressions control for district-threshold fixed effects, and baseline controls. Heteroskedasticity robust standard errors reported in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table H.8: Impact of rural roads on PM 2.5 and infant mortality in NFHS villages located downwind versus all other directions from PMGSY villages for mothers that have resided in their current location since baseline (2001)

	Downwind		Other directions	
	(1)	(2)	(3)	(4)
	PM 2.5	Infant mortality	PM 2.5	Infant mortality
Road built	0.9450***	0.0035**	-0.1605	-0.0009
	(0.2367)	(0.0015)	(0.1420)	(0.0015)
N	113,793	8,432,179	103,471	7,165,257
Control group mean	42.7785	0.0554	43.6213	0.0571

Notes: This table shows the effects of rural roads on PM 2.5 ($\mu g/m^3$) and infant mortality in NFHS-IV clusters – within 50 km of PMGSY villages – located in downwind (columns 1 and 2) versus non-downwind (column 3 and 4) directions. Road completion is instrumented using an indicator for baseline (2001 Census) village population above the program threshold. Downwind locations are NFHS-IV sample locations that lie within 45-degrees on either side along the wind direction vector for each PMGSY village; locations outside of this comprise the other directions sample. The sample consists of all births recorded during 2002 - 2013 in the NFHS-IV locations in downwind and non-downwind directions, born to women who have resided in their current location (location where surveyed) since baseline (2001). The outcome variables for columns (1) and (3) are the mean annual PM 2.5 averaged over all NFHS-IV locations in downwind and non-downwind directions, respectively. The outcome variable "infant mortality" takes the value 1 if the child born died within the first year of birth, and 0 otherwise. All regressions control for district-threshold fixed effects, year fixed effects, and baseline controls. Standard errors in parentheses are clustered at the PMGSY village level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

References

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association*, 103(484): 1481–1495.

Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development." *American Economic Review*, 110(3): 797–823.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform." *The World Bank Economic Review*, 35(4): 845–871.

Giglio, Louis, Wilfrid Schroeder, Joanne V Hall, and Christopher O Justice. 2021. "MODIS Collection 6 and Collection 6.1 Active Fire Product User's Guide Version 1.0."

Hall, Joanne V, Fernanda Argueta, and Louis Giglio. 2021. "Validation of MCD64A1 and FireCCI51 cropland burned area mapping in Ukraine." *International Journal of Applied Earth Observation and Geoinformation*, 102: 102443.

Jagnani, Maulik, and Meera Mahadevan. 2023. "Women Leaders Improve Environmental Outcomes: Evidence from Crop Fires in India." *Mimeo*.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2): 698–714.

Roy, David P, and Sanath Sathyachandran Kumar. 2017. "Multi-year MODIS active fire type classification over the Brazilian Tropical Moist Forest Biome." *International Journal of Digital Earth*, 10(1): 54–84.

Roy, PS, P Meiyappan, PK Joshi, MP Kale, VK Srivastav, SK Srivasatava, MD Behera, A Roy, Y Sharma, RM Ramachandran, et al. 2016. "Decadal Land Use and Land Cover Classifications across India, 1985, 1995, 2005."

Van Donkelaar, Aaron, Randall V. Martin, Michael Brauer, N. Christina Hsu, Ralph A. Kahn, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and David M. Winker. 2016. "Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors." *Environmental Science and Technology*, 50(7): 3762–3772.

Zubkova, Maria, Luigi Boschetti, John T Abatzoglou, and Louis Giglio. 2019. "Changes in fire activity in Africa from 2002 to 2016 and their potential drivers." *Geophysical research letters*, 46(13): 7643–7653.