

Does BMI Predict the Early Spatial Variation and Intensity of Covid-19 in Developing Countries? Evidence from India*

Nidhiya Menon

Version: February 2021

Abstract

This paper studies BMI as a correlate of the early spatial distribution and intensity of Covid-19 across the districts of India and finds that conditional on a range of individual, household and regional characteristics, adult BMI significantly predicts the likelihood that the district is a hotspot, the natural log of the confirmed number of cases, the case fatality rate, and the propensity that the district is a red zone. Controlling for air-pollution, rainfall, temperature, demographic factors that measure population density, the proportion of the elderly, and health infrastructure including per capita health spending and the proportion of respiratory cases, does not diminish the predictive power of BMI in influencing the spatial incidence and spread of the virus. The association between adult BMI and measures of spatial outcomes is especially pronounced among educated populations in urban settings, and impervious to conditioning on differences in testing rates across states. We find that among women, BMI proxies for a range of comorbidities (hemoglobin, high blood pressure and high glucose levels) that affects the severity of the virus while among men, these health indicators are also important, as is exposure to risk of contracting the virus as measured by work propensities. We conduct sensitivity checks and control for differences that may arise due to variations in timing of onset. Our results provide a readily available health marker that may be used to identify and protect especially at-risk populations in developing countries like India.

Key Words: BMI, Covid-19, Spatial Variation, Intensity, SIRD Epidemiological Model, India

JEL Codes: I15, I18, O12, D83

Declarations of interest: none

* Thanks to the Associated Editor and two anonymous referees whose suggestions have substantially improved the paper. Thanks also to Sonia Bhalotra, Pushkar Maitra, and Yana Rodgers for their detailed comments. The usual disclaimer applies. Address for correspondence: Professor of Economics, MS 021, Department of Economics, Brandeis University, Waltham, MA 02453, USA. Tel. 781.736.2230. Email: nmenon@brandeis.edu.

1. Introduction and background

As Covid-19 permeates across the world, there is rising evidence that countries with weak health infrastructure, lack of universal health insurance coverage, higher proportion of vulnerable (elderly) populations, and a greater incidence of non-communicable diseases (diabetes, cardiovascular disease and high blood pressure) bear the brunt of the burden associated with the pandemic. Conditions in developing countries are even starker as in addition to the prevalence of these factors, these countries experience exceptionally high levels of environmental insults and are constrained in terms of resources that may be used to alleviate health and economic hardships (Bansal 2020). Given this, and in the interests of mitigating negative health consequences, is it possible to leverage a health marker that might reasonably predict the spatial variation and severity of the virus in poor countries? This research demonstrates that individual body mass index (BMI) of adults is an important correlate that may be utilized for this purpose as, net of a comprehensive set of characteristics, BMI is significantly associated with early measures of the regional spread and intensity of the pandemic in India, a country that is now second in the world in terms of the overall number of cases.

It has been noted that comorbidities may worsen the impact of the virus (Chatterjee and Jain 2020) and recent biomedical literature finds an association between BMI and Covid-19 risk. In particular, chronic inflammation and suppressed immune responses associated with obesity may worsen the severity of Covid-19, while “mechanical dysfunction” due to obesity can exacerbate respiratory infections and lead to secondary infections (Gao *et al.* 2020). Honce and Schultz-Cherry (2020) further notes that obesity disturbs cellular metabolism and weakens the body’s antiviral response. Yates *et al.* (2020) underlines the debilitating effect of obesity on immune responses while highlighting that weakened pulmonary function in obese populations can also increase the association between this factor and Covid-19. These biomedical studies show the presence of a

robust relationship between BMI and Covid-19 risk; however, as noted in Yates *et al.* (2020), selection bias is a concern given the reduced sample sizes and absence of a comprehensive set of controls for other risk factors that may also taint this association. This study contributes to the economics literature where evidence on the relationship between BMI and Covid-19 risk is scarce. More specifically, we examine BMI as a novel lens to investigate consequences in developing countries using the tools of economic modeling that are relatively more flexible than those in biomedicine are, and use a comprehensive dataset of individual, household and regional characteristics to control for omitted variables that, in their absence, may bias the predictive power of BMI.

The heterogeneity in the spread of the virus across areas may reflect two factors as noted in Desmet & Wacziarg (2020). The first is differences in timing: some regions contract the virus early because they are located near international airports for example, or near borders adjacent to countries where the disease is rampant. However, over time, the disease spreads and most regions will experience similar rates of infections, hospitalizations and mortality. The second factor underlines that heterogeneity in the spread of the virus is linked to variations in regional fundamentals that ensure that area-specific differences persist despite controls for elapsed time since onset. These fundamentals include risk factors such as the incidence of pollution, variations in demographic factors (population density and the proportion of the elderly), health capacity (per capita health expenditures, the number of doctors, and previous experience with respiratory illnesses), and political leaning and efficacy (the number of women elected to state assemblies in the most recent election). We find evidence that supports this second point of view.

Using nationally representative data and validated crowdsourced information on the early spatial variation and intensity of Covid-19 in India, we find that the BMI of adults aged 15-49 significantly predicts the likelihood that a district is denoted a hotspot district (districts with high

growth rates of cases or clusters of cases), the natural log number of confirmed cases at the district-level, the case fatality rate at the district-level, and the likelihood that a district is a red zone (where growth rates are exceptionally high or where there are multiple clusters). The predictive power of BMI is evident conditional on a set of individual and household characteristics that measure socio-economics and risk factors such as sanitation and access to drinking water, as well as differences in state-specific demographic and health infrastructure. We consider impacts by regional locations where the southern states (Kerala in particular) had been particularly successful in containing the pandemic in its early stages. We also consider specifications that control for migration. Other factors that we condition on include differences in testing rates across states and differences in the timing of onset as in Desmet & Wacziarg (2020). BMI continues to remain a significant correlate of the severity and spatial variation of the disease in India.

Since there is evidence of gender differences in various facets related to this pandemic (Galasso *et al.* 2020; Papageorge *et al.* 2020; Scavini and Piemonti 2020), we estimate specifications for adult women and men separately in order to understand whether this is true in our data as well. This is especially important in India as unlike in many other countries, the case fatality rate for women from Covid-19 has been found to be higher than the case fatality rate for men (Joe *et al.* 2020). Moreover, on average, BMI of women is higher than the BMI of men particularly in urban areas, and the country as a whole has the third highest number of overweight or obese people in the world with up to 20% of adults in this category (Dang *et al.* 2019). We find that while BMI is a significant predictor for women, including health indicators such as hemoglobin (HBA), a measure for high blood pressure, and a control for high glucose levels weakens the magnitude of BMI in the women-only sample. This is consistent with the fact that these non-communicable diseases are mechanisms for why BMI matters for women in predicting the spatial incidence and severity of the virus. These health indicators are explanatory for men as well although another factor that seems to

matter in weakening the effect of BMI for them is work propensity which possibly proxies for exposure to risk of contracting the virus. Finally, we find no evidence that BMI matters in any systematic way for children aged 0-14. This could be reflective of evidence that among children, the impacts of the virus has so far been mild (Centers for Disease Control 2020). It could also underline the fact that in child populations, BMI is an erratic indicator of health given growth spurts and rapidly evolving physiology (Vanderwall *et al.* 2017). Our research highlights that adult BMI is a significant correlate of the extent of the virus in India.

2. Empirical framework

2.1. Specifications

We leverage an empirical specification that builds on the commonly used Susceptible-Infectious-Recovered-Deceased (SIRD) epidemiological model that outlines pathways for a specific infectious disease and given population, based on Desmet and Wacziarg (2020). For a given outcome such as the log number of confirmed cases at a point in time, the rate of infection is influenced by a variety of individual and household, district and state-specific characteristics. Since our outcomes are at the district-level, we follow Foster-Johnson and Kromrey (2018) and conduct all empirical analyses at the district-level.¹ Consider the following:

$$y_{jk} = \beta_0 + \sum_{m_1=1}^{n_1} \beta_{m_1} \bar{x}_{jkm_1} + \sum_{m_2=1}^{n_2} \beta_{m_2} x_{jkm_2} + \sum_{m_3=1}^{n_3} \beta_{m_3} x_{km_3} + \varepsilon_{jk} \quad (1)$$

where j denotes a district and k denotes a state, and y_{jk} are the district-level outcomes considered including an indicator for a hotspot district, the natural log number of confirmed cases in the district, the case fatality rate in the district, and an indicator for a red zone district (these outcomes are

¹ In particular, Foster-Johnson and Kromrey (2018) note that when the data structure is of a “micro-macro” nature where the outcome is measured at a more aggregate group level while the explanatory variables are at either the individual or aggregate group level, OLS methods applied to group means with White’s heteroscedasticity correction for standard errors is appropriate and has more power for tests of the aggregate group level variables.

defined in detail below).² \bar{x}_{jkm_1} are the district-level means of the individual and household controls, x_{jkm_2} are other district-level variables, x_{km_3} are state-specific factors (enumerated below), and ε_{jk} is the district specific error term. Equation (1) is a fully saturated model and represents our preferred specification. We build up to this model however by sequentially adding regressors starting from a framework that includes only state fixed-effects, and then district means of individual, household and district characteristics, but no state-specific variables except for state fixed-effects. This specification is:

$$y_{jk} = \beta_0 + \sum_{m_1=1}^{n_1} \beta_{m_1} \bar{x}_{jkm_1} + \sum_{m_2=1}^{n_2} \beta_{m_2} x_{jkm_2} + \gamma_k + \varepsilon_{jk} \quad (2)$$

where γ_k denotes state fixed-effects and ε_{jk} is the corresponding district-level disturbance. We use ordinary least squares to estimate results for all outcomes including those that are binary in nature. Hence, we interpret results from linear probability models (LPMs) in the case of districts that are denoted as hotspots or red zones. We justify our use of LPMs in the case of these dichotomous dependent variables by relying on Caudill (1988) and Wooldridge (2010). As noted here, if there is a variable for group membership (like state fixed-effects), and if there is high or perfect correlation for members of a group for the dependent variable (states are likely to have multiple clusters of hotspot or red zone districts), then using LPMs in the case of dichotomous dependent variables provides an advantage over logit or probit models in that LPMs allow identification of the group variable.

2.2. Timing and sample selection

We estimate equations (1) and (2) on our data as of a specific point in time (April 27, 2020).

An issue to be cognizant of is that in this combined sample, spatial variation and measures of

² Following Desmet and Wacziarg (2020), we consider the natural log of (1+number of confirmed cases) so that we do not lose the extensive margin (districts and states where there are no cases, especially in the early days of the pandemic).

intensity of the disease could be correlated with timing. That is, regions where the disease arrived earlier will naturally advance on a different trajectory as compared to those where the disease arrived relatively more recently. In order to account for this, we include a comprehensive set of state-specific factors that have been noted to importantly influence the evolution of the disease. These include testing rates, political variables, and demographic and health infrastructure variables including the natural log of per capita health expenditure, natural log of the number of doctors, and the proportion of the population that is elderly (60 years and above); the full set of factors is discussed below.

Next, in order to directly address variation in the timing of onset, we consider samples of states that have the *same* elapsed time since onset. Onset is defined as the point (day) when the state reached a certain benchmark in the number of confirmed cases. The benchmark we use is the same as in Desmet and Wacziarg (2020), and which is commonly used in the epidemiological literature that defines onset as when the state reported at least 1 case per 100,000 people on any specific day. We then compare estimates in samples where all states experienced the same number of elapsed days since the threshold was reached. More specifically, we begin by examining the influence of BMI on outcomes in all states just before onset (days since onset = 0). Then we consider different cut-offs for these thresholds and examine results in samples in states one day after onset, five days after onset, and finally, 10 days after onset. Evaluating such samples in which states have exactly the same number of elapsed days minimizes the impact of time in influencing cross-state variation in the intensity and spread of the disease. However, there is a trade-off. Using earlier benchmark cut-offs allows a larger sample for estimation in which selection concerns are few. Using later cut-offs results in smaller samples where selection may be more of an issue since states with earlier onset are more likely to appear. We consider the full range of cut-offs that our data allow and report these results below. In our case, the broad consistency in magnitude of the parameter estimates of BMI

across these thresholds suggests that selection is less of a concern.

3. Data

We use a variety of data sources in this study. First, district-wise information on the number of confirmed cases, the case fatality rate, and testing rates at the state-level are obtained from the crowdsourced data publicly available at <https://www.covid19india.org/>. Other recent papers using this source includes Joe *et al.* (2020) which notes that these data are consistent with official information from the Government of India’s Ministry of Health and Family Welfare as well as international sources on India’s statistics such as those from Johns Hopkins University and online databases like Medicine (<https://ourworldindata.org/coronavirus>).³ Cases are confirmed following the administration of tests, and the testing rate is defined as the number of tests given per 100 people. The case fatality rate is estimated as the ratio of confirmed deaths in total confirmed cases, and is a widely used measure of mortality risk (Joe *et al.* 2020). The natural log of confirmed cases and the case fatality rate are two of the four outcome measures we analyze to understand the early spatial variation and intensity of Covid-19 in India.

In addition to this measure, we use two more variables to estimate the incidence of the pandemic. These include an indicator for a “hotspot” district and an indicator for a “red zone” district. “Hotspot” districts were denoted by the Government of India on April 15, 2020 as those that contributed to more than 80% of the caseload for the state, or districts in which the doubling rate was below 4 days (Ministry of Health and Family Welfare 2020). “Non-hotspot” districts are those where cases are present but lower than the thresholds above. Of India’s 735 districts, 170 were hotspots, 207 were non-hotspots, and 358 districts had no cases as of mid-April 2020 (Khanna and Kochhar 2020). The 170 hotspot districts were further demarcated into “red zones” and “orange zones” where the former denotes districts with a cluster of more than 15 cases or a district that has

³ Roser *et al.* (2020).

multiple clusters. Correspondingly, “orange zones” are hotspot districts with fewer than 15 cases. We analyze red zone districts separately to emphasize differences in intensity within all districts denoted as hotspots. Further, since these classifications of districts are as of April 15, 2020, we use data from the crowd-sourced site on cases and the case fatality rate until April 27, 2020 in order to span a time that is closest to the casting of these definitions.⁴

Our individual and household level determinants are obtained from the National Family Health Surveys of India from 2015-2016 (NFHS-4). We use a variety of individual controls in the adult samples including measures of BMI (constructed as weight in kilograms divided by height in meters squared (kg/m^2)), age, height (as a measure of long-term health), educational level, and number of children below five years. Household characteristics include religion, caste, a wealth index, measures of the age and gender of the household head, household size, rural/urban status, presence of electricity, type of toilet facility, primary sources of drinking water, and a measure of migration (years lived in current place of residence). Since the outcomes are at the district-level, following Foster-Johnson and Kromrey (2018), we use district means of these individual and household characteristics in all models, as noted above.

The socio-economic and demographic variables are from a different time-period than the Covid-19 data – there is no contemporaneous nationally representative survey available for India in the first quarter of 2020 as yet. Hence, is BMI in 2015-2016 a good predictor of BMI during the pandemic in early 2020? In order to answer this, we estimate state-level models where we regress an indicator for overweight or obese status constructed from BMI in NFHS-4 on an indicator for overweight or obese constructed from BMI in NFHS-3 (from 2005-2006) with robust White standard errors.⁵ In both the rural and urban samples, the coefficient on overweight-obese is positive

⁴ Another reason is that the case fatality rate at the district-level is reported only until this date.

⁵ BMI greater than or equal to $25.0 \text{ kg}/\text{m}^2$ denotes overweight or obese.

and highly significant (coefficient = 0.618, p -value = 0.000 in the rural sample; coefficient = 1.119, p -value = 0.000 in the urban sample). That is, BMI in 2005-2006 is a stronger predictor of BMI almost a decade later in 2015-2016. Furthermore, Dang *et al.* (2019) notes that there is strong persistence in these measures across space and time in India; in particular, transition matrices reveal that in the decade between NFHS-3 and NFHS-4, there has been rapid movement into the overweight-obese categories, but hysteresis in movement out of these groupings. Hence, yes, it is very likely that BMI in 2015-2016 is a good predictor of BMI 4-5 years later in the early months of 2020.

In terms of regional characteristics, we include air-pollution at the district-level as measured by PM_{2.5}, particulate matter of size 2.5 micrometers, a widely accepted measure of pollution in developed and developing contexts. This is in response to evidence that mortality risk from such pollution is particularly severe during the pandemic (Cole *et al.* 2020; Wu *et al.* 2020), and because of recent evidence that social distancing and the slowing of economic activity has reduced pollution and premature deaths attributable to air pollution (Dang and Trinh 2020; Muller *et al.* 2011; Cicala *et al.* 2020). The source of the PM_{2.5} data are satellite measurement estimates generated from aerosol optical depth information collected using techniques developed in Dey *et al.* (2012). Given that the impact of air-pollution may be mediated by rainfall and temperature, we include district-level measures of these weather variables in most models. The source is ERA-Interim daily data that is publicly available at <https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>.

Finally, we collect information from the *Handbook of Urban Statistics, 2019* (Ministry of Housing and Urban Affairs 2019) on state-level measures of population density in 2011 (2011 values from the most recent census is what is available) and the proportion of the population that is 60 years and above in 2017. Additional data from the *National Health Profile, 2019* (Ministry of Health and Family Welfare 2019) is obtained on state-level measures of per capita health expenditure in 2015-

2016, the number of doctors in 2018, and the proportion of respiratory cases in 2018. These controls together constitute a set of measures of a state's demographic profile and health capacity. In order to gauge political will and efficacy at the state-level, we include information from the *Election Commission of India*, various year, on the number of women elected to state assemblies in the most recent elections.

As mentioned above, data from the crowdsourced site on the spatial distribution and severity of cases is available at the district-level. Pollution and weather variables are also available at the district-level. These are then merged with individual and household information from NFHS-4 at the district-level. Lastly, the demographic and health capacity measures are merged at the state-level. Since the outcome variables are at the district-level, we take district-level means (constructed using sample weights) of these variables to create the final dataset for analysis as recommended in Foster-Johnson and Kromrey (2018). The total number of districts for which we have data is 631, and given the focus of the NFHS-4 (these are India's premier reproductive and health surveys), the proportion of women is 87.2% (12.8% are men). Sample sizes for the regressions at the district-level vary from these numbers depending on the completeness of information in the control variables that are included.

Figure 1 presents GIS plots for BMI and the natural log number of confirmed cases. It is clear that these variables are positively correlated, as districts with higher levels of BMI also appear to have a higher incidence of confirmed cases (the pair-wise correlation coefficient between these variables is 7.2% with a p -value < 0.01). The summary statistics of the district-level means of variables are presented in Table 1. The table is organized by outcomes, individual and household characteristics, and regional characteristics. Statistics are reported for the aggregate adult sample as well as samples demarcated by gender. We discuss estimates for all adults mainly but note differences between women and men (in column (7)) when these are of particular interest.

The first column indicates that as of end April 2020, about 18.5% of districts were hotspots and among these, 17.7% were red zones. The mean natural log number of confirmed cases is about 2.5 (12.2 cases) in the combined sample. The mean case fatality rate in the aggregate sample is 32.9%. Average BMI is 21.7 kg/m² for adults, which is in the normal range.⁶ These levels translate into 4.3% of women and 2.5% of men being classified as obese. The mean altitude adjusted hemoglobin level (HBA) in the sample is about 12.0 g/dl, the threshold for being categorized as anemic for women. Gender disaggregated values of HBA reveal that adult women are on average anemic in India (whereas men are not – the threshold for men is about 13.0 g/dl). Slightly more than 50% of the aggregate sample has a glucose level that is higher than the median value (relatively higher for men), consistent with the fact that India has one of the highest number of diabetics in the world (Gupta 2016). Other individual health characteristics that we condition on include the proportion of people who are medically diagnosed as having high blood pressure (8.6%), among whom women report somewhat higher values than men (8.9% versus 6.3%).

Considering other individual characteristics, average age is 29.9 years and 70.9% of the sample is married. The proportion of uneducated is 24.8%. Approximately 74.5% of households are Hindus whereas scheduled castes and scheduled tribes together make up 38.5%. About 40.3% of households fall in the poorest or poor categories based on the household wealth index while 21.4% of households are middle class. Most households are headed by men (86.9%), are rural (71.8%), and about 47.4% have access to clean drinking water. The average years lived in the place of residence is approximately 16.9 years consistent with evidence that (permanent) migration rates in India are relatively low (Munshi and Rosenzweig 2009).

The remaining descriptive statistics in Table 1 reveal that average natural log of PM_{2.5} is 3.6 micrograms per cubic meter (a PM_{2.5} level of 37.0 micrograms per cubic meter – more than double

⁶ BMI less than 18.5 denotes underweight, between 18.5 and 25 denotes normal weight, above 25 but below 30 denotes overweight, and 30 or above denotes obese.

the United States Environmental Protection Agency (EPA) standard).⁷ Other summary statistics reported at this level include those for weather (rainfall and temperature). We end by briefly noting descriptive estimates of the state-level measures on testing rates and health indicators. The average testing rate was about 32.7% by end April 2020 with significant variation across states. Kerala stands out for its early success in containing the disease due to its relatively high testing rates (Chatterjee and Jain 2020; Vibhute and Chattopadhyay 2020). Mean log number of women elected to state assemblies in the most recent elections is about 17. Mean log per capita health expenditure in 2016 is approximately Rupees 1,331 (US dollars 20.5), and the mean log number of doctors at the state-level is about 10.5 (this translates into 36,315 doctors but there is a literature on quality of doctors - see Das (2007)). Approximately 4.0% of all states have experienced respiratory cases in 2018 and the mean proportion of the elderly (60 years and above) is 8.3%. In summary, these measures indicate that while Indian states do not have an especially vulnerable population in terms of the elderly, their health infrastructure (as measured by per capita health expenditure and testing rates in particular) in the early days of the pandemic was, with a few exceptions, relatively low (consistent with evidence in Ray (2020)).

4. Results

We discuss results in Tables 2-5 in this section where only the key parameter of interest (BMI) is noted. The full set of results is reported in Appendix Tables 1, 2 and 3.

4.1. Hotspot districts

The association between BMI and districts denoted as hotspots is shown in Table 2. As discussed above, hotspot districts are those with significant numbers or significant growth rate of cases as of mid-April 2020. Panel A reports results for all adults aged 15-49 whereas Panel B and Panel C report results demarcated by gender. Each column in Table 2 reflects the inclusion of

⁷ See <https://www.epa.gov/criteria-air-pollutants/naaqs-table>. Accessed on June 26, 2020.

different sets of characteristics as stated at the bottom of Table 2, with column (8) reporting the fully saturated specification that includes, among others, weather, health and state-specific characteristics.

Focusing on Panel A first, it is clear that adult BMI has a positive and significant influence on the district being denoted as a hotspot. In the most parsimonious model that includes only state fixed-effects in column (1), the coefficient on BMI indicates that a one-unit increase in district mean BMI (about a 5% increase in district mean BMI) results in a 16.5 percentage point rise in the probability that the district is a hotspot. Inclusion of the district specific pollution and weather specific measures increases the magnitude of this effect to 25.1 percentage points, but including controls reduces the estimate to 10.9 percentage points. Restricting the sample to control for migration does not affect this parameter by much (impact is insignificant in the sample of Southern states that are known to have better health infrastructure). This remains true even when we condition on state differences in testing rates and demographic and health infrastructure measures, as well as health conditions such as hemoglobin levels and indicators for high blood pressure and high glucose levels that are often associated with unhealthy levels of BMI. In the fully saturated model of column (8), a one-unit increase in district mean BMI is associated with a 14.4 percentage point increase in the probability that the district is a hotspot.⁸

Panel B reports the results for women aged 15-49 and in general, many of the patterns in Panel A resonate here. The estimate in column (1) indicates that for a unit increase in the district mean BMI (about a 4.5% increase in district mean BMI), the probability that the district is a hotspot is 12.9 percentage points. The coefficient declines in size with the inclusion of controls in column (3) but again note that inclusion of subsequent variables for testing rates and state-level measures of health capacity does not dramatically change the magnitude of this coefficient. Interestingly,

⁸ In addition to the log number of women elected to state assemblies in the most recent election, we included an indicator measuring affiliation of the state's ruling party with the party ruling the central government to proxy for the dedication by which national public health guidelines would have been adhered to at the state-level. The coefficient and significance of BMI remain the same.

inclusion of the individual health conditions for women somewhat reduces the size of the BMI variable in column (8) in comparison to columns (1) and (2) suggesting that for them, these variables are weak mechanisms that explain the association between BMI and hotspot districts at the district-level. Patterns for men in Panel C are similar except that in comparison to column (1), the decline in magnitude of the coefficient on BMI in column (8) is larger than in the case of women. This is suggestive that hemoglobin, high blood pressure and glucose levels in men have greater explanatory power in linking BMI and hotspot districts.

4.2. Log number of confirmed cases

We examine the spatial intensity of Covid-19 by focusing next on the district-level measure of confirmed cases. These results are reported in Table 3 that has an organization structure similar to that in Table 2. Conditioning only on state fixed-effects, the estimate in column (1) of Panel A indicates that for a unit increase in district mean BMI (again, about a 5% increase in district mean BMI), the natural log number of confirmed cases increases by 80.9%. Including the pollution and weather controls increases the magnitude of this association to 83.1%, and conditioning on state testing rates results in a substantial decline to (a marginally significant) 21.4%. The last column of Table 3 in Panel A indicates that with controls for statewide differences in population and age-structure variables as well as measures of health capacity and political efficacy, a unit increase in district mean BMI generates a 36.5% increase in the number of confirmed cases (about 2.5 additional cases at the mean). Given the decline in the size of the coefficient in column (8) relative to column (1), the initial association between BMI and number of confirmed cases in Panel A is likely reflective of state-level differences in demographic and health infrastructure as well as individual level measures of health.

Disaggregating the combined sample by gender reveals that in general, impacts of BMI on number of confirmed cases is somewhat stronger among adult men except in the last two columns.

In particular, in the specification that conditions on test rates and measures of health capacity at the state-level in column (7), a unit increase in district mean male BMI generates a 29.8% increase in the number of confirmed cases (about 2.1 additional cases at the mean). The corresponding estimate for women is 40.1% (2.8 additional cases at the mean). Including the health indicators for hemoglobin, blood pressure and glucose does not absorb the significance of BMI in the women or male regressions in column (8), thus echoing the findings in Table 2.

4.3. Case fatality rate

Next, we examine the influence of BMI on the case fatality rate, which has been argued to be a better measure of disease severity as compared to the mortality rate (Battegay *et al.* 2020). Table 4 reports these results and indicates that estimates are measured with precision only in two columns in the male sample in Panel C. Focusing on column (3) in this panel that includes weather and individual and household characteristics, the reported coefficient indicates that for a unit increase in BMI, the case fatality rate for men rises by 12.8%. Given that the average case fatality rate in India near end April 2020 in our data is 32.9%, this denotes approximately a 38.9% increase. However, including controls for differences in testing rates, demographic and health infrastructure, individual health conditions and the number of women elected renders the coefficient insignificant in column (8). Overall, it is possible that the lack of significance in Table 4 reflects India's age-structure which has a low proportion of elderly people who have been found to be especially susceptible and under-reporting of deaths in the early days of the pandemic (Malani *et al.* 2020). Furthermore, we note that in comparison to the other outcomes, information on the case fatality rate is available for a relatively lower number of districts (203 districts) from the crowdsourced site. The smaller number of observations may contribute to the overall lack of significance in Table 4. Finally, while the case fatality rate is not the same as the death rate, we note that recent evidence for the United States also finds little correlation between BMI (obesity) and death rates (Knittel and Ozaltun 2020).

4.4. Red zones

Table 5 shows results for the association between BMI and red zones. Considering impacts in Panel A and focusing on the fully saturated model in column (8) reveals that a unit increase in district mean BMI generates a 7.5 percentage point increase in the likelihood that the district will be a red zone. Disaggregating by gender reveals that much of this influence arises from the men's sample (although the coefficient is only marginally significant). In comparison to the results in Table 2, again there is little evidence that BMI is an important determinant of red zone status among southern states. Finally, restricting the sample to those who have been resident in the same place for ten or more years indicates that in all three samples, BMI continues to be a precisely measured factor associated with a district that is flagged as a red zone.

In summary, these results underline that net of a comprehensive set of controls, BMI is a significant correlate of the early spatial variation of Covid-19 across districts in India. Conditioning on individual health measures that BMI influences (hemoglobin, high blood pressure, elevated glucose levels), while significant in their own right, do not fully subsume the significance of BMI's effect in the case of the spatial distribution of districts demarcated as hotspots or red zones, or in the case of the natural log number of confirmed cases. Disaggregation by gender reveals that this is true for women and especially for men.

5. Smoking, work propensity, and heterogeneity checks

5.1. Smoking and work propensity

Do other factors play a role in determining the association between BMI and the outcomes we analyze? In order to study this more closely, we consider differences in smoking rates by gender. Estimates reveal that while 32.4% of men smoke, only 1.9% of women do.⁹ Further, the number of cigarettes (and other things) smoked in the last 24 hours is highly correlated with BMI in men

⁹ This includes smoking cigarettes, pipes, cigars, bidis (less sophisticated/domestic form of a cigarette) or other – cigarettes and bidis make up the largest proportion.

(coefficient = 0.028, p -value < 0.01), but uncorrelated with BMI in women. In order to ascertain whether smoking is the omitted variable in the male sample, we re-ran the male regressions including this measure of smoking. Results for men in column (8) of Tables 2 - 5 remain virtually unchanged.¹⁰

If differential smoking rates are uninformative, could exposure to risk as proxied by men's propensity to work be another factor? The data reveal that while 92.1% of men are currently working, the comparable proportion for women is only 26.8%. We find that including this measure of work in the men's sample weakens the significance of the BMI variable, mainly in the case of the natural log of confirmed cases. That is, men have somewhat greater exposure to risk of contracting the virus given their higher work propensities, and in the absence of this control, BMI, which is positively correlated with work for men, reflects these associations.¹¹ There is evidence in favor of gender differentials in other countries as well (Galasso *et al.* 2020; Papageorge *et al.* 2020). These papers document differentials in response behaviors whereas the distinguishing feature of our study is that we note gender variations in an underlying factor that is correlated with the incidence and intensity of the regional spread of this disease.

5.2. Obese indicator and differences in testing rates

The specifications above use a linear form of BMI. In Panel A of Table 6 we report results when an indicator for obesity is utilized instead. We report results for the comprehensive model that includes all controls only. It is clear that this indicator also has significant predictive power when it comes to measuring the spatial variation in the incidence of Covid-19. In particular, the estimate in column (1) indicates that a one-unit increase in the district mean level of obese (about a 32-fold increase) produces a 6-fold increase in the likelihood that the district will be demarcated a hotspot.

¹⁰ These results are available on request.

¹¹ The pair wise correlation coefficient of BMI and an indicator for currently working in the male sample = 0.046 with a p -value < 0.01. The regression results that include the work variable for men are available on request. Including an indicator that the woman is currently working does not change the BMI coefficient.

In keeping with this, a similar increase in obese generates more than a fourteen-fold increase in the number of confirmed cases. Large and positive effects are evident in column (4) for red zones as well.¹² This strong impact of obesity is entirely consistent with recent evidence in the bio-medical literature noted in Gao *et al.* (2020), Honce and Schultz-Cherry (2020), and Yates *et al.* 2020.

A factor that is important in the case of India is variation in state-level testing rates.

Although we condition on the state's testing rate in the models above, we explicitly examine this factor in detail in Panel B of Table 6. We accomplish this by creating an indicator for states that have a testing rate that is at the 90th percentile or higher, and then interacting the BMI variable with this indicator to analyze the differential influence of testing in such states. The results in Panel B of Table 6 indicate that while BMI in of itself continues to exert a significant positive influence on the outcomes we consider, the influence in states that have relatively high testing rates is on net no different.

5.3. Conditioning on days from onset

As discussed before, part of the district-level variation in outcomes may be reflective of timing issues. That is, the severity of the disease appears higher in some districts perhaps because cases began there earlier. To address variation in timing of onset, we consider sample of states that have the same length of elapsed time since onset. As before, we note that using a benchmark just before onset allows a larger number of observations that are less likely to be selected. Alternatively, samples are smaller and more likely to be selected along unobservable dimensions the longer the time window since onset. We present results for adults for various days since onset for the most complete specification in Table 7. Results in Panel A underline the positive influence of BMI on the spatial variation and severity of the disease, and are reflective of those reported earlier. Results in the subsequent panels of Table 7 report coefficient estimates that are in the same ballpark magnitude

¹² Cut-offs are slightly lower for Asian populations (BMI above 27.5 denotes obese) but applying these did not change the results substantially. Please see results reported in Appendix Table 4.

to those in Panel A but measured with more noise given the smaller sample sizes. For example, the parameter in column (1) of Panel A (just before onset) indicates that a unit increase in district mean BMI is associated with a 14.4 percentage point increase in the likelihood that the district is a hotspot. The magnitude of this parameter remains about the same when we consider 1 day since onset or 5 days since onset (measured imprecisely), and falls to 9.2 percentage points on considering 10 days since onset (again, measured imprecisely). The relative stability in the size of the estimate is evident in column (2) across Panels A and B, and only slightly different in column (4) when we analyze red zone before onset (rounded coefficient = 0.1), one day from onset (rounded coefficient = 0.1), and five days from onset (rounded coefficient=0.2). We conclude that while timing may be an issue as in Desmet and Wacziarg (2020), in our case, its influence is of less significance perhaps because we consider samples from a relatively early period of the outbreak in India.

5.4. Lockdown orders

India was ordered into a nation-wide lockdown from March 25, 2020 onwards. Although this was sudden, strict and largely unanticipated, it is hard to identify impacts of this policy legislation as there is no variation in timing across states.¹³ The first phase of the lockdown extended until mid-April, and there were three extensions until end May 2020. India started loosening restrictions in a phased manner from June 1, 2020. However, it is not possible to exploit differentials in lockdown removals to identify impacts either since it is the worst affected regions that were under extended stay-at-home rules. This simultaneity invalidates empirical exercises given the endogeneity inherent in evaluating regions where lockdown orders were lifted. But, given evidence that such laws have resulted in fewer cases and a slower rise in the number of cases in the United States and overseas (Dave *et al.* 2020; Fang *et al.* 2020), we hypothesize that India's nation-wide lockdown too must have resulted in a similar pattern. This implies that we have a conservative

¹³ In terms of its stringency, India's lockdown scored 100/100 in terms of the Government Response Stringency Index developed by the University of Oxford (Chatterjee and Jain 2020).

bias in the estimates reported in Table 2 through Table 5, that is, in the absence of the lockdown, BMI would be an even stronger correlate of the early spatial variation and intensity of the disease.

5.5. Differences by education, caste status, rural/urban

We evaluate differences in the above results by education, caste and rural/urban status. These results are reported in Table 8 and demonstrate that in general, BMI is a strong predictor of the outcomes we consider primarily among those with some level of education living in urban areas. This is as expected since BMI is highest in urban areas of India among those with some level of schooling and thus higher levels of income (Dang *et al.* 2019). BMI in the early days of the pandemic is mostly not measured with precision among the uneducated (except in the case of hotspots) or among those of lower caste status (except in the case of red zones).¹⁴

5.6. Falsification/sensitivity checks

We cannot implement a standard falsification test given the nature of the variables in this study, but we check to ascertain that the predictive power of BMI varies as expected conditional on the relative anchoring point in the underlying distribution of the outcome variables. The outcome we focus on here is the measure for intensity – natural log of the confirmed number of cases. In the absence of omitted variables that are simultaneously correlated with BMI and this measure, the predictive power of BMI should be relatively greater at points in the distribution where intensity is higher. We report results in Table 9 that confirm that this is the case. Column (1) reports the full sample results for the natural log of the confirmed number of cases (this is the same as in Panel A column (8) of Table 3). It is clear that in comparison to the estimate in column (1), the coefficient on BMI increases in the subsequent columns of Table 9 that condition on points in the distribution with higher intensity of the log number of confirmed cases (slight dip in going from the 10th percentile to the lower quartile value, but still in the same approximate vicinity). At the 56th

¹⁴ The impact of BMI in rural areas cannot be identified for the case fatality rate as the sample size is too small.

percentile, the coefficient on BMI is about 11.2% larger than in the full sample. We conclude from these results that the effect of BMI on the measure of intensity varies as expected, thus indicating that the influence of omitted variables is likely small.

5.7. Children's sample (ages 0-14 years)

We note that the same framework of models was applied to a sample of children aged 0-14 years in order to evaluate the effect of BMI on the spatial distribution of Covid-19 in India. In general, estimates were uniformly insignificant.¹⁵ An issue here is that given growth patterns and rapidly changing body weight and height, BMI is not a reliable indicator in these young ages (Vanderwall *et al.* 2017). Further, there is evidence that so far, this disease has largely spared children (Centers for Disease Control 2020).

6. Conclusion

We study BMI as a correlate of the early spatial variation and intensity of Covid-19 across the districts of India and find that net of controls for individual, household, district and state-specific characteristics that measure a wide range of risk factors, BMI significantly predicts outcomes including the likelihood that the district is a hotspot, the natural log number of confirmed cases, the case fatality rate, and the propensity that the districts is a red zone. The predictive power of BMI is especially pronounced among educated populations in urban settings, impervious to conditioning on differences in testing rates across states, and primarily evident among adults.

Disaggregation of adult results by gender reveals that on average, a unit increase in district mean BMI results in about 2.1 additional cases at the mean for men and approximately 2.8 additional cases at the mean for women. We find that for men in particular, BMI is a proxy for a range of comorbidities that predict the incidence of the pandemic including HBA, high blood pressure and high glucose levels. Exposure to risk as proxied by the likelihood of currently working is also an

¹⁵ Results available on request.

explanatory factor for men. Our results remain essentially unaltered when we condition on variations in time elapsed since onset, are robust to inclusion of a variety of measures that control for demographic and health capacity as well as political leaning and efficacy at the state-level, and follow expected patterns in sensitivity tests. Given data limitations of the NFHS, we are restricted to considering populations in the 15-49 age group. Since developing countries have younger demographic age profiles as compared to the developed world, this age limitation is unlikely to bias our estimates in any systematic way. Furthermore, we condition on the proportion of the population that is 60 years and above at the state level in all models. However, if the data used had been from a developed country and we faced similar age limitations, then our results would possibly underestimate impacts in light of the widespread evidence that older populations are more susceptible to this disease.

These results underline that adult BMI is an important predictor of the early spatial evolution and severity of the pandemic across regions of India. We conclude that policy makers may leverage variation in BMI across the landscape of India to identify vulnerable populations. For example, vaccine distribution plans in states such as Massachusetts in the United States and countries such as Germany prioritize people with comorbidities such as obesity. A similar vaccine relief strategy for India seems advisable given our results. As region specific factors also appear to have predictive power in shaping the area-specific incidence of this disease, amelioration policies tailored to local conditions may be more efficient than nation-wide regulations that ignore these nuances. For instance, since the predictive power of BMI is highest among the educated in urban areas, focusing mitigation policies on this group may be more effective than a blanket policy that ignores such distinctions. This is true even though such groups may be, overall, less deserving of relief. As Cheng *et al.* (2020) notes in their survey of global policy responses, obtaining health resources is prioritized by 148 countries and health monitoring has been implemented in 110 countries. The

results of this study offer a readily available health marker that facilitates such monitoring, and that may help to improve the targeting of scarce resources to those who are especially vulnerable.

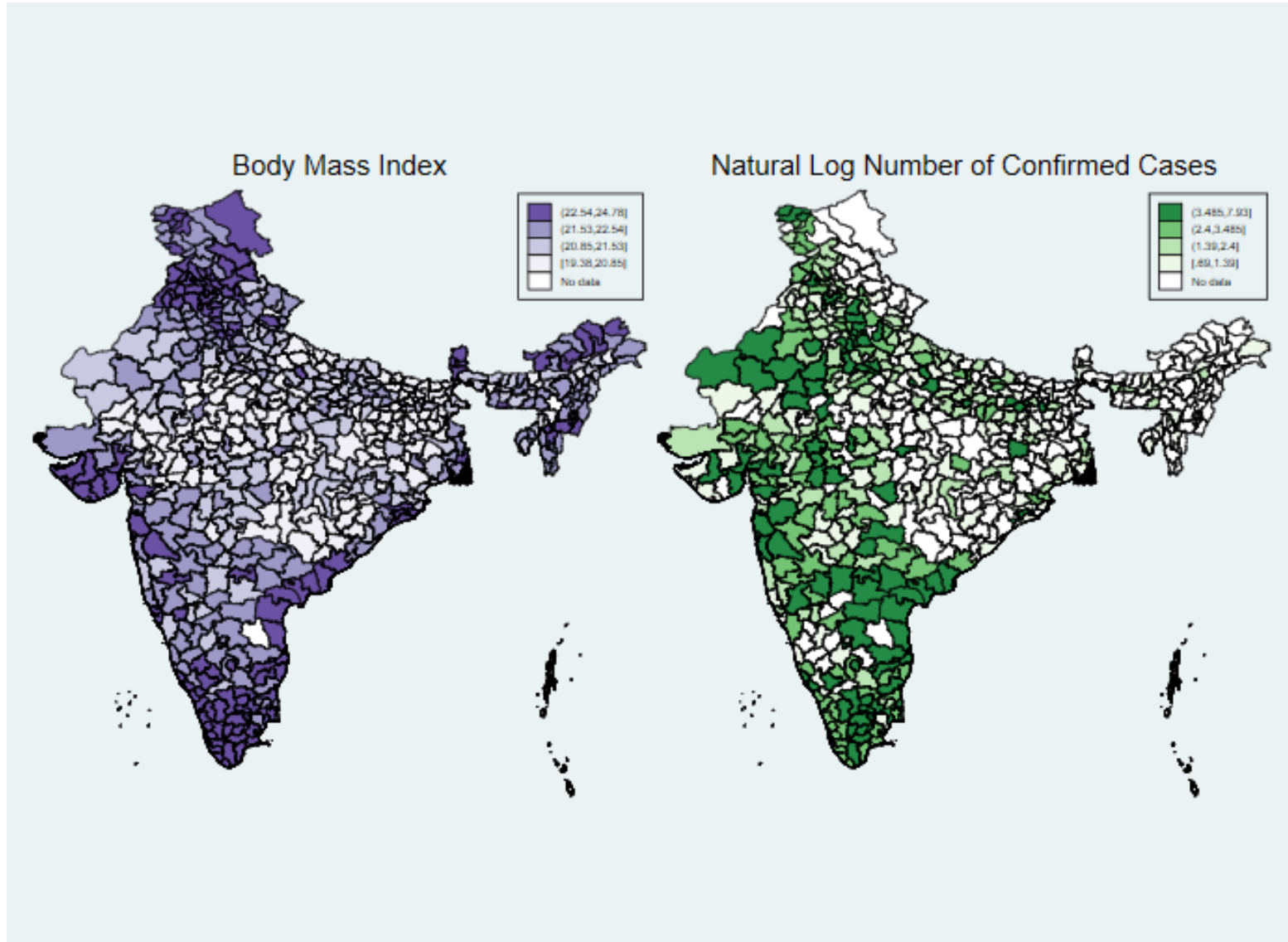
References

- Bansal, M. (2020). Cardiovascular Disease and Covid-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14, 247-250.
- Battegay, M., Richard, K., Tschudin-Sutter, S., and others. (2020). 2019-Novel Coronavirus (2019-nCoV): Estimating the Case Fatality Rate – A Word of Caution. *Swiss Medical Weekly* 150:w20203.
- Caudill, S. (1988). An Advantage of the Linear Probability Model over Probit or Logit. *Oxford Bulletin of Economics and Statistics* 50(4): 425-427.
- Centers for Disease Control. 2020. Coronavirus Disease 2019 in Children – United States, February 12 – April 2, 2020.
- Chatterjee, T., Jain, R. 2020. Is Covid-19 Equally Deadly Across All States? *Ideas for India*.
- Cheng, C., Barcelo, J., Harnett, A., Kubinec, R., & Messerschmidt, L. (2020). CoronaNet: A Dyadic Dataset of Government Responses to the COVID-19 Pandemic. In SocArXiv (dkvxy; SocArXiv). Center for Open Science. <https://ideas.repec.org/p/osf/socarx/dkvxy.html>.
- Cicala, S., Holland, S., Mansur, E., Muller, N., & Yates, A. (2020). Expected Health Effects of Reduced Air Pollution from Covid-19 Social Distancing. NBER Working Paper 27135.
- Cole, M., Ozgen, C., & Strobl, E. (2020). Air Pollution Exposure and COVID-19. IZA Discussion Paper No. 13367.
- Dang, A., Maitra, P., & Menon, N. (2019). Labor Market Engagement and the Body Mass Index of Working Adults: Evidence from India. *Economics and Human Biology* 33, 58-77.
- Dang, H., & Trinh, T. (2020). Does the COVID-19 Pandemic Improve Global Air Quality? New Cross-National Evidence on Its Unintended Consequences. IZA Discussion Paper No. 13480.
- Das, J., & Hammer, J. (2007). Money for Nothing: The Dire Straits of Medical Practice in India. *Journal of Development Economics* 83(1), 1-36.
- Dave, D., Friedson, A., Matsuzawa, K. and others. (2020). Were Urban Cowboys Enough to Control Covid-19? Local Shelter-In-Place Orders and Coronavirus Case Growth. IZA DP No. 13262.
- Desmet, K., & Wacziarg, R. (2020). Understanding Spatial Variation in Covid-19 Across the United States. NBER Working Paper 27329.

- Dey, S., Girolamo, L., Donkelaar, A., Tripathi, S., Gupta, T. & Mohan, M. (2012). Variability of Outdoor Fine Particulate Matter (PM_{2.5}) Concentration in the Indian Subcontinent: A Remote Sensing Approach. *Remote Sensing of the Environment* 127, 153-161.
- Election Commission of India. (Various Years). Election Results - Full Statistical Reports, Statistical Reports of General Election to Lok Sabha. Government of India.
- Fang, H., Wang, L., & Yang, Y. (2020). Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCov) in China. NBER Working Paper 26906.
- Foster-Johnson, L. & Kromrey, J. (2018). Predicting Group Level Outcome Variables: An Empirical Comparison of Analysis Strategies. *Behavior Research Methods* 50, 2461-2479.
- Galasso, V., Pons, V., Profeta, P., Becher, M., Brouard, S., & Foucault, M. (2020). Gender Differences in Covid-19 Related Attitudes and Behavior: Evidence from a Panel Survey in Eight OECD Countries. NBER Working Paper 27359.
- Gao, F., Zheng, K., Wang, X., Sun, Q., and others. (2020). Obesity is a Risk Factor for Greater COVID-19 Severity. *Diabetes Care*. <https://doi.org/10.2337/dc20-0682>.
- Gupta, R. (2016). Health Care Reforms in India. Elsevier Health Sciences.
- Honce, R., Schultz-Cherry, S. (2020). A Tale of Two Pandemics: Obesity and COVID-19. *Journal of Travel Medicine*. <https://doi.org/10.1093/jtm/taaa097>.
- Joe, W., Kumar, A., Rajpal, S., Mishra, U. & Subramanian, S. (2020). Equal Risk, Unequal Burden? Gender Differentials in COVID-19 Mortality in India. *Journal of Global Health Science* 2(1): e17.
- Khanna, M., & Kochhar, N. (2020). Covid-19 Lockdown: What Characterizes India's Hotspot Districts? Scroll.in. Accessed on April 27, 2020.
- Knittel, C. & Ozaltun, B. (2020). What Does and Does Not Correlate with Covid-19 Death Rates. NBER Working Paper 27391.
- Malani, A., Gupta, A., Abraham, R. 2020. Why Does India Have So Few Covid-19 Cases and Deaths? Quartz India.
- Ministry of Housing and Urban Affairs. (2019). Handbook of Urban Statistics, 2019. Government of India.
- Ministry of Health and Family Welfare. (2019). National Health Profile, 2019. Issue 14. Government of India.
- Ministry of Health and Family Welfare. (2020). Letter to States Regarding Containment of Hotspots. Government of India.

- Muller, N., Mendelsohn, R., & Nordhaus, N. (2011). Environmental Accounting for Pollution in the United States Economy. *The American Economic Review*, 101(5), 1649-1675.
- Munshi, K. & Rosenzweig, M. (2009). Why is Mobility in India So Low? Social Insurance, Inequality and Growth. NBER Working Paper 14850.
- Papageorge, N., Zahn, M., Belot, M., van den Broek-Altenburg, E., Choi, S., Jamison, J., & Tripodi, E. (2020). Socio-Demographic Factors Associated with Self-Protecting Behavior During the Covid-19 Pandemic. NBER Working Paper 27378.
- Ray, R. (2020). COVID-19 Exposes India's Weak Public Healthcare. *East Asia Forum*, forthcoming.
- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). Coronavirus Pandemic (COVID-19) [Internet]. <https://ourworldindata.org/coronavirus>. Updated 2020.
- Scavini, M., & Piemonti, L. (2020). Gender and Age Effects on the Rates of Infection and Deaths in Individuals with Confirmed SARS-COV-2 Infection in Six European Countries. *The Lancet*, preprint.
- Vanderwall, C., Clark, R., Eickhoff, J., & Carrel, A. 2017. BMI is a Poor Predictor of Adiposity in Young Overweight and Obese Children. *BMC Pediatrics* 17:135. doi: 10.1186/s12887-017-0891-z.
- Vibhute, V., & Chattopadhyay, A. (2020). On Issues with Covid19 Data and Why Kerala Stands Out in India. International Institute for Population Sciences, Mumbai.
- Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts, London, England, MIT Press.
- Wu, X., Nethery, R., Sabath, M., Braun, D. and others. (2020). Exposure to Air Pollution and COVID-19 Mortality in the United States: A Nationwide Cross-Sectional Study. Harvard T.H. Chan School of Public Health Working Paper.
- Yates, T., Razieh, C., Zaccardi, F., Davies, M. and others. (2020). Obesity and Risk of COVID-19: Analysis of UK Biobank. *Primary Care Diabetes*, forthcoming.

Figure 1. BMI, Natural Log Number of Confirmed Cases at the District-level, and Predicted Correlation in the Adult Sample



Notes: Figures present average values of BMI at the district-level from 2015-2016, and average values of natural log number of Covid19 cases at the district-level as of April 2020. The pair-wise correlation coefficient between BMI and log number of confirmed cases at the district-level in the adult sample is 0.072 with p -value < 0.01 .

Table 1 – Summary statistics of district-level means

Variable	Adults (15-49)		Women (15-49)		Men (15-49)		Diff. (7)
	Mean	SD	Mean	SD	Mean	SD	
	(1)	(2)	(3)	(4)	(5)	(6)	
<i>Outcomes</i>							
Hotspot district	0.185	0.388	0.185	0.388	0.185	0.388	
Natural log number of confirmed cases	2.538	1.461	2.538	1.461	2.538	1.461	
Case fatality rate	0.329	0.425	0.329	0.425	0.329	0.425	
Red zone	0.177	0.373	0.177	0.373	0.177	0.373	
<i>Individual and household characteristics</i>							
Body mass index	21.727	1.133	21.735	1.169	21.690	1.040	
Obese	0.040	0.028	0.043	0.031	0.025	0.022	***
Obese (Asian threshold)	0.088	0.053	0.092	0.056	0.068	0.045	***
Altitude adjusted hemoglobin level (g/dl)	12.000	0.447	11.694	0.458	14.067	0.512	***
Glucose level is greater than median value	0.505	0.081	0.502	0.082	0.531	0.093	***
Told has high BP on >= 2 x by doc./hlth prof.	0.086	0.068	0.089	0.070	0.063	0.064	***
Height in centimeters	153.512	1.961	152.067	1.802	163.354	2.370	***
Age in years	29.921	1.072	29.895	1.098	30.127	1.279	***
Male	0.128	0.029	0.000	0.000	1.000	0.000	
Married	0.709	0.053	0.723	0.055	0.612	0.064	***
Not educated	0.248	0.129	0.267	0.137	0.118	0.079	***
Has some or all primary school	0.133	0.044	0.134	0.044	0.126	0.057	***
Has some secondary school	0.408	0.085	0.397	0.089	0.478	0.076	***
Completed secondary school or higher	0.211	0.100	0.201	0.102	0.278	0.098	***
Number of children below 5 years	0.592	0.18	0.595	0.181	0.574	0.194	**
Hindu	0.745	0.275	0.744	0.275	0.751	0.278	
Muslim	0.127	0.175	0.128	0.175	0.122	0.177	
Christian	0.074	0.211	0.074	0.211	0.074	0.211	
Scheduled tribe	0.192	0.279	0.192	0.279	0.192	0.280	
Scheduled caste	0.193	0.106	0.193	0.106	0.194	0.117	
Other backward caste	0.394	0.213	0.394	0.213	0.398	0.223	
Poorest household	0.189	0.189	0.190	0.190	0.179	0.183	
Poorer household	0.214	0.112	0.214	0.112	0.214	0.118	
Middle income household	0.214	0.088	0.213	0.088	0.218	0.094	
Richer household	0.198	0.099	0.198	0.100	0.201	0.106	
Richest household	0.185	0.174	0.185	0.175	0.188	0.171	
Age of household head	47.526	2.456	47.563	2.453	47.278	2.919	*
Household head is male	0.869	0.059	0.865	0.061	0.899	0.054	***
Household size	5.671	0.771	5.689	0.775	5.555	0.808	***
House has raw wall	0.239	0.232	0.239	0.233	0.234	0.234	
Rural	0.718	0.212	0.721	0.214	0.703	0.206	
Electricity	0.894	0.137	0.894	0.137	0.896	0.139	
Toilet facility: flush toilet	0.512	0.238	0.511	0.239	0.519	0.237	
Source of drinking water: piped water	0.474	0.313	0.474	0.313	0.472	0.317	
Years lived in place of residence	16.904	3.767	16.986	3.783	16.292	4.073	***
<i>District characteristics</i>							

Natural log of PM _{2.5}	3.620	0.542	3.619	0.544	3.623	0.553
Natural log of rainfall in millimeters	3.012	1.598	3.003	1.598	3.081	1.674
Natural log of temperature in centigrade	3.329	0.230	3.328	0.232	3.341	0.224
<i>State characteristics</i>						
Southern states	0.166	0.373	0.166	0.373	0.166	0.373
Testing rate	0.327	0.244	0.327	0.244	0.327	0.244
Natural log of population density in 2011	5.941	1.028	5.941	1.028	5.941	1.028
Natural log of per capita health exp. in 2016	7.194	0.471	7.194	0.471	7.194	0.471
Natural log of number of doctors in 2018	10.491	1.537	10.491	1.537	10.491	1.537
Proportion of respiratory cases in 2018	0.038	0.042	0.038	0.042	0.038	0.042
Prop. of the pop. >= 60 in 2017 (in percent.)	8.316	1.519	8.316	1.519	8.316	1.519
Natural log number of women elected	2.805	0.966	2.805	0.966	2.805	0.966

Notes: Author's calculations from district-level data. Table reports weighted summary statistics. The last column denotes differences in the women and men samples. Sample size is 631 districts. *** Denotes significance at the 1% level, ** at the 5% level and * at the 1% level.

Table 2 – Influence of BMI on hotspots

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Adults (15-49)								
BMI	0.165*** (0.024)	0.251*** (0.061)	0.109*** (0.034)	0.035 (0.155)	0.107*** (0.034)	0.126*** (0.031)	0.137*** (0.035)	0.144*** (0.035)
Observations	631	86	418	70	417	415	391	391
R-squared	0.333	0.524	0.479	0.726	0.474	0.402	0.432	0.442
Panel B: Women (15-49)								
BMI	0.129*** (0.019)	0.209*** (0.052)	0.102*** (0.029)	0.139 (0.143)	0.100*** (0.029)	0.105*** (0.026)	0.120*** (0.029)	0.123*** (0.030)
Observations	631	86	418	70	416	415	391	391
R-squared	0.328	0.520	0.483	0.664	0.479	0.391	0.431	0.447
Panel C: Men (15-49)								
BMI	0.133*** (0.025)	0.220*** (0.067)	0.059** (0.030)	0.153 (0.136)	0.058* (0.030)	0.089*** (0.028)	0.090*** (0.031)	0.112*** (0.032)
Observations	631	85	418	70	414	415	391	391
R-squared	0.300	0.503	0.458	0.670	0.445	0.373	0.404	0.419
Includes PM _{2.5} and rainfall, temperature	NO	YES	YES	YES	YES	YES	YES	YES
Includes controls	NO	NO	YES	YES	YES	YES	YES	YES
Sample restricted to southern states	NO	NO	NO	YES	NO	NO	NO	NO
Sample restricted to non-movers	NO	NO	NO	NO	YES	NO	NO	NO
Include control for testing rate	NO	NO	NO	NO	NO	YES	YES	YES
Include controls for demographic/health infrastructure	NO	NO	NO	NO	NO	NO	YES	YES
Includes controls for HBA, BP and glucose levels	NO	NO	NO	NO	NO	NO	NO	YES
Includes log number of women elected	NO	NO	NO	NO	NO	NO	NO	YES
Includes state fixed-effects	YES	YES	YES	YES	YES	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Southern states include Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Goa. “Non-movers” include those who have been resident in the area for 10 years or more. Testing rate is measured at the state-level. State-level measures on

demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (6)-(8) as these variables are at the state-level. Column (3) does not include $PM_{2.5}$ given number of missing values. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 3 – Influence of BMI on natural log number of confirmed cases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Adults (15-49)								
BMI	0.809*** (0.106)	0.831*** (0.243)	0.160 (0.160)	0.061 (0.380)	0.154 (0.160)	0.214 (0.139)	0.387*** (0.148)	0.365** (0.147)
Observations	390	62	265	63	264	270	266	266
R-squared	0.329	0.460	0.549	0.776	0.554	0.412	0.491	0.503
Panel B: Women (15-49)								
BMI	0.664*** (0.085)	0.699*** (0.202)	0.226 (0.144)	-0.038 (0.369)	0.225 (0.144)	0.227* (0.115)	0.401*** (0.130)	0.355*** (0.132)
Observations	390	62	265	63	264	270	266	266
R-squared	0.333	0.454	0.541	0.714	0.546	0.399	0.474	0.497
Panel C: Men (15-49)								
BMI	0.731*** (0.111)	0.859*** (0.233)	0.172 (0.158)	0.246 (0.339)	0.188 (0.158)	0.217 (0.140)	0.298** (0.149)	0.280* (0.150)
Observations	390	62	265	63	263	270	266	266
R-squared	0.291	0.475	0.516	0.729	0.517	0.386	0.457	0.472
Includes PM _{2.5} and rainfall, temperature	NO	YES	YES	YES	YES	YES	YES	YES
Includes controls	NO	NO	YES	YES	YES	YES	YES	YES
Sample restricted to southern states	NO	NO	NO	YES	NO	NO	NO	NO
Sample restricted to non-movers	NO	NO	NO	NO	YES	NO	NO	NO
Include control for testing rate	NO	NO	NO	NO	NO	YES	YES	YES
Include controls for demographic/health infrastructure	NO	NO	NO	NO	NO	NO	YES	YES
Includes controls for HBA, BP and glucose levels	NO	NO	NO	NO	NO	NO	NO	YES
Includes log number of women elected	NO	NO	NO	NO	NO	NO	NO	YES
Includes state fixed-effects	YES	YES	YES	YES	YES	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Southern states include Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Goa. “Non-movers” include those who have been resident in the area for 10 years or more. Testing rate is measured at the state-level. State-level measures on

demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (6)-(8) as these variables are at the state-level. Column (3) does not include $PM_{2.5}$ given number of missing values. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 4 – Influence of BMI on the case fatality rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Adults (15-49)							
BMI	0.018 (0.038)	-0.167 (0.179)	-0.013 (0.079)	-0.523 (0.000)	-0.012 (0.079)	-0.009 (0.070)	-0.062 (0.072)
Observations	203	34	135	33	134	137	137
R-squared	0.399	0.577	0.561	1.000	0.560	0.389	0.457
Panel B: Women (15-49)							
BMI	0.013 (0.032)	-0.117 (0.140)	-0.040 (0.077)	-0.040 (0.077)	0.030 (0.065)	-0.042 (0.081)	-0.056 (0.083)
Observations	203	34	135	134	137	137	137
R-squared	0.399	0.570	0.539	0.539	0.356	0.456	0.530
Panel C: Men (15-49)							
BMI	0.014 (0.039)	-0.093 (0.130)	0.128* (0.067)	0.125* (0.068)	0.049 (0.055)	0.066 (0.064)	0.060 (0.068)
Observations	203	34	135	134	137	137	137
R-squared	0.399	0.568	0.523	0.516	0.362	0.427	0.464
Includes PM _{2.5} and rainfall, temperature	NO	YES	YES	YES	YES	YES	YES
Includes controls	NO	NO	YES	YES	YES	YES	YES
Sample restricted to non-movers	NO	NO	NO	YES	NO	NO	NO
Include control for testing rate	NO	NO	NO	NO	YES	YES	YES
Include controls for demographic/health infrastructure	NO	NO	NO	NO	NO	YES	YES
Includes controls for HBA, BP and glucose levels	NO	NO	NO	NO	NO	NO	YES
Includes log number of women elected	NO	NO	NO	NO	NO	NO	YES
Includes state fixed-effects	YES	YES	YES	YES	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Southern states include Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Goa, however, do not report estimates for this sample as the number of observations is too small. “Non-movers” include those who have been resident in the area for 10 years or more. Testing rate is measured at the state-level. State-level measures on demographic/health infrastructure include the natural log of population density in 2011,

natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (5)-(7) as these variables are at the state-level. Column (3) does not include PM_{2.5} given number of missing values. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 5 – Influence of BMI on red zones

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Adults (15-49)								
BMI	0.171*** (0.021)	0.248*** (0.063)	0.083** (0.040)	0.220 (0.143)	0.082** (0.040)	0.075** (0.034)	0.091** (0.037)	0.075* (0.038)
Observations	622	86	416	70	415	413	389	389
R-squared	0.252	0.424	0.351	0.533	0.343	0.304	0.329	0.342
Panel B: Women (15-49)								
BMI	0.140*** (0.017)	0.198*** (0.057)	0.056* (0.034)	0.156 (0.147)	0.055* (0.034)	0.041 (0.030)	0.050 (0.032)	0.031 (0.033)
Observations	622	86	416	70	414	413	389	389
R-squared	0.256	0.410	0.319	0.388	0.312	0.286	0.302	0.324
Panel C: Men (15-49)								
BMI	0.140*** (0.022)	0.203*** (0.064)	0.060* (0.035)	-0.068 (0.095)	0.062* (0.035)	0.049 (0.033)	0.052 (0.036)	0.051 (0.037)
Observations	622	85	416	70	412	413	389	389
R-squared	0.216	0.368	0.324	0.527	0.327	0.279	0.301	0.311
Includes PM _{2.5} and rainfall, temperature	NO	YES	YES	YES	YES	YES	YES	YES
Includes controls	NO	NO	YES	YES	YES	YES	YES	YES
Sample restricted to southern states	NO	NO	NO	YES	NO	NO	NO	NO
Sample restricted to non-movers	NO	NO	NO	NO	YES	NO	NO	NO
Include control for testing rate	NO	NO	NO	NO	NO	YES	YES	YES
Include controls for demographic/health infrastructure	NO	NO	NO	NO	NO	NO	YES	YES
Includes controls for HBA, BP and glucose levels	NO	NO	NO	NO	NO	NO	NO	YES
Includes log number of women elected	NO	NO	NO	NO	NO	NO	NO	YES
Includes state fixed-effects	YES	YES	YES	YES	YES	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Southern states include Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Goa. “Non-movers” include those who have been resident in the area for 10 years or more. Testing rate is measured at the state-level. State-level measures on

demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (6)-(8) as these variables are at the state-level. Column (3) does not include $PM_{2.5}$ given number of missing values. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 6 – Influence of obese and differences in testing rates

	Hotspots (1)	Natural log number of confirmed cases (2)	Case fatality rate (3)	Red zones (4)
<i>Panel A: Adults (15-49)</i>				
Obese	6.199*** (1.133)	14.904*** (3.982)	-3.146 (2.855)	2.340* (1.454)
Observations	391	266	137	389
R-squared	0.459	0.511	0.522	0.342
<i>Panel B: Adults (15-49)</i>				
BMI	0.138*** (0.036)	0.382** (0.149)	-0.040 (0.082)	0.068* (0.039)
BMI*high testing rate state	0.009 (0.154)	-0.602 (0.645)	-0.476 (0.456)	0.083 (0.156)
Net effect of BMI in high testing rate states	0.147 [0.335]	-0.219 [0.737]	-0.516 [0.257]	0.152 [0.333]
Observations	391	266	137	389
R-squared	0.431	0.505	0.528	0.337
Includes rainfall and temperature	YES	YES	YES	YES
Includes controls	YES	YES	YES	YES
Include control for testing rate	YES	YES	YES	YES
Include controls for demographic/health infrastructure	YES	YES	YES	YES
Includes controls for HBA, BP and glucose levels	YES	YES	YES	YES
Includes log number of women elected	YES	YES	YES	YES
Includes state fixed-effects	NO	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Testing rate is measured at the state-level. State-level measures on demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (1)-(4) as these variables are at the state-level. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 1% level. *p*-values in square brackets.

Table 7 – Influence of BMI in samples that condition on days from onset

	Hotspots (1)	Natural log number of confirmed cases (2)	Case fatality rate (3)	Red zones (4)
<i>Panel A: Adults (15-49)</i>				
BMI: Sample in which days from onset=0	0.144*** (0.035)	0.365** (0.147)	-0.070 (0.081)	0.075** (0.038)
Observations	391	266	137	389
R-squared	0.442	0.503	0.517	0.342
<i>Panel B: Adults (15-49)</i>				
BMI: Sample in which days from onset=1	0.116** (0.058)	0.208 (0.178)	-0.027 (0.103)	0.144** (0.061)
Observations	209	172	100	208
R-squared	0.521	0.583	0.642	0.446
<i>Panel C: Adults (15-49)</i>				
BMI: Sample in which days from onset=5	0.157 (0.112)	0.049 (0.276)	0.170 (0.200)	0.202* (0.114)
Observations	131	122	64	131
R-squared	0.571	0.598	0.743	0.449
<i>Panel D: Adults (15-49)</i>				
BMI: Sample in which days from onset=10	0.092 (0.082)	-0.002 (0.255)	0.080 (0.177)	0.084 (0.090)
Observations	185	151	79	184
R-squared	0.521	0.608	0.714	0.429
Includes rainfall and temperature	YES	YES	YES	YES
Includes controls	YES	YES	YES	YES
Include control for testing rate	YES	YES	YES	YES
Include controls for demographic/health infrastructure	YES	YES	YES	YES
Includes controls for HBA, BP and glucose levels	YES	YES	YES	YES
Includes log number of women elected	YES	YES	YES	YES
Includes state fixed-effects	NO	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Testing rate is measured at the state-level. State-level measures on demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (1)-(4) as these variables are at the state-level. Robust White standard errors reported. Table reports weighted estimates. Onset day is defined as the day at which the number of confirmed cases reaches 1 per 100,000 people (Desmet and Wacziarg 2020). *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 8 – Influence of BMI by SES and rural/urban

	Hotspots (1)	Natural log number of confirmed cases (2)	Case fatality rate (3)	Red zones (4)
<i>Panel A: Adults (15-49) - Uneducated</i>				
BMI	0.146** (0.067)	0.305 (0.359)	-0.082 (0.496)	0.037 (0.066)
Observations	210	125	50	208
R-squared	0.368	0.553	0.865	0.364
<i>Panel B: Adults (15-49) – Educated</i>				
BMI	0.106** (0.048)	0.046 (0.211)	-0.010 (0.094)	0.113* (0.058)
Observations	181	141	87	181
R-squared	0.566	0.625	0.635	0.416
<i>Panel C: Adults (15-49) – SC/ST/OBC</i>				
BMI	0.090 (0.085)	0.347 (0.466)	0.026 (0.299)	0.152** (0.067)
Observations	195	117	46	193
R-squared	0.459	0.561	0.980	0.309
<i>Panel D: Adults (15-49) - Rural</i>				
BMI	0.123** (0.053)	0.461* (0.265)		0.015 (0.039)
Observations	201	108		200
R-squared	0.271	0.440		0.153
<i>Panel E: Adults (15-49) - Urban</i>				
BMI	0.137*** (0.053)	0.413* (0.225)	0.035 (0.083)	0.126* (0.070)
Observations	190	158	98	189
R-squared	0.542	0.562	0.605	0.470
Includes rainfall and temperature	YES	YES	YES	YES
Includes controls	YES	YES	YES	YES

Include control for testing rate	YES	YES	YES	YES
Include controls for demographic/health infrastructure	YES	YES	YES	YES
Includes controls for HBA, BP and glucose levels	YES	YES	YES	YES
Includes log number of women elected	YES	YES	YES	YES
Includes state fixed-effects	NO	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Testing rate is measured at the state-level. State-level measures on demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (1)-(4) as these variables are at the state-level. Robust White standard errors reported. Table reports weighted estimates. The impact of BMI in rural areas cannot be estimated for the case fatality rate because of insufficient observations. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Table 9 – Falsification/sensitivity checks

	Natural log number of confirmed cases				
	Full sample (1)	Above 10 th percentile (2)	Above 25 th percentile (3)	Above 50 th percentile (4)	Above 56 th percentile (5)
<i>Panel A: Adults (15-49)</i>					
BMI	0.365** (0.147)	0.333** (0.162)	0.273* (0.169)	0.303* (0.171)	0.406** (0.157)
Observations	266	232	211	141	129
R-squared	0.503	0.520	0.495	0.519	0.532
Includes rainfall and temperature	YES	YES	YES	YES	YES
Includes controls	YES	YES	YES	YES	YES
Include control for testing rate	YES	YES	YES	YES	YES
Include controls for demographic/health infrastructure	YES	YES	YES	YES	YES
Includes controls for HBA, BP and glucose levels	YES	YES	YES	YES	YES
Includes log number of women elected	YES	YES	YES	YES	YES
Includes state fixed-effects	NO	NO	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). Testing rate is measured at the state-level. State-level measures on demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (1)-(5) as these variables are at the state-level. Robust White standard errors reported. Table reports weighted estimates. Cut-off values are for the outcome variable in each case. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Online Supplementary Material

Appendix Table 1 – All results for adult sample

Variable	Hotspot district		Natural log number of cases		Case fatality rate		Red zone	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BMI	0.144***	0.035	0.365**	0.147	-0.070	0.081	0.075*	0.038
Altitude adjusted hemoglobin level (g/dl)	-0.027	0.052	0.022	0.315	0.048	0.176	0.013	0.051
Glucose level is greater than median value	-0.361*	0.219	-0.217	1.384	-0.988	0.778	0.128	0.249
Told has high BP on >=2 occ. by doc./health profess.	-0.060	0.357	2.286**	1.092	1.093**	0.436	0.839***	0.316
Height in centimeters	0.008	0.021	0.031	0.091	0.152***	0.046	-0.058***	0.020
Age in years	0.004	0.039	0.094	0.190	-0.056	0.108	-0.033	0.041
Male	-0.939	0.951	3.377	4.154	-2.159	1.674	0.846	1.004
Married	-0.254	0.671	-1.804	3.188	2.692	1.692	-1.159	0.738
Has some or all primary school	0.085	0.558	-4.373*	2.609	1.025	1.806	0.870	0.610
Has some secondary school	-0.688*	0.403	-6.564***	1.552	-0.179	0.912	-0.380	0.354
Completed secondary school or higher	-0.293	0.434	-0.676	2.400	1.501	1.367	0.027	0.480
Number of children below 5 years	0.251	0.271	-0.606	1.327	-0.192	0.684	0.257	0.329
Hindu	0.151	0.176	-0.881	0.962	0.319	0.478	-0.133	0.182
Muslim	0.339	0.239	1.624	1.121	0.508	0.592	-0.084	0.241
Christian	0.276	0.335	-1.146	1.806	0.328	0.861	-0.356	0.340
Scheduled tribe	-0.115	0.194	0.458	1.030	0.671	0.594	-0.210	0.203
Scheduled caste	-0.152	0.269	1.862	1.287	0.241	0.717	-0.477	0.292
Other backward caste	-0.221	0.158	-0.160	0.658	-0.493	0.404	-0.331**	0.139
Poorer	0.492	0.401	2.273	2.120	2.105	1.332	0.076	0.386
Middle	-0.274	0.496	-1.479	2.122	2.705**	1.114	-0.282	0.553
Richer	0.821	0.522	1.585	2.288	1.567	1.250	0.025	0.593
Richest	0.265	0.456	-0.175	2.285	1.226	1.018	0.245	0.520
Age of household head	-0.012	0.017	-0.032	0.070	0.016	0.036	-0.012	0.018
Household head is male	-0.044	0.391	0.276	2.101	-0.948	1.244	0.989**	0.427
Household size	-0.070	0.071	0.195	0.306	0.102	0.190	0.027	0.069
House has raw wall	0.209	0.157	0.072	0.754	0.344	0.553	-0.044	0.167
Rural	-0.665***	0.192	-2.698***	0.703	-0.831***	0.315	-0.417**	0.201

Electricity	-0.159	0.229	0.811	1.342	-0.836	0.755	0.283	0.230
Toilet facility: pit toilet/latrine	-0.166	0.181	-0.083	0.783	-0.386	0.501	0.336*	0.197
Source of drinking water: ground water	-0.020	0.141	0.593	0.502	-0.084	0.339	-0.065	0.137
Years lived in place of residence	-0.008	0.006	-0.055*	0.029	0.018*	0.010	-0.006	0.005
Natural log of rainfall in millimeters	-0.004	0.011	-0.169***	0.058	-0.008	0.031	-0.032**	0.015
Natural log of temperature in centigrade	-0.055	0.091	-0.955	0.841	-0.031	0.413	-0.044	0.092
Testing rate	0.504**	0.237	0.674	0.850	-0.342	0.544	0.364**	0.180
Natural log of population density in 2011	-0.098*	0.052	-0.709***	0.217	0.055	0.129	-0.063	0.043
Natural log of per capita health expenditure in 2016	-0.114	0.120	-1.273**	0.582	0.196	0.293	-0.257**	0.120
Natural log of number of doctors in 2018	0.011	0.040	0.122	0.187	0.142	0.085	0.034	0.038
Proportion of respiratory cases in 2018	-0.596	0.803	-1.019	4.998	-2.565	2.863	0.798	0.896
Proportion of the pop. 60 years and above in 2017	0.048	0.030	0.278**	0.129	0.051	0.065	0.014	0.026
Natural log of number of women elected	0.103**	0.049	0.333	0.228	0.359**	0.149	0.020	0.048
Constant	-1.124	3.430	5.157	16.299	-27.526***	7.649	10.525***	3.321

Notes: OLS regression results of district-level data presented. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Appendix Table 2 – All results for women sample

Variable	Hotspot district		Natural log number of cases		Case fatality rate		Red zone	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BMI	0.124***	0.030	0.358***	0.132	-0.060	0.083	0.031	0.033
Altitude adjusted hemoglobin level (g/dl)	0.055	0.039	0.016	0.206	-0.163*	0.093	-0.023	0.043
Glucose level is greater than median value	-0.297	0.208	0.301	1.317	-1.262*	0.691	0.274	0.258
Told has high BP on >=2 occ. by doc./health profess.	-0.025	0.316	2.319**	0.976	0.887*	0.453	0.811***	0.293
Height in centimeters	-0.193	0.598	-1.397	2.686	4.314***	1.416	-1.655***	0.601
Age	-0.042	0.032	-0.009	0.165	-0.120	0.088	0.018	0.037
Married	-0.857	1.291	-2.180	5.842	2.501	2.903	-0.216	1.180
Has some or all primary school	-0.179	0.443	-3.881**	1.896	0.779	1.268	0.133	0.454
Has some secondary school	-0.592*	0.312	-3.089**	1.281	0.131	0.662	-0.013	0.302
Completed secondary school or higher	-0.609	0.442	-1.565	2.462	1.558	1.190	-0.344	0.464
Number of children below 5 years	0.307	0.261	0.503	1.235	-0.478	0.622	0.216	0.305
Hindu	0.091	0.167	-1.315	0.938	0.256	0.473	-0.118	0.171
Muslim	0.312	0.232	0.964	1.134	0.383	0.590	-0.018	0.235
Christian	0.421	0.364	-0.759	1.751	0.599	0.917	-0.064	0.371
Scheduled caste/scheduled tribe	0.013	0.182	1.112	0.919	0.111	0.542	-0.216	0.190
Other backward caste	-0.111	0.160	0.138	0.658	-0.513	0.357	-0.381**	0.151
Poorer	0.469	0.427	2.329	2.019	1.653	1.291	0.052	0.417
Middle	-0.168	0.469	-0.716	2.020	2.037*	1.122	-0.146	0.530
Richer	0.890	0.549	2.289	2.150	1.281	1.346	0.035	0.616
Richest	0.619	0.451	1.593	2.158	1.047	1.124	0.306	0.519
Age of household head	0.005	0.015	-0.027	0.068	-0.001	0.036	-0.021	0.018
Household head is male	-0.150	0.332	0.299	1.942	-1.594	1.372	0.561	0.376
Household size	-0.144**	0.070	-0.093	0.302	0.139	0.161	0.078	0.074
House has raw wall	0.200	0.149	-0.374	0.728	0.551	0.519	-0.040	0.161
Rural	-0.655***	0.189	-2.613***	0.707	-0.637*	0.338	-0.565***	0.194
Electricity	-0.141	0.223	0.638	1.226	-0.322	0.652	0.281	0.226
Toilet facility: pit toilet/latrine	-0.214	0.188	-0.504	0.788	-0.355	0.490	0.308	0.216
Source of drinking water: ground water	-0.035	0.133	0.670	0.493	-0.103	0.326	-0.100	0.134
Years lived in place of residence	-0.008	0.005	-0.049*	0.026	0.017*	0.009	-0.003	0.005

Natural log of rainfall in millimeters	-0.003	0.012	-0.170***	0.058	-0.017	0.029	-0.028*	0.015
Natural log of temperature in centigrade	-0.053	0.090	-0.948	0.752	-0.184	0.312	-0.031	0.093
Testing rate	0.468**	0.232	0.210	0.811	-0.077	0.508	0.289*	0.165
Natural log of population density in 2011	-0.112**	0.046	-0.717***	0.206	0.036	0.117	-0.073**	0.036
Natural log of per capita health expenditure in 2016	-0.082	0.108	-0.843	0.528	0.096	0.287	-0.159	0.105
Natural log of number of doctors in 2018	0.001	0.039	0.129	0.185	0.223**	0.095	0.033	0.037
Proportion of respiratory cases in 2018	-0.592	0.684	1.834	5.012	-1.799	2.803	0.696	0.777
Proportion of the pop. 60 years and above in 2017	0.047	0.029	0.256**	0.127	0.043	0.057	0.013	0.026
Natural log of number of women elected	0.140***	0.050	0.460**	0.220	0.329**	0.156	0.030	0.047
Constant	1.378	2.209	12.943	9.976	-8.313*	4.969	4.375**	2.117

Notes: OLS regression results of district-level data presented. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Appendix Table 3 – All results for men sample

Variable	Hotspot district		Natural log number of cases		Case fatality rate		Red zone	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BMI	0.112***	0.032	0.280*	0.150	0.060	0.068	0.051	0.037
Altitude adjusted hemoglobin level (g/dl)	-0.042	0.041	-0.096	0.216	0.014	0.128	-0.008	0.044
Glucose level is greater than median value	-0.154	0.208	-0.889	1.063	-0.679	0.570	-0.117	0.226
Told has high BP on >=2 occ. by doc./health profess.	-0.204	0.305	1.574	1.070	0.424	0.435	0.501	0.347
Height in centimeters	0.036***	0.012	0.132**	0.055	0.032	0.031	-0.018	0.012
Age in years	0.005	0.024	-0.067	0.094	-0.004	0.049	-0.020	0.024
Married	-0.641	0.511	-0.916	2.078	-0.604	1.123	-0.462	0.517
Has some or all primary school	-0.273	0.520	-3.529	2.561	0.758	1.464	-0.617	0.495
Has some secondary school	-0.777**	0.350	-5.263***	1.557	0.115	0.890	-0.386	0.349
Completed secondary school or higher	-0.322	0.414	-1.555	2.101	0.600	1.120	-0.382	0.394
Number of children below 5 years	0.059	0.155	-0.592	0.802	0.016	0.528	0.197	0.193
Hindu	0.230	0.169	-0.440	0.880	-0.176	0.476	-0.038	0.177
Muslim	0.355	0.241	1.329	1.026	-0.244	0.520	0.032	0.250
Christian	0.321	0.339	0.286	1.492	-0.666	0.755	-0.222	0.345
Scheduled tribe	-0.110	0.163	1.025	0.773	0.073	0.467	0.002	0.169
Scheduled caste	-0.286	0.220	0.917	1.041	0.581	0.552	-0.052	0.225
Other backward caste	-0.182	0.140	0.247	0.556	-0.329	0.256	-0.196	0.131
Poorer	0.200	0.285	2.244	1.566	0.870	1.099	0.440	0.295
Middle	0.012	0.347	0.396	1.594	0.908	0.999	0.112	0.389
Richer	0.415	0.391	2.335	1.876	0.586	1.030	0.225	0.430
Richest	0.010	0.406	1.312	1.894	0.636	1.055	0.419	0.437
Age of household head	-0.010	0.010	-0.024	0.045	0.008	0.023	-0.005	0.010
Household head is male	0.135	0.390	2.367	1.872	-0.104	1.097	1.309***	0.380
Household size	-0.036	0.045	0.137	0.209	0.050	0.118	0.073	0.048
House has raw wall	0.187*	0.108	-0.119	0.678	0.149	0.437	0.152	0.138
Rural	-0.746***	0.180	-3.108***	0.675	-0.575*	0.329	-0.674***	0.179
Electricity	-0.318	0.204	-1.084	1.063	-0.284	0.720	-0.055	0.194
Toilet facility: pit toilet/latrine	-0.105	0.158	-0.542	0.713	-0.394	0.450	0.407**	0.162
Source of drinking water: ground water	-0.018	0.117	0.583	0.442	-0.248	0.226	-0.043	0.116

Years lived in place of residence	-0.008	0.005	-0.037*	0.022	0.002	0.008	-0.004	0.005
Natural log of rainfall in millimeters	-0.003	0.011	-0.102*	0.054	-0.016	0.029	-0.013	0.014
Natural log of temperature in centigrade	-0.036	0.086	-0.627	0.647	-0.182	0.324	-0.058	0.085
Testing rate	0.549***	0.205	0.072	0.743	0.449	0.473	0.243	0.174
Natural log of population density in 2011	-0.077*	0.042	-0.600***	0.190	-0.065	0.106	-0.085**	0.040
Natural log of per capita health expenditure in 2016	-0.132	0.098	-0.520	0.465	0.180	0.250	-0.123	0.097
Natural log of number of doctors in 2018	-0.008	0.032	0.156	0.180	0.145*	0.075	0.007	0.036
Proportion of respiratory cases in 2018	-0.621	0.581	1.104	4.638	0.200	2.601	-0.271	0.749
Proportion of the pop. 60 years and above in 2017	0.050**	0.024	0.232*	0.128	-0.014	0.063	0.039	0.026
Natural log of number of women elected	0.133***	0.044	0.454**	0.219	0.236*	0.130	0.076*	0.045
Constant	-4.753*	2.550	-12.035	13.116	-8.542	6.631	3.107	2.492

Notes: OLS regression results of district-level data presented. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.

Appendix Table 4 – Influence of obese using the threshold for Asian populations

	Hotspots (1)	Natural log number of confirmed cases (2)	Case fatality rate (3)	Red zones (4)
<i>Panel A: Adults (15-49)</i>				
Obese	3.439*** (0.651)	7.652*** (2.471)	-1.714 (1.518)	1.154 (0.812)
Observations	391	266	137	389
R-squared	0.456	0.507	0.521	0.340
Includes rainfall and temperature	YES	YES	YES	YES
Includes controls	YES	YES	YES	YES
Include control for testing rate	YES	YES	YES	YES
Include controls for demographic/health infrastructure	YES	YES	YES	YES
Includes controls for HBA, BP and glucose levels	YES	YES	YES	YES
Includes log number of women elected	YES	YES	YES	YES
Includes state fixed-effects	NO	NO	NO	NO

Notes: OLS regression results of district-level data presented. Models include a constant term which is not reported. Controls include a set of individual (height, age, marital status, educational level, and number of children less than 5 years in the household) and household characteristics (religion and caste identifiers, wealth index, controls for age of the household head, gender of the household head, household size, type of wall material of the house, rural/urban status, presence of electricity, type of toilet facility, primary source of drinking water and years lived in place of residence). State-level measures on demographic/health infrastructure include the natural log of population density in 2011, natural log of per capita health expenditure in 2016, natural log number of doctors in 2018, proportion of respiratory cases in 2018, and proportion of the population that is 60 years and above in 2017. State fixed-effects cannot be included in columns (1)-(4) as these variables are at the state-level. Robust White standard errors reported. Table reports weighted estimates. *** Denotes significance at the 1% level, ** at the 5% level and * at the 10% level.