Rural-Urban Divide of Covid fatalities in India – Investigating the Role of Lifestyle Disorder Diseases

(Work in progress)

Parantap Basu (Durham University Business School, UK)
(Email: parantap.basu@durham.ac.uk)
Phone +447872040275

Susmita Das (Dept. of Economics, Assam University, Silchar India)
(Email: dsusmita521@gmail.com)

Arnab Dutta Choudhury (Dept. of Economics, Assam University, Silchar India)
(Email: arnabdutta.duttachoudhury@gmail.com)

Ritwik Mazumder (Dept. of Economics, Assam University, Silchar, India)
(Email: ritwiksmazumder@gmail.com, ritwik.mazumder@aus.ac.in)

Acknowledgements

The paper benefitted from the comments of the participants of Allied Social Science Association (ASSA) meeting in 2021. We also acknowledge Mahuya Sengupta for her assistance in editing the paper. The usual disclaimer applies.

Rural-Urban Divide of Covid fatalities in India – Investigating the Role of Lifestyle Disorder Diseases

Abstract

Using data on weekly COVID infections and fatalities at the district level for 23 states and union territories of India, we investigate the determinants of COVID-19 deaths focusing exclusively on the second wave of infections. We include several macroeconomic and structural indicators for districts namely, per capita district domestic product, the degree of urbanization, population density, percentage of aged population, share of agriculture, poverty, amongst several others. Our findings suggest that fatalities have a clear rural-urban divide. Rural agricultural districts with more poor people have experienced less cases and fatalities. Fatalities are more clustered in prosperous and dense industrial and districts. Regions having higher COVID fatalities also have a higher proportion of ageing population with urban life-style disorder related diseases such as obesity, diabetes and hypertension.

JEL Classification Code: I18, I31

Government Policy, Regulation, Public Health, General welfare, Well Being.

Rural-Urban Divide of COVID fatalities in India – Investigating the Role of Lifestyle Disorder Diseases

1. Introduction

A staggering feature of COVID infections in India is its regional disparity. Basu and Mazumder (2021) document that confirmed cases are more concentrated in prosperous and urbanized regions with high population density. On the other hand, poor and less developed regions in India suffer fewer infections. In this present paper, we do a more comprehensive analysis with the district level second wave of COVID data to understand the deep-rooted factors behind this regional variation of fatalities.

There are two principal findings from our study. First, we find a clear rural-urban divide of COVID fatalities in India. Rural districts classified by the degree of urbanization are predominantly poor as per the official poverty indicators and infant mortality rates. These poor rural districts experienced less COVID case fatalities. Fatalities are more clustered in prosperous, urbanized and denser areas with lower poverty. This experience stands in sharp contrast with the experiences of US where poor Hispanic, black and indigenous population were exposed more to COVID infections compared to whites (Stafford, Hoyer and Morrison, 2020, Abedi et al. 2020). Similar pattern is also experienced in UK where black, Asian and Middle Eastern (BAME) groups suffered more infections (ONS 2020). Second, we probe into the reasons for the urban dominance of cases and fatalities. Our study suggests that urban population compared to the rural suffer from diseases of life-style disorders such as obesity, diabetes and hypertension. In addition, the richer and relatively more urban districts have relatively more aged people. These factors together with high population density contributed significantly to COVID cases and fatalities in urban regions of India. The incidence of less cases and fatalities among the poor suggests herd immunity among the poor and it lends support to Strachan's (1989) *hygiene hypothesis* that adults may be more immune to infections if they are exposed to unhygienic environments from early childhood.

In India poor may be immune to various kinds of infections due to unhygienic living conditions from very early childhood while in the US, the basic health infrastructure permits low-income people to access clean and germ-free environment from childhood. We use district level data for all-India analysis of 23 mainland states and union territories leaving out the north-eastern states. Our source of data is the real time database available in www.covid19india.org at the district level, which is by far the most comprehensive dataset for COVID infections and fatalities in different regions of India widely used by researchers. Our principal variables of interest are confirmed cases per million district population, deaths per million, and several regional macro development indicators such as per capita net district domestic product, the degree of urbanization, population density, district-level head count poverty, percentage of aged population (60 years plus), share of district GDP from agriculture and allied activities.

Furthermore, three life-style disorder indicators are chosen that capture the typical urban middle and upper-middle class disease patterns in India. These are hypertension, diabetes and obesity. Most of these state and district level socio-economic and demographic features are drawn from government sources such as the Census of India 2011, the Niti Ayog and the NFHS-5, amongst several others, detailed in the appendix.

The paper is organized in the following sections. We review the related literature in section 2. Section 3 discusses data, measurement and econometric issues. Section 4 reports the results of district level panel data analysis followed by an analysis of the impact of life-style diseases and population ageing on COVID fatalities in section 5. Section 6 concludes.

2. Literature

Several recent studies have reported the trends in COVID-19 infections in India and its regional variations. The most notable study is by Jalan and Sen (2020a) who point out using district level data that all regions of India have not been impacted uniformly by COVID-19, and that there are reasons to argue in favour of more selective lockdown. They find that the severely affected pockets are in metro areas of Delhi, Mumbai, Indore, Jaipur, Chennai, Pune amongst others. The regional disparities in COVID-19 infections in India have also been reported by Mandi, *et al* (2020) where they construct a multi-dimensional Index of vulnerability for districts. Further, Ray and Subramanian (2020) also note regional variations of cases although the objective of their paper aims to provide a critical appraisal of the COVID-19 lockdown in India.

Several studies in recent times have focused on socio-economic and socio-demographic causes of COVID-19 deaths and the literature is evolving rapidly. For US data, Hawkins *et al* (2020) observed that Socio-economic factors play a crucial role in coronavirus disease 2019 (COVID-19) prevalence and mortality. In particular they found that lower education level had the highest association with cases as well as fatalities. Cases and fatalities were higher in proportion among Black residents. COVID-19 fatalities were also correlated with median income and shifts in jobs. Basu, Bell and Edwards (2020) also find that cases are considerably higher in poor districts of UK while residents there also practice less social distancing during the heart of pandemic. In a cross-country study, Sannigrahi *et al* (2020) examined the local and global spatial associations between key social and demographic factors and COVID-19 deaths and cases in the European region using spatial regression models. They documented disparate COVID experiences of different countries where the most affected countries are Italy, Germany, Austria, Slovenia and Switzerland. Yang *et al* (2021) examined the influences of climate, socioeconomic determinants, and spatial distance from Wuhan on the confirmed cases and deaths in the peak phase of COVID-19 in China.

Along similar lines, Amaratunga *et al* (2021) investigated the possible effect of several localised socio-economic factors on the case count and time course of confirmed COVID-19 cases and fatalities across twenty-one counties in New Jersey. Their findings suggest that counties with more dense population proxied by number of restaurants have higher COVID cases. For 401 counties in Germany during the first wave in 2020, using a multivariate spatial model, Ehlert (2021) finds that cases and deaths are significantly positively associated with mean age, population density and the share of people employed in elderly care.

A handful of epidemiological studies are reported in the literature, outlining the potential role of life-style disorders, especially obesity and diabetes in influencing COVID-19 deaths. From a clinical perspective, Albashir (2020) reports that obese patients with high body mass index are at greater risk of complications from viral lung infections and more vulnerable to COVID-19 than non-obese patients because comorbidities associated with obesity are correlated with higher deaths. Wang *et al.* (2021) investigate the global association between lifestyle disorder factors and COVID-19 deaths by means of cross-country regression analysis. Several lifestyle-related indicators, such as obesity and diabetes, are recognised as risk factors behind COVID-19 deaths which together with ageing are associated with increased COVID-19 deaths across countries. Gardiner *et al.* (2021) also provide evidence that a large proportion of the cross-country variation in COVID-19 death rates can be attributed to differences in proportion of obese populations, population health, population density, demographics features, per capita GDP amongst others. For the UK, Tan *et al.* (2020) find increasing evidence lending to the hypothesis that obesity is an independent life-style disorder behind severe infection and even death from COVID-19. Ioannidis *et al.* (2020), Sasson (2021), Cortis (2020), Yanez *et al.* (2020), and Haklai *et al.* (2021), have empirically verified the incidence of higher COVID-19 deaths among old age populations and have demonstrated that the aged populations are significantly more vulnerable to the COVID-19. Basu and Sen (2020) also provide cross country evidence of significant association between ageing and COVID-19 during the onset of the pandemic. Menon (2021) finds that BMI predicts quite significantly the covid hotspots after controlling for several factors. Dang and Gupta (2021) also find evidence of overnutrition and resulting obesity as a determinant of cases and fatalities. Our study complements their finding.

In the backdrop of these studies, we investigate here the role of various socio-economic factors which include development, health, and structural indicators in determining the COVID-19 death differentials across districts of 23 mainland Indian states (listed in Appendix 2B). The interaction between socioeconomic and health indicators in determining regional disparity in COVID fatalities has largely remained unanswered for India. Although Basu and Mazumder (2021) investigated the role of socioeconomic determinants in determining the regional disparity in cases, their work was based on first wave of state level data and did not include fatalities while in this paper, we look at fatalities during the second wave of COVID-19 infections in India with a focus on more disaggregated district level data. Furthermore, we examine the role of life-style diseases in determining regional variations in case fatalities in India which is largely unexplored in the COVID literature. The uniqueness of our study is that we explore the role of interactions between ageing and diseases of life-style disorders in determining the rural-urban divide of COVID case fatalities. Our study is novel because to the best of our knowledge there is no analysis of the determinants of regional variations in

COVID fatalities and infections with district level data. Throughout the text, the terms deaths and fatalities have been alternatively used to mean the same variable.

3. Data

Our key data source for covid-19 related statistics is the national COVID-19 portal for India https://www.covid19india.org/ which has been regularly updated across states and districts of India since the onset of the pandemic in 2020. We take weekly cumulative total COVID-19 figures for both confirmed cases per million (CASES) as well as fatalities per million (DEATHS) across 557 districts covering 23 mainland states and union territories of India leaving out all north-eastern states, the union territories of Ladakh, Andaman and Nicobar, Lakshadweep Islands and also Daman and Diu. Our start date is 23-02-2021 and end data is 27-09-2021, thus covering the second wave of COVID-19 infections in entirety at the district level. This is justified by the fact that the deaths were mostly concentrated in the mainland states of India and that too during the second wave of infections (February to September 2021).

Apart from COVID infections (we call 'cases' in this paper) and deaths, we compile district level development and socio-economic indicators primarily from Census of India, 2011, the Niti Ayog, NFHS-5 and few other sources like state level statistical abstracts (for district level information) compiled from the respective state government portals (listed in Appendix 2 along for precise definition and source of each). These district level indicators are time invariant or fixed factors that vary only across districts but not over time. The district level macroeconomic indicators are PCDDP (per capita district domestic product at constant prices), URBAN (percentage of urban population at the district level), DENSITY (district level density of population, drawn from Census 2011), AGRI (percentage district domestic product from agriculture and allied activities), BPL (percentage of district population lying below the poverty line – basically head count ratio), IMR (the infant mortality rate –district level, drawn from Niti Ayog), ELECT (percentage of district level population living in households having electricity connections, taken from NFHS-5 district level data) and ROADS (sum total of the length of state and national highways at the district level expressed as km/100 square km of district area (compiled from state level statistical abstracts). Finally, the variable AGED represents the percentage of 60 years plus population at the district level which we take as district level old age population. Hypertension (HYPER), diabetes (DIAB) and obesity (OBES) in terms of percentages of state level adult populations are taken from the newly released NFHS-5 statistics (National Family Heath Survey, 5th round, 2019-20) that provide district level cross-sectional data.

Apart from COVID cases and deaths all the developmental and socio-economic indicators including the three life-style disorder variables capture district level fixed effects as every non-COVID variable under this set-up is time invariant over the entire second wave of COVID-19 infections in India. Thus, even in the absence of a fixed effects specification (as in LSDV) the non-covid or district level variables automatically capture fixed (district) effects, a point that must be kept in mind while interpreting the pooled estimators.

3.1. ***Underreporting of fatalities***

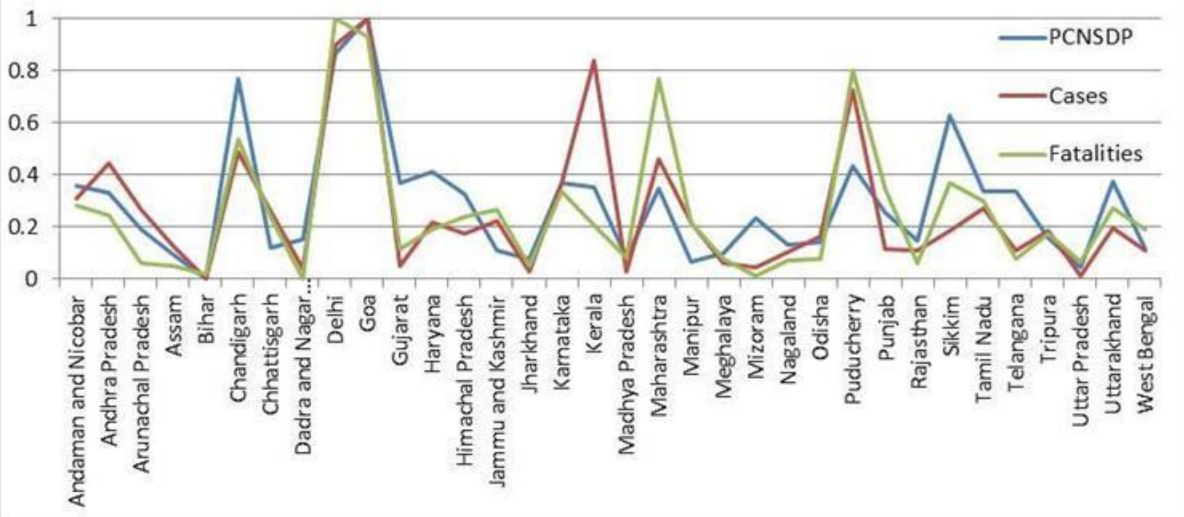
Pursuing COVID research on India, one encounters a formidable problem of underreporting of cases, particularly fatalities. This may quite legitimately raise doubts about the reliability of our regression results. Given that our research focuses on determinants of fatalities, the underreporting typically gives rise to a measurement error issue for the dependent variable. To see it clearly, define $\tilde{y}_{it} =$ reported fatalities at date t in the i^{th} district, y_{it}=actual fatalities and v_{it}, a positive measurement error representing the underreporting. In other words, $\tilde{y}_{it} = y_{it} - v_{it}$. Let x_{it} be the vector explanatory variables. Our true regression equation is, $y_{it} = \alpha + \beta x_{it}$ +u_{it} where u_{it} is the underlying error term which captures all omitted variables. I. The actual regression with observed fatalities as the dependent variable is: $\tilde{y}_{it} = \alpha + \beta x_{it} + e_{it}$ where the composite error term is e_{it} =u_{it}- v_{it} . If u_{it} has a zero conditional mean then E($e_{it}|u_{it}$)) = - E($v_{it}|x_{it}$). The bias then depends on the property of the measurement error. If the error does not depend on the independent variable, and we assume that E($v_{it}|x_{it}$) = μ for some positive constant c for all i and t, then the estimator α is biased because it is $\alpha - \mu$ but the estimator of β is unbiased and consistent.

However, if the measurement error depends on the independent variables, in the sense that E($v_{it}|x_{it}$) changes with x_{it} then we have the usual omitted variable bias problem which makes our estimator of β inconsistent. We need to find a suitable set of instrumental variables (IV) to rectify this bias. In this paper we report both OLS and IV.

4. Empirical Analysis

In figure 1 we plot the per capita NSDP, confirmed cases per million state population and the COVID deaths per million across *states* after expressing each variable in a 0 to 1 scale for the sake of comparability. Seemingly cases and deaths are concentrated in the relatively richer states of India. Motivated by this plot we compute the ordinary correlations between variables of interest at the *district level*. Results are in presented in Table 1. Not surprisingly CASES and DEATHS are significantly correlated. Deaths are positively and significantly correlated with PCDDP, URBAN and DENSITY implying that the COVID deaths in India during the second wave are concentrated in the richer, more urbanised and densely populated districts. The cases are weakly correlated with PCDDP but correlated significantly and positively with URBAN. Almost similar is the correlation pattern for deaths.

Figure 1. Plotting Per Capita NSDP, Confirmed Cases and Fatalities across Indian states (cases and fatalities per million as on 27-09-2021)



Source: Plotted by the authors on the basis of secondary data. Covid statistics are drawn fromcovid19india.org. All variable are expressed in a 0 to 1 scale following the HDI-type attainment index formula.

Few observations are in order. Poor districts classified by BPL are predominantly rural as suggested by the significant negative correlation between BPL and urban and positively correlated with AGRI. Poor districts also have high IMR due to poor health infrastructure and lower population density due to their agricultural base. Cases and fatalities are lower in poor districts as indicated by the significant negative correlations with BPL, AGRI and IMR.

Table 1. District level correlation among variables of interest in the cross-section									
Variables	DEATHS	CASES	PCDDP	URBAN	DENSITY	BPL	AGRI	IMR	ELECT
DEATHS	1.000								
CASES	0.716 (0.000)	1.000							
PCDDP	0.112 (0.057)	0.010 (0.090)	1.000						
URBAN	0.409 (0.000)	0.437 (0.000)	0.287 (0.000)	1.000					
DENSITY	0.242 (0.000)	0.179 (0.002)	0.254 (0.000)	0.376 (0.000)	1.000				
BPL	-0.561 (0.000)	-0.594 (0.000)	-0.261 (0.000)	-0.456 (0.000)	-0.137 (0.019)	1.000			
AGRI	-0.335 (0.000)	-0.374 (0.000)	-0.191 (0.001)	-0.437 (0.000)	-0.195 (0.001)	0.360 (0.000)	1.000		
IMR	-0.412 (0.000)	-0.532 (0.000)	-0.183 (0.001)	-0.305 (0.000)	-0.062 (0.290)	0.686 (0.000)	0.219 (0.000)	1.000	
ELECT	0.699 (0.000)	0.554 (0.000)	0.769 (0.000)	0.737 (0.000)	0.372 (0.001)	-0.221 (0.002)	-0.189 (0.001)	-0.233 (0.000)	1.000
No. of observations = 557									

Source: Computed by the authors on the basis of secondary data.
Notes: Correlations are computed on the basis of a cross-section of 557 districts covering 23 states and union territories of India. Deaths and cases are cumulative totals for 27-09-2021.

All these correlations confirm the finding that poor districts experience less covid cases and fatalities. Moreover, deaths are weakly associated with PCDDP but significantly positively with URBAN. Finally, deaths and cases strongly associate positively with ELECT which is anticipated as urbanised and richer regions of India have better access to household electricity. Arguably ELECT proxies both PCDDP and URBAN in this paper.

Motivated by these correlations, we next run a log-linear cross-district dynamic panel regression to focus on various developmental determinants of COVID-19 deaths across 23 states and union territories of India covering 557 districts. The results are in Table 2. Broadly, in line with the existing literature, we choose PCDDP, URBAN, DENSITY, BPL, AGRI and ELECT as explanatory developmental variables [see for instance, Ehlert (2021); and Canatay *et al* (2021)]. A first order autoregressive term in the form of LOG(DEATHS(-1)) is introduced to adjust for serial correlation in the residuals (all resulting Durbin-Watson statistic values are reasonably close to 2.00). Time and time-squared terms are introduced for capturing the non-linear nature of cumulative deaths.

At first glance, non-industrial states have had fewer fatalities as seen by the significantly negative AGRI coefficient in models 2 and 3. Finally URBAN, PCDDP and DENSITY have significant and positive coefficients. The richer and more densely populated states have a higher chance of COVID related fatalities. These results are broadly consistent with Basu and Mazumder (2021).

Table 2. District level panel regression of weekly cumulative total Covid-19 deaths during second wave (Dependent variable: Log(deaths))

Explanatory Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	0.101** (0.000)	0.018 (0.782)	0.031 (0.084)	-0.052 (0.053)	-1.049** (0.000)
Log(Deaths(-1))	0.961** (0.000)	0.968** (0.000)	0.969** (0.000)	0.965** (0.000)	0.964** (0.000)
Log(PCDDP)		0.006 (0.1583)		0.007** (0.001)	
Log(URBAN)	0.0108** (0.000)		0.014** (0.001)	0.012** (0.001)	0.014** (0.000)
Log(DENSITY)	0.008** (0.000)			0.005* (0.011)	0.004* (0.010)
Log(BPL)	-0.016** (0.000)	-0.008* (0.020)			
Log(AGRI)		-0.010* (0.019)	-0.009* (0.021)		
Log(ELECT)					0.233** (0.000)
Log(URBAN)*Log(BPL)			-0.003** (0.003)		
Time	0.016** (0.000)	0.014** (0.000)	0.014** (0.000)	0.015** (0.000)	0.015** (0.000)
Time-squared	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)
Adjusted R-Square	0.97	0.97	0.97	0.97	0.97
F-Statistic	8636.30**	5829.26**	5787.51**	8438.28**	8580.46**
Durbin -Watson	2.29	2.28	2.28	2.29	2.30
No. of districts = 557, No. of time points = 32; Panel contains 17824 observations					

Source: Estimated by the authors on the basis of secondary data.
Notes: 1. Numbers in the parentheses are p-values where White's diagonally corrected standard errors are used. Here ** means significant at 1% level, * means significant at 5% level. 3. Number of states and UTs = 23, number of districts = 557, number of weeks = 32; panel includes 17824 pooled observations for the second wave only.

We run five different models by adding and dropping district level fixed effect variables. Our rationale for running these various specifications is just to ascertain whether fatalities are consistently lower in less prosperous poor districts which is the key hypothesis of this investigation. Model 1 shows that both DENSITY and URBAN have positive and statistically significant influence on deaths per million, while BPL has a negative influence on the same. Next, model 2 suppresses URBAN but introduces PCDDP. AGRI is also taken while BPL is retained. Income (i.e., PCDDP) fails to explain deaths while partial influences of both AGRI and BPL turn out to be negative and significant. In model 3 the URBAN-BPL interactive term has a negative and significant coefficient although URBAN by itself significantly positively influences deaths, other things unchanged. The negative sign of the interaction term in model 3 suggests that the positive effect of urbanisation on fatalities is partly muted by poverty which accords well with the key result that poor are more immune to COVID fatalities. Among similarly urbanised districts if we move to poorer regions, COVID deaths are expected to go down. AGRI has a death suppressing influence even in model 3. Finally in model 4 when BPL is dropped, PCDDP, URBAN and DENSITY all have positive and significant coefficients strengthening our fundamental hypothesis. Since BPL has strong negative correlation with these three variables, dropping BPL possibly makes the other developmental variables more significant in determining COVID cases and fatalities. On the whole our district level results in table 2 are consistent with our correlation matrix findings in Table 1.

District level cross-sectional regression of fatality-cases ratio is presented in table 3 and it has deeper implications. The explanatory variables chosen are SPATIAL-CASES (which is the simple arithmetic mean of cases per lakh population in the neighbouring districts, representing a spatial autoregressive component), infrastructure related variables such as ELECT and ROADS, HOSPITAL (representing the district level health infrastructure), keeping URBAN as a control factor. The negative hospital coefficient is significant at 1 per cent implying that if physical infrastructure and the level of urbanization is controlled at the district level, COVID-19 deaths have been lower in regions with better health infrastructure. It is noteworthy that SPATIAL-CASES influence deaths positively and significantly. In other words, the larger the cases per lakh in neighbouring districts, the higher the deaths per lakh in the i^{th} district. This finding actually justifies restrictions on physical movements across districts to control COVID-19 fatalities.

Table 3. District level cross-sectional regression of fatality-case ratio on Infrastructure [Dependent variable: Log(deaths/cases)]	
Explanatory Variables	Coefficients (p-values)
Constant	-18.53** (0.000)
LOG(SPATIAL-CASES)	0.457** (0.000)
LOG(ELECT)	0.371** (0.000)
LOG(ROADS)	0.111** (0.000)
LOG(HOSPITAL)	-0.126** (0.006)
LOG(URBAN)	0.294** (0.000)
Adjusted R-Square	0.486
F-Statistic	130.335**
Durbin -Watson	2.187
No. of observations = 557	

Source: Estimated by the authors on the basis of secondary data.
Notes: 1. Numbers in the parentheses are p-values where White's Heteroskedasticity corrected standard errors are used. 2. Here ** means significant at 1% level, * means significant at 5% level. 3. Number of districts = 557. Fatalities (cumulative total deaths per lakh) are for 27-09-2021 for each district.

5. Why did urbanisation and affluence lead to higher fatalities: Exploring the role of lifestyle diseases and population ageing at the district level

We now turn to the question why COVID deaths are more concentrated in the richer and urbanized regions of India? Does the population in the richer and urbanized regions suffer from specific health disorders which are rare in the poorer and rural areas? The NFHS-5 report released recently in India, gives an opportunity to probe deeper. We directly do not have state level estimates of population proportions with comorbidities (which require at least two different diseases or medical conditions simultaneously in the same person), but we do have district level figures for three key lifestyle diseases or disorders taken from the National Family Heath Survey 2019-20 (i.e., NFHS-5).

These are, (i) Diabetes (DIAB) – as measured by the percentage of population above 15 years who have above 140mg/dl blood sugar, (ii) OBESE or obesity as captured by the percentage of 15-49 years population who are obese (i.e., BMI>25kg/m2), and (iii) HYPER – i.e., hypertension, measured by the percentage of 15 years and above population who suffer from elevated blood pressure (Systolic ≥140 mm of Hg and/or Diastolic ≥90 mm of Hg) or taking medicine to control blood pressure). The ordinary correlation coefficients between development variables and the three selected heath indicators that capture lifestyle disorders are presented in Table 4.

Table 4. District level cross-section based Ordinary correlations between Socio-economic and life-style disorder variables

Variables	DEATHS	PCDDP	URBAN	BPL	AGRI	DIAB	OBES	HYPER	AGED
DEATHS	1.000								
PCDDP	0.112 (0.057)	1.000							
URBAN	0.409 (0.000)	0.287 (0.000)	1.000						
BPL	-0.561 (0.000)	-0.261 (0.000)	-0.456 (0.000)	1.000					
AGRI	-0.335 (0.000)	-0.191 (0.001)	-0.437 (0.000)	0.360 (0.000)	1.000				
DIAB	0.173 (0.003)	0.217 (0.000)	0.305 (0.000)	-0.410 (0.000)	-0.145 (0.000)	1.000			
OBES	0.401 (0.000)	0.365 (0.000)	0.444 (0.000)	-0.711 (0.000)	-0.305 (0.000)	0.371 (0.000)	1.000		
HYPER	0.430 (0.000)	0.217 (0.000)	0.289 (0.000)	-0.471 (0.000)	-0.166 (0.005)	0.208 (0.000)	0.682 (0.000)	1.000	
AGED	0.019 (0.018)	0.247 (0.000)	0.670 (0.000)	-0.396 (0.000)	-0.477 (0.000)	0.181 (0.000)	0.403 (0.000)	0.155 (0.000)	1.000
No. of observations = 557									

Source: Computed by the authors on the basis of secondary data.
Notes: 1. Numbers in the parentheses are p-values. 2. Number of districts = 557. Cumulative weekly total deaths (per lakh) are for 27-09-2021 for each district. 3. Correlations are cross-section based.

First, looking at the DEATHS column we find that both obesity (OBESE) and hypertension (HYPER) are significantly and positively associated with DEATHS implying that COVID-19 deaths in India have been more concentrated in districts that suffer more from obesity and hypertension. This is consistent with cross-country observations separately by Wang et al.

(2021) and Gardiner et al. (2021). The ordinary correlation coefficient of DEATHS with each of the life-style diseases is highly significant at the district level cross-section. Moreover, fatalities are significantly and positively associated with ageing.

Second, a glance at the PCDDP and URBAN columns reveals that all three diseases are significantly positively associated with PCDDP, and the degree of urbanization. In addition, all three diseases are significantly negatively associated with BPL and AGRI thereby indicating further that these life-style diseases are more concentrated in the urban and richer regions of India. However, a remarkable observation in table 4 is the highly positive association between AGED and URBAN (correlation being 0.67). It suggests that there are more aged people in urban areas. Motivated by the correlation in Table 4 we run a family of district level panel regressions judiciously choosing the life-style diseases as district level fixed factors along with our usual structural socioeconomic variables. The results are in Table 5.

Table 5. District level panel regression of Covid-19 fatalities, Ageing and lifestyle diseases [Dependent variable: log(deaths)

Explanatory Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.869** (0.000)	-1.009** (0.000)	-1.994** (0.000)	-0.898** (0.000)	-1.147** (0.000)
log(deaths(-1))	0.924** (0.000)	0.923** (0.000)	0.926** (0.000)	0.924** (0.000)	0.922** (0.000)
Log(AGED)	0.036** (0.000)	0.029** (0.000)	0.028** (0.000)	0.021** (0.000)	0.026** (0.000)
Log(HYPER)	0.112** (0.000)				
Log(OBES)		0.047** (0.000)			
Log(DIAB)*Log(HYPER)			0.006** (0.000)		
Log(OBES)*log(HYPER)				0.013** (0.000)	
Log(OBES)*log(DIAB)					0.018* (0.040)
Time	0.016** (0.000)	0.015** (0.000)	0.015** (0.000)	0.016** (0.000)	0.015** (0.000)
Time-squared	-0.0005** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)	-0.0005** (0.000)
Adjusted R-Square	0.954	0.953	0.955	0.965	0.949
F-Statistic	8582.94	8232.53	8110.23	8406.65	8090.43
Durbin -Watson	2.282	2.281	2.291	2.286	2.291
No. of districts = 557, No. of time points = 32; Panel contains 17824 observations					

Source: Estimated by the authors on the basis of secondary data.

Notes: 1. Numbers in the parentheses are p-values where White's diagonally corrected standard errors are used. 2. Here ** means significant at 1% level, * means significant at 5% level. 3. Number of states and UTs = 23, number of districts = 557, number of weeks = 32; panel includes 17824 pooled observations for the second wave only. 4. OBES at the district level is actually female obesity (Source: NFHS-5).

We run five pooled regressions where DEATHS are explained on the basis of percentage of 60 years plus population (we call AGED) and the three life-style diseases including the pairwise interactions. AGED is statistically significant and positive across models implying that everything else equal, the higher the percentage of old age population at the state level, the higher the COVID deaths per million. From table 4, the ordinary correlation between DEATHS and AGED turns out to be significant at 1.8% hinting towards the fact that states with higher old age populations are likely to suffer more covid deaths.

It is noteworthy that obesity and hypertension have significant and positive coefficients implying greater risk of COVID-19 deaths on account of such factors. Hypertension (HYPER) has the highest coefficient value followed by OBES. As a variable DIAB is insignificant on its own (models with insignificant DIAB not reported) but its interaction with HYPER and OBES are both statistically significant and positive. In fact the OBES-HYPER and OBES-DIAB interactions are highly significant. On the whole, obesity in conjunction with hypertension seems to be a very significant factor behind 2nd wave COVID-19 deaths in India. Since all three chosen life-style diseases are primarily urban in nature (in the Indian context), it provides an explanation for incidence of high COVID-19 deaths specifically in urbanized districts. The key outcome from this exercise is that the pair-wise interactions between obesity, diabetes and hypertension have positive and statistically significant coefficients implying that simultaneous incidence of any two of these diseases in the adult population (i.e., the population with comorbidities) can potentially aggravate the fatality risk. In other words, life-style diseases in combinations seem to have a synergistic effect, or else, are potentially more life threatening across India.

Table 6 presents different specifications of models where the principal focus is on the influence of the interactions between URBAN and each of the life-style disorder indicators on COVID-19 deaths. Aged populations are relatively more concentrated in the urbanised districts which is in line with the correlation reported in Table 4. Since these disorders are primarily urban life-style disorders in the Indian context, we attempt an interaction with URBAN, keeping in mind that the likely answer to higher COVID-19 deaths in urban areas could be rooted in URBAN – life-style disorder interactions. To start with, AGED – URBAN interaction coefficient is positive and significant in the first model. Next, across models the coefficients of the interactions between URBAN and each of the three chosen life-style disorder indicators are positive and statistically significant. The central message from table 6 is that, given the levels of incidence of these life-style diseases, as we move to more aged as well as urbanized populations, COVID-19 deaths are expected to rise significantly.

Table 6. Urbanisation and lifestyle disease interactions and consequent Impacts upon Covid-19 deaths– A district level panel regression
Dependent Variable: log(deaths)

Explanatory Variables	Model 1	Model 2	Model 3	Model 4
Constant	-1.003** (0.000)	-1.007** (0.000)	-1.041** (0.000)	-1.507** (0.000)
log(deaths(-1))	0.953** (0.000)	0.954** (0.000)	0.965** (0.000)	0.959** (0.000)
Log(AGED)*Log(URBAN)	0.003** (0.000)			
Log(URBAN)*Log(DIAB)		0.002** (0.000)		
Log(URBAN)*Log(OBES)			0.004** (0.000)	
Log(URBAN)*Log(HYPER)				0.004** (0.000)
Time	0.015** (0.000)	0.016** (0.000)	0.015** (0.000)	0.015** (0.000)
Time-squared	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)	-0.0004** (0.000)
Adjusted R-Square	0.966	0.965	0.965	0.964
F-Statistic	9731.08	9026.08	9097.32	9114.08
Durbin -Watson	2.281	2.288	2.289	2.282
No. of districts = 557, No. of time points = 32; Panel contains 17824 observations				

Source: Estimated by the authors on the basis of secondary data.
Notes: 1. Numbers in the parentheses are p-values where White's diagonally corrected standard errors are used. Here ** means significant at 1% level, * means significant at 5% level. 3. Number of states and UTs = 23, number of districts = 557, number of weeks = 32; panel includes 17824 pooled observations for the second wave only. 4. At the district level OBES is actually female obesity (NFHS-5).

5.1. *Endogeneity*

An important econometric question is whether our key structural-development variables such as PCDDP, URBAN amongst others are exogenous and pass the standard orthogonality tests. The endogeneity is unlikely to arise from reverse causality because it is implausible that COVID fatalities will influence our structural variables within such a short time. However, the endogeneity could arise due to either omitted variables or measurement errors in fatalities correlated with the structural variables. In the presence of such endogeneity the application of ordinary least squares leads to biased and inconsistent estimates of parameters of the regression model. We adopt a standard Two Stage Least Squares (2SLS) approach to test for orthogonality for each of our structural variables. Based on district-level data, the results are reported in Tables A1.1 relegated to appendix (A1). In all regressions, the p-value associated with the J-statistic reveals that IVs are adequate and all our structural (explanatory) variables pass the orthogonality tests.

6. Summary and Conclusions

In this paper we explain the inter-district variations in COVID-19 fatalities in India focusing primarily on the second waves of covid infections data. A striking finding is that fatalities are concentrated in urban and prosperous regions of India with an aged population with diseases of life-style disorders such as obesity, hypertension and diabetes. In addition to this, high population density in these urban industrial districts also contributed to case fatalities while poor in sparsely populated, agricultural regions in India experienced less COVID fatalities. The stark rural-urban and rich-poor divide in covid case fatalities stands in sharp contrast with COVID experiences of the advanced economies and it lends support to Strachan's (1989) hygiene hypothesis that poor, in this case rural India, may be relatively more immune to various infections. Our exercise gives rise to a few vital policy directions. First, given that richer, and denser regions of India have primarily suffered from COVID deaths, there is a clear case for targeted interventions and lockdowns in case of a future outbreak. Second, since aging clearly is a factor explaining deaths, there is a case for temporary separation of the young workers from the old across households, particularly in case of a future outbreak. Third, since incidences of multiple life-style diseases raise the risk of covid deaths especially among the adult and aged population, targeted interventions by the state health department may be needed for testing and vaccinating the vulnerable aged population. A similar targeted intervention is adopted by the National Health Services in the UK.

Appendix 1

Testing for Exogeneity of Regressors – The IV-2SLS Approach

In Table A1.1, to start with we take PCDDP as the only variable that explains deaths (in logarithmic terms). We take 2 instruments, DENSITY and ELECT to explain PCDDP and obtain estimated PCDDP from stage 1 regression. In stage 2, we regress DEATHS on the estimated PCDDP from stage 1 and verify its significance (t-ratio) and overall goodness of fit (R-square). This 2SLS-IV exercise yields a J-statistic that has a p-value of 0.577 leading to acceptance of the orthogonality condition. The null hypothesis here is that PCDDP is exogenous in the DEATHs regression. In a similar fashion we run the 2SLS-IV models for URBAN, DENSITY, AGRI, BPL and ELECT. The tests show that our explanatory variables of interest are indeed exogenous once the instruments are judiciously chosen.

Table A1.1 2SLS-IV Regression of Exogeneity test for explanatory variables
Dependent Variable: Log(DEATHS)

Variable tested For Exogeneity	Instruments Chosen	Instrument Rank	R-square in 2^{nd} stage regression	Prob(J-statistic) & Inference
PCDDP	DENSITY ELECT	3	0.213	(0.577) PCDDP is exogenous
URBAN	ELECT DENSITY IMR	4	0.964	(0.156) URBAN is exogenous
DENSITY	PCDDP LEB IMR	4	0.128	(0.142) URBAN is exogenous
AGRI	IMR PCDDP URBAN	4	0.877	(0.170) AGRI is exogenous
BPL	AGRI PCDDP URBAN	4	0.462	(0.130) AGRI is exogenous
ELECT	AGRI PCDDP URBAN	4	0.951	(0.210) ELECT is exogenous
No. of districts = 557, No. of time points = 32; Panel contains 17824 observations				

Source: Computed by the authors on the basis of secondary data.
Note: The figures in the table are EVIEWS-10 generated during second stage regression of LOG(DEATHS) on first stage OLS estimates of each of the explanatory variables listed in column 1, on the basis of its instruments. HAC adjusted standard errors are used throughout.

Appendix 2(A):

Variable Definitions and Data Sources

AGED – district level percentage of district level population aged 60 years and above taken from Census 2011, (C-14: Population in five year age group by residence and sex, India – 2011) available at https://censusindia.gov.in/nada/index.php/catalog/1541

AGRI – Percentage contribution of District Domestic Product from agriculture and allied activities, Compiled from the District Level Database (DLD) for Indian agriculture and allied sectors published by International Crops Research Institute for the Semi -Arid Tropics (ICRISAT) available at http://data.icrisat.org/dld/src/gdp.html

BPL - Percentage of population below poverty line at the district level (2011-12); Source: Multidimensional Poverty Index Baseline Report based on NFHS-4, NITI Ayog for district level data on head count percentage.

CASES – confirmed cumulative total COVID-19 Infections per lakh district populations (Source: https://www.covid19india.org/for India.

DENSITY –District level population density per sq.km as per 2011 Census, compiled from https://www.census2011.co.in/density.phpfor India (Source: Census of India, 2011), and 2010

DIAB - Diabetes as measured by the percentage of district level population above 15 years who have above 140mg/dl blood sugar, National Family Heath Survey 2019-20.

ELECT –Population living in households with electricity (%); compiled from District Fact Sheets of National Family Health Survey (NFHS-5) 2019-20, published by the Ministry of Health and Family Welfare Government of India available at: chiips.org/nfhs/factsheet_NFHS-5.shtml

HOSPITAL – District level sum of health Sub Centres, PHCs, CHCs, Sub Divisional Hospital, and District Hospitals (per lakh population); compiled from Rural Health Statistics 2020-21 published by Ministry of Health and Family Welfare Statistics Division, government of India, available at https://main.mohfw.gov.in/sites/default/files/rhs20-21_1.pdf

HYPER – Hypertension, measured by the percentage of 15 years and above population at district level who suffer from elevated blood pressure (Systolic ≥140 mm of Hg and/or Diastolic ≥90 mm of Hg) or taking medicine to control blood pressure), National Family Heath Survey 2019-20 (i.e., NFHS-5, available at http://rchiips.org/nfhs/factsheet_NFHS-5.shtml

IMR – Infant Mortality Rate(per 1000 live births, district level) for 2016 obtained from the NitiAyog, Government of India, available at, https://niti.gov.in/content/infant-mortality-rate-imr-1000-live-births (Source: Sample Registration System).

OBES – Obesity as captured by the percentage of 15-49 years population at the district level who are obese (i.e.,BMI>25kg/m2), National Family Heath Survey 2019-20 (i.e., NFHS-5, available at http://rchiips.org/nfhs/factsheet_NFHS-5.shtml

PCDDP– Per capita district domestic product (**PCDDP**); Compiled from state level Directorate of Economics and Statistics, available at respective state government portals.

ROADS- Sum total of the road lengths of state and National Highways at district level expressed as kilometre per 100 square kilometre district area; Compiled from state level Directorate of Economics and Statistics, available at respective state government portals.

URBAN – the degree of urbanisation (%) taken as a percentage of district level urban population based on 2011 Census. (Source: Census of India, 2011).

Appendix 2(B):

List of 23 States and Union Territories considered in this district level study

Andhra Pradesh, Bihar, Chhattisgarh, Chandigarh, Delhi, Gujarat, Haryana, Himachal Pradesh, Jammu & Kashmir, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Puducherry, Rajasthan, Sikkim, Tamil Nadu, Uttar Pradesh, Uttarakhand and West Bengal.

References

Abedi, V., Olulana, O., Avula, V., Chaudhary, D., Khan, A., Shahjouei, S., & Zand, R. (2020). Racial, economic and health inequality and COVID-19 infection in the United States. *Journal of racial and ethnic health disparities*, 1-11.

Albashir, A. A. D. (2020). The potential impacts of obesity on COVID-19. *Clinical medicine*, *20*(4), e109

Amaratunga, D., Cabrera, J., Ghosh, D., Katehakis, M. N., Wang, J., & Wang, W. (2021). Socioeconomic impact on COVID-19 cases and deaths and its evolution in New Jersey. *Annals of Operations Research*, 1-14.

Basu, P. & Sen, K. (2020). Age composition of population and COVID-19. *Ideas for India*, 10 April (Perspectives).

Basu, P., & Mazumder, R. (2021). Regional disparity of covid-19 infections: an investigation using state-level Indian data. *Indian Economic Review*, 1-18.

Basu, P., Bell, C., & Edwards, T. H. (2021). COVID Social Distancing and the Poor: An Analysis of the Evidence for England, *B.E Journal of Economics*, https://doi.org/10.1515/bejm-2020-50.

Brotherhood, L., Cavalcanti, T., Da Mata, D., & Santos, C. (2020). Slums and pandemics. *CEPR Discussion Paper No. DP15131*

Brown, C. S., & Ravallion, M. (2020). Inequality and the coronavirus: Socioeconomic covariates of behavioural responses and viral outcomes across US counties (No. w27549). *National Bureau of Economic Research*.

Canatay, A., Emegwa, T. J., & Talukder, M. F. H. (2021). Critical country-level determinants of death rate during Covid-19 pandemic. *International Journal of Disaster Risk Reduction*, *64*, 102507.

Cortis, D. (2020). On determining the age distribution of COVID-19 pandemic. *Frontiers in public health*, *8*, 202.

Dang, A & Gupta, I. (2021). Obesity and its Impact on COVID Occurrence: Evidence from India, IEG Working Papers 430, Institute of Economic Growth.

Davies, J.B. (2021). Economic Inequality and Covid-19 death rates in the first wave: A cross country analysis. *Covid Economics*. (73), 23 March.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Ehlert, A. (2021). The socio-economic determinants of COVID-19: a spatial analysis of German county level data. *Socio-Economic Planning Sciences*, 101083.

Finch, W. H., & Hernández Finch, M. E. (2020). Poverty and Covid-19: rates of incidence and deaths in the United States during the first 10 weeks of the pandemic. *Frontiers in Sociology*, (*5*), 47.

Gardiner, J., Oben, J., & Sutcliffe, A. (2021). Obesity as a driver of international differences in COVID-19 death rates. *Diabetes, Obesity and Metabolism*, *23*(7), 1463-1470.

Haklai, Z., Aburbeh, M., Goldberger, N., & Gordon, E. S. (2021). Excess mortality during the COVID-19 pandemic in Israel, March–November 2020: when, where, and for whom?*Israel Journal of Health Policy Research*, *10*(1), 1-7

Hawkins, R. B., Charles, E. J., & Mehaffey, J. H. (2020). Socio-economic status and COVID-19 related cases and fatalities. *Public Health*, *189*, 129-134.

Ioannidis, J. P., Axfors, C., & Contopoulos-Ioannidis, D. G. (2020). Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environmental research*, *188*, 109890.

Jalan, J., & Sen, A. (2020a). Understand the method in Covid-19's madness. India doesn't need complete lockdown. *The Print.*10 April, 2020 12:34 pm IST.

Jalan, J., & Sen, A. (2020b). Containing a pandemic with public actions and public trust: the Kerala story. *Indian Economic Review*, *55*(1), 105-124.

Jalan, J., & Sen, A. (2020c). Spread of Covid-19 in India: Across Space and Over Time. https://indiacovidspread.weebly.com.

Kermack, W. O., &McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700-721.

Mandi, J., Chakrabarty, M., & Mukherjee, S. (2020). How to ease Covid-19 lockdown? Forward guidance using a multi-dimensional vulnerability index. *Ideas for India* May 21, 2020.

Menon, N. (2021). Does BMI predict the early spatial variation and intensity of Covid-19 in developing countries? Evidence from India. Economics & Human Biology, 41, 100990.

Ray, D., & Subramanian, S. (2020). India's lockdown: An Interim Report. *Indian Economic Review*, *55*(1), 31-79.

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable cities and society*, *62*, 102418.

Sasson, I. (2021). Age and COVID-19 mortality. *Demographic Research*, *44*, 379-396.

Stafford, K., Hoyer, M., & Morrison, A. (2020). Racial Toll of Virus grows even starker as more data emerges. *AP News*, April 19 https://apnews.com/article/8a3430dd37e7c44290c7621f5af96d6b).

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, *299*(6710), 1259.

Tan, M., He, F. J., & MacGregor, G. A. (2020). Obesity and covid-19: the role of the food industry. *bmj*, *369.*

Wang, J., Sato, T., & Sakuraba, A. (2021). Worldwide association of lifestyle-related factors and COVID-19 mortality. *Annals of medicine*, *53*(1), 1531-1536.

Yanez, N. D., Weiss, N. S., Romand, J. A., & Treggiari, M. M. (2020). COVID-19 mortality risk for older men and women. *BMC Public Health*, *20*(1), 1-7.

Yang, X. D., Su, X. Y., Li, H. L., Ma, R. F., Qi, F. J., & Cao, Y. E. (2021). Impacts of socioeconomic determinants, spatial distance and climate factors on the confirmed cases and deaths of COVID-19 in China. *PLOS One*, *16*(7), e0255239.