



# Rural–Urban Divide of COVID-19 Fatalities in India—Investigating the Role of Lifestyle Disorder Diseases

Parantap Basu<sup>1</sup> · Susmita Das<sup>2</sup> · Arnab Dutta Choudhury<sup>2</sup> · Ritwik Mazumder<sup>2</sup> · Mahuya Sengupta<sup>3</sup>

Accepted: 14 September 2024 / Published online: 26 October 2024  
© The Author(s) 2024

## Abstract

Using data on weekly COVID-19 infections and fatalities at the district level for 19 states and 4 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, among 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 districts. Regions having higher COVID-19 fatalities also have a higher proportion of ageing population with urban life-style disorder related diseases such as obesity, diabetes, and hypertension. Prevalence of respiratory illnesses further aggravates the effects of some of these life-style disorder diseases on COVID-19 fatalities.

**Keywords** Government policy · Regulation · Public health · General welfare · Well-being

**JEL Classification** I18 · I31

## Introduction

During the first two years of coronavirus disease 2019 (COVID-19), numerous studies have highlighted disparities in infection rates and fatalities. Individuals living in poverty exhibited higher rates of mortality compared to their more affluent counterparts. In the United States, it was observed that Hispanic, Black, and indigenous populations were disproportionately affected by COVID-19 infections compared to White individuals (Stafford et al. 2020; Abedi et al. 2020). A similar trend

was observed in the United Kingdom, where Black, Asian, and Middle Eastern (BAME) groups experienced a higher incidence of infections (Windsor-Shellard, & Kaur, 2020). Based on 81 Upper Tier Local Authority (UTLA) districts of UK, Basu et al. (2021) documented that the individuals with lower economic status had difficulty in adhering to the rules of social distancing which led to higher susceptibility to COVID-19 fatalities compared to wealthier individuals.

Moreover, the patterns of infections and fatalities due to the COVID-19 have varied immensely across countries. Surprisingly, richer countries with better healthcare systems have been more affected in terms of higher fatalities compared to poorer nations. The reported death toll remains confined to just a few high-income regions. In fact, lower-middle income and low-income countries (India and those in South Asia) account for just 3% of the global death toll although they carry roughly half of the world's population (Schellekens and Sourrouille 2020). This rich-poor divide in COVID-19 deaths across the globe is well documented.

Understanding the relationship between COVID mortality and poverty is complex because we need to take into account the geographical distribution of the population. If richer people live in urban areas with high population density while poor live predominantly in rural less densely populated areas, poor may be less vulnerable to infections. Second, it has been studied that contact with natural environments enriches the human microbiome, promotes immune balance and protects from allergy and inflammatory disorders. For example, in rural areas, there are domesticated animals which reduce the risk of childhood asthma and other inflammatory disorders through early life zoonotic exposure. These exposures are non-existent in urban slum areas. This concept aligns with the “Missing Microbes” and “Old Friends” hypotheses which suggest that people are not encountering enough microbial stimuli early in life to develop a stronger immune system (Shimojo and Izuhara 2017; Rook 2023). Third, if affluent individuals are predominantly older, they may possess additional co-morbidity factors that increase their vulnerability to infections.

There are serious immunological reasons behind the hypothesis that the urban population across the globe could be more susceptible to COVID-19. A lack of rural upbringing, that includes a long-term and early-life exposure to stables and farm milk, and an elevated acute stress response (induced by adrenaline and cortisol surges) in an urban lifestyle provides a possible mechanism underlying the higher prevalence of chronic inflammatory disorders in urban areas. These ailments include asthma and allergies. Prospective human and mechanistic animal studies reinforce the idea that an exaggerated immune reactivity could play a role in a hyper-reactive immune system and chronic low-grade inflammation (Steinheuser et al 2014; Akdis 2021; Celebi Sozener et al. 2022). With this backdrop, we study the incidence of case fatalities in India during the second wave of COVID-19.

A staggering feature of COVID-19 infections in India is its regional disparity. Basu and Mazumder (2021) document that confirmed cases were more concentrated in prosperous and urbanized regions with high population density. On the other hand, poor and less developed regions in India suffered fewer infections. In this paper, we do a more comprehensive analysis with second wave of COVID-19 data at the district level to understand the deep-rooted factors behind this regional variation of fatalities.

Two principal findings emerge from our study. First, we find a clear rural–urban divide of COVID-19 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-19 case fatalities. Fatalities are more clustered in prosperous, urbanized and denser areas with lower poverty. This experience stands in sharp contrast with the experience of the US where poorer sections viz., Hispanic, Black and indigenous populations were exposed more to COVID-19 infections compared to Whites (Stafford et al. 2020; Abedi et al. 2020). A similar pattern is also experienced in the UK where Black, Asian and Middle Eastern (BAME) groups suffered more infections (Windsor-Shellard, & Kaur, 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 life-style disorders such as obesity, diabetes and hypertension. In addition, the richer and relatively more urban districts have more aged people. Furthermore, urban population in India is likely to suffer more from Chronic Obstructive Pulmonary Disease (COPD) and associated respiratory diseases on account of poor air quality (Ghosh and Mukherji 2014; Maji et al 2018) due to compromised industrial and vehicular emissions. Incidence of respiratory diseases could be an additional factor behind COVID-19 deaths as COVID-19 infections trigger a severe respiratory condition known as Acute Respiratory Distress Syndrome or ARDS (Tzotzos et al 2020). These factors, together with high population density, contributed significantly to COVID-19 cases and fatalities in urban regions of India.

In India, the 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 19 mainland states and 4 union territories leaving out the north-eastern states. Our source of data is the real time database available in [www.COVID19india.org](http://www.COVID19india.org) at the district level, which is by far the most comprehensive dataset for COVID-19 infections and fatalities across India. Our principal variables of interest are confirmed cases per million district population, deaths per million, and several regional macroeconomic and 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 and above), share of district GDP from agriculture and allied activities.

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

The paper is organized in the following sections: We review the related literature in “Literature” section. “Data” section discusses data, measurement and econometric issues. “Empirical Analysis” section reports the results of district level panel data analysis followed by an analysis of the impact of life-style diseases and ageing on COVID-19 fatalities in “Why did Urbanisation and Affluence Lead to Higher

Fatalities: Exploring the Role of Lifestyle Diseases and Population Ageing at the District Level” section. “Summary and Conclusions” section concludes.

## Literature

Several recent studies have reported the trends in COVID-19 infections in India and their regional variations. 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 their paper aims to provide a critical appraisal of the COVID-19 lockdown in India. Using district level data, Jalan and Sen (2020a, b, c) also point out that all regions of India have not been impacted uniformly by COVID-19.

A growing tome of literature reveals several studies that have focused on socio-economic and socio-demographic causes of COVID-19 deaths. For US data, Hawkins et al (2020) observed that socio-economic factors play a crucial role in COVID-19 prevalence and mortality. 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. 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-19 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 21 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, using a multivariate spatial model, Ehlert (2021) finds that cases and deaths have significant positive association with mean age, population density and the share of people employed in elderly care during the first wave of infections in 2020.

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 a greater risk of complications from viral lung infections and more vulnerable to COVID-19 than non-obese patients. This is because co-morbidities 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, demographic features, per capita GDP among 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 the aged. 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-19 hotspots after controlling for several factors. Dang and Gupta (2021) also find evidence of over-nutrition and resulting obesity as a determinant of cases and fatalities. Our study complements their finding.

A growing body of physiological and epidemiological literature reports the association between respiratory conditions and COVID-19. The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) affects the respiratory tract and usually leads to pneumonia in most patients resulting in acute respiratory distress syndrome (ARDS). ARDS, one of the leading causes of lung damage and death in patients with COVID-19, is mainly triggered by elevated levels of pro-inflammatory cytokines, referred to as cytokine storm, that in their ardour to battle a pathogen, impair the respiratory epithelium (Montazersaheb et al 2022). Increased inflammation in urban environments could be due to impaired immune regulation, which is thought to be dependent on reduced exposure, especially during early life, to micro-organisms with which mammals co-evolved (evolutionary symbionts), as proposed by the “biodiversity” hypothesis (Hanski et al 2012), “missing-microbes” hypothesis (Blaser 2017), or “old-friends” hypothesis (Rook et al. 2013). In accordance with the “old friends hypothesis”, early exposure to both pets and farm animals is able to reduce the risk of childhood asthma and other inflammatory disorders (Böbel et al. 2018). In the light 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 and union territories (listed in Appendix 2B).

The interaction between socioeconomic and health indicators in determining regional disparity in COVID-19 fatalities has largely remained unanswered for India. Although Basu and Mazumder (2021) investigated the role of socioeconomic determinants in explaining the regional disparity in cases, their work was based on state level data of the first wave of infections but did not include fatalities. 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-19 literature. The uniqueness of our study is that we explore the role of interactions between ageing and diseases associated with life-style disorders in determining the rural–urban divide of COVID-19 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-19 fatalities with district level data.

Throughout the text, the terms deaths and fatalities have been alternatively used to convey the same meaning.

## 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 19 mainland states and 4 union territories of India leaving out all north-eastern states, the union territories of Ladakh, Andaman and Nicobar Islands, Lakshadweep islands, and Daman and Diu for which district level figures were unavailable on a weekly basis for our study period. The start date for our study is 23-02-2021 and the end date is 27-09-2021, thus covering the second wave of COVID-19 infections in its entirety at the district level. The selection of the region and the period of study are justified by the fact that the COVID-19 deaths were mostly concentrated in the mainland states of India, specifically during the second wave of infections (February–September, 2021).

Apart from tracking COVID-19 infections and deaths, we have compiled district-level development and socio-economic indicators primarily sourced from the Census of India- 2011, Niti Ayog, NFHS-5, and various other references such as state-level statistical abstracts (for district-level information) gathered from the respective state government portals (detailed in Appendix 2 for precise definitions and sources).

Our district-level macroeconomic and development indicators are considered as time-invariant fixed factors. While these factors may vary across districts, they are unlikely to change every week. The district-level indicators include: (i) PCDDP (Per Capita District Domestic Product at constant prices), (ii) URBAN (Percentage of urban population at the district level), (iii) DENSITY (District-level population density, derived from Census 2011), (iv) AGRI (Percentage of district domestic product from agriculture and allied activities), (v) BPL (Percentage of the district population below the poverty line, essentially the headcount ratio), (vi) IMR (Infant Mortality Rate at the district level, obtained from Niti Ayog), (vii) ELECT (Percentage of district-level population residing in households with electricity connections, extracted from NFHS-5 district-level data), (viii) 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).

Additionally, the variable AGED represents the percentage of the population aged 60 years and above at the district level, serving as a measure of the old-age population. Hypertension (HYPER), diabetes (DIAB), and obesity (OBES), presented as percentages of the state-level adult populations, are extracted from the recently released NFHS-5 statistics (National Family Health Survey, 5th round, 2019–20), providing district-level cross-sectional data. Also, district level population per lakh (1 lakh = 1,00,000 or 0.1 million) affected by respiratory diseases (RESP) including Chronic Obstructive Pulmonary Disease (COPD) for 2019–20, (drawn from

Ministry of Health and Family welfare) is taken to represent overall district level respiratory health.

In addition to monitoring COVID-19 cases and deaths, all developmental and socio-economic indicators, including the three lifestyle disorder variables, encapsulate district-level fixed effects. Within the framework of this study, every non-COVID variable remains time-invariant throughout the entire second wave of COVID-19 infections in India. Consequently, we do not use a fixed effects specification, such as in the Least Squares Dummy Variable (LSDV) model, because the non-COVID or district-level variables inherently account for fixed (district) effects. It is crucial to bear this in mind while interpreting the pooled estimators.

### Underreporting of Fatalities

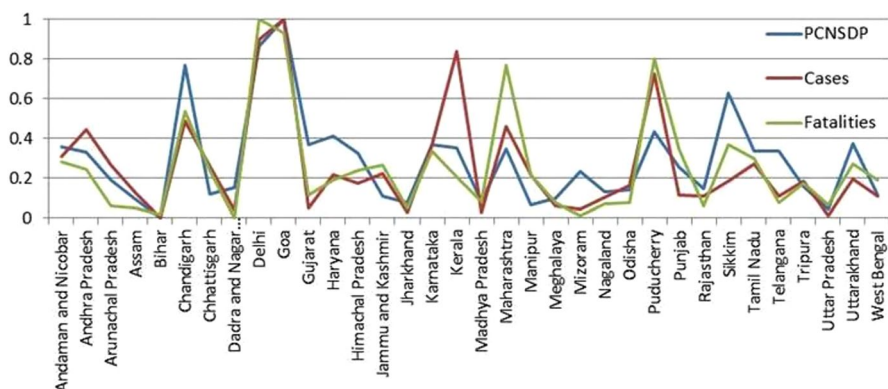
Pursuing COVID-19 research on India, one encounters a formidable problem of underreporting of cases, particularly fatalities. This may quite legitimately raise doubts about the reliability of our regression results. Given that our research focuses on determinants of fatalities, the underreporting typically gives rise to an issue of measurement error 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. 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  $E(e_{it}|x_{it}) = -E(v_{it}|x_{it})$ . 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  $E(v_{it}|x_{it}) = \mu$ , then the estimator  $\alpha$  is biased because it is  $\alpha - \mu$  but the estimator of  $\beta$  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  $\beta$  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.

### Empirical Analysis

In Fig. 1 we plot the per capita NSDP, confirmed cases per million state population and the COVID-19 deaths per million across states after expressing each variable in a 0–1 scale for the sake of comparability. 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 presented in Table 1. Not surprisingly CASES and DEATHS are significantly correlated. Deaths associate positively and significantly with PCDDP, URBAN and DENSITY implying that the COVID-19 deaths in India during the





**Fig. 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 from COVID19 india.org. All variables are expressed in a 0 to 1 scale following the HDI-type attainment index formula

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.

Few observations about the correlations reported in Table 1 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.

All these correlations confirm the finding that poorer districts experience less COVID-19 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 regression to focus on various developmental determinants of COVID-19 deaths across 23 states and union territories of India covering 557 districts taking weekly new death count as the dependent variable. 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)]. 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



**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

**Table 2** District level dynamic panel regression of weekly COVID-19 deaths during second wave [Dependent variable: Dlog(deaths)]

Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.150 (0.070)	-0.246 (0.086)	-0.472** (0.000)	-1.619** (0.001)	-7.611** (0.000)
Dlog(deaths(-1))	0.753** (0.000)	0.798** (0.000)	0.797** (0.000)	0.771** (0.000)	0.776** (0.000)
Log(PCDDP)		0.034** (0.000)		0.069** (0.000)	
Log(URBAN)	0.049** (0.000)		0.178** (0.000)	0.075** (0.000)	0.102** (0.000)
Log(DENSITY)	0.0185 (0.075)			0.015 (0.142)	0.010** (0.008)
Log(BPL)	-0.168** (0.000)	-0.125** (0.000)			
Log(AGRI)		-0.025 (0.094)	-0.019** (0.007)		1.482** (0.000)
Log(ELECT)					
Log(URBAN)*Log(BPL)			-0.035** (0.000)		
Time	0.050** (0.000)	0.019** (0.003)	0.019* (0.003)	0.041** (0.000)	0.038** (0.000)
Time-squared	-0.002** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.002** (0.000)	-0.002** (0.000)
Adjusted R-Square	0.712	0.756	0.756	0.709	0.708
Durbin-Watson	2.261	2.291	2.223	2.269	2.221
F-Statistic	7038.175**	4594.313**	4596.929**	6931.004**	6779.316**
No. of districts = 557, No. of time points = 32; Panel contains 17,824 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. 2. All regressions are pooled

populated states have a higher chance of COVID-19 related fatalities. These results are broadly consistent with Basu and Mazumder (2021).

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) explains deaths while partial influence of both BPL and AGRI 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-19 fatalities. Among similarly urbanised districts if we move to poorer regions, COVID-19 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 and URBAN have positive and significant coefficients strengthening our fundamental hypothesis. Since BPL has strong negative correlation with these variables, dropping BPL possibly makes the other developmental variables more significant in determining COVID-19 cases and fatalities. On the whole our district level results in Table 2 are consistent with our correlations in Table 1.

Time and time-squared are used as regressors to capture the non-linearity in the cumulative growth pattern. White's period standard errors have been used throughout for the sake of robustness although consistency in the signs of the coefficients is the focus rather than statistical precision of the estimates. To adjust for serial correlation in errors we introduce a lagged dependent variable as a regressor throughout.

### **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-19 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? Studies have demonstrated that hospitalized patients younger than 50 with morbid obesity ( $\text{BMI} \geq 40 \text{ kg/m}^2$ ) are more likely to die from COVID-19. Patients with obesity are associated with impaired immune response, endothelial dysfunction, decreased functional residual capacity in the lungs and hypoxemia and expression of ACE2 in adipose tissue which has high affinity to the SARCOV2 virus (Klang et al. 2020; Tibirić and Lorenzo 2020). Moreover, there is sufficient evidence by now that co-morbidity factors like diabetes, hypertension and obesity might be jointly responsible for higher COVID-19 deaths (Escobedo-de la Peña et al. 2021; Mahamat-Saleh 2021).

The NFHS-5 report in India provides an opportunity to delve deeper into health-related insights. While state-level estimates of population proportions with co-morbidities (defined as the presence of at least two different diseases or medical conditions simultaneously in the same person) are not directly available, we do possess district-level data for three crucial lifestyle diseases or disorders from the National Family Health Survey 2019–20 (NFHS-5).

These are namely: (i) Diabetes (DIAB)—measured by the percentage of the population above 15 years with blood sugar levels exceeding 140mg/dl, (ii) Obesity (OBES)—captured by the percentage of the population aged 15–49 years who are obese ( $BMI > 25 \text{ kg/m}^2$ ). It is worth noting that at the district level, the NFHS-5 fact sheet provides data specifically for female obesity, (iii) Hypertension (HYPER)—indicating the percentage of the population aged 15 years and above suffering from elevated blood pressure (Systolic  $\geq 140$  mm of Hg and/or Diastolic  $\geq 90$  mm of Hg) or taking medication to control blood pressure.

Furthermore, it is well-recognized that the elderly population constitutes a significant proportion of COVID-19 mortalities globally (Wang et al. 2021; Yanez et al. 2020). Urban India having a relatively higher proportion of aged individuals could potentially explain the higher urban deaths due to COVID-19, especially when considering the influence of co-morbidity factors. In addition COPD and associated respiratory conditions are likely to be associated with higher COVID-19 deaths because a significant proportion of COVID-19 patients have been observed to develop ARDS (Hsu et al. 2021).

Table 3 presents the ordinary correlation coefficients between development variables and the three selected health indicators, providing insights into the interplay between development factors, lifestyle disorders as well respiratory diseases.

First, looking at the DEATHS column we find that both obesity (OBES) 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 and the incidence of respiratory diseases (RESP).

Second, a glance at the PCDDP and URBAN columns reveal that the three life-style disorder diseases along with respiratory disease variable (RESP) are all significantly positively associated with PCDDP, and the degree of urbanization. Two of the lifestyle disorder diseases, namely hypertension and diabetes, are significantly positively correlated with respiratory diseases across Indian districts implying the coexistence of lifestyle diseases and respiratory diseases in India. 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 noteworthy observation in Table 3 is the highly positive association between AGED and URBAN (correlation being 0.67). It suggests that there are more aged people in urban areas. The statistically significant

**Table 3** District level cross-section based ordinary correlations between Socio-economic and life-style disorder variables

Variables	DEATHS	PCDDP	URBAN	BPL	AGRI	DIAB	OBES	HYPER	RESP	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		
RESP	0.248 (0.000)	0.211 (0.000)	0.289 (0.000)	-0.002 (0.212)	-0.009 (0.193)	0.179 (0.011)	-0.001 (0.119)	0.197 (0.002)	1.000	
AGED	0.119 (0.058)	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 (0.003)	0.203 (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

and positive ordinary correlation coefficient between AGED and RESP suggests that in India, the aged suffer more from respiratory diseases (Divo et al. 2014).

Although these correlations do not necessarily mean any causality among the variables, they certainly motivate further empirical analysis. 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 4. We run three district specific regressions where weekly new 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. We introduce AGED as a control factor throughout the three models. 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-19 deaths per million.

It is noteworthy that the coefficients of age, obesity, hypertension, diabetes and respiratory disease variables are all significant and positive (Model 1), implying greater risk of COVID-19 deaths on account of such factors. In the other interactive models (models 2 and 3), the AGED-HYPER and AGED-DIAB interactions have positive coefficients and are highly significant. In addition, the AGED-RESP interaction is also statistically significant indicating that the simultaneous presence of more aged people and proportion of population suffering from chronic respiratory diseases are likely to raise COVID-19 deaths at the district level. On the whole, ageing in conjunction with hypertension and diabetes seem to be significant

**Table 4** District level dynamic panel regression of weekly COVID-19 fatalities on ageing, lifestyle diseases and their interactions [Dependent variable: Dlog(deaths)]

Explanatory variables	Model 1	Model 2	Model 3	Model 4
Constant	−3.133** (0.000)	−1.798 (0.087)	1.442* (0.011)	−1.323* (0.031)
Dlog(deaths(−1))	0.756** (0.005)	0.756** (0.000)	0.770** (0.000)	0.757** (0.000)
Log(AGED)	0.349** (0.000)	−1.770 (0.060)	−1.059 (0.071)	−1.067 (0.076)
Log(HYPER)	0.429** (0.000)	−1.081 (0.077)		
Log(OBES)	0.077** (0.000)			0.069* (0.041)
Log(DIAB)	0.079** (0.002)		−1.199 (0.076)	
Log(RESPI)	0.113* (0.037)			0.109* (0.039)
Log(AGED)*Log(HYPER)		0.709** (0.000)		
Log(AGED)*log(DIAB)			0.603** (0.000)	
Log(AGED)*log(RESPI)				0.217* (0.047)
Time	0.049** (0.000)	0.049** (0.000)	0.049** (0.000)	0.049** (0.000)
Time-squared	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)
Adjusted R-Square	0.712	0.711	0.711	0.711
F-Statistic	6995.99	6991.66**	6912.81**	6898.39**
Durbin–Watson	2.361	2.361	2.368	2.332

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 Districts = 557, number of time points = 32; panel includes 17,824 pooled observations for the second wave only

co-morbidity factors behind second wave of COVID-19 deaths in India. Since all three chosen life-style disorder 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. In Table 4 the AGED–OBES interaction term is deliberately avoided as OBES covers the age-group 15–49 years only implying two mutually exclusive groups of individuals for AGED and OBES.

The key implication of this exercise is that if the population is comparatively more aged and if it suffers more from life style disorder related diseases (compared to that in poor, agricultural and rural districts) then the fatality risks due to COVID-19 are significantly higher. Respiratory disease both in isolation as well as through its interaction with ageing explains COVID-19 fatality. In other words, co-morbidities resulting from a combination of lifestyle diseases, ageing, and respiratory illnesses are potentially more life-threatening across India concerning COVID-19 deaths. These findings are fairly consistent with Sasson (2021), Ho et al. (2020), and Yanez et al. (2020) who also report higher COVID-19 mortality rates among the elderly and also higher mortality among the elderly with co morbidities.

### **The Interactions Between Urbanization and Urban Life-Style Diseases and Consequent Impact on COVID-19 Fatalities**

Urban life-style diseases in India are concentrated more among the affluent urban population. The percentage of district level urban population (the variable URBAN in this paper) and the percentage of population suffering from life-style disorder diseases can have interactive effects and these interactions can potentially influence the COVID-19 fatality rates.

Table 5 presents different specifications of models where the principal focus is on the influence of the interactions between urbanisation (URBAN) and each of the life-style disorder indicators on COVID-19 deaths. Aged populations in India are relatively more concentrated in the urbanised districts according to the correlation reported in Table 4. Since these disorders are primarily urban life-style disorders, we add an interaction with URBAN, keeping in mind that a likely candidate for explaining higher COVID-19 deaths in urban areas could be rooted in URBAN—life-style disorder interactions. We present a non-interactive model to begin with (Model 1), which simply has the pure effects due to ageing, urbanization and each of the three life-style diseases besides incidence of respiratory diseases. Apart from DIAB, all explanatory variables including respiratory diseases turn out to be statistically significant in explaining weekly COVID-19 deaths. That is, the partial impacts of ageing, urbanization, respiratory disease and the three key urban life-style diseases for India (i.e., obesity, diabetes and hypertension) on COVID-19 deaths are positive and significant. The AGED–URBAN interaction coefficient is positive and significant in the second 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 results in Table 5 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 related deaths are expected to rise significantly.



**Table 5** Urbanisation and lifestyle disease interactions and consequent Impacts upon COVID-19 deaths—a district level dynamic panel regression

Explanatory variables	Dependent Variable: Dlog(deaths)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	-3.008** (0.000)	-1.252** (0.000)	-1.153** (0.007)	-1.142** (0.000)	-1.775** (0.000)	-1.733** (0.000)
Dlog(deaths(-1))	0.757** (0.000)	0.788** (0.000)	0.778** (0.000)	0.769** (0.000)	0.760** (0.000)	0.760** (0.000)
Log(AGED)	0.266** (0.000)	0.305** (0.000)				0.299* (0.032)
Log(URBAN)	0.026 (0.102)	-0.026 (0.182)	-0.111 (0.089)	-0.001 (0.990)	-0.186 (0.117)	-0.113 (0.132)
Log(DIAB)	0.071** (0.006)		-0.091 (0.155)			0.066* (0.044)
Log(OBES)	0.075** (0.001)					
Log(HYPER)	0.437** (0.000)					
Log(RESPI)	0.071* (0.037)			0.128* (0.015)	0.302** (0.009)	0.062* (0.041)
Log(URBAN)*Log(AGED)		0.025 (0.079)				
Log(URBAN)*Log(DIAB)			0.083** (0.001)			
Log(URBAN)*Log(OBES)				0.028** (0.007)		
Log(URBAN)*Log(HYPER)					0.092* (0.014)	
Log(URBAN)*Log(RESPI)						0.020* (0.039)
Time	0.048** (0.000)	0.042** (0.000)	0.043** (0.000)	0.042** (0.000)	0.047** (0.000)	0.047** (0.000)
Time-squared	-0.002** (0.000)	-0.002** (0.000)	-0.002** (0.000)	-0.002** (0.000)	-0.002** (0.000)	-0.002** (0.000)
Adjusted R-Square	0.712	0.709	0.711	0.712	0.711	0.713
F-Statistic	5265.85**	6890.4	6895.40**	6947.56**	7001.76**	6997.72**
Durbin-Watson	2.361	2.384	2.384	2.377	2.379	2.376
No. of districts = 557, No. of time points = 32; Panel contains 17,824 observations						

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

Notes: 1. Numbers in the parentheses are p-values where White's cross-section adjusted standard errors are used. Here \*\* means significant at 1% level, \* means significant at 5% level

In sum, urbanization in India is likely to have aggravated the COVID-19 deaths toll through the life-style disease channel. Our results corroborate the findings by González-Val and Sanz-Gracia (2022), Naudé and Nagler (2022) and Chang et al (2022), who also provide similar evidence across nations where urbanization has been identified as a crucial predictor of COVID-19 mortalities.

### Presence of Co-morbidities and COVID-19 Fatalities in India

A rudimentary approach of the modelling co-morbidity (the presence of at least two diseases or medical conditions in a person—lifestyle and respiratory diseases here) in a regression framework is through the introduction of pair-wise disease interaction terms from our chosen lifestyle disease incidences—namely, (i) OBES–DIAB, (ii) HYPER–OBES, (iii) DIAB–HYPER and similar interactive variables involving ‘RESP’ as an additional co-morbidity factor. In Table 6 we present the dynamic panel data models with weekly deaths count as the dependent variable. Our key objective here is to explore the impact of life-style disease interactions including those involving the incidence of chronic respiratory illness on COVID-19 deaths in India.

The first model shows the pure and partial effects of each of the life-style diseases on deaths along with a single interactive term in the form of HYPER–RESP. Individually, all three life-style diseases and the respiratory disease variable explain COVID-19 deaths both positively and significantly. However, the disease interaction terms that are of special interest, are significant across the six models implying that co-morbidity related factors have significantly aggravated COVID-19 deaths in India during the second wave. This is fairly consistent with the findings of Rana et al. (2023) for the NCR region of India (Delhi), where higher mortality risk and severity due to COVID-19 was observed in patients with co-morbidities like diabetes, hypertension and chronic kidney diseases. Throughout the analysis, models with OBES–RESP interactive term are not reported, as this interaction is found to be insignificant.

Finally, we repeat a somewhat similar set of models with Fatality-Case ratio (FCR for short) as the dependent variable. FCR is the cumulative total weekly deaths divided by cumulative total weekly confirmed cases, times 1000. This is the number of COVID-19 deaths per 1000 confirmed cases (FCR here) representing the fatality ratio. The estimated pooled regression models are in Table 7. Except DIAB, other chosen variables are statistically significant in Model 1. The rest of the four estimated models are largely similar to those presented in Table 6. AGED and URBAN influence the FCR positively and significantly and this is a consistent finding across models in Table 7. Most importantly, the life-style disease interactions are all highly significant in explaining FCR even when respiratory disease is included as an additional co-morbidity factor.

### 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-19 fatalities will influence our structural

**Table 6** Lifestyle disease interactions and consequent Impacts upon COVID-19 deaths—a district level dynamic panel regression [Dependent Variable: Dlog(deaths)]

Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	−2.433** (0.000)	0.052 (0.115)	0.729 (0.228)	2.719** (0.000)	−1.633** (0.000)	−0.059* (0.021)
Dlog(deaths(−1))	0.766** (0.000)	0.771** (0.000)	0.764** (0.000)	0.765** (0.000)	0.765** (0.000)	0.765** (0.000)
Log(DIAB)	0.081** (0.001)	−0.664 (0.073)		−1.894 (0.099)	−0.889 (0.091)	
Log(OBES)	0.129** (0.000)	−0.454 (0.063)	−0.868 (0.087)			
Log(HYPER)	0.421** (0.000)		−0.575 (0.069)	−1.199 (0.081)		−0.996 (0.091)
Log(RESPI)	0.003* (0.049)				0.031* (0.033)	0.029* (0.037)
Log(OBES)*Log(DIAB)		0.262** (0.000)				
Log(HYPER)*Log(OBES)			0.333** (0.000)			
Log(DIAB)*Log(HYPER)				0.667** (0.000)		
Log(DIAB)*Log(RESPI)					0.010* (0.041)	
Log(HYPER)*Log(RESPI)						0.011** (0.004)
Time	0.044** (0.000)	0.041** (0.000)	0.044** (0.000)	0.044** (0.000)	0.044** (0.000)	0.044** (0.000)
Time-squared	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)	−0.002** (0.000)
Adjusted R-Square	0.709	0.708	0.709	0.709	0.709	0.709
F-Statistic	6939.21	6909.44	6948.43	6948.10	6436.23	6623.33
Durbin–Watson	2.371	2.378	2.370	2.372	2.371	2.324
No. of districts = 557, No. of time points = 32; Panel contains 17,824 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. Standard errors are White's heteroscedasticity adjusted

**Table 7** Lifestyle disease interactions and consequent Impacts upon COVID-19 Fatality-Case Ratio—a district level panel regression

Explanatory variables	[Dependent Variable: log(Fatality-Case Ratio)]					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	1.179** (0.000)	1.177** (0.115)	1.127** (0.228)	1.119** (0.000)	1.118** (0.000)	1.122** (0.000)
Log(URBAN)	0.076 (0.051)	0.077* (0.047)	0.079* (0.050)	0.078 (0.052)	0.077 (0.054)	0.074 (0.053)
Log(AGED)	0.051* (0.044)	0.055 (0.082)	0.063 (0.060)	0.054* (0.048)	0.059* (0.029)	0.058 (0.051)
Log(DIAB)	−0.887 (0.322)	−0.004 (0.121)		−0.007 (0.199)		−0.002 (0.112)
Log(OBES)	0.074** (0.000)	−0.006 (0.163)	−0.117 (0.187)		−0.359 (0.099)	
Log(HYPER)	0.029** (0.000)		−0.039 (0.088)	−0.236 (0.100)	0.006 (0.051)	0.005 (0.053)
Log(RESPI)	0.006 (0.066)					
Log(OBES)*Log(DIAB)		0.052** (0.000)				
Log(HYPER)*Log(OBES)			0.031** (0.000)			
Log(DIAB)*Log(HYPER)				0.091** (0.000)	0.136** (0.001)	
Log(HYPER)*Log(RESPI)						
Log(DIAB)*Log(RESPI)						0.176* (0.044)
Time	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Time-squared	−0.0007** (0.001)	−0.0007** (0.001)	−0.0007** (0.002)	−0.0007** (0.001)	−0.0007** (0.001)	−0.0007** (0.001)
Adjusted R-Square	0.397	0.369	0.368	0.369	0.369	0.369
F-Statistic	1999.21	1989.42	1991.77	1993.86	1879.81	1996.72
No. of districts = 557, No. of time points = 32; Panel contains 17,824 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. Standard errors are White's heteroscedasticity adjusted. 2. FCR is the fatality-case ratio expressed as deaths per 1000 confirmed cases

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 Table 8 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.

## Summary and Conclusions

In this paper we explain the inter-district variations in COVID-19 fatalities in India focusing primarily on the second wave of COVID-19 infections and fatalities data. A striking finding is that fatalities are concentrated in urban and prosperous regions of India with a predominant aged population with a prevalence of life-style disorder related diseases such as obesity, hypertension and diabetes. Urbanised and industrially dominant districts have poorer air quality in India, thereby raising the incidence of chronic respiratory diseases. Consequently this adds to the list of co-morbidity factors among urban population. Our findings suggest that chronic respiratory illnesses in conjunction with life-style disorder diseases have aggravated COVID-19 fatalities in India. In addition to this, high population density in these urban industrial districts also contributed to case fatalities while the low-income citizens in sparsely populated agricultural regions experienced lesser COVID-19 fatalities.

Over the course of human history, a close beneficial relationship has developed between our immune system and the microbes that live within us. Exposure to environmental and commensal microbes played a beneficial role in setting up pathways that regulate our immune system. However, dietary changes, overuse of antibiotics, reduction of breastfeeding, improved sanitation, treated drinking water, Caesarean births, and spending more time indoors have reduced our exposure to these microbes, thus compromising our immune system. This explains why we see a clear rural–urban divide in COVID-19 fatalities. The stark rural–urban and rich–poor divide in COVID-19 case fatalities stands in sharp contrast with COVID-19 experiences of the advanced economies and it lends support to the “missing microbes” and “old friends” hypotheses that poor in rural India may be relatively more immune to various infections.

What does this mean for our health and policies? Since wealthy and densely populated areas were hit harder by COVID-19, in case of another outbreak, it might make sense to focus on and prioritize these areas for special attention and possible lockdowns. Since age is a big factor in COVID-19 deaths, it might be a good idea to keep younger and older people in segregated bubbles during an outbreak. As the elderly, and people with existing health issues are more at risk, it might be helpful for the health department to focus on testing and vaccinating these vulnerable sections of the population.

## Appendix

### Appendix 1: Testing for Exogeneity of Regressors—The IV-2SLS Approach

In Table 8, 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 LOG(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 exogenous once the instruments are judiciously chosen.

### Appendix 2: 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).

**DEATHS**—weekly cumulative total COVID-19 deaths at the district level per lakh district population (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.php> for 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 140 mg/dl blood sugar, National Family Health 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](http://chiips.org/nfhs/factsheet_NFHS-5.shtml).

**Table 8** 2SLS-IV regression of exogeneity test for explanatory variables

Dependent variable: Log(DEATHS)				
Variable tested for exogeneity	Instruments chosen	Instrument rank	R-square in 2nd stage regression	Prob(J-statistic) and 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) DENSITY 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 17,824 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. Heteroscedasticity and autocorrelation adjusted standard errors are used throughout

**HYPER**—Hypertension, measured by the percentage of 15 years and above population at district level who suffer from elevated blood pressure (Systolic  $\geq 140$  mm of Hg and/or Diastolic  $\geq 90$  mm of Hg) or taking medicine to control blood pressure), National Family Health Survey 2019–20 (i.e., NFHS-5, available at [http://rchiips.org/nfhs/factsheet\\_NFHS-5.shtml](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  $> 25$  kg/m<sup>2</sup>), National Family Health Survey 2019–20 (i.e., NFHS-5, available at [http://rchiips.org/nfhs/factsheet\\_NFHS-5.shtml](http://rchiips.org/nfhs/factsheet_NFHS-5.shtml))

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

**RESP**—District level population suffering from respiratory diseases for 2019–20, measured by the inpatient—asthma, chronic obstructive pulmonary



disease (COPD) and respiratory infections (expressed per lakh district level population), drawn from Item-wise HMIS report of all States and Districts across months, available at <https://www.data.gov.in/catalog/item-wise-hmis-report-all-states-and-districts-across-months>.

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

### Appendix 3: List of 19 States and 4 Union Territories Considered in this District Level Study

**States:** Andhra Pradesh, Bihar, Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Sikkim, Tamil Nadu, Uttar Pradesh, Uttarakhand and West Bengal.

**Union Territories:** Chandigarh, the NCT of Delhi, Jammu & Kashmir, and Puducherry.

**Acknowledgements** The paper has benefitted from the comments of the participants at the Allied Social Science Association (ASSA) meeting in 2021, and particularly a referee from this journal. Hong Il Yoo is gratefully acknowledged for econometric insights. We also acknowledge Sushmita Maitra for providing insights about the physiological and epidemiological issues related to the COVID-19. The usual disclaimer applies.

**Data availability** All data used in this article are freely available in the public domain.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abedi, V., O. Olulana, V. Avula, D. Chaudhary, A. Khan, S. Shahjouei, and R. Zand. 2020. Racial, economic and health inequality and COVID-19 infection in the United States. *Journal of Racial and Ethnic Health Disparities* 8 (3): 732–742.
- Akdis, C.A. 2021. Does the epithelial barrier hypothesis explain the increase in allergy, autoimmunity and other chronic conditions? *Nature Reviews Immunology* 21 (11): 739–751.
- Albashir, A.A.D. 2020. The potential impacts of obesity on COVID-19. *Clinical Medicine* 20 (4): e109.
- Amaratunga, D., J. Cabrera, D. Ghosh, M.N. Katehakis, J. Wang, and W. Wang. 2021. Socioeconomic impact on COVID-19 cases and deaths and its evolution in New Jersey. *Annals of Operations Research* 317 (1): 5–18.
- Basu, P., C. Bell, and T.H. Edwards. 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-2020-2050>.
- Basu, P., and R. Mazumder. 2021. Regional disparity of COVID-19 infections: An investigation using state-level Indian data. *Indian Economic Review* 56: 215–232.

- Basu, P., and K. Sen. 2020. Age composition of population and COVID-19. *Ideas for India*, 10 April (Perspectives).
- Blaser, M.J. 2017. The theory of disappearing microbiota and the epidemics of chronic diseases. *Nature Reviews Immunology* 17 (8): 461–463.
- Böbel, T.S., S.B. Hackl, D. Langgartner, M.N. Jarczok, N. Rohleder, G.A. Rook, et al. 2018. Less immune activation following social stress in rural vs. urban participants raised with regular or no animal contact, respectively. *Proceedings of the National Academy of Sciences* 115 (20): 5259–5264.
- Canatay, A., T.J. Emegwa, and M.F.H. Talukder. 2021. Critical country-level determinants of death rate during COVID-19 pandemic. *International Journal of Disaster Risk Reduction* 64: 102507.
- Celebi Sozener, Z., B. Ozdel Ozturk, P. Cerci, M. Turk, B. Gorgulu Akin, M. Akdis, S. Altiner, U. Ozbey, I. Ogulur, Y. Mitamura, I. Yilmaz, K. Nadeau, C. Ozdemir, D. Mungan, and C.A. Akdis. 2022. Epithelial barrier hypothesis: Effect of the external exposome on the microbiome and epithelial barriers in allergic disease. *Allergy* 77 (5): 1418–1449.
- Chang, D., X. Chang, Y. He, and K.J.K. Tan. 2022. The determinants of COVID-19 morbidity and mortality across countries. *Scientific Reports* 12 (1): 5888.
- Cortis, D. 2020. On determining the age distribution of COVID-19 pandemic. *Frontiers in Public Health* 8: 202.
- Dang, A., and I. Gupta. 2021. *Obesity and its impact on COVID occurrence: Evidence from India*. IEG Working Papers 430, Institute of Economic Growth.
- Divo, M. J., Martinez, C. H., and Mannino, D. M. 2014. Ageing and the epidemiology of multimorbidity. *European Respiratory Journal*, 44(4), 1055–1068.
- Ehlert, A. 2021. The socio-economic determinants of COVID-19: a spatial analysis of German county level data. *Socio-Economic Planning Sciences* 78: 101083.
- Escobedo-de la Peña, J., Rascón-Pacheco, R. A., de Jesús Ascencio-Montiel, I., González-Figueroa, E., Fernández-Gárate, J. E., Medina-Gómez, O. S., and Borja-Aburto, V. H. 2021. Hypertension, diabetes and obesity, major risk factors for death in patients with COVID-19 in Mexico. *Archives of Medical Research*, 52(4), 443–449.
- Gardiner, J., J. Oben, and A. Sutcliffe. 2021. Obesity as a driver of international differences in COVID-19 death rates. *Diabetes, Obesity and Metabolism* 23 (7): 1463–1470.
- Ghosh, A., and A. Mukherji. 2014. Air pollution and respiratory ailments among children in urban India: Exploring causality. *Economic Development and Cultural Change* 63 (1): 191–222.
- González-Val, R., and F. Sanz-Gracia. 2022. Urbanization and COVID-19 incidence: A cross-country investigation. *Papers in Regional Science* 101 (2): 399–415.
- Haklai, Z., M. Aburbeh, N. Goldberger, and E.S. Gordon. 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.
- Hanski, I., L. von Hertzen, N. Fyhrquist, K. Koskinen, K. Torppa, T. Laatikainen, et al. 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. *Proceedings of the National Academy of Sciences* 109 (21): 8334–8339.
- Hawkins, R.B., E.J. Charles, and J.H. Mehaffey. 2020. Socio-economic status and COVID-19 related cases and fatalities. *Public Health* 189: 129–134.
- Ho, F.K., F. Petermann-Rocha, S.R. Gray, B.D. Jani, S.V. Katikireddi, C.L. Niedzwiedz, and J.P. Pell. 2020. Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. *PLoS ONE* 15 (11): e0241824.
- Hsu, C.Y., C.C. Lai, Y.P. Yeh, C. Chang-Chuan, and H.H. Chen. 2021. Progression from pneumonia to ARDS as a predictor for fatal COVID-19. *Journal of Infection and Public Health* 14 (4): 504–507.
- Ioannidis, J.P., C. Axfors, and D.G. Contopoulos-Ioannidis. 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., and A. Sen. 2020a. Understand the method in COVID-19's madness. India doesn't need complete lockdown. *The Print*.
- Jalan, J., and A. Sen. 2020b. Containing a pandemic with public actions and public trust: The Kerala story. *Indian Economic Review* 55 (1): 105–124.
- Jalan, J., and A. Sen. 2020c. Spread of COVID-19 in India: Across space and over time. <https://indiaCOVIDspread.weebly.com>.
- Klang, E., Kassim, G., Soffer, S., Freeman, R., Levin, M. A., and Reich, D. L. 2020. Severe obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50. *Obesity*, 28(9), 1595–1599.

- Mahamat-Saleh, Y., Fiolet, T., Rebeaud, M. E., Mulot, M., Guihur, A., El Fatouhi, D., and Severi, G. 2021. Diabetes, hypertension, body mass index, smoking and COVID-19-related mortality: a systematic review and metaanalysis of observational studies. *BMJ Open*, 11(10), e052777.
- Maji, S., S. Ghosh, and S. Ahmed. 2018. Association of air quality with respiratory and cardiovascular morbidity rate in Delhi, India. *International Journal of Environmental Health Research* 28 (5): 471–490.
- Mandi, J., M. Chakrabarty, and S. Mukherjee. 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.
- Montazersaheb, S., S.M. HosseiniyanKhatibi, M.S. Hejazi, V. Tarhriz, A. Farjami, F. Ghasemian Sorbeni, et al. 2022. COVID-19 infection: An overview on cytokine storm and related interventions. *Virology Journal* 19 (1): 1–15.
- Naudé, W., and P. Nagler. 2022. COVID-19 and the city: Did urbanized countries suffer more fatalities? *Cities* 131: 103909.
- Rana, R., Ranjan, V., Kumar, N., Chugh, P., Khillan, K., Gogia, A., and Ganguly, N. K. 2023. Association of underlying comorbidities and progression of COVID-19 infection amongst 2586 patients hospitalised in the National Capital Region of India: a retrospective cohort study. *Molecular and Cellular Biochemistry*, 478(1), 149–160.
- Ray, D., and S. Subramanian. 2020. India's lockdown: An interim report. *Indian Economic Review* 55 (1): 31–79.
- Rook, G.A.W. 2023. The old friends hypothesis: Evolution, immunoregulation and essential microbial inputs. *Front Allergy* 4: 1220481.
- Rook, G.A., C.A. Lowry, and C.L. Raison. 2013. Microbial 'old friends', immunoregulation and stress resilience. *Evolution, Medicine, and Public Health* 2013 (1): 46–64.
- Sannigrahi, S., F. Pilla, B. Basu, A.S. Basu, and A. Molter. 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.
- Schellekens, P., and D. M. Sourrouille. 2020. *COVID-19 mortality in rich and poor countries: A tale of two pandemics?* World Bank Policy Research Working Paper (9260).
- Shimojo, N., and K. Izuhara. 2017. Old friends, microbes, and allergic diseases. *Allergy International* 66 (4): 513–514.
- Stafford, K., M. Hoyer, and A. Morrison. 2020. Racial toll of virus grows even starker as more data emerges. *AP News*, April 19 <https://apnews.com/article/8a3430dd37e7c44290c7621f5af96d6b>.
- Steinheuser, V., K. Ackermann, P. Schönfeld, and L. Schwabe. 2014. Stress and the city: Impact of urban upbringing on the (re) activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine* 76 (9): 678–685.
- Tan, M., F.J. He, and G.A. MacGregor. 2020. Obesity and COVID-19: the role of the food industry. *BMJ* 369: m2237.
- Tibirica, E., and De Lorenzo, A. 2020. Importance of the evaluation of systemic microvascular flow and reactivity in critically ill patients with coronavirus disease 2019—COVID-19. *Microvascular Research*, 131, 104028.
- Tzotzos, S.J., B. Fischer, H. Fischer, and M. Zeitlinger. 2020. Incidence of ARDS and outcomes in hospitalized patients with COVID-19: A global literature survey. *Critical Care* 24: 1–4.
- Wang, J., T. Sato, and A. Sakuraba. 2021. Worldwide association of lifestyle-related factors and COVID-19 mortality. *Annals of Medicine* 53 (1): 1531–1536.
- Windsor-Shellard, B., and Kaur, J. 2020. Coronavirus (COVID-19) related deaths by occupation, England and Wales: deaths registered up to and including 20 April 2020. Statistical Bulletin. London: Office for National Statistics.
- Yanez, N.D., N.S. Weiss, J.A. Romand, and M.M. Treggiari. 2020. COVID-19 mortality risk for older men and women. *BMC Public Health* 20 (1): 1–7.
- Yang, X.D., X.Y. Su, H.L. Li, R.F. Ma, F.J. Qi, and Y.E. Cao. 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.

## Authors and Affiliations

**Parantap Basu<sup>1</sup> · Susmita Das<sup>2</sup> · Arnab Dutta Choudhury<sup>2</sup> · Ritwik Mazumder<sup>2</sup> · Mahuya Sengupta<sup>3</sup>**

✉ Parantap Basu  
parantap.basu@durham.ac.uk

Susmita Das  
dsusmita521@gmail.com

Arnab Dutta Choudhury  
arnabdutta.duttachoudhury@gmail.com

Ritwik Mazumder  
ritwikmazumder@gmail.com; ritwik.mazumder@aus.ac.in

Mahuya Sengupta  
senguptamahuya35@gmail.com

<sup>1</sup> Durham University Business School, Durham, UK

<sup>2</sup> Department of Economics, Assam University, Silchar, India

<sup>3</sup> Department of Biotechnology, Assam University, Silchar, India