Role of multiresolution vulnerability indices in COVID-19 spread in India: a Bayesian model-based analysis

Rupam Bhattacharyya, Anik Burman, Kalpana Singh, Sayantan Banerjee, Subha Maity, Arnab Auddy, Sarit Kumar Rout, Supriya Lahoti, Rajmohan Panda, Veerabhadran Baladandayuthapani

ABSTRACT

Objectives COVID-19 has differentially affected countries, with health infrastructure and other related vulnerability indicators playing a role in determining the extent of its spread. Vulnerability of a geographical region to COVID-19 has been a topic of interest, particularly in low-income and middle-income countries like India to assess its multifactorial impact on incidence, prevalence or mortality. This study aims to construct a statistical analysis pipeline to compute such vulnerability indices and investigate their association with metrics of the pandemic growth.

Design Using publicly reported observational socioeconomic, demographic, health-based and epidemiological data from Indian national surveys, we compute contextual COVID-19 Vulnerability Indices (cVIs) across multiple thematic resolutions for different geographical and spatial administrative regions. These cVIs are then used in Bayesian regression models to assess their impact on indicators of the spread of COVID-19.

Setting This study uses district-level indicators and case counts data for the state of Odisha, India.

Primary outcome measure We use instantaneous R (temporal average of estimated time-varying reproduction number for COVID-19) as the primary outcome variable in our models.

Results Our observational study, focussing on 30 districts of Odisha, identified housing and hygiene conditions, COVID-19 preparedness and epidemiological factors as important indicators associated with COVID-19 vulnerability.

Conclusion Having succeeded in containing COVID-19 to a reasonable level during the first wave, the second wave of COVID-19 made greater inroads into the hinterlands and peripheral districts of Odisha, burdening the already deficient public health system in these areas, as identified by the cVIs. Improved understanding of the factors driving COVID-19 vulnerability will help policy makers prioritise resources and regions, leading to more effective mitigation strategies for the present and future.

INTRODUCTION

The outbreak of the highly infectious COVID-19, caused by the SARS-CoV-2, emerged in China and spread widely globally. As of 16 July 2021 India recorded more than 31 million confirmed cases, of which around 425,000 (1.4%) were active; more than 30 million (97.3%) recovered; and more than 412,000 (1.3%) have died. India, with a population of more than 1.34 billion, has a high population density of 454 people/km², almost three times that of China, and faces multiple challenges to tackle COVID-19.

While a lot of research has focused on clinical outcomes, epidemiological modelling and transmission dynamics of the novel COVID-19, less focus has been placed on preponderance of risk and vulnerability to contracting the disease ascribed to the social economic environment. Emerging studies have begun to report on the impacts of social vulnerability on COVID-19 from an incidence and outcome standpoint. However, the resolution of most studies has been at the global or country level, and less attention has been paid to a subregional or subnational level. This is important in the Indian context as the district is the unit for administrative management and delivery of health services.
Vulnerability in the present context of COVID-19 is more than just the risk of contracting the disease. It is described as a dynamic concept—a person or a group might not be vulnerable at the onset of the pandemic but could subsequently become vulnerable, depending on policies and response at the country or state level. The term vulnerability implies a measure of risk or consequences associated with socioeconomic factors and financial, mental and physical coping mechanisms resulting from a system’s ability to cope with the pandemic. Vulnerability manifests itself in different forms. Beyond epidemiological vulnerability to COVID-19 (eg, elderly people and individuals with comorbidities), low-income and middle-income countries (LMICs) such as India are characterised by other concerns: transmission vulnerability (eg, population density, mobility and social structures, household and social structures), health system vulnerability (eg, availability of formal healthcare providers and intensive care) and vulnerability to direct control measures (eg, impact of quarantines, lockdowns, self-isolation and disrupted social interactions). Poor public health infrastructure places India at a strategic disadvantage in this pandemic since it increases the risk of exposure to vulnerability.

To this end, we created an interactive algorithm to estimate the COVID-19 Vulnerability Indices (cVIs) across multiple geographical districts to COVID-19 and inform our understanding of the most potentially vulnerable district(s). We hypothesise that more vulnerable districts are less able and prepared to detect, respond to and prevent COVID-19 spread, whereas more resilient districts are better able to do so. We computed vulnerability indices (VI) across five factors: socioeconomic and demographic composition, housing and hygiene conditions, availability of healthcare facilities, preparedness of COVID-19 and epidemiological factors. These factors were then individually and collectively assessed to identify geographical locations that are disproportionately exposed to the risk of infection and/or disease severity over the pandemic timeline across the districts of Odisha, an eastern state in India which has reasonably managed its influence on resources and possession of assets as a proxy for poverty. In total, four indicators (figure 1 and online supplemental table 1) were used to define the socioeconomic demographics. Furthermore, housing and hygiene conditions, particularly uncontaminated and safe drinking water sources, healthy sanitation behaviour and clean fuel for cooking, are important factors in COVID-19 transmission since they form the first line of defence against the infection. Exposure to smoke associated with the use of unclean or solid fuels is associated with harmful health effects as well and thus constitutes an important domain of vulnerability. We considered a total of three indicators in this group, namely, clean drinking water, non-shared toilet facilities and clean fuel for cooking (figure 1 and online supplemental table 1).

Two indicators were used to define availability of healthcare (figure 1 and online supplemental table 1). The management of an epidemic and the treatment-seeking ability of a population depend on easy and affordable access to well-capacitated healthcare systems and health security, and thus should be included in the VI. The preparedness for COVID-19 was represented by four indicators (figure 1 and online supplemental table 1). The ability of a country/region to respond to the pandemic depends heavily on the healthcare system capacity and preparedness. Besides, there are several known epidemiological factors and underlying medical conditions that might put a population at risk of higher morbidity and mortality from COVID-19 infection and thus merit inclusion in the VI. We considered six such indicators to define epidemiological factors (figure 1 and online supplemental table 1).

District-level COVID-19 count data on confirmed cases, deaths and recoveries were obtained from the COVID-19 India Dashboard.
Computation of relative cVIs

The district-level covariates considered for the analyses varied in their scales and interpretations and covered a range of continuous (e.g., per capita net district domestic product), count (e.g., district-wise population) and percentage/proportion (e.g., percentage of households with improved sanitation) data. The cVIs were computed following a similar methodology to that developed by Flanagan and colleagues, used by the Centers for Disease Control and Prevention to compute social VIs for each census track in the USA.29 For the purposes of downstream interpretations and analyses, we aggregated these variables into unified VIs offering two major advantages. First, transforming them to a common scale of values between 0 and 1, with lower values (closer to 0) indicative of low vulnerability and analogously higher values (closer to 1) indicative of higher vulnerability, allows coherent interpretations across the different indices. Second, the common scale also allows us to extract regression coefficients and other associated metrics at comparable scales from models with more stable fits, as we show further.

We considered a total of 25 district-level indicators grouped into five broad themes, as described in the previous section and in figure 1 and online supplemental table 1. The amount of risk associated with a particular value of an indicator is interpreted as how detrimental its effects are on the population. The risk is higher towards the terminal value of an indicator where it is potentially

---

**Figure 1** Summary of the five themes and the underlying variables used to compute the vulnerability indices. Each primary rectangle connected to the central oVI box contains the name of one theme, and the corresponding sublabels list the indicators included within that theme. All data sources are summarised in online supplemental table 1. ICU, intensive care unit; NDDP, net district domestic product; oVI, Overall Vulnerability Index.

Hypertension
Diabetes
Obesity
Cancer (Oral, Breast and Cervix)
Pneumonia
Plasmodium
Vivax test
HIV

- Population
- Literacy rate
- Work participation
- Per capita NDDP
- Households with improved drinking
- Households using improved sanitation
- Households using clean fuel for cooking
- Total beds capacity (at district level)
- Total ICU beds (at district level) facility
- Temporary medical camps (at district level)
- COVID Hospital testing centres (at district level)
- Hypertension
- Diabetes
- Obesity
- Cancer (Oral, Breast and Cervix)
- Pneumonia
- Plasmodium
- Vivax test
- HIV

Overall Vulnerability Index (oVI)

Available of public hospitals (at district level)
General number of beds (at district level) facility

Socioeconomic and demographic
Housing and hygiene condition
Preparedness of COVID
Epidemiological factors
Availability of health-care facilities
causing a worse influence. In other words, the risk is high for high values of those indicators (covariates) which have a negative influence on the population (eg, disease proportion for districts); likewise, the risk is low for high values of those indicators which have a positive influence (eg, hospital/bed counts). We adapt the algorithm by Acharya and Porwal for the calculation of the relative VIs for each district.\(^\text{10}\) The algorithm consists of three steps, as summarised in the left panel of figure 2 (online supplemental figure 1) and mathematical derivations are provided in online supplemental text section 1.2, equations 1–3. We compute indicator-specific relative cVI, theme-specific cVI and an Overall Vulnerability Index (oVI) for each district. A numerical illustration for the computation of these VIs for a specific district is provided in online supplemental text section 1.3.

Estimation and summarisation of time-varying reproductive number (R)

We refer to the effective reproduction number as ‘R’ throughout, which is similar to the concept of \( R_0 \) in typical compartmental epidemiological models\(^\text{36}\) but different in the sense that \( R_0 \) is a constant that is inherent to the virus and the affected population and is neither time-varying nor impacted by interventions such as social distancing or lockdown, which is the case for R. We compute indicator-specific relative cVI, theme-specific cVI and an Overall Vulnerability Index (cVI) for each district. A numerical illustration for the computation of these VIs for a specific district is provided in online supplemental text section 1.3.

Regression analyses using summaries of time-varying R profiles

We undertake regression analyses using the summaries of R profiles (iR and vR) as response variables and the cVIs (theme-specific and within-theme) as our covariates to identify the associations of these cVIs with the R summaries and estimate the extent to which changes in the cVIs impact them. The cVIs corresponding to the themes socioeconomic and demographic indices (cVISHD), housing and hygiene conditions (cVIHH), preparedness for COVID-19 (cVIPC), epidemiological factors (cVIEF) and availability of healthcare (cVIAH) lead to a final regression model of the following form:

\[
\text{iR} = \beta_0 + \beta_1 \text{cVISHD} + \beta_2 \text{cVIHH} + \beta_3 \text{cVIPC} + \beta_4 \text{cVIEF} + \beta_5 \text{cVIAH} + \text{error}.
\]

To fit this multiple linear regression model, we use a Bayesian model averaging (BMA) procedure implemented via the BMS package in R.\(^\text{37}\) Briefly, BMA performs a search across the weighted posterior probabilities of the plausible models (in terms of possible variable combinations) and selects the model with the highest posterior probability. A joint Zellner g-prior is used for the full set of coefficients, \( \beta_s \), with a standard non-informative prior.
for the dispersion parameter of the error term. Previous studies have shown the extensive utility of these choices in other settings concerning association and variable selection questions. More details regarding the model choices and the fitting procedure are available in online supplemental text section 1.5, equation 6.

For each covariate and its corresponding coefficient, two quantities generated from the Bayesian model are of interest to us, providing two complementary sets of information on the effect of that covariate on the outcome. We use posterior inclusion probabilities (PIPs) provided by the BMS package as estimates of variable importance in the fitted model, summarising the relative contributions of the covariates towards explaining the variability in iR across districts. The PIPs have classically been used in Bayesian machine learning frameworks in context of variable selection. Similar to p values in the context of frequentist models, PIPs provide a numerical option to quantify variable significance both in absolute and relative ways (within the set of variables included in the model) and can be interpreted in the context of vulnerability as evidence in favour of impact of different cVIs on the pandemic growth. The specific values of the estimated coefficients are of interest in terms of quantifying the exact effect of a covariate on the outcome, if any. For example, a value of \( \beta > 0 \) means the VI corresponding to this coefficient has a positive association with the outcome; that is, an increase in vulnerability of one unit (moving from the least vulnerable to the most vulnerable position) in terms of that index results in a \( \beta \) amount of increase in the iR, and hence the pandemic grows.

We also fit similar models using iR as the response variable and within-theme cVIs as the covariates (five models, one for each theme). The details for these models are in online supplemental text section 1.5, equation 7. Correspondingly, we fit overall (online supplemental text section 1.5, equation 8) and theme-specific (online supplemental text section 1.5, equation 9) models using vR as the response. Online supplemental text section 1.6 provides a discussion of the Bayesian model averaging procedure implemented via the BMS package and the computation of the PIPs. We also assess the validity of our Bayesian regression framework via simulation studies. For the simulations, we use covariates taking values in the same range as our real data covariates, that is, the cVIs, and assess selection performance using standard metrics across varying ranges of effect size, signal-to-noise ratio and sample size. The details and results of the simulation studies are presented in online supplemental text section 1.7. Further, we have discussed the unique features and utilities of our pipeline over existing vulnerability-driven computational approaches in online supplemental text section 1.8.

**Patient and public involvement statement**

Our study does not involve the participation of patients or any members of the public, nor does it use patient-level or identifiable data.

**RESULTS**

Using the indicators listed in figure 1 and online supplemental table 1, we developed an area-based composite COVID-19 susceptibility and VIs at the district level for Odisha, with a framework towards providing policy makers with some indication on which districts are likely to be most susceptible or vulnerable to a COVID-19 outbreak and specifically where the government should potentially target its resources and accordingly plan data-driven intervention strategies. The elicited results from these indices are presented further.

We first summarise the case incidence data across the 30 districts of Odisha at both state and district levels between the dates 1 May 2020 and 15 April 2021 in figure 3. Some key takeaways can be obtained from a visual inspection of figure 3 that allow us to understand the pattern of the pandemic growth in Odisha during the timeline of interest and interpret some of the results obtained via further and more nuanced analyses.

- **Figure 3A,B** indicate spikes/risks of cases in May 2020, July–September 2020 and April 2021, most recently in the data. The stable and/or decreasing patterns of cases reported during the other periods could possibly be attributed to either strict lockdown and associated measures or undercounting of cases due to limited testing capacity.

- Some district-specific patterns emerge from figure 3C,E. The SARS-CoV-2 was mostly active in districts such as Cuttack, Deogarh, Dhenkanal, Kalahandi and Puri (showing high values of R in red) throughout the lockdown until the end of Unlock V.4.0.

- All districts showed controlled values of R during Unlock V.5.0–V.7.0, but it tended to increase in the initial months of 2021. **Figure 3D** indicates that as of the first fortnight of April 2021, all the districts experience \( R > 1 \), that is, further growth of the pandemic. **Figure 4** shows that the spatial map of the cVIs along with the epidemic spread across all the districts of the state, with the colour grading of the districts indicating the oVI values and the sizes of the dots indicating iR values. Clearly, all the districts have iR larger than 1. On observation of the spatial plots for overall vulnerability for Odisha, some dissimilarities among the VI and district-wise cases are witnessed. The districts of Malkangiri, Nabarangapur, Rayagada and Mayurbhanj have relatively high oVI values, as seen in figure 4A. Within these, Mayurbhanj shows a relatively low iR value. On the other hand, districts like Sambalpur and Balasore have high values of iR despite having low oVI values. Not all these results are along the direction of our expectations. Hence, we further explore the individual theme-level indices based on figure 4B–F, with the colours of the districts now indicating the theme-specific cVI values.

For \( \text{cVI}_{\text{sp}} \) districts like Ganjam, Balangir and Kendrapara show relatively high cVI with relatively low values of iR compared with the five districts with highest iR. Districts such as Gajapati, Sambalpur, Kendhramal and

Boudh show high values of iR despite having low cVI values for this theme (figure 4B). For cVI_{HI}, districts like Malkangiri, Kandhamal and Boudh have high cVI values as well as high values of iR. On the other hand, districts like Puri, Cuttack and Bargh show low iR values as well as low values of cVI (figure 4C). For cVI_{AH}, we see that the expected pattern of relatively high cVI associated with high iR is generally followed. The districts of Gajapati, Deogarh and Boudh fall in this class. However, districts such as Kendujhar and Nuapada are exceptions (figure 4D). Similarly, for cVI_{PC}, districts like Gajapati, Deogarh and Malkangiri have high values of both cVI and iR. On the other hand, districts like Koraput, Puri and Subarnapur show low values of both metrics. However, there are exceptions where the iR and cVI are in opposite directions, such as the districts of Balasore, Sambalpur and Sundargarh (low cVI and high iR) and Rayagada and Jajpur (high cVI and low iR) (figure 4E). For cVI_{EF}, we see that the districts of Ganjam, Puri, Rayagada and Koraput show relatively low iR despite a high cVI. Boudh, Sambalpur and Balasore, on the other hand, show low cVI with high iR values (figure 4F).

All districts have an iR greater than 1 in the second wave. Figure 4G shows that three clusters of districts with most common vulnerability were identified, where red, blue and green colours indicate high, moderate and low iR districts, respectively. The theme in bold is the most commonly occurring vulnerability in the cluster. Districts with relatively higher iR are more influenced by preparedness of COVID-19, housing and hygiene and availability of healthcare. In other districts with moderate iR, sociodemographic and epidemiological factors have more influence. Districts with relatively lower iR are more influenced by preparedness for COVID-19. District preparedness indicators influence R more, and this is further accentuated by sociodemographic and household variables.

We also observed clustering patterns in the overall as well as individual theme spatial plots. Similar shades of VI values are clustered in regions. Highly vulnerable districts (darker shades) are clustered around the southern and southwestern regions of the state, whereas lower vulnerability districts (lighter shades) are clustered near the central and northern part of state in case of overall
vulnerability. Similar patterns are exhibited by theme-level VIs like epidemiological, socioeconomic and demographic factors. COVID-19 preparedness and housing and hygiene condition themes display different clustering patterns. These two sets of clusters exhibit differing sets of iR values. Nevertheless, the analysis devised from figure 4 is largely correlative, and the interpretations are based on visual inspections. Hence, in order to understand the explicit dependence between the covariates (cVIs) and responses (iR and vR), we present the results of our Bayesian linear regression analysis next.

Figure 5 and online supplemental figure 2 and table 3 summarise the relative ranking for the overall and theme-specific VIs in terms of their PIPs, along with the signs of the estimated regression coefficients with iR as the response. In the absence of any other information, since the cVIs are all interpretable in ‘the higher the worse’ direction, we would largely expect the signs to be positive. Figure 5A (and online supplemental table 3), displaying the relative importance of the different theme-specific cVIs, confirms that cVI_{HH} is ranked the highest (PIP=0.45, $\beta_{HH}=0.18$, SD=0.26), followed closely by cVI_{PC} (PIP=0.42, $\beta_{PC}=0.18$, SD=0.27) and cVI_{EF} (PIP=0.42, $\beta_{EF}=0.17$, SD=0.26). These results indicate that the total posterior probability of explored models where these respective variables were included each exceed 40%. Further, a unit change in

---

**Figure 4** Summary of overall and themed VIs and iR across the 30 districts of Odisha. (A) Overall Vulnerability Index. (B–F) Five themed VIs, as mentioned in the individual panel titles. The size of the red dots is proportional to the corresponding iR on 15 April 2021. (G) Summary of the clusters of districts colour coded and categorised by iR with associated vulnerability themes: SD, HH conditions, PC, EFs and AH. Colours red, blue and green indicate high, moderate and low iR districts, respectively. The theme in bold is the most commonly occurring vulnerability in the cluster. AH, availability of healthcare; cVI, COVID-19 Vulnerability Index; EF, epidemiological factor; HH, housing and hygiene; iR, instantaneous R; PC, preparedness for COVID-19; SD, socioeconomic and demographic; VI, vulnerability index.

---

**Figure 5** Summary of regression models for iR with the VIs as covariates. (A) Themed VIs constituting the Overall Vulnerability Index. (B,C) Indicators within the themes ‘availability of healthcare’ and ‘COVID-19 preparedness’, as mentioned in the individual panel titles. In each case, a Bayesian averaging-based linear regression model is fit using iR as response and the indicators/themes as covariates. The heights of the bars indicate the posterior inclusion probabilities for the covariates in the fitted models, and the labels on top of the bars indicate the signs of the estimated coefficient as obtained in those models. iR, instantaneous R; VI, vulnerability index.
each of these cVIs (shift from the least vulnerable to the most vulnerable district for the corresponding theme) result in approximately 0.18, 0.18 and 0.17 increments, respectively, in the iR, indicating that the pandemic grows when the vulnerabilities increase. Only one out of these five themes yielded cVIs with a negative coefficient –cVI_{SD} (PIP=0.20, \beta=−0.04, SD=0.15). The negative sign of the coefficient is apparently counterintuitive, since this indicates that with an increase in vulnerability in terms of socioeconomic and demographic factors, the pandemic actually slows down and the iR decreases. However, as evident from the PIP, the variable ranks lowest in the model and would not be selected at any sensible cut-off, which means the effect of this variable would actually be negligible on iR.

To further investigate the actual association of the themed cVIs on iR, we also summarise the theme-specific models in figure 5B,C, online supplemental figure 2A–C and online supplemental table 3. Except for cVI_{SD} and cVI_{EF}, most of the indicators in all the other groups show an intuitive (positive) direction of the coefficients. For each theme, a few indicators stand out, such as population (PIP=0.66, \beta=−0.35, SD=0.32) and literacy rate (PIP=0.48, \beta=0.23, SD=0.30) (cVI_{SD}), clean fuel (PIP=0.59, \beta=0.37, SD=0.42) and improved drinking (PIP=0.36, \beta=0.13, SD=0.23) (cVI_{HI}), and public hospitals (PIP=0.62, \beta=0.35, SD=0.36) (cVI_{HC}). Specifically, for figure 5C, all cVI_{RC} values are calculated in a way such that the higher the cVI, the lower the availability of the corresponding resource is for a district and vice versa. As can be seen from the signs of the estimated regression coefficients on top of the bars in figure 5C and online supplemental table 3, for each of these cVIs, an increment in vulnerability (decrease in availability of the corresponding resource) causes the iR, and hence the pandemic spread, to increase. For example, for the cVI with the highest importance in this category, the number of temporary medical camps set up in rural regions, a unit increment in the cVI (movement from least vulnerable to most vulnerable among all the districts) increases the mean of the iR by an estimated quantity of 0.33. This means that compared with the least vulnerable district in terms of temporary medical camps set up in rural regions, the most vulnerable district with respect to the same experiences, on an average, a 33% increase in infection rate per infected individual. The corresponding PIP for this variable tells us that the total posterior probability of the explored models where this variable is included is 0.59. Online supplemental figure 3 and table 4 offer the summaries of the similar regression models with vR as the response.

**DISCUSSION**

Our study makes a novel attempt to examine heterogeneities in the underlying socioeconomic vulnerabilities related to COVID-19 risk across districts in the eastern state of Odisha using a composite index. Using districts in Odisha as an example, we have developed a set of cVIs that can be adopted by the healthcare, public health and data science) communities for designing more effective intervention strategies. Mapped to the districts, the cVIs provide information that is useful for emergency response planning and mitigation at a relatively granular level and can help support response planning for the current and future epidemics (figure 4G).

The heat map (figure 3C) shows if the districts show patterns of ‘red’ or ‘green’ zones, indicating the burden of cases over the pandemic timeline. From the end of September 2020, cases begin to decline, gradually plateauing towards the end of December 2020. From April 2021, all districts show a high burden of cases as well as high values of iR. In the beginning of April 2021, increased infection rates were reported from the western and southern border districts. The pattern of vulnerability identified shows that the majority of the districts that are tribal also as well as vulnerable to natural disasters also show high overall and socioeconomic vulnerability and a high number of cases for COVID-19. Many tribal districts display vulnerability across all themes. Across the state, higher vulnerability districts are mostly clustered in the southern and southwestern regions. Our findings about clustering of districts has been reflected in previous studies which have identified clusters of vulnerable districts to COVID-19 at the national level in India, as well as social and epidemiologically vulnerable clusters at the subcounty level in Kenya. Another study at the subnational level also identified clusters of socially vulnerable districts to the Ebola virus disease in Africa.

**Socioeconomic and demographic compositions**

Poor socioeconomic conditions may be associated with a poorer recovery, reflecting the importance of this theme. In addition to such indicators, access to impaired healthcare services and facilities can also significantly deteriorate immune response and patient recovery. Most districts that report a poor literacy rate as compared with the state average also have a higher iR (PIP=0.48, \beta=0.23, SD=0.30). Compared with the least vulnerable district in terms of literacy rate, the most vulnerable district with respect to the same was seen to experience, on an average, a 28% increase in infection rate per infected individual.

**Housing and hygiene conditions**

The lack of proper home structure and the lack of access to minimum resources, such as water and basic sanitation in districts, can increase the risk of illness due to COVID-19, as observed with other respiratory diseases. For example, poor quality housing is associated with certain health outcomes; damp housing can lead to respiratory diseases such as asthma, while overcrowding can result in higher infection rates. Recent research has identified SARS-CoV-2 in the faeces of patients with COVID-19, suggesting possible ‘faecal–mucosal transmission’ in public toilets or areas with poor sanitation. Most districts which show high number of cases in the second wave have poor coverage of sanitation (Malkangiri, 17%;
Nabarangapur, 16%; Kandhamal, 17%; and Boudh, 16%) as compared with those districts reporting lower cases (Ganjam, 41%; Cuttack, 39%; and Khordha, 47%).

Further, previous research has demonstrated that lack of access to clean water and sanitation facilities leads to exacerbation of viral transmission of viral infections like the respiratory Avian influenza H7N9 virus. As per the National Family Health Survey in 2016, Odisha has one of the of the highest rates (65%) for open defecation. There is a significant urban–rural disparity with the proportion practising open defecation ranging from 28.3% in urban to 72.4% in rural areas in Odisha. Many of the vulnerable districts have large populations living in rural areas where open defecation is common. It is important to generate public awareness about the risk of COVID-19 and precautions to avoid the same. This is further supported by the findings related to safe sanitation practices and exposure to mass media that seem to be associated with lower risk of COVID-19 in other parts of India.

Availability of healthcare facilities
Access to healthcare is lower in disadvantaged and marginalised communities, even in countries which have universal healthcare systems. In Brazil, unequal distribution in resources have resulted in regional and social healthcare inequalities. The Ebola outbreak in Africa (2014–2016) has demonstrated the importance of timely and quality healthcare from resilient health systems to prevent, detect and respond to diseases of pandemic potential. The state of Odisha has 83% population living in rural areas and has a deficit of health infrastructure. The state has a significant shortfall of doctors, specialists and health human resources in rural areas with one subcentre for 5551 people, one primary health centre for 28 822 persons and one community health centre covering a population of 98 469. This deficit is more pronounced in the aspirational districts which have limited coping-up capacities. Our model shows that, in the second wave, certain aspirational districts have recorded a greater number of cases. Our findings reveal that infrastructure measures taken for COVID-19 preparedness had a high variable importance in the overall model (PIP=0.42, β =-0.18, SD=0.27). Compared with the least vulnerable district in terms of overall COVID-19 preparedness, the most vulnerable district with respect to the same was seen to experience, on an average, an 18% increase in infection rate per infected individual. Preparedness for COVID-19 was also an important factor in the 11 districts that have reported lower iR (figure 4G). The government of Odisha adopted a decentralised approach by setting up temporary medical camps for the migrant influx at the Gram Panchayat level to arrest community spread of the virus. Such camps enable early detection and action to contain the spread of infection. Historically, camps offering institutional quarantine facilities have been at the core of multicomponent strategies for controlling communicable disease outbreaks.

Epidemiological factors
In the COVID-19 pandemic, obesity has been identified as a major risk factor and shown to be associated with disease progression. Research has shown that people with underlying uncontrolled medical conditions such as diabetes, hypertension and lung, liver and kidney diseases; patients with cancer on chemotherapy; smokers; transplant recipients; and patients taking steroids chronically were at increased risk of COVID-19 infection. Brazil reported a prevalence of 83% of comorbidities in around 17 752 COVID-19-related deaths, with the leading two comorbidities being chronic heart disease and diabetes. Odisha also had a prevalence of 33.7% of obesity rates. Districts such as Deogarh, Boudh and Nabrangapur that showed a high number of cases in both waves also have large numbers of people living with comorbidities such as obesity, diabetes and hypertension. In our model, within the epidemiological factors, obesity ranked a close second (PIP=0.72, β=−0.41, SD=0.33) following the lung disease tuberculosis. While developing responses to tackle the current epidemic and future pandemics, it would be useful to note that districts that have high epidemiological susceptibility need enhanced vigilance (through special rapid-action teams) to avoid transmissions.

Odisha’s management of the pandemic
There is some evidence that Odisha has done well in comparison to other states in terms of handling the pandemic. Odisha took many proactive measures such as a decentralised community-based approach at the very beginning of the pandemic. At a later stage of the pandemic, Odisha was recognised for its handling of the COVID-19 pandemic by the WHO. By October 2020, that is, after the first wave of the pandemic had subsided, the state had a very low fatality rate (0.42% against the national average of 1.51%). Such metrics put the performance of Odisha relative to other Indian states at a better standing.

CONCLUSION
This paper constructs a multidimensional VI to account for both long-term structural vulnerabilities and the recent weaknesses unearthed by the pandemic. It contributes to the debate on vulnerability measurement by contrasting a narrow focus on epidemiological or environmental vulnerability with a multidimensional approach to assessing vulnerabilities at granular subregional levels. The VI has highlighted the relative vulnerability of certain districts as compared with others. Acknowledging the diversity, varying needs and priorities of the different districts in this state, our study provides information that can help detailed planning at subnational levels. Vulnerability assessment such as the CVIs proposed in this paper can provide an understanding of real data situations and can be used for pandemic planning at a subnational level for other states as well as in other LMICs.
Limitations

Data-related limitations were encountered when computing the VIs due to use of secondary data that are subject to constant variation. Several variables including mobility between districts could not be accessed during the analysis. Also, the most recent data from the national surveys are not yet available and might not have captured vulnerability well in certain districts where rapid changes may have occurred. Further, due to the lack of granular data, the epidemiological indices data are not updated to 2021 and their trends are likely to have changed. Additionally, the known issue of under-reporting of COVID-19 cases due to the limited availability of tests and the capacity of local surveillance services might have resulted in deficiencies in investigating deaths due to the disease, with possible under-reporting there as well. Multiple studies in the past couple of years have touched on the topic of under-reporting in context of COVID-19 cases and deaths worldwide.63 64 Specific to India, epidemiological models and serosurvey-based estimation procedures have indicated 5×–10× under-reporting for deaths and 30×–40× under-reporting for cases.65 66 Finally, the time-varying reproduction numbers and any summaries thereof such as the iR can be impacted by geographical diversity and environmental factors, as has been evidenced by studies both in context of prepanemic infectious diseases and COVID-19.67 68 70 Odisha, in particular, has districts that are located at varying distances from the Indian Ocean and experiences differing patterns of weather which may impact the pandemic growth differently. Therefore, a similar analysis performed on a bigger and possibly more representative sample obtained via large-scale population-level testing, while adjusting for mobility-related, geographical and environmental factors, may have yielded better explanations. Fortunately, our analysis pipeline offers a generalisable Bayesian framework to update the inferences with incoming data providing more granular, detailed and in-depth views of the true scenario, and can be adapted suitably to any other geographical region of interest and additional sets of covariates in the context of the disease.

Policy and public health implications

Overall, we can summarise the implications of our findings from two different directions, namely, the policy and public health perspectives. From a policy perspective, the factors or processes generating vulnerability and their measurement may differ in LMICs, and the CVI can help in capacity building and informing responses to outbreaks in the future, allowing policy makers to develop and implement their response at the district level. From a public health outlook, the findings of this study can enable national authorities and partners including academia, international organisations and donors to better align health emergency planning with broader population health needs and consider strengthening health system components for delivery of both emergency and non-emergency health services in tandem.

Author affiliations

1Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
2Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA
3Hamad Medical Corporation, Doha, Qatar
4Operations Management and Quantitative Techniques Area, Indian Institute of Management Indore, Indore, Madhya Pradesh, India
5Department of Statistics, University of Michigan, Ann Arbor, Michigan, USA
6Department of Statistics, Columbia University, New York, New York, USA
7Indian Institute of Public Health, Public Health Foundation of India, Bhubaneswar, Odisha, India
8Public Health Foundation of India, New Delhi, Delhi, India

Twitter Rupam Bhattacharyya @rupam_mmb


Contributors RP and VB conceptualised the project. RP conceptualised the idea of the index, and led the writing of the Introduction, Discussion and Conclusion sections. VB conceptualised the analyses pipeline, with RB leading the analyses and writing the results along with AB, SB, SM and AA. SL helped in writing the Discussion section. KS helped in accessing and cleaning the data used for constituting the index. SKR reviewed the manuscript and provided important inputs on all sections of the paper. VB and RP act as guarantor for the statistical analyses and data respectively.

Funding VB was supported partially by National Institutes of Health grants R01-CA160736 and NSF grant 1463233 and start-up funds from the U-M Rogel Cancer Centre and School of Public Health. SB was supported partially by DST INSPIRE Faculty Grant 04/2015/002165, and IIM Indore Young Faculty Research Chair Award Grant.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein) or of any geographical or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. All data used in the study, along with the codes used to perform the computations and build the figures and tables, are available in the online repository located at https://github.com/bayesrx/COVID_vulnerability_India.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

Open access

Copyright and re-use

This is a BMJ Publishing Group Limited (BMJ) commercial service available in association with the journal. This tool has been made possible by generous funding from the BMJ Group. The BMJ Group has no financial connection with the developer of this tool. No benefit has been directly received by the developer of the tool or any individuals associated with its development. Please see the limitations associated with the tool and contact the BMJ Group for more information.
REFERENCES


69 Charnley GE, Yennan S, Ochu C. Investigating the impact of social and environmental extremes on cholera time varying reproduction number in Nigeria. medRxiv 2022.
70 Kong JD, Tekwa EW, Gignoux-Wolfsohn SA. Social, economic, and environmental factors influencing the basic reproduction number of COVID-19 across countries. PLoS One 2021;16:e0253273.