Modelling COVID-19 With More Disaggregation and Less Nomothetic Parameterisation: UK and India Examples

Wendy Olsen (wendyolseninmanchester@gmail.com)  
University of Manchester  https://orcid.org/0000-0003-1440-5810

Manasi Bera  
Indian Institute of Dalit Studies

Jihye Kim  
University of Manchester

Arkadiusz Wisniowski  
University of Manchester  https://orcid.org/0000-0002-7567-3600

Purva Yadav  
Centre for the Study of Regional Development, School of Social Sciences, Jawaharlal Nehru University

Research Article

Keywords: Epidemic modelling, India, nomothetic parameter, latent variable, hierarchical model, severe COVID19, SARS-CoV2, Virus pandemic

DOI: https://doi.org/10.21203/rs.3.rs-71372/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Modelling of pandemic vulnerability in a development context can be improved through combining disciplines, combining data, and recognising the many nested levels of the epidemic. Models of transmission have been constructed at national level or for multiple nations. We instead construct a model allowing for social-group differentials in risk, along with conditioning regional factors and lifestyle factors. Severe COVID-19 disease is our innovative key outcome. We use three data sources at once: National Family and Health Survey for India, Indian Census 2011, and COVID-19 deaths. We provide results for 11 states of India, enabling best-yet targeting of policy actions. The future uses of such models are many. COVID-19 deaths in north and central India were higher in areas with older populations and overweight populations, and was more common among those with pre-existing health conditions, or who smoke or live in urban areas. Policy experts may both want to ‘follow World Health Organisation advice’ and yet also use disaggregated and spatially-specific data to improve wellbeing outcomes during the pandemic.

Introduction

In India, state governments manage COVID evidence, including individual cases of infection and the COVID-19-related deaths recorded. Typically, many migrant workers have been living in urban areas separate from their household of origin. Migrants who moved back to a village were generally not tracked, although temporary migrant registration lists were kept because there were demands to manage the transportation of migrants on key routes. In India, all COVID-19 patients are required to attend at government-managed COVID-19 testing sites. Data summaries were provided at District level via third-party organisations not benchmarked by government (How India Lives, 2020). In the UK, by contrast, the government controlled or managed most testing and reporting, along with carefully recorded hospitalisation data, death records, and data release to the public. By mid-March 2020, the UK authorities revealed that many un-tested COVID-19 cases had led to deaths in non-hospital homes, notably in care homes and personal homes, not recorded in the standard ‘COVID-19’ death figures (Dunn et al., 2020). A parastatal (The Health Foundation, see ibid) tracks policy over time. Data thus forms a key part of the machinations of managing a pandemic in both countries. In this paper, we suggest how a mixture of administrative data, random-sample microdata, and COVID-19 deaths data can be analysed. Data for India is modelled based in part on pre-existing data with 2020 updates, using an information-optimising Bayesian estimation routine. Such data combining models are rare so far in the COVID-19 literature. Our simple initial model can open up a wide range of variant models.

The paper begins with a review of some modelling options, including epidemic models and social aspects of health and social care in epidemics. Next the data and methods are introduced for the 11-state analysis of India using June 17, 2020 data. The research question for this section is what social factors contribute to the higher rates of COVID-19 deaths, both directly at the time of this disease and also through underlying long-term processes of exposure to health risks and lifestyle factors that created vulnerability. We develop a vulnerability index linked to district death counts. Next, we discuss implications both at the India level and cross-nationally. The paper closes with reflections on the broad purpose of development research. The aim in this particular case is to reduce deaths and suffering while recognising that policy makers are balancing a wide range of economic and political factors and not just health aspects of policy during the epidemic.

The epidemic progressed differently in the two countries. In the UK both cases and deaths rose exponentially then tailed off during and after a 10-week lockdown (March 24–approx. July 1, 2020). Cumulative deaths were reported at 39,815 by UK government on July 14 (Figure 1; PHE, 2020a; this figure omits the care-home deaths as do most ‘official’ data, even though many care-home deaths are specifically acknowledged to be COVID19-related). Total deaths officially recorded from COVID-19 in England rose from 55 to 65 per 100 K population (8/5/2020-2/6/2020; PHE 2020b). In India, case counts rose from 137 K to 307 K, and deaths from 4,800 to 9,900 (20/5/2020-17/6/2020, see How India Lives, 2020). India’s deaths then grew to 23,700 by 14/7/2020 (time of writing), creating a national cumulative death rate of 1.7 per 100 K population[1]. Compared with the UK lockdown, the Indian lockdown was more draconian running from 25 March – 30 June 2020 (Gov, 2020a). In India, persons over age 65 or below age 10, with comorbidities, or otherwise vulnerable were advised to stay at home (Gov, 2020b). COVID helpline numbers at national, state and district levels and a contract-tracing app were provided for information and emergency response (Gov, 2020c).

Models of Transmission and Severity of SARS-Cov-19

Three kinds of models emerged for the transmission of the virus, all utilising both the infections and deaths. They all utilised variables on ‘cases of COVID-19’ and ‘deaths from COVID-19’, suggesting an ultimate focus on human wellbeing. The ratio of the two variables is the case fatality ratio (CFR), ie deaths per person infected, e.g. Verty et al., 2020). The Lancet editorial of 24/1/2020 (Wang et al., 2020) noted a CFR of 2.9% in the Wuhan epidemic and that for the first 41 cases the rate was 15%. The CFR reported for the UK has been 14%, while India’s is 2.6% (Roser, 2020, 14/07/2020), but the underlying data may not be reliable. The exposure (cases) have not been precisely estimated (e.g. see Russell et al., 2020 for the Princess Diamond cruise ship exception with 100% testing). The CFR and cumulative cases per capita are parameters in models. We can also model the impact of COVID-19 based on models of the spread of the disease (Pellis et al., 2020).

Figure 2 illustrates the SEIR model of the spread of disease. The concepts of susceptibility – exposure - infection – recovery (SEIR) are fundamentals of epidemiology (Brauer et al., 2008). The infection rate in turn can be amended to allow for the reduction of contagion via distancing, hand-washing, lockdown and other policy tools. Observers can use Hill’s online app (2020) to develop their own forecasts. Two more online apps were produced by Pueyo (2020) and Neher (2020a, b). Using these forecasts, by mid-March 2020 it appeared likely that an Intensive Care Unit (ICU) bed shortage could occur in the UK. The UK government used similar models to underpin its decision to expand bed capacities in a number of hospital sites, including new field hospitals (Dunn et al., 2020). This model adapts a traditional epidemiological model to allow for social parameters.
The count of each compartment's people is based on a proportion of the previous compartment, depending on who has been exposed: children, adults, old people; and their body types, home types, past illnesses, and other factors. The basic reproductive number, R0, summarises how far each infected case spreads the virus to further cases (Wu et al., 2020). The model's summation of exposure rates allows modellers to disaggregate each overall parameter. Change over time is expressed in a series of differential equations: the marginal change in each compartment's count depends on the experience of the people in the previous compartment in the past time periods. Using disaggregated data, UK modellers have made it possible to examine how social groups, those living in high risk areas, and age-sex groups have different parameters (Pellis et al., 2020).

To complete the model for forecasting purposes, data were often absent, so estimates are made of key parameters. Data from the Diamond Princess cruise ship showed that children rarely tested positive for COVID-19 (Russell et al. 2002). The age group breakdown of positive tests was biased toward the oldest age groups. There was little sex bias.

The SEIR model has had a lot of close attention, leading to explorations of the time trend elements and how to model them. In epidemiology, the sequence of infections is seen as a series of generations of the virus (Brauer et al. 2008). The models used vary in terms of how they treat the latent period, which is usually thought of as the incubation period without symptoms. The panel data on the virus would be handled differently if the incubation period does or does not involve virus shedding (i.e. virus is growing and is emitted as airborne or urine-borne particles; see Alimentarius, 2012; Amri, 2020; WHO 2020a and b; WHO 2008). Currently, experts disagree on whether the latent period is asymptomatic but involves viral shedding; or whether those people who are asymptomatic while bearing SARS-CoV2 virus are not contagious. Overall, Figure 3 illustrates one likely sequence (see also Brauer et al., 2008: 121-5, 246-250, for alternatives).

Hellewell et al. (2020) carried out a SEIR treatment of transmission scenarios for SARS-COV2 virus using a branching mathematical model to allow for the many people, ranging from 0 to 20 or more, who were infected by each particular COVID-19 patient. This model treats latent and incubation periods carefully. As a purely mathematical model, without empirical microdata, the model tended to validate its own hypotheses. In other words, models based only upon algebra tend to be logically circular. Hellewell et al. concluded that isolation and contact-tracing would not succeed in limiting the virus spread, if the R number was too high. Their R0 assumption was from R0=1.5 to 3.5, based on Wuhan, China, and did not allow for the effects of lockdown. Mathematical models can be black-box style with facts-in-facts-out. Then if one inputs false claims (which we call facts) then you get facts out (Olsen and Morgan, 2005). Nevertheless, models help with thinking through interrelationships, exposing correlations, and creating an openness to new parameters that policy may focus upon.

Applications of the concept of branching processes in India led to concern that living in joint families would lead to high rates of transmission (Singh and Adhikari, 2020). Hill et al. (2010) also showed decisively that social networks matter very much to the way that a virus is passed on. Pujari and Shekatkar (2020) analysed social network patterns in India. Verity et al. (2020:3) said that the cruise ship 'Diamond Princess’ data gave the best insight to the biological processes of exposure and infection. There the 100% testing enabled asymptomatic cases to be included in the case rate data. The nomothetic assumption here (which means assuming all persons are alike, no matter which country) is that biological data from anywhere is valid everywhere. In Verity’s paper, transmission is the overarching term for exposure and infection. Attack rates are the rate of spread of virus from person to person – without infection necessarily occurring. The latent period is a time when the person holds virus particles on or in their body whilst not showing symptoms of the COVID-19 disease. The latent period length became a key variable. This parameter is in turn potentially differentiated by social group, body type, genetic features and immune response. The conceptual framing can be opened up to local, social and lifestyle differentiation – which could obviate the usefulness of the nomothetic assumptions.

The models also show that there can be interactions of the various underlying structural features. In Hill’s model (simplified in Figure 2), the treatment of the mild, medium, and severe infection stages all interacted to influence the rate of recovery from a detectable case of COVID-19. This abstract point holds true for all the countries in Hill’s modelling website. Treatment, hospitalisation, ventilation, intensive care unit stays, and medicine could all increase the recovery rate. Pellis et al. (2020) illustrate the inclusion of a longitudinal model of transmission with an SEIR model of recovery for the UK. A concept of biologically exponential spread is appropriate at an early stage of the virus, but data for another country cannot well inform the forecasting of the effects of distancing or lockdown. Pellis et al. (2020) covered 15 countries but did not discuss the obvious possibility that countries not included in this study could also have different medical, social-health-behavioural or genetic structures.

The normative overlay on most of the above models is that recovery is good, while the rest of the underlying mechanisms are complexly interpreted. One may want to get the disease and fight it off, so that one might become immune. This complex normative possibility raised much discussion: would it be better to expose >50% of the population so that they could become immune? Or would a country be better advised to use non-pharmaceutical interventions, such as lockdown and distancing, until a vaccine could be found – the decision rests in part upon the disease symptoms and the severe impacts? The costs of each case are relevant. Decisions rest in part on the future vision of the possibility of finding a vaccine and being able to produce it (or buy it). Neoliberals tended to imagine it would be easy to eventually produce or buy a vaccine. Most analysts try to reduce the infection rates in order to reduce suffering, improve wellbeing, and decrease the potentially lost lives arising from vulnerable people getting the disease (the neoliberal angle on these ethical issues is discussed further in Olsen, 2020; see also the contrasts in Chang and Grabel, 2014). Thus, the undertone of the expanded SEIR model is that the norms are complex and an informed, and an evidence-based discussion is needed (Olsen, 2007).

Dropkin (2020) also incorporated a SEIR model when modelling the exponential growth stage of the UK epidemic. His forecasts were, in the end, based mainly on scalar estimates for key parameters, not breaking the population into groups. Dropkin used a Bayesian method to fit the data to the exponential curve and differential equations, so that the final estimates are derived, not imposed. This can be thought of as a non-nomothetic approach (i.e. not lawlike; see Olsen, 2019).

An SEIR variant model by Tomas Pueyo (2020) became popular. Pueyo stressed that lockdown lowers the daily rise in cases, but then the epidemic will resurgence (ibid). Pueyo's focused upon cycles of rising and declining case rates. Hill and Dropkin were focused on the exponential growth stage (Dropkin, 2020). A
key problem is that exponential growth is the pattern to which the virus reverts if mitigation measures are relaxed. Yadav and Bhattacharjee (2020) showed India's transport sector slowed down considerably during its March-April lockdown.

Aswi et al. (2020) showed that the correlated geographically-contiguous spread of a different disease, Dengue Fever, meant that a spatial autoregression model would help in forecasting a virus spread pattern. Aswi et al. (2020) showed how a Bayesian Markov Chain Monte Carlo estimation method could make large models tractable without requiring a single maximum-likelihood function to encapsulate all the equations at once. Aswi et al. provides a sample program in the supplemental document. Rajendrakumar et al. (2020) illustrate these Bayesian methods.

An alternative modelling approach is to estimate an exponential trend model in the very short term. A particular example (Deasy et al. 2020) looked at mortality rates and hospitalisation rates per 'case' using data for UK regions. The data, unfortunately, were based on a very restricted COVID19 testing regime. Beds in hospital ICU in the UK were getting very full by the end of March or early April. Their upward-trend forecasts were based on a model that could not turn downward. This model with an obvious limitation in its algebra nevertheless usefully brought into the frame the regional and age-group differences in hospitalisation and ICU use rates. So models can support policy responses.

Ironically, such moves in a fast-changing scene show the value of modelling exercises which may have obvious weaknesses. We can attach positive normative value to the exercise whilst a critical reading and a sense of potential biasedness are also required. The case fatality rate is an example. CFR was estimated at 1.1% of cases for China by Russell et al. (2020: 2), and is widely perceived as having a biological, and hence universal, basis. Verity et al. (2020) saw the CFR fundamentally as a biological fact, calling the cruise ship data "robust" (ibid.: 1, here meaning gathered from administrative data, not from sample-data or biased datasets). They suggest CFR is useful for nomothetic assumptions. Yet these authors also argue in favour of a country-specific CFR due to the differing age-structure of populations. Our view is that the CFR responds to both biological and lifestyle factors and local circumstances. The CFR and SEIR model parameters reflect social, economic, and political, as well as biological causal mechanisms. All parameters may vary regionally and locally.

Usefully, there is a third model type that responds to this issue. It uses agent-based modelling to make a forward trajectory for the parameters corresponding to each social group. The model then blends the disaggregated findings of agent-simulation with the standard compartmentalised SEIR analysis over time (Ferguson et al., 2020, and Flaxman et al., 2020). Baseline characteristics from the northern Italy epidemic were used to set some scalar parameters (Grasselli et al., 2020). Factors in each person's background, such as diabetes, lung problems or tuberculosis, were ignored in this model (c.f. Liu et al., 2020). Its focus was infection rates, not severity.

Ferguson et al. (March 16, 2020) showed that their combination of the SEIR model with parameter changes over time invoking elements of shutting shops and factories, lockdown of schools and homes, and social distancing, including care-home restrictions and over-70s isolation for 12 weeks, would lead to an initial improvement followed by a second wave of viral infections. Ferguson's model was superior in noting social types, social-networking patterns, and social groups (Ferguson et al., 2020: 4), and a wide range of information was brought into play to make the combined SEIR-Agent-Based model (summarised by Adam, April 2020). Adaptations were also made to allow for people being more or less contiguous at each point in time. Ferguson's ICL model is a contagion model. However, as published, it was weak in its coverage of the lockdown impact.

With so many parameters, any of the above models could be tweaked, leading to a lower or higher overall case fatality rate. Criticisms arose because there was a worry that subjective or political factors were entering into scientific modelling. Extrapolations for the UK based purely on Chinese data were especially worrying. For example, social networking patterns in the UK might be different, so the transmission rates might be different (see the comparison of rates for countries in the contact-tracing literature, Kretschmar et al., 2020; and Singh and Adhikari, 2020, covering India, China and Italy). Extrapolations in India at an early stage were inconclusive (Gupta and Pal, 2020). Another study of transmissibility by Imai et al. 2020, used algebraic methods. By contrast, the model in the present paper has no social network information, is focused on severity, and uses probabilistic cross-sectional methods.

Discussion of the Types of Models of SARS-COV2 and COVID-19 Transmission

Models help to organise the thoughts of scientists and experts who are crossing disciplinary boundaries. They expose and make measurable key elements in social and socio-biological processes. Each theory is general, in that it can only handle a certain limited amount of detail. Some efforts to make models super-complex have succeeded in the sense of being socially and politically useful. Other efforts to achieve complexity have failed due to ignoring the existing literature on complexity. The specific additions to knowledge of complexity theory include three: first that each process can be broken down into a series of stages, and progression is path-dependent. Second, that one can pay attention to different types of units at one time point, which can make a huge difference to a wide set of trajectories; this is a disaggregated, multilevel form of path dependence. As a result, qualitative factors matter very much, and trajectory-switching can be tuned to respond to recorded multinomial and modal measures. Latent measures can be used, if there is evidence to enable estimation. Figure 4 illustrates the concept of mediated effects upon COVID-19 disease severity. The mediation occurs when the impact of a variable (like smoking) is both direct and indirect. Moderated effects occur if a local condition like pollution increases the pre-existing diseases and make people more susceptible.

The variable highlighted in red is a possible 'severe COVID-19' status, the risk of which can be reflected in a latent variable. Death counts are used here as evidence about the incidence of severe COVID-19 at each point in time. Models with 2 or more dependent variables are broadly classed as structural equation models.

In development theory and practice, we approach interpreting such a model using complexity theory. It is usual in complexity theory to be aware of nested systems, multiple levels, and the upward and downward effects of each 'layer' or type of entity. Eg. human bodies, hospitals, state policies, and international travel quarantines all affect each other.

Thirdly, complexity theory tells us that there is a difference between 'unobserved' and 'unobservable' features of the world. We can have concepts and algebraic representations of 'unobservable' features whilst making models correspond to observed factors at the same time. The butterfly effect and path
Models of SARS-CoV2 and COVID-19 impact have been a way of engaging in dialogue with experts who, after many past epidemics, were determined to avoid letting the exponential growth process, which has predictable elements, repeat itself. Many experts want to reduce human suffering. In human capabilities theory, we stress the improvement of human lives with a global scope as well as locally situated standpoints for knowledge (Olsen, 2019; Nussbaum, 1999). In capabilities theory, respect for all agents means a role can be played by translating models from scientific expert to lay language. There is also a role of asking for new, better data. Capabilities theory has been adopted by the Sustainable Development Goals campaign, by UK’s Department for International Development (DFID), and by other key actors in world development[1]. Capabilities theory has supplanted the older Gross Domestic Product focus and created a sense of awareness of diverse voices in development. Capabilities theory is consistent with the remit of development journals as it applies transdisciplinarity. Epidemiology and its sub-disciplines of medicine which include managing critical care, public health, and general practitioner services, are all aimed at reducing suffering and improving capabilities. They share the aim of promoting health from the start of life.

The best-yet knowledge rooted in modelling exercises can be presented in tabular, mapping, and graphical formats. We give illustrations from the Indian case. In the next sections, we attempt to disaggregate one impact of COVID-19 for a large part of India using secondary sources.

[1] URL https://www.gov.uk/government/publications/implementing-the-sustainable-development-goals/implementing-the-sustainable-development-goals--2, accessed July 2020; Steiner, 2018.

Data

The model links the district-level number of COVID-19 deaths in India to the individuals’ physical, demographic, and socio-economic status and other risk factors at the district level. The primary dataset of this study is the National Family Health Survey (NFHS)-4 2015/16 that includes individual-level variables such as family wealth, adult health, child health, and reproductive health variables. The NFHS has been fielded since 1992/93 in India, and we use the one of 2015-2016. It has a large sample size: the NFHS-4 surveyed women aged 15–49 (N=699,686) and men aged 15–54 (N=122,122), and household members (household N=601,059, household members’ N=2,869,043). The NFHS-4 provides a couples’ dataset, in which the total number of individual cases is 127,300 for all states. Among them, we choose 11 states spread across the Northern, Central, Western, and northeastern regions (Bihar, Chhattisgarh, Delhi, Gujarat, Haryana, Jharkhand, Madhya Pradesh, Maharashtra, Punjab, Rajasthan, and Uttar Pradesh). The sample size is 70,980 individuals in 11 states. Among them, 1,143 cases lacking body mass index (BMI) information are excluded. Thus, the final N is 69,837 (males: 34,715, females: 35,122).

Also, we use the 2011 Census data to obtain the proportion aged over 65 years in each district. As we regress the current death data as of June 17, 2020, the 2011 Census data on the age distribution and population size should be adjusted for 2020. To project the 2020 population, we calculate the monthly growth rate of the population in each district between the last censuses of 2001 and 2011, applying that to reach a 2020 population. For the new districts formed 2001–2011, the growth rate is calculated by combining the population of both districts.

Individual-level variables include sex, urban residency, scheduled caste (SC or Dalit), or scheduled tribes (ST or Adivasi), smoking, health indicators, obesity, underweight, and the lowest category of wealth index. Urban, SC, and ST are estimated at an individual level, as they show the different social status of individuals. Urban population is almost 30% of the total population in the 11 States. Dalit and Adivasi form 19% and 14% of the total population, respectively. We include several health-related variables, which do not have a correlation problem between them (see Table 1.) Smoking, a binary variable, indicates whether people smoke any of tobacco, cigarettes, pipe, chewing tobacco or snuff, or not. Obesity indicates whether the body mass index (BMI) is 30 or more, or not. If people’s BMI is under 18, they are underweight. We created an index of pre-existing health conditions using confirmatory factor analysis. Manifest variables were five key diseases – blood pressure, diabetes, asthma, heart disease, and cancer. Next quintile groups were assigned categorical indicators 0 to 4 on this scale, higher scores meaning more chronic disease. Similarly, the wealth index provided in the NFHS 2015/16 is calculated by a principal component analysis, basing the score on numerous consumer goods including vehicles along with housing characteristics (drinking water, toilet and flooring; IIPS and ICF, 2017:16). Our dummy variable indicates people belonging to the lowest wealth quintile.

District-level variables include obesity, underweight, the proportion of the aged population (ages 65+), and the percentage of migrants by place of the last residency. Here, obesity and underweight are based on district-level and are obtained from the NFHS household member’s dataset, which has a large samples size. The variable obesity means the proportion of people who have a BMI 30 or more; underweight is the proportion of people who have a BMI of less than 18.

We make separate models using individual obesity and underweight, and compare the results with using those at the district level. The proportion of the population aged 65 or above is obtained from the Indian Census, in consideration of population growth between 2011 and 2020. Also, Census 2011 provides a proportion of migrants per district. In Indian Census 2011, the person was asked if they are a migrant living in the destination. The period is 0–9 years recall from 2011. Migration = 1 refers to a person who is a migrant within or between states. They originate from another place. Migration is a binary whose % at District level is recorded in census tables. We do not use the rural and urban differentiated Migrant data. Table 1 shows the correlation of key variables (deaths and cases per 1000), showing the independent variables are weakly correlated.

Table 1. Correlation Matrix
| Variable         | Min. | Mean | Max. | N  | Description                                                                 |
|------------------|------|------|------|----|------------------------------------------------------------------------------|
| Deaths           | 0    | 24.22| 3,166| 345| Number of deaths due to COVID-19 in the district (median=1)                   |
| Female           | 0    | 0.50 | 1    | 69,837 | male=0, female=1                                                               |
| Urban            | 0    | 0.29 | 1    | 69,837 | person’s residence is rural=0, urban=1                                         |
| SC (Dalits)      | 0    | 0.19 | 1    | 69,837 | other groups=0, scheduled caste (SC or Dalit) =1                               |
| ST (Advasi people)| 0    | 0.14 | 1    | 69,837 | other groups=0, scheduled tribes (ST or Advasi)=1                              |
| Smoke            | 0    | 0.35 | 1    | 69,837 | Non-smoker=0, smoker=1                                                         |
| Ill-Health       | 0    | 0.42 | 4    | 69,837 | confirmatory factor index of blood pressure, diabetes, asthma, heart disease, cancer (0-4) |
| Low Assets       | 0    | 0.23 | 1    | 69,837 | The lowest quintile of wealth index category=1, others=0                       |
| Obesity          | 0    | 0.04 | 1    | 69,837 | BMI>=30 at individual level=1, others=0                                        |
| Underweight      | 0    | 0.14 | 1    | 69,837 | BMI<18 at individual level=1, others=0                                         |
| Over-Age-65      | 0.02 | 0.05 | 0.12 | 345 | the proportion of population of ages 65+ living in the district               |
| Migration        | 0.02 | 0.09 | 0.32 | 345 | the proportion of in-migrants (0-9 years) living in the district              |
| Obesity%         | 0    | 0.04 | 0.14 | 345 | the proportion of people living in the district with BMI≥30                    |
| Underweight%     | 0.05 | 0.20 | 0.37 | 345 | the proportion of people living in the district with BMI<18                     |

Notes: cases_pop/1000= COVID-19 cases per 1000 population, deaths_pop/1000= COVID-19 related deaths per 1000 population.

**Table 2. Summary of the Variables in the Regression Model**

Sources: 1) NFHS 2015/16 couples’ individual data; 2) deaths- howindialives.com (accessed on the 17th of June); 3) over-age-65 and migration – Indian Census 2011, summaries at District level.

**Methods**

The models were constructed using confirmatory methods, i.e. as well-theorised attempts to mimic the real situation. We planned to combine actual measurements of COVID-19-related counts of deaths (by District within State) and individual COVID-19 cases with the latest NFHS data. We planned that by including age-group data and a rural vs urban indicator at the individual level, we would be adjusting for the age structure of the more and less urbanised districts. Further hypotheses based on the medical and epidemiological literature, reviewed in earlier sections, would inform the analysis of the risk of death from COVID-19, both for spatial areas and for specific social groups. We theorised that assets (reflecting both wealth and the overall size of the household, with joint household pooling larger assets like vehicles) would act as a control variable. The asset index would control for wealth, which in turn is associated with getting good health care for each infected patient. At the same time, a low level of assets would act as a proxy for possible deficiencies in nutrients and economic resilience.
We also have predictive objectives. We aimed to predict the death rate for each district given change over time in the rate of infection, age-standardised adjusted vulnerability to severe COVID-19, and multilevel analysis of migrant influx. We did not conduct a time-series prediction but rather a cross-sectional, adjusted-for-confounders prediction. Future applications of the data combining methods can create SEIR time-series predictions combined with various measures of social networks, normal social behaviour during the latent period, and aspects of transport and schooling which might affect both contagion and the severity of disease. (For example, the migrants returning to village homes may have more severe cases of COVID-19.)

We therefore fit a model to three data sources (NFHS, Census data at District level, and COVID-19 death data). Maps 1-2 (Figure 6) illustrate the locations modelled, showing both the observed death rate and the predicted rates in districts of 11 states. However, there is no methodological reason not to create an all-India or even multi-country model. In the 11 states analysed in detail here, the COVID-19 case count was cumulatively 205 K (23 per 100 k population), with 8358 deaths by June 17, i.e. 0.6 per 100k population. The population in the 11 states was 888.3 m. The model with 11 states covered a wide range of both cities, mega-cities and villages. We omitted the Southern and North-eastern regions, where social norms, the social structure and cultural norms are rather distinctive. By doing so, a more nuanced set of results can be presented here.

A key reason for studying only part of India, not the whole, is our desire to publish maps; state and country boundaries being sometimes contested, it seems most suitable to avoid certain boundaries. We examined the whole of the Gangetic plain and several surrounding states that have large cities containing the SARS-COV2 virus entry points.

The specific datasets comprise two matrices. One contains a linked set of district data, one row per district, with NFHS and Census data concatenated. The second is NFHS data at the individual level, with a District variable. In the NFHS data, we have one row per individual for women and a second row for their male spouses.

We use a Poisson regression with the outcome the COVID-19-related deaths in each of 345 districts, for 11 states. The model aims to reveal potential risk factors related to COVID-19 mortality of individuals among women aged 15 to 49 and men aged 15 to 54.

To estimate model parameters and produce forecasts of mortality per district, we use Bayesian inference with weakly informative prior distributions. Bayesian inference permits estimating of complex hierarchical models. We use Hamiltonian Monte Carlo implemented in Rstan software, R interface to the program STAN for model analysis (Carpenter et al. 2017). Because of the large sample size, we used a generated total of 800 iterations with burn-in sample of 300 iterations. The code for analysing the data is available at [github location anonymised].

There are two levels. Let $i = 1,\ldots,N$ denote individuals and $j = 1,\ldots,G$, districts. Deaths in district $j$ follow a Poisson distribution

$$D_j \sim \text{Poisson}(\mu_j \cdot P_j) \quad \text{Eq. 1}$$

where $P_j$ is an offset in form of a population in district $j$ (in thousands), and $\mu_j$ is a death rate due to COVID-19 is calculated as the average of the latent death risk variable (or vulnerability) over individuals $N_j$ in district $j$:

$$\mu_j = \frac{\sum_{i=1}^{N_j} \mu_{ij}}{N_j} \quad \text{Eq. 2}$$

We model the individual latent death risk with a two-level model:

$$\log \mu_{ij} = \alpha_0 + \sum_{k=1}^{K} \alpha_k X_{ijk} + \sum_{l=1}^{L} \beta_l Z_{jl} \quad \text{Eq. 3}$$

where $X_{ijk} = \{\text{female, urban, scheduled caste, scheduled tribes, smoke, ill-health, low-assets}\}$ are individual-level predictors from the NFHS, and $Z_{jl} = \{\text{over-age-65, migration, obesity, underweight}\}$ are district-level predictors. The proportion over age 65 is taken from India’s 2011 Census tables, along with migration (see previous section). The proportion who are obese or underweight are derived as the district mean of individual cases from married couples in NFHS, as described earlier, and are Poisson regression coefficients.

**Results**

The variables that cause potential risks of COVID-19 death are shown in Table 3. At the individual level, living in urban areas is a substantial risk factor. Its mean coefficient across the MCMC samples is 2.88, with 95% Predictive Interval (PI 2.69-3.11), based on Model 1. Figure 5(a) confirms that there is a strong correlation between predicted risks of death and the proportion of the urban populations. Up to June 2020, India’s COVID-19 deaths have been concentrated in large urban areas (Figure 6). Social group differentials in the COVID-19 fatality rate are found after all other factors were controlled for; slope $\beta$ = 3.8 for SC (Dalit) people and $\beta$ = 1.79 for ST people, i.e., the effect is more apparent among Scheduled Castes than Schedules Tribes compared to all other groups. The heightened risk of severe COVID-19 among marginalised social groups could be because of poor living conditions, sanitation, nutritional status and their long-term economic activities in some cases leading them to higher exposure to pollutants (Kim et al., 2020). Females have a negative coefficient, which could result from females’ lower levels of outdoor travel and low economic participation outside homes.

**Table 3. Regression result of the models with the number of COVID-19 deaths**
In Model 2 as a variant we include the variable for wealth, which does not disrupt the findings of Model 1. Asset poverty was not typically associated with the intensity of Bayesian estimation is that the computer sometimes requires 4 to 6 hours to estimate one model.

It is a limitation of the model that transport links are not measured here. Interaction effects would create scope for improving this model. The resource-intensity of Bayesian estimation is that the computer sometimes requires 4 to 6 hours to estimate one model.

In Model 2 as a variant we include the variable for wealth, which does not disrupt the findings of Model 1. Asset poverty was not typically associated with severe COVID-19, but instead is negatively correlated with urban residence, after all other controls.

| Variables | Model 1 | Model 2 |
|-----------|---------|---------|
|           | mean    | median  | sd | 2.50%  | 97.50%   | mean    | median  | sd | 2.50%  | 97.50%   |
| Constant  | -12.87  | -12.85  | 0.27 | -13.45 | -12.36   | -12.46  | -12.47  | 0.29 | -13.01 | -11.91   |
| **Individual-level variables** |         |         |     |        |          |         |         |     |        |          |
| Female    | -3.89   | -3.87   | 0.45 | -4.79  | -3.1     | -4.05   | -4.04   | 0.45 | -4.94  | -3.24    |
| Urban     | 2.88    | 2.88    | 0.11 | 2.69   | 3.11     | 2.11    | 2.11    | 0.09 | 1.95   | 2.28     |
| SC (Dalit)| 3.8     | 3.79    | 0.13 | 3.56   | 4.09     | 4.16    | 4.15    | 0.16 | 3.89   | 4.49     |
| ST (Adivasi)| 1.79  | 1.79    | 0.18 | 1.46   | 2.14     | 2.16    | 2.16    | 0.2  | 1.78   | 2.58     |
| Smoke     | 4.08    | 4.06    | 0.25 | 3.63   | 4.61     | 4.14    | 4.14    | 0.24 | 3.68   | 4.63     |
| Ill-health| -0.26   | -0.26   | 0.03 | -0.33  | -0.2     | -0.29   | -0.29   | 0.04 | -0.36  | -0.22    |
| LowAssets | -3.08   | -3.04   | 0.4  | -4     | -2.42    |
| **District-level variables** |         |         |     |        |          |         |         |     |        |          |
| OverAge65*| 28.37   | 28.39   | 0.62 | 27.15  | 29.52    | 26.77   | 26.76   | 0.46 | 25.88  | 27.69    |
| Migration*| 2.23    | 2.23    | 0.26 | 1.72   | 2.77     | 1.85    | 1.86    | 0.27 | 1.29   | 2.38     |
| Obesity%  | 26.77   | 26.75   | 0.48 | 25.82  | 27.67    | 26.68   | 26.69   | 0.64 | 25.41  | 27.93    |
| Underweight%| 9.49 | 9.5     | 0.36 | 8.8    | 10.21    | 9.71    | 9.72    | 0.34 | 9.07   | 10.37    |

Source: NFHS 2015/16; *data obtained from Census, 2011

Notes: bold letters indicate increased risk of death; units are the number of COVID-19 deaths per 1,000 population. The last four row variables are measured at district level, and the rest at individual level. Ill-health is index of co-morbidities; LowAssets indicates lowest asset index quintile; OverAge65 is the % (scaled 0 to 1) of population in each district who are over age 65. Obesity% is the percentage(scaled 0-1) of district population who are obese (BMI>=30); and underweight% is the percentage(scaled 0-1) of district population who are underweight (BMI<18).

Some demographic variables also show a close association with COVID-19-related mortality. The older age group variables contribute to increased vulnerability. The mean of the population over age 65 is just 5% across 11 states, and the coefficient of the variable, OverAge65 is large (mean =28.37, 95% Pl=27.15–29.52). According to the PLFS 2018-9 data, a larger proportion of elderly (65+) people lives in rural areas (67% in rural vs 33 % in urban areas), which is because of the large rural population in India. However, the relative proportion of older people is higher in urban areas. Furthermore, the percentage of migrants in the district is also positively related to the potential risks of death (β= 2.23, 95% PI =1.72 – 2.77). Districts with a high percentage of in-migrants like Mumbai, Ahmedabad and Delhi have more jobs and movement of people, and accordingly, their risk level increases. Particularly vulnerable are those living in slums and temporary shelters without adequate living space to maintain physical distance and livelihood security.

The regression includes many health-related variables, which are smoke, the ill-health index, obesity and underweight. Smoke is one of the most influential health indicators that represent a high risk regarding COVID-19 mortality. Individuals who smoke are vulnerable to severe COVID-19 (β=4.08, 95% PI=3.63–4.61). COVID-19 is a respiratory virus, which could significantly affect people who have a weak lung function. Also, the obesity ratio in a district helps explain the increased death risk due to COVID-19. People who have BMI equal to 30 or more have higher vulnerability to severe COVID-19. The result shows that the coefficient of the variable, obesity% is 26.77 (25.82–27.67). Obese people are at risk of many other diseases, and so COVID-19 infections might result in significant morbidity among them. Figure 5(b) shows that there is a slight positive correlation between the obesity rate of each district and the predicted number of deaths. This result confirms that lifestyle and health conditions can bring a strong impact on the outcome due to COVID-19.

The underweight% (BMI <18) of the district is found on the increased death rate. Although the covariance table (see Appendix Table A.5) shows that it could have weak negative covariance with death rate, by controlling other variables, it turns out that underweight can be positively related to high risks of severe COVID-19. The result of the model shows it has a positive coefficient value (Mean=9.49, 95% PI=8.8–10.21). People below the normative threshold of BMI=18 are widely at risk of vitamin deficiencies or anaemia, and thus at risk of severe COVID-19. Note that in Models 1 and 2, the variables, obesity% and underweight% are measured at district level. Obesity and underweight at the individual level are tested in Models 3 and 4 (see Appendix Table A.3).

The results show that individuals’ condition of obesity or underweight were so far not a strong risk factor in severe COVID-19 cases compared to the other socio-economic and demographic variables, notably urban and smoking. The impact of obesity or being underweight had a small effect size. The ill-health index is negatively associated with the COVID-19 deaths, after controls, contradicting the correlation which was positive between the ill-health index and the death rate of each district. This implies that other factors have a stronger positive association, and they in turn are mediating a negative association of deaths with prior health conditions.

It is a limitation of the model that transport links are not measured here. Interaction effects would create scope for improving this model. The resource-intensity of Bayesian estimation is that the computer sometimes requires 4 to 6 hours to estimate one model.

In Model 2 as a variant we include the variable for wealth, which does not disrupt the findings of Model 1. Asset poverty was not typically associated with severe COVID-19, but instead is negatively correlated with urban residence, after all other controls.
In Figure 6, maps show the predicted outcome of COVID-19 death vs the recorded number of deaths for 11 states as of 17 June. Higher risks of severe COVID-19 are found in Maharashtra, Gujarat and Delhi. They have high urban populations and more transportation links. The risks of COVID-19 related deaths are strongly related to the number of COVID-19 infected cases. In selected states, Mumbai has the highest risks of severe COVID-19. Its observed number of deaths reaches 110 per 100,000 population, yet the model predicts it as 66. Other districts with the highest predicted death rate within their respective states are South West of the National Capital Region, 8.96; Mumbai suburban in Maharashtra, 6.03; Ahmedabad in Gujarat, 5.26; Rupnagar in Punjab, 1.94; and Ghaziabad in Uttar Pradesh, 0.97.

Discussion

World Health Organisation (WHO) versus Disaggregated Policy Advice

The WHO policy strategy offers a global common-denominator of advice about improved sanitation and hygiene; distancing and isolating/shielding; and avoiding this particular virus. The WHO aim to offer global advice appears at first glance nomothetic, so we give critical comments on three aspects.

First, the smoking issue. The WHO (2020c) argues that smoking is associated with severe COVID-19 and thereby causes increases mortality among hospitalised COVID-19 patients. Our study also provides clear evidence of a positive relation, but there is a strong male bias in smoking in India. This male bias would be specific to India. This male bias might also need to be explored further as a country-specific, and perhaps social-group and social-class specific aspect. Passive smoking has also got to be considered as a health risk. Furthermore, pollution at the district or city level (Kim, et al., 2020) has not yet been put into the model. The pollution aspect was not part of the WHO advice, yet it turns out to be relevant for the severe COVID-19. Thus more research is needed around this topic. For development experts, gender-sensitivity is needed in the policy response to both smoking and pollution issues.

Secondly, the WHO (2020d) also argues urban settings are riskier because of a high density of population and intense transport links. Informal settlements and social marginalisation are related concerns because COVID-19 transmission rates may be high as a result. More support should be given to marginal groups such as in-migrant workers in urban settings. We conclude that separate treatment is needed for managing the risk of infection versus the risk of severe COVID-19 cases and death. While linked, the pathways of cause are distinct (Figure 3).

Thirdly, the issues around obesity and being overweight have had a little discussion in the India case. The WHO advice has not gone into the precursors of either one. These weight issues are distinct, but both are related to social stigma. There is a risk of desirability bias which can cause them to be ignored. Being overweight places long-term pressure on heart and circulation systems. Those who are overweight may be at risk of other conditions like anaemia or calcium deficiency. Yet there are also social-group differences to consider, so that normative conclusions are not easy to derive. Each government, whether Indian or state government, needs to address the weight and lifestyle issues specific to its residents.

At the policy level, to be prepared for potential risks of severe COVID-19, identifying the most vulnerable people, understanding their needs, hearing their authentic priorities, and providing adequate services through national and local collaborations are necessary. An focus on older-age groups seems less justified for India than looking at the diverse sources of risk themselves. Beyond anodyne WHO advice on sanitation, policy targeting may also aim partly to enable safe practices in households, firms, and farms. Dis-aggregating information could help reduce the case fatality ratio.

Conclusions

This paper innovates by demonstrating the feasibility of a method of combining data from random and non-random sources. We applied the method in the case of 11 states of India. The paper also reviewed the range of modelling options, which are being used in the UK and India to examine patterns of contagion and severity of COVID-19 cases.

When applying such methods, a good theoretical foundation offers a solid grounding for choosing variables and testing hypotheses. Development theories range from trickle-down neoliberalism to structuralist democratic socialism, but the theoretical foundation of data combining is at a meta-level. In this case our meta-theory involves making transdisciplinary assertions cutting across medical and health, economics, and socio-cultural disciplines. This is common in three specific development theories: the idea of development as both policy and practice; ideas about development as the evolution of wellbeing (capabilities theory); and human development theory. The models in this paper tended to suggest that GDP per capita and assets were not sufficient indices to predict severe COVID19, but that assets needed to be seen within a broader model of vulnerability. We will comment on each of the three development theories in turn.

First, the idea of development as policy and practice sets up the agents of development as both individuals and corporate bodies. Development is not specific to the global ‘south’ but rather is a common method of improving lives in many countries. This approach coheres well with the WHO, who recommend that we learn from the regions of the world that had the epidemic earlier in the pandemic. For improved public health we would look at improvements in health delivery, review the structure of health services, examine how the society created vulnerabilities to the pandemic, and help local policy makers to improve their area’s resilience to this epidemic. The models we create can be helpful to policy makers anywhere and they aim to inform evidence-based policy.

Second, the idea of development as wellbeing is usually presented as a long-term, multi-dimensional approach to human good lives, not just as conceptions but as lived experience. Here, development is inhibited by multidimensional poverty, and poverty enabled this virus to attack certain minority ethnic groups and slum dwellers, plus in India’s case the many in-migrants who do not live with extended families. The ‘wellbeing’ literature has a strong focus on measuring both subjective and objective aspects of wellbeing in synergy with each other. The new recession will cause further reductions in wellbeing. The capability approach usually breaks up the achievement of wellbeing into various domains: the health domain is one; but housing, jobs, sanitation, being politically active i.e. participating, and other domains are also important. This pandemic has showed that a failure of entitlements in a single domain can create barriers to all the others, and cause a rapid decline in wellbeing. In an epidemic even the well-off people cannot always buy good health. In India, our model showed that
being overweight, and obesity in particular, has been associated with more COVID-19 deaths. It may not be a direct causal link. What is more, worse future patterns in rural areas could be observed, based on the association of COVID-19 vulnerability with being underweight. Underweight people are more prevalent in the rural-dominated states such as UP and Madhya Pradesh, see Figure 6. The link between the wellbeing theory and our model is a close one.

Thirdly, human development theory has built upon the wellbeing argument to state that what constitutes development is the fuller emergence of actualised human beings. The main outcome focus in human development theory is multifaceted wellness, often proxied by the Human Development Index indicators (HDI). The HDI uses people’s achieved education, income per capita, and longevity to gauge the aggregate level of development for a group or a region. During the COVID-19 era all three of these are threatened directly. Human development is going downhill in India. Ameliorating policies are suggested from within all three tracks of the HDI index: getting children back to school or schooled from home; increasing earnings and creating opportunities for microenterprises; and directly improving health care and sanitation to raise longevity back upward. In India, several long-term trends of improvement in these key areas have been set back already by the disease. Our model showed that the areas with more over-age-65 people were indeed the ones with more deaths in the March-to-June 2020 stages. Furthermore, after controlling for that, age itself appeared irrelevant but a range of lifestyle and medical conditions increased or decreased a person's vulnerability to severe COVID-19.

These results might seem anodyne yet they are evidence-based, and the results vary considerably by place and by social group. The pandemic has shown the strengths of the three transdisciplinary approaches to development, and the weaknesses of economic theories centred on commercial behaviour only (GDP per capita, trade flows and foreign investment, and so on). Finally, the UK-India comparison had several aspects in common, suggesting that a theoretical approach that allows for great similarities of the human experience across the world may have value. It may not be right to consider the Indian and other post-colonialised countries experience as ‘special’. Nor does it seem obvious that the UK as an ex-imperial country had any special advantages over India; up to June 2020, UK’s death rate (per recorded case, and also per resident population) was far higher than India’s. It may be worth noting also that the development researcher with experience in multidisciplinary analysis has strengths that can improve health care outcomes all over the world.

The data-combining method can be applied in other developing-country contexts e.g. Pakistan, Sri Lanka etc. Variations on the method include an L-shaped option (one new random survey using a subset of the original national survey, but with additional variables); the Delphi method option; and generally well-informed priors as opposed to flat priors. We discuss each briefly. First, we can construct an L shaped dataset, using fresh random qualitative data on death counts, or using phone-survey data. Non-random data would be less useful; administrative data is highly useful. From this matrix, we improve estimates of the impact of the disease COVID-19. Second, we can use an ‘expert questionnaire’ known as the Delphi method to gather more information for the evidence base. Thirdly, we could also use well-informed priors based on other outside information. In the COVID-19 debate, much of such analysis has rested upon purely China data (e.g. based on purely Wuhan data from Feb-March 2020), or China plus Italy, or the Princess Diamond Cruise ship data. We argued that these outside sources must be used carefully, and not simply via the lawlike nomothetic assumption that each Indian district would have the same social structure or the same relationship of severe-COVID-19 to the age structure of exposure. It is possible to use external data to inform a first run of a model but we then need to try to locate relevant, preferably random-sample based data from inside the country to make further investigations. The best data will be disaggregated, and Census results for 2021 that enable online use, user-driven crosstabulation and perhaps even a record-level download of a random sample of cases would be potentially useful in future for India and other countries where administrative data sources are not typically released to the public.

Declarations

Acknowledgements: The research funder is Global Research Challenges Fund and project title Social-Action Messages to Reduce Transmission of COVID-19 in North India, May-July 2020. Principal Investigator Prof Wendy Olsen and co-investigators Prof A. Dubey, Dr. P. Yadav and Dr. A. Wiśnioswki. Thank you to Zoe Williams, Amaresh Dubey, and Clelia Cascella, who all contributed to the analysis in this paper through discussions.

Statement of No Competing Interests

For all the authors on this article, there are no relevant financial or non-financial competing interests to report. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Statement of Conforming to Ethical and

Data-Management Governance Principles

The project team conforms to the principles of the Committee on Publication Ethics, (COPE), URL https://publicationethics.org/. We broadly also follow the ethical and co-authoring guidelines of the British Sociological Association, URL https://www.britsoc.co.uk/media/24310/bsa_statement_of_ethical_practice.pdf.

On 29 June, 2020, the project received approval from the Indian Institute of Dalit Studies, Delhi (IIDS), confirming that it conforms to the ethical and governance guidelines of the IIDS.

On 23 July, 2020, the project received ethics approval from the University of Manchester under its project title, "Social-Action Messages to Reduce Transmission of COVID-19 in North India", Research Governance, Ethics and Integrity department, number 2020-10055-16152.

Data Availability Statement

The raw data analysed during the current study are available in the following four repositories. Indian Census data, see URL https://censusindia.gov.in/2011census/populationEnumeration.html; National Family and Health Survey data, register via Demographic and Health
Survey website URL: https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf, and see URL: http://rchiips.org/NFHS/nfhs4.shtml; data on Indian Covid19, URL: www.howindialives.com/gram/coronadistricts/ (update date: June 17, 2020).

References

Adam, David (2020, 2 April, corrected 3 April) Special report: The simulations driving the world’s response to COVID-19 How epidemiologists rushed to model the coronavirus pandemic. Nature, www.nature.com, URL: https://www.nature.com/articles/d41586-020-01003-6, accessed April 2020.

Alimentarius, C. (2012) CAC/GL 79-2012 guidelines on the application of general principles of food hygiene to the control of viruses in food. Codex Committee on Food Hygiene. 13.

Amirian, E.S. (2020) Potential Fecal Transmission of SARS-CoV-2: Current Evidence and Implications for Public Health. International Journal of Infectious Diseases.

Aswi, A., Cramb, S., Duncan, E., Hu, W., White, G., & Mengersen, K. (2020) Bayesian spatial survival models for hospitalisation of Dengue: A case study of Wahidin hospital in Makassar, Indonesia. International Journal of Environmental Research and Public Health, 17(3), 878.

Brauer, F., Driessche, P.D. and Wu, J. (eds) (2008) Mathematical Epidemiology, NY: Springer.

Chang, H.J., and Grabel, I. (2014) Reclaiming Development: An Alternative Economic Policy Manual, London: Zed Press.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017) Stan: A probabilistic programming language. Journal of Statistical Software 76(1). DOI 10.18637/jss.v076.i01

Dropkin, G., 2020. COVID-19 UK Lockdown Forecasts and R0. Frontiers of Public Health, 8. Earlier variant models were published as Dropkin, Greg (2020) http://www.labournet.net/other/2003/lockdown.pdf, and later version COVID-19 UK Lockdown Forecasts and R0, MedRxiv April 2020, DOI https://doi.org/10.1101/2020.04.07.20052340.

Dunn, P., Allen, L., Cameron, G., and Alderwick, H. (2020) The Health Foundation, COVID-19 policy tracker: A timeline of national policy and health system responses to COVID-19 in England, https://www.health.org.uk/news-and-comment/charts-and-infographics/covid-19-policy-tracker. Accessed June 2020.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z.U.L.M.A., Cuomo-Dannenburg, G. and Dighe, A. (2020) Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H., Coupland, H., Mellan, T., Zhu, H., Berah, T., Eaton, J., Perez Guzman, P. and Schmit, N. (2020) Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London.

Gupta, R. and Pal, S.K. (2020) Trend Analysis and Forecasting of COVID-19 outbreak in India. medRxiv. DOI: 10.1101/2020.03.26.20044511

Government of India (2020a) PM calls for complete lockdown of entire nation for 21 days, PIB Delhi, https://pib.gov.in/PressReleasePage.aspx?PRID=1608009, accessed July 2020.

Government of India (2020b) New Guidelines to fight COVID-19 to be effective from 1st June 2020, Ministry of Home Affairs, PIB Delhi https://pib.gov.in/PressReleasePage.aspx?PRID=1627965, accessed July 2020.

Government of India (2020c) Aarogya Setu is now open source, Ministry of Electronics & Information Technology, New Delhi, https://static.mygov.in/rest/s3fs-public/mygov_159050700051307401.pdf, accessed July 2020.

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A. (2020) Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. The Lancet. Global health, 8 (4). DOI: https://doi.org/10.1016/S2214-109X(20)30074-7. (28 Feb 2020)

Hill, Alison L., David G Rand, Martin A Nowak, Nicholas A Christakis (2010) Infectious Disease Modeling of Social Contagion in Networks, PLoS Computational Biology, 6:11, doi: 10.1371/journal.pcbi.1000968

Hill, Alison (2020) “Modelling COVID-19 Spread vs Healthcare Capacity” https://alhill.shinyapps.io/COVID19seir/

How India Lives (2020) India: Joe Data Services Pvt. Ltd. URL https://www.howindialives.com/gram/coronadistricts/ , accessed 17 June, 2020.

Imai, N., Cori, A., Dorigatti, I., Baguelin, M., Donnelly, C.A., Riley, S. and Ferguson, N.M., 2020. Report 3: transmissibility of 2019-nCoV. Imperial College London.

International Institute for Population Sciences (IIPS) and ICF (2017) National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS.

Kim, C., Henneman, L.R., Choirat, C. and Zigler, C.M. (2020) Health effects of power plant emissions through ambient air quality. Journal of the Royal Statistical Society: Series A (Statistics in Society).
Kretzschmar, M., Rozhnova, G. and van Boven, M. (2020) Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. Available at SSRN 3562458.

Liu, Y., Bi, L., Chen, Y., Wang, Y., Fleming, J., Yu, Y., Gu, Y., Liu, C., Fan, L., Wang, X. and Cheng, M. (2020) Active of latent tuberculosis increases susceptibility to COVID-19 and disease severity. medRxiv.

Neher, R.A. et al. (2020a) modelling website (URL https://covid19-scenarios.org/, accessed April 2020.

Neher, R.A., Robert D., Valentin D., Emma B. H. and Jan A. (2020b) Potential Impact of Seasonal Forcing on a SARS-CoV-2 Pandemic, 16 March, Swiss Med Wkly 150:w20224. doi:10.4414/swm.2020.20224

Nussbaum, M.C., 1999. Sex and Social Justice. Oxford University Press.

Olsen, W.K. (2007) “Pluralist Methodology for Development Economics: The Example of Moral Economy of Indian Labour Markets”, Journal of Economic Methodology, 14:1, 57-82..

Olsen, W.K. (2019) “Social Statistics Using Strategic Structuralism and Pluralism”, in Contemporary Philosophy and Social Science: An Interdisciplinary Dialogue, eds Michiru Nagatsu and Attilia Ruzzene. London: Bloomsbury Publishing.

Olsen, W.K. (2020) Models and Ethics in Handling the COVID-1919 Pandemic: Models Enable Strategic Reasoned Public Actions, Radical Statistics, Vol. 126: 75-90. URL https://www.radstats.org.uk/no126/Olsen126.pdf.

Olsen, W.K., and Morgan, J. (2005) A critical epistemology of analytical statistics: Addressing the sceptical realist, Journal for the Theory of Social Behaviour, 35:3, September, pages 255-284. (my role 50%) DOI: 10.1111/j.1468-5914.2005.00279.x http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5914.2005.00279.x/full50%

Pellis, L., Scarabel, F., Stage, H.B., Overton, C.E., Chappell, L.H., Lythgoe, K.A., Fearon, E., Bennett, E., Curran-Sebastian, J., Das, R. and Fyles, M., Lewkowicz, H., Pang, X., Vekaria, B., Webb L., House, T., Hall, I. (2020) Challenges in Control of Covid-19: Short doubling time and long delay to effect of interventions, arXiv. 2004.00117v1 [q-bio.PE] 31 Mar 2020 (version 1), and preprint at URL medRxiv preprint doi: https://doi.org/10.1101/2020.04.12.20059972 (version 2), 15 April 2020.

Public Health England, Government of UK (2020a) Weekly Coronavirus Disease 2019 (COVID-19) Surveillance Report: Summary of COVID-19 surveillance systems, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/le/899301/Weekly_COVID19_Surveillance_Report_week_28 accessed 14.7.2020.

Public Health England (2020b) Coronavirus (COVID-19) in the UK, Accessed 04/06/2020. Available from: https://coronavirus.data.gov.uk.

Pueyo, T (2020). Tomas Pueyo. Accessed 16/07/20. Available from: https://medium.com/@tompueyo

Pujari, B.S. and Shekatkar, S.M. (2020) Multi-city modelling of epidemics using spatial networks: Application to 2019-nCov (COVID-19) coronavirus in India. medRxiv. DOI: 10.1101/2020.03.13.20035386, 17 March 2020.

Rajendrakumar, Aravind Lathika, Anand Thakarakkattil Narayanan Nair, Charvi Nangia, Prabhal Kumar Chourasia, Mehul Kumar Chourasia, Syed Mohammed Ghouse, Anu Sasidharan Nair, Arun B Nair, and Sha Fazaludeen Koya (2020) Epidemic Landscape and Forecasting of SARS-CoV-2 in India. medRxiv doi.org/10.1101/2020.04.14.20065151

Roser, Max, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell (2020) - “Coronavirus Pandemic (COVID-19)”. Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/coronavirus, online resource.

Russell, Timothy W., Joel Hellewell, Christopher I Jarvis, Kevin Van Zandvoort, Sam Abbott, Ruwan Ratnayake, Stefan Flasche, Rosalind M Eggo, W John Edmunds, and Adam J Kucharski (2020) Estimating the Infection and Case Fatality Ratio for COVID-19 Using Age-Adjusted Data from the Outbreak on the Diamond Princess Cruise Ship, MedRxiv preprint doi: https://doi.org/10.1101/2020.03.05.20031773.

Singh, R. and Adhikari, R. (2020) Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv preprint arXiv:2003.12055.

Steiner, Achim (2018) Commentary: The Extraordinary Opportunity of the 2030 Agenda for Sustainable Development, European Journal of Development Research (2018) 30, 163–165 doi:10.1057/s41287-0180131-x.

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P, Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P.G., Fu, H. and Dighe, A., (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases.

Wang, C., Horby, P.W., Hayden, F.G. and Gao, G.F., (2020) A novel coronavirus outbreak of global health concern. The Lancet, 395(10223), pp.470-473.

World Health Organization (2020a) Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus. Interim guidance, 1 April 2020.
World Health Organization (2020b) *Water, sanitation, hygiene, and waste management for the COVID-19 virus: interim guidance, 23 April 2020.*

World Health Organization (2020c) Smoking and COVID-19, Scientific brief, 30 June 2020.

World Health Organization (2020d) Strengthening preparedness for COVID-19 in cities and other urban settings: interim guidance for local authorities. Geneva: World Health Organization (WHO/2019-nCoV/Urban_preparedness/2020.1).

World Health Organization (2020e) Community-based health care, including outreach and campaigns, in the context of the COVID-19 pandemic, WHO/2019-nCoV/Comm_health_care/2020.1.

World Health Organization (2008) Viruses in food: scientific advice to support risk management activities: meeting report.

Wu, J.T., Leung, K. and Leung, G.M. (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet, 395*(10225), pp.689-697.

Yadav, P., and A. Bhattacharjee (2020) The Impact of COVID-19 on Mobility in India: A spatial approach, Radical Statistics, Vol. 126: 75-90, URL https://www.radstats.org.uk/no126/Yadavetal126.pdf

**Figures**

**Figure 1**

Cumulative COVID-19 Cases Per 100K population, by local authority, England (box shows London area)

**Figure 2**

A SEIR Model with Compartments (after Wu et al., 2020) Sources: Time-series equations for the model are in Wu et al., 2020: 247-8; Hill (2020). We do not normally allow for a case to be re-infected, but this question is now being raised for SARS-COV2.
Figure 3
Model Illustrating A Disease Transmission Process Over Time Source: Brauer et al., (2008: 121-5, 246-250). This panel illustrates the Crump–Mode–Jagers process, in contrast to alternatives where a latent or asymptomatic period is more prominent (ibid.:247-8). Key: time moves left to right. Cases are generated as shown at the bottom. Index case is the entrant who starts off a group of cases in one area.

Figure 4
Model of COVID19 Contagion and Severity Notes: The multilevel model shown here can be estimated as a cross-section or over time. The model we operationalise is shown in the results section using a subset of the variables. Horizontal=time.

| Plot | Description |
|------|-------------|
| ![Plot](image1.png) | a. Urban % of district and predicted deaths per 100,000 population (a supplementary variable) |
| ![Plot](image2.png) | b. Obesity % of district and predicted deaths per 100,000 population |
| ![Plot](image3.png) | c. Underweight % of district and predicted deaths per 100,000 population |

Figure 5
Predicted number of deaths by the urban, obesity and underweight rate of districts Notes: prediction using Model 2; (a) percentage of district population who are urban residents (scaled 0-1); (b) percentage of district population who are obese (BMI>=30); (c) percentage of district population who are underweight (BMI<18); predicted means of 344 districts; Mumbai(pred=66) is excluded from the plots. @The urban % across a district is used in Panel (a) but individual urban residence was used in the actual regression model.
Maps of the observed and predicted number of deaths Notes: For greater coherence we modelled 11 states rather than all states, dark grey being parts not included; figures greater than 5 are rounded down to 5 here. Source: https://www.howindialives.com/gram/coronadistricts/ (update date: June 17, 2020) Map source: Global Map of India © ISCGM/Survey of India

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableA1.png
- TableA2.png
- TableA3.png
- TableA4Copy.png
- TablesA5and6Copy.png
- FigA1.png