COVID-19 inequalities in England: a mathematical modelling study of transmission risk and clinical vulnerability by socioeconomic status

Authors: Lucy Goodfellow[1*], Edwin van Leeuwen[1,2], Rosalind M Eggo[1]

Affiliations:

[1] Centre for Mathematical Modeling of Infectious Disease, London School of Hygiene and Tropical Medicine, London WC14 7HT, United Kingdom

[2] Modelling and Economics Unit and NIHR Health Protection Research Unit in Modelling and Health Economics, UK Health Security Agency, London NW9 5EQ, United Kingdom

* corresponding author. Email address lucy.goodfellow@lshtm.ac.uk

Abstract

Background: The COVID-19 pandemic resulted in major inequalities in infection burden between areas of varying socioeconomic deprivation in many countries, including England. Areas of higher deprivation tend to have a different population structure - generally younger - which can increase viral transmission due to higher contact rates in school-going children and working-age adults. Higher deprivation is also associated with higher presence of chronic comorbidities, which were convincingly demonstrated to be risk factors for severe COVID-19 disease. These two major factors need to be combined to better understand and quantify their relative importance in the observed COVID-19 inequalities.

Methods: We used UK Census data on health status and demography stratified by decile of the Index of Multiple Deprivation (IMD), which is a measure of socioeconomic deprivation. We calculated epidemiological impact using an age-stratified COVID-19 transmission model, which incorporated different contact patterns and clinical health profile by decile. To separate the contribution of each factor, we also considered a scenario where the clinical health profile of all deciles was at the level of the least deprived. We also considered the effectiveness of school closures in each decile.

Results: In the modelled epidemics in urban areas, the most deprived decile experienced 9% more infections, 13% more clinical cases, and a 97% larger peak clinical size than the least deprived; we found similar inequalities in rural areas. 21% of clinical cases and 16% of deaths in England observed under the model assumptions would not occur if all deciles experienced the clinical health profile of the least deprived decile. We found that more deaths were prevented in more affluent areas during school closures.

Conclusions: This study demonstrates that both clinical and demographic factors synergise to generate health inequalities in COVID-19, that improving the clinical health profile of populations would increase health equity, and that some interventions can increase health inequalities.

Keywords: COVID-19, Mathematical model, Socio-economic inequality, Comorbidities,

England, Demography

Background

The COVID-19 pandemic disproportionately affected people in lower socioeconomic groups around the world [1], [2]. In England, there were large disparities in COVID-19 burden between areas of different relative deprivation, measured by the Index of Multiple Deprivation (IMD). Initial reports by the UK Office for National Statistics (ONS) found that from April-July 2020 the most deprived 10% of areas in England experienced an age-standardised COVID-19-related mortality rate more than twice as high as the least deprived 10% [3]. These disparities were repeatedly observed through the pandemic: between June 2020 and January 2021, the age-standardised mortality rate in laboratory-confirmed cases of COVID-19 was 371.0 per 100,000 (95% Confidence Interval (CI): 334.2 - 410.7) compared to 118.0 (95% CI: 97.7-141.3) in the most vs least deprived quintiles [4]. The inequality in mortality rates seen in the early pandemic exceeded that observed in previous years, indicating that there were further factors exacerbating the 'expected' effects of relative deprivation [5]. Even after adjusting for age, sex, region, and ethnicity, this report found worse outcomes in more deprived areas, but did not adjust for prevalence of comorbidities. Other studies have consistently confirmed an association between comorbidities and more severe COVID-19 outcomes [6], [7].

Morbidity and the presence of underlying health conditions tend to vary greatly by socioeconomic status (SES), and are a significant risk factor for severe infection [8]. Vulnerability to more severe infection has both direct effects, including a greater risk of consequential long-term health complications and greater mortality risk, and indirect population-level effects, such as potentially increased infectiousness of symptomatic cases. In England, before the COVID-19 pandemic, life expectancy was 9.4 years longer for men in the least deprived decile than the most, and 7.7 years longer for women [9]. These gaps continue to widen: female life expectancy in the most deprived decile fell by 4 weeks between 2014-2016

and 2017-2019, but rose by 11 weeks in the least deprived [9], [10]. The prevalence of underlying health conditions that affect quality of life is also consistently correlated with local deprivation levels: men in the most deprived decile could expect to live 18.4 years fewer in good health than those in the least deprived decile; the corresponding gap for women was 19.8 years [9].

It is well established that infectious disease burden is associated with SES [11]–[13]. This is linked to a multitude of complex and interwoven factors including, but not limited to, lack of access to healthcare, poor housing conditions, inability to avoid high-exposure settings such as crowded public places, differences in occupation type, and avoiding restrictions or testing due to mistrust of authorities [14], [15].

Here, we use a novel transmission model to combine differences between socioeconomic groups in their risk of infection with their risk of severe disease on infection to quantify their relative importance in the observed COVID-19 inequalities in England. We consider underlying health conditions as a key determinant of an individual's risk of developing a clinical case of COVID-19, and focus on the impact of IMD-specific health and age structure on infectious disease burden at the population level. By making simplifying assumptions and modelling a synthetic population, we aim to produce a conceptual exploration of the interaction between underlying health and demographic structure.

Methods

We developed an age-stratified dynamic transmission model for SARS-CoV-2, which was further stratified by IMD decile, and by urban or rural classification in England. Here we detail how the model was modified to incorporate the characteristics of each decile and geography.

IMD-specific age structure

Each epidemic was simulated on the population of a given IMD decile in either an urban or rural area, to account for the distinct underlying age structures in these areas. We used 17 age groups (0-1, 1-5, every 5 years to 75, and over 75). The mid-2020 (30 June) age-specific population of each lower layer super output area (LSOA), which is on average 1,500 people, was linked via LSOA codes to their IMD decile and urban/rural classification (where urban is defined as a settlement with over 10,000 residents) [16]–[19]. We calculated the size of each age group, specific to each IMD decile and geography, and used this to determine the average age structure of each IMD- and geography-specific population, $n = (n_1, ..., n_{17})$, where

$$\sum_{a=1}^{17} n_a = 1$$

in each population. We also calculated the median age for each urban and rural IMD decile, and the proportion of each IMD decile residing in urban or rural LSOAs (Supplementary Section 1).

Contact matrices

To define contact between age groups we used age-specific social contact data for the United Kingdom (UK) for physical and conversational contacts accessed via the *socialmixr* R package [20], [21]. The contact matrices are highly age-assortative, with the highest daily contact patterns occurring between individuals in the same age group for those aged 5-19. We projected the contact patterns onto the age structure of each population in 2020, using the

density correction method, by constructing an intrinsic connectivity matrix and scaling this matrix

to match the population's age structure [22].

The intrinsic connectivity matrix was calculated from the 2006 UK contact matrix

$M^{2006} = (M_{ij}^{2006})_{i,j=1,...,17}$ and age structure $N^{2006} = (N_1^{2006}, ..., N_{17}^{2006})$ as

$$\Gamma = (\Gamma_{i,j})_{i,j=1,...,17}$$

$$\Gamma_{ij} = M_{ij}^{2006} \frac{\sum_{a=1}^{17} N_a^{2006}}{N_j^{2006}}$$

The new contact matrix for a population with age group sizes $N = (N_1, ..., N_{17})$ and

proportions $n = (n_1, ..., n_{17})$ had entries

$$M_{ij} = \frac{\Gamma_{ij} N_j}{\sum_{a=1}^{17} N_a} = \Gamma_{ij} n_j$$

Age-specific fraction of COVID-19 cases causing clinical symptoms

We separate infections of SARS-CoV-2 as in [23], into clinical or subclinical cases. Clinical

cases of COVID-19 are infections that lead to noticeable symptoms such that an individual may

seek clinical care. Subclinical infections do not seek care in the model, and are less infectious

than clinical cases. We defined a population's clinical fraction as the probability of an individual

in the population developing a clinical case of COVID-19 upon infection. Here, we extended

previous work by relating an individual's probability of being a clinical case of COVID-19 to the

self-reported health status of their IMD- and age-specific population in England, and then

7

examined how differences in self-reported health status by IMD decile, coupled with differences in age distribution, affect the burden in each IMD decile.

To define health status, we used data from the 2021 Census, specifically the question 'How is your health in general?', with response options of 'Very good', 'Good', 'Fair', 'Bad', and 'Very bad' [24]. This is provided by the Census stratified by IMD and by age. We then defined 'health prevalence' as the proportion of individuals reporting 'Very good' or 'Good' general health, stratified by the same age groups and the deciles of IMD:

$$Health\ prevalence \ = \frac{Number\ in\ 'Very\ good'\ health + Number\ in\ 'Good'\ health}{Number\ in\ all\ health\ statuses} \qquad (1)$$

To map the population health prevalence to the model's age-specific clinical fraction we used locally weighted least squares regression (LOESS) [23]. Any populations with health prevalences outside of the training dataset's range were assigned the most extreme clinical fractions found by Davies et al. [23], to avoid extrapolation outside of observed values. Health prevalence was highest in children, but children have separate risk factors for severe disease (such as smaller airways) and children under 10 have been found to be subject to a higher risk of clinical COVID-19 cases and a greater infection fatality ratio (IFR) [23], [25] (as observed for other infections such as influenza [26]). Therefore, we fixed the clinical fraction of the 0-9 age group at 0.29, matching that found by [23].

COVID-19 transmission model

The transmission model includes a single SARS-CoV-2 variant, no existing immunity in the population, and natural history parameters drawn from the first wave of the pandemic. We did not include vaccination, and the only non-pharmaceutical intervention (NPI) considered was

school closures. We developed an age- and IMD-stratified deterministic compartmental model in R (version 4.3.1) (Figure 1c). There is no mixing between IMD deciles in the model. The aim is to demonstrate the importance of health prevalence and differences in age and social mixing in epidemic impact, rather than to reproduce the COVID-19 epidemic in England.

Individuals are first assumed to be susceptible (*S*), and become exposed (*E*) but not yet infectious after effective contact with an infected individual (Figure 1c). Each exposed individual then progresses to one of two infected states: subclinical infection (*Is*), and clinical infection, which is represented by a pre-symptomatic (but infectious) compartment (*Ip*) followed by a symptomatic compartment (*Ic*). Each individual then moves into the recovered (*R*) or dead (*D*) compartment, at which point they are assumed to no longer be infectious and to be immune to infection. This Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) is an extension of [23], with the addition of a *D* compartment. We ran the epidemic for 365 days, which allowed completion of each epidemic in each decile and geography. Each epidemic was run on a synthetic population of a fixed IMD decile and urban/rural geography, with no births, non-infection-related deaths, or ageing between age groups, as the time-frame of each epidemic was less than a year. The model also assumed that contact patterns remain constant through the epidemic.

The force of infection in age group k is given by:

$$\lambda_k = p \sum_{a=1}^{17} M_{ak} \frac{Ip_k + Ic_k + \xi Is_k}{n_k} \, ,$$

where p is the probability of a contact between an infected and susceptible individual resulting in transmission of infection, M_{ak} is the mean daily number of contacts that an individual in age

9

group *a* has with individuals in age group k, and ξ is the relative infectiousness of subclinical

cases. The age-specific clinical fraction is denoted by π_a and depends on the IMD decile. Rates

of transition from each disease state are given in Table 1.

Parameter	Value	Definition	Source
p	0.06	Transmission probability	[23]
M_{ak}	Varies by age and IMD	Daily age-specific contacts	Based on [20]
ξ	0.5	Relative subclinical infectiousness	Assumption
π_a	Varies by age and IMD	Clinical fraction	Based on [23]
f	3 days	Duration of latent period	[23]
γ	2.1 days	Duration of preclinical infectious period	[23]
r_c	2.9 days	Duration of clinical infectious period	[23]
r_s	5 days	Duration of subclinical infectious period	[23]
μ_{ca}	Varies by age, see Supplementary Table 4	Clinical mortality probability	[27]
μ_{sa}	0 for all age groups	Subclinical mortality probability	Assumption

Table 1. Model parameters.

We assumed the relative subclinical infectiousness (ξ), to be equal to 0.5, and tested this assumption in a sensitivity analysis (see Supplementary Section 12). The transmission probability during a contact was assumed to be $p = 0.06$ as in [23]. Remaining parameter estimates were taken from [23] where possible, to replicate the conditions used to derive the clinical fraction estimates. The mortality probability of subclinical infections was assumed to be 0 for all age groups (a). The age-specific probability of mortality of clinical cases were estimated using age-specific IFRs (ϕ_a) found by Verity et al. in 2020 [27] (Supplementary Table 4). As the IFR is $\phi_a = \pi_a \mu_{ca} + (1 - \pi_a)\mu_{sa} = \pi_a \mu_{ca}$, since $\mu_{sa} = 0$, the age-specific clinical mortality probabilities were estimated by

$$\mu_{ca} = \frac{\phi_a}{\pi_a}$$

where π_a are the age-specific clinical fractions for the general population in [23] (Supplementary Table 4).

We calculated the total infections, clinical cases, and fatalities per 1,000 people, the peak number of clinical cases per 1,000 people, the IFR, and the basic reproduction number (R_0) for each IMD decile in urban and rural areas. We also calculated age-standardised measures of total infections, clinical cases, and fatalities within a specific geography for increased comparability. The age-standardised results were of the form:

$$D^{\text{standard}}(365) = \sum_{a=1}^{17} \frac{D_a(365)n_a^u}{n_a}$$

Where $n^u = (n_1^u, ..., n_{17}^u)$ is the standard urban population, similarly $n^r = (n_1^r, ..., n_{17}^r)$ for rural areas.

11

R_0 of each IMD decile in urban and rural areas was calculated as the absolute value of the largest eigenvalue of the next-generation matrix N:

$$N = (N_{ij})_{i,j=1,...,17}$$

$$N_{ij} = pM_{ij}(\pi_j(\gamma + r_c) + \xi(1 - \pi_j)r_s)$$

Counterfactual scenarios

To determine the epidemic burden attributable to the difference in underlying health status between IMD deciles, we created the counterfactual health prevalence scenario, where all deciles were assigned the age-specific health prevalence of decile 10 (the least deprived). We calculated the total clinical cases and fatalities in each IMD decile under this assumption. In order to reflect the size of each population (while each IMD decile comprises 10% of the population of England, geography-specific IMD deciles vary widely in size, see Supplementary Table 1), we scaled mortality to mid-year 2020 population sizes and totalled over the 20 populations.

We also created the counterfactual scenario of constant age structure, where we held the age structure constant at the average of each geography-specific England population, independent of IMD decile. This allowed us to determine the impact of clinical vulnerability separately from the differences in age distribution in each IMD decile. The health prevalence by age remained at the IMD-specific value.

School closures

To quantify potential differences in impact of school closures in different IMD deciles, we calculated the effect of school closures on R_0 and total fatalities. We removed the school-specific contacts from the contact matrix (retaining contacts in home, work, leisure, transport, or other locations), re-projected onto the 2020 age structure, and recalculated the next-generation matrix, N, and its largest eigenvalue, R_0. While assuming that closure of schools results in a complete subtraction of school-specific contacts may not be realistic (as some contacts would likely be replaced by social interactions in other locations [28]), the results demonstrate the maximum potential impact of school closures.

We simulated closure of schools after a certain cumulative proportion, P, of the population developed clinical COVID-19 cases. The use of cumulative clinical cases as a threshold for implementation is reflective of using total confirmed cases as a measure of the size of an early epidemic. We assumed a value of P = 0.05, but tested different values in sensitivity analyses (Supplementary Section 11).

Results

Self-reported health prevalence is lower in more deprived areas

There was an older age structure in rural areas compared to urban, and a generally younger age structure in more deprived areas (Figure 1a). The relationship between IMD decile and age structure was confirmed by the median age in each population (Supplementary Figure 1); rural areas have consistently higher median ages than urban areas of the same IMD decile. Median age monotonously increased with affluence in urban areas, but peaked in the fourth decile for those living in rural areas.

13

Health prevalence was lower in each IMD decile as age increased, and lower in every age group as relative deprivation increased (Figure 1b). 47% of those aged 65-69 living in the most deprived decile reported living in 'Very good' or 'Good' health, compared to 80% of those in the least deprived decile. Those living in the most deprived decile experienced the same health prevalence (76%) at age 40-44 as those in the least deprived decile did at age 70-74.
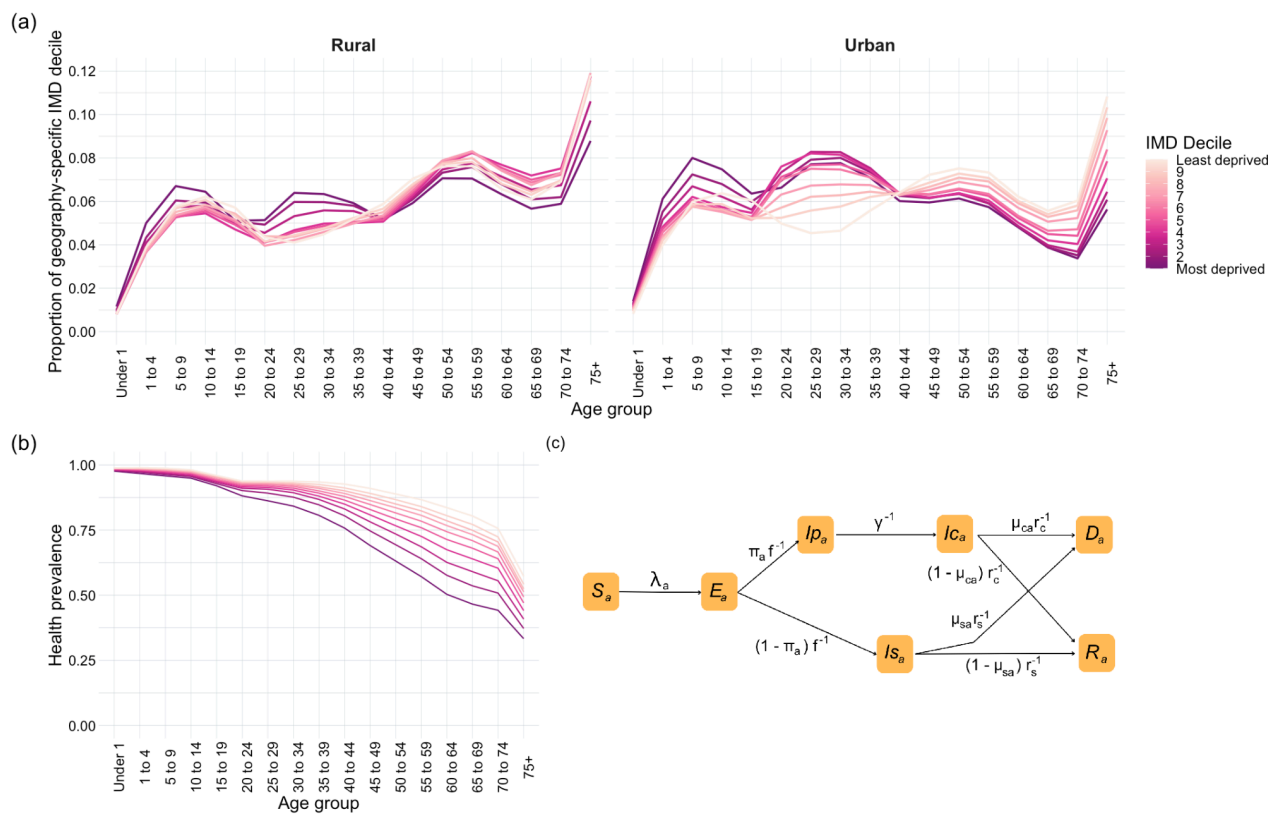

Figure 1. a) Proportion of each geography-specific IMD decile in each age group. **b)** Age- and IMD-specific health prevalence (1 most deprived decile, 10 least deprived). **c)** Age-stratified SEIRD model, specific to IMD decile and geography. Subscript *a* denotes age-specificity, *c* clinical parameters, and *s* subclinical parameters.

Health prevalence was mapped to clinical fraction in the age groups used in [23] as described in

Methods (Figure 2a). Under this assumption, all those over the age of 10 in more deprived

areas had a greater likelihood of developing a clinical case of COVID-19 than in other deciles
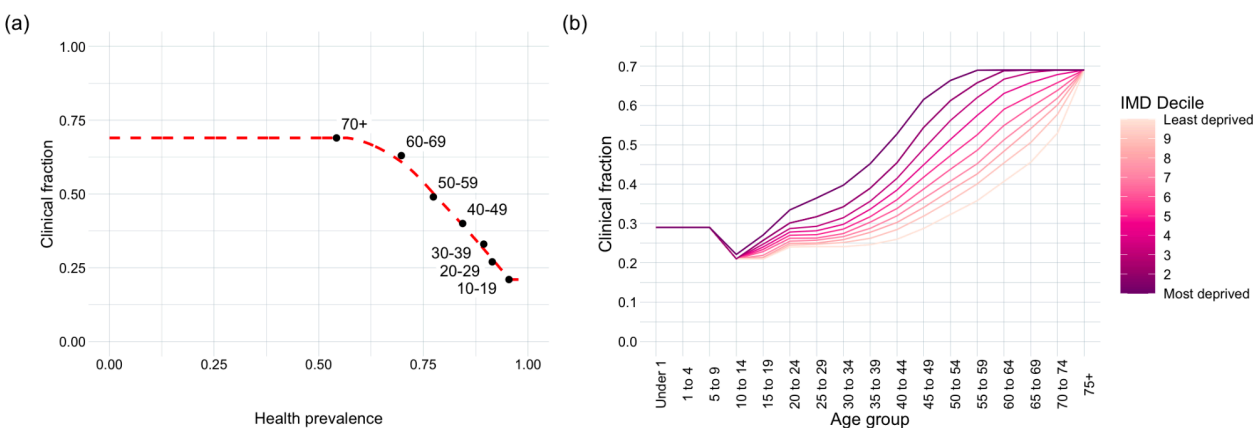
(Figure 2b).

Figure 2. Results of mapping underlying health to clinical vulnerability. a) The training

dataset of age-specific health prevalence and clinical fraction estimates for the general

population of England over age 10, and corresponding LOESS smoother, with linear extensions

outside the domain [0.21, 0.69]. **b)** Resulting age- and IMD-specific clinical fractions (1 most

deprived decile, 10 least deprived).

Epidemic burden increases with relative deprivation

We found that total infections and clinical cases increased with deprivation (Figure 3a, b). In

rural settings, the most deprived decile experienced 72 more crude infections per 1,000

15

population than the least deprived decile; this inequality increased to 90 infections in urban

settings. The inequalities in clinical cases were even larger: in rural areas, the most deprived

decile experienced 147 more clinical cases per 1,000 than the least, and 130 more clinical

cases in urban areas. The peak clinical epidemic size was 97% larger in urban areas of the

most deprived decile than the least deprived decile under these model assumptions, and 91%

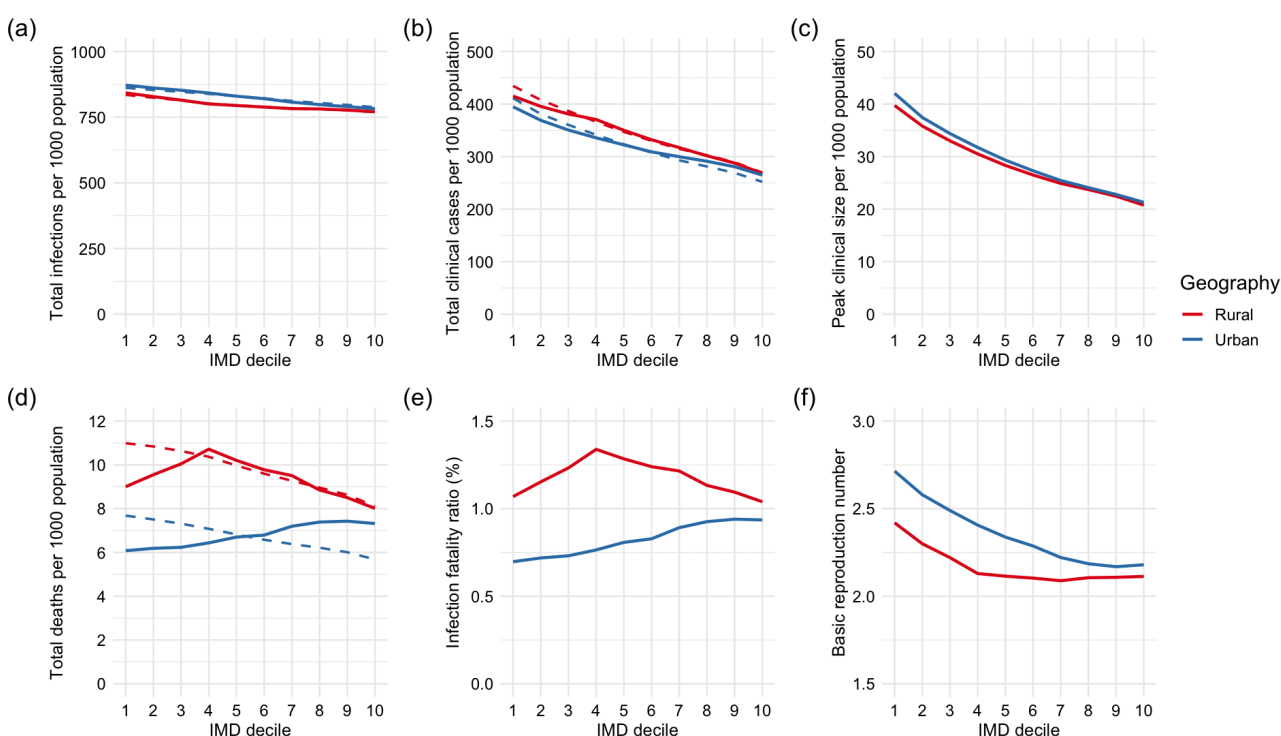larger in rural areas (Figure 3c).

Figure 3. Measures of size of a COVID-19 epidemic in each IMD decile and geography.

Solid lines represent crude measures, and dashed lines those age-standardised by geography.

The most deprived decile is decile 1, and the least is decile 10. **a)** Total infections per 1,000

population. **b)** Total clinical cases per 1,000 population. **c)** Clinical cases per 1,000 population at

the clinical peak of the epidemic. **d)** Total deaths per 1,000 population. **e)** Infection fatality ratio.

f) Basic reproduction number, R_0.

Mortality inequalities differed between the crude and age-standardised results (Figure 3d). The crude total number of deaths by IMD decile and geography closely followed the median age (Supplementary Figure 1). There was a strong positive association between increasing relative deprivation (decreasing decile) and the age-standardised number of deaths (Figure 3d). In urban areas, 2.0 more deaths occurred per 1,000 age-standardised population in the most deprived decile than the least; this inequality increased to 2.9 deaths per 1,000 age-standardised population in rural areas. The IFR followed a very similar pattern to crude mortality (Figure 3e), likely due to a combination of the relative stability of total infections with deprivation compared to the large variation in mortality rates, and the strong relationship between median age and mortality.

R_0 was generally higher in more deprived areas (Figure 3f), and ranged from 2.09 in rural areas of the 7th decile to 2.71 in urban areas of the most deprived decile. The R_0 was not strongly related to median age because the lower clinical fractions in younger populations was counteracted by their higher contact rates.

Rural areas experienced fewer total infections, lower peak clinical sizes, and lower R_0 than urban areas, but more clinical cases and deaths, at all levels of deprivation. This is likely due to the older rural age structure, as older individuals had fewer daily contacts than younger individuals and so produced fewer secondary infections, but were more likely to develop clinical COVID-19 if infected.

Health-attributable deaths occur at all ages

Under the counterfactual health prevalence scenario, 340,532 deaths occurred, compared to the 405,695 under the original assumption. Therefore, 16% of deaths, or over 65,000 fatalities,

17

would have been prevented by achieving health prevalence equity at the level of the least deprived decile. These health-attributable deaths did not only occur in those at older ages: over 29,000 prevented deaths were in individuals aged under 65 (Supplementary Figure 6). At all ages between 30 and 70, over 20% of deaths that occurred under the original model assumptions were attributable to underlying health inequalities (Supplementary Figure 7). We similarly found 21% of clinical cases (3.8 million) to be attributable to inequalities in underlying health under the model assumptions.

Lower clinical infection and mortality rates occurred in the most deprived areas, in both urban and rural geographies in the counterfactual health prevalence scenario (Figure 4). We also found that age-standardised deaths were consistent across IMD deciles in both geographies under the counterfactual health prevalence scenario (Supplementary Figure 10). This result contradicts observed mortality rates [3], [4], providing evidence for the existence of a dependency of clinical vulnerability on IMD and more specifically underlying health. The true relationship between IMD and age-specific clinical fraction may be more complex than the assumptions made in this paper; for example, pre-existing immunity may be dependent on previous exposure to coronaviruses [29], which may be associated with SES but is not considered here.

In the counterfactual scenario of constant age structure, we observed more clinical cases and deaths in more deprived areas (Figure 4). We also considered an underlying age structure independent of IMD decile or geography and found that the most deprived decile experiences 40% higher mortality and a clinical peak 1.88 times larger than the least deprived decile (Supplementary Figure 5), demonstrating the inequality resulting from health prevalence separately from demographic differences. These results indicate that observed inequalities in clinical case numbers and mortality are the result of a complex interaction between

18

comorbidity-related clinical vulnerability and a population's demographic structure, the outcome
of which is not necessarily consistently related to deprivation.

Further sensitivity analyses considering epidemiological parameters show consistent patterns of
age-standardised mortality by deprivation, but a change in the pattern of crude deaths as
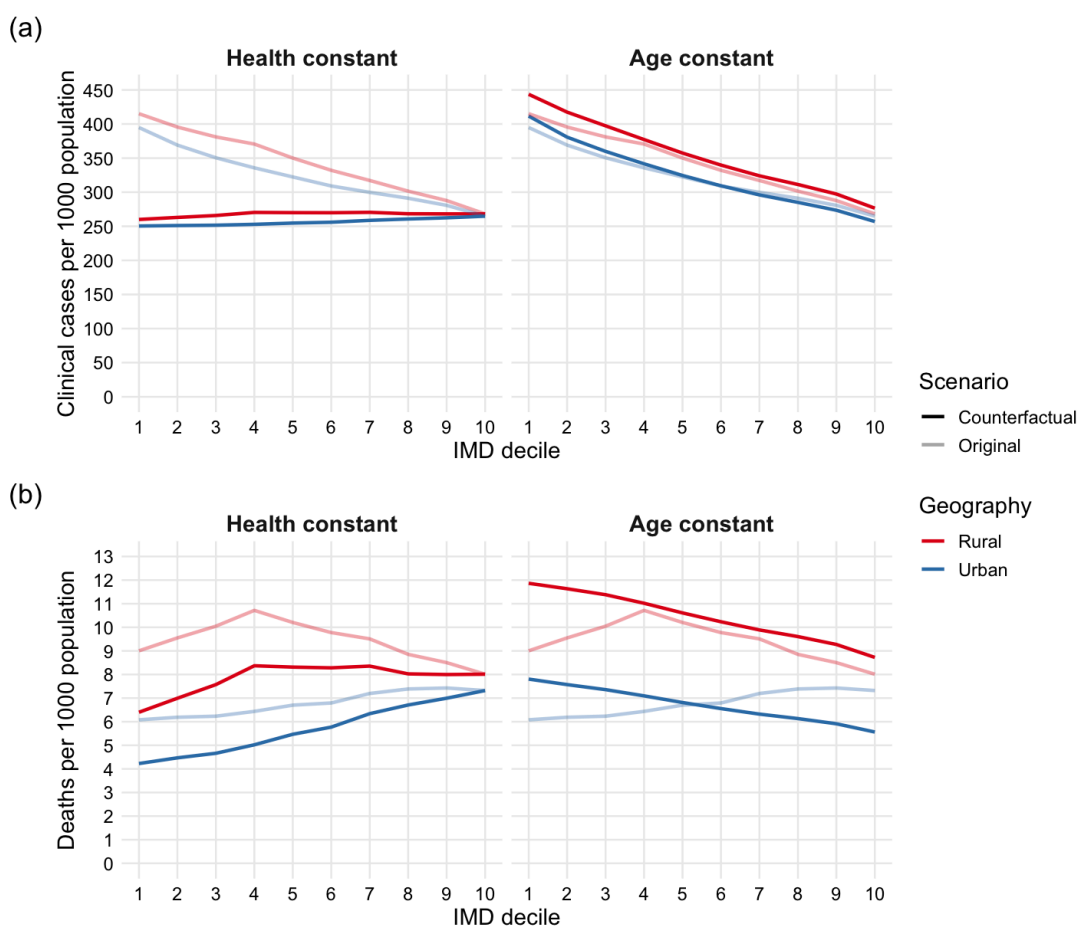epidemiological parameters are varied (Supplementary Section 12).

Figure 4. Epidemiological burden in counterfactual scenarios. a) Total clinical cases per
1,000 population, in geography-specific areas of each IMD decile (1 most deprived decile, 10
least deprived), in the counterfactual health prevalence scenario, and the counterfactual
constant age structure scenario. The original model is shown for comparison in pale lines. **b)**

Total deaths per 1,000 population, in geography-specific areas of each IMD decile, under the same scenarios.

School closures prevent more deaths in less deprived areas

With school closures in place in the model, R_0 decreased for all geographies and IMD deciles, but remained larger in urban than rural areas and was consistently higher in more deprived areas in both geographies (Figure 5a). In urban areas, R_0 was 0.38 higher in the most deprived decile than the least; the equivalent inequality was 0.29 in rural areas. The largest reductions in R_0 occurred in the most and least deprived deciles, with the least impact in the median deciles (Figure 5b). This U-shaped result is likely a product of the age structure of each population, as R_0 is driven by both high daily contact patterns in young individuals and greater clinical vulnerability in older individuals (more detail in Supplementary Section 11). In all IMD deciles, greater reductions in R_0 occurred in urban than rural areas, likely due to the greater proportion of school-aged children and hence larger reduction in contacts. In no scenario was R_0 reduced below 1 (Figure 5a), meaning that school closures were not able to halt COVID-19 transmission in any rural or urban IMD decile and could only reduce epidemic burden under our model assumptions.

By implementing school closures after 5% of the population experienced a clinical case of COVID-19 (P = 0.05), 0.113 more crude deaths were prevented per 1,000 people in the least deprived urban areas than the most deprived, with a corresponding difference of 0.073 deaths per 1,000 people in rural areas (Figure 5c). This is likely due to a combination of more crude deaths occurring in more affluent deciles without intervention, improved health conditions, and older population structures. The deaths prevented when age-standardised by geography, shown as dashed lines, were approximately consistent by IMD. We also investigated the pattern of prevented mortality when changing the school closure implementation threshold

20

(Supplementary Section 11) and found that the effectiveness of school closures in less deprived areas decreased dramatically as P increased.
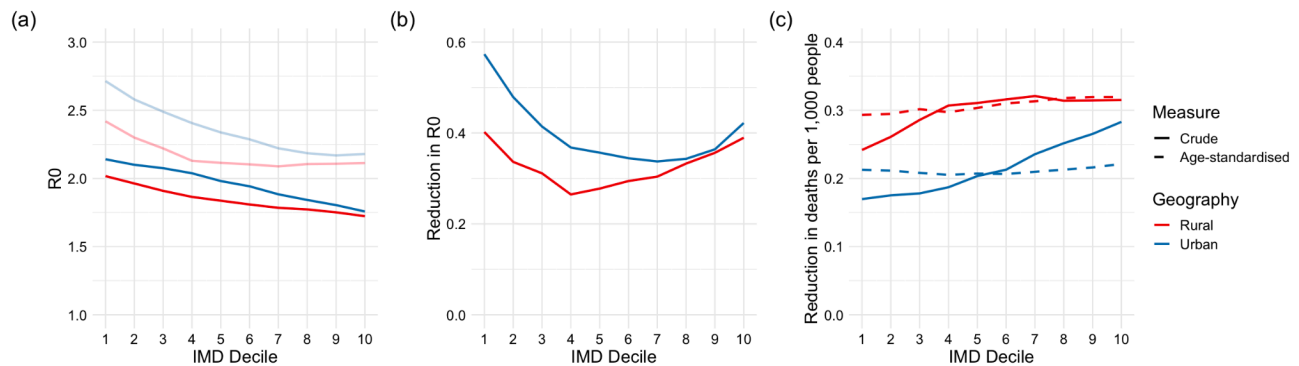

Figure 5. Results of implementing school closures. a) R_0 in each IMD- and geography-specific population (1 most deprived decile, 10 least deprived), before (pale lines) and after school closures. **b)** Reductions in R_0 due to school closures. **c)** Crude (solid lines) and age-standardised by geography (dashed lines) reductions in deaths observed per 1,000 population after implementing school closures at P = 0.05.

Discussion

We have shown that, under the assumption that vulnerability to clinical COVID-19 infection is a direct result of a population's health prevalence, total COVID-19 infections, clinical cases, and age-standardised deaths consistently increased with relative deprivation, therefore exposing those living in the most deprived areas to a greater risk of mortality, as well as more non-fatal consequences such as hospitalisation and long COVID. The peak clinical size of the modelled COVID-19 epidemics, which describes the worst-case scenario hospitals would have to withstand, was approximately twice as large in the most deprived decile than the least deprived. We have found that 16% of the deaths observed under the assumptions of this model, or over 65,000 deaths, would be prevented if every IMD decile experienced the same age-specific

health as the most affluent 10% of the country. We have also shown that school closures, which disproportionately negatively affect children's education and well-being in more deprived areas, may also disproportionately benefit the most affluent in society in terms of epidemiological burden [30], [31].

This study used publicly available data and relied on simplified models of infectious disease transmission; there are hence several limitations to the study. The self-reported nature of the census data means that there may be systematic differences in how health is reported between age groups and levels of deprivation, due to social desirability and the acceptability of self-reporting ill-health varying by demographic, cultural and socioeconomic factors [9]. Census data and the IMD may exclude mobile communities and the over 270,000 homeless individuals in England, who are often among the most vulnerable members of society [32], [33]. Self-reported health in 2021 may include effects of the COVID-19 pandemic, and so pre-emptively confirm the inequities that this model aims to investigate. However, the IMD-specific health prevalence in 2021 (Supplementary Table 3) is very similar to that found in the 2011 UK Census (75.0% health prevalence in the most deprived decile, and 86.9% in the least deprived decile) [34].

Much of the data used in calculating the IMD relate to 2015-2016 [16]. Any changes that have occurred since are therefore not accounted for in the IMD rankings, such as the wider roll-out of Universal Credit, which has been shown to have exacerbated existing inequalities and negatively impacted claimants' well-being [35], [36]. Health is itself a component of the IMD, potentially limiting the IMD as an exposure for studies with health outcomes; a brief analysis confirms that there are associations between domains of deprivation other than health (Supplementary Figure 13). Other studies have also confirmed the relationship between local deprivation and health outcomes when factoring out the health component of IMD [37].

The assumption of a closed population is unrealistic: apart from during the most stringent lockdowns, which are not represented by the contact patterns used in the above work, individuals will interact and transmit infection between LSOAs as well as within them. A major limitation of the contact patterns used is that intrinsic contact patterns are unlikely to be constant across all IMD deciles and urban and rural geographies. Contact patterns also drastically change in an epidemic, to an extent which depends on SES. The more affluent can more readily reduce their mobility and exposures, while many in the most deprived deciles have less control over their mobility and exposure patterns and are more likely to be in public-facing employment [38]. The ability to self-isolate may also depend on SES, for instance through the conditions of sick pay. The assumptions of constant contact patterns were necessary due to a lack of readily available data on IMD- and age-specific contact patterns, both under NPIs and in daily life, and as a consequence this study is likely to have underestimated the socioeconomic inequalities in epidemic burden. SES-specific contact patterns should be incorporated into epidemic models to include the different contacts that for example arise from different occupational prevalences, ability to reduce mobility, household size, and classroom size. To this end, further data should be collected and made accessible for future research.

By restricting clinical fractions between 0.21 and 0.69, clinical fractions converged at the upper bound in deprived deciles over age 60 while health prevalences were still diverging, meaning that the assigned clinical fractions may underestimate the potential difference in vulnerability, and therefore epidemiological burden, between these IMD deciles. The parameters used for the model, taken from [23] and [27], contain some uncertainty which is included in the original papers but not considered in this study.

The presence of drastically worse underlying health conditions in more deprived areas of England has caused, and will continue to cause, dramatic inequities in the burden of infectious disease. This study has quantified the potential inequities in epidemic burden under the assumption that vulnerability to severe infection is a direct result of existing comorbidities. The most effective way to reduce the inequality of epidemic burden caused by socioeconomic health inequalities is to improve socioeconomic equity in health in England. The recommendations made by Health Equity in England: The Marmot Review 10 Years On [10], including maximising empowerment for all, improving standards of living, creating fair employment, and developing healthy communities, would reduce avoidable inequalities in health and by extension avoidable inequalities in epidemic burden.

Abbreviations

CI: Confidence Interval

IFR: Infection Fatality Ratio

IMD: Index of Multiple Deprivation

LSOA: Lower Layer Super Output Area

NPI: Non-Pharmaceutical Intervention

SEIRD: Susceptible-Exposed-Infectious-Recovered-Dead

SES: Socio-Economic Status

UK: United Kingdom

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All analysis code and data are available at https://github.com/1035825/imd-covid.

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

LG and RME conceived the study. LG developed the mathematical model and conducted the analyses. RME and EvL consulted on the analyses. LG and RME wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

[1] G. E. Mena, P. P. Martinez, A. S. Mahmud, P. A. Marquet, C. O. Buckee, and M. Santillana, 'Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile', *Science*, vol. 372, no. 6545, p. eabg5298, May 2021, doi: 10.1126/science.abg5298.

[2] U. Bilal, L. P. Tabb, S. Barber, and A. V. Diez Roux, 'Spatial Inequities in COVID-19 Testing, Positivity, Confirmed Cases, and Mortality in 3 U.S. Cities: An Ecological Study', *Ann. Intern. Med.*, vol. 174, no. 7, pp. 936–944, Jul. 2021, doi: 10.7326/M20-3936.

[3] Office for National Statistics, 'Deaths involving COVID-19 by local area and deprivation.' Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingcovid19bylocalareaanddeprivation

[4] UK Health Security Agency, 'COVID-19 confirmed deaths in England (to 31 January 2021): report', GOV.UK. Accessed: Oct. 04, 2023. [Online]. Available: https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-report

[5] Public Health England, 'Disparities in the risk and outcomes of COVID-19', Aug. 2020, Accessed: Oct. 04, 2023. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908434/Disparities_in_the_risk_and_outcomes_of_COVID_August_2020_update.pdf

[6] K. Bhaskaran *et al.*, 'Factors associated with deaths due to COVID-19 versus other causes: population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform', *Lancet Reg. Health - Eur.*, vol. 6, p. 100109, Jul. 2021, doi: 10.1016/j.lanepe.2021.100109.

[7] K. E. Mason, G. Maudsley, P. McHale, A. Pennington, J. Day, and B. Barr, 'Age-Adjusted

Associations Between Comorbidity and Outcomes of COVID-19: A Review of the Evidence From the Early Stages of the Pandemic', *Front. Public Health*, vol. 9, p. 584182, Aug. 2021, doi: 10.3389/fpubh.2021.584182.

[8] L. Boglione and V. Dodaro, 'Impact of socioeconomic status on the clinical outcomes in hospitalised patients with SARS-CoV-2 infection: a retrospective analysis', *J. Public Health*, vol. 31, no. 10, pp. 1691–1697, Oct. 2023, doi: 10.1007/s10389-022-01730-2.

[9] Office for National Statistics, 'Health state life expectancies by national deprivation deciles, England: 2017 to 2019'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequaliti es/bulletins/healthstatelifeexpectanciesbyindexofmultipledeprivationimd/2017to2019

[10] M. Marmot, J. Allen, T. Boyce, P. Goldblatt, and J. Morrison, 'Health Equity in England: The Marmot Review 10 Years On - The Health Foundation'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.health.org.uk/publications/reports/the-marmot-review-10-years-on

[11] P. D. Rutter, O. T. Mytton, M. Mak, and L. J. Donaldson, 'Socio-economic disparities in mortality due to pandemic influenza in England', *Int. J. Public Health*, vol. 57, no. 4, pp. 745–750, Aug. 2012, doi: 10.1007/s00038-012-0337-1.

[12] X. Ye *et al.*, 'Associations of socioeconomic status with infectious diseases mediated by lifestyle, environmental pollution and chronic comorbidities: a comprehensive evaluation based on UK Biobank', *Infect. Dis. Poverty*, vol. 12, no. 1, p. 5, Jan. 2023, doi: 10.1186/s40249-023-01056-5.

[13] S.-E. Mamelund, C. Shelley-Egan, and O. Rogeberg, 'The association between socioeconomic status and pandemic influenza: Systematic review and meta-analysis', *PLOS ONE*, vol. 16, no. 9, p. e0244346, Sep. 2021, doi: 10.1371/journal.pone.0244346.

[14] K. Morrissey, F. Spooner, J. Salter, and G. Shaddick, 'Area level deprivation and monthly COVID-19 cases: The impact of government policy in England', *Soc. Sci. Med.*, vol. 289, p. 114413, Nov. 2021, doi: 10.1016/j.socscimed.2021.114413.

[15] UK Health Security Agency, 'COVID-19 testing in a deprived local authority: Hackney, London', GOV.UK. Accessed: Oct. 04, 2023. [Online]. Available: https://www.gov.uk/government/case-studies/covid-19-testing-in-a-deprived-local-authority-hackney-london

[16] Ministry of Housing, Communities & Local Government, 'English indices of deprivation 2019', GOV.UK. Accessed: Oct. 04, 2023. [Online]. Available: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

[17] Office for National Statistics, 'Lower layer Super Output Area population estimates (supporting information)'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates

[18] Office for National Statistics, '2011 rural/urban classification'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification

[19] Department for Environment, Food & Rural Affairs., '2011 Rural Urban Classification lookup tables for all geographies', GOV.UK. Accessed: Oct. 04, 2023. [Online]. Available: https://www.gov.uk/government/statistics/2011-rural-urban-classification-lookup-tables-for-all-geographies

[20] J. Mossong *et al.*, 'Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases', *PLoS Med.*, vol. 5, no. 3, p. e74, Mar. 2008, doi: 10.1371/journal.pmed.0050074.

[21] S. Funk, L. Willem, and H. Gruson, 'socialmixr: Social Mixing Matrices for Infectious Disease Modelling'. Accessed: Oct. 04, 2023. [Online]. Available: https://github.com/epiforecasts/socialmixr, https://epiforecasts.io/socialmixr/

[22] S. Arregui, A. Aleta, J. Sanz, and Y. Moreno, 'Projecting social contact matrices to different demographic structures', *PLOS Comput. Biol.*, vol. 14, no. 12, p. e1006638, Dec. 2018, doi:

10.1371/journal.pcbi.1006638.

[23] N. G. Davies *et al.*, 'Age-dependent effects in the transmission and control of COVID-19 epidemics', *Nat. Med.*, vol. 26, no. 8, pp. 1205–1211, Aug. 2020, doi: 10.1038/s41591-020-0962-9.

[24] Office for National Statistics, 'General health by age, sex and deprivation, England and Wales'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellb eing/datasets/generalhealthbyagesexanddeprivationenglandandwales

[25] COVID-19 Forecasting Team, 'Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis', *The Lancet*, vol. 399, no. 10334, pp. 1469–1488, Apr. 2022, doi: 10.1016/S0140-6736(21)02867-1.

[26] 'Influenza (Seasonal)'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[27] R. Verity *et al.*, 'Estimates of the severity of coronavirus disease 2019: a model-based analysis', *Lancet Infect. Dis.*, vol. 20, no. 6, pp. 669–677, Jun. 2020, doi: 10.1016/S1473-3099(20)30243-7.

[28] C. Jackson *et al.*, 'School Closures and Student Contact Patterns', *Emerg. Infect. Dis.*, vol. 17, no. 2, pp. 245–247, Feb. 2011, doi: 10.3201/eid1702.100458.

[29] G. Wang, Z. Xiang, W. Wang, and Z. Chen, 'Seasonal coronaviruses and SARS-CoV-2: effects of preexisting immunity during the COVID-19 pandemic', *J. Zhejiang Univ. Sci. B*, vol. 23, no. 6, pp. 451–460, Jun. 2022, doi: 10.1631/jzus.B2200049.

[30] S. J. Lewis, A. P. S. Munro, G. D. Smith, and A. M. Pollock, 'Closing schools is not evidence based and harms children', *BMJ*, p. n521, Feb. 2021, doi: 10.1136/bmj.n521.

[31] R. Viner *et al.*, 'School Closures During Social Lockdown and Mental Health, Health Behaviors, and Well-being Among Children and Adolescents During the First COVID-19 Wave: A Systematic Review', *JAMA Pediatr.*, vol. 176, no. 4, p. 400, Apr. 2022, doi:

10.1001/jamapediatrics.2021.5840.

[32] A. Moscrop, S. Ziebland, G. Bloch, and J. R. Iraola, 'If social determinants of health are so important, shouldn't we ask patients about them?', *BMJ*, p. m4150, Nov. 2020, doi: 10.1136/bmj.m4150.

[33] Shelter, 'At least 271,000 people are homeless in England today', Shelter England. Accessed: Oct. 04, 2023. [Online]. Available: https://england.shelter.org.uk/media/press_release/at_least_271000_people_are_homeless _in_england_today

[34] Office for National Statistics, 'General Health in England and Wales: 2011 and comparison with 2001'. Accessed: Oct. 04, 2023. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellb eing/articles/generalhealthinenglandandwales/2013-01-30#general-health-and-area-depriva tion-for-small-area-groupings

[35] R. Keen and S. Kennedy, 'Universal Credit changes from April 2016', Nov. 2016, Accessed: Oct. 04, 2023. [Online]. Available: https://commonslibrary.parliament.uk/research-briefings/cbp-7446/

[36] M. Cheetham, S. Moffatt, M. Addison, and A. Wiseman, 'Impact of Universal Credit in North East England: a qualitative study of claimants and support staff', *BMJ Open*, vol. 9, no. 7, p. e029611, Jul. 2019, doi: 10.1136/bmjopen-2019-029611.

[37] J. Adams and M. White, 'Removing the health domain from the Index of Multiple Deprivation 2004—effect on measured inequalities in census measure of health', *J. Public Health*, vol. 28, no. 4, pp. 379–383, Dec. 2006, doi: 10.1093/pubmed/fdl061.

[38] W. D. Lee, M. Qian, and T. Schwanen, 'The association between socioeconomic status and mobility reductions in the early stage of England's COVID-19 epidemic', *Health Place*, vol. 69, p. 102563, May 2021, doi: 10.1016/j.healthplace.2021.102563.