medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Age-dependent effects in the transmission and control of COVID-19 epidemics

Authors: Nicholas G. Davies[1*], Petra Klepac[1^], Yang Liu[1^], Kiesha Prem[1], Mark Jit[1], CMMID COVID-19 working group, Rosalind M Eggo[1*]

The CMMID COVID-19 working group[1] is: Carl A B Pearson, Billy J Quilty, Adam J Kucharski, Hamish Gibbs, Samuel Clifford, Amy Gimma, Kevin van Zandvoort, James D Munday, Charlie Diamond, W John Edmunds, Rein MGJ Houben, Joel Hellewell, Timothy W Russell, Sam Abbott, Sebastian Funk, Nikos I Bosse, Fiona Sun, Stefan Flasche, Alicia Rosello & Christopher I Jarvis. Order of working group determined at random.

[1] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, WC1E 7HT

^ these authors contributed equally

* correspondence to Rosalind M Eggo r.eggo@lshtm.ac.uk or Nicholas G Davies nicholas.davies@lshtm.ac.uk

The COVID-19 pandemic has shown a markedly low proportion of cases among children. Age disparities in observed cases could be explained by assortative mixing patterns and reactive school closures which decrease mixing between children, or by children exhibiting lower susceptibility to infection, or by children having a lower propensity to show clinical symptoms. We formally test these hypotheses by fitting an age-structured mathematical model to epidemic data from six countries, finding

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

strong age dependence in the probability of developing clinical symptoms, rising from around 20% in under 10s to over 70% in older adults. We find that interventions aimed at halting transmission in children may have minimal effects on preventing cases depending on the relative transmissibility of subclinical infections. Our estimated age-specific clinical fraction has implications for the expected global burden of clinical cases because of demographic differences across settings. In younger populations, the expected clinical attack rate would be lower, although it is likely that comorbidities in low-income countries will affect disease severity. Without effective control measures, regions with older populations may see disproportionally more clinical cases, particularly in the later stages of the pandemic.

The outbreak of COVID-19 caused by a novel coronavirus has shown a markedly low proportion of children among reported cases from China[1,2] and other countries[3,4], a feature shared with the 2003 SARS epidemic[5] but not with the 2009 influenza A/H1N1p pandemic[6,7]. There are not only relatively few COVID-19 cases reported in children, but more generally an increased number of cases and risk of severe disease as age increases[8,9]. Understanding the role of age in transmission and disease severity is critical for determining the likely impact of social-distancing interventions for decreasing transmission, especially those aimed at schools, and for estimating the expected global disease burden.

There are at least three hypotheses that could give rise to the age gradient in observed COVID-19 cases in China. First, the contact patterns and demographics of China, and Wuhan in particular, could have resulted in fewer children being infected. The outbreak was linked to the Huanan Market in Wuhan City, China, and early cases were in older adults[10]. Assortative mixing between adults could therefore have reduced transmission to children in the very early stages of the outbreak. The subsequent closure of schools on 12th January

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2020 for the Lunar New Year holiday could have reinforced this effect. Children tend to make more social contacts than adults[11] and hence, all else equal, contribute more to transmission in the community than adults[12,13]. This is why school closures are considered a key intervention for epidemics of respiratory infections[8], and without them, one would expect a higher number of infected children. There is also a low proportion of children in China (21% under 18[14]), which would decrease the relative proportion of cases seen in children. Outside of China, COVID-19 outbreaks may have been initially seeded by working-age travellers entering the country[15,16], producing a similar excess of older individuals in early phases of local epidemics.

Second, there could be age-varying susceptibility to infection by SARS-CoV-2, where children may be less susceptible to becoming infected on contact with an infectious person. This would further reduce cases among children, and potentially lower transmission in the population overall. Decreased susceptibility could result from immune cross-protection from other coronaviruses, or non-specific protection resulting from recent infection by another respiratory virus[17] of which children have higher rates[18,19].

Third, children may be as susceptible to SARS-CoV-2 infection as other individuals, but may more frequently experience milder or no symptoms. Such age-dependent variation in severity has been observed for other respiratory virus infections[20], including SARS[14,15]. For COVID-19, there are indications of age dependence in severity[8] and mortality[21] of reported cases[21], which could extend to severity and likelihood of clinically reportable symptoms on infection. These “subclinical” infections are more likely to remain undetected and would lead to a reduction in reported cases among children, but children could still be capable of transmitting the virus to others, potentially at lower rates than individuals exhibiting clinical

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

infections, as shown for influenza[22]. Subclinical cases are sometimes called "asymptomatic" but very mild symptoms may not be noticed, even though they occur.

Distinguishing which of these hypotheses is most supported by available data has important implications for policies that aim to control transmission[23], especially through interrupting child-driven transmission. For example, the impact of school closures depends on the role and importance of children in the epidemic. Additionally, if the number of infections or cases depends strongly on the role of children, countries with different age distributions could exhibit substantially different epidemic profiles and overall impact of COVID-19 epidemics.

The role of children in transmission in Wuhan

We tested three hypotheses—(1) no age variation in susceptibility or severity, with the age distribution of cases driven by age-dependent contact patterns alone; (2) varying susceptibility to infection by age; and (3) varying clinical fraction by age—using an age-structured dynamic transmission model (see Methods) with heterogeneous contact rates between age groups drawn from social contact surveys in Shanghai[17]. We generated model variants for each hypothesis (**Fig 1a**) and fitted to three data sources from the early epidemic in Wuhan for each hypothesis (**Fig, 1b, 1c**). We included school closures, for which we decreased the school contacts of children. We also estimated the effect of the holiday period, and the travel and movement restrictions in Wuhan, on transmission (**Fig 1d**). We found that under each hypothesis, the basic reproduction number R_0 was 3.2–3.6 initially, was inflated 1.2–1.5-fold during the pre Lunar New Year holiday period, and then fell by 80–95% during restrictions in Wuhan, which brought R_0 below 1 (**Fig 1e**).

All model variants fitted the daily incident number of confirmed cases equally well (**Fig 1f**). However, hypothesis 1 did not reproduce the observed age distribution of cases, overestimating the number of cases in children and underestimating cases in older adults

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

(**Fig 1g**). Hypotheses 2 and 3 fitted the age distribution of cases, but each implied a very different fraction of subclinical infections by age, with much higher numbers of subclinical infections under hypothesis 2 **(Fig 1h)**. Comparison using Deviance Information Criterion[6] (DIC) showed that hypothesis 2 (DIC: 691) and 3 (DIC: 558) were preferred over hypothesis 1 (DIC: 880), and hypothesis 3 was better supported than hypothesis 2.

Under hypothesis 2, where the severity was equal by age, 20% of both clinical and subclinical infections occurred in the 70-100 year old age group (**Fig 1h**). Under hypothesis 3, 20% of clinical cases but less than 5% subclinical cases are in this group. Recent work has demonstrated an age-dependent severity in hospitalised confirmed cases[24,25], which suggests that a high rate of subclinical infection in older adults may not be realistic. Close follow-up of contacts of cases in Shenzhen, China, found that children were infected at the same rate as adults[16], lending more weight to hypothesis 3. Additionally, evidence of increased severity by age[19] suggests that clinical signs are more likely in older adults, which further decreases the plausibility of hypothesis 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

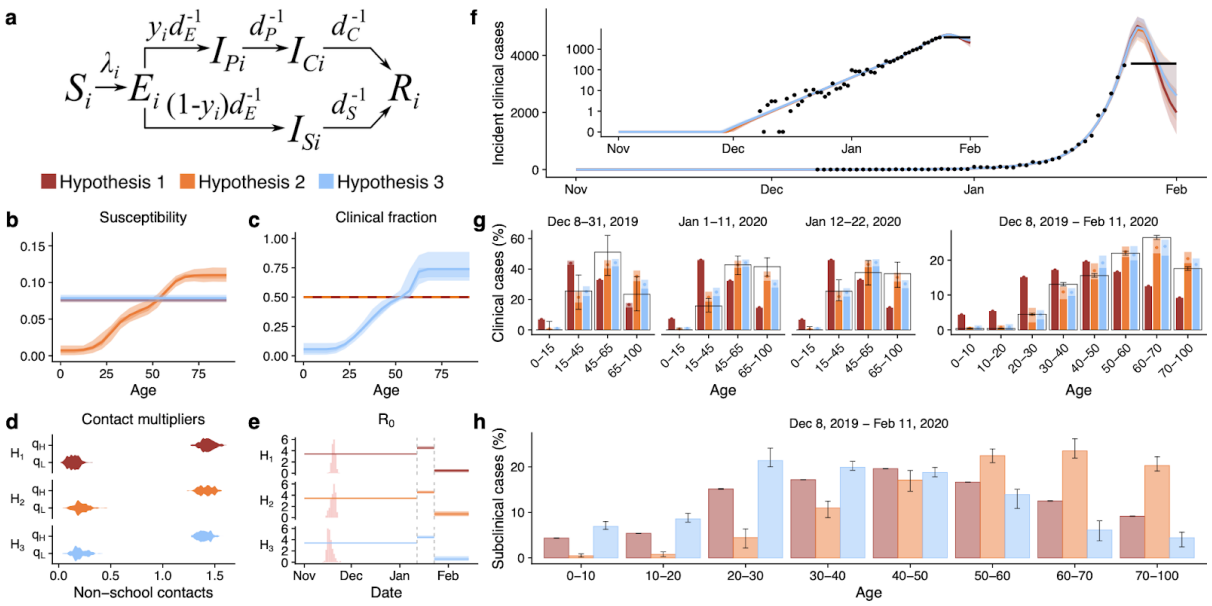

Fig. 1. Comparing the fit of hypotheses 1, 2, and 3 when fitted to data from Wuhan City, China. **(a)** Model diagram showing duration of disease states, where *d* parameters represent the duration of time in each disease state and y_i is the fraction of infections that are clinical in age group *i* (see Methods). **(b)** Susceptibility by age for the three hypotheses. Age-specific values were estimated for hypothesis 2 and all ages had equal susceptibility for hypothesis 1 and 3. Susceptibility is measured as the probability of infection on contact with an infectious person. **(c)** Clinical fraction (y_i) by age for the 3 hypotheses. Age-specific values were estimated for hypothesis 3 and all ages were equal for hypothesis 1 and 2. **(d)** Fitted contact multipliers for holiday and restricted periods for each hypothesis showed an increase in non-school contacts beginning on January 12th (start of Lunar New Year) and a decrease in contacts following restrictions on January 23rd. **(e)** Estimated R_0 values under each hypothesis. The red barplot shows the inferred window of spillover of infection. **(f)** Incident reported cases (black), and modelled incidence of clinical cases for the three hypotheses as fitted to the cases reported by China Centers for Disease Control[1] with onset on or before February 1st, 2020. Line marks mean and shaded window is the 95% highest density interval (HDI). **(g)** Age distribution of cases by onset date as fitted to the age distributions reported by Li et al.[26] Data are shown in the hollow bars, and model predictions in filled bars, where the dot marks the mean posterior estimate. **(h)** Inferred distribution of subclinical cases by age under each hypothesis. Credible intervals on modelled values show the 95% HDIs; credible intervals on data for panels d-f show 95% HDIs for the proportion of cases in each age group.

Estimating the age-specific clinical fraction

Since the initial outbreak in Wuhan, the virus has spread to other regions within China and internationally. Local epidemics have exhibited a less extreme, but still marked lack of reported cases among children. The expected proportion of children infected depends on

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

mitigation measures in place in each region, and is expected to be lower in regions which have closed schools.

Using the best fitting and most biologically plausible hypothesis, hypothesis 3 — age-varying clinical fraction — we estimated the age-specific clinical fraction for 32 settings across six countries by using the stationary distribution of the next generation matrix to reproduce the locally-reported age distribution of cases compiled from a variety of sources **(Fig 2a)**. We used setting-specific demographics, measured contact matrices where possible, and synthetic contact matrices otherwise[27]. The age-dependent clinical proportion was markedly lower in younger age groups in all regions **(Fig 2b),** with 20% of infections in children under 10 resulting in clinical cases, rising to over 70% in adults over 70 in the consensus age distribution estimated across all regions. To determine whether this distribution was capable of reproducing epidemic dynamics, we fitted our dynamic model to the incidence of clinical cases in Beijing, Shanghai, South Korea and Italy (**Fig 2c**). The consensus age-specific clinical fraction was largely capable of reproducing the age distribution of cases, although there are some outliers, for example the 20-30 age group in South Korea. This could be the result of clustered transmission within a church group in this country[4]. The predicted age distribution of cases for Italy is also less skewed towards older adults than reported cases show, suggesting potential differences in age-specific testing in Italy[28]. Locally-estimated age-varying clinical fraction captured these patterns more precisely **(Fig. 2c)**.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

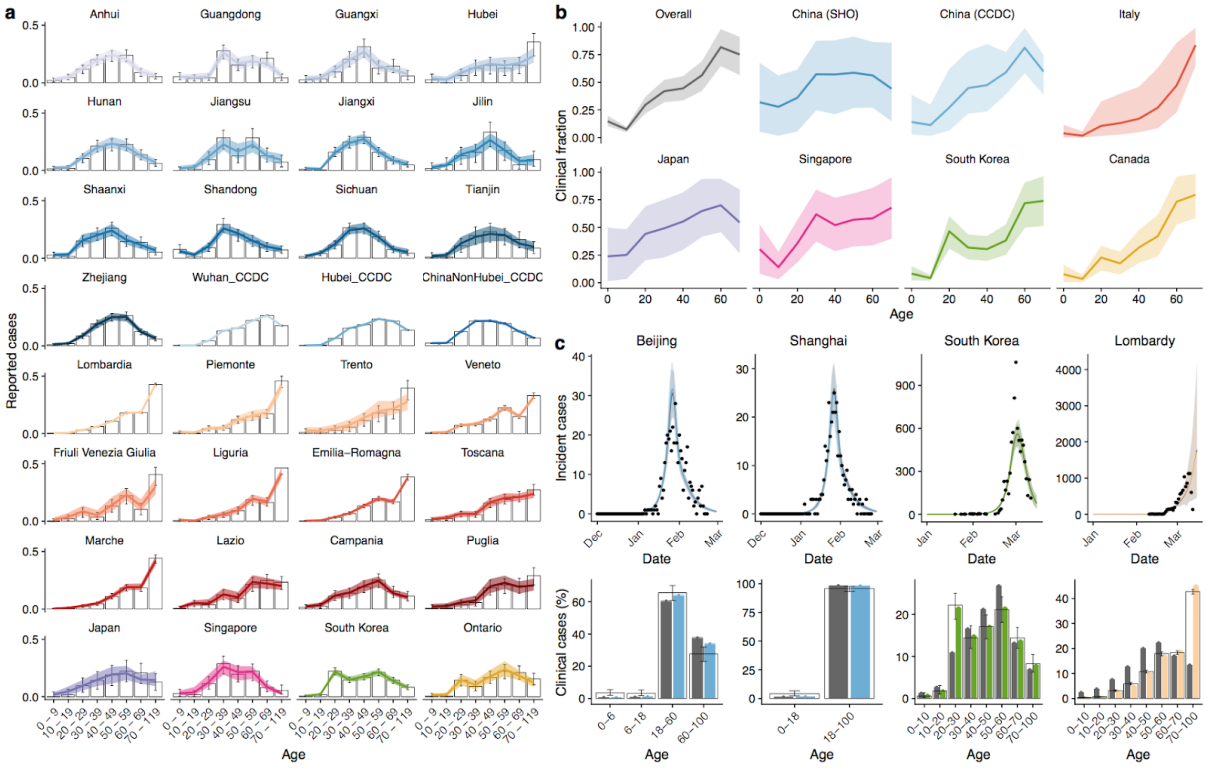

Fig 2. Estimating age-specific symptomatic rate from age-specific case counts for 6 countries. **(a)** Age-specific reported cases from 13 provinces of China, 12 regions of Italy, Japan, Singapore, South Korea, and Ontario, Canada. Hollow bars are data and colour is model fit with 95% HDI. **(b)** Fitted mean and 95% HDI for the age distribution in clinical fraction for all countries. **(c)** Fitted incidence of confirmed cases and resulting age distribution of cases using either the consensus (grey) or country-specific (colour) age-specific clinical fraction from b.

Impact of school closures under different demographics and subclinical infectiousness

School closures during epidemics[29,30] and pandemics[31,32] aim to slow the spread of infections by decreasing transmission amongst children[12]. School closures, which have been used during influenza pandemics[31], can decrease cases in children, but may also have whole-population effects if children play a major role in transmission. The impact will depend on the fraction of the population that are children and the contacts they have with other age groups. Using schematic values[7] for pandemic influenza and our inferred values for COVID-19 **(Figure 3a)** we compared epidemics in three cities with very different demography: Milan (Italy, high median age), Birmingham (UK, intermediate median age),

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

and Bulawayo (Zimbabwe, low median age) **(Fig 3b)**, using measured contact matrices for each country. There were many more clinical cases for COVID-19 than influenza in all cities, with relatively more cases in children in the influenza-like scenario, and more cases in adults in simulated COVID-19 epidemics (**Fig 3c)**. More clinical cases were seen in older adults in Milan compared with the other cities, and a markedly younger age distribution in clinical cases in Bulawayo. Using the same age-dependent clinical fraction drawn from high and upper-middle income countries in low and lower-middle income countries (LMIC) may underestimate clinical cases due to the presence of comorbidities.

To fully explore the effect of school closure we simulated 3 months of school closures with varying infectiousness of subclinical cases, at either 0, 0.25, 0.5 or 0.75 times the infectiousness of clinical cases (**Fig 3d)**. We found that school closures decreased peak incidence slightly for influenza-like infections, and delayed the peak substantially. For COVID-19 epidemics, the delay and decrease of the peak was smaller, and this was especially the case in Bulawayo, which has the highest proportion of children **(Fig 3e)**. Because children exhibit more subclinical cases for COVID-19, school closures were more effective at reducing transmission of COVID-19 when the subclinical infectiousness was assumed to be higher **(Fig 3f)**.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

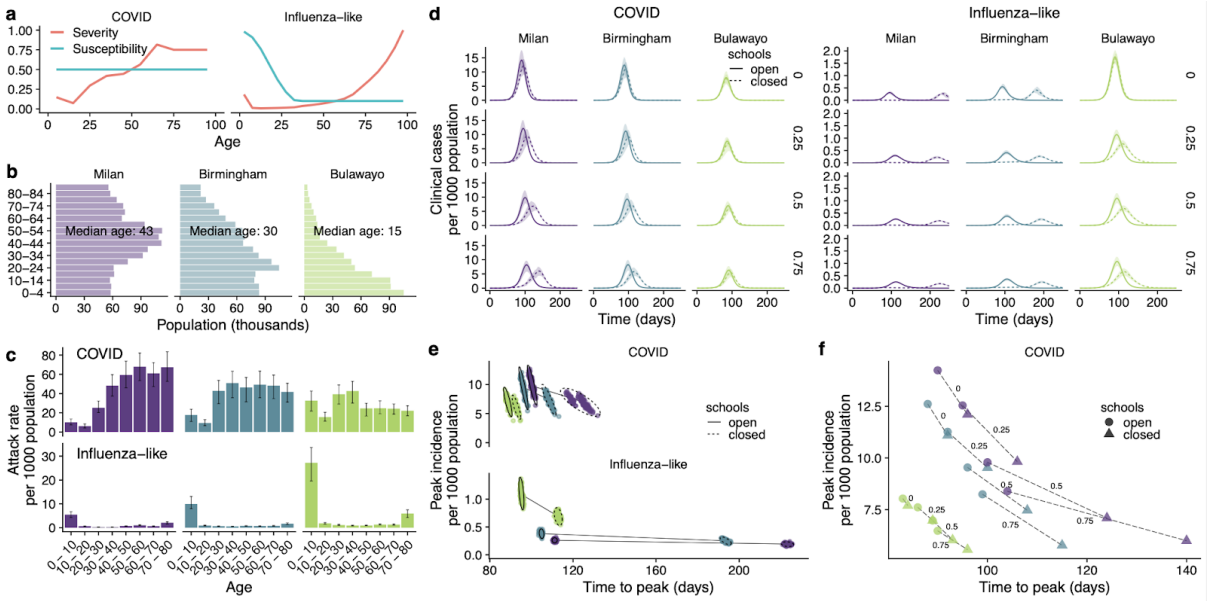

Fig. 3. Effect of school closure under different demographics and subclinical infectiousness. **(a)** Age dependence in clinical fraction (severity) and susceptibility to infection on contact for COVID, and for the schematic influenza-like scenarios (simplified, based on[7]) considered here. **(b)** Age structure for the 3 exemplar cities. **(c)** Age-specific attack rate for COVID-19 and influenza-like infections, assuming 50% subclinical infectiousness. **(d)** Daily incidence of clinical cases in exemplar cities for COVID-19 versus influenza-like infections. R_0 is fixed at 2.4. The rows show the impact of varying the infectiousness of subclinical infections to be 0%, 25%, 50%, or 75% as infectious as clinical cases while keeping R_0 fixed. **(e)** Change in peak timing and peak cases for the three cities, for either COVID-19 or pandemic influenza. **(f)** Change in median COVID-19 peak timing and peak cases for the three cities,depending on the infectiousness of subclinical infections.

Implications for global preparedness

Strong age dependence in the fraction of COVID-19 infections that become clinical cases has implications for the projected global burden. Simulating COVID-19 outbreaks in 146 capital cities, we found that the total expected number of clinical cases in an unmitigated epidemic varied between countries depending on the median age of the population, which is a proxy for the age structure of the population. The total clinical attack rate was higher, and the peak height of the epidemic was greater, in older populations (**Fig 4a**). By contrast, the number and peak of subclinical infections was lower in older populations (**Fig 4b**). The mean estimated basic reproduction number, R_0, was higher in cities with a lower median age (**Fig 4c**), because of the greater proportion of children and the higher number of contacts made by children compared to adults. We applied the same age-dependent clinical fraction to all

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

countries, but the relationship between age and clinical symptoms may be different in different countries, perhaps because of the age distribution of comorbidities, or the presence of other possible comorbidities, such as HIV[33]. If the relationship between clinical fraction and age skews younger in low and lower-middle income countries, there would be higher clinical attack rates in these countries (Supplementary Section 4).

The expected age distribution of cases shifted substantially over time, where in the early phase of the epidemic, the clinical case distribution tended to be skewed to younger ages, and the late phase showed more cases in older individuals (**Fig 4d**). This impacts projections for likely healthcare burdens at different phases of the epidemic, particularly because older individuals tend to have higher healthcare utilisation on infection[1] (**Fig 4e**).

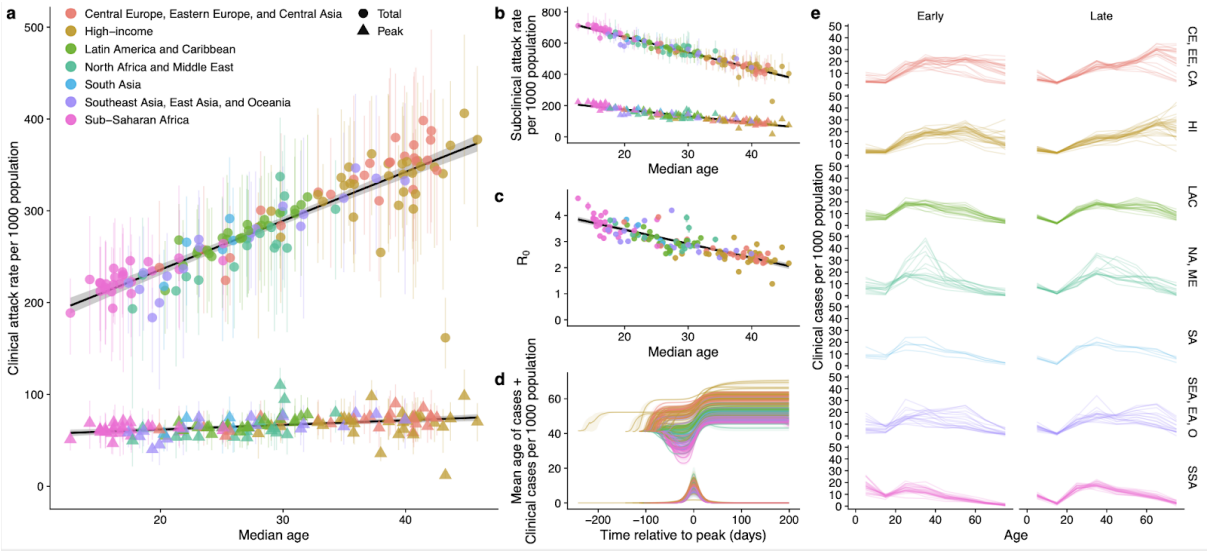

Fig. 4. Implications for global preparedness. (a) Expected clinical case attack rate (mean and 95% HDI), and peak in clinical case incidence for 146 countries in the Global Burden of Disease (GBD) country groupings[34] for an unmitigated epidemic. (b) Expected sub clinical case attack rate, and peak in subclinical cases. (c) Estimated basic reproduction number (R_0) in the capital city of each country assuming age-specific clinical fraction shown in Fig. 2b and 50% infectiousness of subclinically infected people. (d) Mean age of clinical cases on each day of an unmitigated epidemic in the 146 countries, and the epidemic curve for those epidemics. The epidemics are aligned at the peak, and colours mark the GBD groupings in a. (e) Age distribution of the first third and last third of clinical cases for 146 countries in GBD country groupings.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Conclusions

We have shown a strong age dependence in the probability of displaying clinical symptoms for COVID-19, from around 20% in under 10s, to over 70% in older adults. Given evidence of a stark age gradient in severity[8] and mortality[21,28], and recent studies of close follow up of children at risk of infection[16] showing that infection was frequent in all age groups, the plausibility of age-specific severity is higher than age-specific susceptibility to infection. For a number of other pathogens, there is evidence that children (except for the very youngest) have lower rates of symptomatic disease[12] and mortality[26]. For these reasons, we find that age-specific clinical fraction is more supported than age-specific susceptibility to infection. Serological surveys will provide critical information on the true distribution of subclinical infections.

The age-specific distribution of subclinical infection we have found is similar in shape (but larger in scale) to that generally assumed for pandemic influenza. However, for the 2009 influenza A/H1N1p pandemic, the age-specific susceptibility to infection was lower in older individuals compared to COVID-19. These differences have a large effect on how effective school closures may be in limiting transmission, delaying the peak of expected cases, and decreasing the total and peak number of cases. For COVID-19, school closures are likely to be much less effective than for influenza-like infections where children play a more substantial role in transmission.

It is critical to determine how infectious subclinical infections are compared to clinical infections in order to properly assess predicted burdens both with and without interventions. It is biologically plausible that milder cases are less transmissible, for example, because of an absence of cough[28,29], but direct evidence is limited[35]. If those with subclinical infection are similarly efficient transmitters of infection compared to those with clinical infections, the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

overall burden in the groups at risk of severe disease, primarily older populations for COVID-19[21], remains high. If those with subclinical infections are (much) less efficient at transmitting, then the overall contribution to the burden of clinical disease in the population may be proportionally lower. At the same time, lower relative infectiousness would reduce the impact of interventions targeting younger ages, such as school closure. By analysing epidemic dynamics before and after school closures, it may be possible to estimate the infectiousness of subclinical infections, however this will rely on granular data by age and time.

A great deal of concern has been directed toward the expected burden of COVID-19 in low and middle income countries (LMIC), which have lower population median age than many high income countries. Our results show that these demographic differences, coupled with a lower symptomatic fraction in younger ages, can result in proportionally fewer clinical cases than would be expected in higher-income countries with flatter demographic pyramids. This should not be interpreted as few cases in LMIC, because the projected epidemics are still very large, resulting in high numbers infected. Moreover, the particular relationship found with age here is drawn from high income countries, primarily in East Asia, and may reflect not only age, but also the increasing frequency of comorbidities with age. This relationship, therefore, may differ in LMIC for two key reasons: first, the distribution of non-communicable comorbid conditions—which are already known to increase the risk of severe disease from COVID-19[21]may be differently distributed by age, often occurring in younger age groups[34], along with other possible risk factors such as undernutrition[36]; and second, communicable comorbidities such as HIV[33], TB coinfection (which has been suggested to increase risk[37]), and others[38] may alter the distribution of severe outcomes by age. Observed severity and burden in LMIC may also be higher due to a lack of health system capacity for intensive treatment of severe cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

There are some limitations to the study. The true explanation for the age distribution could be a combination of age-specific susceptibility and clinical fraction, although some recent studies indicate children are infected at similar[24], or slightly lower rates[39] than adults, and children are not commonly spared from other coronavirus infections[40,41]. It is not possible to simultaneously estimate both effects from available data, so we were unable to validate a mixture model. While information drawn from the early stages of the epidemic are subject to uncertainty, age-specific information is drawn from several regions and countries, and clinical studies support the hypothesis presented here. We assumed that clinical cases are reported at a fixed fraction throughout the time period, although there may have been changes in reporting. We assumed that subclinical infections were less infectious than clinical infections but were not able to estimate how infectious subclinical infections were, instead testing the sensitivity of our findings to this parameter. We have used mixing matrices from the same country, but not the same location as the fitted data. We used contact matrices that combined physical and conversational contacts. We therefore implicitly assume that they are a good reflection of contact relevant for the transmission of SARS-CoV-2. If fomite, or faecal-oral routes of transmission are important in transmission, these contact matrices may not be representative of transmission risk.

The role of age in transmission is critical to designing interventions aiming to decrease transmission in the population as a whole, and to projecting the expected global burden. Early evidence[24], including presented here, suggests that there is age dependence in the risk of clinical symptoms following infection. Understanding if and by how much subclinical infections contribute to transmission has implications for predicted global burden and the impact of control interventions. This question must be resolved to effectively forecast and control COVID-19 epidemics.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Methods

Transmission model structure used in all analyses

We use an age-structured deterministic compartmental model (Fig. 1a, main text) stratified into 5-year age bands, with time approximated in discrete steps of 0.25 days. We assume that people are initially susceptible (S), and become exposed (E) after effective contact with an infectious person. After an incubation period, exposed individuals either develop a clinical or subclinical infection; an exposed age-*i* individual develops a clinical infection with probability y_i, otherwise developing a subclinical infection. Clinical cases are preceded by a preclinical but infectious (I_P) state; from the preclinical state, individuals develop full symptoms and become clinically infected (I_C). Based on evidence for other respiratory infections[22] we assume that subclinical infections (I_S) are less infectious compared to preclinical and clinical infections, and that subclinical individuals remain in the community until they recover. We use 50% as a baseline for the relative infectiousness of individuals in the subclinical state, and test the impact of other values (Supplementary section 3). Isolated and recovered individuals eventually enter the removed state (R); we assume these individuals are no longer infectious and are immune to reinfection.

The length of time an individual spends in states *E*, I_P, I_C, or I_S is distributed according to distributions d_E, d_P, d_C, or d_S, respectively (Table 1). The force of infection for an individual in age group *i* at time *t* is

$$\lambda_{i,t} = u_i \sum_j c_{ij,t}(I_{Pj} + I_{Cj} + fI_{Sj})/N_j ,$$

where u_i is the susceptibility to infection of an age-*i* individual, $c_{ij,t}$ is the number of age-*j* individuals contacted by an age-*i* individual per day at time *t*, *f* is the relative infectiousness of a subclinical case, and $(I_{Pj} + I_{Cj} + fI_{Sj})/N_j$ is the effective probability that a random age-*j*

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

individual is infectious. Contacts vary over time *t* depending upon the modelled impact of school closures and movement restrictions (see below).

To calculate the basic reproductive number, R_0, we define the next generation matrix as

$$NGM_{ij} = u_i c_{ij,t}(y_j E(d_P + d_C) + (1 - y_j) f E(d_S)) \quad .$$

R_0 is the absolute value of the dominant eigenvalue of the next generation matrix.

We use the local age distribution for each city or region being modelled, and synthetic or measured contact matrices for mixing between age groups (Table 2). The mixing matrices have four types of contacts: home, school, work and other contacts.

Parameter	Description	Applies in fits	Value	Reference
d_E	Incubation period (E to I_P and E to I_S; days)	All	$\sim gamma(\mu = 4.0, k = 5)$	Derived from[42]
d_P	Duration of preclinical infectiousness (days)	All	$\sim gamma(\mu = 2.4, k = 5)$	Derived from[42]
d_C	Duration of clinical infectiousness (I_C to R; days)	All	$\sim gamma(\mu = 3.2, k = 3.7)$	43
d_S	Duration of subclinical infectiousness (days)	All	$\sim gamma(\mu = 7, k = 5)$	Assumed
u_i	Susceptibility for age group *i*	Varies by age in Wuhan hypothesis 2, otherwise all ages equal	Estimated	
y_i	Probability of clinical infection for age group *i*	Varies by age in Wuhan hypothesis 3, otherwise all ages equal	Either fixed (50%) or estimated	44
f	Relative infectiousness of subclinical cases	All	50% (25% and 75% in sensitivity analysis)	Assumed
c_{ij}	Number of age-*j* individuals contacted by an age-*i* individual per day	All	Country-specific contact matrix (sensitivity analysis using synthetic matrices[19])	China[32]; UK[7]; Zimbabwe[34]
N_i	Number of age-*i* individuals	All	Demographic data	14
Δt	Time step for discrete-time simulation	All	0.25 days	
A_{min}, A_{max}	Age range of seed cases	Wuhan	Estimated	

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

t_{seed}	Day upon which seeding of infections starts	All	Estimated	
q_H	Relative change in non-school contacts during lunar new year holidays	Wuhan	Estimated	
q_L	Relative change in non-school contacts following large-scale restrictions	Wuhan, South Korea, Shanghai, Beijing, Italy	Estimated	
t_L	Day upon which large-scale restrictions start	Wuhan, South Korea, Shanghai, Beijing, Italy	Fixed to January 23 for Wuhan; estimated for other settings	

Table 1. Model parameters.

Location	**Mixing matrix details**
Wuhan City, China	We used mixing matrices measured in Shanghai in 2017/2018[45], adapted to the demographics of Wuhan prefecture. This implicitly assumes that Shanghai mixing patterns are representative of large cities in China.
Regions of China: Anhui, Guangdong, Guangxi, Hubei, Hunan, Jiangsu, Jiangxi, Jilin Shaanxi, Shandong, Sichuan, Tianjin, Zheijiang provinces; Beijing, Shanghai.	We used mixing matrices measured in Shanghai in 2017/2018[45], adapted to the demographics of each province / city.
Regions of Italy: Lombardia, Piemonte, Trento Veneto, Friulli Venezia Giulia, Liguria, Emilia-Romagna, Toscana, Marche, Lazio, Campania, Puglia regions; Milan.	We used mixing matrices measured in Italy in 2005/2006[11], adapted to the demographics of each region / city. This assumes that these contact patterns will still be representative of contact patterns in 2020.
Ontario, Canada	We used synthetic contact matrices, generated based on demographic information about the country[27].
Japan	We used synthetic contact matrices, generated based on demographic information about the country[27].
Singapore	We used synthetic contact matrices based on demographic information about the country[27].
South Korea	We used synthetic contact matrices based on demographic information about the country[27].
Birmingham, UK	We used mixing matrices measured in the UK in 2005/2006[11], adapted to the demographics of Birmingham. This assumes that these contact patterns will still be representative of contact patterns in 2020.
Bulawayo, Zimbabwe	We used mixing matrices measured in Manicaland, Zimbabwe in 2013[46], adapted to the demographics of Bulawayo. This implicitly assumes that Manicaland mixing patterns are representative of Bulawayo.
150 capital cities	We used synthetic contact matrices, generated based on demographic information about each country[27].

Table 2. Details on mixing matrices used in the study.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Comparing age hypotheses by fitting to the Wuhan epidemic

We contrasted three possible hypotheses. In hypothesis 1, there were no age-related differences in susceptibility ($u_i = u$) or symptomatic fraction ($y_i = y$). In hypothesis 2, susceptibility varied by age ($u_i = u(i)$), but the proportion of exposed individuals who became clinical cases did not vary by age ($y_i = y$). In hypothesis 3, the clinical case probability varied by age ($y_i = y(i)$), but susceptibility did not ($u_i = u$). Susceptibility and clinical fraction curves were fitted using three control points for young, middle, and old age, interpolating between them with a half-cosine curve (see Methods for details).

We assumed that the initial outbreak in Wuhan was seeded by introducing one exposed individual per day of a randomly drawn age between A_{min} *and* A_{max} for 14 days starting on a day (t_{seed}) in November[30,31]. We used the age distribution of Wuhan City prefecture in 2016[47] and contact matrices measured in Shanghai[32] as a proxy for large cities in China. This contact matrix is stratified into school, home, work, and other contacts. We aggregated the last three categories into non-school contacts and estimated how components of the contact matrix changed early in the epidemic in response to major changes. Schools closed on January 12th for the Lunar New Year holiday, so we decreased school contacts, but the holiday period may have changed non-school contacts, so we estimate this effect by inferring the change in non-school contact types, q_H . Large-scale restrictions started on January 23rd 2020 giving restrictions on travel and movement imposed by authorities, and we inferred the change in contact patterns during this period, q_L . Specifically:

$$c_{ij,t} = school(t) \cdot c_{ij|school} + other(t) \cdot c_{ij|other} \ ,$$

where

$$school(t) = \begin{cases} 1 & t < 12 \text{ January} \\ 0 & t \geq 12 \text{ January} \end{cases}$$

and

$$other(t) = \begin{cases} 1 & t < 12 \text{ January} \\ q_H & 12 \text{ January} \leq t < 23 \text{ January} \\ q_L & t \geq 23 \text{ January.} \end{cases}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

We fitted the model to incident confirmed cases from the early phase of the epidemic in China (December 8, 2019-February 1, 2020) reported by China CDC[1]. During this period, the majority of cases were from Wuhan City, and we truncated the data after February 1st because there were more cases in other cities after this time. We jointly fitted the model to the age distribution of cases at 3 time windows (December 8, 2019 to January 22, 2020) reported by Li et al.[26] and a further time window (December 8, 2019 to February 11, 2020) reported by China CDC[1]. Because there was a large spike of incident cases reported on February 1 determined to have originated from the previous week, we amalgamated all cases from January 25 to February 1, including those in the large spike, into a single data point for the week. We assumed 10% of clinical cases were reported[19]. We used a Dirichlet distribution with a flat prior to obtain 95% HDIs for reported case data stratified by age group for display in figures.

We used Markov-chain Monte Carlo to jointly fit each hypothesis to the two sets of empirical observations from the epidemic in Wuhan City, China. We used a negative binomial likelihood for incident cases and a Dirichlet-multinomial likelihood for the age distribution of cases, using the likelihood

$$L = \left(\prod_{k=1}^{K} NegBinom(C_k | size = 200, mean = c_k)\right)\left(\prod_{m=1}^{M} DirMultinom(A_m | \frac{200}{\|a_m\|} a_m)\right)$$

Above, C_k is the observed incidence on day k while c_k is the model-predicted incidence for day k, for each of K days. A_m is the observed age distribution for time period m (case counts for each age group) while a_m is the model-predicted age distribution for the same period, and $\|a_m\|$ is the total number of cases over all age groups in time period m, measured for M time periods. We set the precision of each distribution to 200 to capture additional uncertainty in data points that would not be captured with a Poisson or multinomial likelihood model.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

We distinguished fitted models using Deviance Information Criterion (DIC)[48].

Parameter	**Description**	**Prior**
u_i	Susceptibility to infection upon contact with an infectious person	Non-age-varying: $u_i \sim normal(\mu = 0.1, \sigma = 0.025, min = 0)$ Age-varying: young, middle, and old age fit as $a_y \sim normal(\mu = 15, \sigma = 15, min = 0, max = 30)$ $a_m \sim normal(\mu = 45, \sigma = 15, min = 30, max = 60)$ $a_o \sim normal(\mu = 75, \sigma = 15, min = 60, max = 90)$ Susceptibility for young, middle, and old age fit as $u_y \sim normal(\mu = 0.1, \sigma = 0.025, min = 0)$ $u_m \sim normal(\mu = 0.1, \sigma = 0.025, min = 0)$ $u_o \sim normal(\mu = 0.1, \sigma = 0.025, min = 0)$ Then $u_i = coss(i \vert a_y, b_y, a_m, b_m, a_o, b_o)$ (see final row)
y_i	Clinical fraction on infection	Non-age-varying: $y_i = 0.5$ Age-varying: young, middle, and old age fit as $a_y \sim normal(\mu = 15, \sigma = 15, min = 0, max = 30)$ $a_m \sim normal(\mu = 45, \sigma = 15, min = 30, max = 60)$ $a_o \sim normal(\mu = 75, \sigma = 15, min = 60, max = 90)$ Susceptibility for young, middle, and old age fit as $y_y \sim normal(\mu = 0.5, \sigma = 0.1, min = 0, max = 0.5)$ $y_m = 0.5$ $y_o \sim normal(\mu = 0.5, \sigma = 0.1, min = 0.5,\ max = 1)$ Then $y_i = coss(i \vert a_y, y_y, a_m, y_m, a_o, y_o)$ (see below)
t_{seed}	Timing of introduction of cases	$t_{seed} \sim normal(\mu = 15, \sigma = 30,\ min = 0,\ max = 30)$
q_H	Multiplicative factor for transmission during holiday period	$q_H \sim beta(\alpha = 2, \beta = 2)\ scaled\ to\ 0 - 2$
q_L	Multiplicative factor for transmission during large-scale restrictions	$q_L \sim beta(\alpha = 2, \beta = 2)$
A_{min}, A_{max}	Age bounds for introduced cases	$A \sim normal(\mu = 60, \sigma = 20, min = 40, max = 80)$ $A_{range} \sim beta(\alpha = 2, \beta = 2)\ scaled\ to\ 0 - 10$ $A_{min} = A - A_{range}$ $A_{max} = A + A_{range}$
$coss(a \vert x_1, y_1, x_2, y_2, x_3, y_3)$	Cosine-smoothing function	For a given age *a* (the midpoint age of age group *i*) the function evaluates to y_1 for $a \leq x_1$, to y_2 for $a = x_2$, and to y_3 for $a \geq x_3$. Values of a between x_1 and x_2 are interpolated between y_1 and y_2, and values of a between x_2 and x_3 are interpolated between y_2 and y_3, where the interpolation takes the shape of a cosine curve between $-\pi$ and π.

Table 2. Details of model fitting

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

	Hypothesis 1	**Hypothesis 2**	**Hypothesis 3**
DIC	880	691	558
Age control points	N/A	Young 13 (3.7-24) Middle 43 (32-55) Old 70 (61-79)	Young 14 (11-18) Middle 55 (49-60) Old 64 (60-68)
u_i	0.077 (0.073-0.08)	Young 0.0076 (0.00081-0.014) Middle 0.062 (0.045-0.077) Old 0.11 (0.1-0.12)	0.078 (0.074-0.083)
y_i	Fixed at 0.5	Fixed at 0.5	Young 0.056 (0.0084-0.11) Middle 0.49 (0.48-0.51) Old 0.74 (0.64-0.89)
t_{seed} , days after Nov 1	19 (17-21)	18 (15-21)	16 (14-20)
$A_{mid\|seed}$	74 (63-80)	60 (34-75)	48 (31-70)
$A_{range\|seed}$	4.5 (0.8-8.1)	0.78 (0.089-1.5)	2.9 (0.84-4.4)
q_H	1.4 (1.3-1.6)	1.4 (1.3-1.5)	1.4 (1.3-1.5)
q_L	0.14 (0.046-0.23)	0.21 (0.091-0.34)	0.2 (0.098-0.33)

Table 3. Posterior estimates and 95% highest density intervals for parameters fitted to data from Wuhan, China.

Analysis of the stationary age distribution of cases

To infer age-specific susceptibility from reported case distributions, we assumed that reported cases follow the stationary distribution of cases reached in the early phase of an epidemic. Using our dynamic model would allow modelling any transient emphasis in the case distribution associated with the age of the individuals who seeded infection in a given region, but since the age of the true first cases is not generally known, we used the stationary distribution instead. Specifically, we used Bayesian inference to fit age-specific susceptibility to the reported case distribution by first generating the expected case distribution k_i from (1) the age-specific susceptibility y_i, (2) the measured or estimated contact matrix for the country, and (3) the age structure of the country or region. We then used the penalised likelihood

$$L = Multinom(c_i \,|k_i) \prod_{i=2}^{n} Normal(y_i - y_{i-1} | \mu = 0, \sigma = 0.25) \quad ,$$

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

where c_i is the observed case distribution, when fitting to data from a single country or region. Above, the product of normal PDFs constrains the age-specific susceptibilities to avoid overly sensitive responses to noisy data. When fitting to a combined set of regions and/or countries, we used the penalised likelihood

$$L = \prod_{j=1}^{m} DirMultinom(c_{i,j} | sk_{i,j})^{w_j} \prod_{i=2}^{n} Normal(y_i - y_{i-1} | \mu = 0, \sigma = 0.25)$$

across countries $j \in \{1, 2, ..., m\}$ with weights w_j such that $\prod_j w_j = 1$. We weighted[49] each of the 13 provinces of China in our data set by 1/13, each of the 12 regions of Italy by 1/12, the three reported case distributions from China CDC by 1/3, and data from South Korea, Singapore and Japan each by 1, then scaled all weights to multiply to 1.

We fitted to the following data sources. For provinces of China, we used age-specific case numbers reported by China CDC[1] as well as linelist data compiled by the Shanghai Observer[50]. For regions of Italy, we used age-specific case numbers reported by the Istitute Superiore di Sanità on March 13, 2020[51]. For South Korea, we used the linelist released by Kim et al. based on data from the Korea Centers for Disease Control and Prevention[16]. For Japan, we used the Open Covid Linelist[52,53]. For Singapore, we use data compiled from Singapore Ministry of Health data by Alex Koh[15].

To validate our line list analysis, we fitted the dynamic model to incidence data from Beijing, Shanghai, South Korea and Lombary, Italy. We fixed the reporting rate for Beijing, Shanghai, South Korea, and Lombardy to 20%. Beijing and Shanghai incidence data were given by case onset, so we assumed no delay between reported and true case onsets. Incidence data for South Korea were given by the date of confirmation only, so we assumed the reporting delay followed a gamma distribution with a 7-day mean. Incidence data for Italy were given separately for case onset and case confirmation, with only a subset of onset

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

dates available; accordingly, we fit the proportion of confirmed cases with onset dates and the delay from onset to confirmation. We adjusted the size parameter of the negative binomial distribution used to model case incidence to 10 to reflect greater variability among fewer data points for these countries than for Wuhan. Beijing and Shanghai were fitted jointly, with separate dates of introduction but the same fitted susceptibility, large-scale restriction date and large-scale restriction magnitude. South Korea and Italy were each fitted separately; we fitted a large-scale restriction date and magnitude for both South Korea and Italy.

For both the linelist fitting and validation, we assumed that schools were closed in China, but remained open in South Korea, Japan, Italy, Singapore, and Canada, as schools were open for the majority of the period covered by the data in the latter five countries.

Quantifying the impact of school closure

To determine the impact in other cities with different demographic profiles we used the inferred parameters from our linelist analysis to parameterise our transmission model for projections to other cities. We chose these to compare projections for a city with a high proportion of elderly individuals (Milan, Italy); a moderate-aged population (Birmingham, United Kingdom); and a city in a low-income country with a high proportion of young individuals (Bulawayo, Zimbabwe). For this analysis, we compared an outbreak of COVID-19, for which the burden and transmission is concentrated in relatively-older individuals, with an outbreak of pandemic influenza, for which the burden and transmission is concentrated in relatively-younger individuals. We assumed that immunity to influenza builds up over a person's lifetime, such that an individual's susceptibility to influenza infection plateaus at roughly age 35, and assumed that the severity of influenza infection is highest in the elderly and in children under 10 years old[7].

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

To model Milan, we used the age distribution of Milan in 2019[54] and a contact matrix measured in Italy in 2006[11]. To model Birmingham, we used the age distribution of Birmingham in 2018[55] and a contact matrix measured in the UK in 2006[11]. To model Bulawayo, we used the age distribution of Bulawayo Province in 2012[56] and a contact matrix measured in Manicaland, Zimbabwe in 2013[46]. We assumed that the epidemic was seeded by two infectious individuals in a random age group per week for 5 weeks to. We scaled the age-specific susceptibility u_i by setting the "target" basic reproductive number R_0 = 2.4 as an illustrative example. We also performed a sensitivity analysis where we scaled u_i to result in R_0 = 2 in Birmingham, and using the same setting for u_i in all three cities, so that the actual R_0 changed depending upon contact matrices and demographics used to model each city. This produced qualitatively similar results (Supplementary Information).

We projected the impact of school closure by setting the contact multiplier for school contacts *school(t)* to 0. Complete removal of school contacts may overestimate the impact of school closures because of alternative contacts children make when out of school[57]. This will however give the maximum impact of school closures in the model to demonstrate the differences.

Projecting the global impact

To project the impact of COVID-19 outbreaks in global cities, we used mixing matrices from Prem et al.[27] and demographic structures for 2020 from World Population Prospects 2019 to simulate a COVID-19 outbreak in 146 global capital cities for which synthetic matrices, demographic structures and total populations were available. For simplicity, we assumed that capital cities followed the demographic structure of their respective countries and took

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

the total population of each capital city from the R package *maps*. For each city, we scaled u_i to result in an average R_0 = 2 in Birmingham, UK, and used the same setting for u_i for all cities, so that the realised R_0 would change according to the contact matrices and demographics for each city. We simulated 20 outbreaks in each city, drawing the age-specific clinical fraction y_i from the posterior of the estimated overall clinical fraction from our line list analysis (Fig. 2), and analysed the time to the peak incidence of the epidemic, the peak clinical and subclinical incidence of infection, and the total number of clinical and subclinical infections. We took the first third and the last third of clinical cases in each city to compare the early and late stages of the epidemic.

Acknowledgements
We acknowledge the following for funding: NGD: National Institutes of Health Research (HPRU-2012-10096). PK, YL, KP, MJ: This research was partly funded by the Bill & Melinda Gates Foundation (INV-003174). YL, MJ: This research was partly funded by the National Institute for Health Research (NIHR) (16/137/109) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care. RME: HDR UK (grant: MR/S003975/1).

The members of the CMMID COVID-19 working group and the funding they acknowledge are: Carl A B Pearson, Billy J Quilty (NIHR 16/137/109), Adam J Kucharski (Wellcome Trust grant: 206250/Z/17/Z), Hamish Gibbs (funded by the Department of Health and Social Care using UK Aid funding and is managed by the NIHR. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and SocialCare (ITCRZ 03010), Samuel Clifford (Wellcome Trust grant: 208812/Z/17/Z), Amy Gimma (Global Challenges Research Fund (GCRF) for the project "RECAP" managed through RCUK and ESRC (ES/P010873/1), Kevin van Zandvoort (supported by Elrha's Research for Health in Humanitarian Crises (R2HC) Programme, which aims to improve health outcomes by strengthening the evidence base for public health interventions in humanitarian crises. The R2HC programme is funded by the UK Government (DFID), the Wellcome Trust, and the UK National Institute for Health Research (NIHR), James D Munday (Wellcome Trust grant: 210758/Z/18/Z), Charlie Diamond (NIHR 16/137/109), W John Edmunds, Joel Hellewell (Wellcome Trust grant: 210758/Z/18/Z), Timothy W Russel (Wellcome Trust grant: 206250/Z/17/Z), Sam Abbott (Wellcome Trust grant: 210758/Z/18/Z), Sebastian Funk (Wellcome Trust grant: 210758/Z/18/Z), Nikos I Bosse, Fiona Sun (NIHR EPIC grant 16/137/109), Stefan Flasche (Wellcome Trust grant: 208812/Z/17/Z), Alicia Rosello (NIHR grant: PR-OD-1017-20002), Christopher I Jarvis (Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1)), RMGJH (European Research Commission Starting Grant: #757699).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Author contributions

RME conceived the study. NGD and RME designed the model with PK, and YL, KP and MJ providing input. NGD designed the software and inference framework and implemented the model. YL processed the data. NGD and RME wrote the first draft of the manuscript. All authors interpreted the results, contributed to writing, and approved the final version for submission.

Data Availability and Code Availability

The data used for fitting are publicly available, but will also be made available with the code in the github repository for the project. Contact matrix data are available at zenodo[21,22].

Competing interests

The authors have no competing interests.

Additional information

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to Rosalind M Eggo or Nicholas G Davies at r.eggo@lshtm.ac.uk or nicholas.davies@lshtm.ac.uk

References

1. Zhang. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chin. J. Epidemiol.* **41**, 145–151 (20200217).
2. Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit. Health* S2589750020300261 (2020) doi:10.1016/S2589-7500(20)30026-1.
3. D, C. *et al.* The early phase of the COVID-19 outbreak in Lombardy, Italy. *ArXiv200309320 Q-Bio* (2020).
4. Shim, E., Tariq, A., Choi, W., Lee, Y. & Chowell, G. Transmission potential and severity of COVID-19 in South Korea. *Int. J. Infect. Dis.* S1201971220301508 (2020) doi:10.1016/j.ijid.2020.03.031.
5. Anderson, R. M. *et al.* Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 1091–1105 (2004).
6. Jhung, M. A. *et al.* Epidemiology of 2009 Pandemic Influenza A (H1N1) in the United

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

States. *Clin. Infect. Dis.* **52**, S13–S26 (2011).

7. Greer, A. L., Tuite, A. & Fisman, D. N. Age, influenza pandemics and disease dynamics. *Epidemiol. Infect.* **138**, 1542–1549 (2010).
8. Dong, Y. *et al.* Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China. *Pediatrics* e20200702 (2020) doi:10.1542/peds.2020-0702.
9. Zhao, X. *et al. Incidence, clinical characteristics and prognostic factor of patients with COVID-19: a systematic review and meta-analysis*. http://medrxiv.org/lookup/doi/10.1101/2020.03.17.20037572 (2020) doi:10.1101/2020.03.17.20037572.
10. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497–506 (2020).
11. Mossong, J. *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Med.* **5**, e74 (2008).
12. Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. & Ferguson, N. M. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–754 (2008).
13. Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E. & Edmunds, W. J. Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. *PLoS Comput. Biol.* **8**, e1002425 (2012).
14. World Population Prospects - Population Division - United Nations. https://population.un.org/wpp/.
15. Koh, A. Singapore COVID-19 Cases. http://alexkoh.net/covid19/.
16. Data Science for COVID-19 (DS4C). https://kaggle.com/kimjihoo/coronavirusdataset.
17. Cowling, B. J. *et al.* Increased Risk of Noninfluenza Respiratory Virus Infections Associated With Receipt of Inactivated Influenza Vaccine. *Clin. Infect. Dis.* **54**,

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

1778–1783 (2012).

18. Tsagarakis, N. J. *et al.* Age-related prevalence of common upper respiratory pathogens, based on the application of the FilmArray Respiratory panel in a tertiary hospital in Greece: *Medicine (Baltimore)* **97**, e10903 (2018).
19. Common cold. *nhs.uk* https://www.nhs.uk/conditions/common-cold/ (2017).
20. Galanti, M. *et al.* Rates of asymptomatic respiratory virus infection across age groups. *Epidemiol. Infect.* **147**, e176 (2019).
21. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* S0140673620305663 (2020) doi:10.1016/S0140-6736(20)30566-3.
22. Van Kerckhove, K., Hens, N., Edmunds, W. J. & Eames, K. T. D. The Impact of Illness on Social Networks: Implications for Transmission and Control of Influenza. *Am. J. Epidemiol.* **178**, 1655–1662 (2013).
23. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies Needed. *N. Engl. J. Med.* **0**, null (2020).
24. Bi, Q. *et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts*. http://medrxiv.org/lookup/doi/10.1101/2020.03.03.20028423 (2020) doi:10.1101/2020.03.03.20028423.
25. Yang, Y. *et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China*. http://medrxiv.org/lookup/doi/10.1101/2020.02.10.20021675 (2020) doi:10.1101/2020.02.10.20021675.
26. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* NEJMoa2001316 (2020) doi:10.1056/NEJMoa2001316.
27. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

contact surveys and demographic data. *PLOS Comput. Biol.* **13**, e1005697 (2017).

28. Onder, G., Rezza, G. & Brusaferro, S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA* (2020) doi:10.1001/jama.2020.4683.
29. Chan, K. P. Control of Severe Acute Respiratory Syndrome in Singapore. *Environ. Health Prev. Med.* **10**, 255–259 (2005).
30. Lau, J. T. F. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. *J. Epidemiol. Community Health* **57**, 864–870 (2003).
31. Cauchemez, S. *et al.* School closures during the 2009 influenza pandemic: national and local experiences. *BMC Infect. Dis.* **14**, 207 (2014).
32. Cauchemez, S. *et al.* Closure of schools during an influenza pandemic. *Lancet Infect. Dis.* **9**, 473–481 (2009).
33. Cohen, C. *et al.* Severe Influenza-associated Respiratory Infection in High HIV Prevalence Setting, South Africa, 2009–2011. *Emerg. Infect. Dis.* **19**, (2013).
34. IHME. Global Burden of Disease. http://www.healthdata.org/gbd.
35. Williams, C. M. *et al.* Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. *Lancet Infect. Dis.* S1473309919307078 (2020) doi:10.1016/S1473-3099(19)30707-8.
36. Murray, J. *et al.* Determining the Provincial and National Burden of Influenza-Associated Severe Acute Respiratory Illness in South Africa Using a Rapid Assessment Methodology. *PLOS ONE* **10**, e0132078 (2015).
37. Liu, Y. *et al. Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity*. http://medrxiv.org/lookup/doi/10.1101/2020.03.10.20033795 (2020) doi:10.1101/2020.03.10.20033795.
38. Cohen, A. L. *et al.* Potential Impact of Co-Infections and Co-Morbidities Prevalent in Africa on Influenza Severity and Frequency: A Systematic Review. *PLOS ONE* **10**, e0128580 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

39. Zhang, J. *et al. Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China*. http://medrxiv.org/lookup/doi/10.1101/2020.03.19.20039107 (2020) doi:10.1101/2020.03.19.20039107.

40. Jevšnik, M. *et al.* Coronavirus infections in hospitalized pediatric patients with acute respiratory tract disease. *BMC Infect. Dis.* **12**, 365 (2012).

41. Lee, J. & Storch, G. A. Characterization of Human Coronavirus OC43 and Human Coronavirus NL63 Infections Among Hospitalized Children <5 Years of Age: *Pediatr. Infect. Dis. J.* **33**, 814–820 (2014).

42. Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* **25**, (2020).

43. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* S1473309920301444 (2020) doi:10.1016/S1473-3099(20)30144-4.

44. Nishiura, H. *et al.* The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. *J. Clin. Med.* **9**, 419 (2020).

45. Zhang, J. *et al.* Patterns of human social contact and contact with animals in Shanghai, China. *Sci. Rep.* **9**, 1–11 (2019).

46. Melegaro, A. *et al.* Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *PLOS ONE* **12**, e0170459 (2017).

47. National Bureau of Statistics. China Statistical Year Book. (2005-2018). http://www.stats.gov.cn/tjsj/ndsj/.

48. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20043018; this version posted March 27, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

49. Varin, C., Reid, N. & Firth, D. An overview of composite likelihood methods. *Stat. Sin.* **21**, (2011).

50. Observer, S. Shanghai Observer. COVID-2019 Linelist. http://data.shobserver.com/www/datadetail.html?contId=1000895.

51. Epicentro. *Bolletino Sorveglianza Integrata COVID-19 12 Marzo 2020 Appendix*. https://www.epicentro.iss.it/coronavirus/bollettino/Bolletino-sorveglianza-integrata-COVID-19_12-marzo-2020_appendix.pdf.

52. COVID19_2020_open_line_list. https://docs.google.com/spreadsheets/d/1itaohdPiAeniCXNIntNztZ_oRvjh0HsGuJXUJWET008/edit?usp=sharing.

53. Xu, B. *et al.* Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect. Dis.* S1473309920301195 (2020) doi:10.1016/S1473-3099(20)30119-5.

54. Milano (Metropolitan City, Italy) - Population Statistics, Charts, Map and Location. http://citypopulation.info/en/italy/admin/lombardia/015__milano/.

55. Age breakdown of the population of Birmingham - Office for National Statistics. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/agebreakdownofthepopulationofbirmingham.

56. Bulawayo (City, Zimbabwe) - Population Statistics, Charts, Map and Location. http://citypopulation.info/php/zimbabwe-admin.php?adm1id=A.

57. Kucharski, A. J., Conlan, A. J. K. & Eames, K. T. D. School's Out: Seasonal Variation in the Movement Patterns of School Children. *PLOS ONE* **10**, e0128070 (2015).