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

Abstract

22 The COVID-19 pandemic has shown a markedly low proportion of cases among

23 children[1,23,4]. Age disparities in observed cases could be explained by children having lower

24 susceptibility to infection, lower propensity to show clinical symptoms, or both. We evaluate

25 these possibilities by fitting an age-structured mathematical model to epidemic data from six

26 countries. We estimate that clinical symptoms occur in 25% (95% CrI: 19–32%) of infections

27 in 10–19-year-olds, rising to 76% (68–82%) in over-70s, and that susceptibility to infection in

28 under-20s is approximately half that of older adults. Accordingly, we find that interventions

29 aimed at children may have a relatively small impact on total cases, particularly if the

30 transmissibility of subclinical infections is low. The age-specific clinical fraction and

31 susceptibility we have estimated has implications for the expected global burden of COVID-

32 19 because of demographic differences across settings: in younger populations, the

33 expected clinical attack rate would be lower, although it is likely that comorbidities in low-

34 income countries will affect disease severity. Without effective control measures, regions

35 with older populations may see disproportionally more clinical cases, particularly in the later

36 stages of the pandemic.

Main

39 COVID-19 shows an increased number of cases and risk of severe disease with increasing

40 age[5,6], a feature shared with the 2003 SARS epidemic [5]. Understanding the role of age in

41 transmission and disease severity is critical for determining the likely impact of social-

42 distancing interventions for decreasing transmission, especially those aimed at schools, and

43 for estimating the expected global disease burden.

45 The age gradients in reported cases, observed from the earliest stages of the pandemic,

46 could be generated by children having decreased susceptibility to infection, decreased

47   probability of showing disease on infection, or a combination of both. A summary of the main

48   findings, limitations and implications of the model for policy makers is shown in Table 1.

49

Background	The distribution of confirmed COVID-19 cases has shown strong age dependence, with notably few cases in children. This could be because younger ages are less susceptible to infection and/or are less prone to showing clinical symptoms when infected. We have used dynamic transmission models fitted to a range of available data on the age distribution of reported cases, and to studies that looked for subclinical infections amongst contacts, to estimate the age-specific susceptibility to SARS-CoV-2 infection, and the age-specific fraction of infections that develop clinical symptoms of COVID-19.
Main findings and limitations	We find that under-20s are roughly half as susceptible to infection as over-20s, and that 75% of infections are subclinical in 10–19-year olds, compared to 24% in 70+-year-olds. As with all modelling studies, further data generated during the epidemic could change our parameter estimates. Population mixing measured in contact surveys may not be representative of contact patterns made during the early phase of local epidemics. However, our estimates are consistent across countries and intervention contexts.
Policy implications	These results have implications for the likely effectiveness of school closures in mitigating SARS-CoV-2 transmission, in that these may be less effective than for other respiratory infections. They also have implications for the global expected burden of clinical cases; countries with a large number of children may need to account for decreased susceptibility and severity in burden projections.

50   Table 1. Policy Summary

51

52   Age-varying susceptibility to infection by SARS-CoV-2, where children may be less

53   susceptible to becoming infected on contact with an infectious person, would reduce cases

54   among children, and potentially lower transmission in the population overall. Decreased

55   susceptibility could result from immune cross-protection from other coronaviruses[8,910], or

56   possibly from non-specific protection resulting from recent infection by another respiratory

57   virus[11], which children experience more frequently than adults[12,13].

58

59   It is also possible that children may experience mild or no symptoms on infection more

60   frequently than adults. Such age-dependent variation in severity has been observed for other

61   respiratory virus infections[14], including SARS[14,15]. For COVID-19, there are indications of age

62   dependence in severity[8] and mortality[15] among reported cases[15], which could extend to

3

63   severity and likelihood of clinically reportable symptoms given infection. "Asymptomatic"

64   cases have no symptoms at all, and "paucisymptomatic" is sometimes used for those with

65   very mild symptoms that may not be noticed or reported, even though they occur. We call

66   these two types "subclinical", which are more likely to remain undetected than clinically

67   apparent cases. If subclinical infections exhibit age dependence there would be lower

68   reported cases among children, but children could still be capable of transmitting the virus to

69   others, potentially at lower rates than individuals exhibiting clinical infections, as has been

70   shown for influenza[16].

71

72   Contact patterns and demographics affect the expected number of cases in each age group.

73   Children tend to make more social contacts than adults[17] and hence, all else equal,

74   contribute more to transmission than adults[18,19]. This is why school closures are considered

75   a key intervention for epidemics of respiratory infections[8], but the impact of school closure

76   depends on the role of children in transmission.

77

78   The particular context of SARS-CoV-2 emergence in Wuhan, China, could have resulted in a

79   skewed age distribution because early cases were in older adults[20], and assortative mixing

80   between adults could have reduced transmission to children in the very early stages of the

81   outbreak, with subsequent closure of schools on 12th January 2020 for the Lunar New Year

82   holiday potentially reinforcing this effect. Outside of China, COVID-19 outbreaks may have

83   been initially seeded by working-age travellers entering the country[21,22], producing a similar

84   excess of older individuals in early phases of local epidemics.

85

86   Determining the role of children in transmission using available data has important

87   implications for policies that aim to control transmission[23], especially through interrupting

88   child-driven transmission. Additionally, if the number of infections or cases depends strongly

89   on the role of children, countries with different age distributions could exhibit substantially

90   different epidemic profiles and overall impact of COVID-19 epidemics.

91

92   We used an age-stratified transmission model with heterogeneous contact rates between

93   age groups to examine varying susceptibility to infection by age; varying clinical fraction by

94   age; and no age variation in susceptibility or clinical fraction (see Methods). We generated

95   model variants (**Fig 1a**) and fitted each to data sources from the epidemic in Wuhan: a time

96   series of reported cases[1] and four snapshots of the age distribution of cases[1][2][4] (**Fig 1;**

97   **Extended Data Figure 1**). We included the observed school closures, which decreased the

98   school contacts of children in the model. We also estimated the effect of the Lunar New Year

99   holiday period, and the travel and movement restrictions in Wuhan, on transmission (**Fig**

100   **1d**). We found that under each hypothesis, the basic reproduction number R_0 was 2.5-2.8

101   initially, was inflated 1.2–1.4-fold during the pre Lunar New Year holiday period, and then fell

102   by 60–70% during restrictions in Wuhan (**Fig 1e**).

103

104   All model variants fitted the daily incident number of confirmed cases equally well (**Fig 1f**),

105   but the model without age-varying susceptibility or clinical fraction could not reproduce the

106   observed age distribution of cases, overestimating the number of cases in children and

107   underestimating cases in older adults (**Fig 1g**). The other two fitted the observed age

108   distribution of cases, but the model assuming no age variation in clinical fraction implied a

109   large number of mild or asymptomatic infections among the elderly **(Fig 1h)**. Comparison

110   using Deviance Information Criterion[6] (DIC) showed that age varying susceptibility (DIC:

111   697) and age-varying clinical fraction (DIC: 663) were preferred over the model with neither

112   (DIC: 976).

113

114   In the model with age-varying susceptibility, 20% of infections occurred in over-70s, with half

115   of these as clinical cases, and the other half as subclinical infections (**Fig 1h**). In the model

116   with age-varying clinical fraction, 20% of infections occurred in over-70s, but less than a

117   quarter of these were subclinical. Recent work has demonstrated an age-dependent severity

118   in hospitalised confirmed cases[25,26], which suggests that subclinical infection in older adults

119   may be rare and supports the clinical fraction increasing with age.

120

121
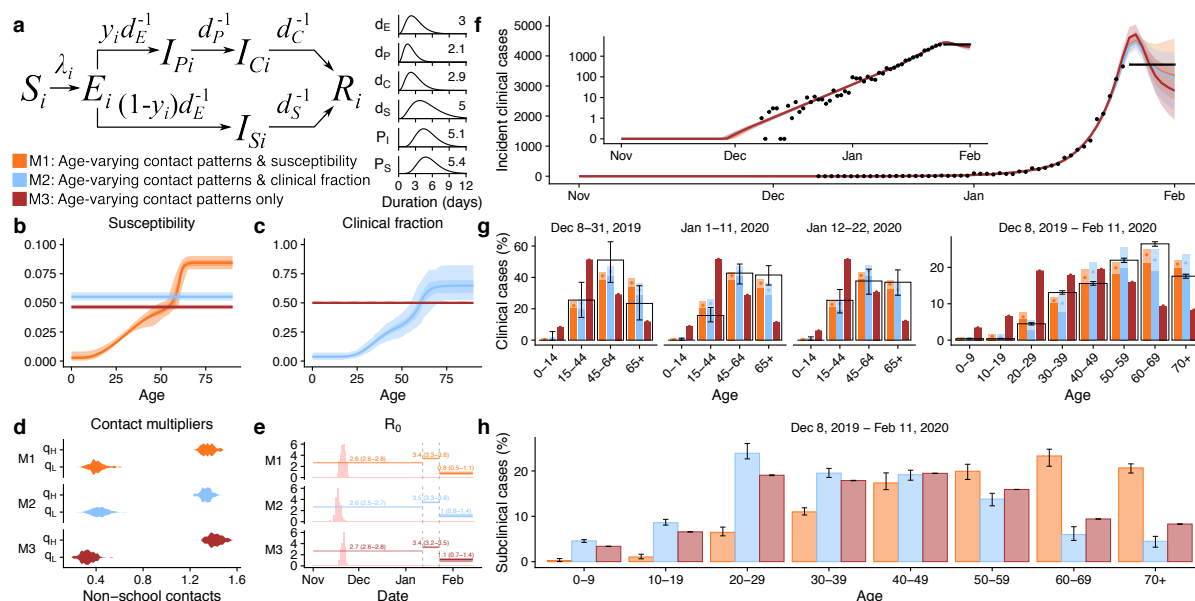
Fig. 1. Comparing the fit of different model variants to data from Wuhan City, China. (a) Model diagram showing duration of disease states in days, where *d* parameters represent the duration of time in each disease state, y_i is the fraction of infections that are clinical in age group *i*, λ_i is the force of infection in age group *i*, P_I is the incubation period and P_S is the serial interval (see Methods). **(b)** Susceptibility by age for the three models. Age-specific values were estimated for model 1 (orange). Susceptibility is defined as the probability of infection on contact with an infectious person. **(c)** Clinical fraction (y_i) by age for the three models. Age-specific values were estimated for model 2 (blue), and fixed at 0.5 for models 1 and 3. **(d)** Fitted contact multipliers for holiday (q_H) and restricted periods (q_L) for each model 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 for each model. The red barplot shows the inferred window of spillover of infection. **(f)** Incident reported cases (black), and modelled incidence of reported clinical cases for the three models fitted to cases reported by China Centers for Disease Control (CCDC)[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[24] (first three panels) and CCDC[1] (fourth panel). Data are shown in the hollow bars, and model predictions in filled bars, where the dot marks the mean posterior estimate. **(h)** Implied distribution of subclinical cases by age for each model. Credible intervals on modelled values show the 95% HDIs; credible intervals on data for panels g and h show 95% HDIs for the proportion of cases in each age group.

142

143

144

145   It is possible that both age-varying susceptibility and age-varying clinical fraction contribute

146   to some degree to the observed age patterns. To investigate this possibility, we fitted a

147   model in which both susceptibility and clinical fraction varied by age, estimating these

148   parameters across 32 settings in six countries. We fitted the stationary distribution of the

149   next generation matrix to reproduce the locally-reported age distribution of cases compiled

150   from a variety of sources **(Fig 2a)** and jointly fitted data from five recent studies giving

151   information on infection rates and symptom severity across ages[25,27–30] (**Extended Data**

152   **Figure 2**). We used setting-specific demographics, measured contact matrices where

153   possible, and synthetic contact matrices otherwise (see Methods)[31]. The age-dependent

154   clinical proportion was markedly lower in younger age groups in all regions **(Fig 2b),** with

155   25% (19–32%) of infections in 10–19-year-olds resulting in clinical cases, rising to 76% (68–

156   82%) in adults over 70 in the consensus age distribution estimated across all regions; the

157   age-specific susceptibility profile suggested that under-20s were half as susceptible to

158   SARS-CoV-2 infection as over-20s (**Extended Data Table 1**). To determine whether this

159   distribution was capable of reproducing epidemic dynamics, we fitted our dynamic model to

160   the incidence of clinical cases in Beijing, Shanghai, South Korea and Italy (**Fig 2c;**

161   **Extended Data Figure 3**). The consensus age-specific clinical fraction was largely capable

162   of reproducing the age distribution of cases, although there are some outliers, for example in

163   the 20–29 age group in South Korea. This could be the result of clustered transmission

164   within a church group in this country[4]. The predicted age distribution of cases for Italy is also

165   less skewed towards older adults than reported cases show, suggesting potential differences

166   in age-specific testing in Italy[32]. Locally-estimated age-varying clinical fraction captured

167   these patterns more precisely **(Fig. 2c)**.
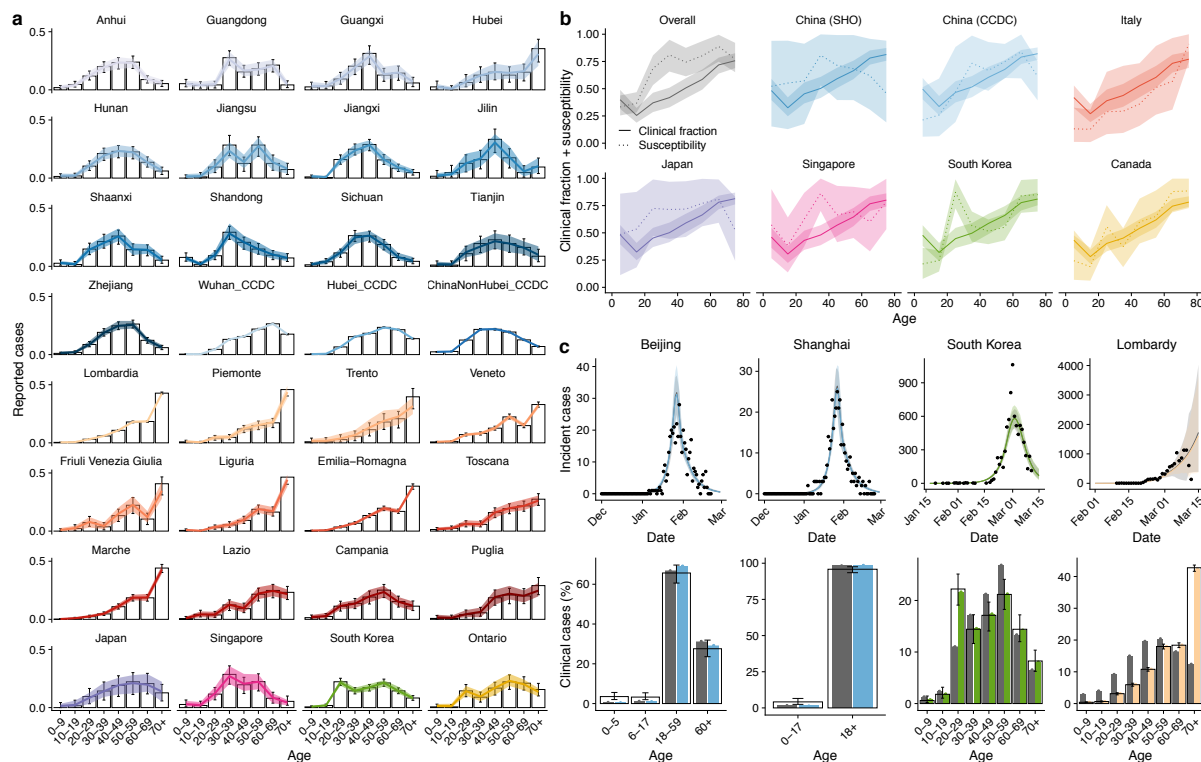
168

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 (solid lines) and age distribution of
susceptibility (dashed lines) for all countries. The overall consensus fit is shown in grey. **(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.

School closures during epidemics[33,34] and pandemics[35,36] aim to decrease transmission

amongst children[18], and 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,

the contacts they have with other age groups, their susceptibility to infection, and

infectiousness if infected. Using schematic values for pandemic influenza[37] 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),

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

(mean attack rate across three cities: 361 per 1000 for COVID-19 versus 23 per 1000 for

influenza), with relatively more cases occurring in under-20s (68%) in the influenza-like

190   scenario compared to COVID-19 (17%) **(Fig 3c)**. More clinical cases were in older adults in

191   Milan compared with the other cities, and a markedly younger age distribution was observed

192   for clinical cases in Bulawayo. The age distribution of clinical cases depends on the

193   demography and mixing in the region.

194

195   To explore the effect of school closure, we simulated 3 months of school closures with

196   varying infectiousness of subclinical infections, at either 0%, 50% or 100% the

197   infectiousness of clinical cases **(Fig 3d)**. For influenza- like infections we found that school

198   closures decreased peak incidence by 17–35% across settings, and delayed the peak by 10-

199   89 days across settings **(Fig 3e)**. For COVID-19 epidemics, the delay and decrease of the

200   peak was smaller (9–18% decrease in peak incidence, 1–6 day delay in peak timing),

201   especially in Bulawayo, which has the highest proportion of children **(Fig 3e)**. Because

202   children have lower susceptibility and exhibit more subclinical cases for COVID-19, school

203   closures were slightly more effective at reducing transmission of COVID-19 when the

204   subclinical infectiousness was assumed to be higher (school closures were 37–53% more

205   effective at reducing peak cases across settings for 100% versus 0% subclinical

206   infectiousness) **(Fig 3f)**.

207
208

209

210
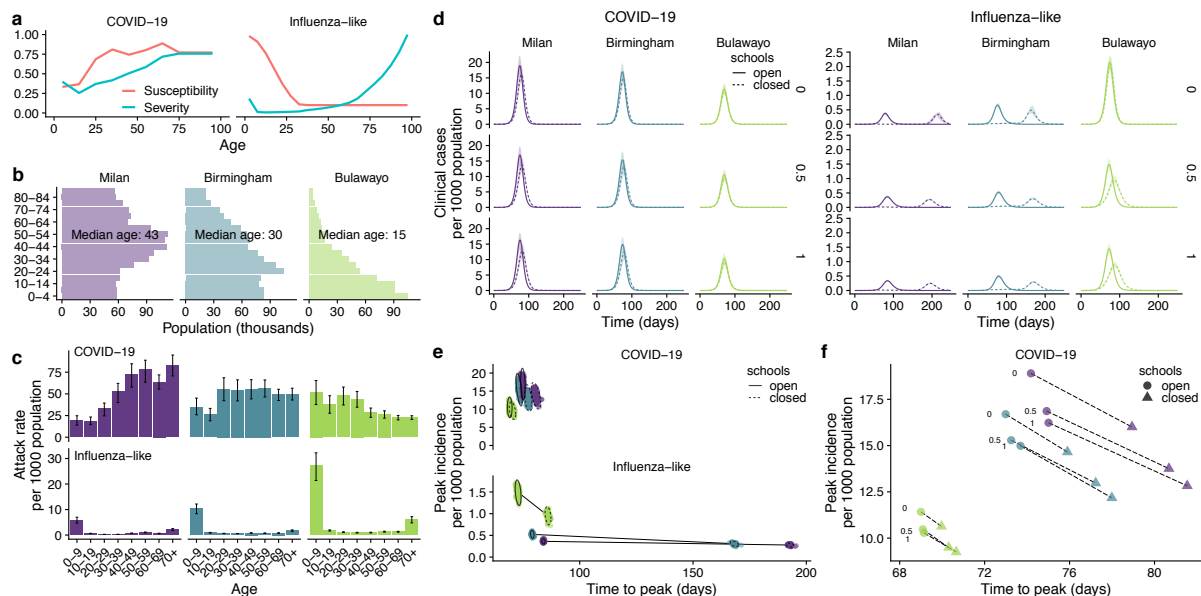
211 **Fig. 3. Effect of school closure under different demographics and subclinical infectiousness.**
212 **(a)** Age dependence in clinical fraction (severity) and susceptibility to infection on contact for COVID-
213 19, and for the influenza-like scenarios (simplified, based on ref. [37]) considered here. **(b)** Age
214 structure for the 3 exemplar cities. **(c)** Age-specific attack rate for COVID-19 and influenza-like
215 infections, assuming 50% subclinical infectiousness. **(d)** Daily incidence of clinical cases in exemplar
216 cities for COVID-19 versus influenza-like infections. R_0 is fixed at 2.4. The rows show the impact of
217 varying the infectiousness of subclinical infections to be 0%, 50%, or 100% as infectious as clinical
218 cases while keeping R_0 fixed. **(e)** Change in peak timing and peak cases for the three cities, for either
219 COVID-19 or influenza-like. **(f)** Change in median COVID-19 peak timing and peak cases for the three
220 cities, depending on the infectiousness of subclinical infections.

221

222

223 Age dependence in susceptibility and clinical fraction has implications for the projected

224 global burden of COVID-19. We simulated COVID-19 epidemics in 146 capital cities and

225 found that the total expected number of clinical cases in an unmitigated epidemic varied

226 between countries depending on the median age of the population, which is a proxy for the

227 age structure of the population. The mean estimated basic reproduction number, R_0, did not

228 substantially differ by median age (**Fig 4c**), because although there was a greater proportion

229 of children the susceptibility of children was lower. We applied the same age-dependent

230 clinical fraction to all countries, but the relationship between age and clinical symptoms may

231 be different in different countries, perhaps because of a different distribution of

232 comorbidities[38], or setting—specific comorbidities such as HIV[39]. If the relationship between

233   clinical fraction and age skews younger in low and lower-middle income countries, there

234   would be higher clinical attack rates in these countries (**Extended Data Figure 4**).

235

236   The expected age distribution of cases shifted substantially during the epidemic, where in

237   the early phase there were more cases in the central age group (20-59), and after the peak a

238   higher proportion on cases in younger and older ages (**Fig 4d**). The size of the shift was

239   higher in countries with higher median age, which impacts projections for likely healthcare

240   burdens at different phases of the epidemic (**Fig 4e**), particularly because older individuals

241   tend to have higher healthcare utilisation if infected[1].

242
243
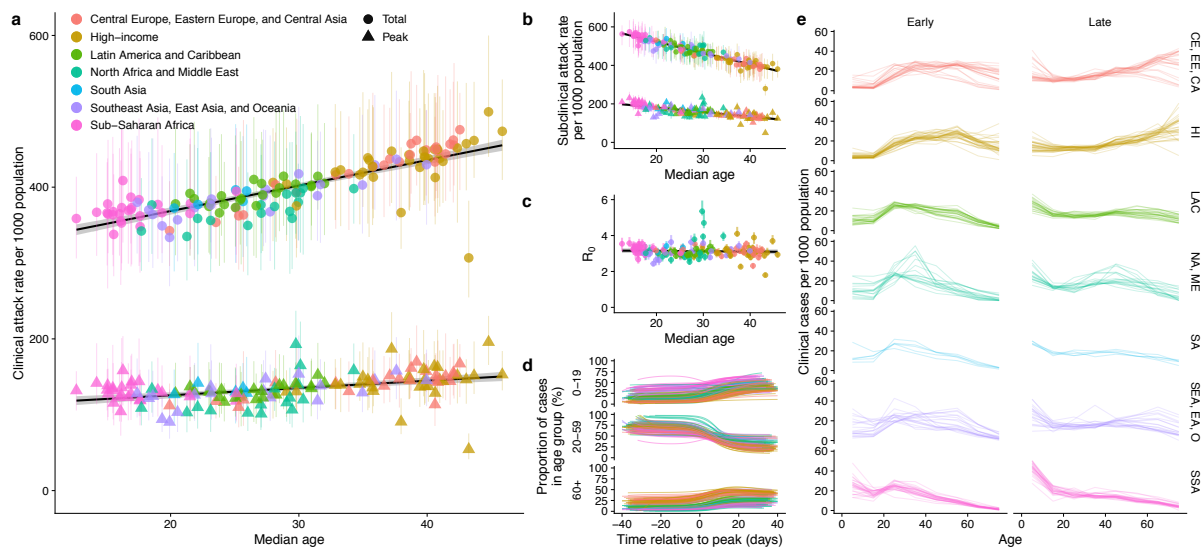
244
245   **Fig. 4. Implications for global preparedness. (a)** Expected clinical case attack rate (mean and 95%
246   HDI), and peak in clinical case incidence for 146 countries in the Global Burden of Disease (GBD)
247   country groupings[40] for an unmitigated epidemic. **(b)** Expected sub clinical case attack rate, and peak
248   in subclinical cases. **(c)** Estimated basic reproduction number (R_0) in the capital city of each country
249   assuming age-specific clinical fraction shown in Fig. 2b and 50% infectiousness of subclinically
250   infected people. **(d)** Proportion of clinical cases in each age group at times relative to the peak of the
251   epidemic. The 146 city epidemics were aligned at the peak, and colours mark the GBD groupings in
252   a. **(e)** Age distribution of the first third and last third of clinical cases for 146 countries in GBD country
253   groupings.

254

255   We have shown age dependence in susceptibility to infection and in the probability of

256   displaying clinical symptoms of COVID-19, from around 20% in under 10s, to over 70% in

257   older adults. For a number of other pathogens, there is evidence that children (except for the

258   very youngest) have lower rates of symptomatic disease[12] and mortality[26], so the variable

259   age-specific clinical fraction for COVID-19 we find here fits with other studies. We have

260   quantified the age-specific susceptibility from available data, and other study types will be

261   needed to build the evidence base for the role of children, including serological surveys, and

262   close follow up of infected households.

263

264   The age-specific distribution of clinical infection we have found is similar in shape (but larger

265   in scale) to that generally assumed for pandemic influenza, but the age-specific susceptibility

266   is inverted. These differences have a large effect on how effective school closures may be in

267   limiting transmission, delaying the peak of expected cases, and decreasing the total and

268   peak number of cases. For COVID-19, school closures are likely to be much less effective

269   than for influenza-like infections.

270

271   It is critical to determine how infectious subclinical infections are compared to clinical

272   infections in order to properly assess predicted burdens both with and without interventions.

273   It is biologically plausible that milder cases are less transmissible, for example, because of

274   an absence of cough[28,29], but direct evidence is limited[41] and viral load is high in both clinical

275   and subclinical cases[30]. If those with subclinical infection are efficient transmitters of infection

276   compared to those with clinical infections, the overall burden is higher than if they are not as

277   infectious. At the same time, lower relative infectiousness would reduce the impact of

278   interventions targeting younger ages, such as school closure. By analysing epidemic

279   dynamics before and after school closures, or close follow up in household studies, it may be

280   possible to estimate the infectiousness of subclinical infections, however this will rely on

281   granular data by age and time.

282

283   A great deal of concern has been directed toward the expected burden of COVID-19 in low

284   and middle income countries (LMIC), which have lower population median age than many

285   high income countries. Our results show that these demographic differences, coupled with a

286   lower susceptibility and clinical fraction in younger ages, can result in proportionally fewer

287   clinical cases than would be expected in higher-income countries with flatter demographic

288   pyramids. This should not be interpreted as few cases in LMIC, because the projected

289   epidemics are still very large, resulting in high numbers infected. Moreover, the particular

290   relationships found between age, susceptibility, and clinical fraction are drawn from high and

291   middle income countries and may reflect not only age, but also the increasing frequency of

292   comorbidities with age. This relationship may therefore differ in LMIC for two key reasons:

293   first, the distribution of non-communicable comorbid conditions—which are already known to

294   increase the risk of severe disease from COVID-19[15] may be differently distributed by age,

295   often occurring in younger age groups[40], along with other possible risk factors such as

296   undernutrition[42]; and second, communicable comorbidities such as HIV[39], TB coinfection

297   (which has been suggested to increase risk[43]), and others[44] may alter the distribution of

298   severe outcomes by age. Observed severity and burden in LMIC may also be higher due to

299   a lack of health system capacity for intensive treatment of severe cases.

300

301   There are some limitations to the study. While information drawn from the early stages of the

302   epidemic is subject to uncertainty, age-specific information is drawn from several regions

303   and countries, and clinical studies support the hypothesis presented here. We assumed that

304   clinical cases are reported at a fixed fraction throughout the time period, although there may

305   have been changes in reporting and testing practices that affected case ascertainment by

306   age. We assumed that subclinical infections were less infectious than clinical infections, and

307   tested the impact of this on our findings (**Extended Data Figures 5 and 6**) but were not able

308   to estimate how infectious subclinical infections were. The sensitivity analyses showed very

309   similar clinical fraction and susceptibility with age, and we demonstrated the effect of this

310   parameter on school closure and global projections (**Fig. 3, Extended Data Figure 6**). We

311   have used mixing matrices from the same country, but not the same location as the fitted

312   data. We used contact matrices that combined physical and conversational contacts. We

313   therefore implicitly assume that they are a good reflection of contact relevant for the

314   transmission of SARS-CoV-2. If fomite, or faecal-oral routes of transmission are important in

315   transmission, these contact matrices may not be representative of transmission risk.

316

317   The role of age in transmission is critical to designing interventions aiming to decrease

318   transmission in the population as a whole, and to projecting the expected global burden.

319   Early evidence[25], including presented here, suggests that there is age dependence in

320   susceptibility and in the risk of clinical symptoms following infection. Understanding if and by

321   how much subclinical infections contribute to transmission has implications for predicted

322   global burden and the impact of control interventions. This question must be resolved to

323   effectively forecast and control COVID-19 epidemics.

324

325   Methods

326

327   **Transmission model structure used in all analyses**

328   We use an age-structured deterministic compartmental model (Fig. 1a, main text) stratified

329   into 5-year age bands, with time approximated in discrete steps of 0.25 days. Compartments

330   in the model are stratified by infection state (S, E, I_P, I_C, I_S, or R), age band, and the number

331   of time steps remaining before transition to the next infection state. We assume that people

332   are initially susceptible (S), and become exposed (E) after effective contact with an

333   infectious person. After a latent period, exposed individuals either develop a clinical or

334   subclinical infection; an exposed age-i individual develops a clinical infection with probability

335   y_i, otherwise developing a subclinical infection. Clinical cases are preceded by a preclinical

336   but infectious (I_P) state; from the preclinical state, individuals develop full symptoms and

337   become clinically infected (I_C). Based on evidence for other respiratory infections[16] we

338   assume that subclinical infections (I_S) are less infectious compared to preclinical and clinical

339   infections, and that subclinical individuals remain in the community until they recover. We

340   use 50% as a baseline for the relative infectiousness of individuals in the subclinical state,

341   and test the impact of other values (**Extended Data Figs. 5 and 6**). Isolated and recovered

342   individuals eventually enter the removed state (R); we assume these individuals are no

343   longer infectious and are immune to reinfection.

344

345   The length of time individuals spend in states E, I_P, I_C, or I_S is distributed according to

346   distributions d_E, d_P, d_C, or d_S, respectively (**Extended Data Table 2**). The force of infection

347   for an individual in age group i at time t is

348       $$\lambda_{i,t} = u_i \sum_j c_{ij,t}(I_{P_j} + I_{C_j} + f I_{S_j})/N_j,$$

349   where u_i is the susceptibility to infection of an age-i individual, $c_{ij,t}$ is the number of age-j

350   individuals contacted by an age-i individual per day at time t, f is the relative infectiousness

351   of a subclinical case, and $(I_{P_j} + I_{C_j} + f I_{S_j})/N_j$ is the effective probability that a random age-j

15

352   individual is infectious. Contacts vary over time t depending upon the modelled impact of

353   school closures and movement restrictions (see below).

354

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

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

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

358

359   We use the local age distribution for each city or region being modelled, and synthetic or

360   measured contact matrices for mixing between age groups (**Extended Data Table 2**). The

361   mixing matrices have four types of contacts: home, school, work and other contacts.

362

363   **Comparing models by fitting to the Wuhan epidemic**

364   We contrasted three models. In model 1, susceptibility varied by age ($u_i = u(i)$), but the

365   proportion of exposed individuals who became clinical cases did not vary ($y_i = y$). In model

366   2, the clinical case probability varied by age ($y_i = y(i)$), but susceptibility did not ($u_i = u$). In

367   model 3, there were no age-related differences in susceptibility or clinical fraction ($u_i = u$,

368   and $y_i = y$). Susceptibility and clinical fraction curves were fitted using three control points

369   for young, middle, and old age, interpolating between them with a half-cosine curve (see

370   below for details).

371

372   We assumed that the initial outbreak in Wuhan was seeded by introducing one exposed

373   individual per day of a randomly drawn age between A_{min} *and* A_{max} for 14 days starting on a

374   day (t_{seed}) in November[30,31]. We used the age distribution of Wuhan City prefecture in 2016[45]

375   and contact matrices measured in Shanghai[32] as a proxy for large cities in China. This

376   contact matrix is stratified into school, home, work, and other contacts. We aggregated the

377   last three categories into non-school contacts and estimated how components of the contact

378   matrix changed early in the epidemic in response to major changes. Schools closed on

379 January 12th for the Lunar New Year holiday, so we decreased school contacts, but the

380 holiday period may have changed non-school contacts, so we estimate this effect by

381 inferring the change in non-school contact types, q_H. Large-scale restrictions started on

382 January 23rd 2020 giving restrictions on travel and movement imposed by authorities, and

383 we inferred the change in contact patterns during this period, q_L. Specifically:

384 $$c_{ij,t} = school(t) \cdot c_{ij|school} + other(t) \cdot c_{ij|other},$$

385 where

386 $school(t) = \{$ 1 t < 12 January

387 0 $t \geq$ 12 January

388 and

389 1 t < 12 January

390 $other(t) = \{$ q_H 12 January $\leq t$ < 23 January

391 q_L $t \geq$ 23 January.

392

393 We fitted the model to incident confirmed cases from the early phase of the epidemic in

394 China (December 8, 2019-February 1, 2020) reported by China CDC[1]. During this period,

395 the majority of cases were from Wuhan City, and we truncated the data after February 1st

396 because there were more cases in other cities after this time. We jointly fitted the model to

397 the age distribution of cases at 3 time windows (December 8, 2019 to January 22, 2020)

398 reported by Li et al.[24] and a further time window (December 8, 2019 to February 11, 2020)

399 reported by China CDC[1]. Because there was a large spike of incident cases reported on

400 February 1 determined to have originated from the previous week, we amalgamated all

401 cases from January 25 to February 1, including those in the large spike, into a single data

402 point for the week. We assumed 10% of clinical cases were reported[19]. We used a Dirichlet

403 distribution with a flat prior to obtain 95% HDIs for reported case data stratified by age group

404 for display in figures.

405

406 We used Markov-chain Monte Carlo to jointly fit each hypothesis to the two sets of empirical

407 observations from the epidemic in Wuhan City, China (**Supplementary Table 1**). We used a

408   negative binomial likelihood for incident cases and a Dirichlet-multinomial likelihood for the

409   age distribution of cases, using the likelihood

410
$$L = \left(\prod_{k=1}^{K} NegBinom(C_k|size = 200, mean = c_k)\right)\left(\prod_{m=1}^{M} DirMultinom(A_m|\frac{200}{||a_m||} a_m)\right)$$

411   Above, C_k is the observed incidence on day k while c_k is the model-predicted incidence for

412   day k, for each of K days. A_m is the observed age distribution for time period m (case counts

413   for each age group) while a_m is the model-predicted age distribution for the same period, and

414    $||a_m||$ is the total number of cases over all age groups in time period m, measured for M

415   time periods. We set the precision of each distribution to 200 to capture additional

416   uncertainty in data points that would not be captured with a Poisson or multinomial likelihood

417   model.

418

419   For all Bayesian inference (i.e. shown in Figs. 1 and 2), we used differential evolution

420   Markov chain Monte Carlo[46], first running numerical optimization to place starting values for

421   each chain near the posterior mode. We then run 2000-3000 samples of burn-in, and

422   generate at least 10,000 samples post-burn-in. Recovered posterior distributions, with prior

423   distributions overlaid, are shown in **Extended Data Fig. 1**. We distinguished fitted models

424   using Deviance Information Criterion (DIC)[47].

425

426   **Analysis of the stationary age distribution of cases**

427   To infer age-specific clinical fraction and susceptibility from reported case distributions, we

428   assumed that reported cases follow the stationary distribution of cases reached in the early

429   phase of an epidemic. Using our dynamic model would allow modelling any transient

430   emphasis in the case distribution associated with the age of the individuals who seeded

431   infection in a given region, but since the age of the true first cases is not generally known,

432   we used the stationary distribution instead. Specifically, we used Bayesian inference to fit

433   age-specific susceptibility and clinical fraction to the reported case distribution by first

434   generating the expected case distribution k_i from (1) the age-specific susceptibility, u_i, (2) the

435   age-specific clinical fraction, y_i, (3) the measured or estimated contact matrix for the country,

436   and (4) the age structure of the country or region. We then used the likelihood

437
$$L = Multinom(c_i \,|k_i),$$

438

439   where c_i is the observed case distribution, when fitting to data from a single country or

440   region. When fitting to a combined set of regions and/or countries, we used the likelihood

441
$$L = \prod_{j=1}^{m} DirMultinom(c_{i,j}\,|sk_{i,j})^{w_j}$$

442   across countries $j \in \{1,2,\ldots,m\}$ with weights w_j such that $\prod_j w_j = 1$. We weighted[48] each of

443   the 13 provinces of China in our data set by 1/13, each of the 12 regions of Italy by 1/12, the

444   three reported case distributions from China CDC by 1/3, and data from South Korea,

445   Singapore, Japan and Ontario each by 1, then scaled all weights to multiply to 1.

446

447   The age-specific susceptibility u_i and age-specific clinical fraction y_i were estimated by

448   evaluating the expected case distribution c_i according to the likelihood functions given

449   above. It is not possible to identify both u_i and y_i from case data alone. Accordingly, we

450   inferred the age-specific clinical fraction, y_i from surveillance data from Italy reporting the

451   age-specific number of cases that were asymptomatic, paucisymptomatic, mild, severe, and

452   critical[29]. We assumed that asymptomatic and paucisymptomatic infections may be

453   underascertained relative to mild, severe, and critical cases, and therefore estimated an

454   "inflation factor" $z > 1$ giving the number of unascertained asymptomatic or

455   paucisymptomatic infections for each reported infection in these data. Accordingly, we

456   applied the likelihood penalty

457
$$P_L = \prod_i Beta\left(\frac{mild_i + sev_i + crit_i}{z(asymp_i + pauci_i) + mild_i + sev_i + crit_i}\middle| \alpha = 10000 y_i, \beta = 10000(1 - y_i)\right)$$

458   when fitting y_i in order to constrain the relative shape of the clinical fraction curve by age.

459   Here, $mild_i$ is the number of mild cases reported in age group i, sev_i the number of severe

460   cases in age group i, etc. Therefore the age-specific clinical fraction reflected the proportion

461   of infections reported by Riccardo et al. as mild, critical, or severe, relative to an estimated

462   proportion of asymptomatic and paucisymptomatic infections.

463

464   In order to estimate a value for the inflation factor z compatible with empirical data on the

465   severity of infections, we applied a further likelihood penalty when estimating the consensus

466   fit for clinical fraction and susceptibility in order to match information on age-specific

467   susceptibility collected from recent contact-tracing studies[25,27,28,30]. A leave-one-out analysis

468   showed that these additional data allowed the model fitting procedure to converge on a

469   consistent profile for both u_i and y_i (**Extended Data Fig. 2**).

470

471   We extracted age-specific case data from the following sources. For provinces of China, we

472   used age-specific case numbers reported by China CDC[1] as well as line list data compiled

473   by the Shanghai Observer[49]. For regions of Italy, we used age-specific case numbers

474   reported by the Istituto Superiore di Sanità on March 13, 2020[50]. For South Korea, we used

475   the line list released by Kim et al. based on data from the Korea Centers for Disease Control

476   and Prevention[22]. For Japan, we used the Open Covid Linelist[51,52]. For Singapore, we used

477   Singapore Ministry of Health data compiled by Koh[21]. For Ontario, we used data compiled by

478   the COVID-19 Canada Open Data Working Group[53].

479

480   To validate our line list analysis, we fitted the dynamic model to incidence data from Beijing,

481   Shanghai, South Korea and Lombardy, Italy (**Extended Data Fig. 3**). We fixed the reporting

482   rate for Beijing, Shanghai, South Korea, and Lombardy to 20%. Beijing and Shanghai

483   incidence data were given by case onset, so we assumed no delay between reported and

484   true case onsets. Incidence data for South Korea were given by the date of confirmation

485   only; we assumed the reporting delay followed a gamma distribution with a 7-day mean.

486   Incidence data for Italy were given separately for case onset and case confirmation, with

487   only a subset of onset dates available; accordingly, we fit the proportion of confirmed cases

488   with onset dates and the delay from onset to confirmation. We adjusted the size parameter

489   of the negative binomial distribution used to model case incidence to 10 to reflect greater

490   variability among fewer data points for these countries than for Wuhan. Beijing and Shanghai

491   were fitted jointly, with separate dates of introduction but the same fitted susceptibility, large-

492   scale restriction date and large-scale restriction magnitude. South Korea and Italy were each

493   fitted separately; we fitted a large-scale restriction date and magnitude for both South Korea

494   and Italy.

495

496   For both the line list fitting and validation, we assumed that schools were closed in China,

497   but remained open in South Korea, Japan, Italy, Singapore, and Canada, as schools were

498   open for the majority of the period covered by the data in the latter five countries.

499

500   **Quantifying the impact of school closure**

501   To determine the impact in other cities with different demographic profiles we used the

502   inferred parameters from our line list analysis to parameterise our transmission model for

503   projections to other cities. We chose these to compare projections for a city with a high

504   proportion of elderly individuals (Milan, Italy); a moderate-aged population (Birmingham,

505   United Kingdom); and a city in a low-income country with a high proportion of young

506   individuals (Bulawayo, Zimbabwe). For this analysis, we compared an outbreak of COVID-

507   19, for which the burden and transmission is concentrated in relatively-older individuals, with

508   an outbreak of pandemic influenza, for which the burden and transmission is concentrated in

509   relatively-younger individuals. We assumed that immunity to influenza builds up over a

510   person's lifetime, such that an individual's susceptibility to influenza infection plateaus at

511   roughly age 35, and assumed that the severity of influenza infection is highest in the elderly

512   and in children under 10 years old[37].

513

514   To model Milan, we used the age distribution of Milan in 2019[54] and a contact matrix

515   measured in Italy in 2006[11]. To model Birmingham, we used the age distribution of

516   Birmingham in 2018[55] and a contact matrix measured in the UK in 2006[11]. To model

21

517   Bulawayo, we used the age distribution of Bulawayo Province in 2012[56] and a contact matrix

518   measured in Manicaland, Zimbabwe in 2013[57]. We assumed that the epidemic was seeded

519   by two infectious individuals in a random age group per week for 5 weeks. We scaled the

520   age-specific susceptibility u_i by setting the "target" basic reproductive number R_0 = 2.4, as a

521   representative example. We also performed a sensitivity analysis where we scaled u_i to

522   result in R_0 = 2.4 in Birmingham, and using the same setting for u_i in all three cities, so that

523   the actual R_0 changed depending upon contact matrices and demographics used to model

524   each city. This produced qualitatively similar results (**Extended Data Figure 7**).

525

526   We projected the impact of school closure by setting the contact multiplier for school

527   contacts *school(t)* to 0. Complete removal of school contacts may overestimate the impact of

528   school closures because of alternative contacts children make when out of school[58]. This will

529   however give the maximum impact of school closures in the model to demonstrate the

530   differences.

531

532   **Projecting the global impact**

533   To project the impact of COVID-19 outbreaks in global cities, we used mixing matrices from

534   Prem et al.[31] and demographic structures for 2020 from World Population Prospects 2019 to

535   simulate a COVID-19 outbreak in 146 global capital cities for which synthetic matrices,

536   demographic structures and total populations were available. For simplicity, we assumed

537   that capital cities followed the demographic structure of their respective countries and took

538   the total population of each capital city from the R package *maps*. For each city, we scaled u_i

539   to result in an average R_0 = 2.4 in Birmingham, UK, and used the same setting for u_i for all

540   cities, so that the realised R_0 would change according to the contact matrices and

541   demographics for each city. We simulated 20 outbreaks in each city, drawing the age-

542   specific clinical fraction y_i from the posterior of the estimated overall clinical fraction from our

543   line list analysis (Fig. 2), and analysed the time to the peak incidence of the epidemic, the

544   peak clinical and subclinical incidence of infection, and the total number of clinical and

545   subclinical infections. We took the first third and the last third of clinical cases in each city to

546   compare the early and late stages of the epidemic.

547

548   **Contact matrices**

549   Wherever possible, we use measured contact matrices (**Supplementary Table 2**). We adapt

550   each of these mixing matrices, using 5-year age bands, to specific regions of the countries

551   they were measured in by reprocessing the original contact surveys with the population

552   demographics of the local regions. The contact matrices and demographics we used for

553   Figs. 1-3 of the main text are shown in **Extended Data Figure 8**.

554

555   The contact survey in Shanghai[59] allowed respondents to record both individual (one-on-

556   one) and group contacts, the latter with approximate ages. While individual contacts were

557   associated with a context (home, work, school, etc.) group contacts were not, and so we

558   assumed that all group contacts which involved individuals aged 0-19 occurred at school.

559   We also assumed that group contacts were lower intensity than individual contacts,

560   weighting group contacts by 50% relative to one-on-one contacts.

561

562   We assumed schools were closed during the epidemic in China (because schools closed for

563   the Lunar New Year holiday and remained closed), but open in Italy, Singapore, South

564   Korea, Japan, and Canada, because we used data from the early part of the epidemics in

565   those countries during which schools were open.

566

567   **Sensitivity analyses**

568   Since the infectiousness of subclinical individuals was not identifiable from data we have

569   available, in Figure 2 we adopted a baseline estimate of 50% relative to preclinical and

570   clinical individuals. In **Extended Data Figure 5**, we performed sensitivity analysis by

571   repeating our model runs with the alternative values for subclinical infectiousness between

572   0% and 100%. We did not find marked difference in the findings or estimates.

573

574   In Figure 2 we fitted the age distributions of cases in 6 countries jointly to findings from

575   recent studies on the susceptibility of children. We tested the sensitivity of our findings to the

576   findings of the other studies by conducting a leave one out sensitivity analysis. The results

577   are given in **Extended Data Figure 2**, and we did not find major changes to the shape of the

578   age-dependence in either susceptibility or clinical fraction.

579

580   In Fig. 3, we showed the epidemic in 3 cities with fixed R_0 at 2.4 to illustrate the impact that

581   demographics alone have on the effectiveness of interventions. This means that higher rates

582   of contact measured in surveys in in Milan and Bulawayo compared to Birmingham were not

583   included. We also tested the sensitivity of findings on school closure for which we fix

584   susceptibility u_i and therefore R_0 varies (**Extended Data Figure 7**). The conclusions

585   regarding the relative effectiveness of school closures for COVID-19 versus influenza are

586   similar.

587

588   In Fig. 4, we assumed that the age-specific clinical fraction was the same across all settings,

589   but we tested the sensitivity of our projections (Figure 4) to the age-specific clinical fraction

590   used in lower-income countries. However, a higher rate of comorbidities in lower-income

591   countries could change the age-specific probability of developing clinical symptoms upon

592   infection. To investigate this possibility, we construct a schematic alternative age-specific

593   profile of clinical fraction by (1) increasing the age-specific probability of developing

594   symptoms by 15% for individuals under the age of 20 and (2) shifting the age-specific clinical

595   fraction for individual over the age of 20 by 10 years older (**Extended Data Figure 4**). We

596   repeated the analyses with these functions and found increased burden in lower-income

597   countries, that could exceed the burden of clinical cases in higher-income countries.

598

599   Finally, we repeated our projections for country-specific burdens of COVID-19 assuming

600   different values for the relative infectiousness of subclinical infections. We found that this

601 had a relatively small impact on the relationship between median age and case burden

602 across countries (**Extended Data Figure 6**).

603

604

605

606

607 **Acknowledgements**

637

638 **Author contributions**

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

644

645 **Data Availability and Code Availability**

646 The data used for fitting are publicly available, but will also be made available with the code

647 in the github repository for the project at https://github.com/nicholasdavies/covid-age.

648 Contact matrix data are available at zenodo[21,22].

649

650 **Competing interests**

651 The authors have no competing interests.

652

653 **Additional information**

654 Supplementary Information is available for this paper. Correspondence and requests for

655 materials should be addressed to Rosalind M Eggo or Nicholas G Davies at

656 r.eggo@lshtm.ac.uk or nicholas.davies@lshtm.ac.uk

657

658

659 **References**

660

661 1. Zhang. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases

662   (COVID-19) in China. *Chin. J. Epidemiol.* **41**, 145–151 (20200217).

663 2. Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019

664   outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit.*

665   *Health* S2589750020300261 (2020) doi:10.1016/S2589-7500(20)30026-1.

666 3. D, C. *et al.* The early phase of the COVID-19 outbreak in Lombardy, Italy. *ArXiv200309320 Q-Bio*

667   (2020).

668 4. Shim, E., Tariq, A., Choi, W., Lee, Y. & Chowell, G. Transmission potential and severity of

669   COVID-19 in South Korea. *Int. J. Infect. Dis.* S1201971220301508 (2020)

670   doi:10.1016/j.ijid.2020.03.031.

671 5. Dong, Y. *et al.* Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus

672   Disease in China. *Pediatrics* e20200702 (2020) doi:10.1542/peds.2020-0702.

673 6. Zhao, X. *et al. Incidence, clinical characteristics and prognostic factor of patients with COVID-19:*

674   *a systematic review and meta-analysis.*

675   http://medrxiv.org/lookup/doi/10.1101/2020.03.17.20037572 (2020)

676   doi:10.1101/2020.03.17.20037572.

677 7. Anderson, R. M. *et al.* Epidemiology, transmission dynamics and control of SARS: the 2002–2003

678   epidemic. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 1091–1105 (2004).

679 8. Nickbakhsh, S. *et al.* Epidemiology of Seasonal Coronaviruses: Establishing the Context for the

680   Emergence of Coronavirus Disease 2019. *J. Infect. Dis.* jiaa185 (2020) doi:10.1093/infdis/jiaa185.

681   9.   Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission

682   dynamics of SARS-CoV-2 through the postpandemic period. *Science* eabb5793 (2020)

683   doi:10.1126/science.abb5793.

684   10. Huang, A. T. *et al. A systematic review of antibody mediated immunity to coronaviruses: antibody*

685   *kinetics, correlates of protection, and association of antibody responses with severity of disease.*

686   http://medrxiv.org/lookup/doi/10.1101/2020.04.14.20065771 (2020)

687   doi:10.1101/2020.04.14.20065771.

688   11. Cowling, B. J. *et al.* Increased Risk of Noninfluenza Respiratory Virus Infections Associated With

689   Receipt of Inactivated Influenza Vaccine. *Clin. Infect. Dis.* **54**, 1778–1783 (2012).

690   12. Tsagarakis, N. J. *et al.* Age-related prevalence of common upper respiratory pathogens, based

691   on the application of the FilmArray Respiratory panel in a tertiary hospital in Greece: *Medicine*

692   *(Baltimore)* **97**, e10903 (2018).

693   13. Common cold. *nhs.uk* https://www.nhs.uk/conditions/common-cold/ (2017).

694   14. Galanti, M. *et al.* Rates of asymptomatic respiratory virus infection across age groups. *Epidemiol.*

695   *Infect.* **147**, e176 (2019).

696   15. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in

697   Wuhan, China: a retrospective cohort study. *The Lancet* S0140673620305663 (2020)

698   doi:10.1016/S0140-6736(20)30566-3.

699   16. Van Kerckhove, K., Hens, N., Edmunds, W. J. & Eames, K. T. D. The Impact of Illness on Social

700   Networks: Implications for Transmission and Control of Influenza. *Am. J. Epidemiol.* **178**, 1655–

701   1662 (2013).

702   17. Mossong, J. *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious

703   Diseases. *PLOS Med.* **5**, e74 (2008).

704   18. Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. & Ferguson, N. M. Estimating the

705   impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–754

706   (2008).

707   19. Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E. & Edmunds, W. J. Measured Dynamic Social

708   Contact Patterns Explain the Spread of H1N1v Influenza. *PLoS Comput. Biol.* **8**, e1002425

709   (2012).

710   20. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan,

711   China. *The Lancet* **395**, 497–506 (2020).

712   21. Koh, A. Singapore COVID-19 Cases. http://alexkoh.net/covid19/.

713   22. Data Science for COVID-19 (DS4C). https://kaggle.com/kimjihoo/coronavirusdataset.

714   23. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies

715   Needed. *N. Engl. J. Med.* **0**, null (2020).

716   24. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected

717   Pneumonia. *N. Engl. J. Med.* NEJMoa2001316 (2020) doi:10.1056/NEJMoa2001316.

718   25. Bi, Q. *et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391*

719   *cases and 1,286 of their close contacts.*

720   http://medrxiv.org/lookup/doi/10.1101/2020.03.03.20028423 (2020)

721   doi:10.1101/2020.03.03.20028423.

722   26. Yang, Y. *et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in*

723   *China.* http://medrxiv.org/lookup/doi/10.1101/2020.02.10.20021675 (2020)

724   doi:10.1101/2020.02.10.20021675.

725   27. Zhang, J. *et al.* Age profile of susceptibility, mixing, and social distancing shape the dynamics of

726   the novel coronavirus disease 2019 outbreak in China. (2020) doi:10.1101/2020.03.19.20039107.

727   28. Gudbjartsson, D. F. *et al.* Spread of SARS-CoV-2 in the Icelandic Population. *N. Engl. J. Med.*

728   NEJMoa2006100 (2020) doi:10.1056/NEJMoa2006100.

729   29. Riccardo, F. *et al. Epidemiological characteristics of COVID-19 cases in Italy and estimates of the*

730   *reproductive numbers one month into the epidemic.*

731   http://medrxiv.org/lookup/doi/10.1101/2020.04.08.20056861 (2020)

732   doi:10.1101/2020.04.08.20056861.

733   30. Lavezzo, E. *et al.* Suppression of COVID-19 outbreak in the municipality of Vo, Italy. (2020)

734   doi:10.1101/2020.04.17.20053157.

735   31. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact

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

737   32. Onder, G., Rezza, G. & Brusaferro, S. Case-Fatality Rate and Characteristics of Patients Dying in

738   Relation to COVID-19 in Italy. *JAMA* (2020) doi:10.1001/jama.2020.4683.

739   33. Chan, K. P. Control of Severe Acute Respiratory Syndrome in Singapore. *Environ. Health Prev.*

740   *Med.* **10**, 255–259 (2005).

741   34. Lau, J. T. F. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10

742   to day 62. *J. Epidemiol. Community Health* **57**, 864–870 (2003).

743   35. Cauchemez, S. *et al.* School closures during the 2009 influenza pandemic: national and local

744   experiences. *BMC Infect. Dis.* **14**, 207 (2014).

745   36. Cauchemez, S. *et al.* Closure of schools during an influenza pandemic. *Lancet Infect. Dis.* **9**,

746   473–481 (2009).

747   37. Greer, A. L., Tuite, A. & Fisman, D. N. Age, influenza pandemics and disease dynamics.

748   *Epidemiol. Infect.* **138**, 1542–1549 (2010).

749   38. Clark, A. *et al. How many are at increased risk of severe COVID-19 disease? Rapid global,*

750   *regional and national estimates for 2020*.

751   http://medrxiv.org/lookup/doi/10.1101/2020.04.18.20064774 (2020)

752   doi:10.1101/2020.04.18.20064774.

753   39. Cohen, C. *et al.* Severe Influenza-associated Respiratory Infection in High HIV Prevalence

754   Setting, South Africa, 2009–2011. *Emerg. Infect. Dis.* **19**, (2013).

755   40. IHME. Global Burden of Disease. http://www.healthdata.org/gbd.

756   41. Williams, C. M. *et al.* Exhaled Mycobacterium tuberculosis output and detection of subclinical

757   disease by face-mask sampling: prospective observational studies. *Lancet Infect. Dis.*

758   S1473309919307078 (2020) doi:10.1016/S1473-3099(19)30707-8.

759   42. Murray, J. *et al.* Determining the Provincial and National Burden of Influenza-Associated Severe

760   Acute Respiratory Illness in South Africa Using a Rapid Assessment Methodology. *PLOS ONE*

761   **10**, e0132078 (2015).

762   43. Liu, Y. *et al. Active or latent tuberculosis increases susceptibility to COVID-19 and disease*

763   *severity*. http://medrxiv.org/lookup/doi/10.1101/2020.03.10.20033795 (2020)

764   doi:10.1101/2020.03.10.20033795.

765   44. Cohen, A. L. *et al.* Potential Impact of Co-Infections and Co-Morbidities Prevalent in Africa on

766   Influenza Severity and Frequency: A Systematic Review. *PLOS ONE* **10**, e0128580 (2015).

767   45. National Bureau of Statistics. China Statistical Year Book. (2005-2018).

768   http://www.stats.gov.cn/tjsj/ndsj/.

769   46. Braak, C. J. F. T. A Markov Chain Monte Carlo version of the genetic algorithm Differential

770   Evolution: easy Bayesian computing for real parameter spaces. *Stat. Comput.* **16**, 239–249

771   (2006).

772   47. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. Bayesian measures of model

773   complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).

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

775   49. Observer, S. Shanghai Observer. COVID-2019 Linelist.

776   http://data.shobserver.com/www/datadetail.html?contId=1000895.

777   50. Epicentro. *Bolletino Sorveglianza Integrata COVID-19 12 Marzo 2020 Appendix.*

778   https://www.epicentro.iss.it/coronavirus/bollettino/Bolletino-sorveglianza-integrata-COVID-19_12-

779   marzo-2020_appendix.pdf.

780   51. COVID19_2020_open_line_list.

781   https://docs.google.com/spreadsheets/d/1itaohdPiAeniCXNlntNztZ_oRvjh0HsGuJXUJWET008/e

782   dit?usp=sharing.

783   52. Xu, B. *et al.* Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect. Dis.*

784   S1473309920301195 (2020) doi:10.1016/S1473-3099(20)30119-5.

785   53. COVID-19 in Canada. https://art-bd.shinyapps.io/covid19canada/.

786   54. Milano (Metropolitan City, Italy) - Population Statistics, Charts, Map and Location.

787   http://citypopulation.info/en/italy/admin/lombardia/015__milano/.

788   55. Age breakdown of the population of Birmingham - Office for National Statistics.

789   https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/agebreakd

790   ownofthepopulationofbirmingham.

791   56. Bulawayo (City, Zimbabwe) - Population Statistics, Charts, Map and Location.

792   http://citypopulation.info/php/zimbabwe-admin.php?adm1id=A.

793   57. Melegaro, A. *et al.* Social Contact Structures and Time Use Patterns in the Manicaland Province

794   of Zimbabwe. *PLOS ONE* **12**, e0170459 (2017).

795   58. Kucharski, A. J., Conlan, A. J. K. & Eames, K. T. D. School's Out: Seasonal Variation in the

796   Movement Patterns of School Children. *PLOS ONE* **10**, e0128070 (2015).

797   59. Zhang, J. *et al.* Patterns of human social contact and contact with animals in Shanghai, China.

798   *Sci. Rep.* **9**, 1–11 (2019).

799

Extended Data

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

Parameter	Age Group	Mean	Quantile 2.5%	Quantile 25%	Quantile 50%	Quantile 75%	Quantile 97.5%
Susceptibility	0-9	0.33	0.25	0.3	0.33	0.36	0.43
Susceptibility	10-19	0.37	0.28	0.33	0.37	0.4	0.47
Susceptibility	20-29	0.69	0.54	0.64	0.69	0.74	0.82
Susceptibility	30-39	0.81	0.65	0.76	0.81	0.86	0.95
Susceptibility	40-49	0.74	0.59	0.7	0.75	0.79	0.86
Susceptibility	50-59	0.8	0.65	0.76	0.81	0.85	0.93
Susceptibility	60-69	0.89	0.72	0.85	0.9	0.94	0.99
Susceptibility	70+	0.77	0.62	0.72	0.78	0.82	0.9
Clinical fraction	0-9	0.4	0.31	0.36	0.4	0.44	0.48
Clinical fraction	10-19	0.25	0.19	0.23	0.25	0.29	0.32
Clinical fraction	20-29	0.37	0.28	0.33	0.37	0.41	0.46
Clinical fraction	30-39	0.42	0.33	0.38	0.42	0.46	0.51
Clinical fraction	40-49	0.51	0.41	0.47	0.51	0.55	0.6
Clinical fraction	50-59	0.59	0.48	0.55	0.59	0.62	0.67
Clinical fraction	60-69	0.72	0.63	0.69	0.72	0.75	0.79
Clinical fraction	70+	0.76	0.68	0.73	0.76	0.79	0.82

Extended Data Table 1. Posterior estimates for the consensus susceptibility and clinical fraction from 6 countries. Note that susceptibility is a relative measure.

2

| Parameter	Description	Applies in fits	Value	Reference
d_E	Incubation period (E to I_P and E to I_S; days)	All	$\sim gamma(\mu = 3.0, k = 4)$	Derived from[1][2]
d_P	Duration of preclinical infectiousness (days)	All	$\sim gamma(\mu = 2.1, k = 4)$	Derived from[2]
d_C	Duration of clinical infectiousness (I_C to R; days)	All	$\sim gamma(\mu = 2.9, k = 4)$	[3]
d_S	Duration of subclinical infectiousness (days)	All	$\sim gamma(\mu = 5, k = 4)$	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	[4]
f	Relative infectiousness of subclinical cases	All	50% (0% and 100% 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	[5]
Δt	Time step for discrete-time simulation	All	0.25 days	
A_{min}, A_{max}	Age range of seed cases	Wuhan	Estimated	
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	

Extended Data Table 2. Model parameters.
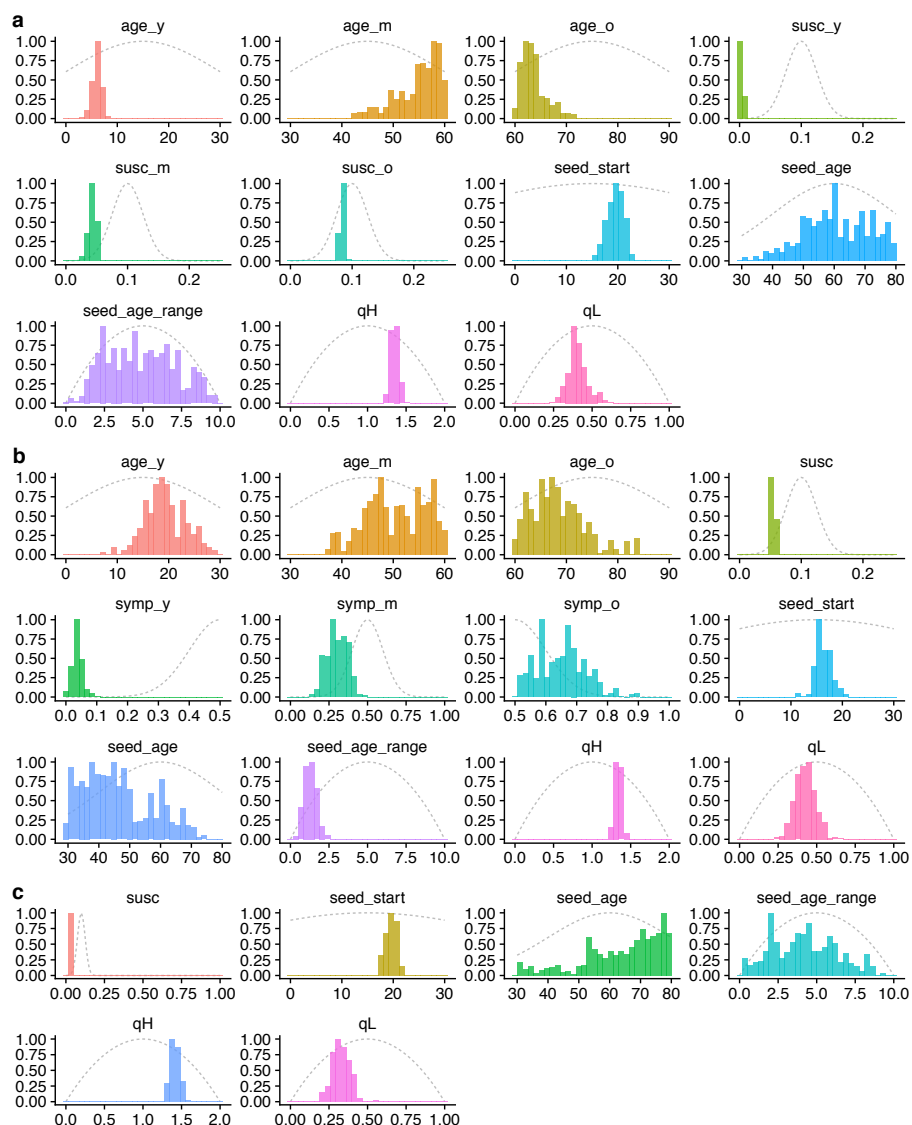
Extended Data Figure 1. Prior distributions (grey dotted lines) and posterior distributions (coloured histograms) for model parameters fitting to the early epidemic in Wuhan (Fig. 1, main text); seed_start is measured in days after November 1st, 2019. **(a)** Model 1 (age-varying contact patterns and susceptibility); **(b)** Model 2 (age-varying contact patterns and clinical fraction); **(c)** Model 3 (age-varying contact patterns only). See also Supplementary Table 3.
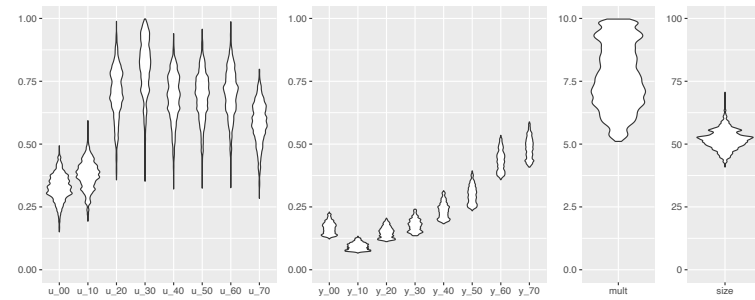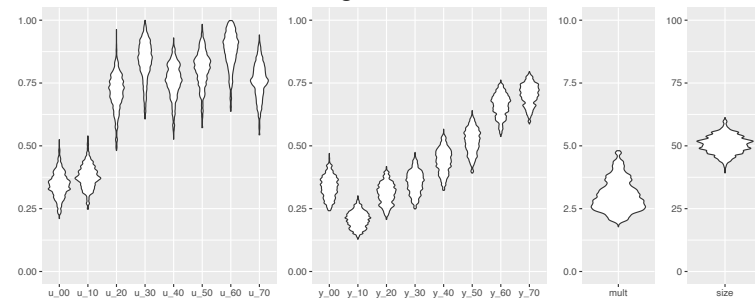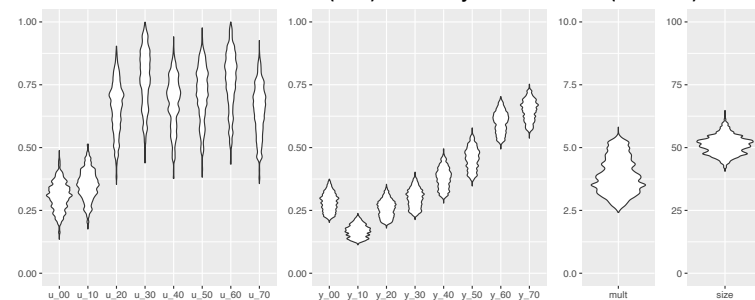
4

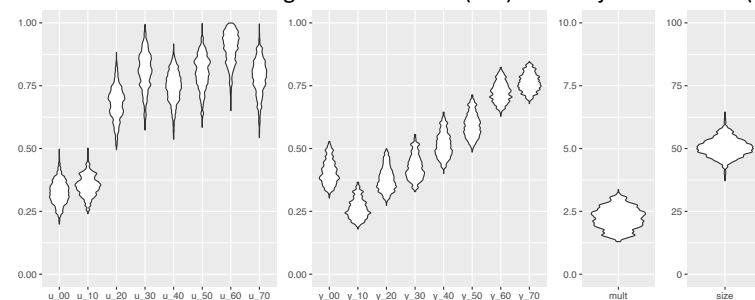**Extended Data Figure 2.** Analysis showing how the inferred age-varying susceptibility (first column) and age-varying clinical fraction (second column) depend upon the additional data sources used.
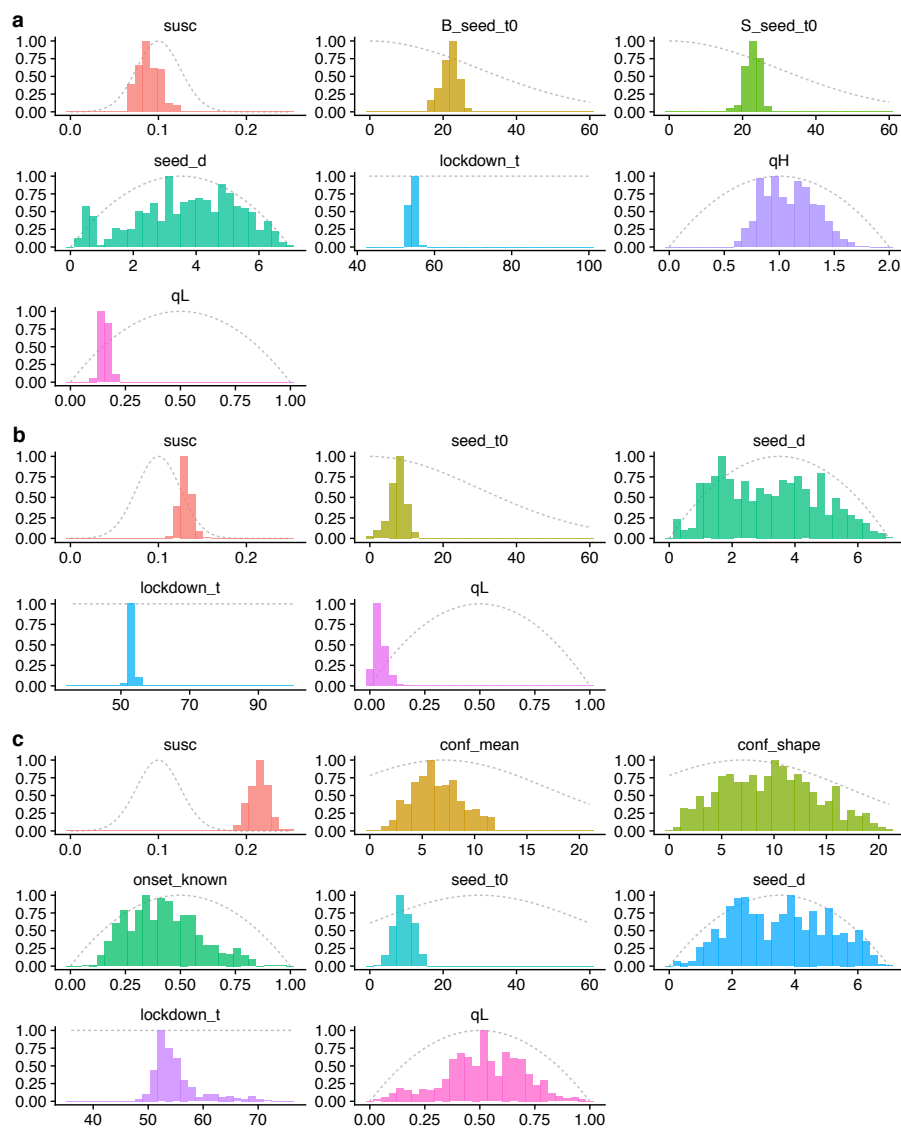
Extended Data Figure 3. Prior and posterior distributions for the epidemics in **(a)** Beijing and Shanghai, **(b)** South Korea and **(c)** Lombardy using the "consensus" fit for age-specific clinical fraction and assuming subclinical infections are 50% as infectious as clinical infections (see Fig. 2c, main text). For **(a),** times are in days after December 1st, 2019; for **(b)** and **(c)**, times are in days after January 1st, 2019. Note, seed_d is the inferred duration of the seeding event. See also Supplementary Table 3.
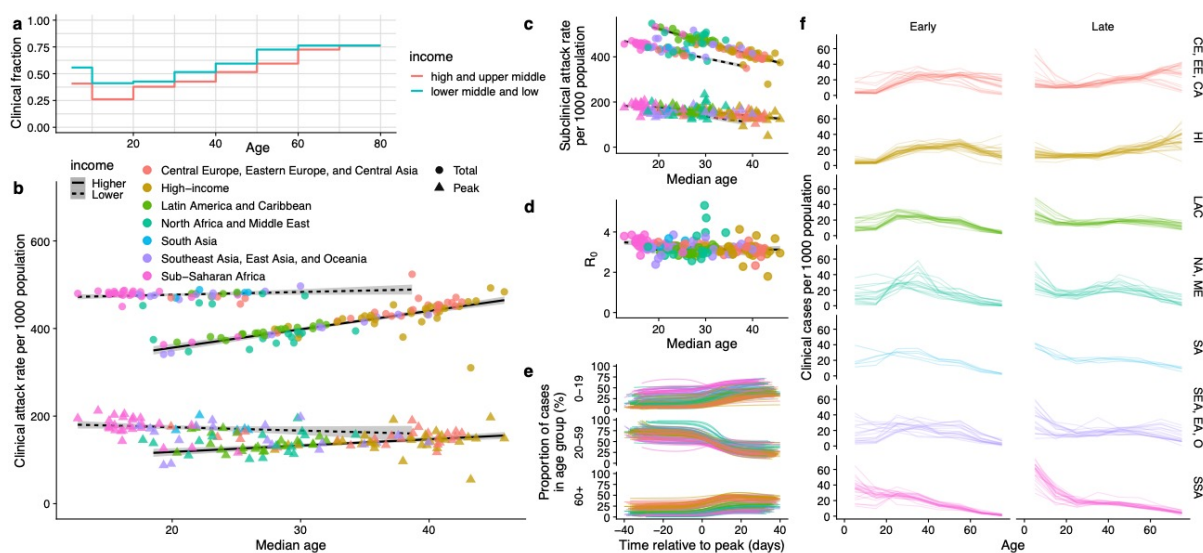
Extended Data Figure 4. Global projections assuming greater severity in lower-income countries. **(a)** Schematic age-specific clinical fraction for higher-income and lower-income countries. **(b-f)** Illustrative results of the projections for 146 capital cities assuming a higher age-varying clinical fraction in lower-income countries. See Fig. 4 (main text) for details.
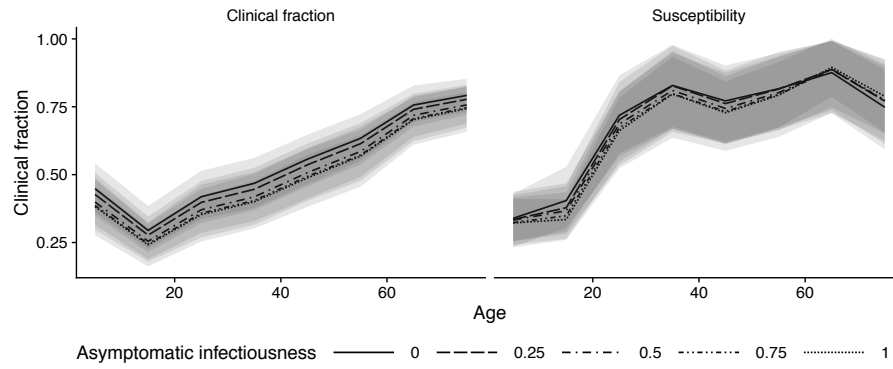
Extended Data Figure 5. Consensus age-specific clinical fraction and susceptibility, assuming subclinical infections are 0%, 25%, 50%, 75%, or 100% as infectious as clinical infections.
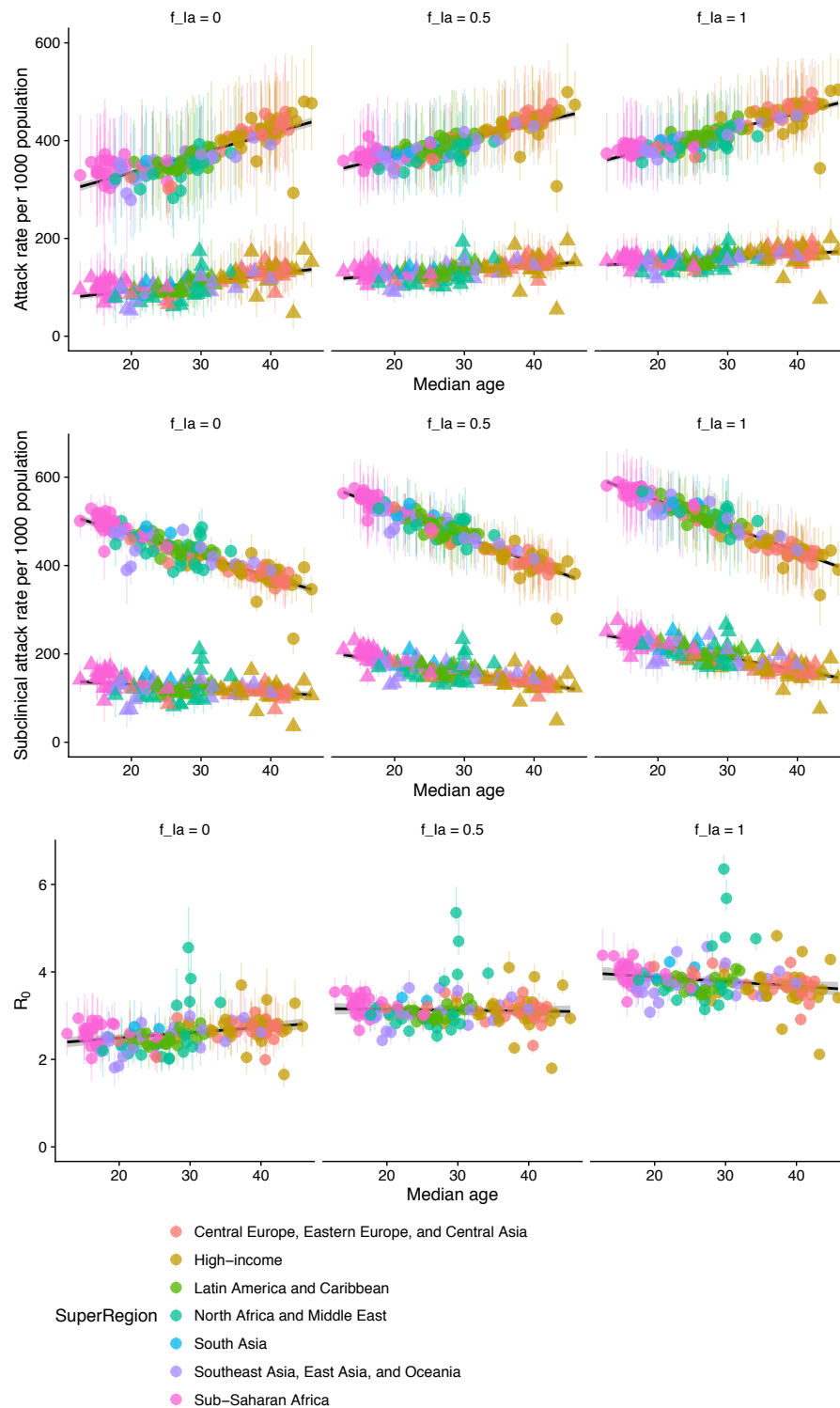
Extended Data Figure 6. (a) Projected total and peak clinical case attack rate for 146 capital cities, under different assumptions for the infectiousness of subclinical infections. **(b)** Projected total and peak subclinical infection attack rate for 146 capital cities, under different assumptions for the infectiousness of subclinical infections. **(c)** Projected differences in R0 among 146 capital cities, under different assumptions for the infectiousness of subclinical infections.
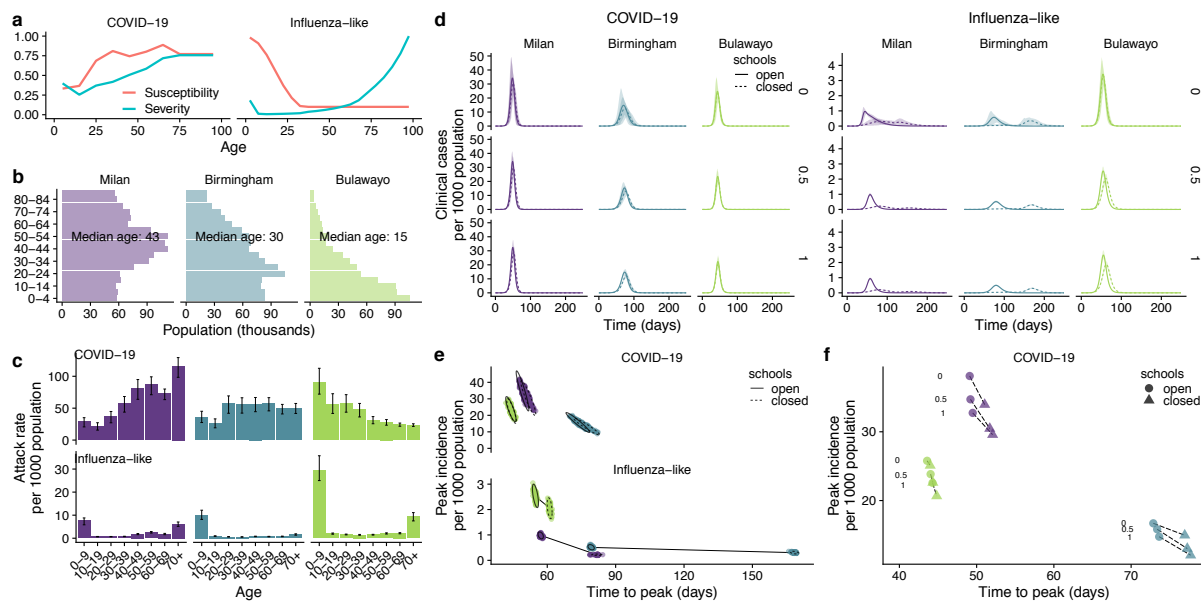
Extended Data Figure 7. Comparison of school closures in three exemplar cities when susceptibility u_i is fixed across settings instead of R0.
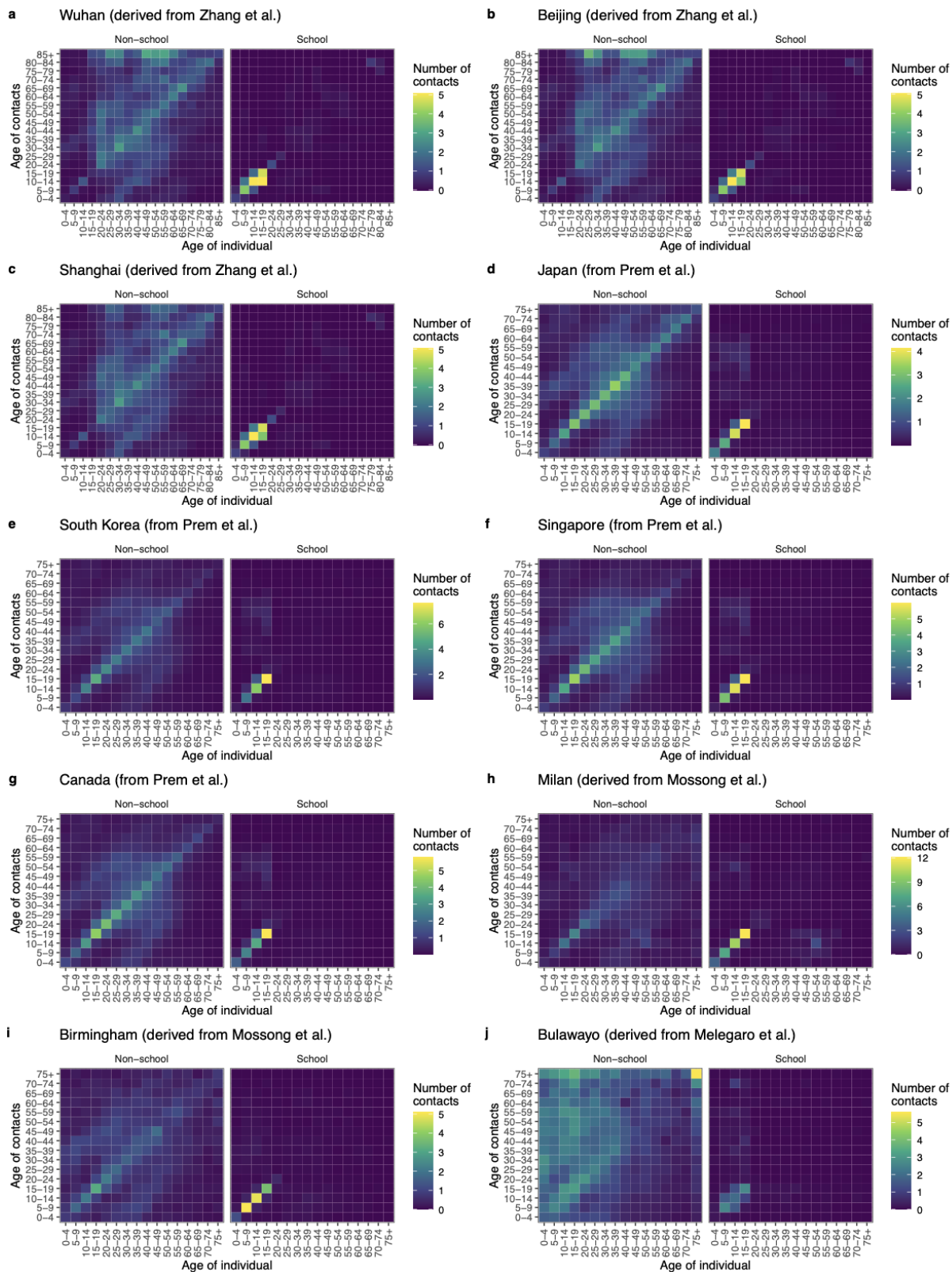
Extended Data Figure 8. Contact matrices used for Figs. 1-3 of the main text. We have not shown matrices for all 12 regions of Italy modelled, nor for all 13 provinces of China modelled, as these show similar patterns to the matrices for Milan and for Wuhan, Beijing and Shanghai, respectively.

11

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

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 $u_i = coss(i|a_y, b_y, a_m, b_m, a_o, b_o)$ (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 $y_i = coss(i|a_y, y_y, a_m, y_m, a_o, y_o)$ (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	$A \sim normal(\mu = 60, \sigma = 20, min = 40, max = 80)$ $A_{range} \sim beta(\alpha = 2, \beta = 2)\ scaled\ to\ 0 - 10$ $A_{min} = A - A_{range}$ $A_{max} = A + A_{range}$
$coss(a|x_1, y_1, x_2, y_2, x_3, y_3)$	Cosine-smoothing function	For a given age a (the midpoint age of age group i) the function evaluates to y_1 for $a \le x_1$, to y_2 for $a = x_2$, and to y_3 for $a \ge x_3$. 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.

Supplementary Table 1. Details of model fitting.

Location	Mixing matrix details
Wuhan City, China	We used mixing matrices measured in Shanghai in 2017/2018[6], 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[6], 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[7], 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[8].
Japan	We used synthetic contact matrices, generated based on demographic information about the country[8].
Singapore	We used synthetic contact matrices based on demographic information about the country[8].
South Korea	We used synthetic contact matrices based on demographic information about the country[8].
Birmingham, UK	We used mixing matrices measured in the UK in 2005/2006[7], 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[9], 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[8].

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

Wuhan: Model 1		
age_y	6	(4.2-7.2)
age_m	55	(46-60)
age_o	64	(60-68)
susc_y	0.003	(0.00014-0.0076)
susc_m	0.044	(0.032-0.054)
susc_o	0.084	(0.079-0.09)
seed_start	19	(16-22)
seed_age	61	(42-79)
seed_age_range	4.9	(1.5-8.9)
qH	1.4	(1.3-1.5)
qL	0.41	(0.3-0.56)
Wuhan: Model 2		
age_y	19	(14-29)
age_m	50	(40-60)
age_o	68	(60-79)
susc	0.055	(0.052-0.059)
symp_y	0.037	(0.0051-0.062)
symp_m	0.3	(0.19-0.42)
symp_o	0.65	(0.52-0.77)
seed_start	16	(14-20)
seed_age	46	(30-67)
seed_age_range	1.3	(0.5-1.9)
qH	1.3	(1.2-1.4)
qL	0.43	(0.31-0.56)
Wuhan: Model 3		
susc	0.046	(0.045-0.048)
seed_start	20	(17-21)
seed_age	64	(37-80)
seed_age_range	4.2	(0.93-8.7)
qH	1.4	(1.3-1.5)
qL	0.33	(0.21-0.42)
Beijing, Shanghai		
susc	0.062	(0.05-0.077)
B_seed_t0	18	(8.7-23)
S_seed_t0	19	(12-25)
seed_d	3.1	(0.74-6.3)
lockdown_t	54	(53-56)
qH	1.3	(0.89-1.8)
qL	0.19	(0.15-0.25)
South Korea		
susc	0.098	(0.087-0.11)
seed_t0	9.2	(4.9-13)
seed_d	3.3	(0.73-6)
lockdown_t	53	(52-54)
qL	0.052	(0.0011-0.1)
Lombardy		
susc	0.084	(0.075-0.096)
conf_mean	7.6	(2.7-13)
conf_shape	11	(3.7-20)
onset_known	0.36	(0.061-0.62)
seed_t0	15	(11-20)
seed_d	3.6	(0.83-6.3)
lockdown_t	50	(47-54)
qL	0.48	(0.28-0.72)

Supplementary Table 3. Posterior means and 95% HDIs from fitting the dynamic transmission model (Figs. 1 and 2, main text).