Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Research paper

Characterizing key attributes of COVID-19 transmission dynamics in China's original outbreak: Model-based estimations

Houssein H. Ayoub a,⁎, Hiam Chemaitelly b,c, China R. Mumtaz d, Shaheen Seedat b,c,e, Susanne F. Awad b,c,e, Monia Makhouli b,c,e, Laith J. Abu-Raddad b,c,e,⁎⁎

⁎⁎ Correspondence to: L. J. Abu-Raddad, Infectious Disease Epidemiology Group, World Health Organization Collaborating Centre for Disease Epidemiology Analytics on HIV/AIDS, Sexually Transmitted Infections, and Viral Hepatitis, Weill Cornell Medicine – Qatar, Cornell University, Qatar Foundation – Education City, Doha, Qatar
⁎ Correspondence to: H. H. Ayoub, Department of Mathematics, Statistics, and Physics, Qatar University, P.O. Box 2773, Doha, Qatar.

a Department of Mathematics, Statistics, and Physics, Qatar University, Doha, Qatar
b Infectious Disease Epidemiology Group, Weill Cornell Medicine-Qatar, Cornell University, Qatar Foundation – Education City, Doha, Qatar
c World Health Organization Collaborating Centre for Disease Epidemiology Analytics on HIV/AIDS, Sexually Transmitted Infections, and Viral Hepatitis, Weill Cornell Medicine–Qatar, Cornell University, Qatar Foundation – Education City, Doha, Qatar
d Department of Epidemiology and Population Health, American University of Beirut, Beirut, Lebanon
e Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York City, NY, USA

A B S T R A C T

A novel coronavirus strain, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in China. This study aims to characterize key attributes of SARS-CoV-2 epidemiology as the infection emerged in China. An age-stratified mathematical model was constructed to describe transmission dynamics and estimate age-specific differences in biological susceptibility to infection, age-assortativeness in transmission mixing, and transition in rate of infectious contacts (and reproduction number \(R_0\)) following introduction of mass interventions.

The model estimated the infectious contact rate in early epidemic at 0.59 contacts/day (95% uncertainty interval – 0.48–0.71). Relative to those 60–69 years, susceptibility was 0.06 in those ≤19 years, 0.34 in 20–29 years, 0.57 in 30–39 years, 0.69 in 40–49 years, 0.79 in 50–59 years, 0.84 in 70–79 years, and 0.88 in ≥80 years. Assortativeness in transmission mixing by age was limited at 0.004 (95% UI = 0.002–0.008). \(R_0\) rapidly declined from 2.1 (95% UI = 1.8–2.4) to 0.06 (95% UI = 0.05–0.07) following interventions’ onset. Age appears to be a principal factor in explaining the transmission patterns in China. The biological susceptibility to infection seems limited among children but high among those >50 years. There was no evidence for differential contact mixing by age.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

An outbreak of a novel coronavirus strain, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified in Wuhan, Hubei province, China, in late December 2019 [1,2]. The outbreak started with identification of four cases of severe pneumonia of unknown etiology, but with symptoms similar to those of Severe Acute Respiratory Syndrome (SARS) and the Middle East Respiratory Syndrome (MERS) [1,3]. Initial cases were linked to exposure at the Huanan Seafood Market, but subsequent infections resulted from rapid community transmission [1–3]. Within about two months, over 80,000 cases and 3000 deaths occurred across China [2,4], amid extreme preventive measures and outstanding healthcare mobilization [2,3]. The resulting disease was named Coronavirus Disease 2019 (COVID-19) by the World Health Organization (WHO) [5], and has been declared a pandemic [6] after affecting tens of countries and territories [4]. For simplicity, we will thereafter refer to this virus as “COVID-19”, though it is the name of the disease form, given its prevalent current use in the public sphere and to avoid confusion between SARS-CoV-2 and SARS-CoV. The aims of this study are to investigate and characterize key attributes of COVID-19 epidemiology as the infection emerged in China including 1) age-specific differences in the biological susceptibility to infection, 2) age-assortativeness in infection transmission, and 3) transition in the rate of infectious contacts (and reproduction number) following the introduction of mass interventions.

Materials and methods

Mathematical model

A deterministic compartmental mathematical model was constructed to describe COVID-19 transmission dynamics in a given
population, and was applied here to the population of China (S1 Fig of Supplementary Information (SI)). The model was designed based on current understanding of the infection’s natural history and epidemiology. Nine age groups were considered, each representing a ten-year age band except for the last category (0–9, 10–19, ..., ≥80 years). For each age group, eight coupled nonlinear differential equations were used to describe population flow across compartments based on infection status, infection stage, and disease stage. The model consisted in total of 72 nonlinear differential equations. Analyses were performed in MATLAB R2019a [7].

Susceptible individuals in each age group are at risk of being exposed to the infection at varying hazard rates, which are age- and time-dependent, to capture the variability in the risk of exposure and the impact of public health interventions. Following a latency period, infected individuals are stratified to develop mild infection followed by recovery, or severe infection followed by severe disease then recovery, or critical infection followed by critical disease and either recovery or disease mortality.

The model parameterized the variation in the rate of infectious contacts through a Woods-Saxon function [8–11] to characterize the transition after China’s robust public health response in terms of its scale or strength, smoothness or abruptness, duration, and the turning point. The model also incorporated an age mixing matrix that allows a range of contact mixing between individuals varying from fully assortative (mixing only with individuals in the same age group) to fully proportionate (mixing with individuals with no preferential bias for a specific age group). The degree of assortativeness in infection transmission mixing by age is a parameter defined through the age mixing matrix that describes the mixing between the different age groups. The latter parameter was estimated through model fitting. Relevant equations pertaining to the age mixing matrix and its components can be found in Section 1 of SI.

Further details on model structure can be found in Section 1 of SI.

Model parameterization

The model was parameterized using current data on COVID-19 natural history and epidemiology. The duration of latent infection was set at 3.69 days based on an existing estimate [12] and reported infection transmission prior to onset of symptoms [14]. The age-stratified proportions of infected individuals that will eventually progress to develop mild, severe, or critical infections were based on the observed distribution of cases across these infection stages in China [3,15,16]. The duration of infectiousness was assumed to last for 3.48 days based on an existing estimate [12] and based on the observed time to recovery in persons with mild infection [3,12] and the observed viral load among infected persons [13,14].

Individuals with severe (or critical) infections develop severe (or critical) disease over a period of 28 days prior to recovery, as informed by the observed duration from onset of severe (or critical) disease to recovery [3]. Individuals with critical disease had the additional risk of disease mortality [17]. The age-specific disease mortality rate was fitted factoring the observed crude case fatality rate in each age group in China as of February 11, 2020 [2,17].

The population size, demographic structure (age distribution), and life expectancy of the population of China, as of 2020, were obtained from the United Nations World Population Prospects database [18].

Further details on model parameters, values, and justifications can be found in S1–S2 Tables and Section 2 of SI.

Model fitting

The model was fitted to the following sources of data: 1) time series of diagnosed COVID-19 cases and of the cumulative number of diagnosed COVID-19 cases and of recovered individuals [3,2] time series of reported COVID-19 deaths and of the cumulative number of COVID-19 deaths [3,3] crude case fatality rate in each age group [2,17], 4) relative attack rate by age, that is the proportion of the population that has already been infected by February 11, 2020 stratified by age [17], 5) proportion of infections in each age group that will progress to be mild infections [3,15,16], 6) proportion of infections in each age group that will progress to be severe infections [3,15,16], 7) proportion of infections in each age group that will progress to be critical infections [3,15,16], China’s reported cases and deaths were adjusted to reflect the change in coronavirus case definition to include, in addition to the laboratory-confirmed cases, those who are clinically-diagnosed [19,20] (Section 2 of SI).

Model fitting was used to estimate the infectious contact rate, nine parameters for the age-stratified susceptibility to the infection, degree of assortativeness in the age group mixing, overall attack rate, overall disease mortality rate, time delay between onset of actual infection and case notification, and between actual death and reported death, and transition in the basic reproduction number $R_0$ (Section 3 of SI). A nonlinear least-square data fitting method, based on the Nelder-Mead simplex algorithm, was used to minimize the sum of squares between data points and model predictions [21]. The model was further used to estimate the susceptibility effect sizes (relative susceptibility) for each age group while accounting for the infection transmission dynamics and the effects of assortativeness in mixing in the population.

Uncertainty analyses

A multivariable uncertainty analysis was conducted to determine the range of uncertainty around model predictions. Five-hundred simulation runs were performed, applying at each run, Latin Hypercube sampling from a multidimensional distribution of the model parameters, where parameter values are selected from ranges specified by assuming ±30% uncertainty around parameters’ point estimates. These parameters included the duration of latent infection, the duration of infectiousness, the duration of severe disease following onset of severe disease, and the duration of hospitalization for critical infection. The model was then refitted to the input data, and the resulting distributions of estimates, across all 500 runs, were used to calculate the model predictions’ means and 95% uncertainty intervals (UIs).

Results

The model fitted the different COVID-19 empirical data such as time-series of diagnosed cases (Fig. 1A), time-series of reported deaths (Fig. 1B), and age-stratified attack rate as of February 11, 2020 (Fig. 1C). The model estimated the epidemic emergence at ~49 days (95% UI: 48–50) prior to January 17, 2020, that is towards the end of November 2019. At the beginning of the epidemic, the predicted infectious contact rate was 0.59 contacts per day (95% UI: 0.48–0.71; S2D Fig of SI). The predicted (average) time delay was 5.4 days (95% UI: 5.2–5.6) between onset of actual infection and reported infection, and 1.6 days (95% UI: 1.5–1.7) between actual death and reported death.

Fig. 2 shows the predicted time evolution of COVID-19 crude case fatality rate (CFR). In the early phase of the epidemic, the crude CFR increased rapidly following the rise in incidence, but plateaued shortly after (towards the end of the first month) and remained so till incidence reached its peak. When incidence started declining (~90 days), the crude CFR grew rapidly, eventually saturating at ~150 days. End of outbreak CFR, estimated through the 500 simulation runs, was 5.1% (95% UI = 4.8–5.4%; S2A Fig of SI).

Fig. 3 features the estimated age-stratified susceptibility profile to COVID-19 infection. Susceptibility was lowest in individuals 0–9 years of age and highest in the 60–69 years age group. Relative to those 60–69 years of age, susceptibility to the infection was only 0.05 in those 0–9 years of age and 0.06 in those 10–19 years of age, but 0.34
in those 20–29 years of age, 0.57 in those 30–39 years of age, 0.69 in those 40–49 years of age, 0.79 in those 50–59 years of age, 0.94 in those 70–79 years of age, and 0.88 in those ≥80 years of age. The uncertainty analysis affirmed these results with narrow uncertainty intervals (S2B Fig of SI).

Fig. 4 illustrates the predicted degree of assortativeness in the age group mixing for each of the 500 simulation runs and model fitting of the uncertainty analysis. The mean degree of assortativeness (mean of the parameter $e_{Age}$ as described in SI) was estimated at 0.004 (95% UI = 0.002–0.008)—there was virtually no assortativeness in infection transmission mixing by age.

Fig. 5 and S2C Fig of SI show the time evolution of $R_0$, and its predicted mean and 95% UI across the 500 uncertainty runs, respectively. In the early phase of the epidemic, $R_0$ was estimated at 2.1 (95% UI: 1.8–2.4), but rapidly declined to 0.06 (95% UI: 0.05–0.07) following the onset of interventions. The sharp transition duration for $R_0$ was estimated at 11.5 days (95% UI: 9.5–13.0).
Discussion

Several key attributes of the epidemiology of COVID-19 have been investigated and estimated. A finding is that the biological susceptibility to the infection appears to vary by age (Fig. 3). Susceptibility to COVID-19 was substantially higher among those >50 years of age compared to those in the younger age groups. For instance, compared to those 60–69 years of age, those ≤19 years of age, 20–29 years of age, and 30–39 years of age were, respectively, 94%, 68%, and 43% less susceptible to being infected. Notably, this age-dependence in the susceptibility to the infection could not be explained by differences in mixing between age groups, as the results indicated limited assortativeness in infection transmission mixing by age (Fig. 4).

These findings support an important role for age in the epidemiology of this infection and affirm other studies suggesting lower susceptibility to the infection at younger age [3,22–29]. Remarkably, the observed attack rate pattern for COVID-19 by age (Fig. 1C) is the inverse (or better complement) of the age-specific cumulative incidence pattern of the 2009 influenza A (H1N1) pandemic (H1N1pdm) infection (S3 Fig of SI) [30]. Presumably, prior recent exposure to other common cold coronaviruses (which are believed to have a similar attack rate pattern to that of H1N1) could be potentially acting as a protective factor (cross immunity) against COVID-19 acquisition (or rapid clearance) at young age [31]. Growing evidence indicates that exposure to other common cold coronaviruses may induce cross-reactive T cell responses and this development of T cell immune memory may be protective against COVID-19 or its severe forms in unexposed individuals [32–34]. An alternative hypothesis has suggested immune imprinting to a similar virus among adults [22]. This being said, the underlying immunological and/or epidemiological factors driving this age effect remain to be investigated with several alternative mechanisms potentially explaining this pattern. For instance, children and young adults may have subclinical infection with low viral load and rapid clearance with or with no transmission potential to others [35]. Of note that the contact tracing data from China suggest that children did not appear to play a significant role in the transmission [3,24,25]. Differences in social behavior and contact patterns by age, which can be very complex [36,37], may also play a role in explaining the variation in the

---

Fig. 3. Model predictions for the age-stratified susceptibility profile to COVID-19 infection, relative to those 60–69 years of age.

Fig. 4. Model predictions for the degree of assortativeness in infection transmission mixing by age across the 500 uncertainty analysis simulation runs. This parameter, defined through the age mixing matrix, describes the mixing between the different age groups. Relevant equations pertaining to the age mixing matrix and its components can be found in Section 1 of Supporting Information.

Fig. 5. Model predictions for the time evolution of the basic reproduction number $R_0$ before and after onset of the interventions in China.
susceptibility to the infection. This being said, additional studies among children exploring viral load, immune response, and social behavior factors are needed before the role of children in community transmission can be ascertained to satisfaction.

Our results indicated that the crude CFR observed in the first three months of the outbreak underestimated the end of outbreak CFR by about 50% (Fig. 2). This is because most infections were still recent infections and have not yet progressed to critical disease or death—death is a late outcome that was estimated to occur 2–8 weeks after onset of symptoms [3].

Another finding of this study is the limited assortativeness in infection transmission mixing by age (Fig. 4), that is equal mixing between the different age groups, which is rather uncommon for respiratory infections [36,37]. This finding is possibly explained by the fact that, with the rapid implementation of stringent lockdown, most transmissions occurred in the context of households rather than of schools, workplaces, or other settings, as supported by existing evidence [3]. Indeed, analysis of 344 clusters in Guangdong and Sichuan provinces indicated that 78–85% of clusters pertained to families [3]. This finding is also supported by analysis of the contacts stratified by age in China [23]. It is, however, important to highlight that while this finding may apply to the Chinese epidemic (given the lockdown measures that may have limited all other transmission pathways such as in schools, workplaces or other settings), it may not necessarily be generalizable to other settings, as existing evidence suggests strong age assortativeness in the transmission dynamics of respiratory infections [36,37]. Indeed, we recently observed strong age assortativeness when we modeled the COVID-19 epidemic in Qatar [38].

The present study affirmed the impact and success of the drastic preventive measures in curtailing infection transmission. R0 was sharply reduced by 97% over a short duration (Fig. 5). At the beginning of the epidemic, on average, each infected person had 0.6 infectious contacts per day, that is each person passed the infection to 0.6 persons per day (S2D Fig of S1). The rate of infectious contacts can be expressed roughly as \( c > p \), where \( c \) is the number of “social” contacts conducive to COVID-19 transmission per day and \( p \) is the transmission probability of the virus in a single contact. While \( p \) is unknown, it is possibly in the order of 1–5%, as suggested by contact tracing data from China—1–5% of contacts were subsequently laboratory-confirmed as COVID-19 cases [3]. This implies that, on average, each infected person had somewhere between 10 and 60 contacts per day at the beginning of the epidemic, but very few contacts after the lockdown. The latter further affirms the role of the lockdown in severely cutting the contact rate, making sustainable infection transmission very difficult. This finding demonstrates how strong action by policymakers can have an immense impact on limiting infection spread and sequelae.

This study has limitations. Model projections are contingent on the quality and representativeness of the input data. For instance, we assumed infection levels to be as officially documented, but evidence suggested that many infections may have been undocumented [12]. Case ascertainment could also be age dependent. The natural history of this infection is not yet firmly established, and the case management protocols have evolved over time [19,20]. Mortality data seem to suggest that the standard of care improved over time especially in recent weeks when the healthcare sector was no longer overwhelmed with a large case load. We conducted this study using early epidemic data for one specific country, but our knowledge of the epidemic has been rapidly expanding, and thus an alternative broader view of the epidemiology may emerge with accumulation of data. For example, differences in susceptibility among adults may not be as large as those estimated based on China’s early epidemic data. We used a deterministic compartmental model, but this type of model may not be representative of the stochastic transmission dynamics when the number of infections is small, such as in the very early phase of the epidemic, thereby adding uncertainty to our estimate for the day of outbreak emergence. Despite these limitations, our parsimonious model, tailored to the nature of available data, was able to reproduce the COVID-19 epidemic as observed in China, and provided insights about infection transmission and disease progression in the population.

In conclusion, age appears to be a principal factor in explaining the patterns of COVID-19 transmission dynamics in China. The biological susceptibility to the infection seems limited among children, intermediate among young adults and those mid-age, but high among those >50 years of age. There was no evidence for considerable differential contact mixing by age, consistent with most transmission occurring in households rather than in schools or workplaces. Further mathematical modeling research is needed to understand the COVID-19 epidemics in other countries, and to draw further inferences about the global epidemiology of this infection.

Data availability

All data generated or analysed during this study are included in this article and its Supplementary Material.

Funding

The modeling infrastructure was made possible by NPRP grant number 9-040-3-008 from the Qatar National Research Fund (a member of Qatar Foundation). GM acknowledges support by UK Research and Innovation as part of the Global Challenges Research Fund, grant number ES/P010873/1. The statements made herein are solely the responsibility of the authors. The authors are also grateful for support provided by the Biostatistics, Epidemiology, and Biomathematics Research Core at Weill Cornell Medicine-Qatar. The publication of this article was funded by Qatar National Library.

Author contributions

Conceptualization, Laith J. Abu-Raddad; Data curation, Houssein Ayoub, Hiam Chemaitelly and Laith J. Abu-Raddad; Formal analysis, Houssein Ayoub and Laith J. Abu-Raddad; Funding acquisition, Laith J. Abu-Raddad; Investigation, Houssein Ayoub; Methodology, Houssein Ayoub and Laith J. Abu-Raddad; Project administration, Laith J. Abu-Raddad; Supervision, Laith J. Abu-Raddad; Writing – original draft, Hiam Chemaitelly; Writing – review & editing, Houssein Ayoub, Hiam Chemaitelly, Ghina Mumtaz, Shaheen Seedat, Susanne Awad, Monia Makhoul and Laith J. Abu-Raddad.

Declaration of Competing Interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2020.100042.

References

[1] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;172:577–82. https://doi.org/10.7326/M20-0504 (PubMed PMID: 32150748).

[2] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020. https://doi.org/10.1001/jama.2020.2646 (PubMed PMID: 32091531).

[3] World Health Organization. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). Available from: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf; 2020.
COVID-19 outbreak live update. Available from: https://www.worldometers.info/coronavirus/; 2020. (Accessed: March 14, 2020).

World Health Organization (WHO). Naming the coronavirus disease (COVID-19) and the virus that causes it. Available from: https://www.who.int/ emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-19)-and-the-virus-that-causes-it; 2020. (Accessed: March 11 2020).

World Health Organization (WHO). WHO Director-General’s opening remarks at the media briefing on COVID-19 - 11 Mar 2020. Available from: https://www.who.int/ dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020; 2020. (Accessed: March 14, 2020).

MATLAB®. The language of technical computing. The MathWorks, Inc; 2019.

World Health Organization (WHO). Age specification for a small proportion of SARS-CoV-2 infections and do not exhibit a higher viral load than adults. Eur J Clin Microbiol Infect Dis. 2020; 39: 1983–7. https://doi.org/10.1007/s10096-020-03900-0 (PubMed PMID: 32845413; PubMed Central PMCID: PMCPMC447989).

Macartney K, Quinn HE, Pillbury AJ, Koestala A, Deng L, Winkler N, et al. Transmission of SARS-CoV-2 in Australian educational settings: a prospective cohort study. Lancet Child Adolesc Health. 2020. https://doi.org/10.1016/S2532-4642(20)30251-0 (PubMed PMID: 32758454; PubMed Central PMCID: PMCPMC398855).

Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. medRxiv. 2020. https://doi.org/10.1101/2020.07.19.20157362 (PubMed PMID: 32743609; PubMed Central PMCID: PMCPMC3786533).