Transmission roles of symptomatic and asymptomatic COVID-19 cases: a modelling study

Jianbin Tan1,*, Yang Ge2,*, Leonardo Martinez3, Jimin Sun4, Changwei Li5, Adrianna Westbrook6, Enfu Chen4, Jinren Pan4, Yang Li7, Wei Cheng4, Feng Ling4, Zhiping Chen4, Ye Shen6 and Hui Huang1

1School of Mathematics, Sun Yat-Sen University, Guangzhou, China; 2School of Health Professions, University of Southern Mississippi, Hattiesburg, Mississippi, USA; 3Department of Epidemiology, Boston University, Boston, USA; 4Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China; 5Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA; 6Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia, USA and 7School of Statistics, Renmin University of China, Beijing, China

Abstract

Coronavirus disease 2019 (COVID-19) asymptomatic cases are hard to identify, impeding transmissibility estimation. The value of COVID-19 transmissibility is worth further elucidation for key assumptions in further modelling studies. Through a population-based surveillance network, we collected data on 1342 confirmed cases with a 90-days follow-up for all asymptomatic cases. An age-stratified compartmental model containing contact information was built to estimate the transmissibility of symptomatic and asymptomatic COVID-19 cases. The difference in transmissibility of a symptomatic and asymptomatic case depended on age and was most distinct for the middle-age groups. The asymptomatic cases had a 66.7% lower transmissibility rate than symptomatic cases, and 74.1% (95% CI 65.9–80.7) of all asymptomatic cases were missed in detection. The average proportion of asymptomatic cases was 28.2% (95% CI 23.0–34.6). Simulation demonstrated that the burden of asymptomatic transmission increased as the epidemic continued and could potentially dominate total transmission. The transmissibility of asymptomatic COVID-19 cases is high and asymptomatic COVID-19 cases play a significant role in outbreaks.

Introduction

Coronavirus disease 2019 (COVID-19), caused by the novel coronavirus (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) [1], is a great threat to human health [2]. COVID-19 patients may present and remain pre-symptomatic, asymptomatic or symptomatic and transmission may occur at each of these disease states [3–5]. Unlike the transmission caused by symptomatic cases, pre-symptomatic and asymptomatic transmission are hard to detect and difficult to measure as many surveillance systems rely on symptom-based population screening [3, 6–8]. Previous case studies suggested that asymptomatic COVID-19 individuals are less infectious than symptomatic cases [9, 10]. However, asymptomatic cases may spread for a longer period due to reduced efficiency in case detection [11]. How much asymptomatic transmission had contributed to past outbreaks is challenging to quantify and not well studied [12, 13]. Several studies investigated undetected transmission of SARS-CoV-2, but have presented contradictory conclusions with estimated burden ranged from 3% to 79% [11, 14, 15]. Without sufficient follow-up time, asymptomatic and pre-symptomatic cases are often indistinguishable [16]. Consequently, studies using population-level data to estimate of age-specific transmission and susceptibility parameters commonly falls short of accuracy which potentially explains for the heterogeneous findings from different studies [17–19]. Common issues were modelling without data on observed asymptomatic infection [11, 14, 15, 20] and the inclusion of pre-symptomatic cases as part of an asymptomatic classification [17–19]. Meanwhile, few studies assessing asymptomatic infectiousness and viral load with limited sample sizes fail to capture the transmission dynamics [9, 10, 21–28]. A comprehensive understanding of the age-specific symptomatic and asymptomatic transmission dynamics at the population level is essential to the evaluation of an epidemic and the creation of responding health policies.

In this study, we used reliable case symptom classification, social contact measures, and susceptibility parameters at the population level to learn the transmission dynamics. We report on a longitudinal cohort of all diagnosed COVID-19 infections, between 8 January and 23 February 2020, from Zhejiang province, China. All patients without initial symptoms were
followed by at least 90 days to distinguish between asymptomatic and pre-symptomatic cases, an essential procedure rarely implemented by previous studies to ensure the reliable classification of case symptoms. We then built age-stratified compartmental models to study the age-dependent population-level transmission roles of symptomatic and asymptomatic COVID-19 cases.

Methods

Data sources

Zhejiang province is an eastern coastal province adjacent to Shanghai city with a population of approximately 54 million individuals [29]. The first and only major wave of the COVID-19 epidemic in Zhejiang began on early January 2020 and continued until late February 2020 after which only sporadic single-case events were observed [30, 31]. We included information from all confirmed cases in this major wave (a total of 1342 cases), as well as a follow-up investigation related to all detected asymptomatic infections to distinguish between asymptomatic and pre-symptomatic cases. All COVID-19 cases were microbiologically confirmed through positive reverse transcriptase–polymerase chain reaction (RT-PCR) test results. Individual-level data related to the symptom onset of symptomatic infections, as well as COVID-19 confirmation dates and ages of both symptomatic and asymptomatic cases were collected. On 23 January 2020, the provincial government changed its infectious disease alert category to the highest level and, on 1 February, began a comprehensive set of interventions [32]. As of 10 April 2020, the date in which we restricted our data for this analysis, no additional outbreak had been observed. Trained health professionals interviewed each confirmed case with a predefined questionnaire by which basic health and demographic information were collected.

Definition of symptomatic and asymptomatic cases

All confirmed cases and their close contacts were isolated or quarantined after being identified through contact tracing. During the isolation/quarantine period, cases and their contacts received regular testing and daily symptom screening for fever, cough, and shortness of breath. Tests for case confirmation were conducted using reverse transcription polymerase chain reaction (RT-PCR) or viral genome sequencing on samples from throat swabs (oropharynx and nasopharynx). If a case or contact had a positive test result but without any symptoms, they would be temporarily classified as an asymptomatic/pre-symptomatic case at the time. All cases were followed for at least 90 days after their initial positive test to distinguish between asymptomatic and pre-symptomatic cases. Among these subjects, those who developed symptoms later would receive a final classification as a symptomatic case. Others who had never developed any symptoms between their initial positive test and first subsequent negative PCR test would be classified as asymptomatic cases.

Model structure

We divided the total population of Zhejiang province into seven age groups (Fig. 1). To consider transmission related to symptomatic and asymptomatic infections among different age groups, our model contained 8 compartments for the ith age group: susceptible population (S), exposed contacts (E), pre-symptomatic cases (Ips, infected but have not yet developed symptoms), symptomatic cases (I symptomatic till confirmed/recovery) and removed/recovery groups (Rps, Rs, Rp). We assumed new infections were driven by transmission from compartments of Ips, I symptomatic and Rps in all age groups. (See Supplementary material for further details).

Asymptomatic cases (I symptomatic) were infections without typical symptoms (e.g. respiratory disease symptoms) which were often untraceable in the clinical survey and, therefore, their contribution to population-level transmission would be underestimated. To account for this, we assumed only a proportion of asymptomatic infections were detected (Rps), while others (Rps) would be unconfirmed. We considered the unconfirmed cases as those infected with SARS-CoV-2 but were not detected and confirmed by tests. Besides, we were able to observe the disease confirmation date but not the date of infection for the period from cases becoming infectious to the diagnosis of COVID-19 (A') in those with confirmed asymptomatic infection. Based on the virus shedding pattern of asymptomatic infections reported in previous studies [8, 10, 33, 34], we assumed that this period should be less than 30 days, after which virus shedding generally ceases, and infection is no longer detectable through pathogen-specific testing.

To identify age-varying transmissibility and susceptibility [20], we assumed a time-varying curve for the average contact numbers of ith age group with jth age group (cij) to characterize the effect of policy interventions, which is estimated with the contact matrix between age groups through surveys conducted in Shanghai [35, 36]. We separate the probability of infection into two..
components: transmissibility (T) and susceptibility (s). We define transmissibility (T) as the infectiousness of one case. Similarly, we define susceptibility (s) as the probability of acquiring infection from an infectious case (T = 1). Therefore, s = 0 corresponds to a situation in which the susceptible individuals are immune to the disease. We assumed that case transmissibility would depend on age and the presence of symptoms. To capture the age-dependent pattern, B splines basis functions were used to model the variability in age-varying transmissibility smoothly, conditioning on a pre-specified informative prior of susceptibility parameters. Finally, under the Bayesian framework, the compartmental model was fitted to the daily new symptomatic and asymptomatic meters. Finally, under the Bayesian framework, the compartmental model was fitted to the daily new symptomatic and asymptomatic cases in Zhejiang province (Fig. S3 in Supplementary material for each age group with Markov Chain Monte Carlo (MCMC) algorithm.

All analyses were implemented in R version 3.5.1. Packages of deSolve [37], extraDistr [38] and splines [39] were used for model fitting. The modelling framework, posterior distributions of some parameters, and a model assessment procedure for fitting data is given in the supplementary material. Unless stated otherwise, the medians of the posterior distributions were used as the point estimators of parameters and simulated numbers. More details of the model are given in the supplementary material.

Results

Transmissibility

The estimated transmissibility presented an age-dependent difference between symptomatic and asymptomatic infections (Fig. 2). While the transmission of symptomatic cases monotonically increased with increasing age, the transmissibility of asymptomatic infection remained low until age 40, after which point it significantly increased with increasing age. The age-varying ratios of the two kinds of transmissibility indicated asymptomatic cases were, on average, 66.72% lower in transmission than symptomatic cases. However, the difference between the two types of infections was not as big in those aged 0–20 and 60+ years old, but became more obvious in the middle-aged group where the ratios were as low as 24.42% and 23.38% for those aged 30–40 and 40–50 years old, respectively.

The proportion of asymptomatic cases

In Figure 3, the proportion of asymptomatic cases ($\frac{R_t}{R_t+R_c}$) estimated by our model was much larger than what was observed in the data. The average proportion of asymptomatic cases was 28.22% (95% CI 22.97–34.56) of the total counts of cases in our model estimation, but was 9.24% in the observed data ($\frac{R_c}{R_t+R_c}$). In our estimation from the empirical data, the highest proportion of asymptomatic cases was among 0–10 (60.18% (95% CI 53.61–66.99)) and 10–20 (57.64% (95% CI 47.45–66.98)) years old groups. For asymptomatic cases, we further estimated the proportion of cases that failed to be detected ($\frac{R_c}{R_t+R_c}$). In the posterior samples, the average proportion of unconfirmed cases in all asymptomatic infections was 74.10% (95% CI 65.85–80.72). The maximum proportion of unconfirmed cases was observed in 20–30 years old at 86.59% (95% CI 73.64–92.19).

Symptomatic and asymptomatic transmission

To explore the impact of symptomatic and asymptomatic transmission, we present several features of the estimated dynamic of the epidemic and the transmission burden caused by symptomatic and asymptomatic cases in Figure 4. The estimated number of daily new transmissions reached a peak around ten days prior to the peak of the daily reported new confirmed cases (Fig. 4a). We estimated a substantial number of undetected asymptomatic cases (109 (95% CI 73–164)) were infected before the first asymptomatic case was diagnosed (27 January) (Fig. 4b). New transmissions were nearly eliminated by 2 February 2020 (Fig. 4a), when a comprehensive set of restrictions had been implemented. The peak of the two types of transmission both occurred between 18 and 22 January (Fig. 4c). The average burden of asymptomatic transmission during the major outbreak period was estimated to be 12.86% (95% CI 7.54–19.27). The burden of asymptomatic transmission increased with time, ranging from 7.77% to 16.03% (Fig. 4d). Simulation studies were conducted to investigate the dynamic changes in the transmission burden over time during a prolonged epidemic (Supplementary Fig. S7). When the duration of the decreasing process of the contact function (represented by ‘m’ in Supplementary Fig. S1) was prolonged by two weeks and each individual’s daily contact number was increased by one person during the outbreak period, we found a slower decreasing trend in daily new cases infected by asymptomatic cases compared with that contributed by symptomatic cases (Supplementary Fig. S7, scenario 1). Additional scenarios were generated demonstrating the possibility of asymptomatic transmission dominating the total transmission under different conditions, especially when the duration between symptom onset and
Fig. 3. The proportion of asymptomatic infections and unconfirmed asymptomatic infections until 22 February 2020, for seven age groups. The estimated proportions of asymptomatic cases, the proportions of cases that failed to be detected among asymptomatic infections (unconfirmed proportions), and the observed proportions of asymptomatic cases are defined as: \( \frac{R_i}{R_i + R_{ic}} \), \( \frac{R_{ih}}{R_{ih} + R_{ihc}} \), and \( \frac{R_{ihc}}{R_{ihc}} \), respectively. The 95% credible intervals for the estimated proportions of asymptomatic cases are shown for each age group.

Fig. 4. The estimated dynamics of the epidemic and the transmission burdens from symptomatic and asymptomatic cases. (a): The estimated numbers of daily new transmissions with 95% credible intervals and the observed numbers of daily reported new confirmed cases from 8 January to 22 February 2020; (b) The observed numbers of daily reported new confirmed symptomatic (\( R_{cs} \)) and asymptomatic cases (\( R_{ca} \)) and the estimated numbers of daily new cases that failed to be detected (\( R_h \)) with 95% credible intervals; (c): The estimated numbers of infected individuals caused by symptomatic and asymptomatic transmission over time, with 95% credible intervals; (d): Corresponding proportions of symptomatic and asymptomatic transmissions over different time periods.
disease confirmation for symptomatic infections was shortened and the asymptomatic infections were not controlled (Supplementary Fig. S7, scenarios 2 and 3).

**Age-depended transmission**

Within each age group, we observed heterogeneous transmission contributions during different time periods (Fig. 5a). Early on in the epidemic, the transmission burden was dominated by persons of 50–60 years old (32.75% from 8 January to 12 January), but the proportion of transmission contribution from people over 60 years old significantly increased over time, surpassing the 50–60 years old and reaching 30.42% by 1 February 2020. The proportion of transmission contribution among varying age groups was distinct between symptomatic and asymptomatic cases (Fig. 5b). The majority of both symptomatic and asymptomatic transmissions were contributed by persons over 30 years old (Supplementary Table S10). Individuals below 30 years old only contributed less than 5% of all symptomatic transmission and approximately 12% of all asymptomatic transmission, respectively, despite representing almost 40% of the entire population. Contributions to asymptomatic transmission among 20–30 and >60 year age groups (9.44% and 31.73%, respectively) were substantially higher than their corresponding contributions to symptomatic transmission (3.77% and 26.55%, respectively). To further understand possible age-dependent vaccination strategies, a simulation of seven scenarios was conducted to assess the percentage decline in different age groups if one age group were to achieve 100% immunity by vaccinations (Supplementary Fig. S8). The results suggested that vaccinations targeting age groups above 30 years are likely to be more effective at the population level, with the most percentage decline of cases from the entire population achieved by targeting the 50–60 years old group. Meanwhile, vaccinating those younger than 30 years old are more likely to benefit their own age groups.

**Discussion**

In our study, we found that asymptomatic cases were over 60% less infectious compared to symptomatic cases. While great efforts
like mass screening and strict contact tracing were conducted, our results suggested that a large proportion of asymptomatic infections were still not detected [40]. The burden of asymptomatic transmission was inferior in the early outbreak but could become higher with the continuous spread of COVID-19. Current evidence suggests that asymptomatic COVID-19 cases are generally less infectious [9] than cases with symptoms. We found that this difference may partially be explained by the patient age. Age may directly impact COVID-19 transmission through virus shedding patterns [10] as discussed in previous studies [41]. Symptoms are commonly mild in children [42] but severe in the elderly [43]. While still debatable [44], higher severity has been associated with increased shedding of the virus [45]. In our study, symptomatic and asymptomatic cases were most infectious in individuals 60 years old or older. In contrary to the monotonic increasing association between age and transmission in symptomatic cases, there was a plateau of a low degree of transmission in young asymptomatic infections. We suspect older adults are not only the most vulnerable to succumb to COVID-19 but also may be more likely to transmit once infected, regardless of symptom status.

Similar to previous studies, our results suggest a small proportion of asymptomatic cases have been detected since the start of the COVID-19 pandemic in our setting [46–48]. Symptom-based screening has limited capability in asymptomatic case detection [12]. One meta-analysis has shown a similar result that the proportion of asymptomatic infection among all confirmed cases in Asia was about 27.58% (95% CI 13.60–41.57) [49]. While mass pathogen or immunological-based testing at the population-level consumes tremendous health resources, and thus is not feasible in most settings. Considering these challenges, age-dependent screening strategies may be more practical. We found that the highest number of undetected asymptomatic cases was among young adults aged 20–30 years old (Supplementary Table S8) and the corresponding transmission contribution was significantly higher than that of symptomatic case (Fig. 5b). The current strategy cannot identify all asymptomatic infections. Therefore, a tailored strategy for better asymptomatic infection detection is needed in the future.

Based on the estimated transmission contributions from symptomatic and asymptomatic infections, roughly 13% of infections were associated with asymptomatic transmission and that percentage continuously increased with a prolonged period. The overall burden of transmission was mainly contributed by symptomatic cases at the beginning of the epidemic, but asymptomatic infections appeared to have increasing percentages of subsequent cases later on. Additional simulations suggested that the transmission burden could even be dominated by asymptomatic transmissions under certain circumstances (Supplementary Fig. S7). Therefore, the spreading potential of asymptomatic cases cannot be ignored, especially in the later stages of the epidemic. Meanwhile, potential differences in transmission burden by age groups, as shown in Figure 5 and Supplementary Figure S8, supports prioritizing age-dependent prevention and control strategies when facing strained resources. As the larger contributor to the transmission of COVID-19, the older age population is not only a highly vulnerable group but should also be the primary target for prevention strategies. There are several limitations in this study. First, we do not have information on SARS-CoV-2 variant status of index cases. Further studies on emerging SARS-CoV-2 variants are warranted. Second, data collection likely missed potential cases of the epidemic, despite intensified efforts devoted by the local investigation team to trace contacts. Due to this, we introduced a compartment in our model ($R_i$) to adjust for poor case ascertainment and missing cases. Third, transmissibility and susceptibility were two factors related to symptomatic and asymptomatic transmission estimation and can be difficult to capture simultaneously. We used the susceptibility estimates from a previous study [20] as priors in our model to account for this parameter identification problem. The used susceptibility parameter is consistent with other observations including the Chinese population that individuals over 20 years-old are roughly twice as susceptible as those below 20 years-old [50]. Additionally, the contact survey data we used in our model were obtained in Shanghai, a city adjacent to Zhejiang province. Although the two regions share a similar culture and modes of social activities, there were potential uncertainties associated with the discrepancies in contact matrices. To address this limitation, we introduced correction parameters in our model for partially adjusting the uncertainties. However, we acknowledge that the adopted contact matrices for the analysis of the epidemic data in Zhejiang province were crucial for the parameter estimation under our modelling framework, and their uncertainties may affect the results. To assess their potential impacts, we proposed several sensitivity analyses in the supplementary material. Overall, our model was heavily driven by the contact data prior to the lockdown period, but less sensitive to the unknown structure of the contact matrix in Zhejiang during the lockdown period. Finally, due to the limited data and sample size, we did not stratify the transmissibility rate by comorbidities and disease severity.

In summary, our results suggest that individual-level transmissibility of COVID-19 increases with patient age. While asymptomatic cases are difficult to trace, the burden of asymptomatic transmission is still sizable and should not be ignored.

**Supplementary material.** The supplementary material for this article can be found at [https://doi.org/10.1017/S0950268822001467](https://doi.org/10.1017/S0950268822001467).

**Acknowledgements.** Not applicable.

**Author contributions.** All authors contributed to the study’s conception and design. Material preparation and data collection were performed by Wei Cheng, Zhiping Chen, Enfu Chen, Jinren Pan, Feng Ling, Jimin Sun. Material preparation and data analysis were performed by Jianbin Tan, Yang Ge, Leonardo Martinez, Changwei Li, Adrianna Westbrook, Yang Li, Hui Huang, Ye Shen. The first draft of the manuscript was written by Jianbin Tan, Yang Ge, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Financial support.** This study was partially supported by the Zhejiang Basic Public Welfare Research Project (Grant No. LGF21H260003, PI: Feng Ling), National Natural Science Foundation of China (Grant No. 11871485, PI: Hui Huang), and the Medical and Health Research Project of Zhejiang Health Commission (Grant No. 2021K1629, PI: Wei Cheng). Yang Ge was supported by the Start-up Grant from the University of Southern Mississippi. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Conflict of interest.** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Ethical standards and Consent to participate.** The research protocol was approved by the institutional review board at the Zhejiang Provincial Center for Disease Control and Prevention. This study was part of surveillance during pandemics. Data was maintained by Zhejiang Provincial Center for Disease Control and Prevention and was shared with the analysis group after de-identification, as such, consent was waived by the institutional review board at the Zhejiang Provincial Center for Disease Control and Prevention. All methods were carried out in accordance with relevant guidelines and regulations.
Consent for publication. Not applicable.

Authors’ information. Sun Yat-Sen University, School of Mathematics, Guangzhou, China: Jianbin Tan, Hui Huang; School of Health Professions, University of Southern Mississippi, Hattiesburg, Mississippi, United States: Yang Ge; University of Georgia, College of Public Health, Department of Epidemiology and Biostatistics, Athens, Georgia, United States: Adrianna Westbrook, Ye Shen; Boston University, Department of Epidemiology, Boston, United States: Leonardo Martínez; Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China: Jumin Sun, Enfu Chen, Jinren Pan, Feng Ling, Wei Cheng, Zhiping Chen; Tulane University School of Public Health and Tropical Medicine, Department of Epidemiology, New Orleans, Louisiana, United States: Changwei Li; Renmin University of China, School of Statistics, Beijing, China: Yang Li.

Data availability statements. The data used in the analysis contains the age and COVID-19 symptom onset/disease confirmation dates of each case. It is possible to reveal the patients’ identities by linking this individual-level information to other sources such as local media coverage or social media release. Therefore, the data were not made publicly available. However, we will provide de-identified data upon request for research purposes (Dr Zhiping Chen, zhuchen@cdc.zj.cn).

References

49. Ma Q et al. (2021) Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis: a systematic review and meta-analysis. JAMA Network Open 4, e2137257.
50. Viner RM et al. (2020) Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. JAMA Pediatrics 175(2), 143–156.