Received: 3 February 2022 | Revised: 16 November 2022 | Accepted: 25 November 2022

DOI: 10.1111/biom.13814

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Age-related model for estimating the symptomatic and asymptomatic transmissibility of COVID-19 patients

Jianbin Tan[1] | **Ye Shen**[2] | **Yang Ge**[3] | **Leonardo Martinez**[4] | **Hui Huang**[1]

[1]School of Mathematics, Sun Yat-sen University, Guangzhou, China

[2]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia, USA

[3]School of Health Professions, University of Southern Mississippi, Hattiesburg, Mississippi, USA

[4]Department of Epidemiology, Boston University, Boston, Massachusetts, USA

Correspondence
Hui Huang, School of Mathematics, Sun Yat-sen University, Guangzhou, China.
Email: huangh89@mail.sysu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 11871485

Abstract

Estimation of age-dependent transmissibility of COVID-19 patients is critical for effective policymaking. Although the transmissibility of symptomatic cases has been extensively studied, asymptomatic infection is understudied due to limited data. Using a dataset with reliably distinguished symptomatic and asymptomatic statuses of COVID-19 cases, we propose an ordinary differential equation model that considers age-dependent transmissibility in transmission dynamics. Under a Bayesian framework, multi-source information is synthesized in our model for identifying transmissibility. A shrinkage prior among age groups is also adopted to improve the estimation behavior of transmissibility from age-structured data. The added values of accounting for age-dependent transmissibility are further evaluated through simulation studies. In real-data analysis, we compare our approach with two basic models using the deviance information criterion (DIC) and its extension. We find that the proposed model is more flexible for our epidemic data. Our results also suggest that the transmissibility of asymptomatic infections is significantly lower (on average, 76.45% with a credible interval (27.38%, 88.65%)) than that of symptomatic cases. In both symptomatic and asymptomatic patients, the transmissibility mainly increases with age. Patients older than 30 years are more likely to develop symptoms with higher transmissibility. We also find that the transmission burden of asymptomatic cases is lower than that of symptomatic patients.

KEYWORDS

age-dependent heterogeneity, asymptomatic transmission, Bayesian epidemic model, COVID-19, ordinary differential equations, shrinkage prior

1 | INTRODUCTION

COVID-19, caused by the novel coronavirus (SARS-CoV-2) (Guan et al., 2020), is a pronounced threat to human health (World Health Organization, 2020). Understanding the shedding ability of persons infected with COVID-19 (i.e., transmissibility) is essential to nonpharmaceutical interventions such as social distancing, case isolation, and contact quarantines (Ge et al., 2021; Lewnard & Lo, 2020; Prem et al., 2020). COVID-19 patients may present as pre-symptomatic, asymptomatic, or symptomatic, and transmission may occur in each of these states (Hoehl et al., 2020; Rothe et al., 2020). Previous studies suggest that asymptomatic COVID-19 individuals are less infectious than symptomatic cases (Gao et al., 2020; Kim et al., 2020) and such infectiousness may be associated with both age and clinical manifestation (Zhou et al., 2020). However, the age-dependent transmissibility of COVID-19

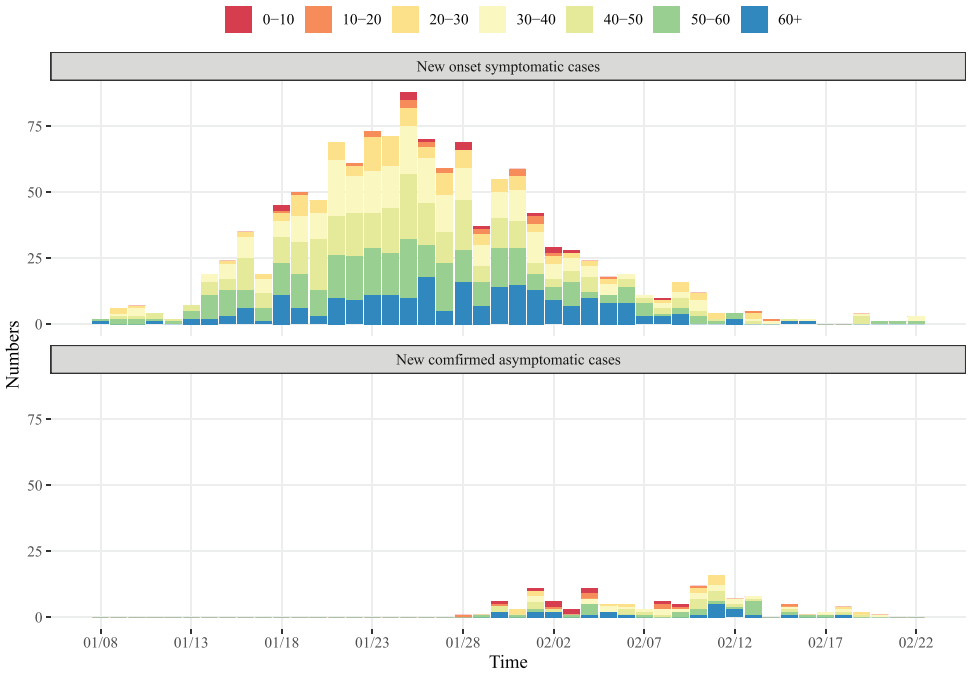

FIGURE 1 The observed daily new onset symptomatic cases and new confirmed asymptomatic cases in seven age groups from January 8 to February 22. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.

patients, especially those that are asymptomatic, is yet to be quantified.

The transmission rate from infectious patients to susceptible individuals is the number of infections per unit of time given contact with their social network, which is associated with both the total contact number per unit of time and the transmission probability (the probability that a susceptible individual gets infected from his/her contact) (Bjørnstad et al., 2020; Rock et al., 2014). Since the transmission of COVID-19 is age-dependent (Gao et al., 2020; Kong et al., 2020; Verity et al., 2020), Prem et al. (2020) proposed a susceptible-exposed-infectious-removed (SEIR) model with age-stratified contact patterns; however, this model did not account for age heterogeneity in transmission probabilities for the constriction of transmission rates. To adjust for this issue, Davies et al. (2020) and Keeling et al. (2021) presumed age-dependent transmission probabilities in their SEIR models by considering age-varying susceptibility. In these studies, however, the transmission probabilities were not assumed to be related to the susceptible contact and the infectious host; age-dependent transmissibility and susceptibility thus could not be captured simultaneously. Such simplifications in transmission mechanism are inconsistent with current empirical evidence (Davies et al., 2020; Viner et al., 2021; Zhou et al., 2020), making model outputs difficult to interpret and largely deficient to age-specific variability.

Nevertheless, there are several challenges in addressing this issue even when an age-structured epidemic dataset (as shown in Figure 1) is available. First, most existing surveillance systems cannot reliably distinguish between symptomatic and asymptomatic COVID-19 cases due to the need for long-term follow-up (Hu et al., 2020; Huang et al., 2020; Meng et al., 2020). Second, age-dependent transmissibility needs to be identified based on known susceptibility patterns and contact strength. Finally, evaluating the age-dependent transmissibility of COVID-19 patients requires the consideration of transmission mechanisms of infectious states among different symptoms and age groups.

To overcome the above challenges, we use a dataset in which all diagnosed patients were followed for at least 90 days to distinguish between asymptomatic and symptomatic cases. This study design is uncommon in previous studies and ensures a reliable classification of case symptoms. In addition, we modify a Bayesian age-related SEIR model to integrate patient data, social contact data, and other epidemic information from the literature. In our model, we assume an age-dependent transmissibility and susceptibility mechanism among contacts in different age groups and, therefore, transmission probabilities of contacts are age-dependent for both infector and their infectees. This mechanism allows greater flexibility in capturing complex heterogeneity in data. Furthermore, for robustly estimating age-dependent transmissibility, we assume smooth changes across age groups and impose shrinkage priors to penalize roughness. We divide compartments from the SEIR model into multiple

states to consider the distinction between symptomatic and asymptomatic transmissibility. A time-varying transmission mechanism within the symptom-age-stratified compartments is then constructed based on an ordinary differential equation framework, with a closed-form approximation for the latent trajectories to avoid computationally intensive numerical calculations.

This paper is organized as follows. We introduce our data and propose the Bayesian age-related SEIR model in Section 2. Comparisons of our model with more-parsimonious basic models are made through simulation studies in Section 3 to motivate the use of the age-related SEIR model. We then apply our proposed method to the real data in Section 4. The conclusions and limitations of our model are discussed in Section 5.

2 | METHODOLOGIES

2.1 | Data source

The first major wave of the COVID-19 epidemic in Zhejiang province in China began on January 8, 2020, and continued until February 22, 2020, after which only sporadic single-case events were observed. We collected information from all confirmed patients in this wave (a total of 1,342 individuals), as well as a follow-up investigation related to all patients to distinguish between asymptomatic and pre-symptomatic cases; see Part A of Supporting Information for more details. Collected data included individual-level symptom onset dates of symptomatic infections, confirmation dates of symptomatic and asymptomatic infections, and ages of symptomatic and asymptomatic patients. We calculated the individual-level symptomatic transmission periods of symptomatic infections by subtracting their symptom onset dates from their confirmation dates. After obtaining the data on symptomatic transmission periods, we only focus on the symptom onset dates of symptomatic infections and the confirmation dates of asymptomatic infections, which are illustrated in Figure 1. The peak of confirmed symptomatic patients occurred around January 25, 2020, while asymptomatic infections were reported after January 28, 2020. It is worth noting that all patients in this period are infected by early types of SARS-CoV-2 virus.

For social contact data, we obtained contact matrices between age groups from surveys conducted in a city that is geographically close to the province of study (Zhang et al., 2019, 2020). Since the age structures of the population in these two regions are different, we also collected census data of distinct age groups in Zhejiang province to adjust contact matrices accordingly; see Part B.1 of Supporting Information for details.

2.2 | Susceptibility–transmissibility–contact decomposed transmission rate

When modeling transmission dynamics, a key factor is the construction of the transmission rate. The transmission rate in this paper is calculated as the product of the transmission probability (ρ) within a contact network and their total contact number per unit of time (c). To study age-dependent transmission dynamics, we divide the total population into seven age groups: 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, >60, and define an age-stratified network containing all age groups as nodes and their contacts as edges. Distinct from previous studies (Davies et al., 2020; Keeling et al., 2021; Prem et al., 2020), we allow varying transmission probabilities of infectors and infectees across age groups, which are denoted as ρ^{ij}, where i and j represent the age groups of infectee and infector, respectively.

In our model, we separate ρ^{ij} into two components: transmissibility (β_j) and susceptibility (κ_i). We define β_j as the infectiousness of a patient in the jth age group, or more particularly, the probability of disease transmission from an infector in the jth age group to all possible susceptible contacts. Accordingly, susceptibility κ_i is the probability of acquiring infection for a susceptible individual in the ith age group from any possible infector in his/her contacts. We assume that

$$\rho^{ij} = \kappa_i \cdot \beta_j. \tag{1}$$

Transmissibility is often related to virus shedding ability of a COVID-19 case, which may also be associated with the presence of symptoms (Gao et al., 2020; Kim et al., 2020; Luo et al., 2020). We further denote the transmissibility of symptomatic cases in the jth age group as $\beta_{j,s} \in [0,1]$, and define $\beta_{j,a} := \gamma\beta_{j,s} \in [0,1]$ as the transmissibility of asymptomatic infection for the jth age group, where $\gamma \geq 0$ is the transmissibility ratio. The transmission probability of a susceptible individual in the ith age group contacting with a symptomatic or asymptomatic infection in the jth age group (denoted as ρ_s^{ij} and ρ_a^{ij}, respectively) then can be calculated as $\rho_s^{ij} = \kappa_i\beta_{j,s}$ or $\rho_a^{ij} = \kappa_i\gamma\beta_{j,s}$.

To consider the dynamic social contact across different age groups, we assume a time-varying curve for the average contact numbers of the ith age group with the jth age group. This curve is estimated with the contact matrices between age groups through surveys (Zhang et al., 2019, 2020). The surveys provided two different levels of contact strength in Shanghai, an adjacent city of the Zhejiang province. The first survey from 2017 to 2018 was viewed as a baseline of the strength of contact. The second survey, with

the same design, was conducted during the epidemic (from February 1 to 10, 2020). Considering the stringent intervention policies implemented in the Zhejiang province (Ge et al., 2021), we construct a monotone trend of a contact curve (Tian et al., 2021) by using these contact matrices

$$c^{ij}_{m,\tau}(t) = \frac{c^{ij}_{\text{base}} - \tau c^{ij}_{\text{outb}}}{1 + \exp\left\{2\log(\frac{1-\varepsilon}{\varepsilon}) \cdot \frac{t - m/2}{m}\right\}} + \tau c^{ij}_{\text{outb}}, \quad (2)$$

where t, m, τ, and $\varepsilon \geq 0$, $c^{ij}_{m,\tau}(t)$ denotes the contact numbers (per unit of time) of the ith age group with the jth age group at time t ($t = 0$ denotes the first day of the collected data, which is January 8, 2020), and c^{ij}_{base} and c^{ij}_{outb} are the provided contact numbers of the ith age group with the jth age group based on the previous contact surveys in the baseline period and the outbreak period. Following Tian et al. (2021), we choose ε as 0.01 so that $c^{ij}_{m,\tau}(0) \approx c^{ij}_{\text{base}}$ and m roughly represents the duration of the changing process of the contact number from c^{ij}_{base} to $\tau c^{ij}_{\text{outb}}$, where τ is the correction parameter for the contact numbers in the outbreak periods accounting for potential deviations from different intervention policies between Zhejiang province and Shanghai. See Part B.1 of Supporting Information for more details on the calculation of c^{ij}_{base} and c^{ij}_{outb} and the roles of parameters m and τ. For simplicity, we suppose that $c^{ij}_{m,\tau}(t) \equiv c^{ij}_{m,\tau}(k)$ for $t \in (k, k+1]$ for any non-negative integer k. Moreover, we assume $m \in [0, 30]$ (i.e., the social contact level is stabilized within one month) and $\tau \in [0, 2]$.

Based on the above discussions, $\rho^{ij}_s c^{ij}_{m,\tau}(t)$ and $\rho^{ij}_a c^{ij}_{m,\tau}(t)$ are the transmission rates of a susceptible individual in the ith age group contacting with a symptomatic or asymptomatic infection in the jth age group, respectively. We estimate $\beta_{j,s}$'s, γ, m and τ in the transmission rates by fusing the susceptibility parameters in Davies et al. (2020) and the contact data in Zhang et al. (2019) and Zhang et al. (2020) into the epidemic model mentioned below.

2.3 | Age-related susceptible-exposed-infectious-removed model

The SEIR model and its various extensions have been applied to study transmission dynamics of many infectious diseases with various incubation periods, including flu (Dukic et al., 2012), Ebola (Frasso & Lambert, 2016; Lekone & Finkenstädt, 2006), and COVID-19 (Chong et al., 2020; Keeling et al., 2021; Tian et al., 2021). Based on the SEIR modeling framework, we divide the total population into eight compartments in seven age groups, which include the susceptible class (S^i), exposed class (E^i), pre-symptomatic infectious class (I^i_{ps}), symptomatic infectious class (I^i_s), asymptomatic infectious class (I^i_a), confirmed/recovery symptomatic class (R^i_{cs}) and asymptomatic class (R^i_{ca}, R^i_h) in the ith age group. Detailed definitions of these compartments are given in Part B.2 of Supporting information.

Within the SEIR model stratified by age groups, new infections are driven by the transmission from I^i_{ps}, I^i_s, and I^i_a. Asymptomatic patients (I^i_a) are infectors without typical symptoms. Some asymptomatic patients may have been quarantined through contact tracing, and others may have remained unnoticed throughout the entire epidemic. To account for this discrepancy, we assume that only a proportion of asymptomatic patients are detected (R^i_{ca}), while others (R^i_h) remain undiscovered and eventually self-recovered.

Let $\boldsymbol{X}_i(t) := (S^i(t), E^i(t), I^i_{\text{ps}}(t), I^i_s(t), I^i_a(t), R^i_{\text{cs}}(t), R^i_{\text{ca}}(t), R^i_h(t))^T$ be the random numbers of the corresponding compartments at time t in the ith age group, and $\boldsymbol{X}(t) := (\boldsymbol{X}_1(t), \ldots, \boldsymbol{X}_7(t))^T$. Given the initial numbers $\boldsymbol{X}(0)$, we define $\overline{\boldsymbol{X}}(t) := \mathbb{E}\{\boldsymbol{X}(t)|\boldsymbol{X}(0)\}$ with $\overline{\boldsymbol{X}}(0) = \boldsymbol{X}(0)$, and $(\overline{S}^i(t), \overline{E}^i(t), \overline{I}^i_{\text{ps}}(t), \overline{I}^i_s(t), \overline{I}^i_a(t), \overline{R}^i_{\text{cs}}(t), \overline{R}^i_{\text{ca}}(t), \overline{R}^i_h(t))^T$ are the expectations conditioning on $\boldsymbol{X}(0)$ for the numbers of corresponding compartments. Based upon the setting above, we assume that $\overline{\boldsymbol{X}}(t)$ is a smooth multiple curve w.r.t. the variable t, and define an age-related SEIR model by ordinary differential equations to consider the susceptibility–transmissibility–contact decomposed transmission rate. Specifically, we assume that the transmission occurrence follows:

$$\frac{\mathrm{d}\overline{S}^i(t)}{\mathrm{d}t} = -\overline{S}^i(t) u_i(t), \quad (3)$$

where $u_i(t)$ is the rate of infection for the susceptible individuals in the ith age group at time t associated with the transmission probabilities ρ^{ij}_s and ρ^{ij}_a and the dynamic contact curve $c^{ij}_{m,\tau}(t)$. We construct $u_i(t)$ as

$$u_i(t) = \sum_{j=1}^{7} \left(\rho^{ij}_s c^{ij}_{m,\tau}(t) \frac{\overline{I}^j_{ps}(t) + \overline{I}^j_s(t)}{N_j} + \rho^{ij}_a c^{ij}_{m,\tau}(t) \frac{\overline{I}^j_a(t)}{N_j} \right) \quad (4)$$

$$= \kappa_i \sum_{j=1}^{7} \beta_{j,s} c^{ij}_{m,\tau}(t) \left(\frac{\overline{I}^j_{ps}(t) + \overline{I}^j_s(t)}{N_j} + \gamma \frac{\overline{I}^j_a(t)}{N_j} \right).$$

The above equation is essential to consider the time-varying transmission within the symptom-age-stratified infectious states. Accordingly, to reflect the transmission nature of COVID-19, the subsequent transmission process

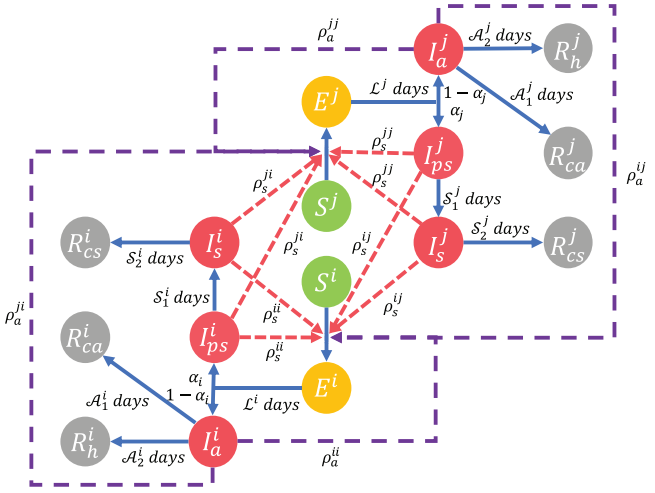

FIGURE 2 Susceptible-exposed-infectious-removed (SEIR) model with the age-related transmissions, where "j" represents another age group different from "i" for the compartments. The solid blue lines indicate the transition direction, and the dotted lines represent the age-related transmission (the dotted red lines denote the pre-symptomatic and symptomatic transmission, and the dotted purple lines denote the asymptomatic transmission, respectively). This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.

is assumed as

$$
\begin{aligned}
\frac{d\overline{E}^i(t)}{dt} &= \overline{S}^i(t)u_i(t) - \frac{\overline{E}^i(t)}{\mathcal{L}^i},\\
\frac{d\overline{I}^i_{\mathrm{ps}}(t)}{dt} &= \frac{\overline{E}^i(t)}{\mathcal{L}^i}\alpha_i - \frac{\overline{I}^i_{ps}(t)}{\mathcal{S}^i_1},\\
\frac{d\overline{I}^i_s(t)}{dt} &= \frac{\overline{I}^i_{ps}(t)}{\mathcal{S}^i_1} - \frac{\overline{I}^i_s(t)}{\mathcal{S}^i_2},\\
\frac{d\overline{I}^i_a(t)}{dt} &= \frac{\overline{E}^i(t)}{\mathcal{L}^i}(1-\alpha_i) - \frac{\overline{I}^i_a(t)}{\mathcal{A}^i_1} - \frac{\overline{I}^i_a(t)}{\mathcal{A}^i_2},\\
\frac{d\overline{R}^i_{\mathrm{cs}}(t)}{dt} &= \frac{\overline{I}^i_s(t)}{\mathcal{S}^i_2},\\
\frac{d\overline{R}^i_{\mathrm{ca}}(t)}{dt} &= \frac{\overline{I}^i_a(t)}{\mathcal{A}^i_1},\\
\frac{d\overline{R}^i_h(t)}{dt} &= \frac{\overline{I}^i_a(t)}{\mathcal{A}^i_2}.
\end{aligned}
\tag{5}
$$

The state transitions of the above equations are given in Figure 2.

Let Ω be the collection of all parameters, which includes the age-dependent parameters $\{\alpha_i, \beta_{i,s}, \kappa_i, \mathcal{L}^i, \mathcal{S}^i_1, \mathcal{S}^i_2, \mathcal{A}^i_1, \mathcal{A}^i_2, N_i;\ i = 1, \ldots, 7\}$, the transmissibility ratio $\gamma \in [0, \infty)$, and the contact-related parameters $m \in [0, 30]$ and $\tau \in [0, 2]$. Among $\alpha_i \in [0, 1]$, $\mathcal{L}^i$, $\mathcal{S}^i_1$, $\mathcal{S}^i_2$, $\mathcal{A}^i_1$, $\mathcal{A}^i_2$ are the symptomatic transition probability, and the average lengths of the latent period, the pre-symptomatic period, the symptomatic transmission period, the period between the beginning of asymptomatic transmission and the diagnosis of asymptomatic infection, and the period from the beginning of asymptomatic transmission to self-recovery for an asymptomatic infection in the ith age group, respectively. Besides, N_i is total population in the ith age group. It can be shown that Equations (3) and (5) uniquely determine $\overline{\boldsymbol{X}}(t)$ for all $t > 0$ given the parameters Ω and the initial numbers $\overline{\boldsymbol{X}}(0)$.

We introduce epidemic information from the existing literature for model identification and computation reduction. First, we set that $\mathcal{L}^i = 3$ and $\mathcal{S}^i_1 = 3$ (Kong et al., 2020; Lauer et al., 2020; Wei et al., 2020) for all i. In addition, we assume a length of 17 days for $\mathcal{A}^i_2$, based on estimated parameters from previous studies (Kim et al., 2020; Long et al., 2020; Lee et al., 2020). For the susceptibility parameters, according to Davies et al. (2020), we set κ_i as 0.40, 0.38, 0.79, 0.86, 0.80, 0.82, 0.81 for different age groups. These reference values are consistent with another previous observation that individuals over 20 years old are roughly twice as susceptible as those below 20 years old (Viner et al., 2021). The population number N_i is taken as the census data from Zhejiang Provincial Bureau of Statistics (2014). Besides, we estimate $\mathcal{S}^i_2$ by the sample mean $(\sum_{l=1}^{L_i} \mathcal{W}^i_l)/L_i$, where L_i is the number of symptomatic cases in the ith age group, and $\mathcal{W}^i_l$ is the length of symptomatic transmission period for the lth individual in the ith age group. Furthermore, we assume that $\mathcal{A}^i_1$ is age-invariant, that is, $\mathcal{A}^i_1 \equiv \mathcal{A}_1$. The parameters, including $\{\alpha_i, \beta_{i,s}; i = 1, \ldots, 7\}$, γ, $\mathcal{A}_1$, m, and τ are estimated from our age-structured data as illustrated in Figure 1. We use $\Omega_1 \subset \Omega$ to denote the parameters needed to be estimated and define $\Omega_2 := \Omega \setminus \Omega_1$.

2.4 | Bayes procedure

We follow a Bayesian paradigm to account for uncertainties in the unknown parameters Ω_1 conditioning on $\overline{\boldsymbol{X}}(0)$ and Ω_2. Let R^i_{ds} be an additional compartment including all patients who have developed symptoms, and define $R^i_{\mathrm{ds}}(t) := I^i_s(t) + R^i_{\mathrm{cs}}(t)$ and $\overline{R}^i_{\mathrm{ds}}(t) := \mathbb{E}\{R^i_{\mathrm{ds}}(t)|\boldsymbol{X}(0)\}$. For $t \in \{0, \ldots, T-1\}$ with T being the total length of collected data, we assume that $Y^{i,s}_t := R^i_{ds}(t+1) - R^i_{ds}(t)$ is the number of daily new symptom-developed individuals at time t in the ith age group, and $Y^{i,a}_t := R^i_{\mathrm{ca}}(t+1) - R^i_{\mathrm{ca}}(t)$ is the number of daily new confirmed asymptomatic

infected individuals at time t in the ith age group, as shown in Figure 1. Similarly, we define

$$\overline{Y}_t^{i,s}(\overline{\boldsymbol{X}}(0),\Omega) := \overline{R}_{\mathrm{ds}}^i(t+1) - \overline{R}_{\mathrm{ds}}^i(t),$$
$$\overline{Y}_t^{i,a}(\overline{\boldsymbol{X}}(0),\Omega) := \overline{R}_{\mathrm{ca}}^i(t+1) - \overline{R}_{\mathrm{ca}}^i(t), \quad (6)$$

for $t = 0,\dots,T-1$, where $\overline{R}_{\mathrm{ds}}^i(t)$ and $\overline{R}_{\mathrm{ca}}^i(t)$ are determined by Equations (3) and (5) given the initial numbers $\overline{\boldsymbol{X}}(0)$ and the parameters Ω. The values of $\overline{\boldsymbol{X}}(0)$ are specified by the initial symptomatic and asymptomatic cases on January 8, 2020, in Zhejiang province. In detail, we first specify the initial numbers of the compartments I_s^i, R_{cs}^i and R_{ca}^i and set the initial numbers of the compartments E^i, I_{ps}^i, I_a^i, R_h^i as zero for all i. After that, we calculate the initial numbers of the compartment S^i by N_i subtracting the total initial numbers of other compartments in the ith age group. We use negative binomial masses as the sample distributions of $Y_t^{i,s}$ and $Y_t^{i,a}$ for a flexible adjustment of overdispersion (Bliss & Fisher, 1953; Lindén & Mäntyniemi, 2011). Define $\boldsymbol{Y} := \{Y_t^{i,s}, Y_t^{i,a}; i = 1,\dots,7, t = 0,\dots,T-1\}$, the composite marginal likelihood for the increased numbers $\boldsymbol{Y}$ is then constructed as:

$$\mathcal{L}\left(\Omega_1, \upsilon \mid \boldsymbol{Y}, \overline{\boldsymbol{X}}(0), \Omega_2\right) = \prod_{t=0}^{T-1}\prod_{i=1}^{7}\prod_{g\in\{s,a\}} \mathrm{NegB}(Y_t^{i,g} \mid \overline{Y}_t^{i,g}(\overline{\boldsymbol{X}}(0),\Omega), \upsilon), \quad (7)$$

where $\mathrm{NegB}(*|\,\mu,\upsilon) = \binom{*+\upsilon-1}{*}(\frac{\upsilon}{\upsilon+\mu})^{\upsilon}(\frac{\mu}{\upsilon+\mu})^{*}$ represents the mass function of a negative binomial distribution parameterized with mean μ and dispersion parameter υ.

Due to imbalanced sample sizes for different age groups (see Figure 1), age-group-specific transmissibility may be noisily estimated with further complexities from dynamic contact strength and age-dependent susceptibility. To solve this problem, we impose shrinkage priors for transmissibility parameters to borrow information from adjacent age groups. Denoting $\widetilde{\beta}_{i,s}$ as $\log\{\beta_{i,s}/(1-\beta_{i,s})\}$ to transform $\beta_{i,s}$ into the interval $(-\infty,\infty)$, we assume that the transformed transmissibility $\widetilde{\beta}_{i,s}$ varies smoothly among age groups. Subsequently, we utilize a shrinkage prior for $\widetilde{\beta}_{i,s}$ under the fully Bayesian framework to penalize their differences, that is,

$$\pi(\widetilde{\beta}_{1,s},\dots,\widetilde{\beta}_{7,s}|\lambda) \propto \exp\left\{-\frac{\lambda}{2}\sum_{i=1}^{6}(\widetilde{\beta}_{i+1,s}-\widetilde{\beta}_{i,s})^2\right\}, \quad (8)$$

where $\pi(\cdot \mid \lambda)$ is the prior distribution of $\widetilde{\beta}_{1,s},\dots,\widetilde{\beta}_{7,s}$ with a parameter $\lambda \geq 0$. To avoid mixing problems, the prior distribution of λ is chosen as a relatively flat density that converges to 0 sufficiently fast as $\lambda \to \infty$ (Park & Casella, 2008). Particularly, we use the Gamma distribution with shape parameter 1 and scale parameter 0.01 as the prior distribution of λ. Besides, $\mathcal{A}_1$ is set as ∞ during the early period (before January 28) when no asymptomatic infections were identified. Flat priors are used for the remaining parameters. See Table 2 in Supporting Information for details of parameters and their prior distributions (or reference values).

To improve computational efficiency, we assume that the total population of susceptible individuals remains roughly unchanged before the epidemic is under control (i.e., $\overline{S}^i(t) \equiv N_i$ a.s. for all t in Equations (3) and (5)), consistent with what we have observed in the real data. As such, the right side of Equations (3) and (5) could be represented as

$$\frac{d\overline{\boldsymbol{X}}(t)}{dt} = \mathbf{A}(t \mid \Omega)\,\overline{\boldsymbol{X}}(t), \quad (9)$$

where $\mathbf{A}(t \mid \Omega)$ is a 56×56 matrix determined by t and Ω. Noting that $\mathbf{A}(t \mid \Omega) \equiv \mathbf{A}(k \mid \Omega)$ when $t \in (k, k+1]$ for $k = 0,\dots,T-1$, we obtain a closed-form expression for $\overline{\boldsymbol{X}}(t)$ as

$$\overline{\boldsymbol{X}}(t) = \mathbf{e}^{\mathbf{A}(t-1|\Omega)}\,\overline{\boldsymbol{X}}(t-1) \quad (10)$$

for $t = 1,\dots,T$, where $\mathbf{e}^{\boldsymbol{M}}$ denotes the matrix exponential defined as $\mathbf{e}^{\boldsymbol{M}} = \sum_{z=1}^{\infty}\boldsymbol{M}^z/z!$ (Moler & Van Loan, 2003). Accordingly, given $\overline{\boldsymbol{X}}(0)$ and Ω, we use the matrix exponential to efficiently calculate $\overline{Y}_t^{i,s}(\overline{\boldsymbol{X}}(0),\Omega)$ and $\overline{Y}_t^{i,a}(\overline{\boldsymbol{X}}(0),\Omega)$ in the likelihood (7). With all these materials, the posterior distributions of $\Theta := \{\Omega_1, \upsilon, \lambda\}$ given $\{\boldsymbol{Y}, \overline{\boldsymbol{X}}(0), \Omega_2\}$ are then calculated as

$$\pi\left(\Theta \mid \boldsymbol{Y}, \overline{\boldsymbol{X}}(0), \Omega_2\right) \propto \mathcal{L}\left(\Omega_1, \upsilon \mid \boldsymbol{Y}, \overline{\boldsymbol{X}}(0), \Omega_2\right)\cdot\pi(\Theta), \quad (11)$$

where $\pi(\cdot)$ and $\pi(\cdot\,|*)$ are the prior and posterior distributions of parameters, respectively. We use a random walk Metropolis-within-Gibbs algorithm (Andrieu et al., 2003; Givens & Hoeting, 2012) with an adaptive Metropolis step (Haario et al., 2001) to sample the posterior distributions of the parameters Θ. After the convergence of MCMC, the post-burn-in samples are collected to construct credible intervals (CI). Unless stated otherwise, posterior medians are used as the point estimates for parameters.

2.5 | Basic models and model comparison

To show the superiority of the proposed age-related SEIR model, we compare our model with two basic epidemic models similar to those proposed in Prem et al. (2020) and

Keeling et al. (2021). For the first basic model (denoted as basic model I), we assume that the transmission probabilities ρ_s^{ij} and ρ_a^{ij} are identical for all i and j. To implement this, we suppose that $\kappa_i = \kappa$ for each i with a known κ taken as the average of the age-dependent susceptibilities from Davies et al. (2020). The transmission probabilities for symptomatic and asymptomatic cases are subsequently defined as $\rho_s := \kappa\beta_s$ and $\rho_a := \kappa\gamma\beta_s$, where β_s is the transmissibility of symptomatic cases, which is assumed to be age-invariant and estimated from data. Accordingly, we replace ρ_s^{ij} and ρ_a^{ij} in Equation (4) by ρ_s and ρ_a, respectively, that is, the age-related transmission probabilities are not considered in the basic model I. For the second basic model (denoted as basic model II), we assume the transmission probabilities are constructed by age-invariant transmissibility and age-dependent susceptibility. Using the notations above, we calculate ρ_s^{ij} and ρ_a^{ij} in Equation (4) as $\kappa_i\beta_s$ and $\kappa_i\gamma\beta_s$, respectively, that is, the age-dependent transmissibility is not considered in the basic model II.

Apparently, the model complexity is reduced in these basic models. We use the deviance information criterion (DIC) to evaluate the overall model generalizability for epidemic data in balancing overfitting and the model complexity (Spiegelhalter et al., 2002, 2014). Let $\mathcal{M}$ be a model with the parameters Θ conditioning on $\boldsymbol{D} := \{\boldsymbol{Y}, \overline{\boldsymbol{X}}(0), \Omega_2\}$, and $\mathcal{L}(\Theta \mid \boldsymbol{D})$ be the likelihood function conditioning on $\boldsymbol{D}$, we define

$$\mathrm{DIC}(\mathcal{M}) = \mathrm{Res}(\mathcal{M}) + \mathrm{df}(\mathcal{M}), \tag{12}$$

the first term of which measures the fitting residual for the model $\mathcal{M}$ and is calculated as $-2\mathbb{E}_{\Theta|\boldsymbol{D}}\log\{\mathcal{L}(\Theta \mid \boldsymbol{D})\}$, where the expectation is taken over the posterior distribution of parameters. The second term $\mathrm{df}(\mathcal{M})$ is the model complexity penalty of $\mathcal{M}$, which is specified as $-2[\mathbb{E}_{\Theta|\boldsymbol{D}}\log\{\mathcal{L}(\Theta \mid \boldsymbol{D})\} - \log\{\mathcal{L}(\tilde{\Theta} \mid \boldsymbol{D})\}]$ with $\tilde{\Theta}$ being pseudotrue values of Θ. We use posterior samples of the parameters to approximate $\mathbb{E}_{\Theta|D}\log\{\mathcal{L}(\Theta \mid \boldsymbol{D})\}$ and take $\tilde{\Theta}$ as posterior medians of the corresponding parameters, as suggested in Spiegelhalter et al. (2002). Considering the overfitting issue of DIC (Ando, 2011; Spiegelhalter et al., 2014), we also present the predictive information criterion (PIC; Ando, 2011): $\mathrm{PIC}(\mathcal{M}) = \mathrm{Res}(\mathcal{M}) + 2\mathrm{df}(\mathcal{M})$ for model comparison, where the second term is twice that in DIC to increase the weight of the model complexity penalty for model comparison.

3 | SIMULATION STUDY

We compare the age-related SEIR model with two basic models in a simulation study. Following the age-related SEIR model, we first generate the epidemic data $\boldsymbol{Y}$. According to Section 2.3, we assume that $\mathcal{L}^i$'s, $\mathcal{S}_1^i$'s, $\mathcal{S}_2^i$'s, $\mathcal{A}_2^i$'s, κ_i's, and N_i's are taken the values suggested from previous studies. Besides, $\mathcal{A}_1$, m, and τ are taken as 14, 21, 0.5, respectively, and c_{base}^{ij}, c_{outb}^{ij}, and $\overline{\boldsymbol{X}}(0)$ are specified as discussed in earlier sections. Moreover, we sample the symptomatic transition probabilities α_i's independently from a Beta distribution with the shape parameters both chosen as 2. For the transmissibility parameters, we set $\gamma = 0.5$ and generate $\beta_{i,s}$ as $\beta_{i,s} = 0.4/[1 + \exp\{-\sum_{h=1}^{4} a_h B_h(i)\}]$, where $B_h(\cdot)$'s are chosen as the B-spline functions to introduce smoothness for the transmissibility, and a_h's are independently generated from a standard normal distribution. Given the parameters Ω and $\overline{\boldsymbol{X}}(0)$, we generate $\overline{Y}_t^{i,s}(\overline{\boldsymbol{X}}(0), \Omega)$ and $\overline{Y}_t^{i,a}(\overline{\boldsymbol{X}}(0), \Omega)$ for $i = 1, \ldots, 7$ and $t = 1, \ldots, 60$ through the Runge–Kutta scheme (Soetaert et al., 2010) based on Equations (3) and (5), and sample $Y_t^{i,s}$ and $Y_t^{i,a}$ from the negative binomial masses with means $\overline{Y}_t^{i,s}(\overline{\boldsymbol{X}}(0), \Omega)$ and $\overline{Y}_t^{i,a}(\overline{\boldsymbol{X}}(0), \Omega)$, and dispersion parameter $\upsilon = 100$.

We apply the age-related SEIR model and the basic models I and II separately to fit the generating data, and extract the medians of posterior distributions as the point estimates of parameters. Let $\widehat{\theta}^u$ be the point estimate of the parameter θ in the uth simulation. We define several measures to assess the performance of different models

$$\begin{aligned}
\mathrm{RBIAS}(\theta) &= \frac{|\overline{\theta} - \theta_0|}{\theta_0} \cdot 100\%, \\
\mathrm{RSDE}(\theta) &= \frac{\sqrt{\frac{1}{U}\sum_{u=1}^{U}(\overline{\theta} - \widehat{\theta}^u)^2}}{\theta_0} \cdot 100\%, \\
\mathrm{RRMSE}(\theta) &= \frac{\sqrt{\frac{1}{U}\sum_{u=1}^{U}(\widehat{\theta}^u - \theta_0)^2}}{\theta_0} \cdot 100\%,
\end{aligned} \tag{13}$$

where $\overline{\theta} = \frac{1}{U}\sum_{u=1}^{U}\widehat{\theta}^u$, θ_0 is the true value of θ, and U is the total replicates of simulations. RBIAS, RSDE, and RRMSE stand for the relative bias, the standard deviation error, and the square root of the mean square error of the estimated parameters, respectively. We conduct 100 replicates of simulations and present the results in Table 1.

The results show that the RBIASs from the age-related SEIR model and the basic model II are significantly lower than those from the basic model I, indicating that the consideration of the age-related transmission probabilities significantly reduces the estimation bias of model parameters. Furthermore, while some of the RSDEs in the age-related SEIR model are slightly larger than those in the basic model II, the smaller RBIASs in the age-related SEIR model result in overall smaller RRMSEs for the parameters estimated by the age-related SEIR model. The larger

TABLE 1 The RBIAS, RSDE, and RRMES of the parameters needed to be estimated in Ω for three models

	Age-related SEIR model			Basic Model I			Basic Model II		
Parameters	**RBIAS (%)**	**RSDE (%)**	**RRMSE (%)**	**RBIAS (%)**	**RSDE (%)**	**RRMSE (%)**	**RBIAS (%)**	**RSDE (%)**	**RRMSE (%)**
α_1	0.15	3.63	3.63	36.45	2.73	36.55	4.47	3.98	5.98
α_2	0.09	2.24	2.24	38.28	1.14	38.30	2.18	2.14	3.06
α_3	0.52	5.07	5.10	46.72	9.20	47.61	2.44	5.06	5.62
α_4	0.01	0.99	0.99	5.43	0.82	5.49	0.99	1.03	1.43
α_5	0.44	3.23	3.26	47.71	6.12	48.10	2.76	3.44	4.41
α_6	0.05	2.06	2.06	13.97	2.21	14.15	0.04	2.08	2.08
α_7	0.05	2.93	2.93	6.08	2.82	6.70	5.71	3.21	6.55
$\beta_{1,s}$	7.31	14.08	15.86	120.64	7.56	120.88	27.83	5.87	28.45
$\beta_{2,s}$	16.99	14.45	22.31	144.32	8.37	144.57	41.55	6.51	42.06
$\beta_{3,s}$	15.32	15.66	21.90	131.33	7.93	131.57	34.03	6.16	34.58
$\beta_{4,s}$	12.67	12.12	17.54	100.35	6.86	100.58	16.07	5.33	16.94
$\beta_{5,s}$	0.03	5.23	5.23	70.16	5.83	70.40	1.42	4.53	4.75
$\beta_{6,s}$	1.97	6.54	6.83	48.97	5.10	49.24	13.69	3.97	14.26
$\beta_{7,s}$	4.64	6.64	8.10	37.94	4.73	38.24	20.08	3.67	20.42
γ	15.67	18.45	24.21	99.03	0.39	99.03	16.31	18.29	24.51
$\mathcal{A}_1$	0.29	5.95	5.96	113.14	13.79	113.98	4.53	6.62	8.02
m	0.63	1.09	1.26	1.87	1.28	2.27	1.05	1.04	1.48
τ	2.14	2.45	3.25	3.54	1.44	3.82	3.63	2.23	4.26

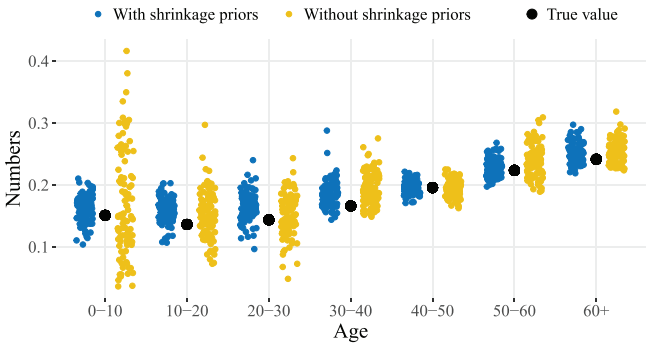

FIGURE 3 The true values of the transmissibility parameters in all age groups and their point estimates fused with or without shrinkage priors for each simulation. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.

variance of the estimated parameters in the age-related SEIR model is likely due to the higher level of model complexity. Despite the increased model complexity, our method further reduces the estimation bias and estimation error by considering age-dependent transmissibility.

To further examine the effectiveness of the shrinkage procedure in Equation (8), we also estimate the age-dependent transmissibility based on the age-related SEIR model without imposing shrinkage priors. The true values of the transmissibility parameters and their point estimates fused with or without shrinkage priors are given in Figure 3. Our result shows that the transmissibility parameters estimated without the shrinkage priors significantly scatter from their true values than those estimated with the shrinkage priors. While there exists a slight estimation bias, our shrinkage procedure improves the estimation behavior of the age-dependent transmissibility by borrowing information from adjacent age groups.

4 | CASE STUDY

In this section, we use the age-related SEIR model to estimate the symptomatic and asymptomatic transmissibility of COVID-19 patients from our dataset. First, we fit our epidemic data from January 8 to February 22, 2020, by the age-related SEIR model and the other two basic models. Their fitting residuals, model complexities, DIC's, and PIC's are compared in Table 3 of Supporting Information. Although the complexity of our proposed model is higher than the two basic models, its fitting residuals are substantially smaller, leading to smaller DIC and PIC. Therefore, we conclude that the proposed age-related SEIR model is a better representation of our epidemic data in terms of DIC and PIC.

We present the box-plot of posterior samples of transmissibility in Figure 4. The point estimate of the transmissibility ratio γ indicates that asymptomatic cases are, on average, 76.45% (with a CI (27.38%, 88.65%)) less infectious than symptomatic cases. Besides, the transmissibility of symptomatic or asymptomatic cases monotonically

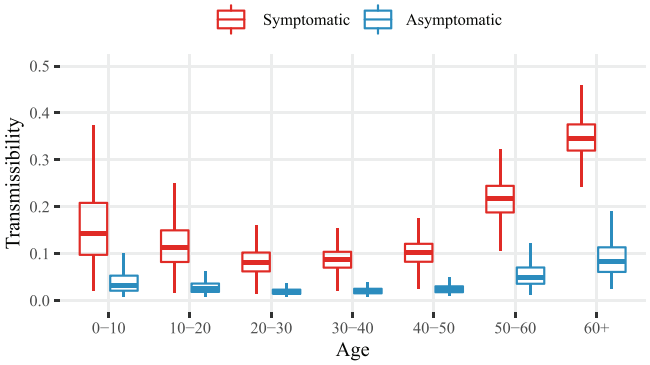

FIGURE 4 The box-plot of posterior samples of transmissibility for symptomatic and asymptomatic cases in each age group. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.

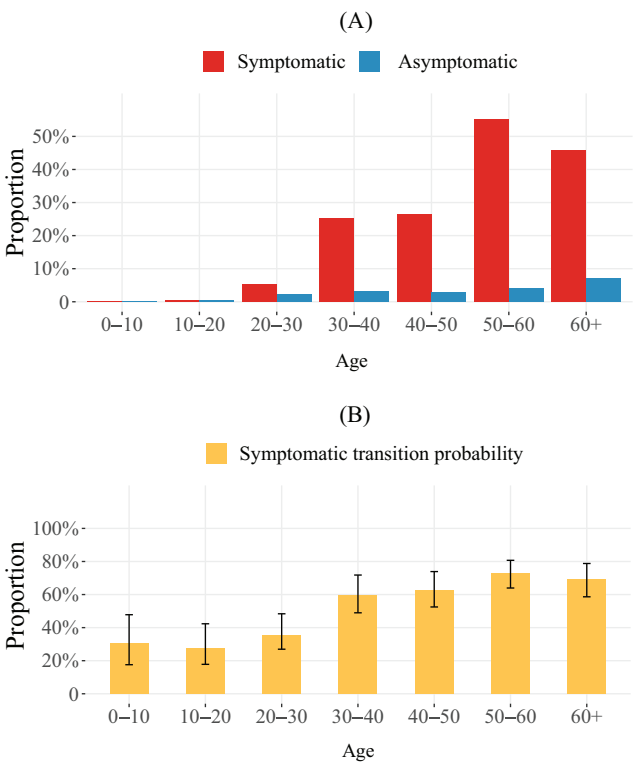

FIGURE 5 (A) The percentage reduction in all infections (up to February 22) when the transmissibility of symptomatic or asymptomatic cases is reduced by 50% for each age group. (B) The estimated symptomatic transition probability α_i for each age group with 95% credible intervals. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.

increases with age for those of age 20 and over. We suspect that older adults are not only the most vulnerable to be succumbed to COVID-19 but also may be more likely to transmit the virus once infected. Interventions that attempt to suppress transmissions, such as mask-wearing and vaccination, could prioritize older adults if interventions for the entire population are not feasible.

To verify the robustness of the estimated transmissibility with shrinkage priors, we propose sensitivity analyses for the estimated parameters under certain disturbances of susceptibility parameters, contact matrices, lengths of stay in states, and daily increased numbers in Part C.2 of Supporting Information. Essentially, these parameters or data are important to estimate age-dependent transmissibility, and changes in values would certainly influence the estimates. However, in these sensitivity analyses (Figures 2–5 in Supporting Information), we find that the trend of transmissibility by age groups is similar to those in Figure 4. It suggests certain robustness of our estimated age-dependent transmissibility.

To illustrate the effect of the age-dependent transmissibility within the transmission dynamics, we calculate the percentage reduction in all infections when the transmissibility of symptomatic ($\beta_{i,s}$) or asymptomatic patients ($\beta_{i,a}$) is reduced by 50%, that is, the virus shedding ability of the corresponding patients is reduced by 50% (e.g., wearing a mask). We present these results in Figure 5A and provide further details in Part C.3 of Supporting Information. We find that the average percentage reduction for reducing the asymptomatic transmissibility (2.81%) is significantly smaller than that for reducing the symptomatic transmissibility (22.63%), and the benefit of reducing the transmissibility reaches its peak in the 50–60 age group for symptomatic infection, rather than increasing with age. The highest reductions are obtained in age groups 50–60 (55.15%) and >60 (7.05%) for reducing symptomatic and asymptomatic transmissibility, respectively. Such differences in transmission burden by age groups support prioritizing age-dependent prevention and control strategies when facing strained resources.

Additionally, we find a considerable benefit to reducing transmissibility for persons over 30 years old. Reducing transmissibility of individuals < 30 years old by 50% results in a reduction of less than 5% of overall infections (see Table 4 in Supporting Information for the specific values), despite being almost 40% of the entire population. The estimated symptomatic transition probabilities given in Figure 5B also indicate that the symptomatic transition probabilities (α_i's) for those aged over 30 years old (59.76%, 62.86%, 72.66%, 69.53%, for 30–40, 40–50, 50–60, >60, respectively) are substantially higher than those of the younger groups (30.75%, 27.77%, 35.36% for 0–10, 10–20, and 20–30, respectively). Identified as the primary contributors to the transmission of COVID-19, older age groups (> 30) are not only more susceptible to the disease but also more likely to develop symptoms leading to higher transmissibility. Interventions prioritizing those aged 30 years

old and older are expected to be more cost-effective at the population level.

5 | DISCUSSION

In this study, we propose a Bayesian age-related SEIR model with shrinkage priors that allow for robustly identifying the transmissibility of symptomatic and asymptomatic patients in distinct age groups. Through simulation, we find that the proposed method reduces the estimation bias of model parameters, and the use of shrinkage priors improves the estimation behavior of age-dependent transmissibility. Superior DIC and PIC values justify the choice of the age-related SEIR model in fitting our epidemic data of COVID-19 with complex age structures.

Based on our results, the infectiousness of asymptomatic patients is significantly lower than those of symptomatic patients, and the transmissibility of COVID-19 mainly increases with age. Middle-aged adults and the elderly have higher transmissibility compared to younger participants. Prioritizing these groups in prevention strategies is likely more effective in reducing the population-level transmission burden. Infectiousness profiling of the quantified age-dependent symptomatic and asymptomatic patients reported in our study can assist policy decisions on pandemic control.

Our findings are consistent with current empiric evidence (Gao et al., 2020). We further find that age may directly impact COVID-19 transmission through virus shedding patterns. Similar results have been discussed in previous studies (Esposito & Principi, 2020; Kim et al., 2020). Symptoms are often mild in infected children (Dong et al., 2020) but are generally more severe among the elder (Zhou et al., 2020). Higher symptom severity has arguably been associated with increased shedding of the virus (He et al., 2020; Liu et al., 2020).

There are several limitations in this study. First, since susceptibility parameters and contact survey data were obtained from the literature, a lack of accounting for population heterogeneity could lead to biased estimations of the age-dependent transmissibility parameters. To address this limitation, we introduced shrinkage priors for transmissibility parameters to adjust for the uncertainties around such discrepancies. Second, our statistical analysis is based on epidemic data collected during the first wave of the COVID-19 pandemic, which may not be generalizable for understanding the transmissibility of emerging variants (Lopez Bernal et al., 2021). Nonetheless, our proposed modeling framework is generally applicable in modeling infectious disease transmission dynamics with complicated heterogeneity related to age-dependent transmissibility and contact patterns.

ACKNOWLEDGMENT

This research was partially supported by National Natural Science Foundation of China (Grant No. 11871485).

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available in the Supporting Information in this article.

ORCID

Jianbin Tan https://orcid.org/0000-0002-3264-1086
Ye Shen https://orcid.org/0000-0002-6662-0048
Hui Huang https://orcid.org/0000-0001-5703-8788

REFERENCES

Ando, T. (2011) Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31(1–2), 13–38.

Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M.I. (2003) An introduction to MCMC for machine learning. *Machine Learning*, 50(1), 5–43.

Bjørnstad, O.N., Shea, K., Krzywinski, M. & Altman, N. (2020) Modeling infectious epidemics. *Nature Methods*, 17(5), 455–456.

Bliss, C.I. & Fisher, R.A. (1953) Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2), 176–200.

Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M. & Eggo, R.M. (2020) Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8), 1205–1211.

Dong, Y., Mo, X., Hu, Y., Qi, X., Jiang, F., Jiang, Z., & Tong, S. (2020) Epidemiology of COVID-19 among children in China. *Pediatrics*, 145(6), e20200702.

Dukic, V., Lopes, H.F. & Polson, N.G. (2012) Tracking epidemics with Google flu trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107(500), 1410–1426.

Esposito, S. & Principi, N. (2020) To mask or not to mask children to overcome COVID-19. *European Journal of Pediatrics*, 179(8), 1267–1270.

Frasso, G. & Lambert, P. (2016) Bayesian inference in an extended SEIR model with nonparametric disease transmission rate: an application to the Ebola epidemic in Sierra Leone. *Biostatistics*, 17(4), 779–792.

Gao, M., Yang, L., Chen, X., Deng, Y., Yang, S., Xu, H., Chen, Z. & Gao, X. (2020) A study on infectivity of asymptomatic SARS-CoV-2 carriers. *Respiratory Medicine*, 169, 106026.

Givens, G.H. & Hoeting, J.A. (2012) *Computational statistics*, vol. 703. John Wiley & Sons.

Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., Liu, L., Shan, H., Lei, C.-L., Hui, D. S.C., Du, B., Li, L.-J., Zeng, G., Yuen, K.-Y., Chen, R.-C., Tang, C.-L., Wang, T., Chen, P.-Y., & Xiang, J. (2020) Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*, 382(18), 1708–1720.

Haario, H., Saksman, E. & Tamminen, J. (2001) An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.

He, X., Lau, E.H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W.,

Hu, F., Zhang, Q., Zhong, M., Wu, Y., & Zhao, L. (2020) Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5), 672–675.

Hoehl, S., Rabenau, H., Berger, A., Kortenbusch, M., Cinatl, J., Bojkova, D., Behrens, P., Boddinghaus, B., Gotsch, U., Naujoks, F., Neumann, P., Schork, J., Tiarks-Jungk, P., Walczok, A., Eickmann, M., Vehreschild, M. J. G. T., Kann, G., Wolf, T., Gottschalk, R., & Ciesek, S. (2020) Evidence of SARS-CoV-2 infection in returning travelers from Wuhan, China. *New England Journal of Medicine*, 382(13), 1278–1280.

Hu, Z., Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., Ma, H., Chen, W., Lin, Y., Zheng, Y., Wang, J., Hu, Z., Yi, Y., & Shen, H. (2020) Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China Life Sciences*, 63(5), 706–711.

Huang, L., Zhang, X., Zhang, X., Wei, Z., Zhang, L., Xu, J., Liang, P., Xu, Y., Zhang, C., & Xu, A. (2020) Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside Wuhan and characteristics of young patients with COVID-19: a prospective contact-tracing study. *Journal of Infection*, 80(6), e1–e13.

Ka, C. C., Wei, C., Shi, Z., Feng, L., Kirran, N. M., Maggie, H. W., Benny, C. Y. Z., Lei, W., Xi, X., Hengyan, L., Jingxuan, W., & Enfu, C. (2020) Monitoring disease transmissibility of 2019 novel coronavirus disease in Zhejiang, China. *International Journal of Infectious Diseases*, 96, 128–130.

Keeling, M.J., Hill, E.M., Gorsich, E.E., Penman, B., Guyver-Fletcher, G., Holmes, A., Holmex, A., Leng, T., McKimm, H., Tamborrino, M., Dyson, L., & Tildesley, M. J. (2021) Predictions of COVID-19 dynamics in the UK: short-term forecasting and analysis of potential exit strategies. *PLoS Computational Biology*, 17(1), e1008619.

Kong, D., Zheng, Y., Wu, H., Pan, H., Wagner, A.L., Zheng, Y., Gong, X., Zhu, Y., Jin, B., Xiao, W., Mao, S., Lin, S., Han, R., Yu, X., Cui, P., Jiang, C., Fang, Q., Lu, Y., & Fu, C. (2020) Pre-symptomatic transmission of novel coronavirus in community settings. *Influenza and Other Respiratory Viruses*, 14(6), 610–614.

Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A. S., Reich, N. G., Lessler, J. (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172(9), 577–582.

Lee, S., Kim, T., Lee, E., Lee, C., Kim, H., Rhee, H., Park, S. Y., Son, H.-J., Yu, S., Park, J. W., Choo, E. J., Park, S., Loeb, M., Kim, T. (2020) Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the Republic of Korea. *JAMA Internal Medicine*, 180(11), 1447–1452.

Lekone, P.E. & Finkenstädt, B.F. (2006) Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62(4), 1170–1177.

Lewnard, J.A. & Lo, N.C. (2020) Scientific and ethical basis for social-distancing interventions against COVID-19. *The Lancet Infectious Diseases*, 20(6), 631–633.

Lindén, A. & Mäntyniemi, S. (2011) Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7), 1414–1421.

Liu, Y., Yan, L.-M., Wan, L., Xiang, T.-X., Le, A., Liu, J.-M., Peiris, M., Poon, L. L. M., Zhang, W. (2020) Viral dynamics in mild and severe cases of COVID-19. *The Lancet Infectious Diseases*, 20(6), 656–657.

Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., Su, K., Zhang, F., Gong, J., Wu, B., Liu, X.-M., Li, J.-J., Qiu, J.-F., Chen, J., & Huang, A.-L. (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine*, 26(8), 1200–1204.

Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., Stowe, J., Tessier, E., Groves, N., Dabrera, G., Myers, R., Campbell, C. N. J.., Amirthalingam, G., Edmunds, M., Zambon, M., Brown, K. E., Hopkins, S., Chand, M., & Ramsay, M. (2021) Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *New England Journal of Medicine*, 385(7), 585–594.

Luo, L., Liu, D., Liao, X., Wu, X., Jing, Q., Zheng, J., Liu, F., Yang, S., Bi, H., Li, Z., Liu, J., Song, W., Zhu, W., Wang, Z., Zhang, X., Huang, Q., Chen, P., Liu, H., & Cheng, X. (2020) Contact settings and risk for transmission in 3410 close contacts of patients with COVID-19 in Guangzhou, China: a prospective cohort study. *Annals of Internal Medicine*, 173(11), 879–887.

Meng, H., Xiong, R., He, R., Lin, W., Hao, B., Zhang, L., Lu, Z., Shen, X., Fan, T., Jiang, W., Yang, W., Li, T., Chen, J., Geng, Q. (2020) CT imaging and clinical course of asymptomatic cases with COVID-19 pneumonia at admission in Wuhan, China. *Journal of Infection*, 81(1), e33–e39.

Moler, C. & Van Loan, C. (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1), 3–49.

Park, T. & Casella, G. (2008) The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681–686.

Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Flasche, S., Clifford, S., Pearson, C. A. B., Munday, J. D., Abbott, S., Gibbs, H., Rosello, A., Quilty, B. J., Jombart, T., Sun, F., Diamond, C., Gimma, A., & van Zandvoort, K. (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5), e261–e270.

Rock, K., Brand, S., Moir, J. & Keeling, M.J. (2014) Dynamics of infectious diseases. *Reports on Progress in Physics*, 77(2), 026602.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., & Hoelscher, M. (2020) Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*, 382(10), 970–971.

Seong, E. K., Hae, S. J., Yohan, Y., Sung, U. S., Soosung, K., Tae, H. O., Uh, J. K., Seung-Ji, K., Hee-Chang, J., Sook-In, J., & Kyung-Hwa, P. (2020) Viral kinetics of SARS-CoV-2 in asymptomatic carriers and presymptomatic patients. *International Journal of Infectious Diseases*, 95, 441–443.

Soetaert, K., Petzoldt, T. & Setzer, R.W. (2010) Solving differential equations in R: package desolve. *Journal of Statistical Software*, 33, 1–25.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van der Linde, A. (2014) The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3), 485–493.

Tian, T., Tan, J., Jiang, Y., Wang, X. & Zhang, H. (2021) Evaluate the risk of resumption of business for the states of New York, New Jersey and Connecticut via a pre-symptomatic and asymptomatic transmission model of COVID-19. *Journal of Data Science*, 19(2), 178–196.

Tian, T., Tan, J., Luo, W., Jiang, Y., Chen, M., Yang, S., Wen, C., Pan, W., & Wang, X. (2021) The effects of stringent and mild interventions for coronavirus pandemic. *Journal of the American Statistical Association*, 116(534), 481–491.

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubel~, Z., FitzJohn, R., & Gaythorpe, K. (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6), 669–677.

Viner, R.M., Mytton, O.T., Bonell, C., Melendez-Torres, G., Ward, J., Hudson, L., Waddington, C., Thomas, J., Russell, S., van der Klis, F., Koirala, A., Ladhani, S., Panovska-Griffiths, J., Davies, N. G., Booy, R., & Eggo, R. M. (2021) 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.

Wei, W.E., Li, Z., Chiew, C.J., Yong, S.E., Toh, M.P. & Lee, V.J. (2020) Presymptomatic transmission of SARS-CoV-2-Singapore, January 23–March 16, 2020. *Morbidity and Mortality Weekly Report*, 69(14), 411.

World Health Organization (2020) WHO Coronavirus Disease (COVID-19) Dashboard. Available from: https://covid19.who.int [Accessed 12 December 2020].

Yang, G., Zhiping, C., Andreas, H., Leonardo, M., Qian, X., Changwei, L., Enfu, C., Jinren, P., Yang, L., Feng, L., Ye, S. (2021) The impact of social distancing, contact tracing, and case isolation interventions to suppress the COVID-19 epidemic: a modeling study. *Epidemics*, 36, 100483.

Zhang, J., Klepac, P., Read, J.M., Rosello, A., Wang, X., Lai, S., Li, M., Song, Y., Wei, Q., Jiang, H., Yang, J., Lynn, H., Flasche, S., Jit, M., & Yu, H. (2019) Patterns of human social contact and contact with animals in Shanghai, China. *Scientific Reports*, 9(1), 1–11.

Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., Wu, Q., Merler, S., Viboud, C., Vespignani, A., Ajelli, M., & Yu, H. (2020) Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 368(6498), 1481–1486.

Zhejiang Provincial Bureau of Statistics (2014) Zhejiang Provincial Bureau of Statistics: Sixth Census Data. 2014; published online Sept 3. Available from: http://tjj.zj.gov.cn/art/2014/9/3/art_1530851_20980968.html (Accessed 30 July 2020].

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229), 1054–1062.

SUPPORTING INFORMATION

Web Appendices, Tables, Figures referenced in Sections 1–4 and the codes to reproduce our results are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Tan, J., Shen, Y., Ge, Y., Martinez, L. & Huang, H. (2023) Age-related model for estimating the symptomatic and asymptomatic transmissibility of COVID-19 patients. *Biometrics*, 79, 2525–2536. https://doi.org/10.1111/biom.13814