An age-of-infection model with both symptomatic and asymptomatic infections

Fan Bai

Received: 28 August 2022 / Revised: 27 February 2023 / Accepted: 3 April 2023 / Published online: 25 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
We formulate a general age-of-infection epidemic model with two pathways: the symptomatic infections and the asymptomatic infections. We then calculate the basic reproduction number $R_0$ and establish the final size relation. It is shown that the ratio of accumulated counts of symptomatic patients and asymptomatic patients is determined by the symptomatic ratio $f$ which is defined as the probability of eventually becoming symptomatic after being infected. We also formulate and study a general age-of-infection model with disease deaths and with two infection pathways. The final size relation is investigated, and the upper and lower bounds for final epidemic size are given. Several numerical simulations are performed to verify the analytical results.

Keywords Mathematical epidemiology · Deterministic compartmental models · Age of infection models · Integro-differential equations · Final epidemic size estimation

Mathematics Subject Classification 92D30 · 45J05

1 Introduction
Asymptomatic infections are recorded for COVID-19 (Coronavirus SARS-CoV-2) pandemic and have been studied from many different aspects. Evidence suggests that asymptomaticity has significant impacts on the development of COVID-19 (Rothe et al. 2020; McCulloh et al. 2020). Furthermore, because of its “silently transmissible” nature, the pathway of asymptomatic infection is considered as a key component.
for understanding the mechanisms of transmission and evaluating the efficiency of intervention measures. However, quantification of asymptomatic infections and further inference of value of symptomatic ratio \( f \) is tricky due to the unreliable recorded numbers of symptomatic patients and asymptomatic patients at any specific time post epidemic outbreak. For COVID-19, it is suggested that the asymptomatic ratio is around 35.1\% (Sah et al. 2021) and it is highly age dependent. Other studies based on daily incidence data from certain geographic locations and using various parameter estimation techniques indicate very different values of symptomatic/asymptomatic ratio (e.g., McCulloh et al. 2020; Nishiura et al. 2020; Li et al. 2021). The purpose of this paper is to formulate the epidemic models with two infection pathways, and to obtain the relation between the ratio, the basic reproduction number \( R_0 \) and the final epidemic size. It is possible that the symptomatic ratio \( f \) has different formats. Both analytical and numerical results are provided for the case that \( f \) is a constant. Only numerical results are presented for the case that \( f \) is a piece-wise function of time, due to the difficulty of model analysis. Another focus of this paper is to compare the final sizes of symptomatic infections and asymptomatic infections.

The Age-of-infection model was originally derived by Kermack and McKendrick in their well-known paper (Kermack and McKendrick 1927), it is the general form of epidemic models (Brauer 2005b, a; Brauer and Castillo-Chavez 2012; Brauer 2008a; Yang and Brauer 2008; Feng et al. 2007). This type of epidemic model is especially useful to study the epidemic models in which periods are far from exponential distributed, e.g., the current COVID-19 pandemic. The formulation of the basic reproduction number for the age-of-infection model was well studied in Yang and Brauer (2008) and the final size relation has been established in Brauer (2008a). The age-of-infection model with disease-induced mortality was also studied in Brauer (2008a), an inequality of final size relation was developed. The general model was also extended for the heterogeneous mixing population in Brauer and Watmough (2009). In this paper, we formulate and study an age-of-infection model for epidemics with both symptomatic infections and asymptomatic infections. The derivation and comparison of the basic reproduction number \( R_0 \) and the formulation of final size relations are the main focus of the study. We also consider the age-of-infection model with two infection pathways and with disease-induced mortality.

This paper consists of five parts. In Sect. 2, we firstly review the basic general age-of-infection model in a homogeneously mixing population. In Sect. 3, we formulate the age-of-infection model with two pathways, the symptomatic infection and the asymptomatic infection. We then calculate the basic reproduction number and obtain the corresponding final size relation. We will focus on the ratio of total symptomatic cases and total asymptomatic cases in this study, which entirely depends on the symptomatic ratio \( f \). As an example, we study a special case of Susceptible-Exposed-Infectious-Asymptomatic-Removed (SEIAR) model, by assuming exponentially distributed disease stages. In Sect. 4, we further consider the disease-induced deaths for the age-of-infection model with two infection pathways and investigate the final size relations. In Sect. 5, we perform several sets of numerical simulations to verify our theoretical results. In Sect. 6, we summarize the investigation and also propose some interesting future work in the direction of age-of-infection modeling.
2 The general age-of-infection model

We have \( S(t) \) denote the number of susceptibles at time \( t \) and have \( \phi(t) \) be the total infectivity at time \( t \). \( \phi(t) \) is defined as the sum of products of the number of infected members with each infection age and the mean infectivity for that infection age (Yang and Brauer 2008). The infection age can be interpreted as the time units after initial infection for infected individuals who are still not recovered. We assume that on average each member of the population make \( a \) contacts per unit time. The population size is \( N \) and there are no demographic effects (births, deaths, migration) in the population. We further define \( B(\tau) \) the fraction of infected members remaining infected at infection age \( \tau \) and \( \pi(\tau) \) the mean infectivity at infection age \( \tau \). \( B(\tau) \) is a non-increasing function and satisfies that

\[
B(0) = 1, \quad B(\infty) = 0.
\]

The general age-of-infection model is

\[
\begin{align*}
S' &= -\frac{a}{N} S \phi \\
\phi(t) &= \phi_0(t) + \int_0^t \frac{a}{N} S(t - \tau) \phi(t - \tau) B(\tau) \pi(\tau) d\tau,
\end{align*}
\]

where \( \phi_0(t) \) is the total infectivity of members of the population who were infected at the initial time, at time \( t \). The following Fig. 1 demonstrates the dynamics of the general model. It is worth mentioning that the number of infectives is described as

\[
I(t) = I_0(t) + \int_0^t \frac{a}{N} S(t - \tau) \phi(t - \tau) B(\tau) d\tau.
\]

It is straightforward to derive the equation about the change of number of infectives (Yang and Brauer 2008)

\[
I'(t) = \frac{a}{N} S \phi + \int_0^t \frac{a}{N} S(\tau) \phi(\tau) B'(t - \tau) d\tau,
\]

**Fig. 1** The diagrammatic framework of the basic age-of-infection model (1). The solid arrows represent the transition from the susceptible compartment to the infectious compartment and the transition from the infectious compartment to the recovered compartment, respectively. The dashed arrow represents the contacts between susceptible individuals and infectious individuals and the occurrences of new infections.
where $B'$ reflects the recovery of infected individuals.

It has been well studied in Brauer (2008a) and Yang and Brauer (2008), that the basic reproduction number for model (1) is

$$R_0 = a \int_0^\infty B(\tau)\pi(\tau)d\tau. \quad (2)$$

The final size relation for model (1) is

$$\log \frac{S_0}{S_\infty} = R_0 \left[1 - \frac{S_\infty}{N_0}\right] - \frac{a}{N_0} \int_0^\infty [(N_0 - S_0)B(t)\pi(t) - \phi_0(t)]dt. \quad (3)$$

The uniqueness of $S_\infty$ has been confirmed in Brauer (2008a). If the initial infectivity is small and neglectible, the final size relation can be simplified as

$$\log \frac{S_0}{S_\infty} = R_0 \left[1 - \frac{S_\infty}{N_0}\right]. \quad (4)$$

3 The age-of-infection model with both symptomatic and asymptomatic infections

3.1 The model formulation

We extend the general age-of-infection model (1) by considering that the infected individuals are possibly asymptomatic or symptomatic. It is assumed that once an individual is infected, there is a probability of $0 < f < 1$ that this individual eventually becomes symptomatic, and the probability of becoming asymptomatic is $(1 - f)$. We denote $\phi_i(t)$ the sum of products of the symptomatic individuals with each infection age and the mean infectivity for that infection age, $\phi_a(t)$ the sum of products of the asymptomatic individuals with each infection age and the mean infectivity for that infection age. Thus, the total infectivity is

$$\phi(t) = \phi_i(t) + \phi_a(t). \quad (5)$$

$B_i(\tau)$ and $B_a(\tau)$ represent the fraction of symptomatic and asymptomatic individuals remaining infected at infection age $\tau$, respectively; $\pi_i(\tau)$ and $\pi_a(\tau)$ represent the mean infectivity of symptomatic and asymptomatic individuals at infection age $\tau$, respectively. $A_i(\tau) := B_i(\tau)\pi_i(\tau)$ is the mean infectivity of a symptomatic individual at infection age $\tau$, while $A_a(\tau) := B_a(\tau)\pi_a(\tau)$ is the mean infectivity of an asymptomatic individual at infection age $\tau$.

The dynamics is depicted in Fig. 2. It is noticed that the recovered symptomatic/asymptomatic patients are theoretically counted as one stage of the infection process, but with no infectivities.
The age-of-infection model is

\[ S'(t) = -a \frac{S}{N} (\phi_i + \phi_a) \]
\[ \phi_i(t) = \phi_{i,0}(t) + \int_0^t f \frac{a}{N} S(t-\tau)(\phi_i(\tau) + \phi_a(\tau)) B_i(\tau) \pi_i(\tau) d\tau, \]
\[ \phi_a(t) = \phi_{a,0}(t) + \int_0^t (1 - f) \frac{a}{N} S(t-\tau)(\phi_i(\tau) + \phi_a(\tau)) B_a(\tau) \pi_a(\tau) d\tau. \]

(6)

It can be proved that model (6) is an extension of the general age-of-infection model (1). We first calculate the basic reproduction number for model (6),

\[ R_0 = a \left[ f \int_0^\infty A_i(\tau) d\tau + (1 - f) \int_0^\infty A_a(\tau) d\tau \right]. \]

(7)

This form is consistent with Theorem 3.1 in Yang and Brauer (2008) that the basic reproduction number \( R_0 \) depends only on the mean period in each infective stage, regardless of its distribution. We now generate the final size relation for model (6). We first write

\[ -\frac{S'}{S} = a \frac{1}{N} (\phi_{i,0}(t) + \phi_{a,0}(t)) + a \frac{1}{N} \int_0^t \left[ -S'(t-\tau) \right] (f A_i(\tau) + (1 - f) A_a(\tau)) d\tau. \]

Integration with respect to \( t \) from 0 to \( \infty \) yields

\[ \log \frac{S_0}{S_\infty} = a \frac{1}{N} \int_0^\infty (\phi_{i,0}(t) + \phi_{a,0}(t)) dt \]
\[ + a \frac{1}{N} \int_0^\infty \int_0^t \left[ -S'(t-\tau) \right] (f A_i(\tau) + (1 - f) A_a(\tau)) d\tau d\tau. \]
\[
\frac{a}{N} \int_0^\infty \left( \phi_{i,0}(t) + \phi_{a,0}(t) \right) dt
+ \frac{a}{N} \int_0^\infty \left( f A_i(\tau) + (1-f) A_a(\tau) \right) \int_\tau^\infty \left[ -S'(t - \tau) \right] dt d\tau
= \frac{a}{N} \int_0^\infty \left( \phi_{i,0}(t) + \phi_{a,0}(t) \right) dt + \frac{a}{N} \left[ S_0 - S_\infty \right] \int_0^\infty \left( f A_i(\tau) \right.
\left. + (1-f) A_a(\tau) \right) \left. d\tau \right)
+ \mathcal{R}_0 \frac{S_0 - S_\infty}{N}.
\]

For each infection case, the probability of proceeding to the symptomatic state is \( f \) and the probability of proceeding to the asymptomatic state is \( (1-f) \). Thus, it is intuitive to conclude that the total number of symptomatic patients is \( f(S_0 - S_\infty) \) and the total number of asymptomatic patients is \((1-f)(S_0 - S_\infty)\). We now perform some calculations to verify this conjecture. Firstly, the numbers of symptomatic patients \( I(t) \) and asymptomatic patients \( A(t) \) are

\[
I(t) = I_0(t) + \int_0^t \frac{a}{N} S(t - \tau) \phi(t - \tau) B_i(\tau) d\tau,
A(t) = A_0(t) + \int_0^t (1-f) \frac{a}{N} S(t - \tau) \phi(t - \tau) B_a(\tau) d\tau.
\]

We now focus on \( I(t) \) and calculate the derivative of \( I(t) \),

\[
I'(t) = f \frac{a}{N} S \phi + \int_0^t \frac{a}{N} S(\tau) \phi(\tau) B_i'(t - \tau) d\tau.
\]

In Eq. (9), the first term indicates the rate of new symptomatic infections and the second term represents the transition from infected stage to recovery stage for symptomatic patients. Thus, we have

\[
R_i'(t) = - \int_0^t \frac{a}{N} S(\tau) \phi(\tau) B_i'(t - \tau) d\tau.
\]

We then integrate both sides of the equation with respect to \( t \) from 0 to \( \infty \),

\[
\int_0^\infty R_i'(t) dt = - \int_0^\infty \int_0^t \frac{a}{N} S(\tau) \phi(\tau) B_i'(t - \tau) d\tau dt
R_i(\infty) = - \int_0^\infty \int_\tau^\infty f [-S'(t - \tau)] B_i'(\tau) d\tau dt,
= - \int_0^\infty f B_i'(\tau) \int_\tau^\infty [-S'(t - \tau)] d\tau dt
= - \int_0^\infty f B_i'(\tau) [S_0 - S_\infty] d\tau
\]
\[
= -f[S_0 - S_{\infty}][B_i(\infty) - B_i(0)] \\
= f[S_0 - S_{\infty}].
\] (10)

Similarly, we have
\[
R_a(\infty) = (1 - f)[S_0 - S_{\infty}].
\]

We state the following theorem to summarize all calculations.

**Theorem 3.1** If the population size \( N \) is large and the number of initial infection cases is small, the final size of an age-of-infection model with both symptomatic infections and asymptomatic infections is
\[
\log \frac{S_0}{S_{\infty}} = R_0 \frac{S_0 - S_{\infty}}{N},
\] (11)

with \( R_0 \) being expressed in Eq. (7). Moreover, the total estimated cases of symptomatic infections are \( f(S_0 - S_{\infty}) \), while the total estimated cases of asymptomatic infections are \( (1 - f)(S_0 - S_{\infty}) \).

If the initial infection cases are not negligible, we have the following inequality,
\[
\log \frac{S_0}{S_{\infty}} > R_0 \frac{S_0 - S_{\infty}}{N}.
\] (12)

It is possible that the ratio \( f \) is not a constant, but a function of the time \( t \). This may be caused by the mutations of responsible viruses. The model then becomes
\[
S' = -a \frac{S}{N} (\phi_i + \phi_a) \\
\phi_i(t) = \phi_{i,0}(t) + \int_0^t f(\tau) \frac{a}{N} S(t - \tau)(\phi_i(t - \tau) + \phi_a(t - \tau)) A_i(\tau) d\tau, \\
\phi_a(t) = \phi_{a,0}(t) + \int_0^t (1 - f(\tau)) \frac{a}{N} S(t - \tau)(\phi_i(t - \tau) + \phi_a(t - \tau)) A_a(\tau) d\tau.
\] (13)

The basic reproduction number \( R_0 \) for model (13) depends on initial ratio \( f(0) \) and is given by,
\[
R_0 = a \left[ f(0) \int_0^{\infty} A_i(\tau) d\tau + (1 - f(0)) \int_0^{\infty} A_a(\tau) d\tau \right].
\] (14)

\( f(t) \) may take the form of a Heaviside function (with one or multiple switch times) or its smooth approximations (generalized logistic functions). Explicit final size relation for model (13) is impossible to obtain, but can be approximated by Eq. (3) in Theorem 3.1 if \( \max(f(\tau)) - \min(f(\tau)) \) is small or the switch time(s) is late or close to the end of the epidemic. Some numerical simulations will be performed to show that the final sizes can not be accurately predicted, but be appropriately approximated.
3.2 The effect on final epidemic size estimation by using incorrectly inferred ratio

In practice, the counts of symptomatic patients are relatively accurate. While the counts of asymptomatic patients basically rely on serology reports and surveys. Thus, it is generally a challenging task to correctly infer the value of $f$ based on epidemiological data. This leads to an interesting problem of the comparison of outcomes by using the actual $f$ and the inaccurately inferred $f'$.

First we have a similar formula for $R'_0$ based on $f'$:

$$R'_0 = a \left[ f' \int_0^\infty A_i(\tau) d\tau + (1 - f') \int_0^\infty A_a(\tau) d\tau \right].$$  \hspace{1cm} (15)

We denote

$$R_i := a \int_0^\infty A_i(\tau) d\tau,$$

$$R_a := a \int_0^\infty A_a(\tau) d\tau$$ \hspace{1cm} (16)

to represent the average secondary infection cases caused by a symptomatic patient or an asymptomatic patient, when being introduced into a wholly susceptible population. The basic reproduction number is the linear combination of $R_i$ and $R_a$. The difference between $R'_0$ and $R_0$ depends on the relation between ratios $f$, $f'$, $R_i$ and $R_a$,

$$R_0 - R'_0 = (f - f')(R_i - R_a).$$ \hspace{1cm} (17)

Therefore, the relation between $f$ and $f'$ is insufficient to determine how incorrect fraction will affect the assumed $R_0$. It is critical to compare the values of $R_i$ and $R_a$. In practice, it is reasonable to assume that the inferred ratio $f'$ is larger than the real ratio $f$, due to the fact that the asymptomatic cases are more easier to be missed and the asymptomatic infections are underestimated. The other situation is that the asymptomatic infections are overestimated, if the presymptomatic cases are incorrectly categorized as asymptomatic infections. This case indicates a smaller inferred $f'$. However, the comparison of $R_i$ and $R_a$ is difficult to obtain in real life. The incorrectly inferred values of the ratio, $R_i$ and $R_a$ will lead to the biased estimation of the basic reproduction number $R_0$ for the epidemic. In Bai (2020), it was discussed that the misleading value of $R_0$ will also affect the prediction of the final epidemic size.

3.3 A special case: exponential distributions of disease stages

A special case of the general age-of-infection model (6) is the compartmental epidemic model with exposed stages, for both infection pathways. We start from the original model (6) and introduce two functions $\mu_i(\tau)$ and $\mu_a(\tau)$ to denote the fractions of symptomatic and asymptomatic patients with infection age $\tau$ who are not yet infectious, respectively. Similarly, we introduce functions $\nu_i(\tau)$ and $\nu_a(\tau)$ the fractions
of symptomatic and asymptomatic patients with infection age $\tau$ who are infectious, respectively (Brauer 2008a; Brauer and Castillo-Chavez 2012). For the symptomatic infection path ($S \rightarrow E_i \rightarrow I \rightarrow R$), $\mu_i(\tau)$ and $v_i(\tau)$ are assumed to satisfy the ordinary differential equation system

$$
\begin{align*}
\mu_i'(\tau) &= -\kappa_i \mu_i(\tau), \\
v_i'(\tau) &= \kappa_i \mu_i(\tau) - \alpha_i v(\tau),
\end{align*}
$$

with initial condition $\mu_i(0) = 1$ and $v_i(0) = 0$. $\kappa_i$ and $\alpha_i$ are transition rates. We are able to obtain the solution of the ODE system (18)

$$
\begin{align*}
\mu_i(\tau) &= e^{-\kappa_i \tau}, \\
v_i(\tau) &= \frac{\kappa_i}{\kappa_i - \alpha_i} [e^{-\alpha_i \tau} - e^{-\kappa_i \tau}].
\end{align*}
$$

The change of fractions $\mu_i(\tau)$ and $v_i(\tau)$ with a chosen set of parameter values ($\kappa_i = 0.07$ and $\alpha_i = 0.2$) is visualized in Fig. 3.

It is noticed that the function $B_i(\tau)$ is implicitly expressed by two functions $\mu_i$ and $v_i$ in Eq. (19). It is straightforward to have the averaged infectivities for both exposed patients and symptomatic patients if $\pi_E = 0$ and $\pi_I = 1$. Based on the definition of $A_i(\tau)$, we have

$$
A_i(\tau) = \frac{\kappa_i}{\kappa_i - \alpha_i} [e^{-\alpha_i \tau} - e^{-\kappa_i \tau}].
$$

Similarly, the other asymptomatic infection path ($S \rightarrow E_a \rightarrow A \rightarrow R$) can be analyzed in the same manner. It implies

$$
A_a(\tau) = \frac{\kappa_a}{\kappa_a - \alpha_a} [e^{-\alpha_a \tau} - e^{-\kappa_a \tau}],
$$
Fig. 4 The diagrammatic framework of the SEIAR model (20). The solid arrows and the dashed arrows are applied to represent the transitions between different compartments, and the occurrences of new infections, respectively.

where $\kappa_a$ and $\alpha_a$ are transition rates defined in asymptomatic infection pathway. Now, we are able to rewrite the general age-of-infection epidemic model (6) as the following extended SEIAR epidemic model,

\[
\begin{align*}
S' &= -a \frac{S}{N} (I + A) \\
E_i' &= af \frac{S}{N} (I + A) - \kappa_i E_i \\
E_a' &= a(1 - f) \frac{S}{N} (I + A) - \kappa_a E_a \\
I' &= \kappa_i E_i - \alpha_i I \\
A' &= \kappa_a E_a - \alpha_a A \\
R' &= \alpha_i I + \alpha_a A.
\end{align*}
\]

(20)

The basic dynamics of this SEIAR model (20) is described in Fig. 4. More details regarding the derivation of compartmental models from the general age-of-infection model can be found in Bai (2016), Brauer (2005a), Brauer (2008a), Brauer et al. (2008) and Brauer and Castillo-Chavez (2012).

From the formula for the basic reproduction number $R_0$ in Eq. (7), we have

\[
R_0 = a \left[ f \int_0^{\infty} A_i(\tau)d\tau + (1 - f) \int_0^{\infty} A_a(\tau)d\tau \right] \\
= a \left[ f \int_0^{\infty} \frac{\kappa_i}{\kappa_i - \alpha_i} (e^{-\alpha_i \tau} - e^{-\kappa_i \tau})d\tau \\
+ (1 - f) \int_0^{\infty} \frac{\kappa_a}{\kappa_a - \alpha_a} (e^{-\alpha_a \tau} - e^{-\kappa_a \tau})d\tau \right] \\
= a \left( \frac{f}{\alpha_i} + \frac{(1 - f)}{\alpha_a} \right).
\]

(21)
This result can also be verified by using the next generation matrix approach (van den Driessche and James 2002). If we consider the average relative infectivity for asymptomatic patients is $\epsilon < 1$, compared to that of symptomatic patients, it can be calculated that $A_a(\tau) = \epsilon e^{-\kappa_i \tau} + \frac{\kappa_d}{\kappa_a - \alpha_a} [e^{-\alpha_a \tau} - e^{-\kappa_a \tau}]$ and further we have

$$R_0 = a \left( \frac{f}{\alpha_i} + \frac{(1 - f) \epsilon}{\alpha_a} \right).$$

The final size relation for model (20) can be generated by Theorem (3.1). We can also use the alternative method (Brauer 2008b) to verify this result. Firstly, we denote

$$f(\infty) := \lim_{t \to \infty} f(t), \quad \hat{f} := \int_0^{\infty} f(t) dt,$$

if function $f$ is a non-negative integrable function defined on $0 \leq t < \infty$. It follows the lemma (see Brauer 2008b),

**Lemma 3.1** If $f(t)$ is a non-negative monotone nonincreasing continuously differentiable function, then as $t \to \infty$, $f(t) \to f_\infty \geq 0$ and $f'(t) \to 0$.

Because we have

$$(S + E_i + E_a + I + A)' = -\alpha_i I - \alpha_a A \leq 0,$$

this leads to $I(\infty) = 0$ and $A(\infty) = 0$. We further obtain the number of all infection cases

$$S(0) - S(\infty) = \alpha_i \hat{I} + \alpha_a \hat{A}. \quad (22)$$

We then add the second equation and the fourth equation, the third equation and the fifth equation in (20), respectively,

$$(E_i + I)' = a f \frac{S}{N} (I + A) - \alpha_i I,$$

$$(E_a + A)' = a (1 - f) \frac{S}{N} (I + A) - \alpha_i I.$$

To integrate the above equations with respect to $t$ from 0 to $\infty$ implies

$$E_i(\infty) + I(\infty) - E_i(0) - I(0) = a f \frac{1}{N} \int_0^{\infty} S(I + A) dt - \alpha_i \hat{I},$$

$$E_a(\infty) + A(\infty) - E_a(0) - A(0) = a (1 - f) \frac{1}{N} \int_0^{\infty} S(I + A) dt - \alpha_a \hat{A}.$$

Because the numbers of individuals in all infection related compartments are 0 when $t = 0$ and $t = \infty$, the above system of equations can thus be simplified as the linear
relation between \( \hat{I} \) and \( \hat{A} \),

\[
\hat{I} = \frac{f \alpha_a}{(1-f)\alpha_i} \hat{A},
\]

(23)

Equations (22) and (23) imply that

\[
\hat{I} = \frac{f}{\alpha_i} (S(0) - S(\infty)),
\]

\( \hat{A} = \frac{1-f}{\alpha_a} (S(0) - S(\infty)) \).

(24)

Subsequently, from the first equation in model (20), we have

\[
-S' = \frac{a}{N} I + \frac{a}{N} A,
\]

\[
\log \frac{S(0)}{S(\infty)} = \frac{a}{N} \hat{I} + \frac{a}{N} \hat{A}.
\]

We then substitute equations (24) into the above equation and obtain the final size relation for model (20),

\[
\log \frac{S(0)}{S(\infty)} = a \left( \frac{f}{\alpha_i} + \frac{(1-f)}{\alpha_a} \right) \frac{S(0) - S(\infty)}{N},
\]

\[
\log \frac{S(0)}{S(\infty)} = R_0 \frac{S(0) - S(\infty)}{N}.
\]

(25)

The linear relation between \( \hat{A} \) and \( \hat{I} \) in Eq. (23) provides another point of view to investigate the ratio of accumulative numbers of symptomatic patients and asymptomatic patients. In Eq. (23), \( \hat{I} \) and \( \hat{A} \) can be interpreted as the sum of lengths of infection periods for all symptomatic and asymptomatic patients, respectively. If we define the length of infection period for symptomatic patient \( j \) (\( j \in [1, 2, \ldots, \text{final size of symptomatic patients}] \)) is \( \frac{1}{\alpha_i} \), the law of large numbers indicates that

\[
\hat{I} := \sum_{j=1}^{\text{final size of symptomatic patients}} \frac{1}{\alpha_i} \approx \text{final size of symptomatic patients} \times \frac{1}{\alpha_i}.
\]

Same argument can be made for asymptomatic infections. Apparently the second part of the above equation holds when the final size of symptomatic patients is relatively large, and \( \frac{1}{\alpha_i} \) is the pre-defined mean length of infection period. Therefore, we have the final size of symptomatic patients is \( \hat{I} \alpha_i \) and the final size of asymptomatic patients is \( \hat{A} \alpha_a \). This also implies that the ratio of sizes of symptomatic and asymptomatic patients is \( \frac{\hat{I}}{1-f} \), which is consistent with Theorem 3.1. In stochastic epidemic modeling with only symptomatic path (the Sellke stochastic model), a similar term \( \frac{\hat{I}}{N} \) is defined.
to measure the total accumulative infection pressure on a given susceptible individual during the course of the epidemic (Sellke 1983; Britton 2010).

Remark For two pathways of symptomatic infection and asymptomatic infection, the structures (functions $B_i$ and $B_a$) can be varied significantly. The stage of pre-symptomatic in the pathway of symptomatic infection has been frequently included in modeling COVID-19 pandemic. For an individual who is pre-symptomatic, he or she can infect other susceptible individuals. The infectivity is likely to be even higher than the patients with symptoms (Centers for Disease Control and Prevention 2021). Thus, for upper infection pathway in Fig. 4, it is reasonable to add one more compartment “pre-symptomatic” to describe the intermediate infectious stage. The corresponding function $B_i$ can be constructed straightforwardly. It is also possible to consider that the stays in the exposed periods (or in other compartments) follow the Gamma distributions and assume multiple exposed sub-stages (Bai et al. 2018; Brauer et al. 2019; Subramanian et al. 2021).

4 The age-of-infection model with disease deaths

If we consider there are disease deaths, the total population size does not remain constant. We denote $N(t)$ is the population size over time and $N_0$ is the initial population size. We further assume that a fraction $\mu_1$ ($\mu_2$) of symptomatic (asymptomatic) infectives dies of disease, and $\mu_1 > \mu_2$. The contact rate $a(N)$ is a density dependent saturating function (Brauer 2008a; Brauer and Castillo-Chavez 2012), where $a(N)$ is a non-decreasing function of $N$ and $\frac{a(N)}{N}$ is a non-increasing function of $N$. The age-of-infection epidemic model with disease deaths is

$$ S' = -a(N) \frac{S}{N} (\phi_i + \phi_a) $$

$$ \phi_i(t) = \phi_{i,0}(t) + \int_0^t f [-S'(t - \tau)] A_i(\tau) d\tau, $$

$$ \phi_a(t) = \phi_{a,0}(t) + \int_0^t (1 - f)[-S'(t - \tau)] A_a(\tau) d\tau. $$

(26)

Because of the disease deaths, the exact final epidemic size for model (26) is impossible to obtain. As the basic reproduction number $\mathcal{R}_0$ only depends on the initial states of the epidemic and the population, it can be explicitly defined as (Brauer et al. 2008; Brauer 2005b, 2008a)

$$ \mathcal{R}_0 = a(N_0) \left[ f \int_0^\infty A_i(\tau) d\tau + (1 - f) \int_0^\infty A_a(\tau) d\tau \right]. $$

(27)

We now establish the final size relation for model (26) with respect to the value of $\mathcal{R}_0$ in (27). Based on the properties of functions $a(N)$ and $\frac{a(N)}{N}$ (Brauer 2008b, a),
∀t ∈ [0, ∞), we have:

\[
\frac{a(N_0)}{N_0} \leq \frac{a(N(t))}{N(t)} \leq \frac{a(N_\infty)}{N_\infty} \leq \frac{a(N_0)}{(1 - \max(\mu_1, \mu_2))N_0},
\]

(28)

where \(N_\infty\) is the final size of individuals who survive after the epidemic ends. We now integrate the first equation in model (26) for \(\frac{S}{N}\) and obtain

\[
\log \frac{S_0}{S_\infty} = \int_0^\infty \frac{a(N(t))}{N(t)} (\phi_{i,0}(t) + \phi_{a,0}(t)) dt + \int_0^\infty \frac{a(N(t))}{N(t)} f \int_0^t (-S'(t - \tau))A_i(\tau) d\tau + (1 - f) \int_0^t (-S'(t - \tau))A_a(\tau) d\tau \] dt.

(29)

Because \(\frac{a(N(t))}{N(t)}\) is bounded and continuous for \(t \in [0, \infty)\), there exists the constant \(N_*\) such that \(N_0 \geq N_* \geq N_\infty\), and satisfy

\[
\log \frac{S_0}{S_\infty} = \frac{a(N_*)}{N_*} \left( \int_0^\infty (\phi_{i,0}(t) + \phi_{a,0}(t)) dt \right) + \int_0^\infty \left[ f \int_0^t (-S'(t - \tau))A_i(\tau) d\tau + (1 - f) \int_0^t (-S'(t - \tau))A_a(\tau) d\tau \right] dt
\]

\[
= \frac{a(N_*)}{N_*} \left( \int_0^\infty (\phi_{i,0}(t) + \phi_{a,0}(t)) dt \right) + \int_0^\infty \left[ f A_i(\tau) \int_\tau^\infty (-S'(t - \tau)) d\tau + (1 - f) A_a(\tau) \int_\tau^\infty (-S'(t - \tau)) d\tau \right] d\tau
\]

\[
= \frac{a(N_*)}{N_*} \left( \int_0^\infty (\phi_{i,0}(t) + \phi_{a,0}(t)) dt \right) + (S_0 - S_\infty) \int_0^\infty (f A_i(\tau) + (1 - f) A_a(\tau)) d\tau.
\]

(30)

Similarly, if we consider both the initial symptomatic cases and initial asymptomatic cases are zero or neglectible, the integral term \(\int_0^\infty (\phi_{i,0}(t) + \phi_{a,0}(t)) dt\) can be omitted in Eq. (30). The estimate of final epidemic size is

\[
\log \frac{S_0}{S_\infty} = \frac{a(N_*)N_0}{N_* a(N_0)} R_0 \frac{S_0 - S_\infty}{N_0}.
\]

(31)

It is impossible to explicitly calculate the values of \(N_*\) and \(a(N_*)\), because \(N_*\) depends on many factors, such as the shapes of functions \(A_i(t)\) and \(A_a(t)\). However, we can obtain a relatively accurate estimates by employing the inequalities in (28),

\[
R_0 \frac{S_0 - S_\infty}{N_0} \leq \log \frac{S_0}{S_\infty} \leq \frac{1}{1 - \max(\mu_1, \mu_2)} R_0 \frac{S_0 - S_\infty}{N_0}.
\]

(32)
It has been investigated in Bai (2020) that, $S_\infty$ is uniquely determined by the value of $R_0$. Indicated by Lemma 2.1 and Lemma 2.2 in Bai (2020), we have the lower and upper bounds for $S_\infty$. The first part of inequality in Eq. (32) implies that $S_\infty$ is less or equal to the final susceptible population size, when the basic reproduction number is taken as $R_0$ and disease-induced deaths are not considered; the second part of inequality in Eq. (32) shows that $S_\infty$ is larger or equal to the final susceptible population size, when the basic reproduction number is taken as $1 - \max\{\mu_1, \mu_2\}R_0$ and there are no disease deaths. It is clear to see that, if $\max\{\mu_1, \mu_2\} \to 0$, the elementary squeeze theorem indicates that

$$\log \frac{S_0}{S_\infty} \approx R_0 \frac{S_0 - S_\infty}{N_0}.$$ 

We still argue that, if the ratio $f$ is a constant, the number of total symptomatic cases is $f(S_0 - S_\infty)$ and the number of total asymptomatic cases is $(1 - f)(S_0 - S_\infty)$. This is due to the fact that the inclusion of disease-induced deaths does not affect the whole infection process in our model. We have the following theorem to summarize the analysis.

**Theorem 4.1** Assume the epidemic model has both symptomatic infections and asymptomatic infections, and the disease induced death proportions are $\mu_1$ and $\mu_2$, respectively. If the basic reproduction number is $R_0$, then the final susceptible population size $S_\infty$ has the lower bound of limiting susceptible population size with basic reproduction number $1 - \max\{\mu_1, \mu_2\}R_0$ and with no disease death; the upper bound is the limiting susceptible population size with basic reproduction number $R_0$ and without disease deaths. The ratio of the total symptomatic cases to the total asymptomatic cases is $\frac{f}{1 - f}$.

This theorem is the complement of Theorem 4.1 in Brauer (2008a), it provides the upper bound for the estimate of limiting final susceptible population size with disease deaths considered. It also indicates that when the disease induced death proportions are larger, it is more challenging to accurately predict the outcome of the epidemic.

5 **Numerical experiments**

In this section, we perform two sets of numerical simulations to verify the analytical results in previous sections. The size of population is $N = 10^5$ and the assumed symptomatic ratio $f = 0.6$ which indicates that the probability for an infected individual becoming symptomatic is 60%. We further assume the contact rate $a = 0.5$ (/day). Regarding the infected stages, we have $\kappa_i = 0.07$ (/day) and $\alpha_i = 0.2$ (/day) for symptomatic infection path; and have $\kappa_a = 0.03$ (/day) and $\alpha_a = 0.1$ (/day) for parallel asymptomatic infection path. It is therefore calculated that the basic reproduction number $R_0 = 3.5$, and $R_i = 2.5$ and $R_a = 5$. Clearly, if the symptomatic ratio $f$ is overestimated, the estimated $R_0$ is less than the actual basic reproduction number. Assumptions of standard incidence and exponentially distributed periods are made to simplify the numerical simulations.
5.1 Epidemic models without disease deaths

We first consider there are no disease-induced deaths for both symptomatic infections and asymptomatic infections. If the ratio $f$ is a constant, we are able to obtain the exact final epidemic size. The total numbers of symptomatic patients and asymptomatic patients during the course of the epidemic are 5796 and 3863, respectively. It can be calculated that the ratio of two numbers is $\frac{f}{1-f} = 1.5$. The dynamics is shown in the following Fig. 5.

We next consider the case that the ratio $f$ is a function of time. Different scenarios are considered for numerical experiments:

1. $f$ is a piecewise function and satisfies

\[
    f(t) = \begin{cases} 
        0.6 & \text{if } t \leq 150, \\
        0.5 & \text{if } t > 150.
    \end{cases}
\]  

(33)

2. $f$ is a piecewise function and satisfies

\[
    f(t) = \begin{cases} 
        0.6 & \text{if } t \leq 150, \\
        0.1 & \text{if } t > 150.
    \end{cases}
\]  

(34)

3. $f$ is a piece-wise function and satisfies

\[
    f(t) = \begin{cases} 
        0.6 & \text{if } t \leq 150, \\
        0.9 & \text{if } t > 150.
    \end{cases}
\]  

(35)
4. $f$ is a piece-wise function and satisfies

$$f(t) = \begin{cases} 0.6 & \text{if } t \leq 300, \\ 0.5 & \text{if } t > 300. \end{cases}$$

The simulations are presented in Figs. 6, 7, 8, 9 and the simulated outcomes are summarized in Table 1.

If we consider the simulation results for epidemic model with constant ratio as the baseline. When the switch of ratio occurs near the end of the epidemic, it is observed in Fig. 9 that the final outcome can be approximated by the baseline. However, if the switch occurs earlier, the final epidemic can be significantly different and it is entirely unpredictable.
5.2 Epidemic models with disease deaths

We now consider the epidemic models with identical parameters but with a larger population size $N = 10^6$ and with disease induced deaths. Firstly, we assume the death proportions are $\mu_1 = 0.2\%$ and $\mu_2 = 0.1\%$ and the dynamics of the model is simulated in Fig. 10. We are also able to obtain the number of total symptomatic cases $R_i(\infty) + D_i(\infty) = 57,966$ and the number of total asymptomatic cases $R_a(\infty) + D_a(\infty) = 38,643$. Therefore the final size of epidemic is $S(0) - S(\infty) = 96,609$. The ratio is 1.50. The lower bound for final size estimation can be obtained by simulating epidemic models without disease deaths, and we have $S(0) - S(\infty) = 96,597$. If we consider another epidemic model without disease death and with the basic reproduction number be $R_0/(1 - 0.2\%)$, the final epidemic size is 96623 and this number is the upper bound for the estimation of final size for the original model with disease deaths.
Table 1  Summary of the final epidemic sizes

| Model type                  | Number of total symptomatic cases | Number of total asymptomatic cases | Ratio |
|-----------------------------|-----------------------------------|-----------------------------------|-------|
| Model with constant ratio   | 5796                              | 3963                              | 1.50  |
| Model with ratio function (33) | 5550                              | 4133                              | 1.34  |
| Model with ratio function (34) | 4516                              | 5248                              | 0.86  |
| Model with ratio function (35) | 6497                              | 3081                              | 2.11  |
| Model with ratio function (36) | 5795                              | 3865                              | 1.49  |

Fig. 10  The progression of the epidemic with disease death proportions $\mu_1, 2 = 0.2\%, 0.1\%$ and the accumulation of death cases

It can be observed that, since the death proportions are very low, two bounds are both reasonable approximations.

As a comparison, we next consider the epidemic model with larger death proportions $\mu_1 = 2\%$ and $\mu_2 = 1\%$. The final epidemic size is 96,708, while the upper bounds and lower bounds are 96,724 and 96,597. If the population size is larger, it becomes more difficult to accurately predict the final epidemic size.

6 Conclusion and future work

We have formulated an age-of-infection model to study an epidemic with symptomatic infections and asymptomatic infections. Two separate pathways are initially determined by the symptomatic ratio $f$, which is the probability that an individual being infected and eventually becoming symptomatic. The basic reproduction number $R_0$ is calculated and its value depends on the symptomatic ratio $f$, the values $R_i$ and $R_a$. 
The incorrect inferred value of $f$ leads to the wrongly predicted final epidemic size. It is however a challenging task to infer these three values in practice. We then proved that, for the age-of-infection model with two infection pathways, the final epidemic size is uniquely determined by $R_0$. The ratio of total numbers of symptomatic patients and asymptomatic patients is $\frac{f}{1-f}$. For the similar age-of-infection epidemic model with disease deaths, the calculation of basic reproduction number $R_0$ is not affected. And the final epidemic size can not be explicitly generated, but can be approximated by the given lower bounds and upper bounds. If the death proportions for both symptomatic and asymptomatic patients are smaller, the final epidemic size can be more accurately predicted. Two sets of numerical simulations were performed for these two types of models.

The age-of-infection epidemic model is the general structure for different types of compartmental epidemic models (Brauer 2016) and it can be extended for other scenarios, such as the disease spread in an age-structured population with different activity levels, the impact of vaccination in the population and the behavioral changes in the population. With respect to the fundamental analysis, the complete analysis of corresponding characteristic equation for such models is still very challenging (Brauer 2016, 2005a, 2017). The final epidemic size which is derived from the general age-of-infection model can be applied to study different types of epidemics with different forms of functions $B(\cdot)$ and $\pi(\cdot)$, this is beneficial in the predictions of outcomes for epidemics in the long term. However, the final epidemic size relation strongly depends on the value of estimated basic reproduction number $R_0$. In practice, many factors such as the symptomatic ratio will affect the inference of $R_0$, and further lead to incorrect predictions. Therefore, an important problem of interest is the statistical inferences of $B(\cdot)$, $\pi(\cdot)$ and other model parameter values by collected time-series data, such as daily incidence data. The limitation of publicly available case numbers for epidemiological modeling was presented in Raimúndez et al. (2021). Thus, the development of new parameter estimation techniques for epidemiological models is crucial, especially for age-of-infection models. It has been discussed that the final epidemic size relation can be served as a tool to study the long-term behavior of the epidemic. On the other hand, the short-term analysis such as the estimation of the probability of a major outbreak for such type of models is difficult. This is due to the fact that it is technically challenging in the formulation of corresponding stochastic epidemic models. Another problem of interest is to investigate how initial accumulation of incidences is connected with shape function $B(\cdot)$. The initial accumulation of incidences is highly sensitive to $B(\cdot)$, so one reasonable method is to consider $B(\cdot)$ as a kernel function and then express the solution in terms of $B(\cdot)$. The regularity of solution may reveal the pattern of initial accumulations of incidences and infectivity, which is important in understanding the outbreak stage of an epidemic.

Acknowledgements The work is dedicated to the author’s mentor Professor Fred Brauer (1932.02.03–2021.10.17). This work is based on numerous discussions between the author and Fred in the year of 2020 and the first half of year 2021. The author thanks one reviewer for very constructive suggestions to improve this work. The work was supported in part by the BNU-HKBU UIC startup research grant UICR0700037-22 and by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, project code 2022B1212010006 and UIC research Grant R0400001-22.
Declarations

Conflict of interest  None.

References

Bai F (2016) Vaccination models in infectious diseases. PhD thesis, University of British Columbia
Bai F (2020) Evaluating different epidemiological models with the identical basic reproduction number \( R_0 \). J Biol Dyn 14(1):849–870
Bai F, Huff KES, Allen LJS (2018) The effect of delay in viral production in within-host models during early infection. J Biol Dyn 13(sup1):47–73
Brauer F (2005a) Age of infection in epidemiology models. Electron J Differ Equ (EJDE) [electronic only] pp 29–37
Brauer F (2005b) The Kermack–McKendrick epidemic model revisited. Math Biosci 198(2):119–131
Brauer F (2016) Age of infection epidemic models. In: Mathematical and statistical modeling for emerging and re-emerging infectious diseases. Springer, Berlin, pp 207–220
Brauer F (2008a) Age-of-infection and the final size relation. Math Biosci Eng 5(4):681–690
Brauer F (2008b) Epidemic models with heterogeneous mixing and treatment. Bull Math Biol 70(7):1869–1885
Brauer F (2017) Mathematical epidemiology: past, present, and future. Infect Dis Modell 2(2):113–127
Brauer F, Castillo-Chavez C (2012) Mathematical models in population biology and epidemiology. Springer, New York
Brauer F, Watmough J (2009) Age of infection epidemic models with heterogeneous mixing. J Biol Dyn 3(2–3):324–330
Brauer F, Castillo-Chavez C, Feng Z (2019) Mathematical models in epidemiology. Springer, New York
Brauer F, van den Driessche P, Wu J (2008) Mathematical Epidemiology. Springer, Berlin. https://doi.org/10.1007/978-3-540-78911-6
Britton T (2010) Stochastic epidemic models: a survey. Math Biosci 225(1):24–35
Centers for Disease Control and Prevention. Symptoms of COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html (2021)
Feng Z, Xu D, Zhao H (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control. Bull Math Biol 69(5):1511–1536
Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 115(772):700–721
Li C, Zhu Y, Qi C, Liu L, Zhang D, Wang X, She K, Jia Y, Liu T, He D, Xiong M, Li X (2021) Estimating the prevalence of asymptomatic COVID-19 cases and their contribution in transmission—using Henan Province, China, as an example. Front Med. https://doi.org/10.3389/fmed.2021.591372
McCulloh I, Kiernan K, Kent T (2020) Inferring true COVID-19 infection rates from deaths. Front Big Data. https://doi.org/10.3389/fdata.2020.565589
Nishiura H, Kohayashi T, Miyama T, Suzuki A, Jung S, Hayashi K, Kinoshita R, Yang Y, Yuan B, Akhmetzhanov AR, Linton NM (2020) Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis 94:154–155
Raimúndez E, Dudkin E, Vanhoefer J, Alamousi E, Merkt S, Fuhrmann L, Bai F, Hasenauer J (2021) COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modeling. Epidemics 34:100439
Rothé C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C, Guggemos W, Seilmaier M, Drostén C, Vollmar P, Zvirglmaier K, Zange S, Wölfel R, Hoelscher M (2020) Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med 382(10):970–971
Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadam SM, Singer BH, Galvani AP (2021) Asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis. Proc Natl Acad Sci 118(34):e2109229118
Selkirk T (1983) On the asymptotic distribution of the size of a stochastic epidemic. J Appl Probab 20(2):390–394
Subramanian R, He Q, Pascual M (2021) Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. Proc Natl Acad Sci 118(9):e2019716118
van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180(1–2):29–48
Yang CK, Brauer F (2008) Calculation of R0 for age-of-infection models. Math Biosci Eng 5(3):585–599

Publisher’s Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.