**Methods**

We adopted a method in our previous study regarding the age-varying susceptibility to the Delta variant (doi:10.1001/jamanetworkopen.2022.3064 (2022)).

*Overview*

We built an age-structured compartment model stratified into 5-year age bands. We assumed that people are initially susceptible (*S*) and become exposed (*E*) after an effective contact with an infectious person. After a latent period, exposed individuals become infectious, either with a pre-symptomatic state ($I\_{presym}$) followed by symptomatic infection ($I\_{sym}$), or with an asymptomatic state ($I\_{asym}$). After the infectious period, individuals enter the removed state due to isolation (*Q*).



**Figure. A schematic plot of different time periods of the transmission of SARS-CoV-2.**

In a compartmental model, the force of infection $λ$, the rate at which susceptible individuals are infected (i.e exposed), is a crucial factor. The age-specific force of infection $λ\_{i}$ in age group $i$ at discrete time $t$ could be written as:

$$λ\_{i} = μ\left(q\_{i}\right)= \frac{q\_{i}\sum\_{j=1}^{A}ϕ\_{ij}\left(t\right)\left(I\_{presym}^{j}\left(t-1\right)+ I\_{sym}^{j}\left(t-1\right)+ 0.5I\_{asym}^{j}\left(t-1\right)\right)}{n\_{i}} (S1)$$

where $q\_{i}$ is the probability that a contact between a susceptible in age group $i$ and infectious person leads to infection, $ϕ(t) = ϕ\_{ij}(t)$ means contact matrix at discrete time $t$ ($ϕ\_{ij}(t)$ is the number of contacts an individual of age group $j$ makes with those of age group $i$ per unit time at discrete time $t$), $n\_{i}$ is the number of individuals in age group $i$. We suppose that the relative infectiousness of the $I\_{asym}$is half of $I\_{presym} $or $I\_{sym}$. Here $i$ is in age group $\overline{A}$ $= \{1,2,\cdots ,A\}$.

We try to estimate $q\_{i}$ during the 3rd, 4th, and 5th waves in South Korea and denote it as $θ\_{i}$ hereafter. If we could observe the number of exposed individuals at time $t$ for age group $i$ and the number of total infectious individuals, the likelihood of the parameters could be easily derived. However, what we could observe is the number of diagnosed (i.e quarantined) individuals for age group $i \in \overline{A}$ at time $t, $only. In addition, asymptomatic infection which is the notable feature of COVID-19 should also be reflected in the model. To resolve these difficulties, we use a Bayesian approach. In particular, we develop an efficient MCMC (Markov Chain Monte Carlo) algorithm in which the exposed date, symptom onset date and transmission onset date for all quarantined individuals are imputed with an assumption that there are pre-determined proportion of asymptomatic individuals. We explain the details of our Bayesian method in the following three subsections.

*Data, Model and Posterior*

The data we used in the analysis is daily numbers of quarantined individuals for each age group from 15 October to 22 December 2020 (3rd wave), from 27 June to 21 August 2021 (4th wave), and from January 1 to January 31, 2022 (5th wave).

To estimate $θ\_{i}$, we are going to impute the exposed dates, symptom onset dates and transmission onset dates of all quarantined individuals conditional on given quarantined dates. For this purpose, we need a probability model which relates the exposed dates, symptom onset dates and transmission onset dates to the quarantined dates.

|  |  |
| --- | --- |
| Symptomatic cases | Asymptomatic cases |
| Symbol | **Definition** | **Symbol** | **Definition** |
| *E* | Exposed date of an individual | ***E*** | Exposed date of an individual |
| *Y* | Incubation period | ***L*** | Latent period |
| *I* | Transmission onset time relative to the symptom onset | ***C*** | Infectious period suspended by quarantine |
| *D* | Diagnostic delay (i.e Quarantined) from the symptom onset | ***R*** | Infectious period |

For each symptomatic individual, the quarantined date is sum of the exposed date ($E$), the incubation period ($Y$) and the period for symptom onset to diagnostic delay ($D$). In addition, these individuals start infecting other susceptibles from the transmission onset date ($E+Y+I$).

For each asymptomatic individual, the quarantined date is sum of exposed date ($E$), latent period ($L$) and period for transmission onset to quarantined date ($C$). We assume that the latent period distribution of asymptomatic individuals is same as that of symptomatic individuals. Also, the quarantined time distribution of $C(f\_{C})$ of asymptomatic individuals is set to be the exponential distribution with mean 1/1.7, which satisfies $P(C>R)=0.01$ where $E+L+R$ is defined as the recovered date.

Finally, the quarantined date $T$ is defined as

$$T = \left(E+Y+D\right)I\left(Δ=0\right)+ \left(E+L+C\right)I\left(Δ=1\right) (S2)$$

where $Δ$ is equal to 0 when the individual is symptomatic and 1 when asymptomatic. We set $P(Δ=1)=a\_{i}$, which denotes the proportion of asymptomatic cases in age group $i$.

For individual $k$, let $W\_{k} = (E\_{k},Y\_{k},I\_{k},L\_{k},C\_{k},D\_{k})$ and $V\_{k} = (Y\_{k},I\_{k},L\_{k},C\_{k},D\_{k})$. Let $D$ be the observed data which consist of the daily numbers of quarantined individuals. Our strategy to estimate $θ\_{i}$ is to generate $W\_{k}$ and $θ$ iteratively from their conditional posterior distributions $P(θ\_{i}|W,\overline{D})$ and $P(W\_{k}|W\_{(-k)},θ,\overline{D})$ respectively, where $W =\{W\_{k}\}$ and $W\_{(-k)}$ denotes $W$ except $W\_{k}$.

We could describe as,

$$P\left(W\_{\left(-k\right)},θ\_{i\_{k}},\overline{D}\right)= P\left(W\_{\left(-k\right)},θ\_{i\_{k}},D\right)I\left(Δ\_{k}=0\right)+ P\left(W\_{\left(-k\right)},θ\_{i\_{k}},D\right)I\left(Δ\_{k}=1\right) (S3)$$

$$= P(E\_{k}|V\_{\left(1:N\right),}θ\_{i\_{k}},D)(f\_{Y}(Y\_{k})f\_{I}(I\_{k})f\_{D}(D\_{k})I(Δ\_{k}=0) +f\_{L}(L\_{k})f\_{C}(C\_{k})I(Δ\_{k}=1)) (S4)$$

*Generating* $θ$ *and* $W$ *from their conditional posterior distributions*

*Generating* $θ$

For the prior distribution of $θ\_{i}$ we use a diffuse gamma distribution $Gamma(0.001,0.001)$ for all $i$. Then

$$p(θ\_{i}|W\_{\left(1:n\_{i}\right)}^{i},\overline{D}) ∝ p(θ\_{i})p(W\_{\left(1:n\_{i}\right)}^{i},\overline{D}|θ\_{i}), i\in \overline{A} (S5)$$

when $W^{i} = \{W\_{k} | The age group of k is i\}$, $n\_{i}$ is the population for age group $i$.

In turn, $P(W\_{1}^{i},\cdots ,W\_{n\_{i}}^{i}, \overline{D}|θ\_{i})$ can be expressed as

$$P\left(θ\_{i}\right)=P\left(V\_{\left(1:N\right)},θ\_{i}\right)P\left(V\_{\left(1:N\right)}\right)= \prod\_{k=1}^{n\_{i}}P(E\_{k}^{i}\left| V\_{\left(1:N\right)},θ\_{i}\right)P\left(V\_{\left(1:N\right)}\right) (S6)$$

$P(E\_{k}^{i}| V\_{(1:N)},θ\_{i})$ = $P(E\_{k}^{i}| I\_{total}^{j}\left(t\right), t<E\_{k}^{i} , θ\_{i}) = p(E\_{k}^{i} | θ\_{i})\prod\_{t=1}^{E\_{k}^{i}-1}(1-p(t|θ\_{i})) (S7)$

Since $p(E\_{k}^{i}|V\_{(1:N)},θ\_{i})$ is the probability of an individual $k$ to be infected at discrete time $E\_{k}^{i}$ (implying the individual $k$ has not been infected before).

$$P\left(V\_{\left(1:N\right)}\right)= \prod\_{∆\_{k}=0}^{}(1-a\_{i\_{k}})f\_{Y}\left(Y\_{k}\right)f\_{I}\left(I\_{k}\right)f\_{D}\left(D\_{k}\right)\prod\_{∆\_{k}=1}^{}a\_{i\_{k}}f\_{L}\left(L\_{k}\right)f\_{C}\left(C\_{k}\right) (S8)$$

where $N =\sum\_{i}^{}n\_{i}$ (Total population size), $V\_{k} = (Y\_{k},I\_{k},L\_{k},C\_{k},D\_{k})$ and $i\_{k}$ is the age group of individual $k$. By applying the following approximation,

$$P(E\_{k}^{i}|V\_{\left(1:N\right)},θ\_{i\_{k}}) = P(E\_{k}^{i}| I\_{total}^{j}(t) , t<E\_{k}^{i} , θ\_{i\_{k}}) = p(E\_{k}^{i} | θ\_{i\_{k}})\prod\_{t=1}^{E\_{k}^{i}-1}(1-p(t|θ\_{i\_{k}}))≈ p(E\_{k}^{i} \left| θ\_{i\_{k}}\right)e^{-\sum\_{t=1}^{E\_{k}^{i}-1}p\left(θ\_{i\_{k}}\right)} (S9)$$

we have

$$θ\_{i} | W, D ∼ Gamma\left(0.001 + \left|ℇ\_{i}^{\*}\right|, 0.001 + \sum\_{k \in ℇ\_{i}^{\*}}^{}\frac{\sum\_{t\_{1}}^{E\_{k}}ϕ\_{ij}\left(t\right)I\_{total}^{j}\left(t-1\right)}{n\_{i}}+\sum\_{k \in \left\{s : E\_{s}>t\_{2} or E\_{s}=NA\right\}}^{}\frac{\sum\_{t\_{1}}^{t\_{2}}ϕ\_{ij}\left(t\right)I\_{total}^{j}\left(t-1\right)}{n\_{i}}\right) (S10)$$

where $ℇ\_{i}^{\*}$ is set of individuals in the age group $i$ exposed during the 3rd wave $[t\_{1},t\_{2}]$ (or 4th and 5th waves).

*Generating* $W$

We generate $W$ by generating $W\_{k}$ from $P(W\_{k}|W\_{(-k)},θ,\overline{D})$ iteratively, and generate $W\_{k}$ through the Metropolis-Hasting (MH) algorithm. To sample from the posterior $P(W\_{k}|W\_{(-k)},θ,\overline{D})$ by the MH algorithm, we use the following proposal distribution:

$$Q\left(E\_{k},V\_{k}\right)= Q\left(V\_{k}\right)Q\left(V\_{k}\right) (S11)$$

$$Q\left(V\_{k}\right)= p\left(\overline{D}\right) (S12)$$

$$Q\left(V\_{k}\right)= δ\left(T\_{k}-Y\_{k}-D\_{k}\right)I\left(Δ\_{k}=0\right)+ δ\left(T\_{k}-L\_{k}-C\_{k}\right)I\left(Δ\_{k}=1\right) (S13)$$

Putting the above together, the sampling procedure of $W\_{k}$ is summarized in Alogorithm S1.

**Alogorithm S1. Bayesian Inference**

**---------------------------------------------------------------------------------------------------------------------------------------**

Input : $W\_{k}^{(0)}$ for $k = 1, 2, \cdots , N$

 1. Sample $θ^{(0)} = (θ\_{i}^{(0)})$ from prior.

 2. **for** $m = 1:M$(number of iteration) **do ► Gibbs sampling**

3. **for** $k = 1:N$do **►** **MCMC**

 4. Sample $W\_{k}^{(m)}$ from $Q(E\_{k},V\_{k})$

 5. $α \leftarrow \frac{P(E\_{k}^{(m)},E\_{(-k)}^{(m-1)}| V\_{k}^{(m)}, V\_{(-k)}^{(m-1)}, θ^{(m)} ,\overline{D})}{P(E\_{k}^{(m-1)},E\_{(-k)}^{(m-1)}| V\_{k}^{(m-1)}, V\_{(-k)}^{(m-1)}, θ^{(m)} ,\overline{D})}$ **► Acceptance ratio**

 6. $W\_{k}^{(m)} \leftarrow \left\{\begin{array}{c}W\_{k}^{(m)} α \geq 1\\W\_{k}^{(m)} with prob \\W\_{k}^{(m-1)} else\end{array}\right.α$

 7. Sample $(θ\_{i}^{(m)})$ for $i = 1,2,\cdots ,A$ from $p(θ\_{i}|W^{(m)})$

**---------------------------------------------------------------------------------------------------------------------------------------**