An interpretation of COVID-19 in Tokyo using a combination of SIR models

By Koichiro Maki*1,†

(Edited by Toshimitsu Yamazaki, M.I.A.)

Abstract: A year and a half has passed since the outbreak of the COVID-19 pandemic. Mathematical models to predict infection are expected and many studies have been conducted. In this study, a new interpretation was created that could reproduce the daily positive cases in Tokyo using only a simple SIR model. In addition, the data on the ratio of transfer to delta variants could also be simulated. It is anticipated that this interpretation will be a basis for the development of forecasting methods.

Keywords: COVID-19, SIR model, complete mixing, basic community

1. Introduction

Over the past year and a half, COVID-19 has caused significant damage to the health of people and economic activities around the world.1) As a countermeasure, technology to forecast the spread of infection using mathematical models is indispensable. The SIR model (susceptible, infected, recovered) has been proposed as a mathematical approach,3)–5) and further development, such as the SEIR model, has been advanced.6)–8)

Despite the simplicity and effectiveness of SIR-based models in predicting a variety of other infectious diseases, communities in many countries have reported inconformity in their applicability to COVID-19.9)–15) Therefore, it is necessary to further improve the SIR model and to develop a prediction tool for common infectious diseases, including COVID-19.

In this study, the applicability of the SIR model was scrutinized using data from newly positive individuals in Tokyo. A noteworthy point of view is the condition that the system in which the model holds is completely mixed.16)

2. Characteristics of the SIR model and its arguments

In the basic SIR model, differential equations of the susceptible fraction \( S(t) \) infected fraction \( I(t) \) and recovered fraction \( R(t) \) for each time are shown as follows:

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI, \\
\frac{dI}{dt} &= \beta SI - \gamma I, \\
\frac{dR}{dt} &= \gamma I,
\end{align*}
\]

where \( \beta \) is the infection rate and \( \gamma \) is the recovery rate. Further, the sum of fractions is equal to 1 for each time:

\[ S + I + R = 1. \]

R0 is the basic reproduction number, as follows:

\[ R_0 = \frac{\beta}{\gamma}. \]

Furthermore, the fraction of new positives (daily positive cases) is denoted by \( P(t) \), and since it is the reduction of the susceptible fraction per day, it can be obtained by the following equation:

\[ P = -\frac{dS}{dt} = \beta SI. \]

From Eqs. [1] and [3], a formula for the relationship between \( S \) and \( R \) is obtained as:
\[ S = \eta \exp \left( -\frac{\beta}{\gamma} R \right), \quad [7] \]

where \( \eta \) is the initial value of \( S \) when \( R = 0 \), and its corresponding time is the start of the infection, \( t = 0 \). Then, \( \eta \) has the form

\[ \eta = 1 - I(0). \quad [8] \]

The solution of the SIR model can be obtained numerically as a relationship between \( R \) and time from Eqs. [3], [4] and [7].

If the spread of infection starts with one person, the initial condition \((I(0))\) in a fractional representation requires:

\[ I(0) = \frac{1}{N}. \quad [9] \]

An important point is found in Eq. [6]. It is that the number of new positive cases \((P)\) is proportional to the number of infected \((I)\). The fact that the proportionality coefficient \((\beta S)\) has the same value around all infected persons is an approximation using a constant \((\beta)\) and an average ratio of the susceptibility. In the region where this approximation using the average value (mean approximation) holds, the calculated values of the theory should correspond closely to the measured values. Furthermore, since the range of this region is not infinite, it must be closed at the boundary of a region with a lower average infection rate. If, on the other hand, the average value of the external region is high, it will be one of the fluctuations in the fast and large external spread of infection. The same argument has to be repeated for the larger region. Within a closed region, there should be a saturation effect as the spread of infection progresses. In the mathematical model, this saturation effect is expressed by the proportionality coefficient \((S)\).

In order to determine the population size of the region, it is necessary to determine the characteristics of such a closed system from real data. The characteristics are as follows: exponential rising due to a constant infection rate and exponential falling due to a saturation effect. Each closed region with these characteristics is a kind of “subcommunity”, and its population can be estimated from the mean approximation model. If the mean approximation is adequate, this indicates that there is sufficient mixing in the region. Next, how far the region of sufficient mixing extends is important information to be acquired, which is to be inferred from actual data. Therefore, applying the SIR model to an unreasonably large system makes it difficult to reconcile the shapes, and it is worthwhile only in a regional range where the recovered and infected are completely mixed. It is the condition of “Complete Mixing”\(^{16}\) that the SIR model requires. We define the people included in this complete mixed region as a “Basic Community”.

Eventually, a larger population region should be an aggregation of these perfectly mixed subcommunities. To be precise, it should be possible to express the shape of the data over a long period of time in large population regions by a linear combination of functions of the SIR model solution. This method allows us to extract basic information about the community that is required to adapt the mathematical model to the data of a new positive case: the size of the population \((N)\), the number of communities present at the same time \((m)\), the infection rate \((\beta)\), and the time when the infection started (1st day).

According to the above concept, the actual number of daily positive cases in the basic community is represented as

\[ \psi(t, N, \beta, \eta) = N \cdot P(t, \beta, \eta), \quad [10] \]

where \( R_0 \) can be used instead of \( \beta \), because \( \gamma \) should have the same value for all communities.

Figure 1 shows the time variation of the number of \( I(t) \) in each parameter. As the population of the basic community increases, the number of infected people increases proportionally, and at \( R_0 = 2.7 \), 26,000 infected people (hospitalized patients) are estimated for a total community of 100,000 people. For \( R_0 = 2.7, 2.0 \) and 1.5, there is a steep slope in the

![Fig. 1. Time variation of the number of infected people for each R0, N and \( \gamma = 0.1 \).](image-url)
spread of infection after about 2, 3 and 6 months, respectively.

Figure 2 shows a profile of new positive patients. The maximum number of new positive cases in a community with a population of 100,000 at $R_0 = 2.7$, 2.0, 1.5 is 3340, 1750 and 660, respectively. Corresponding to each parameter, the peak position of $P$ is shifted by about 8 days earlier than $I$.

3. Result of adaption to daily positive cases in Tokyo

The solution $(P)$ of the SIR model for the $i$-th basic community is expressed by the following function:

$$\psi_i = \psi(t, N_i, R_0, \eta_i),$$  \hspace{1cm}  \text{[11]}

where $t$, $N$, $R_0$ and $\eta$ are values of individual communities. Since the spread of an infection in a general community can be represented by a set of basic communities, it is a form of a simple combination of $\psi$ to reproduce the overall infection data,

$$\Psi = \sum_i \psi_i. \hspace{1cm} \text{[12]}$$

Figure 3 shows the data for daily positive cases in Tokyo (available from Tokyo Metropolitan Government’s new coronavirus countermeasure website (https://stopcovid19.metro.tokyo.lg.jp/, in Japanese and English) and NHK website (https://www3.nhk.or.jp/news/special/coronavirus/data/pref/tokyo.html, in Japanese)) and the theoretical curve $\Psi$ that reproduces the data. Each parameter that makes up the theoretical curve $(\Psi)$ is shown in Table 1. As a result, Fig. 3 shows a high degree of agreement between the data for Tokyo and the calculated values of the summed SIR model solution.

In Table 1, the second category “1st day” means the first day of the outbreak, and $m$ is the number of “basic communities” that are simultaneously spreading the infection. The notable point is that the first day of the outbreak is during the peak period of the previous expansion wave. At this time, it indicates that the buds that will cause the spread of infection a few months later have metastasized to other basic communities. In the table, solutions $A_3$ and $A_4$ are not identified as large waves in Japan.

The solutions $(P)$ of $\psi_7$ and $\psi_8$ in Table 1 are the two spreads of infection included in the fifth wave. Using the value of $R_0$ in the table and the size of the basic community, the ratio of the delta mutant strains issued by the Tokyo Metropolitan Govern-

![Fig. 2. Daily positive cases ($P$) for each parameter.](image1)

![Fig. 3. Daily positive cases in Tokyo and theoretical curve $\Psi$. The SIR solution ($\psi_i$) of the basic community used to combine $\Psi$ is shown in the figure.](image2)
The reported targets of analysis so far seem to be large and complex infectious situations, such as national and local governments. In such a large system, the SIR model is impossible to adapt without a convincing mean approximation. The nonconformity of the SIR model reported in Italy, China, and Iran is attributed to the inability to represent the variation in details due to the large population, N. In order for the SIR model to decompose the data, some of the features of the model need to be exposed in the shape of the data, and it is easier to find exposures in small communities than in large ones.

The overall spread of an infection, even in a large system, is explainable by a composite model that combines several basic communities. It is, for example, to create a mathematical model of Japan as a whole by superimposing the surrogate functions of the SIR solution constructed for individual prefectures.

Table 1. Parameter set for reproducing the data of daily positive cases in Tokyo. Equation [12] was fitted to the Tokyo data with these parameters (see Fig. 3).

| Solution No. | 1st day   | β (R0 = β/γ) | N   | m   | Remarks |
|-------------|-----------|--------------|-----|-----|---------|
| ψ1          | 2020/2/5  | 0.22         | 6000| 1   | 1st     |
| ψ2          | 2020/4/24 | 0.17         | 5000| 5   | 2nd     |
| ψ3          | 2020/7/20 | 0.21         | 5000| 1   | none    |
| ψ4          | 2020/8/10 | 0.16         | 5000| 9   | 3rd     |
| ψ5          | 2020/11/28| 0.31         | 8000| 4   | 3rd     |
| ψ6          | 2021/1/11 | 0.18         | 20000| 3  | 4th     |
| ψ7          | 2021/5/6  | 0.2          | 6500| 5   | 5th     |
| ψ8          | 2021/5/28 | 0.235        | 40000| 4  | 5th, δ (L452R) |

Fig. 4. Proportion of positive cases of the L452R(delta) variant compared to other strains. The staircase data are the actual values of the one-week average in Tokyo. The solid line shows the ratio of the calculated values of the other strains (ψ) to the delta strains (ψ₈).

4. Discussion

The reported targets of analysis so far seem to be large and complex infectious situations, such as national and local governments. In such a large system, the SIR model is impossible to adapt without a convincing mean approximation. The nonconformity of the SIR model reported in Italy, China, and Iran is attributed to the inability to represent the variation in details due to the large population, N. In order for the SIR model to decompose the data, some of the features of the model need to be exposed in the shape of the data, and it is easier to find exposures in small communities than in large ones.

The overall spread of an infection, even in a large system, is explainable by a composite model that combines several basic communities. It is, for example, to create a mathematical model of Japan as a whole by superimposing the surrogate functions of the SIR solution constructed for individual prefectures.

The parameter set given in Table 1 is one of the solutions of the mathematical model for understanding the infection situation in Tokyo, a city with a population of 14 million. The communities that actually caused the spread of infection are classified into eight (ψ₁ to ψ₈) locations by the start time. It is estimated from the corresponding peaks that there are tens or hundreds of thousands of people. The members forming this basic community are connected by face-to-face relationships in human behavior, which does not imply a community separated by physical space. Although such a community is constantly recombining, it is regarded as if it had always existed as a mathematical model of infection.
In this picture, Tokyo appears to be as follows: there are many basic communities with high infection rates scattered among a huge community with low infection rates spread all over the place, as well as several neighboring communities with high infection rates which were infected in a chain reaction. It's just as if piles of dead grass dotting a lawn are ignited sequentially.

The fact that the first day of infection is near to the preceding peak, the probability of “the fire jumping to the next pile” is highest at the time of the most intense burning of the previous pile. As found in Fig. 2, the time to reach the peak is three to four months for R0 = 2.0, and the half-width of the peak is within two months. With this feature, the spread of the infection is discrete, like a series of waves that are triggered sequentially.

One basic community is sufficient to explain the fifth wave, yet, in addition, two communities, ψ7 and ψ8, are provided to explain the data on the delta strain (L452R). Consequently, the infection spread appears to be one basic community sometimes becomes a combination of several communities as more detailed data is acquired.

In regards to the peak of the fifth wave, which produced a large number of infections, the infection rate (β = 0.235) is not much different from the previous peak, so the conspicuous difference is the large population of one basic community.

The basic community is a closed system that is instantaneously affected by recovered and infected people under completely mixed conditions. The time required for that complete mixing is at least until the peak of the infection is reached, since it is sufficient that the peak is formed due to the saturation effect. Within this closed system, the number of infections rises exponentially, reaches a maximum value, and then drops off exponentially. If the population of the basic community is even larger, the time dependence of the number of infected people will exhibit steeper ups and downs, even for the same infection rate. The reason why the mathematical model fits so closely with the data on the number of infections in Tokyo is due to the simplicity of the model, which consists of a few isolated large well-formed waves.

Ultimately, any analysis of the actual situation of an epidemic should focus on elucidating the underlying basic communities and their trends. It will then be possible to forecast infections, and to avoid diffusion of the disease among communities as a preventive measure.

References

1) Sohrabi, C., Alsafi, Z., O’Neill, N., Khan, M., Kerwan, A., Al-Jabir, A. et al. (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int. J. Surg. 76, 71–76.
2) Wang, L.-S., Wang, Y.-R., Ye, D.-W. and Liu, Q.-Q. (2020) A review of the 2019 novel coronavirus (COVID-19) based on current evidence. Int. J. Antimicrob. Agents 55, 105948.
3) Kermack, W.O. and McKendrick, A.G. (1927) A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A 115, 700–721.
4) Brauer, F., Van den Driessche, P.V. and Wu, J. (2008) Mathematical Epidemiology. Springer, Berlin.
5) Giuseppe, G. (2020) A simple SIR model with a large set of asymptomatic infectives. arXiv:2003.08720v4 (revised and augmented version).
6) Carcione, J.M., Santos, J.E., Bagaini, C. and Ba, J. (2020) A simulation of a COVID-19 epidemic based on a deterministic seir model. Front. Public Health 8, 230.
7) Chowell, G., Sattenspiel, L., Bansal, S. and Viboud, C. (2016) Mathematical models to characterize early epidemic growth: a review. Phys. Life Rev. 18, 66–97.
8) Weinstein, S.J., Holland, M.S., Rogers, K.E. and Barlow, N.S. (2020) Analytic solution of the SEIR epidemic model via asymptotic approximant. Physica D 411, 132633.
9) Moein, S., Nickaeen, N., Rooiintan, A., Borhani, N., Heidary, Z., Javanmard, S.-H. et al. (2021) Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. Sci. Rep. 11, 4725.
10) Hu, Z., Ge, Q., Jin, L. and Xiong, M. (2020) Artificial intelligence forecasting of Covid-19 in china. arXiv:2002.07112.
11) Maier, B.F. and Brockmann, D. (2020) Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 368, 742–746.
12) Postnikov, E.-B. (2020) Estimation of COVID-19 dynamics “on a back-of-envelope”: does the simplest SIR model provide quantitative parameters and predictions? Chaos Solitons Fractals 135, 109841.
13) Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippo, A.-D., Matteo, A.-D. et al. (2020) Modeling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. 26, 855–860.
14) Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S. et al. (2020) A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int. J. Infect. Dis. 93, 211–216.
15) Roda, W.C., Varughese, M.B., Han, D. and Li, M.Y. (2021) Why is it difficult to accurately predict
the COVID-19 epidemic? Infect. Dis. Model. 5, 271–281.

16) Seno, H. (2012) Reproduction numbers of infectives for a time-discrete epidemic population dynamics model. RIMS Kôkyûroku 1789, 35–45 (in Japanese with English abstract).

17) Postnikov, E.B. (2021) Reproducing country-wide COVID-19 dynamics can require the usage of a set of SIR systems. PeerJ 9, e10679.

(Received Oct. 11, 2021; accepted Nov. 30, 2021)