

Modeling and analysis of COVID-19 infected persons during repeated waves in Japan

Koichi Hashiguchi

Correspondence : khlambda-2020@memoad.jp

Abstract

A model for estimating the number of COVID-19 infected persons (infecteds) is proposed based on the exponential function law of the SIR model. This model is composed of several equations expressing the number of infecteds, considering the onset after an incubation period, infectivity, wavy infection persistence with repeated infection spread and convergence with insufficient quarantine. This model is applied to the infection in Japan, which is currently suffering from the 5th wave, and the initial number of infecteds and various related parameters are calculated by curve fitting of the cumulative number of infecteds to the total cases in the database. As a minimum boundary of the number of infecteds for the infection continuation up to the 5th wave, the initial number of infecteds at the outbreak of infection is more than an order of magnitude higher than the actual initial cases. A convergence ratio (cumulative number of infecteds / total cases) at the end of the first wave is obtained as 1.5. The quarantine rate and social distancing ratio based on the SIQR model are evaluated, and the social distancing ratio increases sharply just after the government's declaration of emergency. The quarantine rate closely equals the positive rate by PCR tests, meaning that the number of infecteds, which had been unknown, is on the order of almost the same as the number of tests.

Introduction

A new type of coronavirus, COVID-19, which took off in Wuhan, China, at the end of 2019 and quickly spread all over the world, has not been settled even after one year and a half, with repeated waves of spread and convergence of the infection.

The SIR model ¹⁾ proposed about 100 years ago is known as a mathematical model for describing the epidemic process of infection. It is a simple model that expresses the increase/decrease in the number of persons between groups S (Susceptible), I (Infected), and R (Recovered) with differential equations under the population conservation law. This model reproduced the plague epidemic of the early 20th century and is considered to help understand the current COVID-19 epidemic ²⁾. Furthermore, an improved SIQR model ³⁾ was also proposed, in which the recovered group, R in the

SIR model was divided into Q (quarantined), and R (recovered). Odagaki ⁴⁾ proposed SIQR model with two important parameters, the quarantine rate and the social distancing ratio. These explicitly introduced parameters are related with the exponent λ of an exponential function for the daily variation in the number of new infection-positive persons and contribute to the evaluation of infection control measures.

In the previous papers ^{5,6)} intended to analyze these parameters, mathematical analyses between these relationships were conducted on Taiwan and South Korea with measures of mass PCR tests in the early stages of infection, and Western countries with measures of lockdown, including Japan, and clarified the difference in the quarantine rate-social distancing ratio relationship caused by these countermeasures.

In this paper, a model estimating the number of infected persons, which has been unknown in the previous papers, is proposed with consideration of an incubation period after infection, subsequent infection rate, and the convergence ratio at the stage of wavy infection convergence. This model also presupposes the exponential spread of infection common to the SIR and SIQR models under the law of conservation of population. This model is applied to a database of infection status, and the daily variation of the infected persons is evaluated by fitting the calculated cumulative number of infected persons to the actual value of the cumulative number of newly detected infection-positive persons. Using the results, the variation of infected persons, quarantine rate, and social distancing ratio are analyzed during five waves in Japan.

Basic equations in the model

Basic equations in this estimation model are described separately for the derivation of the cumulative number of infected persons, and the derivation of the quarantine rate / social distancing ratio. Since the latter parameters have been described in detail in the previous papers, only the resulting equations are described.

Basic equations for the cumulative number of infected persons

Assuming the cumulative number of infected persons (hereinafter referred to as cumulative infecteds) Z_i on d_i day increases exponentially, Z_i is expressed by the exponential function (1) using an exponent α_i .

$$Z_i = Z_{i-1} \cdot e^{\alpha_{i-1}} \quad (1)$$

The number of newly infected persons ΔZ_i is expressed by Eq. (2) as the increase in cumulative infecteds.

$$\Delta Z_i \equiv Z_i - Z_{i-1} = Z_{i-1}(e^{\alpha_{i-1}} - 1) \quad (2)$$

After infection, a latent disease declares itself after an incubation period. Assuming an incubation period follows the gamma distribution $f(d)$ with maximum onset on the 5th day after infection and ending on the 16th day ⁷⁾, which gives the shape parameter $a = 5$ and the scale parameter $b = 1.25$ in the gamma distribution formula. This gamma distribution curve is discretized, normalized so that the sum of $f(d)$ for 16 days is 1, and the ratio of onset on day d_j is expressed by a function gm_j in Eq. (3).

$$gm_j = f(d_j) / \sum_{j=1}^{16} f(d_j) \quad (3)$$

The number of infected persons ΔZ_i on day d_i is divided among 16 days d_{i+1} to d_{i+16} according to the gamma distribution, the number of onsets on day d_j is $\Delta Z_i \cdot gm_j$. Therefore, the number of onsets S_i on day d_i is the sum of the number of persons who have been infected from 16 days before and have completed the incubation period, Eq. (4) is obtained.

$$S_i = \sum_{j=1}^{16} \Delta Z_{i-j} \cdot gm_j \quad (4)$$

The number of infected persons (hereinafter referred to as infecteds) X_i on d_i day is expressed in Eq. (5), as the sum of infecteds X_{i-1} on a previous day and the number of cases S_i that completed the incubation period with subtracting the number of newly detected and quarantined persons y_i (new cases).

$$X_i = X_{i-1} - y_i + S_i \quad (5)$$

The cumulative infecteds Z_i is given by Eq. (6).

$$Z_i = Z_{i-1} + S_i \quad (6)$$

In these basic equations, setting the initial infecteds at the outbreak of infection as X_0 , which is equal to the cumulative infecteds Z_0 on the infection outbreak, and this initial value is substituted into Eq. (2). By repeating the sequential calculation from equations (4) to (6), the infecteds and the cumulative infecteds during a predetermined period are obtained. The basic equations mentioned above are all linear form, and Z_i in Eq. (6) increases and decreases linearly corresponding to the increase and decrease in the initial value Z_0 . The calculated cumulative infecteds are analyzed by referring to the transition data of the cumulative number of new cases (total cases z_i) in the database. In these basic expressions, uppercase variables such as Z_i , X_i , S_i are expressed as calculated values related to infecteds, and lowercase variables such as y_i , z_i are expressed as actual values related to cases from the database.

Infection progresses in a wavy way in which the number of new cases repeatedly increases and decreases, and in the calculation process for each wave, two parameters, IR (Infection Ratio) and CR (Convergence Ratio), are introduced to ensure consistency between waves. The former is defined as the proportion of infecteds capable to infect others after the incubation period, and ΔZ_i in Eq. (2) is replaced by Eq. (2').

$$\Delta Z_i = IR \cdot Z_{i-1} (e^{\alpha_{i-1}} - 1) \quad (2')$$

The latter CR is defined as a ratio of the calculated cumulative infecteds Z_i to the total cases z_i at the end of each wave. When $CR = 1$, excess infecteds becomes 0, which means all the cumulative infecteds are detected and quarantined as total cases during the period, resulting in termination of infection. Otherwise, excess infecteds with CR larger than 1 bring about the next spread of infection. By definition, both parameters can be varied within a range of $IR = 0 \sim 1$, $CR \geq 1$.

Basic equations for quarantine rate and social distancing ratio

The new cases y_i is expressed by Eq. (7) using an exponent λ .

$$y_i \propto \exp(\lambda_i \cdot d_i) \quad (7)$$

As already mentioned in the previous papers, the quarantine rate QR and the social distancing ratio SD (q and x , respectively in the previous papers) are shown by equations (8) and (9), respectively. The relationship of these parameters and λ_i in Eq. (7) is expressed by the following equations.

$$QR_i \equiv y_i / X_i \quad (8)$$

$$SD_i \cdot \beta N + QR_i = \lambda_{max} - \lambda_i = \Delta \lambda_i \quad (9)$$

$$\lambda_{max} = \beta N - \gamma \quad (10)$$

Both parameters are constrained in the range of $0 \leq SD_i, QR_i \leq 1$ by definition. βN is the transmission coefficient β multiplied by population N , which is an index of the degree of transmission due to contact between uninfected and infected persons. γ is the cure rate calculated as 0.04 reciprocal of the healing period set to 25 days²⁾. However, γ is set as $-\lambda_{min}$, which is obtained by substituting Eq. (9) into the constraint of $0 \leq SD_i \leq 1$, because 0.04 for γ may give results deviating from upper constraint for SD . The unit of λ is (1 / day), QR and βN have the same units, and SD is dimensionless. βN associated with the maximum λ in Eq. (10) is regarded as invariant throughout the infection period.

Analysis of COVID-19 infecteds in Japan

This model was applied for the analysis of the infection status in Japan. Daily variation of new cases, y_i and total cases, z_i from the "Our World in Data" (January 1, 2020-August 4, 2021) database⁸⁾ were converted to per million population, as well as the number of PCR tests described later. As a preliminary preparation, the daily α_i and λ_i were calculated by the logarithmic regression of Eq. (1) and (7) using those converted data. In the regression calculation, the exponent was calculated as the moving-average logarithmic slope for 13 days, and gm_j of Eq. (3) was also calculated. The variation of new cases, total cases, and the corresponding exponents λ and α are shown in Fig. 1a,

and b. In Fig. 1a, the new cases repeatedly increase and decrease exhibiting four waves, and currently is rapidly increasing as the fifth wave. In the figure, the converging points with minimum new cases between peakss are indicated by red dots. The total cases exhibit a gentle stepwise increase near the converging points with a rapid increase at the maximum points of new cases in each wave. The variation of exponents α_i and λ_i shown in Fig. 1b are mathematically correlated as differential and integral calculus. The major difference is that λ can be positive and negative, however, α is always 0 or positive.

Using the above basic equations, the cumulative infecteds Z_i is calculated sequentially day by day for a given initial value $X_0(=Z_0)$ at the outbreak of infection. Any calculated Z_i curve above the transition z_i curve in Fig. 1a is acceptable, however, the procedure for obtaining the optimum solution is performed by finding the lower limit for fitting the calculated value Z_i to the actual value z_i transition data. In the following calculation, the infection ratio IR_n and convergence ratio CR_n were kept constant in the same wave and varied for each wave n ($n = 1$ to 4). Since $Z_i = CR_n \cdot z_i$ is set at a converging point of each wave, increase in CR_n increases Z_i at the converging point, while decrease in IR_n decreases Z_i by Eq. (2'). Therefore, the increase/decrease in both parameters has an adjusting function of Z_i between waves at the converging point of each wave. Since there was no termination of infection during this calculation period, cumulative infecteds is always greater than or equal to the actual value ($Z_i \geq z_i$), and infecteds is restricted to be greater than or equal to the new cases ($X_i \geq y_i$).

- First wave

The outbreak day of infection is set on January 27, 2020, as a day of $\lambda_{max} = 0.272$, and an arbitrary value is set for the initial value $X_0 (=Z_0)$ on this day. The daily variation of Z_i are sequentially calculated using α_i and y_i in Fig. 1b based on Eq. (2') to Eq. (6) via daily ΔZ_i , S_i , X_i to the converging point of the first wave. The convergence calculation is applied to find X_0 , which minimizes the difference between Z_i and $CR_1 \cdot z_i$ at the converging point of 1st wave. On this

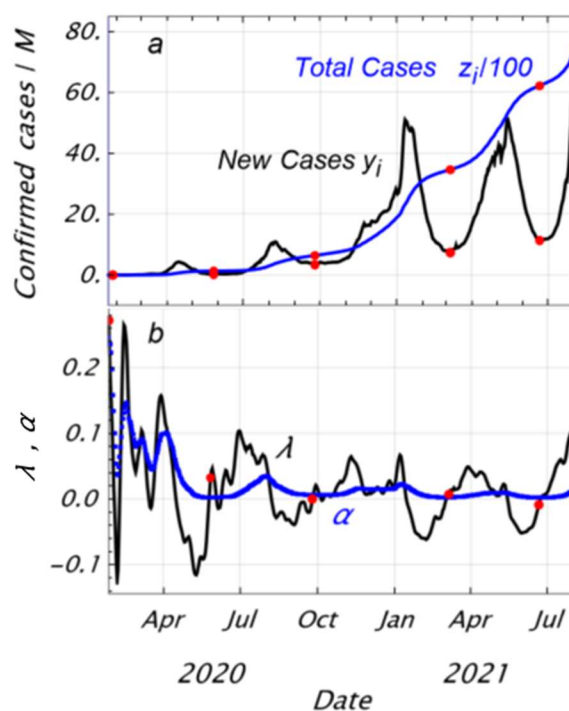


Fig. 1 Variation of new cases, total cases plotted in scale 1/100 (a) and corresponding exponents of λ and α (b) of COVID-19 infection in Japan, data from “Our World in Data”. Red dots showing converging points for each wave.

calculation, CR_1 value was varied in 0.1 step within the range of 1 to 2.3 and the calculation was executed to obtain X_0 for each CR_1 value. In the 1st wave, all infected persons are assumed to be infectivity, then IR_1 was set to be 1.

- After the 1st wave

The same calculating procedure is adopted for subsequent waves setting the last X_i value of previous wave as the initial value for the next wave. For each CR_1 value of the first wave, the infection ratio IR_n ($n = 2 \sim 5$) is varied in a range of 0.1~1 by 0.01 step in the new wave. Among those calculations, the value IR_n minimizing $Z_i - z_i$ (> 0) at the converging point of each wave is chosen as the best fit and CR_n is calculated as $CR_n = Z_i/z_i$. This process is repeated for the subsequent waves. There is some constraint for the allowable range of the initial value of infecteds X_0 in the first wave due to the linearity between X_0 and Z_i as shown in Fig. 2. A higher value of X_0 gives higher X_i , which is acceptable. On the other hand, lower X_0 gives lower X_i , then the term of the new cases y_i in Eq. (5) prevails compared to X_{i-1} and S_i , resulting in a large and rapid decrease in X_i into negative value.

Based on these considerations, the calculation up to the 5th wave was carried out with the condition that the cumulative infecteds Z_i curve up to the final wave best fits the total cases z_i curve. This gives the lower limit of Z_i . As a condition for the best fit, an index that minimizes the difference between Z_i and z_i over the entire period expressed by Eq. (11) is introduced.

$$dv = \sqrt{\sum_{j=1}^i (Z_j - z_j)^2} / z_i \quad (11)$$

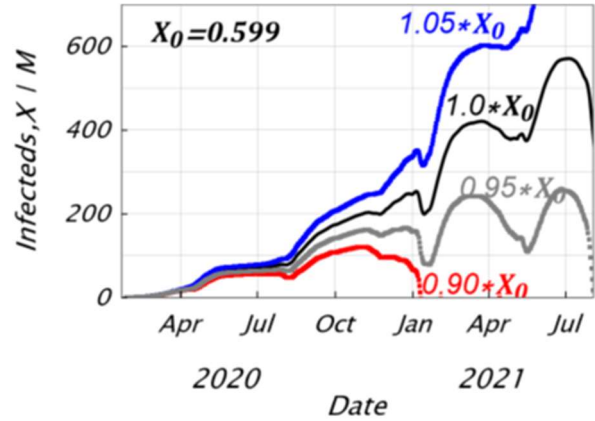


Fig. 2 Effects of initial value of infected persons at outbreak on variation in number of infected persons (infecteds). Lower initial value leads to early termination.

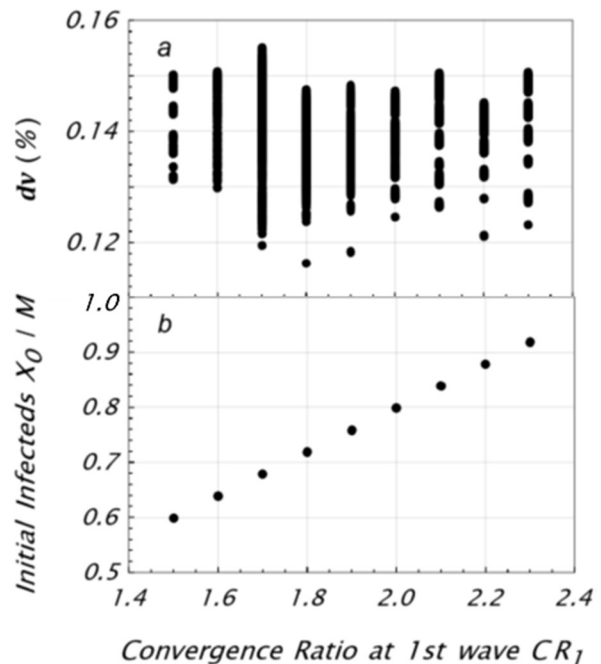


Fig. 3 Relationship between 1st wave convergence ratio, CR_1 and best fit index, dv over the entire period (a), and initial value of infecteds, X_0 (b). Initial value varies linearly with convergence ratio.

The relationship between dv (%) in Eq. (11) and CM_1 value in the first wave is shown in Fig. 3a. dv values obtained are all little over 0.1%, showing very good fittability. As shown in Fig. 3b, the initial value X_0 of infecteds increases linearly with the increase in the convergence ratio CR_1 of the first wave. In the range of $CR_1 < 1.5$, X_0 becomes lower and X_i decreases sharply as shown in Fig.2 before reaching the 5th wave. Thus X_0 is determined as a condition necessary for the continuation of infection up to the fifth wave. In addition, for the purpose of searching for the lower limit of the cumulative infecteds and reducing the number of calculations, the analyses were limited in a range of dv lower than 0.15.

The variation of infecteds X_i and cumulative infecteds Z_i are shown in Fig. 4a, b for each CR_1 value of 1.5, 1.9, and 2.3 with minimum dv . In the case of a lower convergence ratio CR_1 of the first wave, infecteds X_i stay lower in the early wave and higher in the latter wave compared to the higher CR_1 case. The cumulative infecteds in Fig. 4b exhibits the best fittability with the small deviation from the total cases in the latter wave. Table 1 shows the CR_n and IR_n values for each wave including conditions other than those in Fig.4. An increase in CM_1 results in a decrease in IR_2 , and subsequent convergence ratios remains close to 1.

For the case of CR_1 value 1.5, the quarantine rate QR_i and the social distancing ratio SD_i are calculated using the equations (8) and (9) and shown in Fig. 5a, b. By definition, peaks of quarantine rate correspond to the peak of new cases in each wave and reach up to about 0.25 in the third wave, staying low in other waves. Higher CR_1 makes QR_i increase in the later wave due to decrease in X_i in later wave.

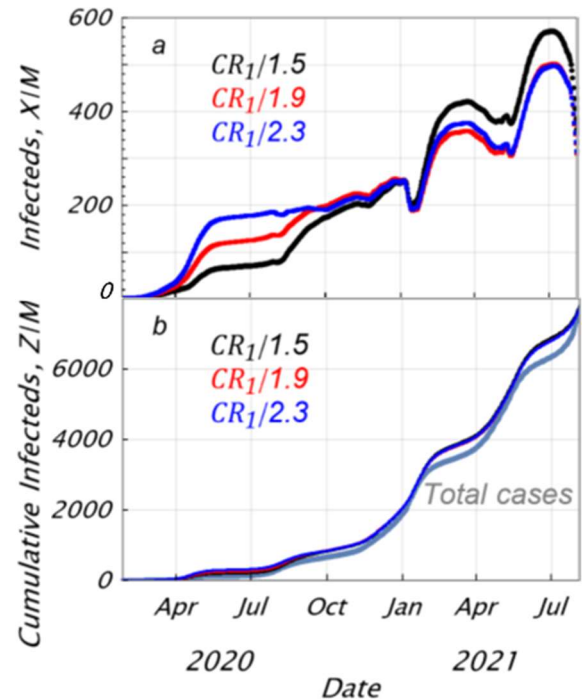


Fig. 4 Variation of infecteds X_i (a) and cumulative infecteds Z_i (b) for each CR_1 value of 1.5, 1.9, and 2.3. Cumulative infecteds best fit to total cases.

Table 1 CR_n value, IR_n value in each wave

	W1	W2	W3	W4	W5					
X_0	CR_1	IR_1	CR_2	IR_2	CR_3	IR_3	CR_4	IR_4	CR_5	IR_5
0.599	1.5	1.0	1.25	1.00	1.11	0.99	1.09	0.99	1.01	0.96
0.639	1.6	1.0	1.28	0.97	1.11	0.97	1.08	0.99	1.01	0.99
0.679	1.7	1.0	1.26	0.91	1.11	0.98	1.07	0.98	1.00	1.00
0.719	1.8	1.0	1.29	0.88	1.10	0.96	1.07	1.00	1.01	1.00
0.758	1.9	1.0	1.29	0.84	1.10	0.96	1.07	1.00	1.00	0.99
0.798	2.0	1.0	1.26	0.78	1.11	0.99	1.07	0.97	1.01	1.00
0.838	2.1	1.0	1.30	0.77	1.10	0.96	1.08	0.99	1.00	0.97
0.878	2.2	1.0	1.29	0.73	1.10	0.96	1.07	1.00	1.00	0.99
0.918	2.3	1.0	1.28	0.69	1.10	0.97	1.07	0.99	1.01	1.00

The social distancing ratio exhibits the opposite tendency to the quarantine rate and tends to be maximum near the minimum quarantine rate point. This is because the sum of the quarantine rate QR_i and the social distancing ratio SD equals $\Delta\lambda_i$ in Eq. (9), and the variation in $\Delta\lambda_i$ is small in a short period. In this figure, the dates of declaration and cancellation of the government's state of emergency is indicated by red and blue dots, respectively. In each wave, the social distancing ratio increases sharply on the declaration, peaks with a time lag of about one month, and then the declaration is lifted after the small decrease in social distancing ratio. The social distancing ratio at the first declaration reached nearly 1, and this high level lasted for a long period, indicating a greater declarative effect, however, the peaks after the second wave become a little bit lower and shorter than the first wave, showing weakening declarative effect.

The loci of the relationship between the quarantine rate and the social distancing ratio are shown in Fig. 6 separately for each wave. In each wave, the social distancing ratio decreases as the quarantine rate increases during infection spread, and vice versa at the stage of infection convergence. These loci show a tendency similar to the loci analyzed in the previous paper.

The variation of positive rate by PCR test as the measures of quarantine is shown in Fig. 7a together with the variation of quarantine rate. The positive rate varies at almost the same timing and same level as the quarantine rate. Both numerators have the same number of new cases, which means that both denominators, the number of tests and the number of infecteds X_i might be nearly equal. As shown in Fig. 7b, the number of PCR tests and the number of infecteds X_i vary at almost the same level. This means that the number of tests is considerably insufficient for detecting and quarantining newly infected persons. Fig. 8 shows the effects of

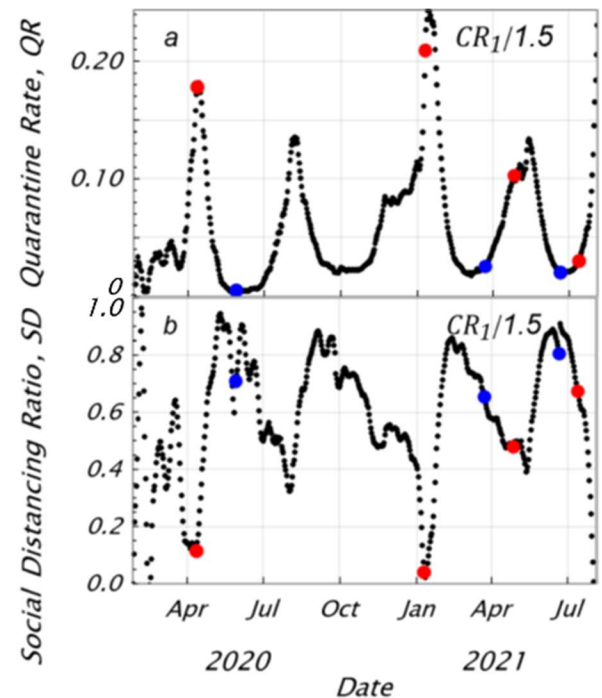


Fig. 5 Transition of quarantine rate QR (a) and social distancing ratio SD (b), red dots: declaration of emergency, blue dots: cancellation of emergency

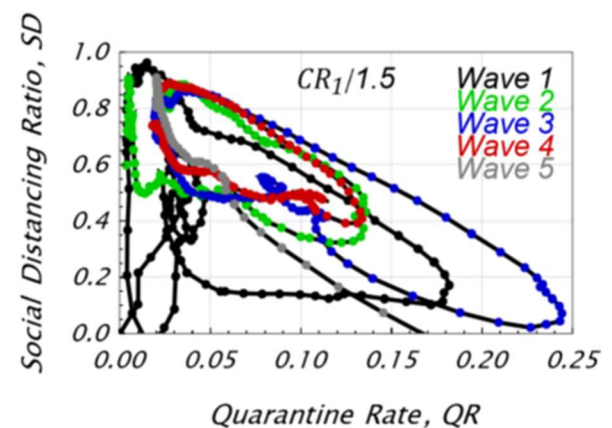


Fig. 6 Relationship between quarantine rate and social distancing ratio for each wave

varying the quarantine rate in Fig. 5a between 0.95 and 1.1 times on the transition of the infecteds by simulation. An increase in the quarantine rate decreases the infecteds, and 10% increase in the quarantine rate may terminate the infection after the third wave. Increasing the quarantine rate also reduces the exponent α together with CR_1 , which may further accelerate the termination of infection.

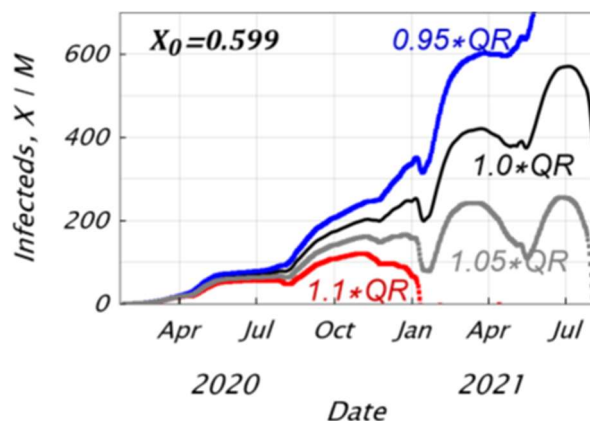


Fig. 8 Effect of quarantine rate, QR on variation of infecteds; increase in QR accelerates convergence of infection.

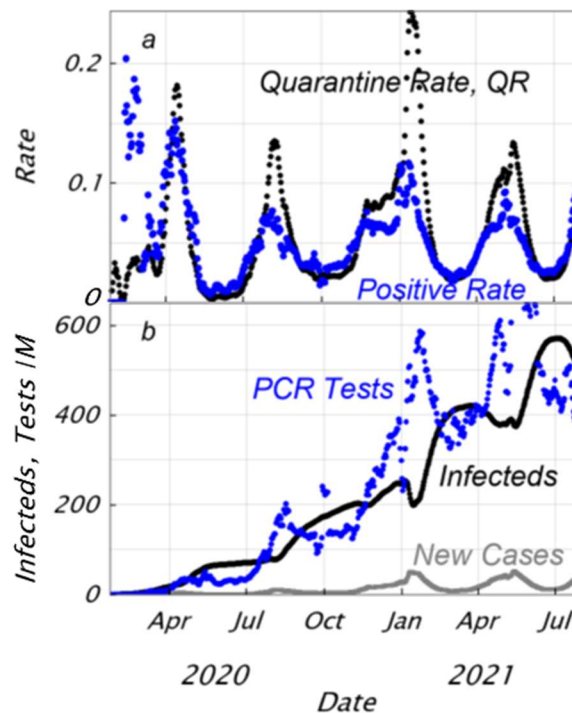


Fig. 7 Variation of positive rate in PCR tests and quarantine rate (a), and comparison of PCR tests and infecteds(b); QR and positive rate are varying in the same level.

The calculation for the best fit of the cumulative infecteds were executed for a lot of combinations of parameters, X_0 , CR_1 for the 1st wave, and IR_n for each wave. Apart from this method, there is also a deductive way, using the linearity between X_0 and X_i or Z_i as shown in Fig. 3b. Relational expressions between the initial value X_0 and Z_i are formulated for each IR_n for each wave, and the optimum solution is obtained while matching the waves using these relational expressions. However, it is necessary to pay attention to the case where the incubation period expressed by Eq. (3) spans two waves.

Conclusions

A model for estimating the number of COVID-19 infected persons (infecteds) was proposed based on the exponential function law of the SIR model. This model is composed of several equations expressing infecteds in the infection process, considering the onset after an incubation period, infectivity, wavy infection persistence with repeated infection spread and convergence while insufficient quarantine. Furthermore, the model was applied to each wave, where parameters were

introduced to guarantee the continuity and compatibility between neighboring waves. This model was applied to the infection in Japan, which is currently up to the 5th wave, and the initial infecteds and various parameters were calculated by curve fitting of the cumulative infecteds to the total cases in the database.

- As a best-fit solution of the cumulative infecteds up to the 5th wave to the total cases, the initial infecteds of 0.6 / million population at the start of infection is more than an order of magnitude higher than the actual initial cases of 0.03 / million population. A convergence ratio 1.5 (cumulative infecteds / total cases) at the end of the first wave was obtained.

- This result gives a minimum value of infecteds for the infection continuation up to the 5th wave. Otherwise, the initial infecteds less than this lower limit may result in the termination of infection in the second or the third wave.

- Although a convergence ratio of more than 1.5 in the first wave is allowed, the number of excess infecteds, in that case, is compensated by the decrease in the infection ratio (infectivity ratio of infecteds after the incubation period) in the second wave.

- The isolation rate and social distancing ratio based on the SIQR model proposed by Odagaki were evaluated, and the social distancing ratio increases sharply just after the government's declaration of emergency. The quarantine rate closely equals the positive rate by PCR tests, meaning that the infecteds, which had been unknown, is the same level as the number of tests.

Acknowledgment

The author would like to thank Professor Emeritus Ikuo Yoshihara of the University of Miyazaki for his detailed suggestions and discussions.

References

- 1) W. O. Kermack and A. G. McKendrick, Proc. Ry. Soc. A115(1927) 700
- 2) J. Makino, Kagaku (Iwanami Shoten) 90(2020) 428, in Japanese publication
- 3) H. Hethcote, M. Zhién and L. Shengbing, Math. Biosciences 180, 141-160 (2002)
- 4) T. Odagaki, *Infect. Dis. Model.* **5**, 691-698 (2020). <https://doi.org/10.1016/j.idm.2020.08.013>
- 5) K. Hashiguchi, <https://medrxiv.org/cgi/content/short/2020.08.04.20167882v1>
- 6) K. Hashiguchi, <https://medrxiv.org/cgi/content/short/2020.09.01.20185611v1>
- 7) H. Nishiura, H. Inaba, Suuri-Toukei54,2,461-480(2006), in Japanese publication
- 8) Our World in Data COVID-19 dataset, <https://ourworldindata.org/coronavirus-source-data>, (Accessed: 2021.8.5)