Investigating factors behind the outbreak of the 6th and the 7th waves of COVID-19 in Tokyo

Yoshihiko Takase *
November 23, 2022

Abstract

The 6th wave of COVID-19 in Tokyo continued for the longest period of infection (about 190 days from late Nov. 2021), and the 7th wave, which occurred in mid-May 2022, was the largest wave ever (approximately 1.7 million people in total)[1].

We have regarded the spread of COVID-19 infection as a phase transition and been analyzing it using the Avrami equation. The advantage is that a straightforward model of nucleation and growth can be assumed, the model can be described by a single equation[3], and then it is relatively easy to apply the least-squares method to determine the optimum values of the parameters[4, 5].

Based on the random nucleation and subsequent linear domain-growth model, the 1st to the 5th waves of COVID-19 in Japan were analyzed and the parameters were determined[6][7][8]. The accuracy of the simulated cumulative infection $D$ was as good as $95\% CI/D < 2.5\%$ after the 2nd wave when $D$ exceeded 10,000[7].

Since this method is highly accurate, the factors behind the infection may be elucidated by simulating the infection waveform in detail. This time, we will focus on the infection wave in Tokyo, the capital of Japan, where various data have been published.

The purpose of this study is to simulate the detailed structure of the infection waveform in Tokyo by using the Avrami equation and to elucidate the factors behind it.

1 Introduction

The 6th wave of COVID-19 in Tokyo continued for the longest period of infection (about 190 days from late Nov. 2021), and the 7th wave, which occurred in mid-May 2022, was the largest wave ever (approximately 1.7 million people in total)[1].

The outbreak of the main component of the 7th wave occurred by the coupling of the invasion of the new virus variant Omicron BA.5 and multiple social factors, followed by human interaction during the July holidays. Based on the results that the domain growth rate $K$ and the infection rise time $t_{on}$ were almost independent of the initial susceptible $D_s$, dense nucleation followed by a near growth model was deduced. The quantity $K \cdot t_{on}$ was considered to represent the infectivity of the virus.

[arXiv:2211.12083v1 [q-bio.PE] 22 Nov 2022]
2 Waveform analysis of the COVID-19 infection in Tokyo

2.1 Simulation characteristics of the 5th to the 7th waves

The 5th, 6th, and 7th waves in Tokyo were simulated by superposition of 5 waves (5.1th to 5.5th waves), 6 waves (6.1th to 6.6th waves), and 5 waves (7.1th to 7.5th waves), respectively.

Since the magnitude of each wave differs greatly, the number of time-dependent daily new infections \( J(t) \) (hereinafter referred to as the daily infections) and the number of time-dependent cumulative daily new infections \( D(t) \) (hereinafter referred to as the total infections) for the 5th to the 7th waves (June 1, 2021 to November 1, 2022) are shown in semi-logarithmic representation in Fig. 1, where \( t \) is the time. The dark green and the red markers are the detected daily and total infections, respectively. The blue curve is the theoretical value of the daily infections for each wave and the dark blue curve is the sum of them, \( J(t) \). The brown curve is the theoretical value of the total infections for each wave and the dark brown curve is the sum of them, \( D(t) \). The accuracy of the simulated total infections \( D(t) \) was as good as 95% CI/\( D(t) < 0.8% \) after the 5th wave.

Start day and end day of the duration of each wave to which the least-squares method is applied to find the parameters for the nucleation and subsequent two-dimensional domain growth model is shown in Table 1. These durations were determined by iteratively applying the least-squares method to improve simulation accuracy [7].

2.2 Determined parameters for each wave

The parameters determined for each wave, \( D_s, K^2, K, \) and \( t_{on} \) are shown in Table 2 where \( D_s \) is the initial susceptible, \( K \) the domain growth rate, and \( t_{on} \) the rise time defined as the time for the \( D(t) - t \) characteristic to change from its 10% value to its 90% value, originally an electronics definition. The value of the nucleation decay constant \( \nu \) was assumed

![Figure 1: The semi-logarithmic representation of the detected and the simulated values of the daily infections \( J(t) \) and the total infections \( D(t) \) for the 5th, the 6th, and the 7th waves in Tokyo.](image)

| Wave  | Start day | End day  | Duration [day] |
|-------|-----------|----------|----------------|
| 5.1th | 02-Jun-2021 | 04-Jul-2021 | 182            |
| 5.2th | 03-Jul-2021 | 22-Jul-2021 | 182            |
| 5.3th | 15-Jul-2021 | 26-Sep-2021 | 182            |
| 5.4th | 23-Aug-2021 | 24-Oct-2021 | 182            |
| 5.5th | 10-Oct-2021 | 01-Dec-2021 | 182            |
| 6.1th | 24-Nov-2021 | 26-Dec-2021 | 182            |
| 6.2th | 23-Dec-2021 | 03-Jan-2022 | 182            |
| 6.3th | 02-Jan-2022 | 21-Feb-2022 | 182            |
| 6.4th | 08-Feb-2022 | 22-Mar-2022 | 182            |
| 6.5th | 12-Mar-2022 | 24-Apr-2022 | 182            |
| 6.6th | 15-Apr-2022 | 29-May-2022 | 182            |
| 7.1th | 21-May-2022 | 20-Jun-2022 | 182            |
| 7.2th | 22-Jun-2022 | 18-Jul-2022 | 182            |
| 7.3th | 08-Jul-2022 | 11-Aug-2022 | 182            |
| 7.4th | 30-Jul-2022 | 11-Sep-2022 | 182            |
| 7.5th | 27-Aug-2022 | 10-Oct-2022 | 182            |
to be constant ($\nu = 0.0090 \ [1/day]$) for the reason described in the previous report[8].

3 Waveform analysis of the COVID-19 6th infection in Tokyo

3.1 Simulation characteristics of the 6th wave

Figure 2 shows the time-dependent characteristics of the detected and the simulated values of the daily infections $J(t)$ and the total infections $D(t)$ for the 6th wave in Tokyo from January 1, 2022 to June 1, 2022. The meaning of each marker and line is the same as that in Fig. 1. The broken line is the 95%CI calculated from the moving standard deviation with 7-days sliding window.

Figure 2: The detected and the simulated values of the daily infections $J(t)$ and the total infections $D(t)$ of the 6th wave in Tokyo from January 1, 2022 to June 1, 2022. Symbols: refer to Fig. 1. The broken line is the 95%CI.

3.2 The correlation between the 6th wave and the BA.1 variant ratio

Shortly before the start of the 6th wave, Delta variant of the virus was replaced by the new virus variant, Omicron BA.1[9]. Figure 3 shows the characteristics of the 6th wave with the percentage $p_{BA1}(t)$ of the suspected Omicron BA.1 variant from the (84th) Tokyo Metropolitan Government Novel Coronavirus Infectious Diseases Monitoring Meeting materials[10] inserted in the upper part. The value of $p_{BA1}(t)$ is also well approximated by the Avrami equation. The 6.1th and the 6.2th waves are too small to be seen in Fig. 3, but can be seen in Fig. 1.

Figure 4 shows the relationship between $D_{6.1}, D_{6.2}, D_{6.1} + D_{6.2}, D_{6.3}$ and $p_{BA1}$ to quantitatively understand the virus mutation effect. The relationship between $D_{6.1}$ and $p_{BA1}$ (brown marker and solid line) is almost linear. The relationship between $D_{6.2}$ and $p_{BA1}$ (orange marker and solid line) and that between $D_{6.1} + D_{6.2}$ and $p_{BA1}$ (red marker and solid line) are almost linear in the region from 0 to 87.1% of $p_{BA1}$.

Table 2 shows the respective sample correlation coefficient $\gamma$. The value of $\gamma$ for $D_{6.1} + D_{6.2}$ is 0.990,
Figure 3: The characteristics of the 6th wave with the data of \( p_{BA1}(t) \). Symbols: refer to Fig. 2.

Table 3: The sample correlation coefficient \( \gamma \) of the relationship between each of \( D_{6.1}, D_{6.2}, D_{6.1} + D_{6.2}, D_{6.3} \) and \( p_{BA1} \).

| Wave | \( D_{6.1} \) | \( D_{6.2} \) | \( D_{6.1} + D_{6.2} \) | \( D_{6.3} \) |
|------|--------------|--------------|----------------|--------------|
| \( \gamma \) | 0.923        | 0.977        | 0.990          | 0.889        |

indicating the strongest correlation. In the region where \( p_{BA1} \) exceeds 87.1%, \( D_{6.2} \) and \( D_{6.3} \) deviate from the straight line and increase rapidly.

### 3.3 The correlation between the BA.1 variant ratio and the infections among people in their 20s

A characteristic increase in infections among people in their 20s can be seen in the above Monitoring Meeting materials[10]. Table 4 shows the percentage \( p_{20s} \) of the total infections \( D_{20s} \) among people in their 20s to \( D_{all} \) of all people on \( t_1 =17\text{-Dec-2021}, \ t_2 =24\text{-Dec-2021}, \ t_3 =31\text{-Dec-2021}, \) and \( t_4 =07\text{-Jan-2022} \). Days are expressed as medians.

Since this increase appears to correspond to an increase in \( p_{BA1}(t) \), Fig. 5 shows the relationship between \( p_{20s} \) and \( p_{BA1} \). The data from 17-Dec-2021 to 07-Jan-2022 are almost linear, with a high sample

Table 4: The percentage \( p_{20s} \) of the total infections \( D_{20s} \) among people in their 20s to \( D_{all} \) of all people on \( t_1 =17\text{-Dec-2021}, \ t_2 =24\text{-Dec-2021}, \ t_3 =31\text{-Dec-2021}, \) and \( t_4 =07\text{-Jan-2022} \).

| Day | \( t_1 \) | \( t_2 \) | \( t_3 \) | \( t_4 \) |
|-----|--------|--------|--------|--------|
| \( p_{20s} \) | 18.8 | 25.9 | 28.0 | 40.5 |


Figure 5: The relationship between the 20s age infection ratio $p_{20s}$ and $p_{BA1}$.

correlation coefficient of 0.950.

### 3.4 Factors behind outbreak of main component of the 6th wave

During the 6th wave, two events occurred in which the main epidemic virus was replaced by a more infectious variant.

As shown in Fig. 3, the infection cases with the Omicron variant were confirmed in Tokyo in mid-December 2021, when the 5th wave was converging. Since the Omicron variant was more infectious than the Delta variant, infections began to increase, and its ratio increased in proportion to the total number of new infections, forming relatively small waves 6.1 and 6.2 (see Fig. 1). This is the beginning of the first replacement.

There were New Year holidays (31-Dec-2021 to 03-Jan-2022) during the 6.2th wave. In New Year holidays, as is the custom of the Japanese people, there is a great increase in the exchange of people. Table 4 shows that the number of new infections among people in their 20s increased significantly around the New Year holidays, which is thought to be the result of active interaction among people in their 20s.

As the number of infections increased significantly, it was newly regarded as the outbreak of the 6.3th wave (see Fig. 2). When the 6.3th wave occurred (02-Jan-2022), the value of $p_{BA1}(t)$ was 9.1%, but by 20-Jan-2022, when the number of daily infections reached 50% of its maximum, it was 97.1%, reaching 99.6% on 05-Feb-2022, about a month later.

From this process, it was thought that the coupling of increased human interaction due to the New Year holidays and the invasion of the new virus variant Omicron BA.1 caused the nucleation and domain growth of the Omicron variant in places and among people where nucleation was difficult to occur with the Delta variant, and then the outbreak of the dominant 6.3th wave occurred.

### 3.5 Factors behind the outbreak of the three side-waves of the 6th wave

After the dominant 6.3th wave, the second replacement of virus variant occurred.

Figure 6 shows the characteristics of the 6th wave with the percentage $p_{BA2}(t)$ of the suspected Omicron BA.2 variant from the (96th) Tokyo Metropolitan Government Novel Coronavirus Infectious Diseases Monitoring Meeting materials inserted in the upper part.

In order to know the quantitative correlation, the relationship between $D_{6.3}, D_{6.4}, D_{6.5}$ and $p_{BA2}$ is shown in Fig. 7 and sample correlation coefficient are shown in Table 5. The relationship between $D_{6.4}$ and $p_{BA2}$ was almost linearly approximated, and the sample correlation coefficient was 0.986, indicating the strongest correlation.

As shown in Fig. 2, the 6.4th wave occurred around the February holidays (11 to 13-Feb-2022). The value of $p_{BA2}(t)$ increased in proportion to the total infections of the 6.4th wave, and exceeded 95% around 20-

Table 5: The sample correlation coefficient $\gamma$ of the relationship between each of $D_{6.3}, D_{6.4}, D_{6.5}$ and $p_{BA2}$.

| Wave   | $D_{6.3}$ | $D_{6.4}$ | $D_{6.5}$ |
|--------|----------|----------|----------|
| $\gamma$ | 0.717    | 0.986    | 0.853    |
Figure 6: The characteristics of the 6th wave with the data of \( p_{BA2}(t) \). Symbols: refer to Fig. 2.

Figure 7: The relationship between \( D_{6.3}, D_{6.4}, D_{6.5} \) and \( p_{BA2} \).

Figure 8: The detected and the simulated values of \( J(t) \) and \( D(t) \) of the 7th wave in Tokyo from June 1, 2022 to November 1, 2022. Symbols: refer to Fig. 2.

4 Waveform analysis of the COVID-19 7th infection in Tokyo

4.1 Simulation characteristics of the 7th wave

Figure 8 shows the detected and the simulated values of \( J(t) \) and \( D(t) \) for the 7th wave in Tokyo from June 1, 2022 to November 1, 2022. The meaning of the markers and each line is the same as that in Fig. 2.
4.2 The correlation between the 7th wave and the BA.5 variant ratio

Shortly before the start of the 7th wave, the BA.2 variant was replaced by the more infectious BA.5 variant [15].

Figure 9 shows the characteristics of the 7th wave with the percentage of the suspected Omicron BA.5 variant from the (96th) Monitoring Meeting materials [11] inserted in the upper part. The 7.1st wave occurred around 20-May-2022. The value of $p_{\text{BA5}}(t)$ increased almost in proportion to the total infections during the 7.1st wave, reaching 92.8% around 22-Jul-2022, about two months later. There are no consecutive holidays in June, so there is no consecutive holiday effect.

Figure 10 shows the relationship between $D_{6.6}, D_{7.1}, D_{7.2}, D_{7.1} + D_{7.2}$ and $p_{\text{BA5}}$. The correlation between $D_{7.1}$ and $p_{\text{BA5}}$, and between $D_{7.1} + D_{7.2}$ and $p_{\text{BA5}}$ (brown and red markers and solid lines, respectively) are almost linear. The relationship between $D_{7.2}$ and $p_{\text{BA5}}$ (dark green marker and solid line) is almost linear in the region from 0 to about 60% of $p_{\text{BA5}}$.

Table 6 shows the respective sample correlation coefficient $\gamma$. The value of $\gamma$ for $D_{7.1}$ is 0.977, indicating the strongest correlation. In the region where $p_{\text{BA5}}$ exceeds 60%, $D_{7.2}$ deviates from the straight line and increases rapidly.

4.3 The correlation between the 7th wave and the daily maximum temperature

The daily maximum temperature $T_{\text{max}}(t)$ was approximately 20°C in early June, but exceeded 35°C in late June and reached 37.0°C on July 1st. When the temperature rises, air conditioners will be used indoors, and there will be an increase in the tendency for families to gather in one room without wearing masks, especially at home.

Figure 11 shows the characteristics of the 7th wave with $T_{\text{max}}(t)$ inserted above. Comparing Fig. 9 and Fig. 11, it is expected that the characteristics of $p_{\text{BA5}}(t)$ and $T_{\text{max}}(t)$ are highly correlated.

The relationship between $p_{\text{BA5}}$ and $(T_{\text{max}} - 18.0)$
is shown in Fig. 12. Although the temperature fluctuates, the value of $\gamma$ is 0.922, showing a strong correlation.

### 4.4 High infection rate among young people

Table 7 shows the percentage of new infections by age group. The combined infection rate for under-10s and teenagers is about 22%, higher than the highest other age group, about 20%, for those in their 20s.

| Age | <10s | 10s | 20s | 30s | 40s | 50s |
|-----|------|-----|-----|-----|-----|-----|
| (a) | 10.3 | 13.1| 20.1| 16.9| 15.5| 11.8|
| (b) | 9.8  | 11.0| 19.3| 17.0| 17.2| 12.8|

Table 7: The percentage of new infections by age group on the period of (a) 19-Jul-2022 to 25-Jul-2022 and (b) 26-Jul-2022 to 01-Aug-2022.

4.5 High rate of infection from cohabitants

About 75% of newly infected people have unknown contact history. Table 8 shows the percentage of newly infected people by infection route whose infection route was clear. Infection from people who live together is the most common, about 68%.

| Route | (1) | (2) | (3) | (4) | (5) | (6) |
|-------|-----|-----|-----|-----|-----|-----|
| (a)   | 68.7| 15.2| 6.6 | 2.2 | 0.0 | 7.4 |
| (b)   | 67.9| 14.3| 7.4 | 1.7 | 0.0 | 8.7 |

Table 8: The percentage of newly infected people by infection route, (1) living together, (2) facilities, (3) workplace, (4) dining together, (5) dining together with entertainment, and (6) others on the period of (a) 19-Jul-2022 to 25-Jul-2022 and (b) 26-Jul-2022 to 01-Aug-2022.

4.6 Low vaccination coverage among younger age groups

Table 9 shows the percentage $p_v$ of the third vaccination by age group in Tokyo (updated on August 8, 2022). In the 12-19 age group, it is the lowest at 37.4%. Vaccination for the 5-11 age group is said to be carried out, but no data is shown.
Table 9: The percentage $p_v$ of the vaccination by age group.

| Age  | 12-19 | 20s  | 30s  | 40s  | 50s  | 60s  |
|------|-------|------|------|------|------|------|
| $p_v$| 37.4  | 47.5 | 54.7 | 62.6 | 79.0 | 86.8 |

4.7 Complex factors behind the outbreak of the 7th wave

The coupling of multiple factors indicated above and the invasion of the new virus variant BA.5 was thought to lead to the 7th wave outbreak in the following procedure.

First, the nucleation of the BA.5 variant occurred predominantly over the BA.2 variant in young people. Infection spread among people who were in daily contact with those nuclei. This infection formed a large initial 7.2th wave. Furthermore, the nucleation of the BA.5 variant coupled with the consecutive holidays in July (16 to 18-Jul-2022, refer to Fig. 8) spread the infection to middle-aged and elderly people in addition to young people, causing the equally large 7.3th wave. The superposition of these two waves formed the main component of the 7th wave, the largest ever (see Fig. 8).

5 Discussions

5.1 Dense nucleation followed by a near growth model

The present analysis is based on the random nucleation and subsequent two dimensional linear domain growth model\[7, 8\]. The values of $K$ in Table 2 are almost constant, about 0.10 for the 5.1th to the 6.3th waves (A group), and about 0.13 for the 6.4th to the 7.5th waves (B group).

The relationship between $K$ and $D_s$ is shown in Fig. 13 for A and B groups. The value of $K$ tends to increase slightly with increasing $D_s$ for A group. The value of $K$ is less than 1.3 times greater even when the value of $D_s$ increases by a factor of about 900 (refer to Table 2). For B group, the value of $K$ is almost constant without depending on $D_s$ (refer to Table 2).

Figure 14 shows the relationship between the turn-on time $t_{on}$ and $D_s$ from the 6.2th and the 7.5th waves. Even if the value of $D_s$ increases by about 40 times, the value of $t_{on}$ changes by 1.1 times at most and is about 31 days (refer to Table 2).

Based on these characteristics, a model in which a few nuclei are generated and grow over a wide area is denied, and a model in which a lot of nuclei are generated randomly and subsequent domain growth reaches the growth limit in a relatively short period of time is deduced.

Large waves are mainly caused by nucleation in many places, resulting in large $D_s$. Although there is a tendency for $D_s$ to increase as the growth rate increases, its contribution is small.

Data on vaccination coverage and number of new PCR positives during the 7th wave are also consistent with this model. Table 10 shows A: population by age group\[14\], B: the 3rd vaccination percentage\[13\], $C = A \times (1-B)$: number of unvaccinated people, $E$: number of new PCR positives during the 7th wave period (29-Jun-2022 to 02-Aug-2022)\[15\], and $F = E/C [\%]$: the percentage of the number of new PCR positives to the number of unvaccinated people.

The number of people who have not been vaccinated for the third time, $C$, is the number of peo-
Table 10: The third vaccination percentage and number of new PCR positives by age group.

| Age  | 12-19  | 20s   | 30s   |
|------|--------|-------|-------|
| A    | 9,010,292 | 12,819,569 | 14,372,705 |
| B    | 35.6   | 48.6  | 52.3  |
| C    | 5,802,628 | 6,589,258 | 6,855,780 |
| E    | 537,335 | 562,450 | 549,759 |
| F    | 9.3    | 8.5   | 8.0   |

These data favor the dense nucleation followed by a near growth model, where a lot of nuclei are generated randomly at places where people gather, such as homes, facilities, working areas, and various gathering places. However, the domain growth is limited within the proximity of the nucleation.

5.2 A quantity to represent the infectivity of the virus

Figure 14 shows the relationship between the quantity $K \cdot t_{on}$ and $D_s$ from the 6.2th and the 7.5th waves.

Figure 15 shows the relationship between $K \cdot t_{on}$ and $K$ for all waves from the 1.1th to the 7.5th waves.
each group of this quantity is about 3.2 for the 1.1th to the 4.2th waves, about 3.6 for the 5.1th to the 6.3 waves, and about 4.0 for the 6.4th to the 7.5th waves. Since these values are almost independent of the magnitude of the infection wave, they are thought to represent the infectivity of the virus and are likely quantities related to the basic reproduction number.

6 Conclusion

The 6th and the 7th waves of COVID-19 in Tokyo were analyzed and well simulated by using the Avrami equation. As a result, the factors that caused the 6th wave to last the longest and the 7th wave to be the largest were elucidated.

The main component of the 6th wave was formed by the coupling of increased human interaction due to the New Year holidays and the invasion of the new virus variant Omicron BA.1. After that, side waves were formed by the coupling of the invasion of the new virus variant Omicron BA.2 and the human interaction in the consecutive holidays in February, March, and May. These side waves caused the 6th wave not to converge for a long time.

The outbreak of the main component of the 7th wave occurred by the coupling of the invasion of the new virus variant Omicron BA.5 and multiple social factors, followed by human interaction during the July holidays.

Based on the results that the domain growth rate $K$ and the infection rise time $t_{on}$ were almost independent of the initial susceptible $D_s$, dense nucleation followed by a near growth model was deduced.

The quantity $K \cdot t_{on}$ was considered to represent the infectivity of the virus.

References

[1] https://stopcovid19.metro.tokyo.lg.jp/
[2] M. Avrami, J. Chem. Phys. 8, 212 (1940).
[3] Crystallization of Polymers, Leo Mandelkern (1966).
[4] Y. Takase, A. Odajima and T. T. Wang, J. Appl. Phys. 60, 2920 (1986).
[5] Y. Takase, J. W. Lee, J. I. Scheinbeim and B. A. Newman, Macromolecules 24, 6644-6652 (1991).
[6] Y. Takase, arXiv:2008.01690
[7] Y. Takase, arXiv:2109.12472
[8] Y. Takase, arXiv:2110.13449
[9] https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/
[10] https://www.bousai.metro.tokyo.lg.jp/taisaku/saigai/1013388/1021271.html
[11] https://www.bousai.metro.tokyo.lg.jp/taisaku/saigai/1021348/1021950.html
[12] https://www.data.jma.go.jp/obd/stats/etrn/view/daily_s1.php?prec_no=44&block_no=47662&year=2022&month=08&day=&view=pl
[13] https://www.kantei.go.jp/jp/content/kenbetsu_nenrei_kakuryo_kyubetsu_data.pdf (The old Aug. 8, 2022 version. It’s not shown now.)
[14] https://www.kantei.go.jp/jp/content/kenbetsu_nenrei_kakuryo_kyubetsu-vaccination_data.pdf
[15] https://covid19.mhlw.go.jp/extensions/public/index.html