

プログラム RView 説明書 (第 2b 版)

S. Yamauchi

2020 年 9 月 11 日

目次

| | | |
|-----|---------------------------|---|
| 1 | プログラムの目的と機能 | 2 |
| 2 | 理論と計算方法 | 2 |
| 2.1 | 時系列データから増加率を求める | 2 |
| 2.2 | 増加率から再生産数を求める | 3 |
| 3 | プログラムの使用法 | 5 |
| 3.1 | メニュー | 5 |
| 3.2 | 入力データ | 5 |
| 3.3 | グラフィック画面表示 | 5 |
| 3.4 | パラメータ変更 | 6 |
| 3.5 | ファイル出力 | 7 |
| 4 | 適用例 | 8 |
| 5 | 使用にあたっての注意とヒント | 8 |

改訂記録

| | | |
|------|----------------|------------------------|
| 初版 | 2020 年 6 月 9 日 | |
| 2 版 | 7 月 30 日 | 発症間隔分布に 5 種類の分布形を採用した。 |
| 2b 版 | 9 月 11 日 | グラフ表示の時間軸表示法を整理した。 |
| | 12 月 20 日 | マニュアル誤記訂正。 |
| | 2021 年 1 月 1 日 | マニュアル誤記訂正。 |

1 プログラムの目的と機能

このプログラムは、任意の感染症について、ある集団の過去の日ごとの累積感染者数データから再生産数(実効再生産数)の時間推移を求めることを目的とする。

累積感染者データはファイルから入力するか、またはキーボードから直接入力する。求めた再生産数の推移はグラフィック画面に表示すると共に、指定するファイルへ数値として書き出すこともできるものとする。

ある集団での感染症の拡大には、一人の平均的感染者から発生する二次感染者数(再生産数)と、それらの二次感染が生じるまでの日数(発症間隔、世代間隔^{*1})が関与するので、観測データから求まる再生産数は、その感染症の発症間隔分布により相違する。

このプログラムでは発症間隔分布として、ワイブル分布、ガンマ分布、対数正規分布、指数分布(SIRモデル)、SEIRモデル相当分布を対象とする。

2 理論と計算方法

プログラム内の主な処理は、次の2点である。

- (1) 感染数の時間的推移の観測データから、その指数関数的増加率を求める。
- (2) 増加率から仮定した発症間隔分布に応じた再生産数を求める。

2.1 時系列データから増加率を求める

毎日($t = t_0, t_1, t_2, \dots$)の新規感染者数 $y = y_0, y_1, y_2, \dots$ が与えられたとする。第 i 日目前後の値を指数関数 $y = Ae^{rt}$ で、つまり $\ln y = \ln A + rt = A' + rt$ で近似する($A' = \ln A$ および r は、この間では一定値であると仮定)。その時の $\ln y$ 推測の誤差は $[\Delta(\ln y)]_i = (A' + rt_i) - \ln y_i$ であるが、元の y の値に換算した誤差は

$$\{\Delta y\}_i = \left\{ \frac{dy}{d(\ln y)} \right\}_i \{\Delta(\ln y)\}_i = y_i(A' + rt_i - \ln y_i)$$

となると評価できる。

第 i 日前後の $2m + 1$ 日間の観測値 $y_{i-m}, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_{i+m}$ に対して、その誤差の2乗和

$$E^2 = \sum_i \{\Delta y\}_i^2 = \sum_i y_i^2 (A' + rt_i - \ln y_i)^2 \quad (1)$$

を最小にするように指数 r および A' を定めて(最小2乗法)、その値を中央の時刻(日) t_i における値として採用する。ただし、 \sum_i の表記は、 $i = (i-m) \sim (i+m)$ の $2m + 1$ 個の点での値の総和を表すものとする。

$\frac{\partial}{\partial A'} E^2 = 0$ および $\frac{\partial}{\partial r} E^2 = 0$ より、

$$\sum_i y_i^2 (A' + rt_i - \ln y_i) = 0$$

^{*1} (serial interval) このような概念は、数理人口学(人口統計学)の分野に由来しているため、親子の年齢差つまり世代間隔と呼ばれてきた。感染症分野では、親元の一次感染者と二次感染者との感染時間の差(日数)の意味である。さらに、感染から症状が発現するまでには日数(潜伏期間)に差があるであろうから、発症間隔というのも正確な表現ではないと思われるが、疫学の分野では、「発症間隔」「世代間隔」で通されているようである。

$$\sum_i y_i^2 t_i (A' + r t_i - \ln y_i) = 0$$

未知の値 A' および r の係数を整理して、次の連立方程式が得られる。

$$a_{00} A' + a_{01} r = b_0 \quad (2)$$

$$a_{10} A' + a_{11} r = b_1 \quad (3)$$

ただし、

$$a_{00} = \sum_i y_i^2, \quad a_{01} = a_{10} = \sum_i y_i^2 t_i, \quad a_{11} = \sum_i y_i^2 t_i^2, \\ b_0 = \sum_i y_i^2 \ln y_i, \quad b_1 = \sum_i y_i^2 t_i \ln y_i$$

である。これより、第 i 日 ($t = t_i$) における r の値

$$r = \frac{b_1 a_{00} - b_0 a_{10}}{a_{00} a_{11} - a_{01} a_{10}} \quad (4)$$

が求まる (A' は不要)。

このような最小 2 乗近似に用いる区間を順次ずらすことにより、指数 r の時間変化を求めることができる。 m の値が小さい場合は日々の観測値の変動の影響が強く残るが、大きい場合は r の変化に対して鈍感になるであろう。両者を考慮して、妥当な m の値を採用するものとする。観測範囲の両端付近では範囲をはみ出ないように m の値を順次小さく変更し、また両端点では前後の日の値で代用することにする。

2.2 増加率から再生産数を求める

増加率 r と再生産数 R の関係はやや複雑であり、別紙資料（「発症間隔分布にもとづいた増加率・再生産数間の関係について」）にその詳細を示す。ここでは、その要点と結果のみを示す。

ある平均的な個人がある感染症に感染したとき、他の人に対するその感染力は日ごとに変化する。感染から t 日後のある 1 日で他の人を感染させる人数（二次感染の確率）を $n(t)$ とすると、その感染力は図 1 のような曲線を描くと予想される。この感染力を表す曲線 $n(t)$ は二次感染者を生じるまでの日数（発症間隔）の分布を表すと考えられるので、この全面積が 1 となるように正規化した関数

$$g(t) = \frac{n(t)}{\int_a^\infty n(a) da} \quad (5)$$

は発症間隔分布を表す一種の確率密度関数となる。この確率で感染した二次感染者たちが、その感染時以降で同様の発症間隔分布で第三次感染を引き起こし、その第三次感染者たちが同様に第四次感染を引き起こす、等々を繰り返して、結果的に指数関数的増加を引き起こす。

（疫学的調査または何らかの仮定を基にして、）この $g(t)$ が既知であれば、次式のように指数増加率 r と再生産数 R を関係づけることができる。

$$\frac{1}{R} = \int_0^\infty e^{-rt} g(t) dt \quad (6)$$

いくつかの代表的な発症間隔分布 $g(t)$ に対して、増加率 r と再生産数 R の間の関係等を下表に示す。

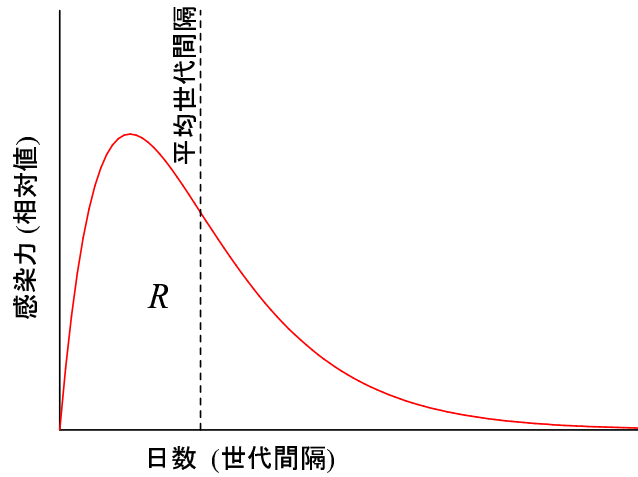


図1 二次感染曲線 $n(t)$ の模式図

| 分布の種類 | $g(t)$ | 平均値 μ | 分散 σ^2 | R | 備考 |
|-------------------|---|-------------|-----------------|--------------------|---|
| ワイブル分布 | | | | | 本表の欄外に示す |
| ガンマ分布 | $\frac{1}{\Gamma(m)\eta} \left(\frac{t}{\eta}\right)^{m-1} e^{-t/\eta}$ | $m\eta$ | $m\eta^2$ | $(1+r\eta)^m$ | $r > -1/\eta$ |
| 対数正規分布 | | | | | 本表の欄外に示す |
| 指数分布 (SIR モデル) | $\frac{1}{T_I} e^{-t/T_I}$ | T_I | T_I^2 | $1+rT_I$ | $r > -1/T_I$ |
| SEIR モデル 相当分布 | $\frac{e^{-t/T_I} - e^{-t/T_E}}{T_I - T_E}$ | $T_E + T_I$ | $T_E^2 + T_I^2$ | $(1+rT_E)(1+rT_I)$ | T_E : 平均潜伏期間 T_I : 平均感染性期間 $r > -1/T_E, r > -1/T_I$ |

ワイブル分布：

$$g(t) = \frac{m}{\eta} \left(\frac{t}{\eta}\right)^{m-1} e^{-(t/\eta)^m} \quad \mu = \eta \Gamma\left(1 + \frac{1}{m}\right)$$

$$\sigma^2 = \eta^2 \Gamma\left(1 + \frac{2}{m}\right) - \eta^2 \left\{ \Gamma\left(1 + \frac{1}{m}\right) \right\}^2 \quad (R \text{ は式 (6) により数値計算})$$

対数正規分布：

$$g(t) = \frac{1}{\sqrt{2\pi\sigma'^2} t} \exp\left\{-\frac{(\ln t - \mu')^2}{2\sigma'^2}\right\} \quad \mu = e^{\mu' + \frac{\sigma'^2}{2}}$$

$$\sigma^2 = e^{2\mu' + \sigma'^2} (e^{\sigma'^2} - 1) \quad (R \text{ は式 (6) により数値計算})$$

本プログラムでは、ここに掲載した ワイブル分布、ガンマ分布、対数正規分布、指数分布 (SIR モデル)、SEIR モデル相当分布の中から、適当な分布を選択して使用できるものとする。

3 プログラムの使用法

3.1 メニュー

プログラムのメニュー一覧を Table 1 に示す。

累積感染者データは "File" > "Load from file" メニューを選択してファイルから読み込むか、または "Keyin data" メニューを選択してキーボードから入力する。累積感染者データが読み込まれると、それをもとに日ごとの新規感染者数、指数増幅率、再生産数を計算して、新規感染者数と再生産数の推移をグラフィック画面に表示する。

3.2 入力データ

累積感染者データを読み込ませるファイル名は In*.dat とし (* は任意文字列)、Windows の標準ダイアログからファイルを指定する。ファイルの書式は下記のとおりとする。

```
第 1 行目 : %Zenkoku0215(NHK) (先頭に % を付して任意のコメント)
          2      : 0   53   (第 2 行目以降にデータを記載する)
          3      : 1   59   (第 1 列 に任意の整数値。任意の整数で可)
          4      : 2   66   (第 2 列 に累積感染者数。実数であるが、整数でも可)
```

第 1 行目はコメント記入用であり、当該データの種類、条件等を記入するために用いる (最大 256 文字)。先頭に % がなければ、データ行と見なされる。データ行には 2 つの数値を 1 個以上のスペースを挟んで並べる。1 列目は整数値であり、通し日数のつもりであるが、読み飛ばすので整数でさえあれば任意である。2 列目は当日までの累積感染者数 (新規感染者ではない) を実数として読み込むが、整数値でもよい。前日より小さい数値を読み込むと、前日の数値に変更する旨のメッセージを表示し、OK を返答すればその数値を変更して計算する (ファイルのデータは変更しない)。2 列目以降には、空白以外の文字は記入せず改行すること。1 行ごとの値を日ごとのデータとして順次読み込み、ファイルの終りで読み込みを終了する。また、途中で数値として読み込めない行があれば、それまでに読み込んだデータを有効として、読み込みを終了する。

"Keyin data" メニューを選択すると、ダイアログボックスから日ごとの累積感染者数データを順に入力することができる。前日より小さい値を入力した場合は、前日と同じ値に置き換える。ダイアログの Cancel を選択 (ESC キー) すると、それまでに入力したデータを有効として、キー入力を終了する。

3.3 グラフィック画面表示

ファイル読み込みまたはキー入力により、1 日以上データが入力されると、必要な計算を行って、日ごとの新規感染者数、指数増加率 r および再生産数 R の推移をグラフィック画面に表示する。横軸の日数はデータの日数に応じて、それより少し長い切りの良い日数になる。また縦軸の表示サイズは、新規感染者数 0 ~ 500 人、増幅率 $-0.2 \sim +0.2$ 、再生産数 0 ~ 5 となるので、不都合であれば、グラフィック表示後に "Scale" メニューを選択して、表示範囲を変更すればよい。"Scale" メニューによる変更は表示範囲の変更のみであり、計算結果には何の影響も及ぼさない。

表 1 メニュー一覧

| 主メニュー | サブメニュー | 機能 |
|-------------------|----------------|------------------------------|
| File | Load from file | ファイル In*.dat より累積感染者データを読み込む |
| | Save to file | ファイル Out*.dat へ計算結果を書き出す |
| | Print | グラフィック画面をプリンタへ印刷する |
| | pRinter setup | プリンタの選択・設定 |
| | eXit | プログラムの終了 |
| Keyin data | | 累積感染者データをキーボードから入力する |
| paraMeters change | | 計算パラメータ (発症間隔分布を含む) を変更する |
| sCale change | | グラフ表示範囲のみを変更する |
| Help | | (なし) |

表 2 グラフィック画面時間軸の選択 (単位: 日)

| t_max | t_min | 主目盛 | 副目盛 |
|-------|---------------|------|-----|
| 10 | 0 | 5 | 1 |
| 20 | 0 | 5 | 1 |
| 30 | 0 | 5 | 1 |
| 50 | 0 | 5 | 1 |
| 100 | 0 | 10 | 2 |
| 200 | 0, 100, 150 | 20 | 5 |
| 300 | 0, 100, 200 | 20 | 5 |
| 500 | 0, 200, 300 | 50 | 10 |
| 1000 | 0, 500, 700 | 100 | 20 |
| 2000 | 0, 1000, 1500 | 200 | 50 |
| 3000 | 0, 1000, 2000 | 200 | 50 |
| 5000 | 0, 2000, 3000 | 500 | 100 |
| 10000 | 0, 5000, 7000 | 1000 | 200 |

グラフィック画面時間軸は表 2 のとおりとする (v. 2b での変更)。データを読み込むと、読み込んだデータを含む最小の範囲が選択されている。"Scale" メニューで変更すれば、変更後の範囲を含む最小の範囲が選択される。

"File" > "Print" メニューを選択すれば、現在のグラフ画面を標準プリンタへ再描画する。

3.4 パラメータ変更

"paraMeters change" メニューを選択すると、まず発症間隔分布を選択するダイアログが表示される。5 種類のうちの一つを選ぶと、次にその平均値および標準偏差入力、週単位平滑化処理の有無、および再生産数 R

表 3 入力パラメーター一覧

(a) 発症間隔分布の選択

| プロンプト | 分布の種類 | デフォルト |
|-------------|-----------------|-------|
| Weibull | ワイブル分布 | |
| Gamma | ガンマ分布 | |
| Log. Normal | 対数正規分布 | |
| Exponential | 指数分布 (SIR 相当分布) | |
| SEIR model | SEIR モデル相当分布 | |

(b) 分布パラメータ & 計算パラメータ

| プロンプト | パラメータの機能 | 許容範囲 | デフォルト値 | 単位 |
|------------------------|-------------------|-------------------------|--------|----|
| mean(mu) | 分布の平均値 μ | $\mu > 0$ | 4.8 | 日 |
| s.d.(sigma) | 分布の標準偏差 σ | $\sigma > 0$ | 2.3 | 日 |
| Weekly averaging | 週単位平滑化処理の有無 | On / Off | On | |
| Smoothing points (odd) | R 評価日数 $2m + 1$ | $3 \leq 2m + 1 \leq 21$ | 11 | 日 |

を評価する期間の日数 ($2m + 1$) 設定のダイアログが表示される。

入力した平均値および標準偏差に応じて、その分布形のパラメータ (m 、 η 等) を設定する。指数分布の場合には、入力した標準偏差は無視されて、 $\sigma = \mu$ が代入される。SEIR 相当分布の場合、許容範囲を超えた σ を入力すれば、もっとも近い値に変更される。評価日数としては奇数を念頭に整数値を入力するが、偶数を入れると一つ大きい奇数に変更される。これらは表 3 のデフォルト値が既に設定されているので、必要であれば許容範囲内の値に変更すれば、再度計算を行って結果を表示する。許容範囲外の値を入力すると、適当に変更される。

週単位平滑化処理の有無は、新規感染者数に 7 日周期の周期性が認められる場合を考慮した処理である。この周期性は、おそらく PCR 検査数自体に周期性があることが原因であると推察されるが、日々の新規感染者数変動の大きな原因となっている向きがあり、 R の推測に支障をきたす。このため、ある日の新規感染者数として、その日を含むそれまでの 7 日間の平均 (つまり、当日までの累積値と 7 日前までの累積値の差の $1/7$) を採用することにした。観測データの最初の 6 日間については平滑化する日数を少なくして処理する。元の平滑化しない値を用いる場合には、これを Off にすれば再計算される。

また、"Help" メニューは準備されていない。

3.5 ファイル出力

"File" > "Save to file" メニューを選択して、計算結果を Out*.dat ファイルへ書き出すことができる。書き出した一例を下記に示す。

```
%Zenkoku0215 127,140,000
```

```
% Weibull-d.: mean = 4.800 sdev = 2.300 m = 2.203 eta = 5.420 Weekly averaging ON
```

| % day | y[i] | dy[i] | r (11) | R (11) |
|-------|-----------|----------|---------|---------|
| 0 | 53.00000 | 0.00000 | 0.08004 | 1.44498 |
| 1 | 59.00000 | 6.00000 | 0.08004 | 1.44498 |
| 2 | 66.00000 | 6.50000 | 0.08498 | 1.47662 |
| 3 | 74.00000 | 7.00000 | 0.08737 | 1.49209 |
| 4 | 84.00000 | 7.75000 | 0.11427 | 1.67495 |
| 5 | 94.00000 | 8.20000 | 0.10164 | 1.58709 |
| 6 | 109.00000 | 9.33333 | 0.09567 | 1.54680 |
| 7 | 135.00000 | 11.71429 | 0.09499 | 1.54232 |
| 8 | 147.00000 | 12.57143 | 0.08948 | 1.50587 |
| - - | - - - - | - - - - | - - - - | - - - - |

出力ファイルの第 1 行は入力ファイルの第 1 行そのままであり (Keyin 入力の場合は "Keyin data" と表示)、第 2 行目に計算モデル (SEIR モデル) のパラメータを示す。4 行目以下が、元データの累積感染者数 $y[i]$ 、新規感染者数 $dy[i]$ 、増加率 r 、再生産数 R であり、第 3 行目の $r(11)$ および $R(11)$ の数字 11 は、それらの評価に用いた最小 2 乗近似区間の日数 (奇数) である。データの最初と最後の近くではこの区間日数を 2 ずつ減らし、先頭と最後の日の r と R は、その前後の値をそのまま用いている。

4 適用例

2020 年 2 月 15 日以降、9 月 10 日までの、日本全体の新型コロナウイルス感染状況について、このプログラムを適用した例を以下に示す。この間のおもな出来事を表 4 に示す。2 月 15 日を第 0 日とした時の日数 (時間軸) を、表中に合わせて示す。

この間の日本全体の新規感染者数 dy 、指数関数的増加率 r 、および再生産数 R の時間推移を図 2 に示す。感染者数は、厚生省発表日 (確認日) をもとにして、NHK 新型コロナウイルス特設サイト [2] にまとめられた値を用いた。計算の諸条件は、図中に示されている。

5 使用にあたっての注意とヒント

- (1) 感染拡大の状態を把握するには、日々の新規感染者数そのもの、または指数関数的増加率 r で十分である。再生産数 R を得たとしても、感染状況についてそれ以上の情報が得られるわけではない。
- (2) 感染者数データから再生産数 R を得るには、二次感染の発症間隔分布が既知 (または推定可能) でなければならない。次のステップとして感染伝播のシミュレーションを行う際には、同じ発症間隔分布を用いて求めた再生産数 R を用いなければならない。同じ分布を用いる限り、途中の R の値は正確でないかもしれないが、シミュレーション結果には大きな影響はないと考えられる。
- (3) 取り上げた分布のうち、ワイブル分布、ガンマ分布および対数正規分布は比較的類似の分布になる (ただし尖り度がこの順に高くなる) ので、それらの間に大きな差は生じないと思われる。
- (4) 発症間隔分布の平均値 μ は R の推測値に大きく影響するので、できる限り正確な値を用いる必要がある。標準偏差 σ が大きくなると、発症間隔が短くなるのと同様の効果があり、 R を小さく算出する。
- (5) SEIR モデルは標準偏差に下限があるので、場合によっては不適當となる。COVID-19 に対してもやや不適當である。指数分布 (SIR モデル) は、標準偏差が大きくなる (平均値に等しくなる) ので、 R の

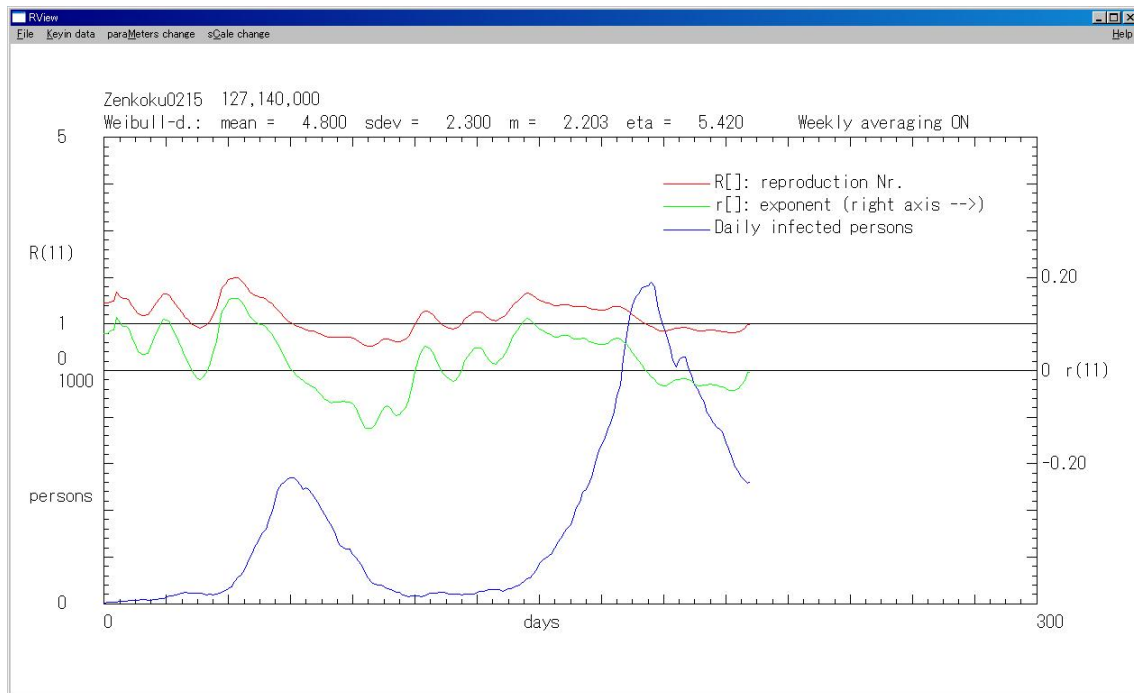


図 2 日本全体の新規感染者と再生産数の推移

値を求めるには不適當である。

- (6) 当然であるが、用いるデータは大きく影響する。感染確認日（公表日）でなく感染日（推定日）をもとにしたデータが望ましい。または、感染確認日を用いざるを得ない場合の、データの扱い方を工夫する必要がある。

参考文献

- [1] S.Yamauchi, "発症間隔分布に基づいた指数増幅率・再生産数間の関係について", 2020.7.30.
 [2] URL, "NHK 特設サイト新型コロナウイルス",
<https://www3.nhk.or.jp/news/special/coronavirus/data-all/>,
<https://www3.nhk.or.jp/news/special/coronavirus/data/>, (2020年7月23日)

表 4 日本全体と東京都における COVID-19 感染の経過

| 月日 | おもな出来事 | 起点よりの日数 (日数) |
|----------|-------------------|-----------------|
| 1月16日 | 全国で最初の感染者 | -30 |
| 1月29日 | 武漢よりチャーター機第1便帰国 | -17 |
| 2月8日 | ダイヤモンド・プリンセス号横浜着岸 | -7 |
| 2月13日 | 東京都内で最初の感染者 | -2 |
| 2月15日 | (起点日) | 0 |
| 2月27日 | 首相が全国の学校へ休校要請 | 12 |
| 3月10日 | WHO がパンデミックと認定 | 24 |
| 4月7日 | 7都府県へ緊急事態宣言発出 | 52 |
| 4月16日 | 全国へ緊急事態宣言拡大 | 61 |
| 5月4日 | 緊急事態宣言を延長 | 79 |
| 5月14日 | 39県で緊急事態宣言を解除 | 89 |
| 5月21日 | 関西2府1県で緊急事態宣言を解除 | 96 |
| 5月25日 | 首都圏・北海道の緊急事態宣言解除 | 100 |
| 6月19日 | 都府県間自粛要請、全国解除 | 125 |
| 7月22日 | Go To キャンペーン開始 | 158 |
| 8月7日 | お盆前 | 174 |
| 8月8-10日 | 3連休 | 175-7 |
| 8月13-16日 | お盆 | 180-183 |
| 8月24日 | 専門家会議、2波ピークアウトを発表 | 191 |
| 8月28日 | (安倍首相突然辞意表明) | 195 |
| 9月10日 | (観測値の最終日) | 208 |
| 9月11日 | GoTo 東京解除発表 | 209 |
| 10月1日 | GoTo 東京解除 | 229 |