

プログラム RView 説明書

S. Yamauchi

2020 年 6 月 9 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	時系列データから成長率を求める	2
2.2	成長率から再生産数を求める	3
3	プログラムの使用法	4
3.1	メニュー	4
3.2	入力データ	4
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	6
4	適用例	7
4.1	日本全体 (2020 年 1 月 16 日 ~ 5 月 31 日)	8
4.2	東京都 (2020 年 2 月 13 日 ~ 5 月 31 日)	8

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(世代間隔)が関与するので、観測データから求まる再生産数は、その感染症の世代間隔分布により相違する。このプログラムでは世代間隔分布が指数分布(またはその複合)で表される SIR モデルおよび SEIR モデルを対象とする。

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、その指数関数的成長率を求める。
- (2) 成長率から仮定した世代間隔分布に応じた再生産数を求める。

2.1 時系列データから成長率を求める

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、

$$\begin{aligned} \sum_i y_i^2 (A' + rt_i - \ln y_i) &= 0 \\ \sum_i y_i^2 t_i (A' + rt_i - \ln y_i) &= 0 \end{aligned}$$

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。

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

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

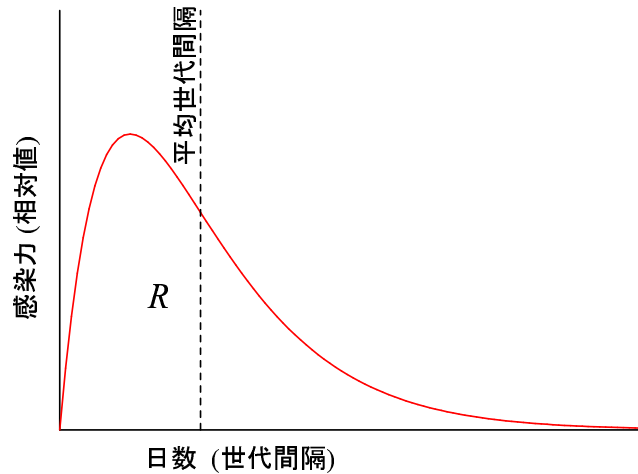


図1 二次感染曲線 $n(t)$ の模式図

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' は必要ない)。

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、また両端では r を求めないか、または前後の日の値で代用する等の工夫を要する。

2.2 成長率から再生産数を求める

成長率 r と再生産数 R の関係はやや複雑であり、別紙資料（「世代間隔分布にもとづいた成長率・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（世代間隔）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

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

は世代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世代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 (成長率 r の) 指数関数的増加を引き起こす。

(医学的実験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幅率 r と再生産数 R を関係づけることができる (詳細は別紙を参照)。

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

いくつかの代表的な世代間隔分布 $g(t)$ に対して、成長率 r と再生産数 R の間の関係等を下表に示す。

分布の種類	$g(t)$	平均値	分散	R	備考
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	$\frac{1}{T_I - T_E} (e^{-t/T_I} - e^{-t/T_E})$	$T_E + T_I$	$T_E^2 + T_I^2$	$(1 + rT_E)(1 + rT_I)$	T_E : 平均潜伏期間 T_I : 平均感染性期間 $r > -1/T_E, r > -1/T_I$
ガンマ分布	$\frac{b^n}{\Gamma(n)} t^{n-1} e^{-bt}$	$\frac{n}{b}$	$\frac{n}{b^2}$	$(1 + \frac{r}{b})^n$	$r > -b$
正規分布	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{r\mu - (1/2)r^2\sigma^2}$	
デルタ分布	$\delta(t - \mu)$	μ	0	$e^{r\mu}$	

このうちデルタ分布は二次感染が常に一定日数 T_c 後に生じる場合であり、平均世代間隔 (表の平均値列) が等しい全ての分布に対して R の上限となる。本プログラムでは、SEIR モデル、およびその一部分としての SIR モデルを対象とする。

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 : % Tokyo from 2020.02.12 to 2020.05.31 (先頭に % を付して任意のコメント)
 2 : 0 0 (第 2 行目以降にデータを記載する)
 3 : 1 3 (第 1 列 に任意の整数値。任意の整数で可)

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		計算パラメータを変更する
sCale change		グラフ表示範囲のみを変更する
Help		(なし)

4 : 2 5 (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と再生産数 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、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。

"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。

表2 入力パラメーター一覧

入力プロンプト	パラメータの機能	許容範囲	デフォルト値	単位
Latent period(T_E)	平均潜伏日数 T_E	$0 \leq T_E$	4.6	日
Infectious period(T_I)	平均感染日数 T_I	$0 < T_I$	5.0	日
Weekly averaging	週単位平滑化処理の有無	有 (On) / 無 (Off)	On	
Smoothing points (odd)	R 評価日数 $2m + 1$	$3 \leq 2m + 1 \leq 21$	11	日

3.4 パラメータ変更

"paraMeters change" メニューを選択すると、SEIR モデルにおける平均潜伏期間日数 T_E 、平均感染性期間日数 T_I 、週単位平滑化処理の有無、および再生産数 R を評価する期間の日数 ($2m + 1$) を変更することができる。 R 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2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。 $T_E = 0$ とすれば SIR モデルに一致する。

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。

```
% Tokyo from 1020.02.12 to 1020.05.31
% SEIR model   T_E =   4.600   T_I =   5.000   Weekly averaging ON
% day          y[i]          dy[i]          r (11)          R (11)
   0           0.00000        0.00000        0.40547         8.67371
   1           1.00000        1.00000        0.40547         8.67371
   2           3.00000        1.50000        0.09775         2.15823
   3           5.00000        1.66667        0.04159         1.43902
   4           6.00000        1.50000        0.05046         1.54296
   5           7.00000        1.40000        0.04414         1.46854
   6          10.00000        1.66667        0.03460         1.35964
```

表 3 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	
		全国	東京
1月15日	(日本全体の起点日)	0	
1月16日	全国で最初の感染者	1	
1月29日	武漢よりチャーター機第1便帰国	14	
2月8日	ダイヤモンド・プリンセス号横浜着岸	24	
2月12日	(東京都の起点日)	28	0
2月13日	東京都内で最初の感染者	29	1
2月27日	首相が全国の学校へ休校要請	43	15
3月10日	WHO がパンデミックと認定	55	27
4月7日	7都府県へ緊急事態宣言発出	83	55
4月16日	全国へ緊急事態宣言拡大	92	64
5月4日	緊急事態宣言を延長	110	82
5月14日	39県で緊急事態宣言を解除	120	92
5月21日	関西2府1県で緊急事態宣言を解除	127	99
5月25日	首都圏・北海道の緊急事態宣言解除	131	103
5月31日	(観測値の最終日)	137	109

7	13.00000	1.85714	0.03555	1.37038
8	14.00000	1.85714	0.02906	1.29842
- - -	- - - - -	- - - - -	- - - - -	- - - - -

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 $dy[i]$ 、成長率 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

新型コロナウイルス感染症 (COVID-19) には、日本では、中国武漢より帰国した神奈川県男性が1月16日に最初に発症した。それ以降の感染に関する主な出来事を表3に示す。以下では、日本全体および東京都での感染拡大に対してこのプログラムを適用するが、その便宜のために、最初に感染者が発生した日を第1日とした時の当日までの日数 (時間軸) を、表3の中に合わせて示す。

4.1 日本全体 (2020年1月16日～5月31日)

この間の日本全体の新規感染者数(元は累積数)と再生産数 R の時間推移を、図2、図3に示す。図2のパラメータはデフォルトのままであり、図3はその表示座標のみを変更したものである。感染者数は厚生省発表日(確認日)をベースとしており、クルーズ船などを含まないものである。朝日新聞デジタル^{*1}による集計結果を用いた(以下同様)。

厚生省のCOVID-19対策専門家会議が5月29日に公表した提言^{*2}の中に、図4の分析結果が示されている。専門家会議の資料は特定した、または推測した感染日(未公表)をベースとしているので、新規感染者数のデータそのものが相違しているが、おおむね一致していると判断できるであろう。ただし、新規感染者数が少ない期間(0～45日または1月、2月)では、曲線近似に無理があるので R の値の信頼性は低い(図4には95%信頼区間が薄青で示されている)。

週単位平滑化処理をOffにして、日々の新規感染者数をそのまま処理した結果を図5に示す。感染者数データだけでなく R の変動も大きく、傾向がつかみにくい。これらのデータに対しては、週単位平滑化処理を行うことが必要であることがわかる。

また、週単位平滑化処理はOnに戻して、 R 評価の区間の日数を11から5または21に変更して計算した結果を、図6と図7に示す。いずれが適当かの根拠は薄いですが、評価日数を11とした図3の結果が図4の専門家会議の曲線に最も近いようである。確認日でなく感染日をベースにしたデータを使うことができれば、7、5程度のより少ない日数で近似しても良いかもしれない。

4.2 東京都 (2020年2月13日～5月31日)

東京都では2020年2月13日に最初の感染者が確認された(感染は1月18日)。これ以降、第1波がほぼ終わったと思われる5月31日までの109日間について、都内居住者の新規感染者数と再生産数 R の推移を図8に示す。感染者数は厚生省発表日(確認日)をベースとしており、出所は日本全体の場合と同じである。

5月29日公表の厚生省専門家会議による R の分析結果を、図9に示す。おおむね似た形になっているが、図8では40～45日目あたりで R の大きな山が見られるが、図9の対応する3月末～4月初めにはそのような山は見られない。この期間は感染確認者が3から5人前後から13～40人へと急速に増加しており、確認日ベースによるデータそのものに起因していると思われる^{*3}。

^{*1} <https://www.asahi.com/topics/word/コロナウイルス.html>

^{*2} 新型コロナウイルス感染症対策専門家会議「新型コロナウイルス感染症対策の状況分析・提言」(令和2年5月29日)
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00093.html

^{*3} 感染者数が比較的少ない期間である感染者の該当日が移動すると、曲線の形に大きく影響する。

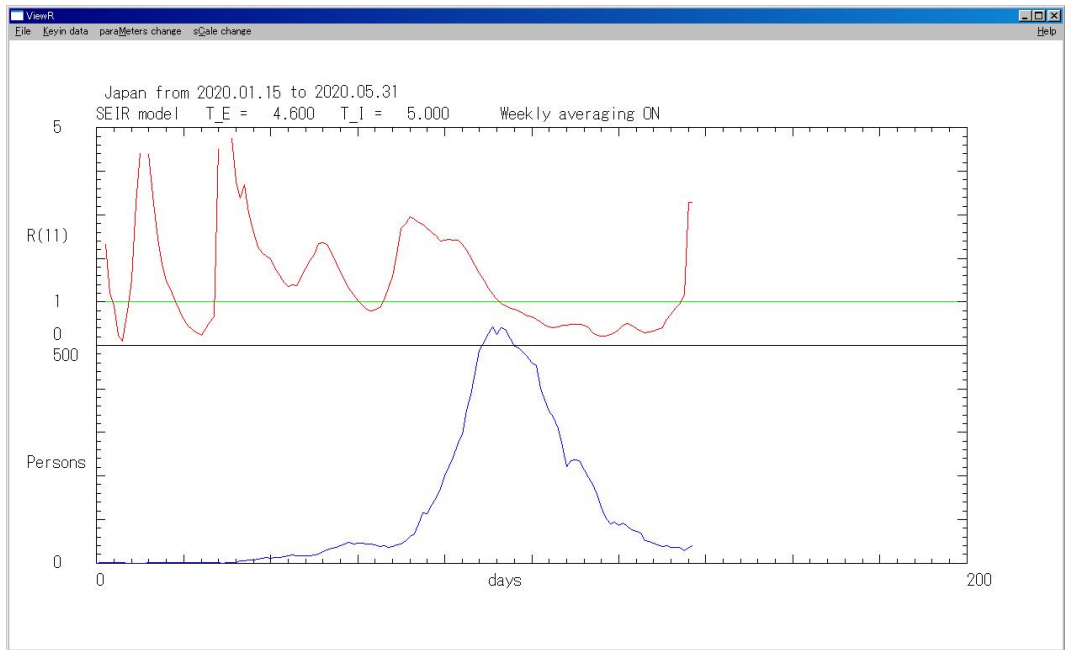


図 2 日本全体の新規感染者と再生産数の推移 (デフォルトのパラメータのまま)

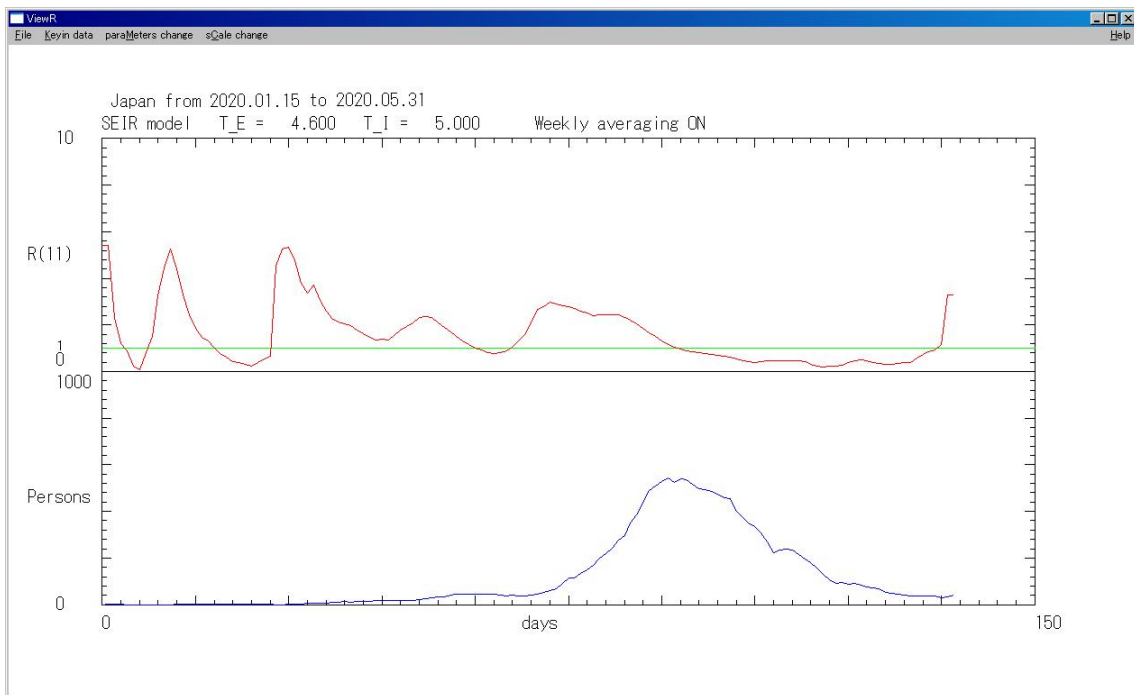


図 3 日本全体の新規感染者と再生産数の推移 (図 2 の座標のみ変更)

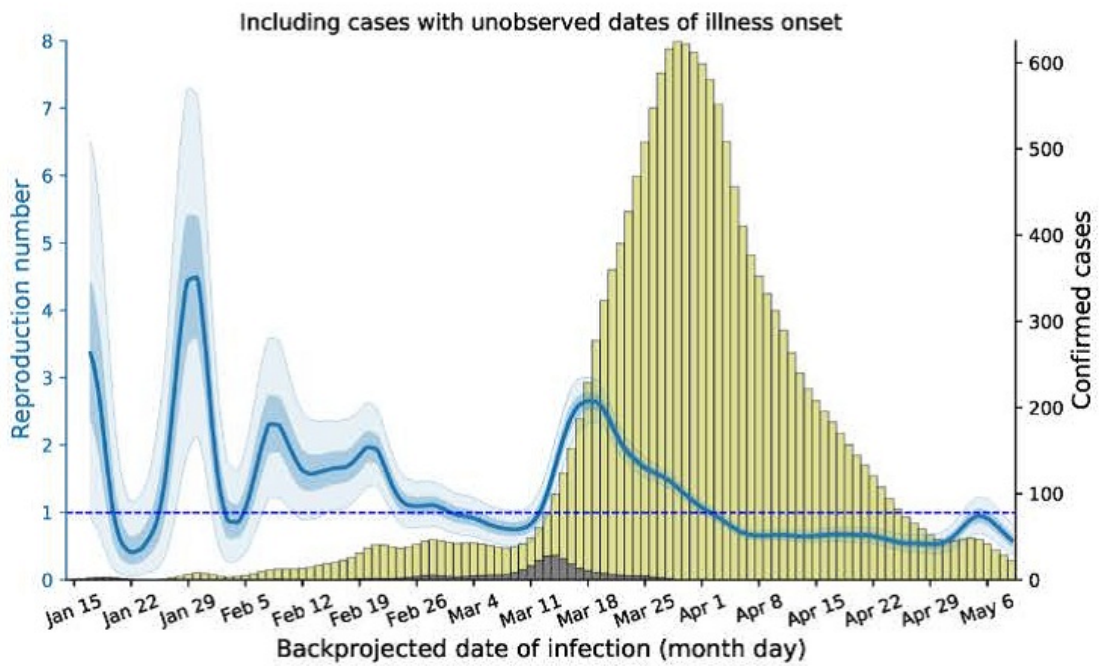


図4 5月29日専門家会議による日本全体の新規感染者と再生産数の推移

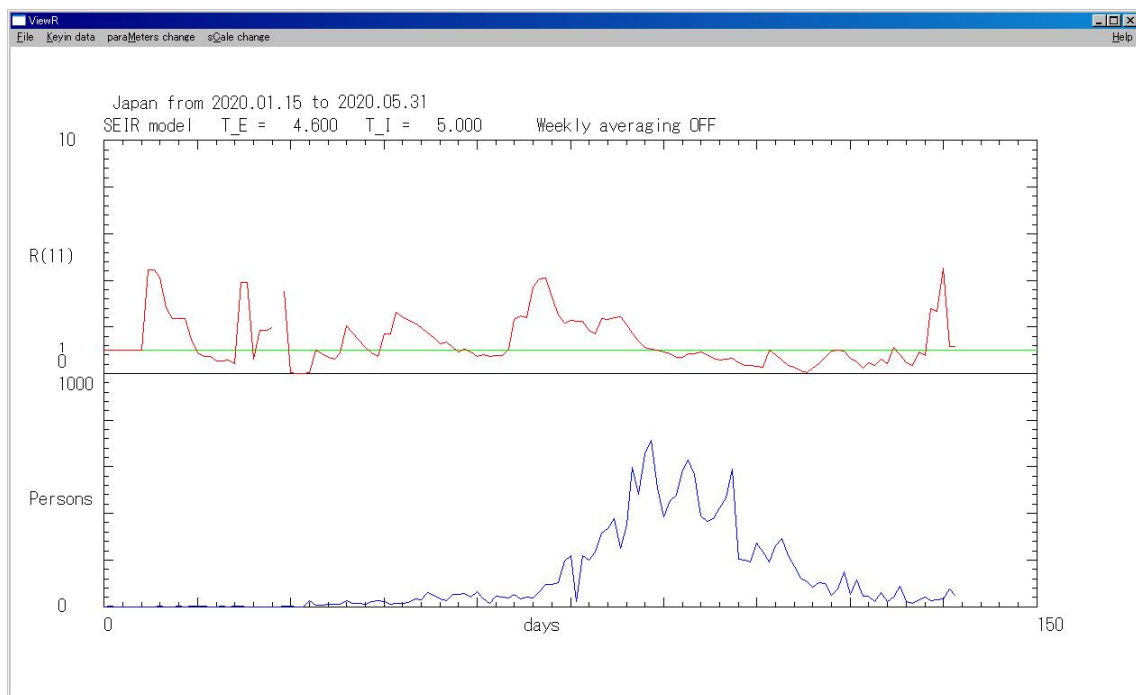


図5 日本全体の新規感染者と再生産数の推移 (週単位平滑化処理を Off にした場合)

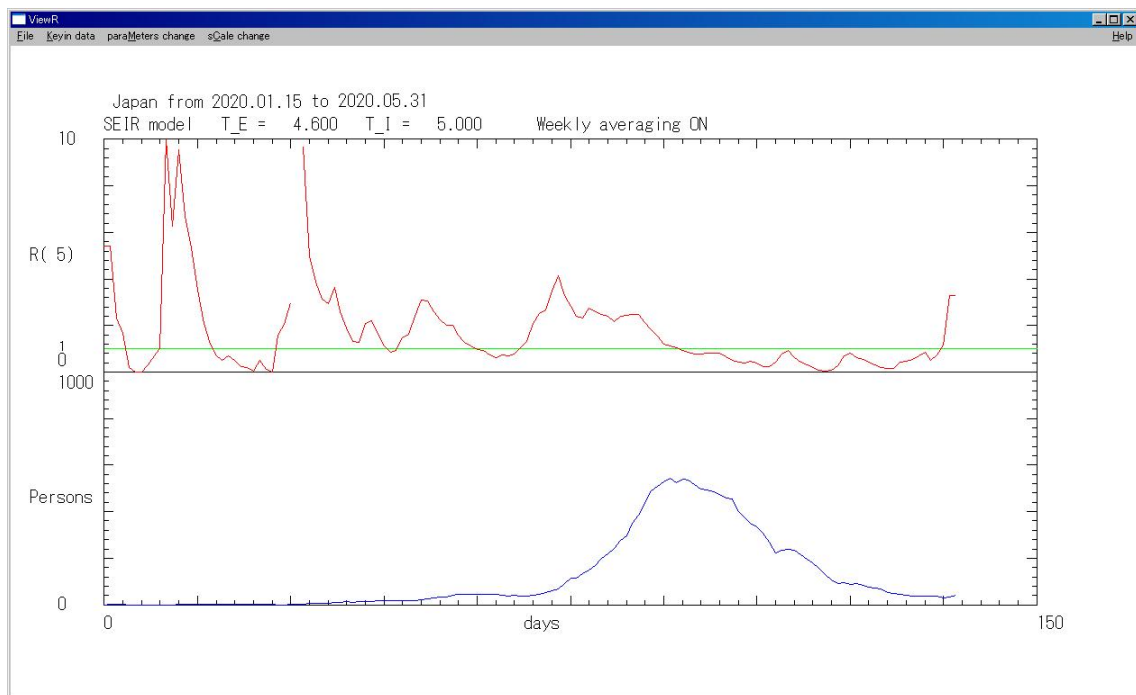


図 6 日本全体の新規感染者と再生産数の推移 (R 評価の区間の日数を 5 にした場合)

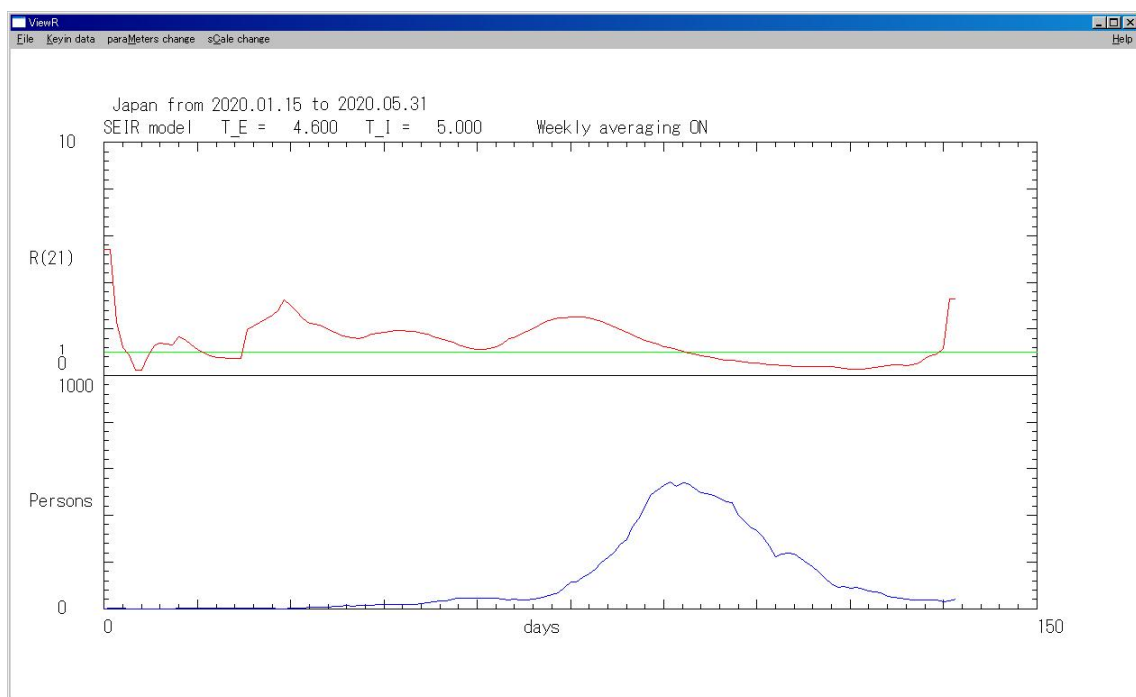


図 7 日本全体の新規感染者と再生産数の推移 (R 評価の区間の日数を 21 にした場合)

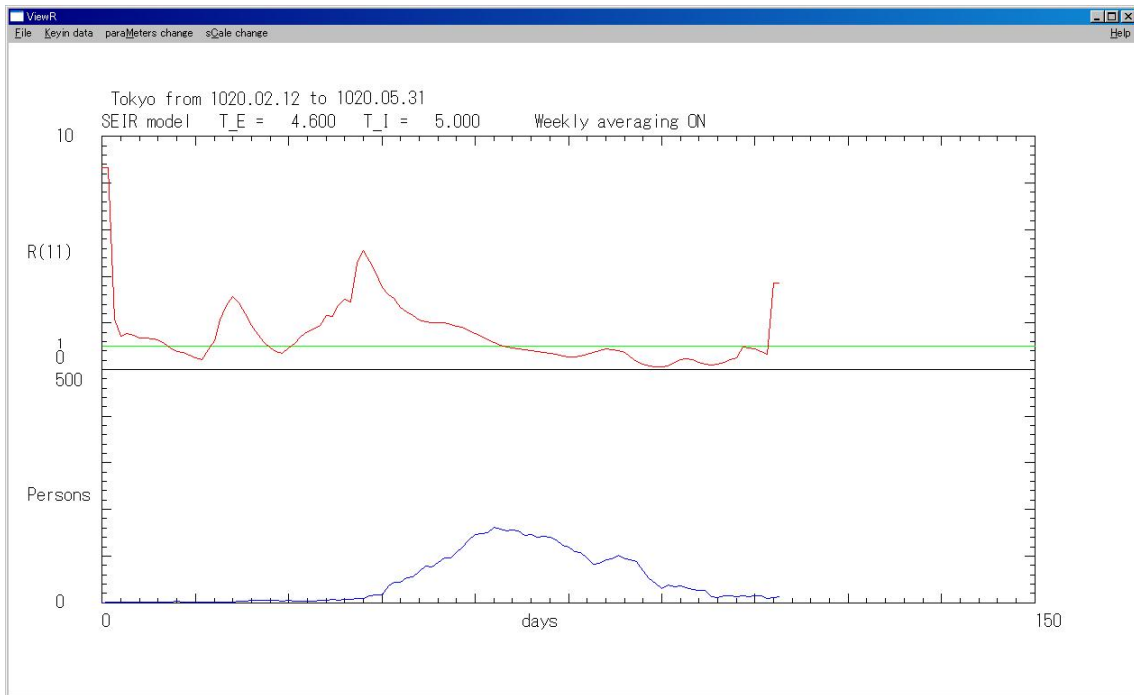


図 8 東京都の新規感染者と再生産数の推移 (座標値以外デフォルトのまま)

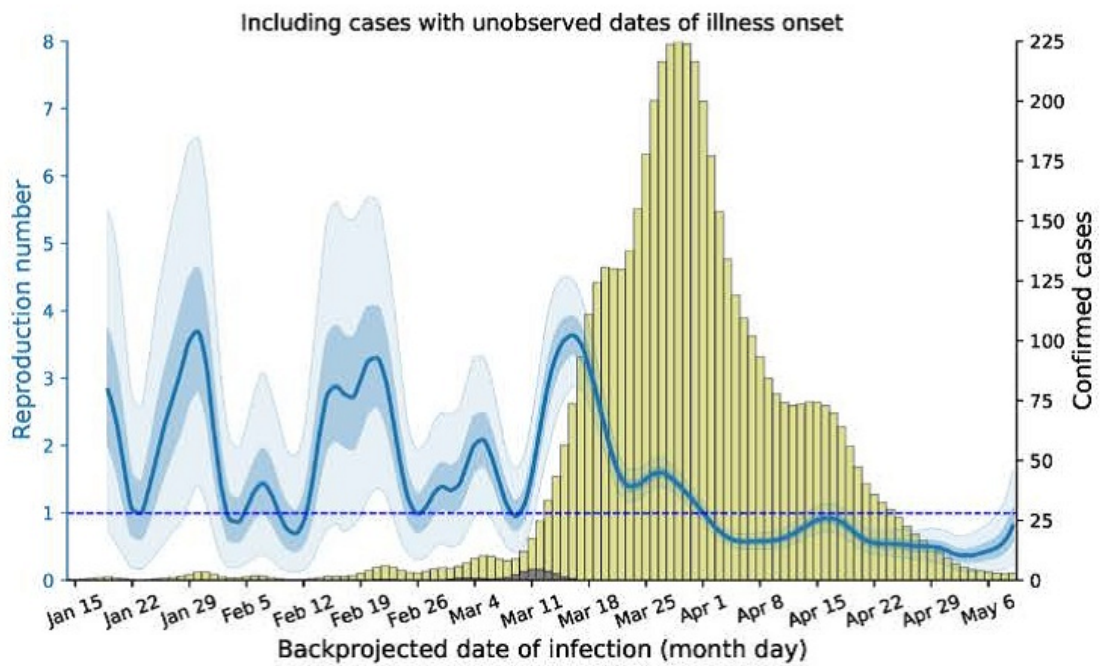


図 9 5月29日専門家会議による東京都の新規感染者と再生産数の推移

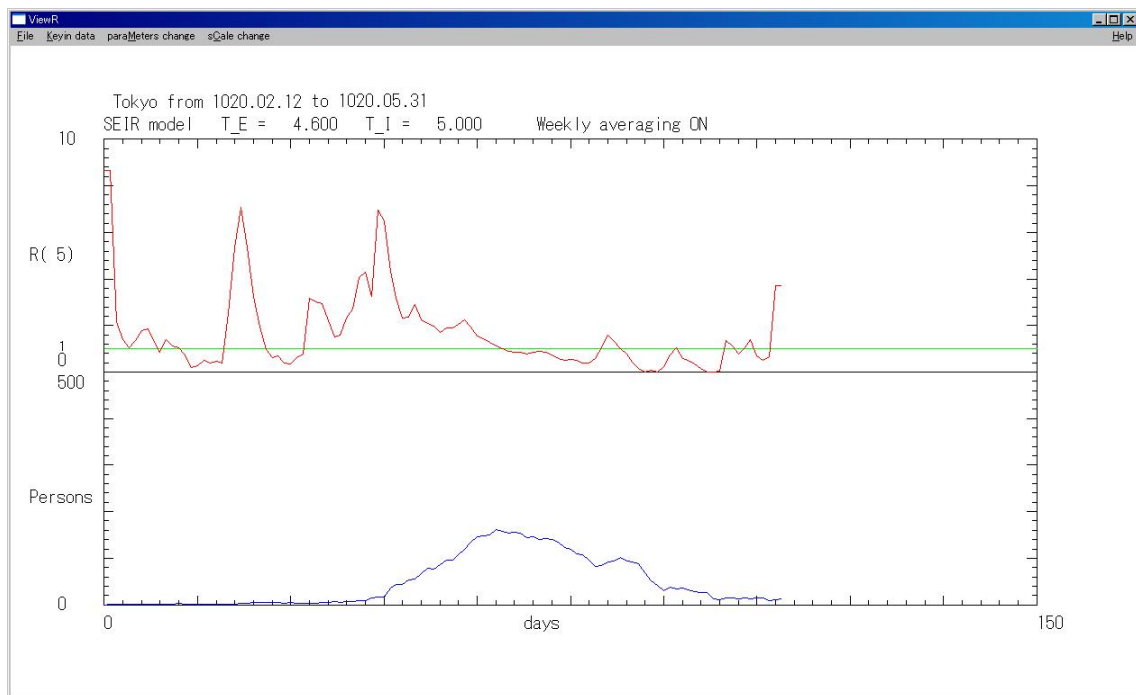


図 10 東京都の新規感染者と再生産数の推移 (R 評価の区間の日数を 5 にした場合)

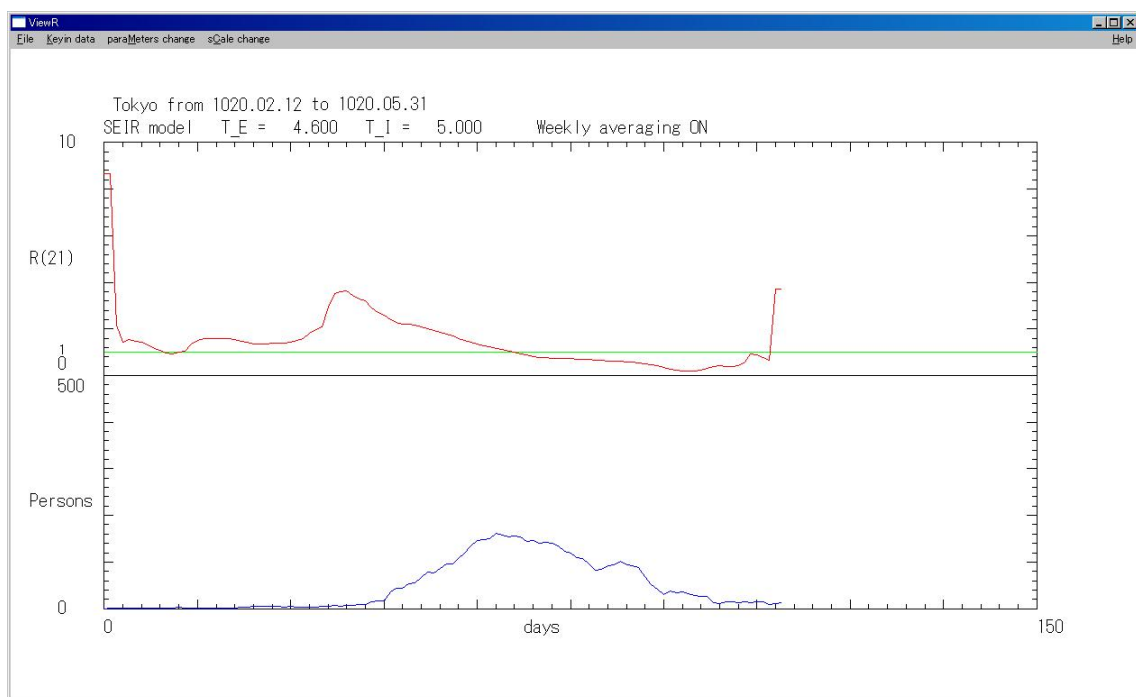


図 11 東京都の新規感染者と再生産数の推移 (R 評価の区間の日数を 21 にした場合)