



Assessing the spread of the novel Coronavirus in the absence of mass testing

Journal:	<i>International Journal of Clinical Practice</i>
Manuscript ID	IJCP-05-20-0427.R2
Wiley - Manuscript type:	Original Paper
Date Submitted by the Author:	21-Oct-2020
Complete List of Authors:	Miles, David; Imperial College London, Business School Dimdore-Miles, Oscar; Oxford University, Atmospheric, Oceanic and Planetary Physics
Specialty area:	

SCHOLARONE™
Manuscripts

Assessing the Spread of the Novel Coronavirus In The Absence of Mass Testing

Abstract

Background: Assessing why the spread of the COVID 19 virus slowed down in many countries in March through to May of 2020 is of great significance. The relative role of restrictions on behaviour ("lockdowns") and of a natural slowing for other reasons is difficult to assess when mass testing was not widely done. This paper assesses the evolution of the spread of the COVID 19 virus over this period when there was not data on test results for a large, random sample of the population.

Method: We estimate a version of the SIR model applied to data on the numbers who were tested positive in several countries over the period when the virus spread very fast and then its spread slowed sharply. Up to the end of April 2020 test data came from non random samples of populations who were overwhelmingly those who displayed symptoms. Using data from a period when the criteria used for testing (which was that people had clear symptoms) was relatively consistent is important in drawing out the message from test results. We use this data to assess two things: how large might be the group of those infected who were not recorded; how effective were lockdown measures in slowing the spread of the infection.

Results: We find that to match data on daily new cases of the virus, the estimated model favours high values for the number of people infected but not recorded.

Conclusions: Our findings suggest that the infection may have spread far enough in many countries by April 2020 to have been a significant factor behind the fall in measured new cases. Government restrictions on behaviour - lockdowns - were only one factor behind slowing in the spread of the virus.

1. Introduction

There is significant uncertainty about the degree to which the novel coronavirus (COVID19) has spread. This is particularly true over the period

1
2
3
4
5
6
7
8 when, in many European countries, those testing positive for the infection
9 rose fast and subsequently slowed markedly. Over the period between Febru-
10 ary and May 2020 in most countries there was no mass testing. Although
11 testing capacity did rise, it did so along with the rise in infections and testing
12 was largely confined to those with symptoms. This has very significant policy
13 implications - [1] shows that different policies aimed at controlling the virus
14 can have very different effects on the numbers who become infected and show
15 symptoms depending on the proportion of those who are asymptomatic. It
16 is those with symptoms who are at risk of death from the virus and so the
17 relative size of the populations of symptomatic to asymptomatic amongst
18 the infected is of enormous significance to welfare, including mortality rates,
19 and to policy [2].

22 The degree of uncertainty about that asymptomatic rate is large enough
23 to mean that neither 0.3 or 0.9 is outside the range of plausible values, though
24 the implications of those two numbers are very different. [3] estimate that
25 86 per cent of all infections were undocumented prior to the Wuhan travel
26 shutdown (on January 23, 2020). In contrast, estimates based on infections
27 amongst passengers on the cruise ship Diamond Princess put the proportion
28 of asymptomatic (or near asymptomatic) cases at around 50 per cent. [4]
29 report enormous ranges for the possible values of the infection rates in Illinois,
30 New York and Italy. As of April 6th 2020 these ranges are estimated as [0.001,
31 0.517], [0.008, 0.645], and [0.003, 0.510] respectively.

34 While large scale testing of a random sample of the population would
35 narrow the range of plausible values for the asymptomatic proportion of the
36 infected, [1], such testing did not come in many countries until after the end
37 of April. That makes it harder to judge now why the spread of the infection
38 seemed to slow so sharply in March and April. In most countries, including
39 the USA and the UK, testing up to the end of April 2020 was concentrated
40 on those who displayed symptoms or were at high risk; it was certainly not
41 random. However, testing capacity did rise strongly over this period; in the
42 UK testing went from under 10,000 a day in early March to almost 100,000
43 a day by early May (according to Office for National Statistics data). Until
44 late April UK testing was overwhelmingly done in hospitals on those who
45 displayed symptoms of the virus. So called "pillar 2" tests have subsequently
46 increased greatly and are carried out on the wider population outside of
47 hospitals.

51 In this paper we implement a version of the SIR (susceptible-infected-
52 recovered) model to estimate the numbers infected from data on the non-
53

1
2
3
4
5
6
7
8 random sample of those tested. We use data from the UK, where relatively
9 few of those showing no symptoms had been tested up to the end of April
10 2020, to provide estimates of the spread of the virus over that period.

11 We also apply the model to the US, Italy, Spain, France and Sweden. In
12 doing so we allow the estimated parameters of the model to vary thereby
13 accounting for a range of factors that are different between countries and
14 which are likely to affect the spread of the virus. Population density, climate,
15 age-structure, working patterns, family structure and living arrangements
16 are likely to vary across countries and have an impact on how the virus is
17 transmitted. Our statistical technique allows for these factors that differ
18 between countries, but which are relatively stable within countries over the
19 months we focus on.
20
21

22 The results for these countries are similar to the UK - the numbers of
23 those untested but infected that seems to best fit the data is very high; far
24 higher than is estimated based on the limited amount of results from more
25 widespread testing which went beyond those who showed symptoms. But
26 other studies using UK data do find evidence of wider spread of the virus
27 in March and April. An Oxford University research team used death data
28 to estimate the proportion of the population who might have built up some
29 form of immunity before the UK lockdown was introduced in mid-March.
30 They put that fraction at around 60% [5]. [6] used data on differences in the
31 spread of the infection across English regions to assess how many might have
32 been infected and put that fraction at similarly high levels, as does [7].
33
34
35
36

37 2. The Model

38
39 We use a version of the SIR model which closely follows [1]. At each
40 point in time the population is made up of three distinct groups: those
41 who are currently infected (I_t); those who are susceptible (S_t) and those
42 who have recovered (R_t). We assume a constant population and that the
43 death rate is low enough to mean that this is reasonable. At each point
44 in time only some fraction of those infected are tested and show a positive
45 result. Over the period we consider, it is likely that many of those who
46 were infected but not tested had mild, or no, symptoms. We use the symbol
47 π_a to denote the fraction of those infected who are not recorded. There
48 is some evidence that the degree to which the asymptomatic are infectious
49 may be different from those who have symptoms [8], but we will initially
50 assume that the transmission rates are the same for those infected, whether
51
52
53
54

1
2
3
4
5
6
7
8 tested or not. We denote the number of people infected at time t by I_t .
9 We distinguish within this group between those who have tested positive
10 (denoted I_{st}) and those who were not tested or incorrectly tested negative
11 (I_{at}) such that $I_t = I_{st} + I_{at}$. (We use the subscripts s and a for these groups
12 because those who were tested were disproportionately those with symptoms
13 while those who were infected but not tested were likely to have had a higher
14 proportion of the asymptomatic). The evolution of S_t , I_t and R_t in discrete
15 time is given by the dynamic system:

$$16 \quad \Delta S_t = -\beta_t I_{t-1} \frac{S_{t-1}}{N} \quad (1)$$

$$17 \quad \Delta R_t = \gamma I_{t-1} \quad (2)$$

$$18 \quad \Delta I_t = \beta_t I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1}, \quad (3)$$

19 ΔS is the change in the population of the susceptible; N is the total
20 population, β_t is the transmission rate of the virus at a time t (the mean
21 number of people an infectious person will infect per unit time) and γ is
22 the rate of recovery. The initial infection rate over the infectious period, the
23 reproduction number, is defined as $R_0 = \frac{\beta_t}{\gamma}$. Initially We shall assume that
24 π_a is a constant so that

$$25 \quad I_{st} = (1 - \pi_a) I_t \quad (4)$$

26 and

$$27 \quad I_{at} = \pi_a I_t. \quad (5)$$

28 The number of new cases at time t (y_t) can be calculated as

$$29 \quad y_t = \Delta I_t + \gamma I_{t-1}. \quad (6)$$

30 New cases are the sum of the change in the number of outstanding cases
31 plus the numbers recovered. The number of new recorded cases (y_{st}) is

$$32 \quad y_{st} = (1 - \pi_a)(\Delta I_t + \gamma I_{t-1}) = (1 - \pi_a) \left(\beta_t I_{t-1} \frac{S_{t-1}}{N} \right). \quad (7)$$

33 The strategy that we pursue is to use the data on the numbers of new cases
34 who test positive for the virus. We then seek the values of the parameters

of the model - and in particular π_a - that give a predicted y_{st} that matches the data. We use data on the numbers of those who test positive (in the UK and in other countries) as the variable we are trying to match; other studies ([9],[5]), use the number of deaths. There would seem to be significant ambiguity over assignment of the cause of death to the virus, perhaps more than over whether a positive test is reliable or not.

We use data on tests up the end of April 2020 by which time nearly 500,000 had been tested in the UK and around 175,000 had tested positive (according to data from the Office of National Statistics). It is over this period that new cases testing positive first rose dramatically and then began to fall sharply a few weeks after lockdown began. It is the main purpose of this paper to identify the contribution to this slowdown of behavioural changes and of the natural dynamics of infections arising from a shifting stock of those who have been infected. This is why we focus on the February-April period.

To implement the estimation of the model we need to make assumptions about the transmission rate of the virus β_t and the recovery rate γ . The transmission rate will not have been constant because of policy measures introduced to slow the spread of the infection and because of behavioural changes that were happening even in the runup to the lockdown. In the UK "lockdown", which began on March 23rd, was strict and social distancing was already happening just before this date; both will likely have brought β down significantly. Similar policies were adopted at various times in March 2020 in other countries. We assume a constant value of β_t before the lockdown date (of β_0), followed by a gradual reduction in the β_t value after this date to simulate the effect the measures have on transmission. The initial value of β_0 is derived from the value of the initial reproduction rate R_0 and the recovery rate γ , using the relation $\beta_0 = \gamma R_0$. We try all values for an initial reproduction rate ranging from 2.2 up to 3.9 at intervals of 0.005. We try three values of the recovery rate implied by half lives of the period of infectiousness - that is the number of days it takes for half an initial number of infected people to recover - of 4 days, 6 days (as used by [1]) and 8 days. The corresponding three values of γ are 0.159, 0.109 and 0.0833.

We assume that after the lockdown date there is a lag until the value of β_t starts to change from β_0 . The lag is between the lockdown measures starting and the impact on the numbers testing positive for the virus. That lag reflects several distinct factors: it must include the lag in the impact on new infections, the lag before symptoms show, the lag before testing the

1
2
3
4
5
6
7
8 symptomatic and finally the lag before results are known and recorded in the
9 daily measure. We set the overall lag at 14 days, but also assess sensitivity of
10 results to shorter lags in part because social distancing was already happening
11 just before lockdown. After this lag, β decays exponentially towards a value
12 of β_L , the post lockdown asymptotic β . The time path for β_t can be expressed
13 as
14

$$15 \beta_t = \begin{cases} \beta_0, & \text{if } t \leq t^* \\ \beta_0 - (\beta_0 - \beta_L)(1 - e^{-(t-t^*)\lambda}) & \text{if } t > t^*, \end{cases} \quad (8)$$

16 where t^* is the lockdown time plus the 14 day lag period and λ is the
17 speed of adjustment in β after lockdown measures begin to take effect. We
18 assume that once the lockdown does begin to affect numbers testing positive
19 it quite quickly reaches its full effectiveness, bringing the transmission rate
20 down so that half of its long run impact on β comes through in 3 days,
21 implying that $\lambda = 0.231$.
22

23 For given values of γ , β_0 and λ we search for the values of the two free
24 parameters - β_L and π_a - so as to maximise the fit of the model. We choose
25 those two free parameters to minimise the root mean squared (RMS) devi-
26 ation between the daily data on the numbers of new positive tests for the
27 virus and the model prediction of that number (y_{st}). The parameters we fit
28 are a measure of how effective the lockdown is in bringing down the infection
29 rate (measured by how much lower β_L is relative to β_0) and the ratio of those
30 infected but not tested to the total population of the infected (π_a).
31

32 These key parameters are the ones which best match each country's data
33 on test results. Separate estimation of these parameters for each country
34 allow for cross-country differences in characteristics that might affect the
35 spread of the virus.
36

37 3. Sensitivity to key assumptions and calibration

38 Before showing results we stress that our model relies on a number of key
39 assumptions.
40

41 In its simplest version, we assume π_a is a constant but we then allow for
42 it to vary with changes in testing capacity. Overall, relatively few in the
43 UK with no symptoms had been tested up to the end of April 2020. For
44 the other countries, we are also focusing on a period over which testing was
45 largely confined to those with clear symptoms. This is important because
46 it means that the results from tests over this period can be expected to
47
48
49
50
51
52
53
54

1
2
3
4
5
6
7
8 reflect the spread of the virus much better than if we extend the period to
9 one where testing became far more widely available to those with few or no
10 symptoms. Nonetheless there was some variability in the criteria for testing
11 over the period up to the end of April 2020. In the UK testing was ramped
12 up strongly in late April and as it was, the ratio of daily positive tests to
13 total tests fell. We take account of the possible impact of that below but
14 first continue with the assumption of a fixed π_a .

15
16 For the UK the model is initialised on data from the 31st of January,
17 the date on which the first non zero value of positive test cases is recorded.
18 At this time, testing was only applied to those who had travelled to certain
19 regions of China and presented with symptoms and therefore data in the first
20 week or so may not be fully representative of all symptomatic cases.

21
22 We rely on estimates of R_0 and of γ to generate a value for β_0 . There
23 is considerable uncertainty about both. At the lower end of the ranges of
24 values used in simulations are those chosen by [8], who assume a value of
25 2.4, and [5] who take figures centred around 2.25 or 2.75. [1], who draws on
26 estimates using data from Wuhan, uses a much higher figure for simulations
27 with a pre-shutdown value for R_0 of 3.8. The range of estimates of R_0 from
28 several studies is between 2.2 and as high as 3.9. A team at the London
29 School of Hygiene and Tropical Medicine found 11 published estimates of R_0
30 for Covid-19, which averaged 2.68 with a standard deviation of 0.57 (see Paul
31 Taylor, London Review of Books, May 2020, vol 42, no 9). The range we
32 use for simulations is 2.2 to 3.9 - values outside this range gave a poor fit to
33 the data for all countries we analysed for any values of the other parameters.
34 For our estimate of γ we assume the half life of the infection as x days and
35 therefore that γ satisfies the equation $(1 - \gamma)^x = 0.5$. We take x as 4, 6 or 8
36 days - a range which encompasses those used in several studies.

37
38 As noted above we assume that once lockdown begins beta is reduced so
39 that it declines asymptotically towards a value that would then be maintained
40 as long as the lockdown remains in place (β_L in our equations). Our choice
41 of the speed with which beta declines towards its steady-state value, after
42 the initial lag, is such that the transition is fairly rapid, corresponding to a
43 half life of 3 days ($\lambda = 0.231$).

44
45 The data we try to fit is the number of new infections recorded. Testing
46 of people with no symptoms was (up to late April 2020) relatively small
47 scale in the countries we analyse and to a large extent limited to those at
48 high risk. We use a grid search to find the values of the two unknown and
49 free parameters (β_L and π_a) to minimise the root mean squared deviation
50
51
52
53

between the observations and (y_{st}) , given the choice of other parameters.

4. Results

Figure 1 shows the data on new cases of those testing positive for the virus in the UK. The data start on January 31st. The data is from the Office for National Statistics. (The spike in reported new cases on 11/04/2020 coincides with an expansion in testing capacity).

Figure 2 shows the best fit of the model when we set the half life of the infection to 6 days ($\gamma=0.109$). The best fit for this value of γ was when $R_0 = 2.5$ and β_L and π_a are 0.1928 and 0.996 respectively. These values imply that the transmission rate started to turn down sharply by the end of first week of April, some 2 weeks after the lockdown began. The value for π_a is very high - implying that, in the period up to the end of April, there were around 250 people with the infection for every person who tested positive. If that were true then by April 20th - by which time around 120,000 had tested positive for the virus (and the great majority of whom had shown symptoms) close to 45% of the UK population might have had the virus.

Figure 3 shows the root mean square error of the model for all combinations of parameters π_a and β_L . The area on the far right of the figure shows the fit of the model in cases where the lockdown had limited effect (i.e. β_L is little different from β_0). The model fits the data very poorly in this region - as illustrated by the darker shading which reflects a high value of the root mean square error - suggesting the lockdown had a significant impact.

Figure 4 shows the model fit to the UK data when we set the half life of the virus at 8 days ($\gamma =0.0833$). The best fit here was with a value of R_0 of 2.95 and β_L and π_a of 0.1598 and 0.996 respectively. Once again the best fit value of π_a is very high and it implies that approximately 45% of the UK population may have been infected by late April. Figure 5 shows the parameter combinations that have a goodness of fit within 10% of the best pair of values; once again these are bunched fairly closely around the best fit values with all such pairs generating a value of π_a close to 0.996. The fit of the model deteriorates so sharply when we set the half life of the virus to be only 4 days that we do not show those results. The parameter space in figure 5 shows the best fit parameters (red dot) and also the parameter combinations that generate a root mean squared error within 10% of the best value. This is illustrative of the degree of uncertainty around the best fit values of the two parameters.

However, it is difficult to construct precise confidence regions around the best-fit parameter estimates. In the appendix we discuss robust methods for calculating confidence intervals on these parameter estimates. Using these intervals, our results suggest that one could reject the hypothesis that π_a is below 0.9 at the 10% level of statistical significance, but not at higher levels. In short, one cannot be very sure that the main reason why test cases of those newly infected turned down was because a large fraction of the population had already been infected (very high π_a) rather than a low value of β_L (a very effective lockdown). Nonetheless, as we describe in more detail below, we consistently find the best fit for the data (for both the UK and other countries) is for a very high value of π_a .

5. Allowing π_a to Vary with Testing Capacity

Testing capacity rose significantly in the UK particularly in the last 2 weeks of April 2020. It is possible that along with this rise in testing capacity there was a significant change in the proportion of new infections that were recorded as positive tests; this would invalidate the assumption of a constant π_a . However, the fact that test capacity rose significantly need not mean that recorded cases rise as a proportion of all infections. If test capacity simply rises in line with new cases of the infection, π_a may well be approximately constant.

We assess whether the evidence suggests that π_a varies significantly with testing capacity. We do so by allowing for the value of π_a to systematically vary with the fraction of test results that are positive. The idea is simple: if testing capacity has substantially moved ahead of the rise in new infections then more people who would not have been tested when capacity was used largely on those with clear symptoms will now be tested. As this happens the fraction of tests which are positive is likely to decline along with a fall in the fraction of new infections that are not recorded. We now allow the value of π_a to vary with p_t , the fraction of positive tests to total tests at time t . We allow for a flexible non linear form of this relation. This is given by

$$\ln\left(\frac{\pi_{at}}{1 - \pi_{at}}\right) = b_1 + b_2 \left[\ln\left(\frac{p_t}{1 - p_t}\right) \right], \quad (9)$$

where π_{at} is π_a at time t , that is the fraction of unrecorded new cases to all new infections. b_1 and b_2 are the parameters we estimate to best fit the data. The natural log is denoted \ln . We expect b_2 to be non negative.

1
2
3
4
5
6
7
8 If b_2 is positive then as the proportion of positive tests to total tests falls
9 (which would happen if testing capacity rises much faster than the numbers
10 becoming infected) the fraction of unrecorded new cases of the virus also falls.
11 A special case of (9) is when $b_2 = 0$ and π_{at} is constant, the assumption we
12 made in estimates shown in the previous section. Now we carry out a grid
13 search over values b_1 , b_2 and β_L to minimise the root mean square (RMS)
14 error of the model. We calculate the series p_t as the ratio of daily new cases
15 to daily total tests using data from the Office for National Statistics.
16

17
18 When we estimate these parameters, we find that the value of b_2 that best
19 fitted the data was 0. At first glance, this seems surprising because testing
20 capacity was clearly not constant. But what matters is testing capacity
21 relative to the scale of infection which was likely much less variable than the
22 number of tests undertaken. Nonetheless there are small positive values of b_2
23 which fit the data almost as well as $b_2 = 0$ and which give notably different
24 variations in π_{at} . Figure 6 illustrates the model when we set $b_2 = 0.1$. In
25 this case the the fit of the model deteriorates by just over 25% relative to
26 when $b_2 = 0$ and when π_{at} is constant - though it could not be rejected at
27 standard significance levels. The implied path of π_{at} is far from constant and
28 significantly lower than when we assume it is a constant (figures 3 and 5).
29 This value of b_2 would imply that in April 2020 when π_{at} was around 0.93
30 there were around 14 unmeasured infections for every measured one - a far
31 lower ratio than if π_{at} was at 0.99 or higher when the unrecorded cases would
32 be at least 100 times as great as the recorded cases. At values of b_2 above
33 0.1, the quality of fit of the model deteriorates significantly and such values
34 can be rejected using similar tests to those outlined in appendix 1.
35
36
37
38

39 6. Results for France, US, Italy, Spain and Sweden

40
41 We described above that the model that best fits the UK data was one
42 with constant π_{at} . We estimate this constant π_a model for other countries
43 where up to the end of April 2020 testing had largely been confined to those
44 with symptoms or those at high risk. Data for those testing positive comes
45 from the Johns Hopkins data bank. Dates at which measures to reduce the
46 spread of the virus became severe (the lockdown date) were taken from the
47 Blavatnik Centre at Oxford university which has constructed an index of the
48 severity of measures. We choose the date at which that index rises most
49 sharply to be our starting date for lockdown measures. The dates used are
50
51
52
53
54

outlined in table 1. For the US, the date is problematic because actions vary substantially across states.

Country	Lockdown Date
France	16 March
Spain	10 March
Italy	23 February
Sweden	19 March (partial lockdown)
USA	16 Mar (localised lockdown)

Table 1: Lockdown dates for various countries used for simulations.

The estimated impact of the lockdown measures, and of the speed of the spread of the virus pre-lockdown, are estimated separately for each country. Since lockdown measures differ significantly across countries (mild in Sweden; severe in France) we expect estimates of the difference between β_L and β_0 could be substantial across countries. More generally, country characteristics will likely have influenced the path the virus took. The statistical technique we use allows for this. That is because the parameters that we estimate for the spread of the virus are estimated separately for each country so that the impact of cross country differences that affect the spread of COVID 19 (like population density, age structure, working patterns, living conditions etc) are already allowed for. We chose parameters at the level of the country to best fit the spread of the virus in that country. We also allow the estimated effectiveness of lockdowns (which might also depend on density, travel patterns, labour force participation etc) to vary across countries and it is estimated in a way that best fits the progress of the virus in each country.

Figure 7 show the fit of the model for each country. The most striking result is that the values of π_a that best fit the national data on positive tests for the virus are consistently at very high levels - generally around 0.995 (though lower for the USA). As with the UK results, taken at face value this would mean that there are 200 or so people who have had the virus but whose infection was not recorded for every recorded case.

But what is equally striking, and much less reassuring, is that these best-fit estimates for π_a are much higher than those implied by other methods for assessing the spread of the virus up to April 2020. Some cross country studies

1
2
3
4
5
6
7
8 based on deaths associated with the virus (for example [9]) suggest that
9 only between 5-10% of populations in the countries studied here have been
10 infected. Results from tests for antibodies in samples of the population from
11 a range of European countries suggest a similar proportion of the population
12 has been infected. However, It is possible that serology testing for past
13 COVID 19 infection based on the presence of antibodies are not picking up
14 cases where the infected had very few symptoms and not identifying others
15 who are nonetheless not susceptible to the virus. There is evidence that such
16 tests are not reliable [10].
17
18

19 20 **7. Interpretation, caveats and implications**

21
22 Our estimates suggest that in the UK, and in the other countries whose
23 data we've analysed, there were, in the period up to the end of April, very
24 many unrecorded infections. For the UK, the best fit of the model implies
25 that there were as many as 200 or so unrecorded infections for every recorded
26 case. If that were true, then much of the decline in the spread of the virus
27 would have been due to a degree of immunity having built up in the pop-
28 ulation. Our model also implies that R fell from 2.95 in early February to
29 just under 1 by the end of April. Our estimate of the parameter β_L implies
30 that a significant part of this was due to behavioural changes which reduced
31 β by enough to bring R down to approximately 1.9. But a slightly larger
32 reduction in R came from the build up in immunity implied by the estimate
33 of π_a . At face value, our results suggest that both factors played a major
34 role in bring down the spread of the infection sharply by the end of April.
35
36

37
38 But why should results of the SIR model designed to fit the UK test data
39 (and which also seem to best fit data from Italy, Spain, France, Sweden and
40 the US) suggest a much higher rate of the spread of the virus than antibody
41 test data from countries that have done more widespread (closer to random)
42 testing including the UK and Spain? One answer is purely mechanical: if
43 one wants to fit a model that tracks the data on positive tests it must be
44 one where the number of infections rises very fast early on (a relatively high
45 R_0 and β_0). Is it possible that we have made assumptions which force the
46 model to explain much more of the slowdown in new positive test cases by
47 a fast rise in the immune population (which implies a very large group have
48 had the virus with few symptoms) rather than attribute it to a very effective
49 lockdown? One factor may be significant: we have assumed a 14 day delay
50 between the start of the lockdown and its beginning to affect the rate of new
51
52
53
54

1
2
3
4
5
6
7
8 positive tests for the virus. If that lag were much smaller, more of the turn
9 around in new cases might be attributed to the lockdown and correspondingly
10 less to a rise in mass immunity. But in fact, when we halve the lag between
11 the start of the lockdown and its effect on β we still find that the value of π_a
12 that best fits the data remains very close to 1.
13

14 There is, however, one assumption that does have a significant impact
15 on our results. This is that β is the same for all infected people, whether
16 recorded or not. It is likely that there are more asymptomatic people amongst
17 the unrecorded and it is possible that the spread of the virus for this group is
18 lower than that for the symptomatic. If the rate at which the asymptomatic
19 infect people is significantly lower than for the symptomatic, the best way
20 for our SIR model to explain the UK data is to have a much lower number of
21 unrecorded cases (π_a). If the transmission rate of the unrecorded is one half
22 that of the recorded, but the weighted average of the two keeps the overall β
23 as it was, π_a falls to approximately 0.5. But, the fit of the model deteriorates
24 and the RMS error is around 16% higher than the lowest value obtained in
25 simulations with identical transmission rates.
26
27

28 There is limited evidence that the transmission of the virus is weaker for
29 those with few symptoms [3]. But, it is clear that it matters for modelling
30 the spread of the virus [11]. The influential Imperial College study [8] does
31 assume a lower asymptomatic transmission rate (by 50%). The analysis of
32 Gupta and her team [5], designed to explain the early spread of the virus in
33 the UK, appears to assume a common transmission rate amongst the infected.
34 That study suggested that the asymptomatic were a very high proportion of
35 the infected and that the virus had spread very widely by early March. Our
36 study suggests that estimates of the spread of the virus that best account for
37 the data are sensitive to whether the transmission rate is assumed to be the
38 same for asymptomatic and symptomatic groups.
39
40

41 We have found that when trying to match data in the period up to the end
42 of April on the recorded cases of the virus our model appears to favour high
43 values of π_a (the unrecorded proportion of the total infected people). This
44 is a consistent finding across a number of scenarios where we vary the mean
45 transmission rate, the recovery rate and lockdown measures. It is only when
46 the transmission rate for the unrecorded is much lower than for the (largely
47 symptomatic) recorded cases that the best fitting estimate of π_a is reduced.
48 These two facts lead to two conclusions: First, that previous estimates of
49 π_a near 0.9 [3], or even higher, are consistent with versions of a simple SIR
50 model designed to track results of tests for the virus in the UK and other
51
52
53
54

1
2
3
4
5
6
7
8 countries; but we do not make the stronger claim that the evidence clearly
9 proves such a high value. Second, that reliable modelling of the evolution of
10 the spread of the virus requires accurate measurement of transmission rates
11 for symptomatic and asymptomatic groups and is sensitive to whether these
12 are different.
13

14 In all of the countries whose data we analysed, the best fit to that data
15 suggests that there have been a very large number of unrecorded infected
16 cases for each recorded case. But the data by no means overwhelmingly
17 reject the hypothesis of a value of π_a lower by enough to mean that the main
18 cause of the slowdown (and then reversal) in the arrival of new positively
19 tested cases of the virus in March and April were the measures taken to curb
20 it. But our results suggest that a factor that several other studies simply
21 ignore - namely that the virus had spread fast enough to itself generate
22 substantial immunity by the Spring of 2020 which slowed the spread - was
23 indeed significant. While the results do not show that a degree of immunity
24 clearly was more powerful than lockdowns, the results (which are consistent
25 across countries) do show that it is likely to have been a significant factor.
26
27
28

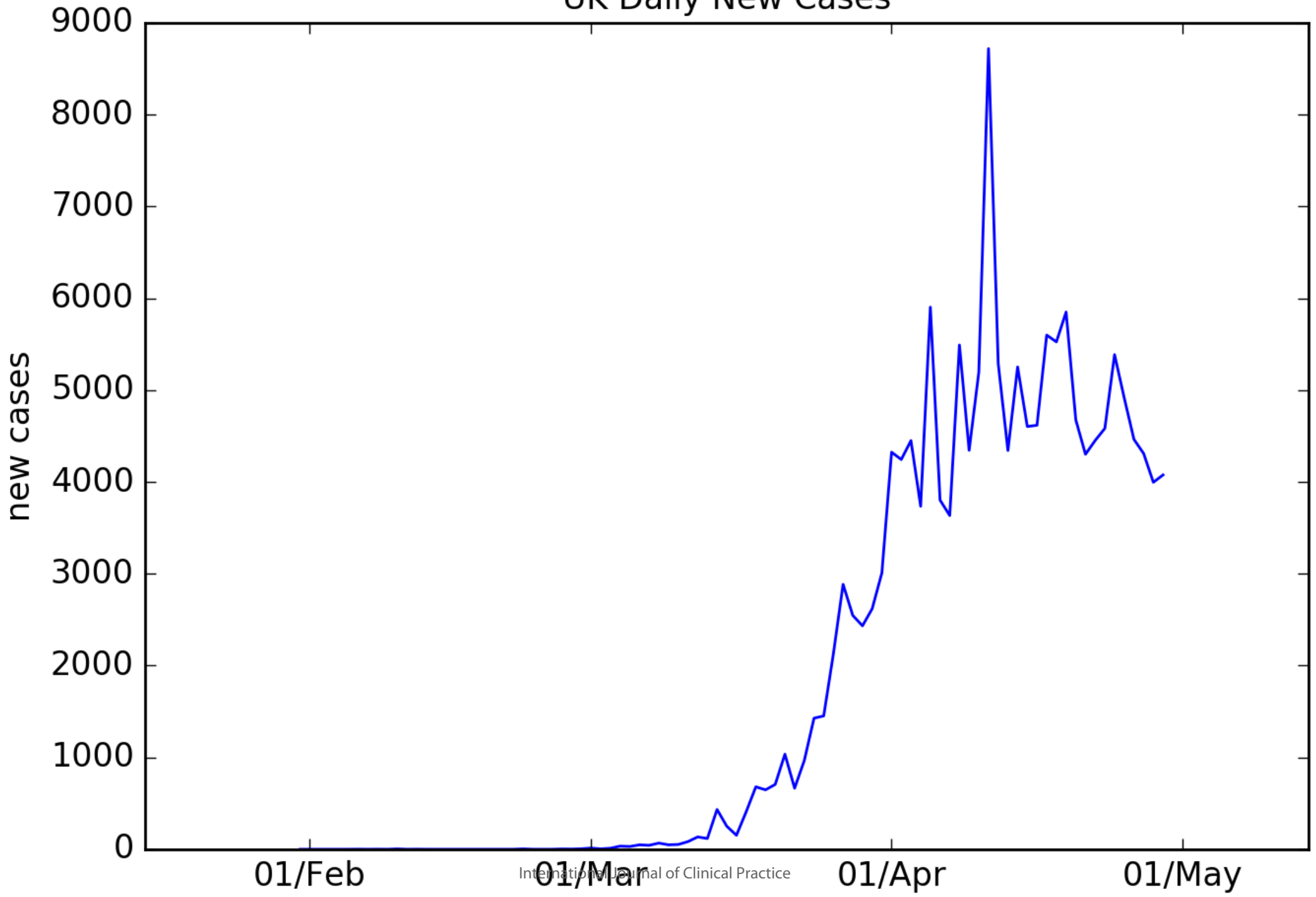
29 References

- 30
31 [1] James H Stock. Data gaps and the policy response to the novel coron-
32 avirus. Technical report, National Bureau of Economic Research, 2020.
33
34 [2] Anthony S. Fauci, H. Clifford Lane, and Robert R. Redfield. Covid-
35 19 — navigating the uncharted. *New England Journal of Medicine*,
36 382(13):1268–1269, 2020. PMID: 32109011.
37
38 [3] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang,
39 and Jeffrey Shaman. Substantial undocumented infection facilitates the
40 rapid dissemination of novel coronavirus (sars-cov2). *Science*, 2020.
41
42 [4] Charles F Manski and Francesca Molinari. Estimating the covid-19
43 infection rate: Anatomy of an inference problem. Technical report,
44 National Bureau of Economic Research, 2020.
45
46 [5] Jose Lourenco, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig
47 Thompson, Peter Simmonds, Paul Klenerman, and Sunetra Gupta. Fun-
48 damental principles of epidemic spread highlight the immediate need for
49 large-scale serological surveys to assess the stage of the sars-cov-2 epi-
50 demic. *medRxiv*, 2020.
51
52
53

- 1
2
3
4
5
6
7
8 [6] Mike Stedman, Mark Davies, Mark Lunt, Arpana Verma, Simon G Anderson, and Adrian H Heald. A phased approach to unlocking during the covid-19 pandemic—lessons from trend analysis. *International Journal of Clinical Practice*, page e13528, 2020.
- 9
10
11
12
13
14 [7] Gustav W Delius, Benedict J Powell, Martin A Bees, George WA Constable, Niall J MacKay, and Jonathan W Pitchford. More prevalent, less deadly? bayesian inference of the covid19 infection fatality ratio from mortality data. *medRxiv*, 2020.
- 15
16
17
18
19 [8] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.
- 20
21
22
23
24
25
26 [9] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, pages 1–5, 2020.
- 27
28
29
30
31
32
33 [10] Mayara Lisboa Bastos, Gamuchirai Tavaziva, Syed Kunal Abidi, Jonathon R Campbell, Louis-Patrick Haraoui, James C Johnston, Zhiyi Lan, Stephanie Law, Emily MacLean, Anete Trajman, et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *British Medical Journal*, 370, 2020.
- 34
35
36
37
38
39
40 [11] Sang Woo Park, Daniel M Cornforth, Jonathan Dushoff, and Joshua S Weitz. The time scale of asymptomatic transmission affects estimates of epidemic potential in the covid-19 outbreak. *medRxiv*, 2020.
- 41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

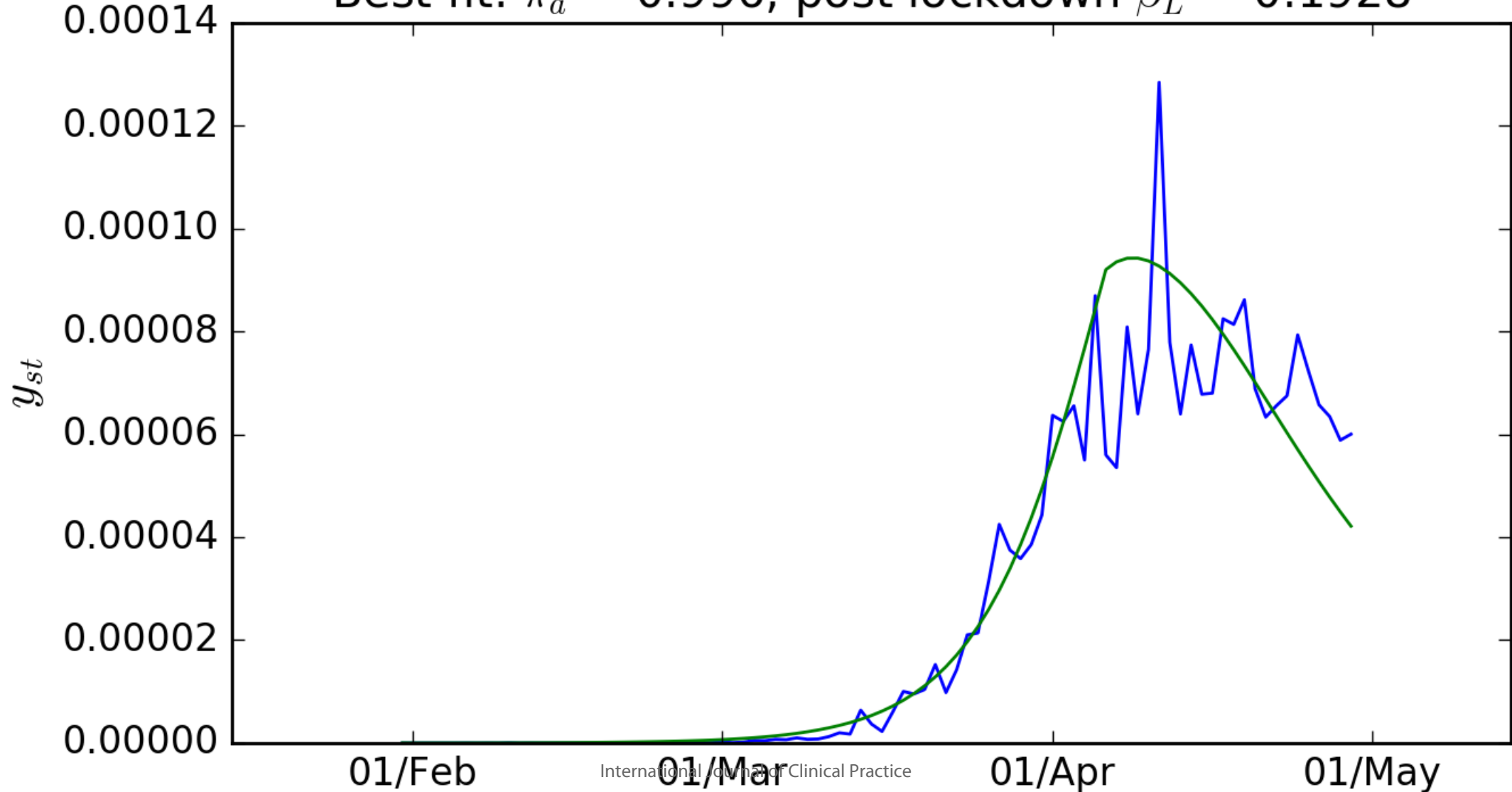
UK Daily New Cases

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

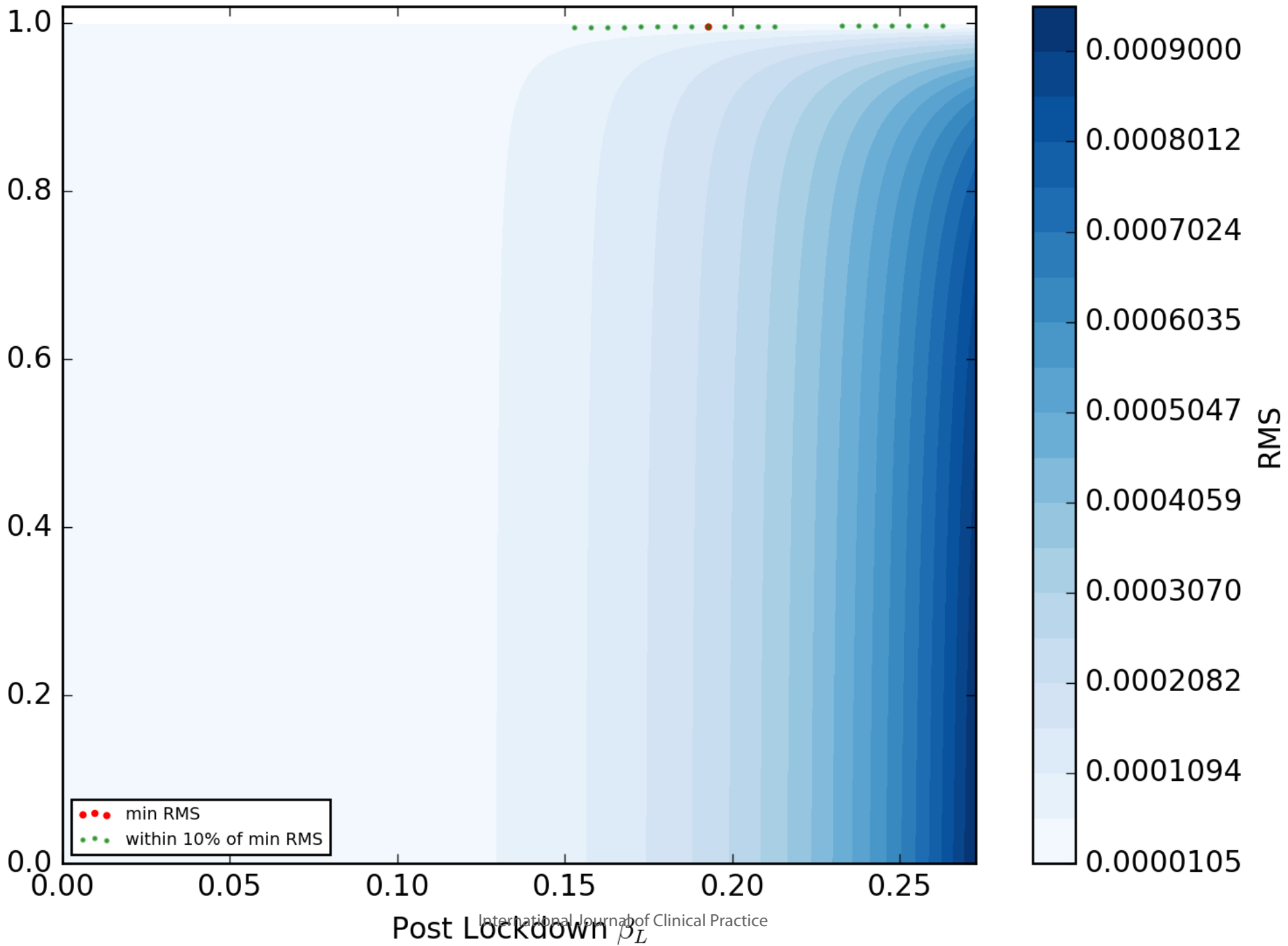


Best fit: $\pi_a = 0.996$, post lockdown $\beta_L = 0.1928$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

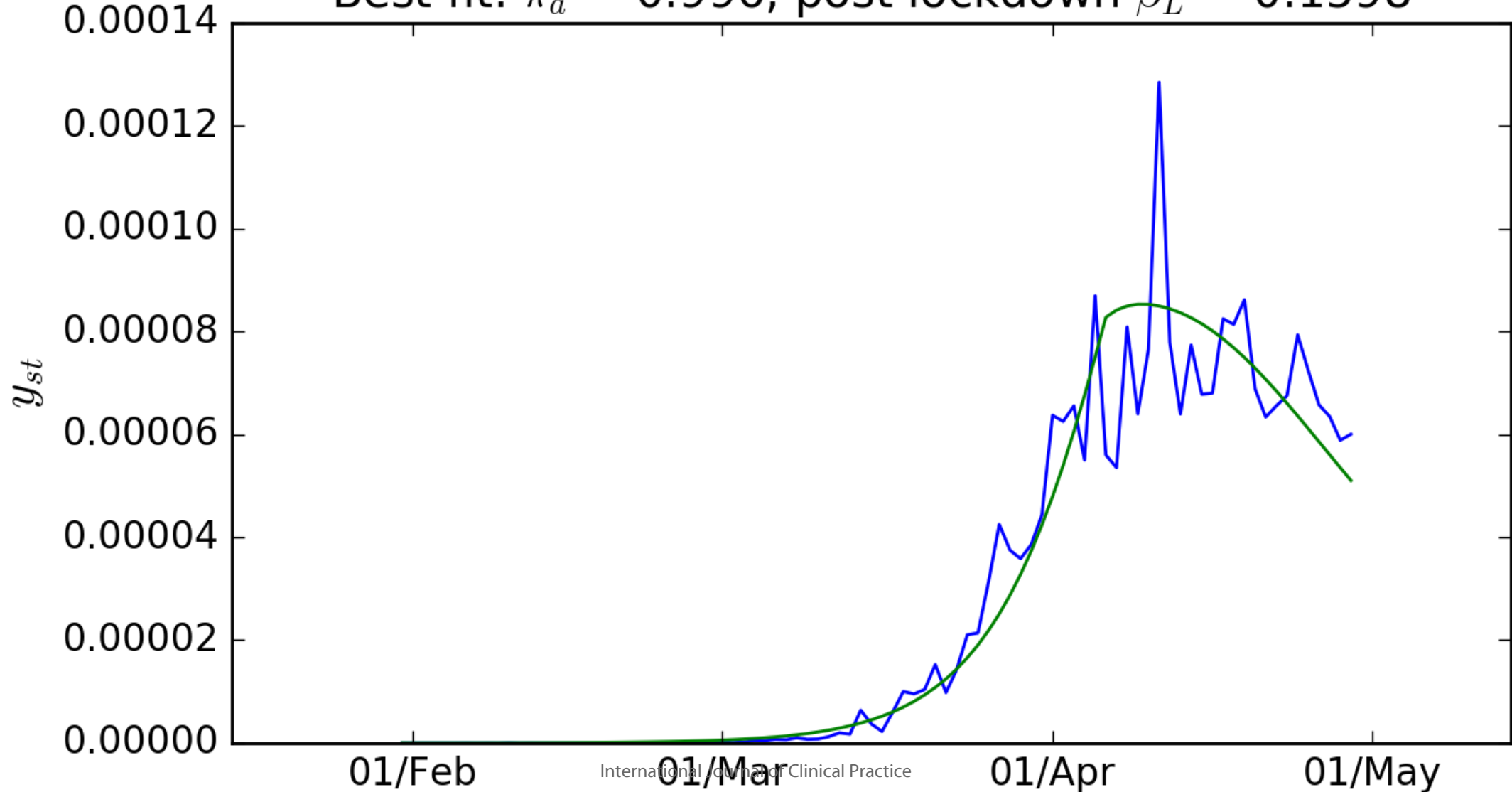


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46



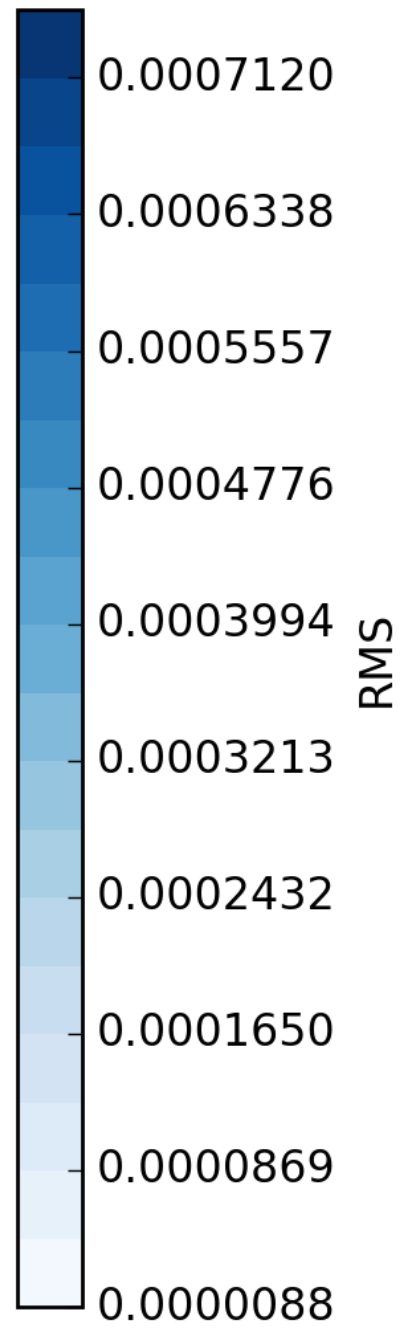
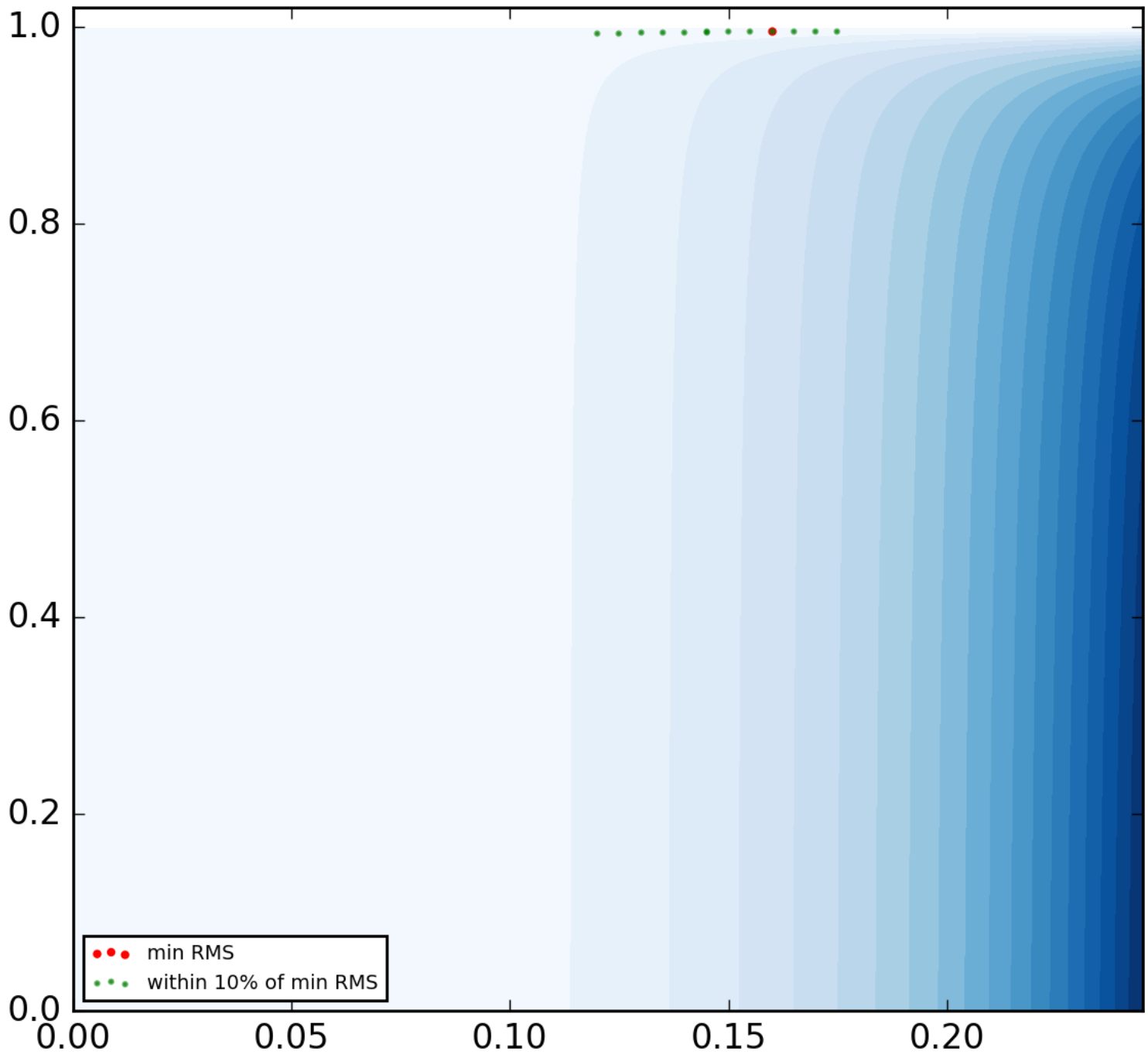
Best fit: $\pi_a = 0.996$, post lockdown $\beta_L = 0.1598$

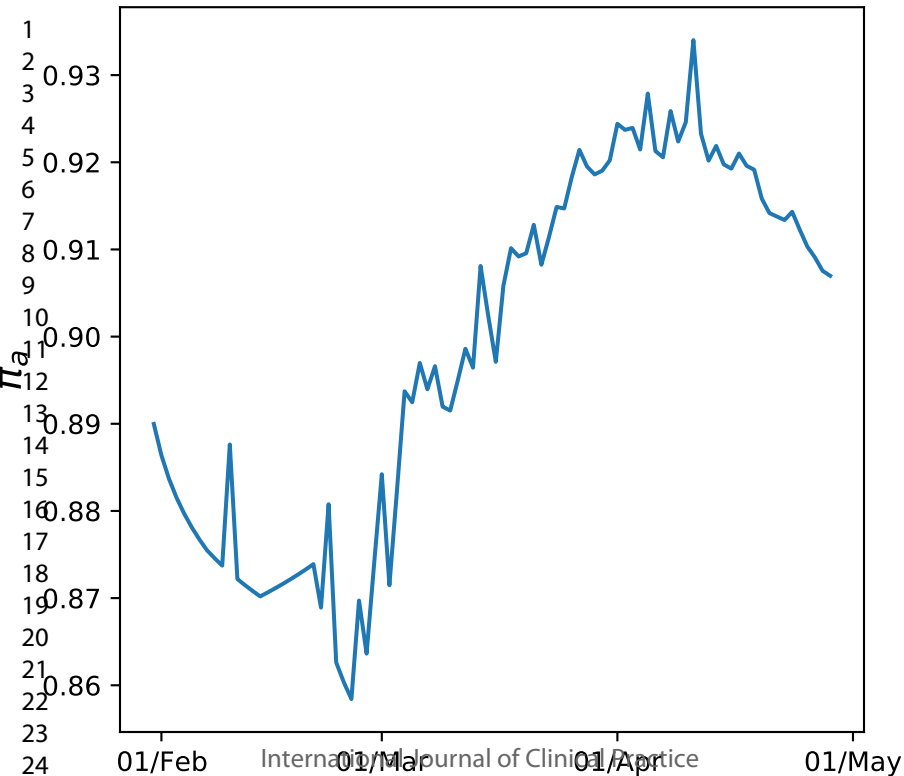
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36



$$R_0 = 2.95$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46





23
24
25
26

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

