An efficient benchmark for COVID-19 pandemic testing effectiveness

Dimitris Nikoloudis[1], Dimitrios Kountouras[1] and Asimina Hiona[1]

[1] Center for Preventive Medicine & Longevity, Bioiatriki Healthcare Group, 11525 Athens, Greece

Abstract

Testing for COVID-19 is an important tool that health administrations dispose to adequately monitor and respond to the pandemic, but it is still unclear at which point the number and strategies of testing become effective for these purposes. The percentage of tests that return a positive result is a metric currently used both as a benchmark of testing adequacy and for assessing the viral spread. However, since the former is a prerequisite for the latter, the interpretation is often conflicting, especially during times of testing scaling-up, or during phases of increasing viral spread. We propose as a benchmark for COVID-19 testing effectiveness a simple metric that creates a link between the cases detected and tests performed, with specific observable outcomes that are actively being monitored in most countries, such as the number of new Intensive Care Unit (ICU) admissions and the number of deaths in the community. This new metric, named 'Severity Detection Rate', or SDR, represents the current number of daily needs for new ICU admissions, per 100 cases detected (t-i) days ago, per 10,000 tests performed (t-i) days ago. Based on the announced COVID-19 monitoring data in Greece from May 2020 until end of January 2021, we show that beyond a threshold of daily testing number, SDR reaches a plateau of weak variability that begins to reflect testing adequacy. Because of this stabilization, it was possible to predict with great accuracy the daily needs for new ICU admissions, 12 days ahead of each testing data point, over a period of 6 months that included the second wave of the pandemic in the country, with Pearson $r = 0.99$ ($p = 10^{-180}$), RMSE = 4,34. We suggest the further study of the metric with data from more countries in order to confirm the proposed functionality and utility.

Introduction

Although no country knows at any point the true total number of COVID-19 cases, it is crucial for public health administrations to be confident that the daily testing performed is stably representative of that number. An effective testing provides health professionals and officials with a clear picture of SARS-CoV-2 spread within community, as well as of the dynamics of COVID-19 pathology, and guides them for the prompt and adequate interventions towards containment of the pandemic, at local and national level.

The percentage of tests that return a positive result, also known as "positivity rate", is an important outcome of testing which is used both as a benchmark for testing adequacy and as a metric for assessing the current spread of the virus (Dowdy & D'Souza, 2021). However, this dual usage presents an inherent drawback in entrusting the metric in any one of the two possible ways: is a high positivity rate due to a high number of infected individuals, or due to a low number of tests performed? A rule of thumb says that a positivity rate of 5% is too high and WHO suggested that the positivity rate should rest below that threshold for a length of at least two weeks before officials decide to progressively reopen professional and social activities (Dowdy & D'Souza, 2021). Another evidence-based perception suggests that positivity rate must remain below 3% to be sure that surveillance is broad and accurate enough (Siddarth et al., 2020). However, these rules may only cover either the virus spread surveillance criterion, or that of testing adequacy, but not both. Indeed, officials often respond to a high positivity rate both with an increase in testing and with measures to restrict virus transmission, such as social distancing and soft or hard lockdowns. But by doing so, it is expectedly hard to timely assess the true rate of the virus spreading out, or being contained, as the new higher levels of testing must be stabilized for a length of time before allowing again to reliably follow the pandemic dynamics. In such a scenario, if health officials rely only on positivity rate metric, the timing of response would be in lag and thus almost invariably suboptimal.

Fundamentally, a metric that would serve as a benchmark for the effectiveness of COVID-19 testing should not concurrently be used for assessing the evolution of the pandemic, as the former is a prerequisite for the latter and therefore the interpretation would be conflicting; indeed, the health administrations of a country should be confident that a sufficient number of tests is performed, in order to effectively track the virus spread. However, if such a metric also implemented measurable outcomes

of the pandemic in the community (e.g., number of deaths, number of ICU admissions, etc.), they could introduce by their more factual nature a link between expectation and actuality, since the outcomes of COVID-19 are inherently tied to the virus' pathogenesis. Therefore, such a link could, in theory, introduce a benchmarkable step of convergence towards a soft cap (threshold) that would in turn reflect testing adequacy, *e.g.*, usually a maximized or minimized value, or a state of minimized variation. In this report we present an easy to implement metric that we developed while independently monitoring and analyzing COVID-19 pandemic evolution in Greece, which considers outcomes that are already monitored in most countries, such as the number of human losses, the number of COVID-19 patients in ICU (Intensive Care Units) and the number of patients that are being discharged from ICU. We show that this metric displays remarkable output stability when a certain threshold of daily testing is reached, which to our view clearly reflects testing adequacy. Furthermore, we validated its benchmarking efficiency by prospectively predicting, not only with high accuracy but also great precision, the total daily needs for new ICU admissions, roughly two weeks in advance, over a period of 6 months.

Methods

The national monitoring data for the evolution of COVID-19 pandemic in Greece were retrieved from the Hellenic National Public Health Organization (NPHO, 2021) and Greek Government's official daily announcements (Greek Government's official community on Viber network, 2021). Specifically, the daily official announcements included the following parameters: (a) number of new COVID-19 cases detected, (b) number of deaths due to COVID-19, (c) total number of COVID-19 ICU patients, (d) total number of COVID-19 patients discharged from ICU, (e) total number of SARS-CoV-2 PCR tests performed, and (f) total number of SARS-CoV-2 rapid antigen tests performed.

Based on the available data, we defined as number U, *the daily needs for new COVID-19 ICU admissions*:

$$U = (x_0 - x_{t-1}) + d + e \qquad (1)$$

where:

- Today's deaths due to COVID-19: d
- Today's number of COVID-19 patients discharged from ICU: e

- Today's total number of COVID-19 ICU patients: x_0
- Yesterday's total number of COVID-19 ICU patients: x_{t-1}

This number U represents the actual daily new COVID-19 ICU admissions, plus those patients who died in the community (not in ICU), whom we theorize to have required ICU admission, hence the definition of *the daily need for new COVID-19 ICU admissions*.

Next, we defined as Severity Detection Rate with a time lag (t-i) (SDR_i), a metric that represents the percentage of patients that require ICU admission, per new cases, detected (t-i) days ago, per 10,000 tests, performed (t-i) days ago:

$$SDR_i = (U*100/c_{t-i})/(n_{t-i}/10{,}000) \Rightarrow SDR_i = U*1{,}000{,}000/(c_{t-i}* n_{t-i}) \quad (2)$$

where:

- Today's rolling 7-day average of new daily needs for COVID-19 ICU: U
- Rolling 7-day average of detected COVID-19 cases, (t-i) days ago: c_{t-i}
- Rolling 7-day average of total number of COVID-19 tests, (t-i) days ago: n_{t-i}

Finally, for the prospective prediction of the rolling 7-day average daily needs for ICU admissions, we inversed equation (2) to forecast the number U_{t+i}, (t+i) days ahead, as follows:

$$U_{t+i} = (SDR_{MEDIAN}*c_0* n_0)/1{,}000{,}000 \quad (3)$$

where:

- Today's rolling 7-day average of detected COVID-19 cases: c_0
- Today's rolling 7-day average of total number of COVID-19 tests: n_0
- Median Severity Detection Rate of a specified period: SDR_{MEDIAN}

For a more accurate prediction, we also added a correction factor to equation (3), which involves the 7-day percent change of the rolling 7-day average number of daily tests performed:

$$U'_{t+i} = U_{t+I} + U_{t+i}*((n_{t-7} - n_0)/n_0) \quad (4)$$

The correction factor accounts for a portion of the variability of SDR and proved beneficial to forecast, in every case.

The dataset was locked on 31st of January 2021.

Results

For exploration, the lag of Severity Detection Ratio was initially set to 14 days, which means that the current day's critical outcomes of COVID-19 (*i.e.*, ICU admission, or death in the community) were attributed to Covid-19 cases detected 14 days ago. The daily evolution of SDR_{14}, from the 11th of May 2020 onwards, was traced for observation versus the positivity rate on the same days, as well as versus the corresponding number of testing samples (*i.e.*, transposed by 14 days for alignment to SDR) (Figure 1). The SDR metric shows a remarkable stabilization past the time mark around 20/8/2020, compared to the previous segment. From that point forward the positivity rate and the testing rate continue to fluctuate independently and considerably, however without considerably perturbing SDR stabilization.

The rate of daily testing in Greece has been scaled up significantly on two occasions, approximately (a) on 29/7/2020, and (b) on 27/10/2020. As the new testing levels were preserved after each scale-up, it is possible to define three distinct periods of testing intensity thus far, during the COVID-19 pandemic in Greece. We characterized the SDR number and the rates of testing for each one of the following time intervals: (i) 1/5/2020 - 28/7/2020, (ii) 29/7/2020 – 16/10/2020, and (iii) 17/10/2020 – 31/1/2021 (Table 1). Tripling the average daily rate of testing (from 4K to 12K) in the second (ii) interval brought an almost equivalent decrease in the CV (Coefficient of Variation) of SDR (0.97/0.36 ~ 2.7), and a 7-fold lower average value of SDR (20.1% / 2.7% ~ 7.4). Further doubling of the previous average daily number of tests (from 12K to 24K) in the third (iii) interval brought again an equivalent decrease in the CV of SDR (0.36/0.19 ~ 1.9), although the average value of SDR was now only moderately diminished by approximately 30% (2.7% / 2.1% ~ 1.29), indicating a tendency towards stabilization of the SDR value and continuous reduction of the standard deviation. The rolling 7-day averages of the two measures display an exceptionally strong correlation employing power regression (Spearman $r = -0.86$, $p = 10^{-77}$, $N = 262$) and suggest that beyond a threshold of daily tests performed, SDR becomes significantly stabilized (Figure 2A). The same effect is observed for a SDR lag between 7 to 21 days, with Spearman r ranging between -0.76 and -0.86 (max r value at $i = 14$).

Next, we correlated the number U with the product of (c_{t-1}* n_{t-i}) with a lag of 7 to 21 days (*i.e.*, the numerator and denominator of SDR), in order to pinpoint the optimal lag for tracking the SDR (1/5/2020 to 31/1/2021); in other words, to find which days' announced detected cases and performed tests best represent the very severe outcomes that are observed in the future. The best fitting linear regression was obtained for a lag 12 days, with Pearson r = 0.99 ($p = 10^{-239}$) (Figure 2B). Subsequently, in order to take advantage of the previous finding (SDR stabilization after a threshold of daily tests), we applied equation (4) to forecast the rolling 7-day average daily needs for new ICU admissions, 12 days ahead of each data point of daily announced cases and tests. As a representative value of Severity Detection Rate (SDR_{MEDIAN}), we selected the value of 2.04%, which is equal to the median of SDR in the 3rd examined period (17/10/2020 – 31/1/2021) (Table 1). Both the median and average values of SDR were very close to each other (2.04 vs 2.12, respectively), but the use of the median resulted in the lowest root mean square error (RMSE). We compared the RMSE and coefficients of correlation of predicted versus observed U, for a range of (t-i) between 8 and 21 days and, as expected, the best results were obtained for i=12. The observed values of number U since 1st of August 2020 correlated very strongly with values that were predicted 12 days ahead of time using equation (4), with Pearson r = 0.99 ($p = 10^{-180}$), RMSE = 4,34 (Observed U[max]=125, U[average]=37), indicating an almost complete agreement between predicted and observed values. When only considering the 3rd examined period of our study, which comprises specifically the 2nd wave of the pandemic in the country, the correlation remained equally strong, with Pearson r = 0.99 ($p = 10^{-97}$), RMSE = 5.46 (Observed U[max]=125, U[average]=60) (Figure 3A). Finally, we attempted to obtain an alternative forecast by applying the linear regression equation from the correlation of number U with the product of (c_{t-1}* n_{t-i}) with a lag of 12 days, as can be seen in Figure 2B. The same correction factor that was previously used in equation (4), was also used in this case. This alternative forecast proved practically the same (i.e., correlated R^2 ~ 1) to the forecast presented previously.

Discussion

We have shown that beyond a threshold of daily tests performed, the percentage of daily needs for new COVID-19 ICU admissions, per new cases detected t-i days ago, per 10,000 tests performed t-i days ago, reaches a plateau that displays very weak variation. This threshold appears roughly around the

10,000 daily samples mark in Greece, a country of approximately 11 million people, but this number is expected to vary greatly from country to country depending on total population, rural density, societal particularities, population's immune profile, and sampling strategies. Reaching that threshold should not mean that there is no need for further increase in the number of daily tests, as it is strongly suggestive that the more tests a country performs, the more informative the results are about the actual viral spread in community, and consequently health administrations are in better position to respond accordingly. In terms of Severity Detection Rate, specifically, more daily tests appear to further decrease its variation (Table I). The weaker its variation, the more accurately we can predict the number of daily needs for new ICU admissions, t+i days in advance. As a direct consequence of this potential of predictability, when SDR establishes a plateau, we consider that the bulk of daily tests is returning a set of positive cases that is stably representative of the current spread of the virus. Therefore, the SDR metric constitutes a benchmark of testing effectiveness. The metric is possibly efficient at a local level as well, if cases that require delocalization (*e.g.*, due to lack of available ICU, locally) are effectively tracked and taken into account. As the full segmentation of the necessary data was not available at local level, it was not possible to assess the effects of viral spread uniformity across the country and, more specifically, the metric's behavior due to potentially disproportionate testing intensities, locally, e.g., higher number of tests in districts with small viral load, and relatively lower numbers of daily tests in districts with bigger true viral load.

We called this new metric Severity Detection Rate, as its representation of the percentage of very severe COVID-19 outcomes is modulated by the number of tests performed. It is essentially a standardization of the very severe cases ratio over the infected individuals, with the rate of daily testing. In other words, SDR becomes representative of the proportion of people that need ICU out of the total cases once a sufficient threshold of daily testing rate (hence 'detection rate') is achieved. The metric's median value is expected to decrease monotonically and with decreasing variation as daily tests increase, unless the virus's lethality becomes enhanced with time, *e.g.*, due to the prevalence of a new more pathogenic variant, in which case the SDR's median will increase disproportionately and beyond its expected variability. Accordingly, SDR's median value may decrease more rapidly for a number of reasons, *e.g.*, gradual containment of the virus, or immunization of the population, thanks to an efficient vaccination program, or the improvement of therapeutic protocols that reduce the number of very severe cases, or even the significant decrease of the age average of infected individuals, due to the efficient protection of the more elderly. If, in theory, the total number of tests became equal to the

entire population of a country, then this metric would practically represent the true percentage of critical patients per infected individuals.

The positivity rate metric remains complementary but is less reliable on its own in tracking the virus spread, especially in cases where the number of daily tests varies strongly and concurrently with the number of actual total positive cases. In comparison, number U, representing the daily needs for new ICU admissions, seems a more reliable metric of the pandemic evolution, as it depends exclusively on reported observed outcomes. Of course, it can be argued that the observed number U is tied to the testing results that took place 12 days before, and therefore offers a view of the pandemic evolution with an equivalent lag. However, as soon as SDR stability is established, the number U(t+i) can be reliably forcasted 12 days ahead of each data point, therefore nullifying the related lag. Importantly, it provides an almost 2-week outlook on the pandemic evolution and the upcoming needs for new ICU admissions (Figure 3B; observed U=23 on 31/1/21, forecast U=37 on 12/2/21).

The forecasting of number U should be viewed as complementary to the currently employed epidemiological tools, i.e., the positivity rate, the efficient contact tracing for determination of the basic reproduction number R_0 (McDonald, 1952), and the wastewater-based surveillance (Polo et al., 2020). The Severity Detection Rate metric introduces the goal to minimize its variation, by means of sufficient number of daily tests and adequate sampling strategy. Once this goal is achieved, then the accurate forecasting of daily needs for new ICU admissions becomes possible. We strongly believe that the explicit tracking of this new metric enhances the visibility of viral spread and dynamics and may procure an accurate outlook of the upcoming needs for ICU admissions, which should serve as an early warning system for covid-19 health establishments and resources. We therefore suggest the further study of Severity Detection Rate with data from more countries, as well as at a more local level wherever possible, in order to confirm the proposed functionality and utility of the metric.

Table 1. Characterization of the Severity Detection Rate and the number of daily tests for each one of the three time-intervals of distinct testing levels in Greece.

intervals	1/5/2020 - 28/7/2020		29/7/2020 – 16/10/2020		17/10/2020 - 31/1/2021	
	Severity Detection Rate	**Daily tests**	**Severity Detection Rate**	**Daily tests**	**Severity Detection Rate**	**Daily tests**
max	92.0%	7309	7.0%	20310	3.2%	31602
average	20.1%	4051	2.7%	12861	2.1%	24039
median	14.3%	3992	2.6%	12453	2.0%	24361
min	0.1%	1400	1.1%	9706	1.3%	17315
sd	19.6%	1260	1.0%	2439	0.4%	3198
cv	0.97	0.31	0.36	0.19	0.19	0.13

Figure 1. Comparison of trendlines of Severity Detection Rate, Positivity rate, and number of Daily Tests, in the period from 7/5/2020 to 31/1/2021. All numbers were calculated from rolling 7-day averages and were normalized by their maximum value in the examined period.

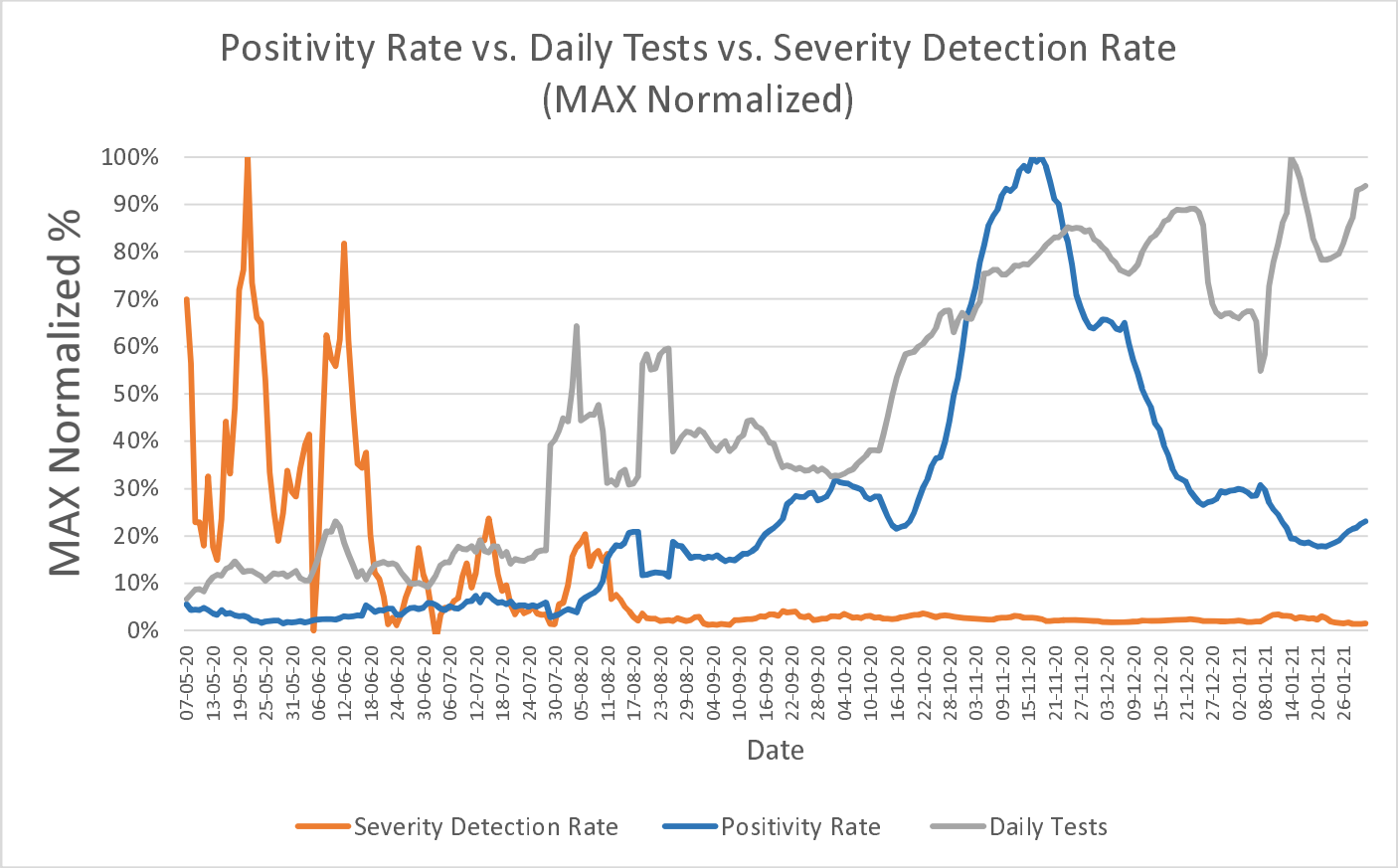


Figure 2A. Correlation between the rolling 7-day averages of number of daily tests and Severity Detection Rate, with Spearman r = -0.86, p = 10^{-77}, N = 262. Numbers of daily tests derived from the period from 15/5/2020 to 31/1/2021.

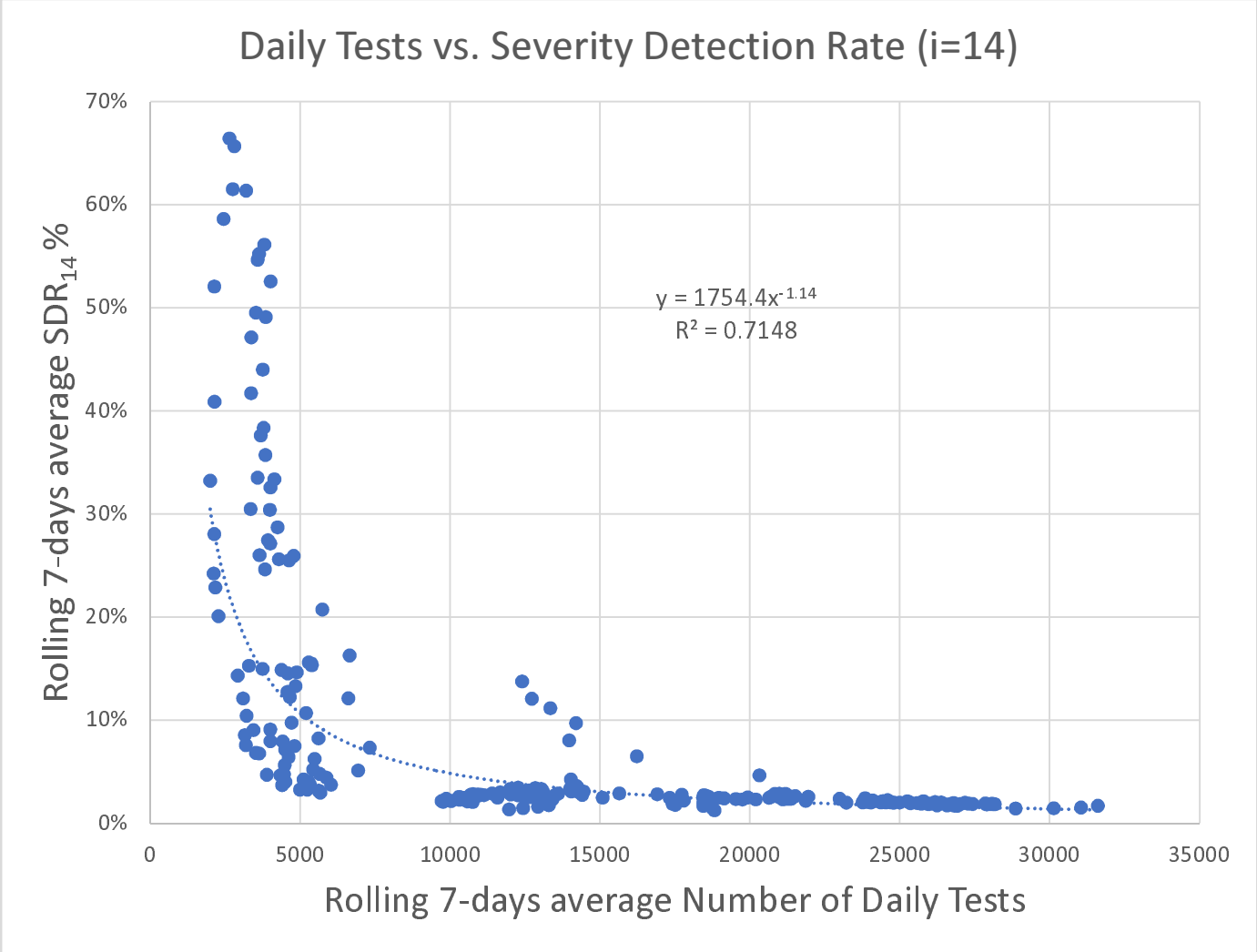


Figure 2B. Correlation of the numerator and denominator of SDR, *i.e.*, number U versus the product of ($cases_{t-12}$* $tests_{t-12}$), with a lag of 12 days. with Pearson r = 0.99, p = 10^{-239}. Numbers of daily tests and detected cases derived from the period from 1/5/2020 to 31/1/2021.

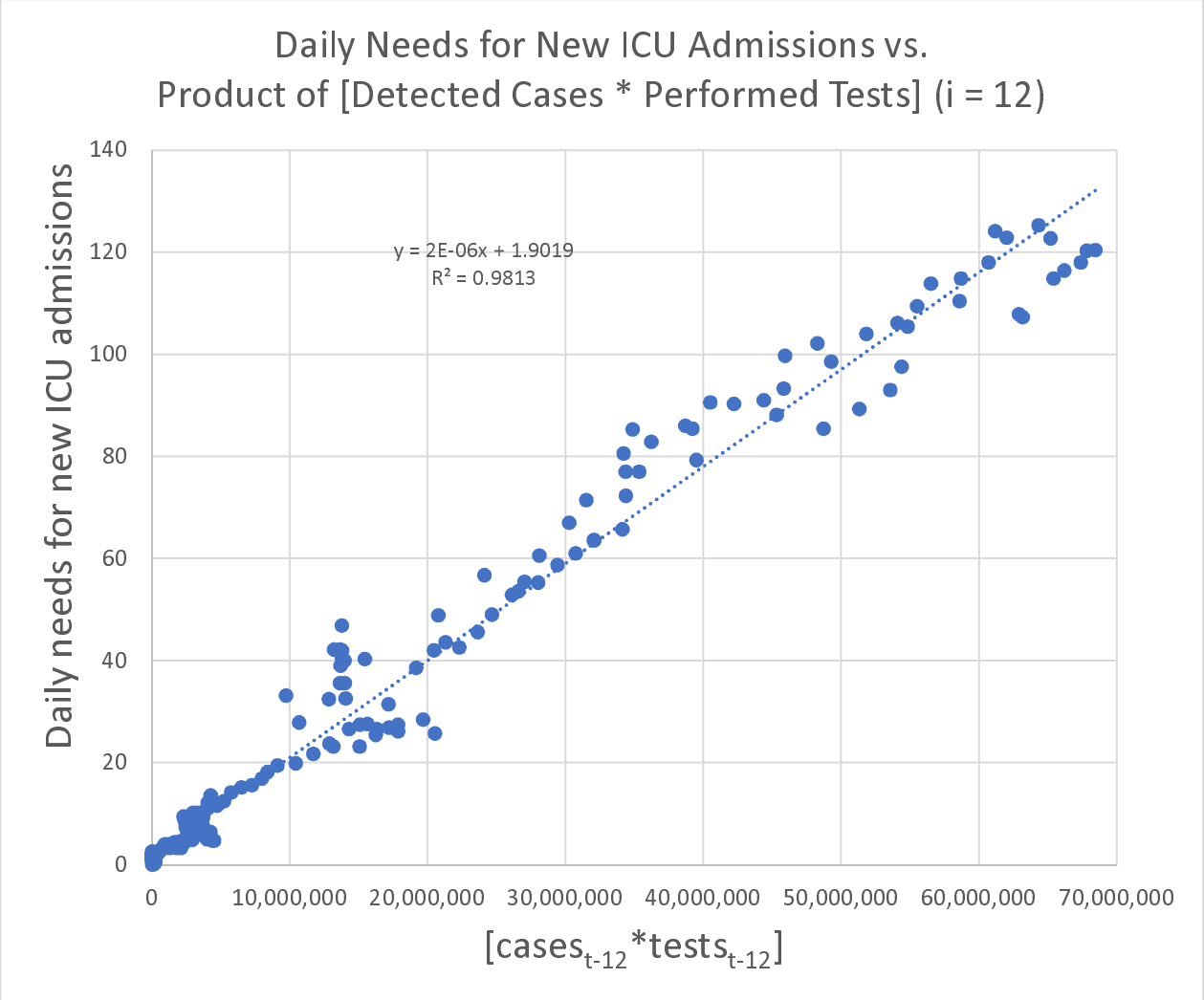


Figure 3A. Correlation between observed and predicted daily needs for new ICU admissions for the period between 17/10/2020 and 31/1/2021, with Pearson r = 0.99 ($p = 10^{-97}$), RMSE = 5.46.

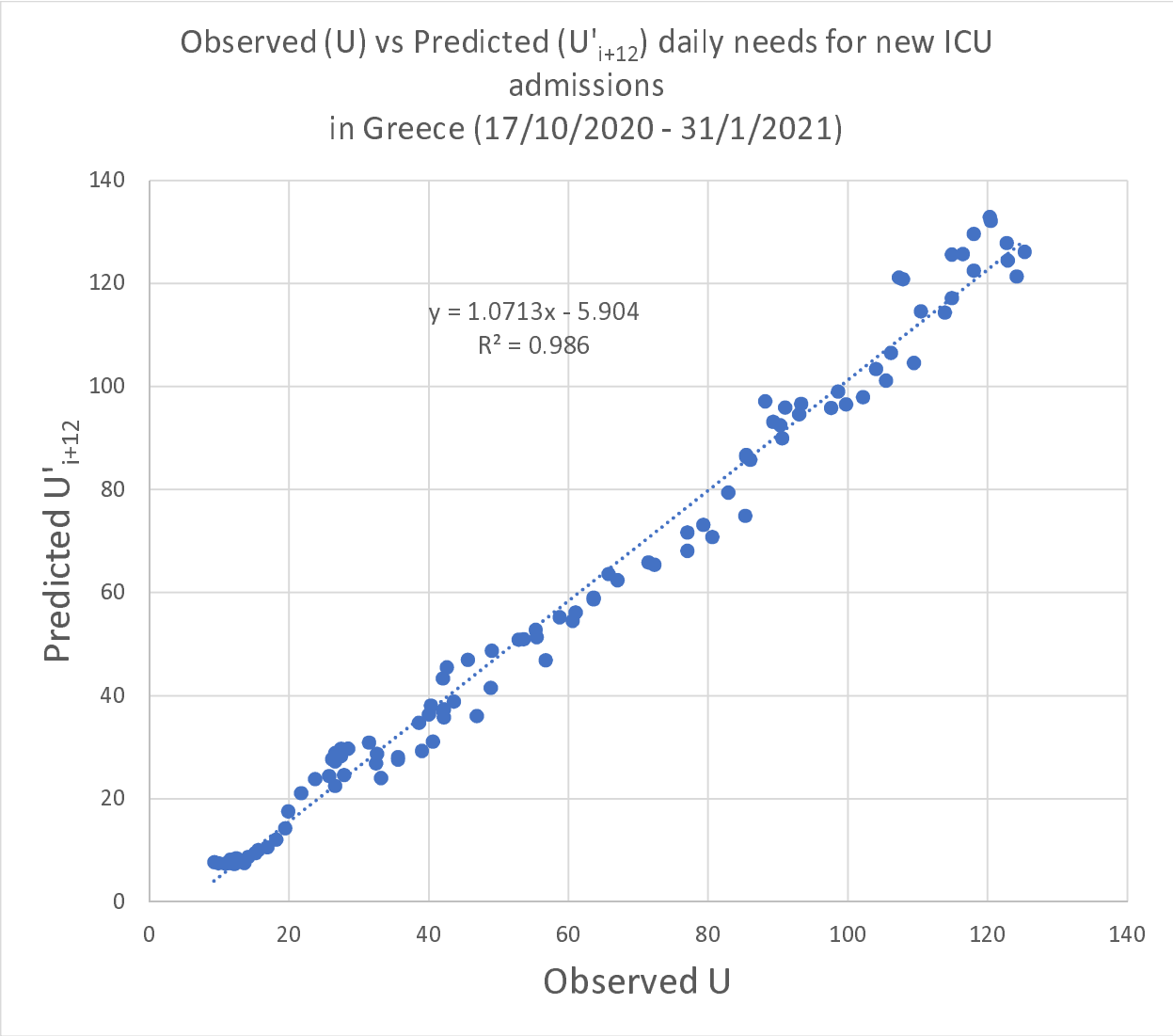
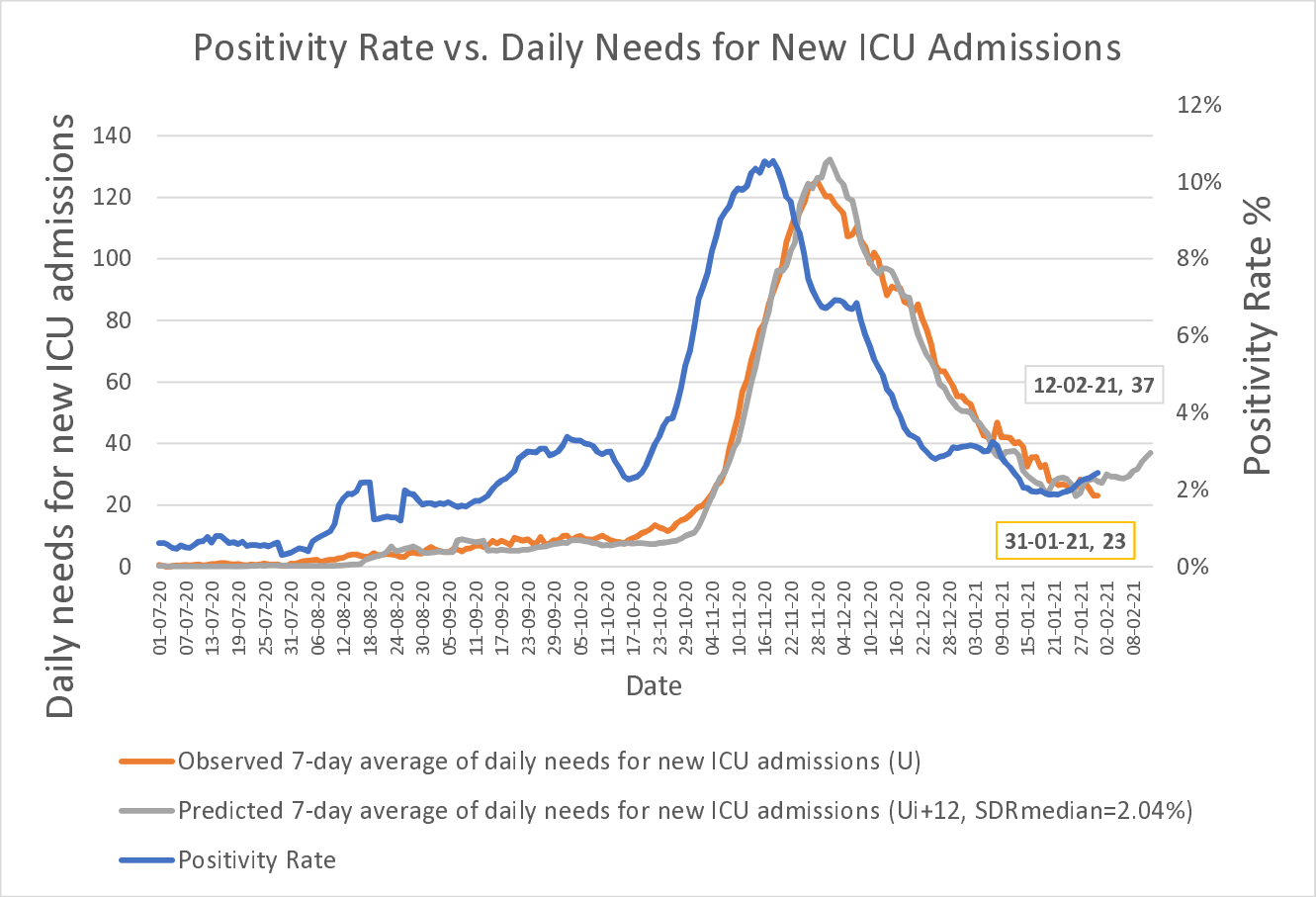


Figure 3B. Comparison of Positivity Rate and the number of daily needs for new ICU admissions (Observed and Predicted), for the period between 1/7/2020 and 31/1/2021.



References

Dowdy D, D'Souza G. COVID-19 Testing: Understanding the "Percent Positive". Johns Hopkins Bloomberg School of Public Health. Accessed January 31, 2021. Available from: https://www.jhsph.edu/covid-19/articles/covid-19-testing-understanding-the-percent-positive.html

Greek Government's official community on Viber network. Official COVID-19 update. Accessed January 31, 2021. Available from:

https://invite.viber.com/?g2=AQAVNiDVlfjrIEtaYLf1s2sUzRrpfLVlfLVg4J8wkdNKMUSnUcQWJxnXH0Os1heH&lang=el

Siddarth D, Katz R, , Graeden E, Analytics T, Allen D, Tsai T. 2020. Evidence Roundup: Why positive test rates need to fall below 3%. Harvard Global Health Institute. Accessed January 31, 2021. Available from: https://globalhealth.harvard.edu/evidence-roundup-why-positive-test-rates-need-to-fall-below-3

Hellenic National Public Health Organization. Daily Reports COVID-19. Accessed January 31, 2021. Available from: https://eody.gov.gr/epidimiologika-statistika-dedomena/ektheseis-covid-19/

Macdonald, G. (September 1952). "The analysis of equilibrium in malaria". Tropical Diseases Bulletin. 49 (9): 813–829. ISSN 0041-3240. PMID 12995455.

Polo D, Quintela-Baluja M, Corbishley A, Davey LJ, Andrew CS, David WG, Jesús LR. 2020. Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. Water Research, 186, 116404. doi: 10.1016/j.watres.2020.116404.