

Mauro Gaspari*

The impact of test positivity on surveillance with asymptomatic carriers

<https://doi.org/10.1515/em-2022-0125>

Received July 10, 2022; accepted July 7, 2023; published online July 26, 2023

Abstract

Objectives: Recent studies show that Test Positivity Rate (TPR) gains a better correlation than incidence with the number of hospitalized patients in COVID-19 pandemic. Nevertheless, epidemiologists remain sceptical concerning the widespread use of this metric for surveillance, and indicators based on known cases like incidence rate are still preferred despite the large number of asymptomatic carriers, which remain unknown. Our aim is to compare TPR and incidence rate, to determine which of the two has the best characteristics to predict the trend of hospitalized patients in the COVID-19 pandemic.

Methods: We perform a retrospective study considering 60 outbreak cases, using global and local data from Italy in different waves of the pandemic, in order to detect peaks in TPR time series, and peaks in incidence rate, finding which of the two indicators has the best ability to anticipate peaks in patients admitted in hospitals.

Results: On average, the best TPR-based approach anticipates the incidence rate of about 4.6 days (95 % CI 2.8, 6.4), more precisely the average distance between TPR peaks and hospitalized peaks is 17.6 days (95 % CI 15.0, 20.4) with respect to 13.0 days (95 % CI 10.4, 15.8) obtained for incidence. Moreover, the average difference between TPR and incidence rate increased to more than 6 days in the Delta outbreak during summer 2021, where presumably the percentage of asymptomatic carriers was larger.

Conclusions: We conclude that TPR should be used as the primary indicator to enable early intervention, and for predicting hospital admissions in infectious diseases with asymptomatic carriers.

Keywords: asymptomatic carriers; COVID-19; incidence; infectious diseases; infectious diseases surveillance; test positivity rate

Introduction

Test Positivity Rate (TPR), e.g., the percentage of positive tests over total tests, is one of the metrics used for public health surveillance in infectious diseases. Typical applications include estimating the prevalence of diseases in the population, see Boyce et al. (2016) for malaria disease and Chiu and Ndeffo-Mbah (2021) for COVID-19, or establishing levels of community transmission from sentinel sites in the COVID-19 pandemic, see World Health Organization (2021). Over the past two years, thanks to the large amount of data collected in the COVID-19 pandemic, new application domains were explored to guide epidemiologic policy-making (Fasina et al. 2021; Furuse et al. 2021; Hittner and Fasina 2021) using TPR to assess epidemic dispersal mediated by asymptomatic carriers. Recent studies have highlighted a correlation of TPR with the number of patients admitted in hospitals (Gaspari 2021; Lopez-Izquierdo, del Campo, and Eiros 2021), which increases with respect to other indicators like incidence Farrugia and Calleja (2021) or the daily number of positive cases Al Dallal, Al Dallal, and Al Dallal (2021). This kind of correlation, was exploited to forecast – two weeks in advance – variations

*Corresponding author: Mauro Gaspari, Department of Computer Science and Engineering, University of Bologna, Via Mura Anteo Zamboni, 7, Bologna, 40126, Italy, Phone: +390512094875, Fax: +390512094520, E-mail: mauro.gaspari@unibo.it. <https://orcid.org/0000-0002-2845-3904>

on the number of patients admitted in hospitals on the basis of TPR variations Fenga and Gaspari (2021), or to define a severity detection rate, to predict Intensive Care Unit (ICU) admissions Nikoloudis, Kountouras, and Hiona (2021).

Despite these promising results, the large percentage of asymptomatic carriers in COVID-19 (Yu and Rongrong 2020; Zhao et al. 2020), and the risk of hampering control efforts Chisholm et al. (2018), epidemiologists remain sceptics concerning the widespread use of TPR for surveillance, and other indicators based on known cases like incidence (7-day incidence rate per 100,000 inhabitants) are generally preferred. One of the motivations of this choice is that the calculation of TPR is more critical, and there is still little agreement on the method to be used, see Calculating SARS-CoV-2 Laboratory Test Percent Positivity (2021). For example, whether or not rapid antigen tests should be considered.

Our first goal is to clarify TPR calculation issues to identify which method behaves better for surveillance purposes. Namely, which method has the best performance in anticipating the trend of hospitalized patients. Then, our aim is to compare TPR and incidence rate to determine which of the two has the best characteristics for the same purpose.

The metric that we use for this comparison is based on the intuition that an optimal indicator for surveillance purposes should be able to track the progress of infection. More precisely, the number of infection cases occurring day by day in a pandemic: if the number of infection cases increases, an optimal indicator should also increase, while vice-versa, if the number of infection cases decreases, an optimal indicator should decrease. This aspect cannot be modelled by R_t (the reproduction number) only, because prevalence levels should be considered to estimate the volume of infection cases. Indeed, R_t only represents the number of secondary infections generated daily from a case, and not the total number of infection cases.

Intuitively, the peaks of an indicator that models the trend of infection should precede the peaks in hospitalized patients time series. Indeed, hospital admission always occurs several days after exposure, including incubation period, symptom onset and patient testing, about 17 days for COVID-19, possibly more. As an example, 11.5 days on average from exposure to symptoms onset was estimated in Lauer et al. (2020), and 5.6 days on average from symptoms onset to hospitalization (also including time for diagnosis) was reported in Faes et al. (2020). Similar considerations hold for admission in intensive care units, which in general occurs a few days after hospitalization.

By comparing the peaks of TPR and incidence with those of hospitalized and ICU patients time series, we can estimate how much the trend of these indicators anticipates the trend of the occupancy of beds in the health-care system. Peak comparison is often used as a metric in infectious diseases Dailey, Watkins, and Plant (2007), for example in the COVID-19 pandemic: to discriminate over time the delay of the infected time series with respect to emergency calls and twitter trends Riviaccio et al. (2021); to investigate the correlation between excess mortality in Italy and the occurrence of COVID-19 waves Roccetti (2023); or to investigate the hypothesis for seasonal periodicity Cappi et al. (2022). For the purpose of this study, peaks indicate the days in which the healthcare system was particularly stressed, but, most importantly, peaks also represent comparable changepoints in all the involved time series. For TPR and incidence, which aim to represent the progress of infection, peaks indicate the days in which the number of infection cases began to drop, followed by peaks in hospitalized and ICU time series, after which bed occupancy began to fall. Other changepoints could be considered to estimate these time lags, for example individuating points where a significant increase in a time series can be observed, like in Casini and Roccetti (2020), where a window sliding search method is used for the infected time series. However, the comparison of these changepoints in different time series requires more complex calibration strategies, and may be affected by varying conditions leading to inaccurate results Truong, Oudre, and Vayatis (2020).

Once we have identified an indicator that better models the progress of infection, anticipating the others, we can use statistical approaches like the one presented in Fenga and Gaspari (2021) for predicting the trend of hospitalized patients and hospital overloading.

In this paper, we start with a discussion of TPR calculation issues, then, using as a metric the distance in days from the peaks of hospitalized patients time series, we compare the incidence rate with different TPR calculation methods, including standard rolling average with or without antigen tests, and a new proposal based on a two-level approach. We consider 60 different outbreak cases using global and local data from Italy in four waves of

the pandemic, starting from the second wave, where data on antigen tests were made available for some regions (Fall 2020), and the successive Alpha, Delta and Omicron waves.

In summary, we aim to answer some key questions concerning COVID-19 and infectious diseases in general: is the incidence rate the most appropriate indicator to track infection when many asymptomatic carriers are present? Would TPR have similar or better performance in the same context? Which TPR calculation method should be used? Should antigen tests be used in TPR calculation?

TPR calculation issues

The main issue concerning TPR calculation is that conducted tests have different goals: they include both diagnostic tests administered with the goal of discovering new cases, and control tests, addressed to infected individuals, to monitor the course of the disease or to check the healing process. While the positive percentage of the former can be used for surveillance modelling the progress of the pandemic, the positive control tests should be used for different purposes, for example to evaluate the length of quarantine Landon et al. (2022). Regretfully, these different types of tests are not classified in the relevant statistics, and only the total number of administered tests is reported daily. Due to this lack of information, test positivity is usually computed as the ratio between new positive cases and the total number of tests done Calculating SARS-CoV-2 Laboratory Test Percent Positivity (2021), being only an approximation of the actual positivity rate.

Another open issue is whether or not antigen tests should be part of the calculation. For example, CDC (US Centers for Disease Control and prevention) computes test positivity as the percentage of all SARS-CoV-2 Nucleic Acid Amplification Tests (NAAT) conducted that are positive, while they recommend to collect antigen tests as separate data, see Calculating SARS-CoV-2 Laboratory Test Percent Positivity (2021). On the contrary, several available statistics use both of them as denominators, such as those presented in the Coronavirus Testing web site Hasell et al. (2020). Moreover, when antigen tests are used in TPR calculation, an additional problem arises: healthcare guidelines may recommend that positive antigen tests should be confirmed by NAAT tests because the latest ensure better accuracy. These repeated diagnostic tests done for the same positive individuals should also be removed by the denominator Gaspari (2021).

Furthermore, there are several problems related to data collection in different regions or countries. For example, in most of the Italian regions, the number of tests reported on Monday is lower than the those reported in other days of the week, and includes a lower ratio of antigen tests. From this it follows that the computed TPR is in general higher on Monday with respect to the other days of the week. In other words, in Italy there is a form of weekly seasonality in TPR time series, and similar problems also arise in other countries.

Finally, a negative correlation between TPR and the number of administered tests was evidenced in some studies (Fasina et al. 2021; Nikoloudis, Kountouras, and Hiona 2021; Vong and Kakkar 2020). Basically, when the number of conducted tests increases, test positivity tends to decrease, and thus some variations of TPR may be linked to an increased volume of testing. This phenomenon is more evident when the number of administered tests per million inhabitants is low, and tends to mitigate when a large number of tests are conducted. For example, in fall 2021, when the “green pass” for workers became mandatory in Italy, and the number of the tests performed almost doubled in a few days, no significant variation was observed in the TPR time series. To deal with this issue, an adjusted TPR calculation method was proposed in Vong and Kakkar (2020) to be used in countries where capacity of testing is limited.

Methods

As in most of the countries Hasell et al. (2020), COVID-19 tests administered for different purposes are not classified in Italian official data (only NAAT and antigen tests are distinguished, see COVID-19 Italia (2022), in this study we will approximate TPR as the ratio

between new positive cases and the total number of tests done. This solution is a common approach, also recommended from CDC, see Calculating SARS-CoV-2 Laboratory Test Percent Positivity (2021).

Starting from this basic calculation method, we compare 7 different versions of TPR, two of them are based on NAAT nasopharyngeal swab only, while the others exploit antigen tests. The first 6 versions are obtained by computing the rolling average of the last 7 days of the following ratios:

- N1: New positive cases /NAAT tests only.
- N2: New positive cases detected with NAAT tests only /NAAT tests only.
- A1: New positive cases /NAAT + Antigen tests.
- A2: New positive cases /NAAT + Antigen tests – Estimated repeated tests.
- A3: New positive cases /NAAT + Antigen tests – Number of healed patients.
- A4: (New positive cases /NAAT + Antigen tests) * (growth rate of cases /growth rate of tests).

Version A1 is the usual 7 days rolling average based on all the tests done Hasell et al. (2020), while versions A2 and A3 are attempts to improve its accuracy by removing those tests that are not devoted to the diagnosis of new cases from the denominator, namely:

A2 removes an estimation of repeated diagnostic tests (NAAT tests conducted to confirm positive Antigen tests) computed using the approach presented in Fenga and Gaspari (2021).

A3 removes the number of healed people, assuming that at least one test was administered to each of them Fenga and Gaspari (2021). This data is usually reported daily in most of the countries.

Version A4 is the adjusted TPR presented in Vong and Kakkar (2020) which deals with the negative correlation between TPR and the number of administered tests.

More formally, let P^d and N^d be respectively the new positive cases and cases detected with NAAT tests only for each day d ; let T^d and A^d be respectively the number of NAAT tests and the number of Antigen tests done for each day d ; let R^d and Pr^d be respectively the number of healed patients and an estimation of repeated tests for each day d , and let the notation \overline{X}_N^d indicate last N days average of all these time series at time d , we can define the above TPR versions computed in day d as follows:

$$\tau_{N1}^d = \frac{\overline{P}_7^d * 100}{T_7^d} \quad (1)$$

$$\tau_{N2}^d = \frac{\overline{N}_7^d * 100}{T_7^d} \quad (2)$$

$$\tau_{A1}^d = \frac{\overline{P}_7^d * 100}{(T^d + A^d)_7} \quad (3)$$

$$\tau_{A2}^d = \frac{\overline{P}_7^d * 100}{(T^d + A^d)_7 - Pr^d} \quad (4)$$

$$\tau_{A3}^d = \frac{\overline{P}_7^d * 100}{(T^d + A^d)_7 - R^d} \quad (5)$$

The estimation of the average number of repeated tests presented in Fenga and Gaspari (2021) is computed as follows: let $\frac{A}{Pr} = \alpha^A$ be the antigen positivity rate, and $\frac{T}{p-Pr} = \alpha^T$ the NAAT positivity rate, given that, in general, $\alpha^T = \psi * \alpha^A$ where $\psi > 1$ holds (namely the positivity rate for antigen tests is lower, see for example the results presented here Turcato et al. 2021); if we set $\psi = 2$, which is a reasonable approximation observing data available for Italy, we obtain:

$$Pr^d = \frac{(A^d * P^d)_7 * 2}{(T^d + A^d)_7 * 2} \quad (6)$$

Finally, the adjusted TPR Vong and Kakkar (2020) is defined by multiplying observed TPR (A1), with a factor which expresses the growth rates of cases and tests as follows:

$$\tau_{A4}^d = \left(\frac{\overline{P}_7^d * 100}{(T^d + A^d)_7} * \frac{(C^d - C^{d-1}) / C^{d-1}}{(Y^d - Y^{d-1}) / Y^{d-1}} \right)_7 \quad (7)$$

where C^d and Y^d are the cumulative number of cases and tests (including antigen tests) at time d respectively.

In addition, we also evaluate a new version of TPR (A5) based on a two-level approach which aims to model the progress of infection with better accuracy, thus generating more stable time series.

A two-level TPR calculation method

A criticism to methods based on standard 7 days rolling average, which compute the TPR of a given day as the average of the preceding 7 days, is that they mainly deal with data collection issues, without considering modelling and epidemiological issues. Indeed, from a knowledge modelling perspective, estimating the TPR of a given day as the average of the preceding and following days will be more appropriate and correct. Moreover, 7 days are necessary to deal with week seasonality issues, but significant variations in the progress of infection could be captured with better accuracy by considering less than 7 days. For example, the length in days of the incubation period, which represents the minimal number of days after which changes due to infection can be observed.

To get through these issues, we have devised a two-level approach for calculating TPR: the first level addresses modelling issues, and the second epidemiological issues. Let t_1, t_2, \dots, t_n be the daily TPR time series, TPR at time i , when $i > 3$ and $i < n - 3$ can be modelled by computing the trend as follows:

$$\bar{t}_i = \frac{t_{i-3} + \dots + t_{i+3}}{7} \quad (8)$$

Namely, TPR in a given day is modelled by computing the average value of the days preceding and following it. This formula cannot be computed for the last three days, but it can be approximated by computing averages on the available days only as follows: $\bar{t}_{n-2} = \frac{t_{n-5} + \dots + t_n}{6}$, $\bar{t}_{n-1} = \frac{t_{n-4} + \dots + t_n}{5}$, $\bar{t}_n = \frac{t_{n-3} + \dots + t_n}{4}$, where t_n be the daily TPR of the last day.

However, this simple approach, which assumes that the average TPR value will not have significant changes in the remaining days, does not work well due to the seasonality of the daily TPR time series. Figure 1(a) illustrates this problem considering the region of Emilia Romagna: the highest values of daily TPR occur on Monday, while on Tuesday the daily TPR is well below average. Thus, if the last day were Monday, the TPR would be overestimated, and if it were Tuesday, it would probably be underestimated.

To solve this problem, modelling seasonality, we compute the difference between the TPR trend (computed with formula (8)) and daily TPR for each day in the last k weeks. Let $(s_1, s_2, s_3, s_4, s_5, s_6, s_7)$, be the list of average values of these differences for each day in the week (where s_7 is the mean difference associated to the last element of the trend time series \bar{t}_{n-3}), the TPR of the last 3 days is obtained by adding the weighted differences of the days that are missing to the previous formulas (e.g., the days not used for computing formula (8); see Figure 1(b)): $\bar{t}_{n-2} = \frac{t_{n-5} + \dots + t_n}{6} + s_4/6$, $\bar{t}_{n-1} = \frac{t_{n-4} + \dots + t_n}{5} + s_4/5 + s_5/5$, $\bar{t}_n = \frac{t_{n-3} + \dots + t_n}{4} + s_4/4 + s_5/4 + s_6/4$. The experiments we did show that this approach avoids the effects of seasonality, thus improving TPR accuracy in the last 3 days. More precisely, when analysing time series history (more than 16,000 cases), setting $k=4$ weeks: \bar{t}_{n-2} improves the naive approach on average of 0.0004, \bar{t}_{n-1} of 0.0022, and \bar{t}_n of 0.0032. Increasing or reducing k worsens these results. It should be noted that this issue does not have any influence on offline peak detection methods, thus it does not interfere with the metric used in the study.

Starting from the trend time series computed as above, we introduced a second level to compute the final TPR value at day d considering epidemiological issues. The TPR value is obtained as the average of the last μ days of the trend time series, where μ is the incubation period.

$$\tau_{AS}^d = \frac{(\bar{t}_d + \bar{t}_{d-1} \dots \bar{t}_{d-\mu})}{\mu} \quad (9)$$

Where $\mu = 5$ on the onset Lauer et al. (2020), and $\mu = 3$ in the Omicron outbreak (Brandal et al. 2021; Jansen et al. 2021).

We will show how this approach effectively improves predictive properties and regularity of the TPR time series.

Data collection

The data used for this study were made available by the Italian Department of Civil Protection COVID-19 Italia (2022) for all the Italian regions, throughout the whole of the pandemic. This site contains all the relevant time series needed for our analysis, namely: new positive cases; NAAT tests; antigen tests; patients admitted in hospitals and ICUs; recovered patients. However, the structure of the dataset was changing over time, and data fundamental for our analysis were not always available and/or reliable, therefore, for each wave of the pandemic, some of the regions were discarded due to lack of data or known reliability issues.

The first wave of the pandemic starting from February 2020 was not considered in the study because antigen tests were not used in Italy. In the second wave (from the 1st of October 2020 to the 10th of January 2021) antigen tests were used in some regions, but they still were not available in official data: they were reported by text notes in the dataset, or published in news for five regions Gaspari (2021). However, since some of these data are uncertain, we only considered Toscana and Piemonte in the second wave, where data on antigen tests were continuously published starting from October 2021. All the other regions were excluded.

From the 15th of January 2021 onwards, data on antigen tests were made available for all the regions of Italy. In spite of this, in the successive Alpha wave (from the 10th of February 2021 to the 1st of May 2021) which started right after (February 2021), some regions were still excluded, because the number of hospitalized patients had started to grow before the beginning of the observed period, where data on tests were still unreliable. The excluded regions are: Abruzzo, Umbria, Molise, and Basilicata.

As regards the Delta wave (from the 1st of July 2021 to the 10th of October 2021), all Italian regions were considered except Lazio, where data on tests had been corrupted due to a hacker attack.

Successively, in the last Omicron outbreak (from the 24th of December 2021 to the 18th of February 2022): Sardinia was excluded since the TPR peak was at the end of the observed period (the 16th of February); Valle d'Aosta was excluded due to an error in the

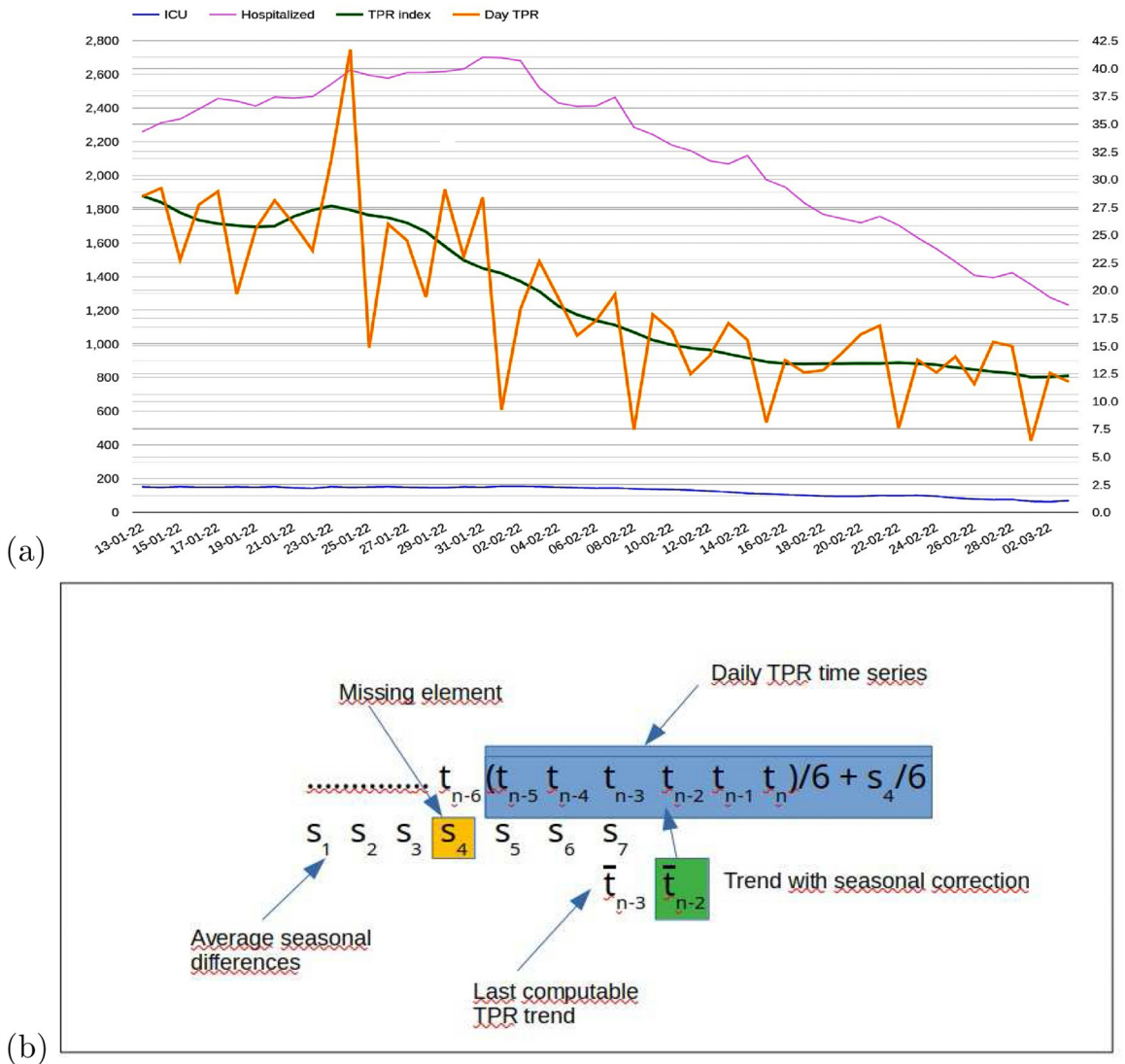


Figure 1: Dealing with seasonality: (a) Daily TPR seasonality in Emilia Romagna in January/February 2022; (b) computation of the TPR trend in the last 3 days by adding seasonal corrections: \bar{t}_{n-2} example.

reported data concerning positive cases detected with NAAT test only; the province of Bolzano was excluded because a large amount of antigen tests were not reported in this period.

Finally, global data for Italy in the last 3 waves were also considered as further case studies, obtaining a total of 60 different outbreak cases: 2 regions in the second wave, 17 regions and whole Italy in the third wave, 20 regions and whole Italy in the Delta wave, and 18 regions and whole Italy in the Omicron wave.

Although all the data come from a single country, Italy, we believe that the collected sample is sufficiently general and heterogeneous in order to draw valid conclusions. Indeed, Italian regions range from little territories with less than 500 thousand inhabitants to larger regions with over 10 million inhabitants. Moreover, Italian regions have their own health departments and different organizations, and, as a consequence, heterogeneous data collection policies for the administration of diagnostic tests.

Data analysis

For all the analysed outbreak cases, we compute the number of days that occur between peaks of the above indicators and peaks of hospitalized people, considering both patients admitted in non critical areas and in intensive care units.

Peaks are identified by finding the local maxima in all the studied time series for all outbreaks. This simple method gives reliable results in this offline study, because each time series in each outbreak has only one peak. Moreover, anomalies are smoothed in health indicators like TPR and incidence rate, which are not raw data time series, and local maxima in hospitalized time series are actually the days in which the healthcare system was more stressed.

We analyse the generated samples by computing average values and standard deviation, also studying differences between indicators. We use a simple non-parametric Bootstrap method Hesterberg (2011) with Monte-Carlo simulation performing 5,000 iterations, to compute confidence intervals for all the obtained average values.

Moreover, we use sample entropy Richman and Moorman (2000) as a measure of the regularity of these indicators Namdari and Zhaojun (2019). This is a measure of the probability that two segments of the analysed time series within a given period minimize discontinuities. Smaller values of sample entropy indicate a greater probability that a set of TPR values will be followed by similar values, while a larger value indicates major irregularities. We assume 2 as the embedded dimension and the Chebyshev distance as a metric.

Results and discussion

The results we obtained considering all the selected 60 outbreak cases are summarized in Table 1. An initial analysis of these results allows us to draw preliminary conclusions about the ability to anticipate hospitalized peaks of different versions of TPR, specifically: N2, which uses positives cases determined with NAAT tests only as a numerator, outperforms N1, which uses all the new positives; A1, e.g., the standard 7 days rolling average including both NAAT and antigen tests, outperforms versions A2, A3 and the adjusted TPR A4. Basically, all the attempts to improve TPR accuracy by removing tests from the denominator, or considering changes in the number of administered tests, seem to fail. For example, removing tests used to check the healing process, apparently provides a better approximation during the growth phase of the curve, but it tends to create artificial peaks, before the phase of descent, when the number of recovered patients is high.

Last but not least, the new version of TPR (A5) anticipates all the other indicators including incidence and has a better regularity, considering both standard deviation and sample entropy. Thanks to these properties, a decrease observed after a growth phase has more chance of identifying a peak, rather than an anomalous effect, and the same principle holds for detecting new pandemic waves. Figure 2(a) presents a comparison between A5 and naive 7 days rolling average (A1) illustrating the benefits of having a better sample entropy.

These considerations for TPR versions N1, A1, and A5 hold in all the analyzed waves, thus we focus the subsequent discussion on these indicators only comparing them with incidence. Similar results hold for both patients admitted in non critical areas and in intensive care units. However, considering ICUs it is clear that the results do not allow to reach meaningful conclusions, because standard deviation is quite high, thus a deeper analysis is needed for ICUs.

Table 1: This table summarizes the average distances between peaks of the studied indicators with respect to those of patients admitted in non-critical areas and ICUs, considering all the 60 outbreak cases. We present average values, standard deviation and sample entropy for each indicator.

Indicator	Hospitalized		Sample Entropy	ICU	
	Avg Dist.	Std Dev.		Avg Dist.	Std Dev.
Incidence	13.067	10.734	0.305	10.417	13.631
TPR N1	11.7	12.17	0.34	9.05	14.835
TPR N2	13.833	11.649	0.421	11.183	12.614
TPR A1	16.083	11.279	0.324	13.433	15.692
TPR A2	16.033	11.309	0.327	13.383	15.662
TPR A3	13.65	11.828	0.349	11.0	16.813
TPR A4	16.033	11.237	0.334	13.383	15.192
TPR A5	17.633	10.66	0.225	14.983	14.871

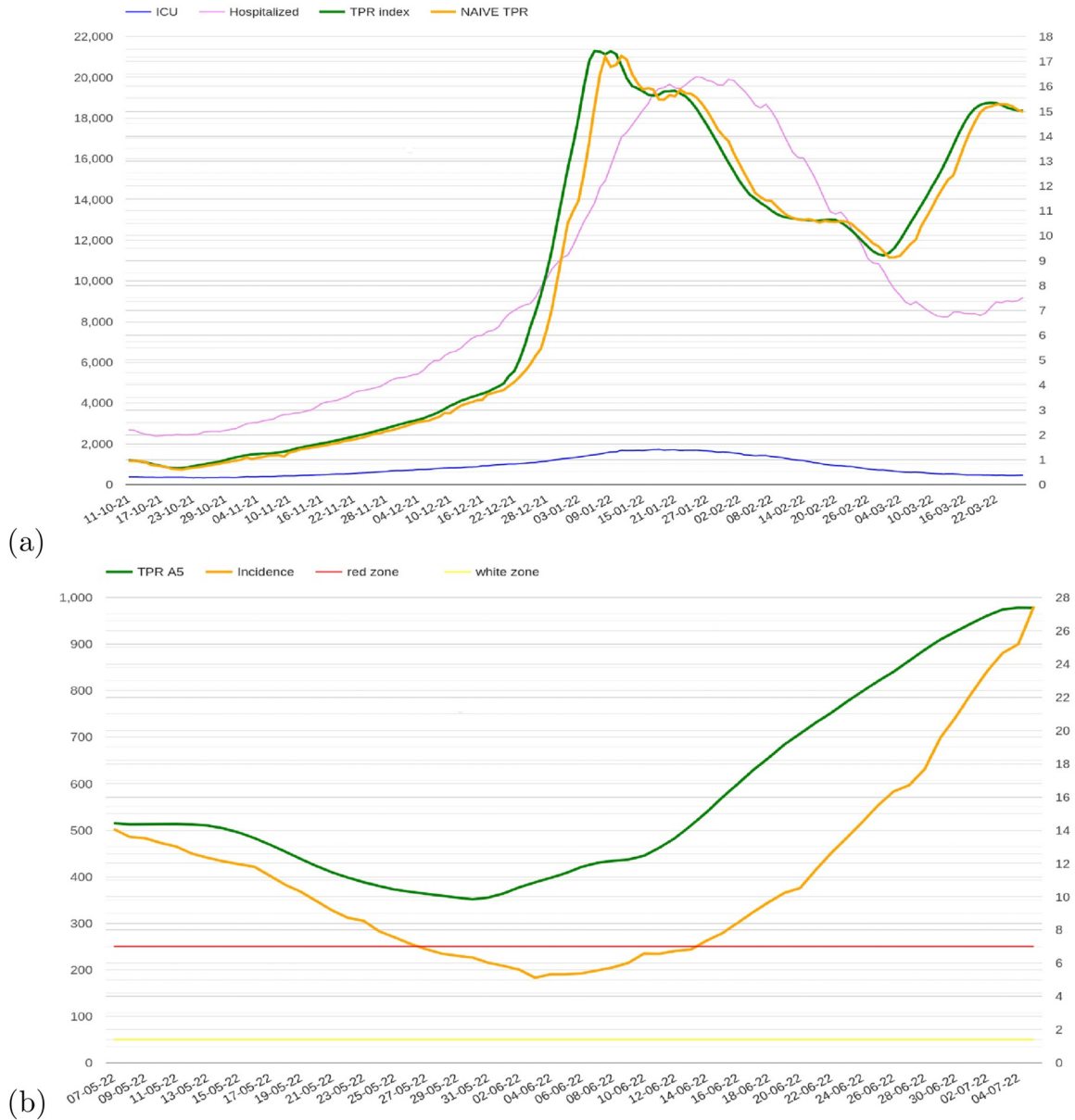


Figure 2: Examples using global Italian data: (a) TPR (A5) compared with standard rolling average (version A1) in the first Omicron Outbreak (January 2022). (b) TPR (A5) compared with incidence in the last Omicron 5 outbreak (June 2022), the incidence rate thresholds for red (above 250) and white (below 50) zones are indicated in red and yellow. TPR captures the effect of the new Omicron variant about 6 days before incidence.

If we consider patients admitted in non critical areas instead, it appears that TPR (A5) anticipates incidence. Indeed, the average distance between TPR peaks and hospitalized peaks is 17.6 days with respect to 13.0 days only, obtained for incidence. In practice, the best TPR-based approach anticipates incidence of about 4.6 days. These preliminary observations are certainly strengthened by the fact that we obtained similar results computing average values of TPR and incidence in all the considered waves, that is, TPR anticipates incidence in all the analyzed waves, and the results have similar proportions. However, more considerations are needed to draw definitive conclusions.

Let Δ_{TPR} and Δ_I be respectively the sets of distances between TPR (A5) and incidence peaks with respect to peaks of hospitalized time series, and $\Delta_{\text{TPR}-I}$ be the set of differences between Δ_{TPR} and Δ_I case by case.

Figure 3(a)–(c) shows how the elements of these sets are distributed. Since, the analysis of single outbreak cases has shown that extreme cases are always outlier, and that the considered sets do not belong to heavy-tailed distributions, we use a non-parametric Bootstrap method with Monte-Carlo simulation to estimate the confidence interval of the averages of the analyzed distribution (Figure 3(d)–(f)). The experiments we made show that 5,000 iterations are enough to converge generating stable values; the resulting confidence intervals are: Δ_{TPR} : average: 17.6, 95 % CI 15.0, 20.4; Δ_I : average 13.0, 95 % CI 10.4, 15.8; $\Delta_{\text{TPR}-I}$: average 4.6, 95 % CI 2.8, 6.4.

Given that confidence intervals estimated for TPR and incidence overlap, we analyze all single outbreak cases where incidence anticipates TPR. As Figure 3(c) shows, incidence anticipates TPR of more than 2 days in 5 out of 60 cases only. Nonetheless, in all these cases which include: Lazio (18 days), Campania (4) and Sardinia (5) in the 3rd wave, Lombardy (8) in the Delta wave, and Puglia (6) in the Omicron wave, the trends of TPR and

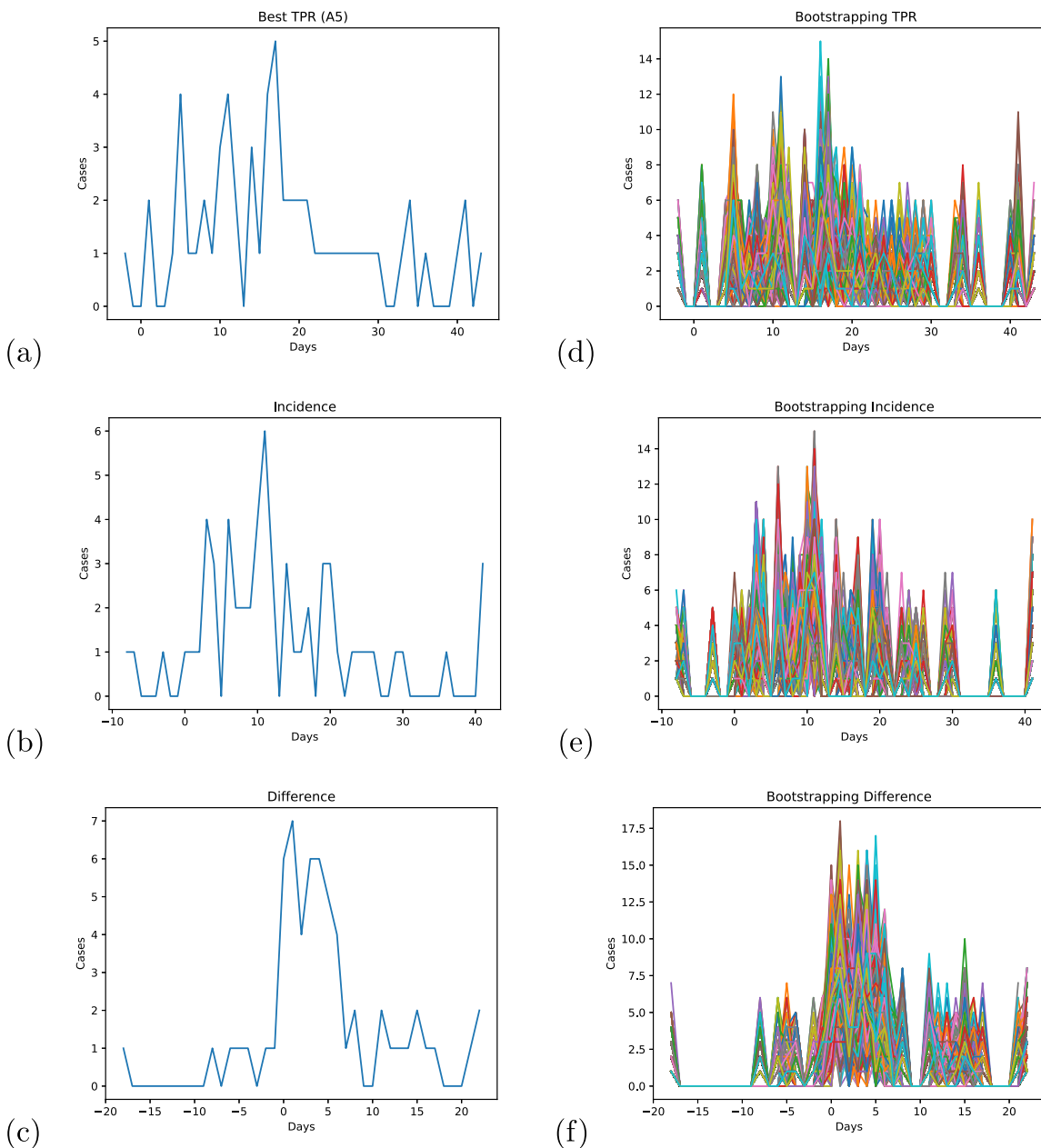


Figure 3: The figure shows how the values in the sets Δ_{TPR} , avg: 17.6, sd: 10.7 (a), Δ_I , avg: 13.1, sd: 10.7 (b), and $\Delta_{\text{TPR}-I}$, avg: 4.6, sd: 7.1 (c), are distributed, and the results obtained bootstrapping them with 5,000 iterations (d)–(f).

incidence remain similar, both of them reaching the top of a plateau nearly at the same time, and the delay of TPR peaks is caused by small variations in the plateau.

On the other hand, when analyzing the relationship of TPR and incidence with hospitalized patients, TPR peaks follow peaks of hospitalized patients in one case only (Friuli in the Delta wave), while this happens three times for incidence (Friuli, Valle d'Aosta, and Marche in the Delta wave).

In summary, although there are a few outbreak cases where incidence anticipates TPR, the computed 95 % confidence intervals are almost disjoint, and it never happens in practice that incidence really anticipates TPR, with a possible negative impact on surveillance. In other words, in the detected outliers, techniques that exploit TPR for estimating variations in hospital admissions two weeks in advance, such as those presented in Fenga and Gaspari (2021), will work anyway. We can therefore conclude that TPR has a better predictive capacity than incidence for surveillance purposes. This also means that other changepoints can be detected in advance. For example, Figure 2(b) compares TPR (A5) and incidence in the recent Omicron 5 outbreak at the beginning of June 2022 using global data from Italy, showing that TPR has allowed detecting the beginning of this new wave about 6 days in advance. The rise of this new wave has been also confirmed observing the growth of TPR which occurred at the same time in almost all Italian regions.

The insight of this research is that indicators based on known cases are not able to model the progress of infections in infectious diseases with asymptomatic carriers with sufficient accuracy, while TPR also accounts for unknown cases, and thus modelling under-ascertainment Russell et al. (2020) does.

Concerning intensive care units an in-depth analysis of the obtained results shows that they were negatively influenced by low values found for the Omicron variant (19 outbreak cases), as presented in Table 2(a). Indeed,

Table 2: TPR and incidence average peaks distances with respect to peaks of patients admitted in non critical areas and ICUs considering different waves and seasons. For each indicator we present average values, standard deviation and sample entropy.

Indicator	Hospitalized		Sample	ICU	
	Avg dist.	Std dev.	Entropy	Avg dist.	Std dev.
(a) Omicron wave (19 cases)					
Incidence	12.895	8.687	0.332	2.579	13.019
TPR N2	15.263	12.9	0.524	4.947	15.582
TPR A1	15.474	10.475	0.237	5.158	14.612
TPR A5	18.158	10.137	0.2	7.842	14.091
(b) All the cases excluding Omicron wave (41 cases)					
Incidence	13.146	11.56	0.292	14.049	12.317
TPR N2	13.171	10.959	0.373	14.073	9.694
TPR A1	16.366	11.622	0.365	17.268	14.662
TPR A5	17.39	10.885	0.236	18.293	14.037
(c) Delta wave, summer 2021 (21 cases)					
Incidence	14.762	14.573	0.32	16.952	15.37
TPR N2	14.19	13.971	0.372	16.381	11.902
TPR A1	20.048	13.899	0.359	22.238	17.765
TPR A5	21.143	12.631	0.252	23.333	16.915
(d) Cold seasons: non-critical areas (39 cases); ICUs excluding the Omicron outbreak (20 cases).					
Incidence	12.154	7.781	0.296	11.0	6.693
TPR N2	13.641	10.177	0.447	11.65	5.695
TPR A1	13.949	8.869	0.306	12.05	7.493
TPR A5	15.744	8.871	0.21	13.0	6.986

if we exclude the Omicron wave, and we compute the average distances between peaks for the remaining 41 cases, the results get closer to those obtained for patients in non-critical areas (see Table 2(b)). More precisely, the best version of TPR anticipates incidence of about 4.2 days, in detail the average distance between TPR and ICU peaks is about 18.3 and 14 days only for incidence, while the values of standard deviation for ICUs remain high. On the contrary, for patients hospitalized in non-critical areas in the Omicron outbreak, average values conform to those obtained in the other waves.

A plausible explanation of the short-anticipation obtained for ICUs in the Omicron outbreak comes from the concatenation of Delta and Omicron variants. A reasonable hypothesis is that the observed peaks for ICUs, which occurred in the first half of January, mostly depend on Delta peaks presumably occurring at the end of December 2021. Subsequently, when Omicron became the dominant variant, there was a significant reduction of critical cases confirmed by an evident decrease of the ratio between critical cases admitted in ICUs and patients in non-critical areas, which in Italy globally dropped by half from the end of December (about 13 %) to the first half of January (about 6 %).

Another aspect which may give further guidance on the relationship of the studied indicators with asymptomatic carriers is whether significant changes can be observed in different seasons of the year, for example during the summer, when the number of asymptomatic carriers is presumably higher. Table 2(c) presents the results of the analysis for the Delta outbreak at the beginning of July 2021. If we compare these results with those obtained in cold seasons presented in Table 2(d), we can observe that the average time lags between TPR (A5) peaks and hospitalized time series increase from 15.7 to 21.1 days. This effect depends on at least two factors: (1) the course of the diseases which is presumably longer during the summer, (2) the percentage of asymptomatic carriers which increases with hot weather. The first factor is supported by the fact that all the indicators increase, the second factor most likely depends on loss of accuracy of incidence in modelling the progress of infection due to a large proportion of asymptomatic carriers. It is worth noticing that during the winter (Table 2(d) excluding the Omicron outbreak) the results obtained for ICUs better conform to those obtained in other studies like (Fenga and Gaspari 2021; Nikoloudis, Kountouras, and Hiona 2021), and standard deviation has acceptable values.

Conclusions

The results we have discussed allow us to answer the questions asked at the beginning of this study with sufficiently convincing arguments.

First, there is an evidence that the incidence rate is not the best indicator for surveillance in infectious diseases when a considerable percentage of asymptomatic carriers is present, as for the COVID-19 pandemic, test positivity should be used instead. The ability to detect peaks earlier would allow health professionals, by means of statistical methods, such as in Fenga and Gaspari (2021), to early predict the trend of hospital admissions, and potential hospital overloads. Similarly, other significant change-points can be detected in advance, like the beginning of a new wave. This result holds for patients admitted in non critical areas, while more investigations are needed for ICUs.

Second, further support is given to the hypothesis that antigen tests should be used in TPR calculation. The performance of the best PCR-based TPR calculation method is similar to those for incidence: the average value is slightly higher, but standard deviation gets worse. In other words, TPR outperforms incidence only if antigen tests are considered in the calculation.

Other key practical implications of this research are the following: data collection procedures should be improved to make TPR calculation as accurate as possible; TPR-based approaches to compute epidemiological parameters, like R_t , should be investigated more deeply.

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

Data availability: The data used in this study were provided by the Italian Civil Protection Department, and are available here: <https://github.com/pcm-dpc>. Upon reasonable request, we also provide the used TPR and incidence rate time series using a Google graph html format, which can be visualized using a simple web browser.

References

- World Health Organization. 2021. *Considerations for Implementing and Adjusting Public Health and Social Measures in the Context of COVID-19. Interim guidance*, 14 June 2021. Strategic Health Operations, WHO Headquarters (HQ). *WHO Reference Number: WHO/2019-nCoV/Adjusting_PH_measures/2021.1*. <https://www.who.int/publications/i/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance> (accessed July 7, 2022).
- Al Dallal, A., U. Al Dallal, and J. Al Dallal. 2021. "Positivity Rate: An Indicator for the Spread of COVID-19." *Current Medical Research and Opinion* 37 (12): 2067–76.
- Boyce, R. M., R. Reyes, M. Matte, M. Ntaro, E. Mulogo, F. C. Lin, and M. J. Siedner. 2016. "Practical Implications of the Non-Linear Relationship between the Test Positivity Rate and Malaria Incidence." *PLoS One* 11 (3): e0152410.
- Brandal, L. T., E. MacDonald, L. Veneti, T. Ravlo, H. Lange, U. Naseer, S. Feruglio, K. Bragstad, O. Hungnes, L. E. Ødeskaug, F. Hagen, K. E. Hanch-Hansen, A. Lind, S. V. Wattle, A. M. Taxt, M. Johansen, L. Vold, P. Aavitsland, K. Nygård, and E. H. Madslie. 2021. "Outbreak Caused by the SARS-CoV-2 Omicron Variant in Norway, November to December 2021." *Euro Surveillance* 26 (50): 2101147.
- Chiu, W. A., and M. L. Ndeffo-Mbah. 2021. "Using Test Positivity and Reported Case Rates to Estimate State-Level COVID-19 Prevalence and Seroprevalence in the United States." *PLoS Computational Biology* 17 (9): e1009374.
- Chisholm, R. H., P. T. Campbell, Y. Wu, S. Y. C. Tong, J. McVernon, and N. Geard. 2018. "Implications of Asymptomatic Carriers for Infectious Disease Transmission and Control." *Royal Society Open Science* 5 (2): 172341.
- Calculating SARS-CoV-2 Laboratory Test Percent Positivity. 2021. CDC Methods and Considerations for Comparisons and Interpretation. <https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/calculating-percent-positivity.html> (accessed July 7, 2022).
- Cappi, R., L. Casini, D. Tosi, and M. Rocchetti. 2022. "Questioning the Seasonality of SARS-COV-2: A Fourier Spectral Analysis." *BMJ Open* 12 (4): e061602.
- Casini, L., and M. Rocchetti. 2020. "A Cross-Regional Analysis of the COVID-19 Spread during the 2020 Italian Vacation Period: Results from Three Computational Models Are Compared." *Sensors* 20 (24): 7319.
- COVID-19 Italia 2022. *Monitoraggio situazione, Italian Civil Protection Department*. <https://github.com/pcm-dpc> (accessed July 7, 2022).
- Dailley, L., R. E. Watkins, and A. J. Plant. 2007. "Timeliness of Data Sources Used for Influenza Surveillance." *Journal of the American Medical Informatics Association* 14 (5): 626–31.
- Fasina, F. O., M. A. Salami, M. M. Fasina, O. A. Otegunrin, A. L. Hoogesteijn, and J. B. Hittner. 2021. "Test Positivity—Evaluation of a New Metric to Assess Epidemic Dispersal Mediated by Non-Symptomatic Cases." *Methods* 195: 15–22.
- Furuse, Y., Y. K. Ko, K. Ninomiya, M. Suzuki, and H. Oshitani. 2021. "Relationship of Test Positivity Rates with COVID-19 Epidemic Dynamics." *International Journal of Environmental Research and Public Health* 18 (9): 4655.
- Farrugia, B., and N. Calleja. 2021. "Early Warning Indicators of COVID-19 Burden for a Prosilient European Pandemic Response." *The European Journal of Public Health* 31 (4): iv21–6.
- Fenga, L., and M. Gaspari. 2021. "Predictive Capacity of COVID-19 Test Positivity Rate." *Sensors* 21 (7): 2435.
- Faes, C., S. Abrams, D. Van Beckhoven, G. Meyfroidt, E. Vlieghe, and N. Hens. 2020. "Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients." *International Journal of Environmental Research and Public Health* 17 (20): 7560.
- Gaspari, M. 2021. "COVID-19 Test Positivity Rate as a Marker for Hospital Overload." medRxiv 01.26.21249544. <https://doi.org/10.1101/2021.01.26.21249544>.
- Hittner, J. B., and F. O. Fasina. 2021. "Statistical Methods for Comparing Test Positivity Rates between Countries: Which Method Should Be Used and Why?" *Methods* 195: 72–6.
- Hasell, J., E. Mathieu, D. Beltekian, B. Macdonald, C. Giattino, E. Ortiz-Ospina, M. Roser, and H. Ritchie. 2020. "A Cross-Country Database of COVID-19 Testing." *Scientific Data* 7: 345.
- Hesterberg, T. 2011. "Bootstrap." *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (6): 497–526.
- Jansen, L., B. Tegomoh, K. Lange, K. Showalter, J. Figliomeni, B. Abdalhamid, P. C. Iwen, J. Fauver, B. Buss, and M. Donahue. 2021. "Investigation of a Sars-Cov-2 B. 1.1. 529 (Omicron) Variant cluster—Nebraska, November–December 2021." *Morbidity and Mortality Weekly Report* 70 (5152): 1782–4.
- Lopez-Izquierdo, R., F. del Campo, and J. M. Eiros. 2021. "Influence of Positive SARS-CoV-2 CRP on Hospital Admissions for COVID-19 in a Spanish Health Area." *Medicina Clínica* 156 (8): 407–8.

- Landon, E., A. H. Bartlett, R. Marrs, C. Guenette, S. G. Weber, and M. J. Mina. 2022. High Rates of Rapid Antigen Test Positivity after 5 Days of Isolation for COVID-19. *MedRxiv* 2022-02, <https://doi.org/10.1101/2022.02.01.22269931>.
- Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler. 2020. “The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application.” *Annals of Internal Medicine* 172 (9): 577–82.
- Nikoloudis, D., D. Kountouras, and A. Hiona. 2021. “A Novel Benchmark for COVID-19 Pandemic Testing Effectiveness Enables the Accurate Prediction of New Intensive Care Unit Admissions.” *Scientific Reports* 11: 20308.
- Namdari, A., and L. Zhaojun. 2019. “A Review of Entropy Measures for Uncertainty Quantification of Stochastic Processes.” *Advances in Mechanical Engineering* 11: 6.
- Rivieccio, B. A., A. Micheletti, M. Maffeo, M. Zignani, A. Comunian, F. Nicolussi, S. Salini, G. Manzi, F. Auxilia, M. Giudici, G. Naldi, S. Gaito, S. Castaldi, and E. Biganzoli. 2021. “CoViD-19, Learning from the Past: A Wavelet and Cross-Correlation Analysis of the Epidemic Dynamics Looking to Emergency Calls and Twitter Trends in Italian Lombardy Region.” *PLoS One* 16 (2): e0247854.
- Rocchetti, M. 2023. “Excess Mortality and COVID-19 Deaths in Italy: A Peak Comparison Study.” *Mathematical Biosciences and Engineering* 20 (4): 7042–55.
- Russell, T. W., N. Golding, J. Hellewell, S. Abbott, L. Wright, C. A. Pearson, K. van Zandvoort, C. I. Jarvis, H. Gibbs, Y. Liu, W. J. Edmunds, A. J. Kucharski, A. K. Deol, C. J. Villabona-Arenas, T. Jombart, K. O’Reilly, J. D. Munday, S. R. Meakin, R. Lowe, A. Gimma, A. Endo, E. S. Nightingale, G. Medley, A. M. Foss, G. M. Knight, K. Prem, S. Hué, C. Diamond, J. W. Rudge, K. E. Atkins, M. Auzenberg, S. Flasche, R. M. G. J. Houben, B. J. Quilty, P. Klepac, M. Quaife, S. Funk, Q. J. Leclerc, J. C. Emery, M. Jit, D. Simons, N. I. Bosse, S. R. Procter, F. Y. Sun, S. Clifford, K. Sherratt, A. Rosello, N. G. Davies, O. Brady, D. C. Tully, and G. R. Gore-Langton. 2020. “Reconstructing the Early Global Dynamics of Under-ascertained COVID-19 Cases and Infections.” *BMC Medicine* 18: 332.
- Richman, J. S., and J. R. Moorman. 2000. “Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy.” *American Journal of Physiology - Heart and Circulatory Physiology* 278 (6): H2039-49.
- Truong, C., L. Oudre, and N. Vayatis. 2020. “Selective Review of Offline Change Point Detection Methods.” *Signal Processing* 167: 107299.
- Turcato, G., A. Zaboli, N. Pfeifer, L. Ciccarliello, S. Sibilio, G. Tezza, and D. Ausserhofer. 2021. “Clinical Application of a Rapid Antigen Test for the Detection of SARS-CoV-2 Infection in Symptomatic and Asymptomatic Patients Evaluated in the Emergency Department: A Preliminary Report.” *Journal of Infection* 82 (3): e14–6.
- Vong, S., and M. Kakkar. 2020. “Monitoring COVID-19 where Capacity for Testing is Limited: Use of a Three-Step Analysis Based on Test Positivity Ratio. WHO South East Asia.” *Journal of Public Health* 9 (2): 141–6.
- Yu, X., and Y. Rongrong. 2020. “COVID-19 Transmission through Asymptomatic Carriers is a Challenge to Containment.” *Influenza and Other Respiratory Viruses* 14 (4): 474–5.
- Zhao, H., X. Lu, Y. Deng, Y. Tang, and J. Lu. 2020. “COVID-19: Asymptomatic Carrier Transmission is an Underestimated Problem.” *Epidemiology and Infection* 148: e116.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/em-2022-0125>).