

Modelling the SARS-CoVid-2 outbreak in Italy: development of a robust statistical index to track disease dynamics.

Mariano Bizzarri (✉ mariano.bizzarri@uniroma1.it)

Sapienza University. Rome, Department of Experimental Medicine, Systems Biology Group Lab;
Sapienza University, Rome, Italy.

Mario Di Traglia

Department of Public Health and Infectious Diseases (Biostatistics Section), Sapienza University, Rome, Italy.

Alessandro Giuliani

Istituto Superiore di Sanità, Environment and Health Department, Rome, Italy.

Annarita Vestri

Department of Public Health and Infectious Diseases (Biostatistics Section), Sapienza University, Rome, Italy.

Valeria Fedeli

Sapienza University. Rome, Department of Experimental Medicine, Systems Biology Group Lab;
Sapienza University, Rome, Italy.

Alberto Prestininzi

NHAZCA Srl, SpinOff-Sapienza University, Rome, Italy.

Case Report

Keywords: Covid-19; healed/infected ratio (RI); Epidemics Modeling; Intensive Care Units (ICU); Backward Forecast.

Posted Date: July 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-49880/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

COVID-19 pandemic in Italy had a spatial distribution that made the tracking of its time course quite difficult. The most relevant anomaly was the marked spatial heterogeneity of pandemics. Lombardia region accounted for around 60% of fatal cases (while hosting 15% of Italian population); moreover, 86% of fatalities concentrated in four Northern Italy regions. The 'explosive' outbreak of COVID-19 in Lombardia at the very beginning of pandemic fatally biased the R-like statistics routinely used to control the disease dynamics. To (at least partially) overcome this bias, we propose a new index $RI = dH/dI$ (daily derivative ratio of H and I, given H=Healed and I=Infected), corresponding to the ratio between healed and infected patients relative daily changes. The proposed index is less affected than R by the uncertainty related to the estimated number of infected persons and allows to follow (and possibly forecast) epidemic dynamics in a largely model-independent way. To analyze the dynamics of the epidemic, starting from the beginning of the virus spreading - when data are insufficient to make an estimate by adopting a SIR model - a "sigmoidal family" model was introduced. That approach allowed in estimating the epidemic peak using the few data gathered even before mid-March. Based on this analysis, the peak had been expected to occur by end of April. Later on, real data of the epidemic evolution have demonstrated to fit with the predicted values.

The methodology of analysis of the dynamics of the epidemic we are proposing herein aims to forecast the time and intensity of the epidemic peak (forward prediction), while allowing identifying the (more likely) beginning of the epidemic (backward prediction). Finally, we established a relationship between hospitalization in intensive care units (ICU) versus deaths daily rates by avoiding the necessity to rely on precise estimates of the infected fraction of the population (the most difficult, uncertain and expensive data that can confidently be acquired, especially in presence of an elevated number of asymptomatic patients). The joint evolution of the above parameters over time allows for a trustworthy and unbiased estimation of the dynamics of the epidemics.

Introduction

After a novel strain of coronavirus - SARS-CoV-2 - was identified in Wuhan (China)¹, Italy resulted to be one among the most affected countries by Covid-19 epidemic². Noticeably, despite the existence of an already established program to deal with such an outbreak, Italian government was taken by surprise. Moreover, despite warning coming from previous, disregarded clinical reports, clinical and administrative countermeasures lasted until March 2020 before to be adopted in a systematic way.

Predictive mathematical models for epidemics are fundamental to forecast the course of the epidemic and to plan effective control strategies. Unfortunately, it is likely that unreliability of the preliminary data as well as inadequacy/lack of proper descriptive models played a relevant role in undermining political decisions of the Italian administration.

Several models have been developed to describe the Sars-CoV-2 pandemic, by reframing the SEIR model³, using the discrete-time SIR model (which includes dead individuals)⁴, or a control-oriented SIR model that stresses the effects of delays and compares the outcomes of different containment policies⁵. Stochastic transmission models have been applied as well⁶.

The above-mentioned approaches stem from compartmental diffusion processes routinely used in physical chemistry and pharmacokinetics. SEIR stands for Susceptible-Exposed-Infected-Recovery that denote four states that each patient traverses going from susceptible (S), to the exposed (E) and infected (I) conditions ending up into the recovery (R), a compartment that collects both recovered and dead patients. The apparent oddity of putting into the same compartment healed and dead persons comes from the physical-chemistry origin of the model in which R compartment collects all the entities (e.g. molecules) arriving at the final step of the process (end products, excreted or adsorbed substances).

So far so good, but the problem (when the model is applied to real processes) is hidden in the statement 'on average', that embeds the physical condition of 'ergodicity'. In simple terms, this condition implies stochasticity of the contacts and (as for any averaging process) the homogeneity of the process.

Both these conditions are violated in the Italian case. Data highlighted differences of one order of magnitude in both infection dimension and spreading rate between Lombardia (accounting for around 60% of total fatalities and infections) and the rest of Italy. We will not discuss the possible causes of this discrepancy (that, even if less dramatic, involves the whole Northern Italy, in which 86% of the overall Italian fatalities have been recorded), as this task would require a specific clinical/epidemiological inquiry. Instead, we focus on the possible ways to get a sufficiently accurate estimate of the epidemics process when in presence of huge violation of SEIR model premises. A statistical index different from the SEIR- derived ones, could be very useful to assess the entity of an epidemics outbreak.

The need of a different statistical index is still more cogent in the Italian case, considering that the Lombardia outbreak happened earlier than the appearance of Covid-19 cases in other regions, and administrative officials were asked to take decisions based on limited information, which – at that time – was mainly restricted to Lombardia singularity. This situation could be not so rare, thus the need of a more robust statistical index is not limited to the Italian situation. Indeed, during the initial stages of the infection, there is not enough information to predict epidemic temporal development using analytical models, but some useful hints could be extracted as “hidden” information from simple statistics based on very preliminary data. This is why, before entering into the new proposed index derivation, it is worth checking if the epidemic data collected during the first months, affected by both uncertainty and heterogeneity biases, do allow for a reliable forecasting by adopting less demanding models, with respect to SEIR ones. We achieved that goal by means of the Verhulst- Pearle model (sigmoid), which was able to predict the occurrence of the peak of infection, when fed with very early data. We estimated the parameters by regression methods, applying the geometric mean as a function of truly infected individuals. The initial condition for the Bernoulli differential equation was determined applying the back-propagation method⁷.

Then we obtained a novel index (RI) derivation, by splitting the R compartment into healed and dead persons, ending up into the dH/dI (RI) index with dH and dI being the daily ($\text{day}(i)-\text{day}(i-1)$) differential of healed and infected people, respectively. This choice allowed us to detect a clear tipping point in epidemic dynamics in Italy (without considering data from Lombardia), thus recognizing the onset the decline in the epidemic cycle. Such tipping point has emerged only later in Lombardia, and this finding is consistent with the anomalous dynamics of the epidemics in that region.

We complemented the investigation by a concomitant estimate of the clinical severity of Covid-19 disease, which should allow grasping both the potential lethality of Covid-19 and the effectiveness of hospital care. Underestimation of asymptomatic (positive) Covid-19 bearing subjects and uncertainty in detecting the true number of infected people, due to limitation in both availability and reliability of oral-pharyngeal swab data during the first months of the epidemic, severely undermined the trustworthiness of raw lethality data (Case Fatality Ratio, CFR, corresponding to the death/infected ratio), mostly because true incidence resulted underestimated. CFR values relative to different regions convey information on both age distribution (COVID-19 fatal cases are almost completely restricted to patients >70 years old) and efficiency of public health systems⁸. In fact, Italy has the oldest population in Europe and the second oldest population in the world, after Japan. COVID-19 has shown a strong dependence on age due to the severity of the infection and the risk of death⁸.

However, the CFR in Japan is much lower (3.2%) than that recorded in Italy (13.7%), despite their similarities. Therefore, to obtain a more reliable evaluation of clinical evolution, we looked for an independent parameter that could free itself from the estimate of infected individuals, which is affected by a high level of uncertainty as abovementioned. Thus, we focused on the daily number of accesses of Covid-19 patients to the intensive care unit (ICU). Here we show that the number of accesses to the intensive care unit has a very high linear correlation with the number of deaths so that the slope of this relationship is a good faith estimate of the effectiveness of hospital care as well as an unbiased index of the clinical evolution of the epidemic.

The shared rationale of the above-mentioned strategies is the 'blessing of delay': in order a patient is actually registered as 'healed' it is necessary to wait for three negative tests; this creates a time delay with respect to the ascertainment of infection cases. This delay (together with the latent but unescapable correlation between infected and healed frequencies) makes the ratio between daily changes in healed and infected compartments a proxy of the time derivative of infection dynamics only marginally affected by infected number uncertainty. Similar considerations hold for intensive care unit hospitalization and fatalities that in turn do not include any infection number estimate.

Methods

RI index. We rely on official data from the Italian Health Administration, recorded daily for each Italian province (in Italy each region is further divided into provinces)⁹. The time window taken into consideration went from the end of February to 12 May 2020. The division of this time window into "fields" derives both from the need to compare different regional entities and to be consistent with different stages of epidemic curve. Accordingly, the average value of the RI (dH/dI) index was calculated from the healed and infected curves within each field (Table 1). The fields correspond to T0 = starting point, with a very low number of infected and almost no healed patients. T1 = initial exponential increase in the epidemic phase, featuring a marginal number of infected compared to healed patients. T2 = inflection of the curve of infected patients, corresponding to the beginning of the exponential increase of fatal cases. T3 = plateau, the phase of the curve of infected patients showing a trend toward reduction with simultaneous exponential increase in the number of healed individuals. T4 = drastically negative drop among infected patients tendency of the infected and exponential increase in the healed. The time windows were calibrated on the total of the Italian regions (excluding Lombardia data), while Lombardia was singularly considered. It is worth noting the Italian dynamics of RI was super- imposable to the dynamics relative to other countries (data not shown).

It is worth remarking that the infected, dead and healed compartments are non- overlapping, as each patient is considered to be healed only after three negative PCR results performed across a period of three weeks. The robustness to the uncertainty of the infected number of the proposed index derives from both the use of the daily differential (instead of the total number of infected patients), and the long delay between the infection and the declaration of "complete recovery", which is obtained only after three PCR negative test. In the initial stages of the infection, there is not enough information to predict epidemics temporal development using too demanding models. Thus, we approached the forecast of time and intensity of the peak of infection through the Verhulst-Pearle sigmoid model (Fig.3). The derivative of the sigmoid indicates (in the peak) the inflection point of the epidemic curve (Fig.3a) and describes the change of speed in the spread of the epidemic. This change, in Italy, occurred around the end of March. The parameters of the model were estimated by applying the geometric mean as a function of actually infected. The parameters, a and b , of this function are estimated by regression methods. The initial condition for the Bernoulli differential equation, $N(0) = 224$ and $N(l) = 8514$, (where l is the time delay) was determined applying the back-propagation method ⁷

Forward forecasting model. The model relies upon the following estimated equation:

$$N(t) = (8514 * a) / (\exp(-a * t) + b * 224)$$

Being $N(t)$ = number of actually infected at time t , 224= number of infected peoples at February 24, and 8514 the infected peoples at March16.

The model shows a delay effect (autocorrelation) of 18-22 days in the epidemic curve (224= number of infected peoples at February 24 and 8514 the infected peoples at March16. Model fitting was based on

the data relative to the interval of time February 24 – March 16, and the coefficients (with 95% confidence limits) were:

$$a = 0.1466 (0.1401, 0.1531)$$

$$b = 5.173e-05 (4.605e-05, 5.741e-05)$$

Goodness of adaptation:

Square R: 0.9307 Adjusted square R: 0.9294 RMSE: 1.057e + 04

The model fitted with the data until 24 March was demonstrated to correctly predict the onset of the end of April peak. The confidence interval, $(1-\alpha) = 0.95$, indicates the fluctuation of the real amount of daily infected people during the peak (LI = 100000, UI

= 110000). The real data was around 108000 infected people (as predicted by the model).

Backward prediction model. The first available data on the number of daily infected in Italy was available on February 24, 2020. In the first ten days of registration, epidemiological data resulted highly “unstable” and untrustworthy. Nevertheless, they can still capture the dynamics of the phenomenon. To eliminate excessive fluctuations, a four terms moving median was applied to the data. Furthermore, on the levelled series, the geometric mean of the rates of change was calculated. This average is 0.02354 (net rate). Finding the evolutionary dynamics of the number of infected, at the beginning of the exponential branch and coming from the asymptotic branch, the formula of compound capitalization is applied. We indicate with t the day

February 24 and with j the number of days before t . The daily infected peoples at time $t-j$ is given as follows:

$$N(t-j) = N(t) / (1+a)^j$$

Where:

$$N(t-j) = 202 / (1 + 0.02353808)^j$$

$N(t)$ = levelled number of infected at day t , α = geometric mean of the in the first 10 days and j =numbers of days before t .

For $j = 130$, we get $N(t-j) = 10$; from February 24, subtracting 130 days we arrive in mid-October.

This apparently paradoxical early beginning of epidemics obtained an independent biological proof by the genetic distances among different Sars-CoV2 isolated strains¹⁰.

Results

Epidemics trend in Italy and Lombardia region.

Table 1 shows the dynamics of the RI index (dH/dI , calculated as the average value at different time intervals both in Lombardia and in the rest of Italy). The different time intervals ("fields") correspond to the peculiar behaviour of infectious dynamics, analyzed for Lombardia and for the rest of other Italian regions. Raw data have been provided by daily sampling made by Italian Health Administration and are affected by biases due to non random sampling and test uncertainty (especially important in the first phases of the epidemic).

With a simple glance at Table 1, it is possible to appreciate the differences between Lombardia and the rest of the Italian regions. The difference is evident from the average values of $RI = dH/dI$ calculated for each field, between T_0 and T_5 . The first field highlights an RI greater in Lombardia with respect to the rest of Italy tracking the start of the epidemic process in this region. Consistently, the much higher percentage of patients healed in Lombardia region in the initial periods suggests the need to bring the actual start of the process back in time compared to the official registration of the first cases. Going to more recent times (successive intervals) is evident as the index reaches a plateau at window T_3 for 'rest of Italy', indicating the end of the cycle of epidemic, while still floating in Lombardia, with a large dispersion of the RI value. This is still more evident in Figures 1A and 1B.

Figure 1A reports daily changes since 24 February 2020 (T_0), of infected, and healed in Italy without Lombardia. Top panel reports RI (dH/dI) index while bottom panel refers to the temporal trend of healed and infected cases. The analysis of single sections of the time course suggest some interesting observations. The average RI value in Field 1 T_0 - T_1 is near to zero; there is only a slow increase of infected (it must be remembered the low number of swap analyses performed in this period) with almost non-detectable increase of healed cases. The average RI value in the Field 2 T_1 - T_2 is on the rise, but lower than 1 due to the temporal delay between healed and infected cases generation. The infection rate rises, and so does healed curve but at a lower rate. The average RI value reaches a 'tipping point' marking a phase transition¹¹ of the dynamics at Field 3 (T_2 - T_3). The healed daily variation overcomes the infection rate and the RI has a huge variance, another signature of phase transition¹¹. In this time interval, the Infection rate reaches its plateau while healed curve displays a neat acceleration. The average value RI assumes an average value >1 in the Field 4 (T_3 - T_4), where a significant departure of the rate in between healed and infected patients can be observed. Finally, in the Field 5 (T_4 - T_5), the dH/dI ratio assumes an average value <0 . From this point onward, dH/dI is no more relevant given the negative rate of infected persons and thus the negative values of dH/dI index. This is a further proof of the phase transition started in the previous intervals and identified by infection peak. It is worth noting how the two extremes of the section (3 and 16 April) are coincident with the reduction of the absolute number of infected persons. This behaviour coincides with the plateau of our forward model based on a completely independent approach.

Figure 1B relates to the Lombardia region and tells us a different story than the rest of Italy. First, we note that, in contrast to the rest of Italy, there is no clear "critical point". The RI does not assume the average

value >1 for the entire range of data examined (82 days from T0 corresponding to 12 May 2020), while in the rest of Italy the number of recovered people is very high, with $RI > 1$ starting from the 40th day (April 2, 2020). The Lombardia region continues to oscillate around an average index of $RI=0.65$, without reaching a stable plateau or going towards stable negative territories. All in all, the proposed index tells us of the reaching of a tipping-point at day 45 from the beginning of analysis (in the rest of Italy) that marks the end of epidemic cycle being the fatalities after that point the 'echo' of previous infections. Lombardia differs from the previous sketched scenario as it displays a completely different dynamics of the index.

Parameters for clinical severity estimation

The aforementioned Japan/Italy paradox, where two nations with largely super-imposable demographic structures have huge differences in lethality trends, asks for something different from the usual lethality indexes. We decide to remove the principal cause of uncertainty, i.e. the rate of infected people, by focusing on different observables. That goal was achieved by relying on the relation between the number of accesses in intensive care units (ICU) and the number of deaths. Both these variables are known with no (or very little) uncertainty and they give an immediate estimate of the severity of the disease (Fig.2A, B).

Figures 2C and 2D reports the evolution of ICU as well as the Covid-19-related fatalities recorded on a daily basis in Italy, excluding (Fig.2C) and including Lombardia data (Fig. 2D). Figure 2C shows the parallel behavior of the number of accesses to intensive care and the number of deaths for the entire country. The two parameters are strictly correlated (Pearson $r = 0.94$), and the slope of the curve is equal to 0.19 pointing to an 81% of success for intensive care (Figure 2a). This relation points to a constant 'clinical relevance' of the disease throughout the entire time course.

The figure refers to the same day for both death numbers and ICUs and is worth noting that the fit comes from the shared trend of the two variables both registering the same epidemic spread. As a matter of fact the 'lagged correlations' (i.e. the correlations between number of deaths and ICUs, three, four, five and six days before) displayed a significantly lower correlation coefficient. The fact that ICUs- Fatalities relation is driven by the epidemics dynamics (the relation practically disappears when lagged correlations are computed) tells us that this relation can be considered as a proxy of the epidemics dynamics more than a purely clinical descriptor. Figures 2C and 2D describe the simultaneous evolution over time of the two parameters – ICUs and death – evidencing how the rise and the subsequent decline of the epidemics can aptly be depicted with the help of these observables. Keeping in mind that fatalities have a variable delay from hospitalization, the ICUs- Dead number function is equivalent to a temporal differential rate of infection scaled by a largely constant intrinsic fatality rate. At odds with other strategies, the ICUs- Dead number function does not imply any direct infection estimate.

Figure 2B refers to Lombardia, and again we observe a strict relation between the number of daily accesses to intensive care units and deaths. It is worth noting that while for the rest of Italy the slope was 0.19, the slope relative to Lombardia is quite higher (0.28) pointing to a decreased success of care dropping from 81 to 72%. By a closer look at Fig.2B is evident how the right part of the plot

(corresponding to the peak of the epidemics with higher number of both fatalities and intensive care accesses) has a higher displacement of observations from the regression line with a tendency to score a higher number of fatalities.

This asymmetric displacement could be the image in light of the initial overcrowding of the hospitals in Lombardia that first faced the initial impact of the epidemic. In any case, the relation between hospital accesses and deaths can be considered as a useful parameter for monitoring disease severity. The corollary of this fact is that the number of healed individuals properly reflects the epidemics evolution in a clinically reasonable way, allowing overcoming bias provided by uncorrected data collection about incidence and lethality. Accesses to critical care units reach a peak at mid- March and then regularly decreases, reaching in these days (July 2020) very limited numbers (<50 patients in respect to ~4000 at mid March). We hypothesize that this behaviour reflects the improvement in treatment of patients that are followed at home or in standard hospital regimen, and does not imply any change in the viral pathogenicity, at least in the short period we have investigated.

These findings bring some relevant consequences: a) despite infection and death rates values being untrustworthy, recording of critical care units accesses rates can provide a useful estimate of epidemic dynamics, especially if we aim to appreciate the most relevant medical outcomes; b) decrease in critical care units accesses rates mirrors the ones we observed for death rates.

Given the reduction in hospitalization in intensive care structures is mostly ascribed to improved efficiency of medical treatments delivered at home or during standard hospitalization, it can be reliably surmised that early and efficient pharmacological therapy can successfully reduce both accesses to critical care units as well as death rates. While the reasons behind the (apparent) high incidence of fatality rates in Lombardia still remain uncovered, we can exclude that differences in virus pathogenicity could provide an affordable answer. The impressive trend we observed in Lombardia from the outset (end of February 2020) simply demonstrated that, since the beginning, the number of (asymptomatic) infected individuals reached a “critical” mass that ultimately caused a sudden “eruption”. Indeed, almost half of the Covid-19 cases registered in Italy occurred in Lombardia.

The Verhulst-Pearle sigmoid model (sigmoid) (Fig. 3A, B) allowed us to reconstruct the time and intensity of the epidemics peak and to highlight the “hidden” initial period of the epidemic (backward prediction). Namely, we have been able in dating back the beginning of epidemics some months before the first ascertained case. This occurrence, that recently has been acknowledge by new investigations¹², tells about an ‘hidden circulation’ of the virus in Lombardia long ago before any official evidence that could be one of the concurring causes of Lombardia singularity. On a more methodological ground it is worth noting how the sigmoidal model was able to correctly estimate the time course of the derivative of epidemic spread (the same information conveyed by SEIR-based R-like statistics) .when fed of only early (and very uncertain) data before 24 March 2020 (Figure 4).

Conclusion

As aptly remarked in a paper recently appeared in Nature 20, R-like statistics are in many cases imprecise estimates of epidemic dynamics that both rest on unverified assumptions and do not capture the current status of an epidemic spiking up and down when case numbers are low. Moreover, they are out of scope in situations with marked spatial heterogeneity. Other measures (trends in new infections rate, deaths and hospital admissions, cohort surveys) could provide a more reliable estimate of how many people in a population currently have the disease (or have had it). Overall, R index is “too large, and is being used for purposes for which it was never intended”²⁰. Here we set some feasible alternatives fixing the above fallacies by checking some of the above-mentioned 20 alternative strategies.

Indeterminacy of the estimate of affected people, by means of a general underestimation of the asymptomatic pool, might have biased the models on which official statistics usually rely. Furthermore, the fact that data on the infection spreading have been mostly obtained through Real-Time RT-PCR diagnostic – a tool burdened by a huge rate of unreliable results 13,14 – is an additional source of uncertainty. It is noteworthy that serological-based survey have evidenced incidence rates 40-80 fold higher than those suggested by previous estimates, based on RT-PCR¹⁵.

Another source of bias is linked to the fact that the beginning of the epidemic has been overlooked by missing a non-negligible number of affected patients (especially in the time lapse of November 2019 – January 2020). Tracing back the early starting point of the Sars-CoV-2 outbreak is a difficult task, besides some attempts have been done in that field¹⁰; our results suggest a sensible way to achieve this goal by means of sigmoidal models much less demanding than classical SEIR ones.

The tracing back the beginning of the epidemic to October 2019. Beside the genetic distance analysis ¹⁰, indirect evidence has been recently provided by mass media relating the emergence of “strange” interstitial pneumonitis during the fall and evidence is mounting that Covid-19 spreading should be backdated also in other countries, including China. This reconstruction casts more than a shadow upon the current narrative and require an in depth investigation of the true origins of Covid-19 epidemic.

Uncertainty affects even lethality rate estimation. Lethality (or Case Fatality Ratio, CFR) is a measure critically dependent on the accuracy of incidence estimation. It is commonplace to explain differences recorded in fatality rate registered in different countries by advocating differences in demographic structure or in the efficiency of public health systems. Indeed, Italy has the most elderly population in Europe and the second most elderly population in the world after Japan. COVID-19 has a strong age dependence for the severity of the infection and the risk of death. However, lethality rate in Japan is far lower (3,2 %) than that recorded in Italy (13,7 %)¹⁶, the similarities in between the two countries notwithstanding. Moreover, the overall burden of casualties ascribed to Sars-CoV-2 might be easily flawed by the choice of exclusion/inclusion criteria. Indeed, it is difficult to differentiate between deaths in which Covid-19 is detectable from deaths caused by SARS-CoV-2 infection because the vast majority of patients who died in Italy had one or more major pathologies (98.8% with at least 1 comorbidity, and 48.6% having 3 or more diseases) that contributed to the death¹⁷. Additionally, the death trend also reflects a number of intertwined factors and represents the ultimate terminus of a complex process that started much

earlier. Therefore, the simple daily death recording unlikely could provide a significant picture of the epidemic dynamics. Overall, these factors are among those that contributed in undermining the effectiveness of epidemic management¹⁸. Indeed, in “the absence of prevalence and incidence data, including the results of serology testing, it is difficult to predict the effects of specific major public health decisions”¹⁹.

To derive easy to grasp but reliable estimates of the time course of epidemics, here we propose the use of a new index (RI) that allows to detect tipping point (and then the change of pace of epidemics going to the end of its cycle) in a reliable way.

In the same time, the relation between fully recovered individuals and deaths could represent a trustworthy and unbiased estimator of the epidemics process. It is worth of interest that, when the number of accesses is plotted against the number of Covid-19-related casualties, we found a near to unity relation between these two parameters largely invariant in time and space (i.e., Italian regions display the same trend without appreciable differences).

This strong and invariant relation makes trends of care unit’s occupancy predictive of epidemic extinction, with an accuracy higher than that offered by rate of infected or recovering patients, which are biased by uncertainty of collected data and regional- related fluctuations.

Declarations

ACKNOWLEDGMENTS

We wish to thank prof. John Ioannidis from Stanford University for his continued interest to our work and his very useful suggestions.

All authors (Mariano Bizzarri, Mario Di Traglia, Alessandro Giuliani, A. Vestri, Valeria Fedeli and Alberto Prestininzi) that authored the manuscript “Modelling the SARS-CoVid-2 outbreak in Italy: development of a robust statistical index to track disease dynamics” consented to publish: No approval of ethic committee is needed, according to the rules of our institutions. No patients were enrolled in this study
Competing interests: The authors declare no competing interests.

References

1. Velavan, T. P. & Meyer, C. G. The COVID-19 epidemic. *Trop. Med. Int.* **25**, 278–280 (2020).
2. Remuzzi, A. & Remuzzi G. COVID-19 and Italy: what next?. **395**(10231):1225-1228 (2020).
3. Lin, Q., Zhao, S., Gao, D. et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis.* **93**:211-216 (2020).

4. Anastassopoulou, C., Russo, L., Tsakris, A. & Siettos, C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*. **15**(3):e0230405 (2020).
5. Casella, F. Can the COVID-19 epidemic be managed on the basis of daily data? Preprint at <https://arxiv.org/abs/2003.06967> (2020).
6. Wu, J. T., Leung, K., Bushman, M. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med*. **26**(4):506-510 (2020).
7. Snyman, J. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New GradientBased Algorithms*. Springer Science & Business Media (2005).
8. Banerjee, A., Pasea, L., Harris, S. et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet*. **395**(10238):1715-1725 (2020).
9. Epicentro (ISS). <https://www.epicentro.iss.it/en/coronavirus/>
10. van Dorp, L., Acman, M., Richard, D. et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol*. **83**:104351 (2020).
11. Scheffer, M., Carpenter, S. R., Lenton, T. M. et al. Anticipating critical transitions [published correction appears in *Science*. *Science*. **338**(6105):344-348 (2012).
12. Cereda, D. et al. The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv:2003.09320v1* (2020).
13. Gallagher, J. Are coronavirus tests flawed? <https://bbc.com/news/health-51491763> (2020).
14. Carver, C. & Jones, N. Is there any significant difference in sensitivity of COVID-19 virus (SARS-CoV-2) tests based on swabs from oropharyngeal (OP) vs nasopharyngeal (NP) sampling vs both? *CEBM research* (2020).
15. Bendavid, E. et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv*. doi: <https://doi.org/10.1101/2020.04.14.20062463> (2020).
16. Registered at May 3, 2020
17. Characteristics of COVID-19 patients dying in Italy. Istituto Superiore di Sanità. <https://epicentro.iss.it/en/coronavirus/sars-cov-2-analysis-of-deaths>. Accessed April 1, 2020.
18. Ioannidis, J. P. A. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. *Eur J Clin Invest*. e13223 (2020).
19. Boccia, S., Ricciardi, W. & Ioannidis, J. P. A. What Other Countries Can Learn From Italy During the COVID-19 Pandemic. *JAMA Intern Med*. (2020).
20. . Adam, D. A Guide to R-The Pandemic's Misunderstood Metric. *Nature*., **583**: 346-348 (2020).

Tables

Lombardia region				Italy without Lombardia region			
Field	Time intervals	RI Index average value	Note	Field	Time intervals	RI Index average value	Note
0	T_0-T_1	0,04	$I>H$	0	T_0-T_1	0,01	$H\geq 0$
1	T_1-T_2	0,46	$I>H$	1	T_1-T_2	0,07	$I>>H$
2	T_2-T_3	0,47	$I>H$	2	T_2-T_3	0,38	$I>H$
3	T_3-T_4	0,29	$I>H$	3	T_3-T_4	2,21	$H>I$
4	T_4-T_5	0,62	$I>H$	4	T_4-T_5	<0	$H>>I$
5	$>T_5$	0,65	final data refer to May 12, 2020, 82th day from the look down	5	$>T_5$	<0	final data refer to May 12, 2020, 82th day from the look down

Figures

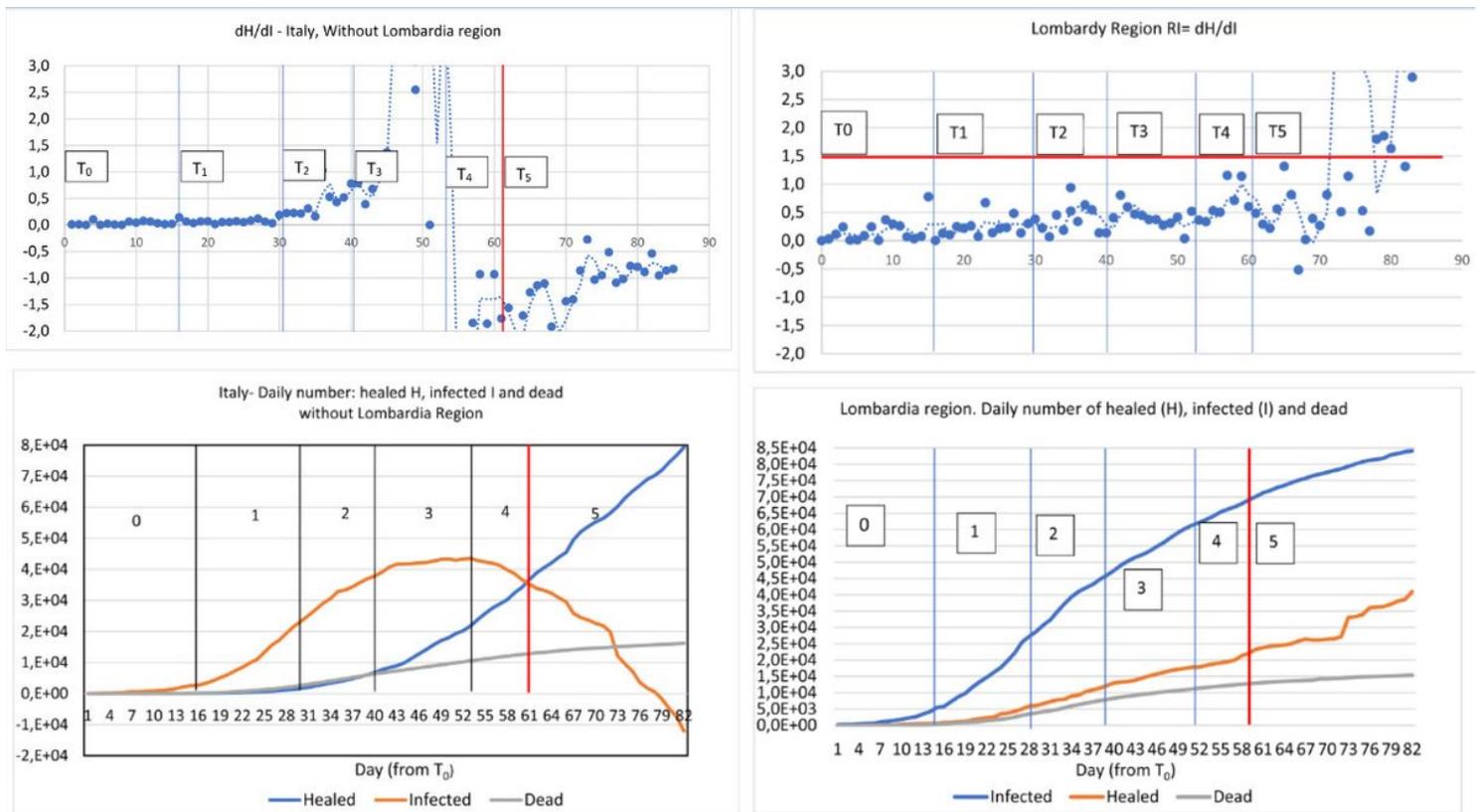


Figure 1

(Left) A reports daily changes since 24 February 2020 (T0), of infected, and healed in Italy without Lombardia. Top panel reports RI (dH/dI) index while bottom panel refers to the temporal trend of healed and infected cases. The analysis of single sections of the time course suggest some interesting observations. (Right) B relates to the Lombardy region

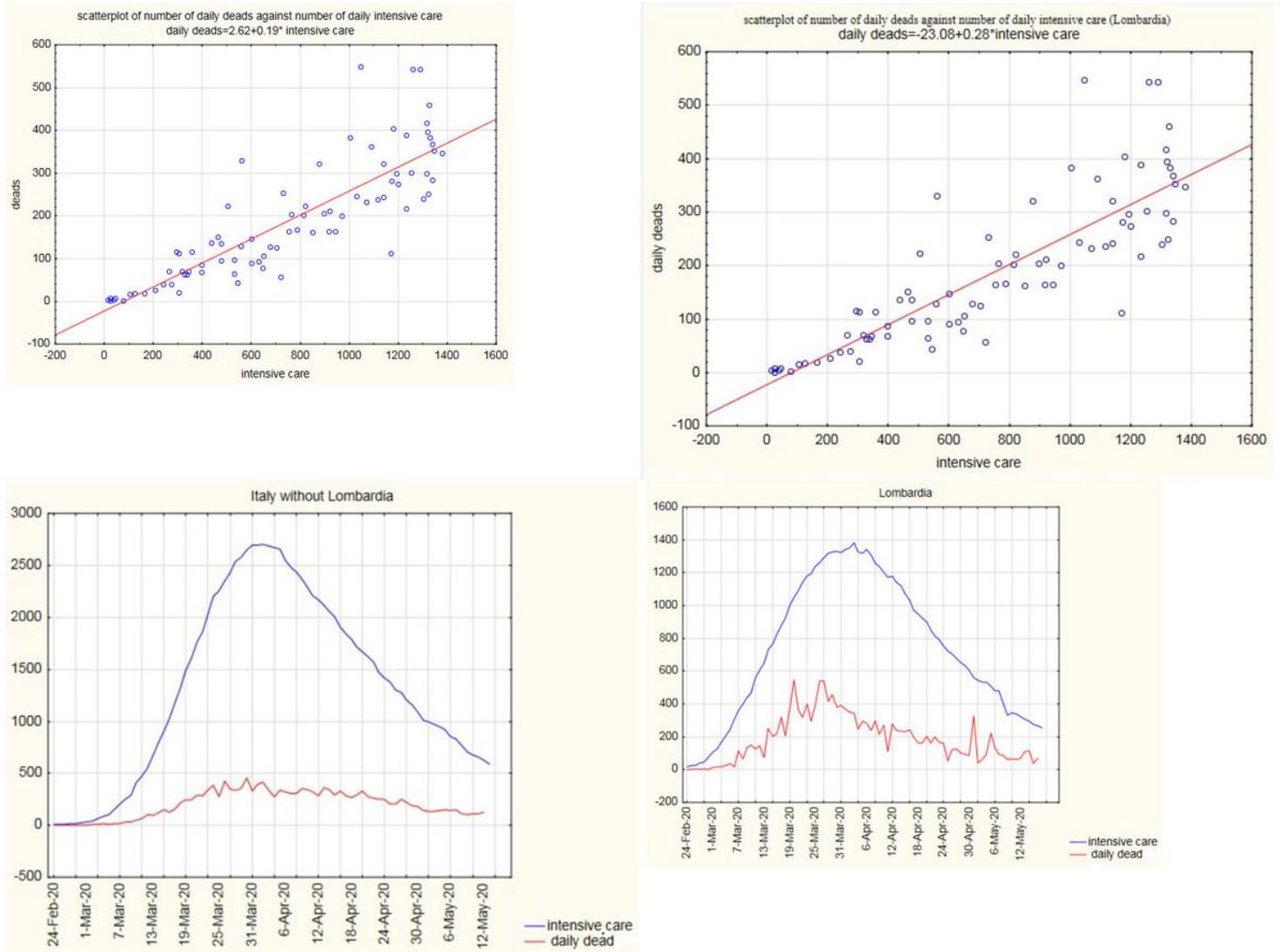


Figure 2

A (top left), B (top right) Estimates the severity of the disease, C (bottom left), D (bottom right) reports the evolution of the ICU

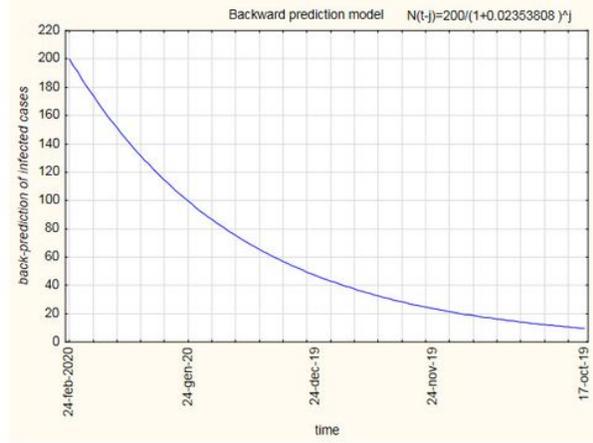
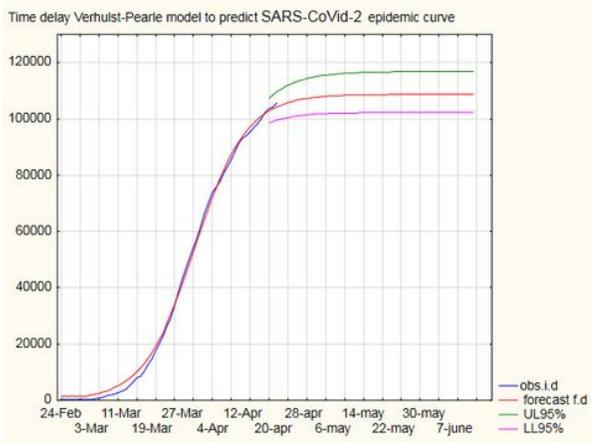


Figure 3

A (left), B (right) The Verhulst-Pearle sigmoid model illustrates the time and intensity of the epidemics peak.

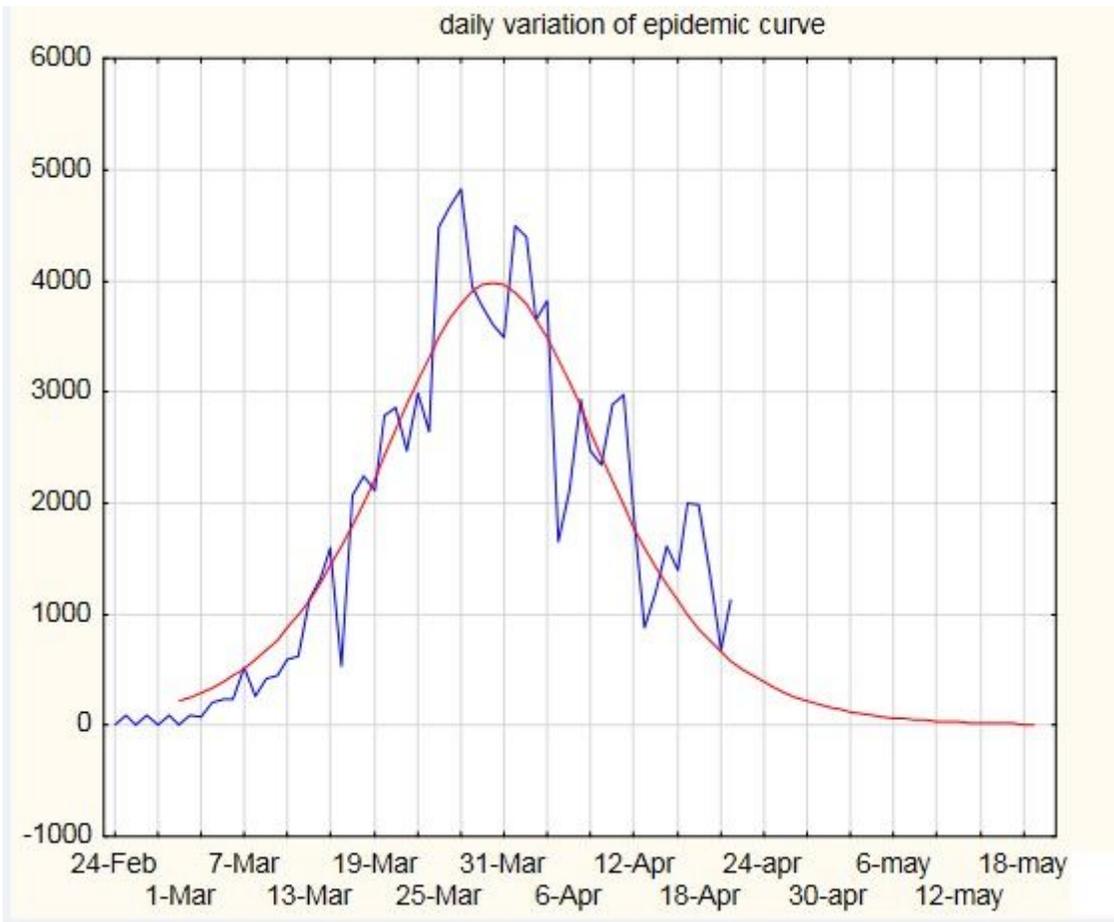


Figure 4

Early data before March 24, 2020