

## Covid-19 Outbreak Progression in Italian Regions: Approaching the Peak by March 29<sup>th</sup>

Cosimo Distante<sup>1</sup>, Prisco Piscitelli<sup>2,3</sup> and Alessandro Miani<sup>3,4</sup>

<sup>1</sup>CNR ISASI, Institute of Applied Sciences and Intelligence Systems, Lecce, Italy

<sup>2</sup> Euro Mediterranean Scientific Biomedical Institute (ISBEM), Bruxelles, Belgium

<sup>3</sup> Italian Society of Environmental Medicine (SIMA), Milan, Italy

<sup>4</sup> Department of Environmental Science and Policy, University of Milan, Milan, Italy

### Corresponding Author:

Cosimo Distante CNR ISASI Unit of Lecce, Institute of Applied Sciences and Intelligence Systems, Via Monteroni 72100 Lecce, Italy; Tel.08321975300; email: [cosimo.distante@cnr.it](mailto:cosimo.distante@cnr.it)

### Abstract

**Background:** Italy and especially the Lombardy region is experiencing a heavy burden of Covid-19 infection. The peak of the epidemics has not yet been reached and it is expected to be delayed in Central and Southern Italian regions compared to Northern ones. We have modeled the Covid-19 outbreak progression in Italian Regions vs. Lombardy. **Methods.** In our models, we have estimated the basic reproduction number ( $R_0$ ) -which represents the average number of people that can be infected by a person who has already acquired the infection - both by fitting the exponential growth rate of the infection across a 1-month period and also by using day by day assessment, based on single observations. We used the susceptible–exposed–infected–removed (SEIR) compartment model to predict the spreading of the pandemic in Italy. **Results.** The two methods provide agreements of values, although the first method based on exponential fit should provide a better estimation, being computed on the entire time series. Taking into account the growth rate of the infection across a 1-month period, in Lombardy each infected person has involved other 5 people (4.94 base on data of March 22<sup>nd</sup> vs. 5.07 based on data of March 19<sup>th</sup>) compared to a value of  $R_0 = 2.68$  reported in the Chinese city of Whuan. According to our model and Piedmont, Veneto, Emilia Romagna, Tuscany and Marche reach an  $R_0$  value up to 4. The  $R_0$  is 3.7 for Lazio and 3.6 for Campania region, where this latter shows the highest value among the Southern Italian regions, followed by Apulia (3.5), Sicily (3.4), Abruzzo (3.4), Calabria (3.1), Basilicata (2.5) and Molise (2.4). The value of  $R_0$  is decreasing in Lombardy and Northern Regions, while it is increasing in Central and Southern Regions. **Conclusion.** The expected peak of SEIR model can be forecast by the last week of March at national level, and by the first weeks of April in Southern Italian Regions. These kind of models can be useful for adoption of all the possible preventive measures, and to assess the epidemics progression across Southern regions as opposed to the Northern ones.

### Epidemiological figures

According to the Italian National Institute of Health (ISS), at the date of March 19<sup>TH</sup> in Italy there were 35731 people who were tested positive to Wuhan novel coronavirus (2019-nCoV). Men were 20686 and women were 14378. Women are the majority of cases in the age group 20-29 (n=732 females vs. 603 males) and >90 years old (n=694 females vs. 397 males); women and men are equally distributed between 30-39 and 40-49 years old (n=1202 females vs. 1261 males, and n=2123 females vs. 2186 males, respectively), while men represent the majority of cases in all the other age groups (n=118 vs 84 from 0 to the age of 9; n=150 vs. 117 between 10 and 19 years old; n=3775 vs. 2926 from 50 to 59 years of age; n=4221 vs. 2025 from 60 to 69; n=4702 vs. 2290 from 70 to 79; n=3158 vs. 2100 from 80 to 89).[7] Regional figures are available up to March 23<sup>rd</sup> and show that about 18910 out of 50418 currently

positive people were detected only in Lombardy (the region of Milan), followed by Emilia Romagna (n=7220), Veneto (n=4986), Piedmont (n=4529), Marche (n=2358), Tuscany (n=2301), Trentino Alto Adige (n=1602), Liguria (1553), Lazio (n=1414), Campania (n=929), Apulia (n=862), Friuli Venezia Giulia (n=771), Sicily (n=681), Abruzzo (n=605), Umbria (n=556), Val d'Aosta (n=379), Sardinia (n=343), Calabria (n=280), Basilicata (n=89) and Molise (n=50). [7]

A total of 20692 symptomatic people were hospitalized at the same date of March 23<sup>st</sup> in Italy, with Lombardy accounting for 9266 hospital admissions, followed by Emilia Romagna (n=2846), Piedmont (n=2194) and Veneto (n=1206). Only Tuscany (n=838), Marche (n=882), Liguria (n=761) and Lazio (n=718) recorded more than 250 hospital admissions at regional level, while other regions remain still lower. At the present, 3204 patients are assisted in intensive care units (1183 in Lombardy, 343 in Piedmont, 281 in Veneto, 276 in Emilia Romagna, 238 in Tuscany, 148 in Marche region and 133 in Liguria. Except Campania (n=110) and Lazio (n=96), all the other regions of Central and Southern Italy, at the moment have less than 50 patients admitted to the ICUs of their regional healthcare systems. [7]

On March 23<sup>rd</sup>, total deaths were 6077 at national level (more than the overall 3264 deaths observed in China), with 3776 in Lombardy, 802 in Emilia Romagna, 315 in Piedmont, 212 in Liguria, 203 in Marche region, 192 in Veneto, 109 in Tuscany, 70 in Trentino Alto Adige, 63 in Lazio, 54 in Friuli Venezia Giulia, 49 in Campania, 38 in Abruzzo, 37 in Apulia, and less than 15 in the other six regions (Figure 1). [7] Lethality rates seems to increase with age and it is higher in males: 0% from 0 to 29 and 0,5% between 30 and 49 years of age; 1.2% in the age group 50-59 (0.6% in women and 1.7% in men); 4.9% from 60 to 69 years old (2.8% in women and 6.0% in men); 15.3% from 70 to 79 (10.7% in females and 17.8% in males); 23.2% from 80 to 89 years old (19.1% and 26.4% in men). [7]

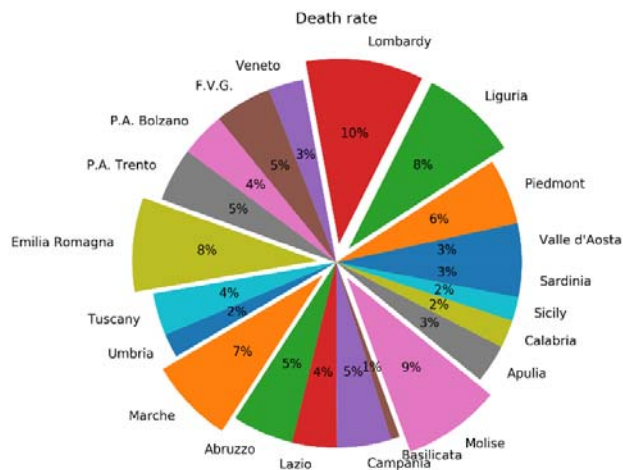


Figure 1 Death rate in Italy per region at the date of March 23<sup>rd</sup>.

Based on these figures, it is clear that the Covid-19 outbreak is now putting overwhelming pressure mainly on Lombardy and Northern regions of the Po Valley (Padana plain), but the peak of the epidemics has not yet been reached. Until now, Southern Regions seem to be less affected by the Covid-19 infection although a huge number of people - mainly students attending Universities in Northern Italy - came back from Po Valley to their families in the South just in the middle of the outbreak, thus representing a potential factor able to accelerate the spreading of the viral infection.

Here we present an attempt to predict the peak of the outbreak in Italy, which is expected at national level by the end of March, and the different progression of the epidemics in Southern Italian Regions compared to Lombardy.

### **Modeling the Covid-19 outbreak progression in Southern Italian Regions vs. Lombardy**

The basic reproduction number ( $R_0$ ) is an indicator that resumes the average number of people that can be infected by a person who has already acquired the infection.  $R_0$  is a metric of how contagious is the disease and its correct estimation is extremely important for epidemiologists, especially when facing new diseases like COVID-19.  $R_0$  can be computed in different ways. In our models, we have estimated the basic reproduction number ( $R_0$ ) both by fitting the exponential growth rate of the infection across a 1-month period and also by using day by day assessment, based on single observations [1]. This study makes use of the susceptible–exposed–infected–removed (SEIR) compartment model [4] to predict the spreading of the pandemic in Italy. Our efforts could be helpful in the adoption of all the possible preventive measures, and to study of the epidemics progression across Southern regions as opposed to the national trend. This metric can be biased by the optimal estimation of the basic reproductive number  $R_0$  (pronounced R-nought). It must be said that  $R_0$  is important if correlated with weather conditions and that reproductive index is reduced as the air temperature and relative humidity increase, [5] according to the formula:

$$R = 3.968 - 0.0383 \text{ Temperature} - 0.0224 \text{ RelativeHumidity}.$$

This means the transmission of COVID-19 can decrease with the warmer season, and that at least some specific figures of the outbreak in Lombardy and Po Valley can be explained taking into account climatic variables.

### **Modeling basic reproductive number $R_0$ Exponential framework estimation**

The exponential estimation is based on the work of Wu et al [7], which essentially was based on that of Zhao and colleagues [8], where the epidemic curve obey to the exponential growth.

At the date of this study (March 23<sup>rd</sup>2020), the epidemic growth was still near exponential growth, and the fitted model has many inliers data points. The method is based on non-linear least square framework for intrinsic growth estimation, in order to obtain with  $M$  the Laplace transforming the probability distribution of the serial interval of the infection. The estimation is obtained with 100% susceptibility for 2019-nCoV at the early stage in Wuhan as reported in [7]. In Figure 2, the number estimates are computed for the Southern Italian regions and for the initial outbreak region (Lombardy). According to our model, in Lombardy, each infected person has involved other 4.9 people (5.07). The  $R_0$  lowers to 3.6 for Campania region, which shows the highest value among the Southern Italian regions, followed by Apulia (3.5), Sicily (3.4), Abruzzo (3.4), Calabria (3.2), Molise (2.4), and Basilicata (2.2).

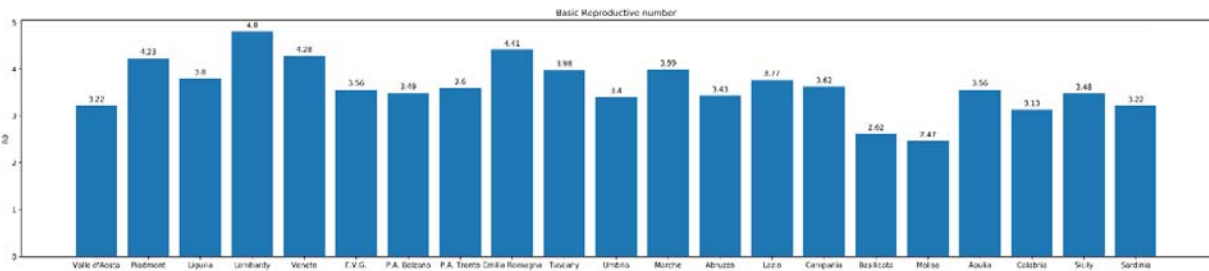


Figure 2. Basic Reproductive index computed with exponential fitting method along the entire time series (1-month period).

### Daily basis estimation of the reproductive number

$R_0$  is an average value, but it can be also computed day by day to monitor the transmission of the infection. Being an average value, it can be skewed by super-spreader events. A super-spreader is an infected individual who infects an unexpectedly large number of people. In Italy this event can be also generated not necessarily by an individual, but from the perturbation of a susceptible population, as it happened in Apulia and Sicily with uncontrolled large group of people coming from outbreak areas. For a “super spreader” individual, such events are not necessarily a bad sign, because they can indicate that fewer people are perpetuating an epidemic. Super-spreaders may also be easier to identify and contain, since their symptoms are likely to be more severe. In short,  $R_0$  is a moving target. Tracking every case and the transmission of a disease is extremely difficult, so the estimation of  $R_0$  is a complex and challenging issue: estimates often change as new data becomes available. If we define the  $Y(t)$  as the number of infected people with symptom at time  $t$ , the exponential growth rate is  $\lambda$ . Let us consider  $\tau$  the generation time (i.e. the serial interval) and  $\tau_0$  the latent or incubation time (values taken from [6]). The infectious time  $\tau_i$ , and the ratio of exposed period to generation time is  $\tau_0/\tau$ . The basic reproductive number can be approximated to:

In order to estimate  $R_t$ , it is important to find  $I$  and then the number of infected people:

Suspects correspond to the number of individuals screened with test which have been confirmed.

Figure 3 and Table 1 show the  $R_t$  estimated values, computed on a daily basis for the Italian regions and for the initial outbreak region Lombardy, where about 40 cases were confirmed out of 100 suspects (Figure 4). The two methods provide agreements of values, although the first method based on exponential fit should provide a better estimation, being computed on the entire time series. From Figure 3, it comes an important aspect with respect to the Wuhan  $R_t$  as reported in [6].

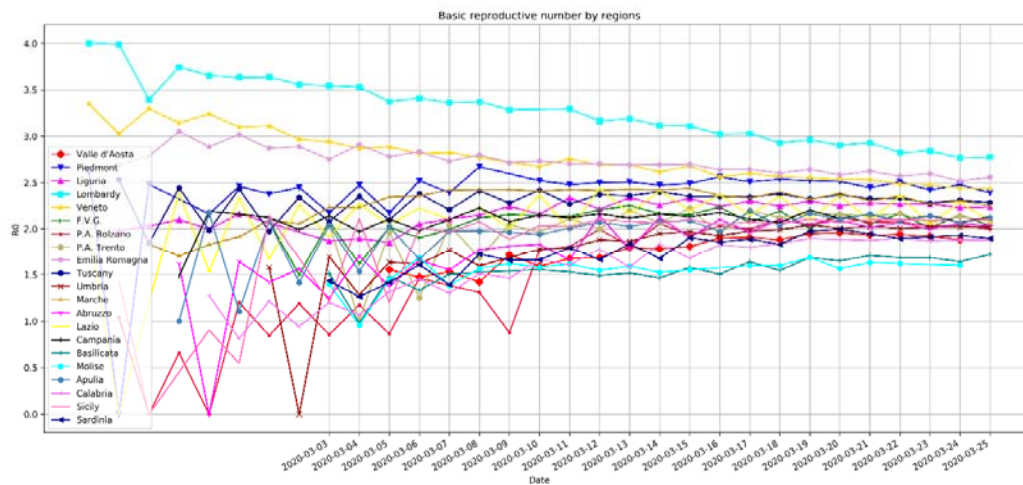


Figure 3 Basic reproductive index computed on a daily basis with the second method investigated in this letter.

Table 1  $R_t$  values of the analyzed regions. Given the trend, min values are usually referred to the beginning of the infection

Region	Estimated	
	min	max
Valle d'Aosta	1.57	2.11
Piedmont	2.36	2.98
Lombardy	2.94	4.27
Liguria	2.02	2.47
P.A. Bolzano	0.84	2.26
P.A. Trento	1.00	2.38
Friuli .V.G.	1.74	2.52
Veneto	2.60	3.64
Emilia Romagna	2.68	3.26
Marche	1.96	2.65

Region	Estimated	
	min	max
Umbria	2.11	2.16
Tuscany	2.11	2.84
Lazio	2.10	3.11
Abruzzo	1.29	2.36
Campania	1.62	2.44
Molise	1.06	1.78
Apulia	1.00	2.32
Basilicata	1.19	1.73
Calabria	0.99	2.02
Sicily	1.45	2.22
Sardinia	1.42	2.12

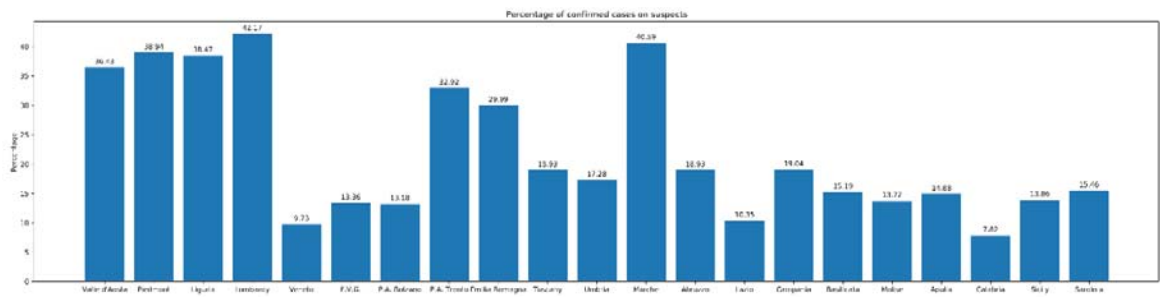


Figure 4 Percentage of suspects confirmed at the time of March 23<sup>rd</sup>. In one day, Lombardy increased by 10% with respect to the previous 24 hours.

### Modeling transmission in Italy

We used the susceptible-exposed-infectious- recovered (SEIR) model [4] to simulate epidemic since it was established on January 2020. It is based on a previous model SIR which was based on three compartments, but since the infection has an incubation period, the compartment E (Exposed) is included. This compartments are modeled over the time, and capture the changes in the population. Let us say that given N the total population, then  $N=S+E+I+R$ , where:

- “S” Susceptible is the portion of population that does not have any vax coverage or immune;
- “E” exposed: is the portion of the population that have been infected, but are in the incubation period that do not infect others;
- “I” Infectious: is the portion of N that is infectious and may infect others, they become dead or may recover;
- “R” Recovered: number of infectious people who have been healed, and become immune.

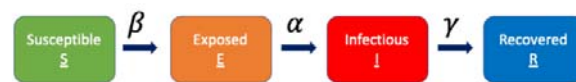


Figure 2 SEIR model with the four compartments and their relationships

This model captures dynamics of these compartments over the time by four ordinary differential equations. One of the most important aspects of these ordinary differential equations is equilibrium, who is achieved by setting to zero their derivatives along the time t. The two equilibriums are: disease-free equilibrium (DFE) and endemic equilibrium (EE).

Besides equilibrium, stability is an issue who is correlated with the basic reproductive number, where if  $R_0 < 1$  DFE is stable; while when  $R_0 > 1$  DFE is unstable and EE is stable. The four equations are:

$$\begin{aligned}\frac{\partial S}{\partial t} &= -\frac{\beta SI}{N} \\ \frac{\partial E}{\partial t} &= -\frac{\beta SI}{N} - \alpha E \\ \frac{\partial I}{\partial t} &= \alpha E - \gamma I \\ \frac{\partial R}{\partial t} &= \gamma I\end{aligned}$$

Where  $\beta = R_0/T_i$ ,  $\alpha = 1/T_l$  and  $\gamma = 1/T_r$  with  $T_i$  and  $T_l$  as defined above, the serial and incubation period respectively. The contact rate  $\beta$  is the rate of infection from an infected individual to one of their susceptible contacts on the unitary time step  $dt$ . The number of individuals transferred from Susceptible state to Exposed state is  $\frac{\beta \cdot S \cdot I}{N} \Delta t$ . The force of infection is defined as  $\beta * S(t)/N$  which is the number of new infections divided the Italian population. At the same time step, there are  $\alpha E(t)\Delta t$  number of cases that are transferred from Exposed to Infectious compartment, and  $\gamma I(t)\Delta t$  number of cases transferred from Infectious compartment to Removed. It is important to state that we assume a closed population, which means the population is fixed, then no births, no deaths or introduction of new individuals. From the above ODE system,  $\frac{d}{dt} [S(t) + E(t) + I(t) + R(t)] = 0$  which means that the population  $N$  is constant at any time step  $t$ :  $S(t) + E(t) + I(t) + R(t) = N$  for any  $t \geq 0$ . The individuals in exposed state is infected but not yet infectious. The population is well mixed, and the model assumes that latent and infectious times of the pathogen are exponentially distributed. In this letter, contact rate  $\beta$  is changing over time as it happens in COVID-19, which increase in early stage due to public unawareness of the disease, then decrease with government control policy measures. The contact rate follow a logistic function trend by estimating it day by day [9]:

$$\beta(t) = \frac{C e^{-\sigma(t+b)}}{(1 + e^{-\sigma(t+b)})^2}$$

With  $t$  the number of days after January 31<sup>st</sup> (the first found cases in Italy),  $\sigma$  a regularization parameter, and  $b$  the bias. A training procedure is performed on the observable data in order to find optimal  $(C, \sigma, b)$ . We have considered as Exposed people, a number of twice Infected people, after lockdown which is in line with the predictions and the observed values. As shown in Figure 6 (red curve), the expected peak of SEIR model is around March 29<sup>th</sup> at national level. It is expected that Southern Italian Regions could reach the peak later, in the first days of April.

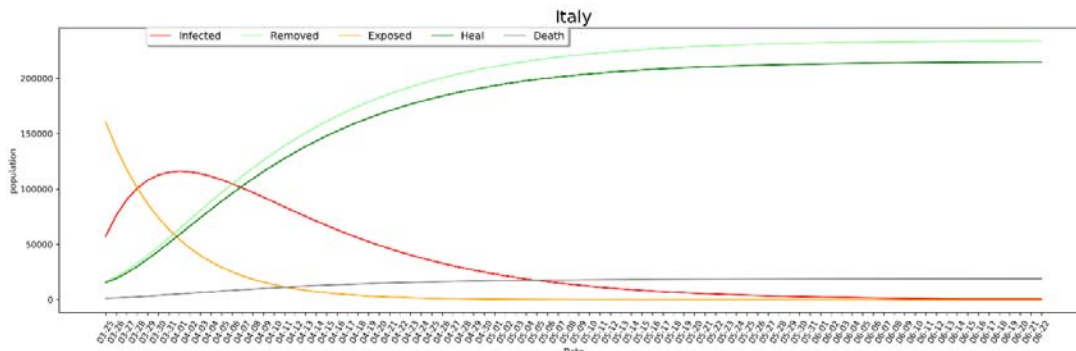


Figure 3 Prediction of the Exposed, Infected and death people. Peak of more than 58K infected people is expected around March 25<sup>th</sup>.

## References

- [1] Italian National Institute of Health (ISS), Coronavirus Surveillance Bulletin March 19/20, available at [https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino%20sorveglianza%20integrata%20COVI-D-19\\_19-marzo%202020.pdf](https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino%20sorveglianza%20integrata%20COVI-D-19_19-marzo%202020.pdf)
- [2] Italian Ministry of Health, daily bulletin Covid-19 outbreak in Italy, available at [http://www.salute.gov.it/imgs/C\\_17\\_pagineAree\\_5351\\_24\\_file.pdf](http://www.salute.gov.it/imgs/C_17_pagineAree_5351_24_file.pdf)
- [3] Zhou T, Liu Q, Yang Z, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *J Evid Based Med*. 2020;13(1):3–7. doi:10.1111/jebm.12376
- [4] Pastor-Satorras R, Castellano C, Mieghem PV, et al. Epidemic processes in complex networks. *Review of Modern Physics* 2015, 87(3): 925-979.
- [5] Wang, Jingyuan, et al. "High Temperature and High Humidity Reduce the Transmission of COVID-19." *Available at SSRN 3551767* (2020).
- [6] Wu, Joseph T., Kathy Leung, and Gabriel M. Leung. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study." *The Lancet* 395.10225 (2020): 689-697.
- [7] Zhao, Shi & Lin, Qianyin & Ran, Jinjun & Musa, Salihu & Yang, Guangpu & Wang, Weiming & Lou, Yijun & Gao, Daozhou & Yang, Lin & He, Daihai & Wang, Maggie. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*. 92. 10.1016/j.ijid.2020.01.050.
- [8] Zhao S, Musa SS, Fu H, He D, Qin J. Simple framework for real-time forecast in a data- limited situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. *Parasites Vectors* 2019;12(1):344.
- [9] Kucharski, Adam J., et al. "Early dynamics of transmission and control of COVID-19: a mathematical modelling study." *The Lancet Infectious Diseases* (2020).