ORIGINAL ARTICLE

COVID-19 outbreak in Italy: estimation of reproduction numbers over 2 months prior to phase 2

Mattia Allieta[1] · Andrea Allieta[2] · Davide Rossi Sebastiano[3]

Abstract

Purpose Two months after its first COVID-19 case, Italy counted more than 190,000 confirmed positive cases. From the beginning of April 2020, the nationwide lockdown started to show early effects by reducing the total cumulative incidence reached by the epidemic wave. Here we provide the reproduction number estimation both in space and in time from February 24 to April 24, 2020 over 2 months into the epidemic.

Methods The aim of the present work was to provide a systematical mapping of the SARS-CoV-2 transmission dynamics spread to all regions of Italy. To do so, we estimated the basic reproduction number (R_0), by using the maximum likelihood estimation method in the early stage of the epidemic. In addition, we determined time evolution of this parameter across the 2 months of the observational period. Finally, we linked R_t, with two indices, the first representing the number of contagious people and the latter the density of susceptibiltiy to infection of people in a region as recorded on April 24, 2020.

Results Our estimates suggest a basic reproduction number averaged over all the regions of 3.29. Based on the SARS-CoV-2 transmission dynamics reported here, we gave a quantitative evaluation of the efficiency of the government measures to lower the reproduction number below 1 (control regime). We estimated that the worst-hit regions in Italy reached the control regime level ($R_t < 1$) in about a month.

Conclusion Our work was carried out in the period between April and July,2020. We found that the mean value of time to reach the control regime across the whole country was about 31 days from February 24, 2020. Moreover, we highlighted the interplay between the reproduction number and two epidemiological/demographic indices to evaluate the "state of activity" of the epidemic, potentially helping in challenging decisions to continue, ease, or tighten restrictions.

Keywords Covid-19 · SARS-CoV-2 · Reproduction number · Time-dependent epidemiology · Transmission dynamics

Introduction

After the first COVID-19 case was diagnosed in Lombardy, Italy, on February 20, 2020 (Carinci 2020), the novel coronavirus rapidly spread across the country leading to a dramatic spike in the number of new positive cases and deaths. To minimize the likelihood that people who were not infected might come into contact with people who had contracted the disease, the Italian government imposed a series of progressively more strict social distancing measures which culminated in a national lock-down announced on March 11, 2020 (Guzzetta et al. 2020).

Around 2 months after the first case and more than 190,000 confirmed positive cases later, from the beginning of April the effect of the nationwide lockdown started to achieve some level of success and the number of new infections began to smoothly decrease (Sebastiani et al. 2020; Sjödin et al. 2020). These early signs of a slowdown of the COVID-19 pandemic in Italy provide a comforting picture of the outbreak's stabilization which is driving the government to periodically review its lockdown measures in view of the so-called "Phase 2" (Sjödin et al. 2020), i.e., the period during which citizens will have to live together with the virus as all the industrial sector, including non-essential economic activities, will start to

✉ Davide Rossi Sebastiano
davide.rossi@istituto-besta.it

1   Ronin Institute, Montclair, NJ 07043, USA

2   Università del Piemonte Orientale "A. Avogadro" Dipartimento di Medicina Traslazionale Anestesia e Rianimazione, Corso Mazzini 18, 28100 Novara, Italy

3   Fondazione-IRCCS-Istituto Neurologico "Carlo Besta", Unità di Neurofisiopatologia, via Celoria 11, 20133 Milan, Italy

reopen. However, since the regional differences in the number of new positive cases have been reported to be huge, with the Northern regions of Italy (namely Lombardia) being most affected, the establishment of the proper precautions to plan "Phase 2" is a truly complicated task.

The planned restrictions and permissions that will be applied could thus vary from region to region. In this context, the systematical estimation of key epidemiological parameters for each region can provide insight into the speed at which the disease had spread and will give a useful tool to figure out if a differential approach at the regional level on the measures to apply for "Phase 2" is feasible to keep down the transmission of SARS-COV2. At the beginning of the epidemic and during the lockdown phase, the Italian Government and the mainstream media have emphasized the relevance of the basic reproduction number (R_0), i.e., the average number of secondary cases generated by a single primary case in a theoretically fully susceptible (100%) population, as the most important and informative parameter to monitor the epidemic trends. Obviously, R_0 has an undoubted relevance since when $R_0 > 1$ the infection may spread in the population, and the more R_0 is large, the deeper would be the interventions needed to control the epidemic. On the other hand, if $R_0 < 1$, on average the infectious individual infects less than one person and the epidemic falls in a so-called "control regime" where it will not be sustained, and it will die out. Nevertheless, R_0 is not the only parameter that affects the impact and the spreading of the disease over a population which may largely result even from several demographic and epidemiological factors.

In this communication, we have provided an estimation of the basic reproduction number R_0 for all the Italian regions by the cumulative confirmed COVID-19 cases continuously updated and made public at the website of Dipartimento della Protezione Civile. In addition, we have estimated the time-dependent reproduction number R_t, which is the average number of secondary cases generated by an infectious individual at time t. We linked R_t related to the last date of our period of observation (April 24, 2020), with two demographic and epidemiologic indices in a simple three-dimensional array in order to highlight the "state of activity" of the epidemic for each Italian region. This provides a useful tool in the management of "Phase 2", potentially helping in challenging decision to continue, ease, or tighten up restrictions.

Material and methods

Demographic and epidemiological data

The official demographic data of the resident population (RP), the surface (S) and the population density (PD) updated on January 1 2019, for each Italian region and Italy were taken from the Italian National Institute of Statistics (Istituto

Nazionale di Statistica, ISTAT) and reported in Table S1 (Supporting Information, data available at http://dati.istat.it/Index.aspx?QueryId=18460). The official data for the COVID-19 epidemic in Italy was taken from the task force of the Dipartimento della Protezione Civile. Cumulative data are available at various aggregation levels, namely national, regional, and provincial, and are accessible on Github (data available at https://github.com/pcm-dpc/COVID-19). Data for the analysis were considered from February 24 to April 24, 2020.

In this period, we collected the daily cumulative number of confirmed positive cases (N), the number of "active" confirmed positive cases (NA), i.e., number of ongoing affected people, and the "density of susceptible-to-infection people" (DS), i.e., the density of people who can still be infected, calculated as (RP-N)/S and expressed, as with the PD, as number of persons/km^2, for each Italian region and Italy.

Estimation of the reproduction number R_0 and estimation of time-dependent reproduction number R_t

To obtain the estimation of reproduction number we use the maximum likelihood estimation (ML) method which assumes that the number of secondary cases caused by an index case is Poisson distributed with an expected value R. Given then observation of (N_0, N_1,..., N_t) incident cases over consecutive time units, R is estimated by maximizing the following log-likelihood function (Obadia et al. 2012)

$$LL(R) = \sum_{t=1}^{T} log\left(\frac{e^{-\mu_t}\mu_t^{N_i}}{N_t!}\right) \quad \text{where} \quad \mu_t = R\sum_{i=1}^{t} N_{t-1}\varphi_i \qquad (1)$$

φ_i is the distribution of the generation time corresponding to the distribution of the serial interval, i.e., the time between when a person gets infected and when they subsequently infect another other people, calculated at time i within the assumption that the incubation period does not change over the course of the epidemic (Britton and Scalia Tomba 2019). We consider that the distribution of the serial interval was expected to follow a gamma distribution with mean (±SD) of 6.50 ± 4.03 days as reported by the Imperial College COVID-19 Response Team Report 13 (March 30, 2020). We note that this value agrees very well with gamma distribution with mean 6.6 days (95% CI, 0.7 to 19) recently determined from the analysis of individual serial intervals in clusters in Lombardia (Italy) (Micheli et al. 2020).

To estimate $R = R_0$, the $LL(R)$ function must be calculated over a period where epidemic curves showed exponential growth. As a first guess, to select this time window we used the simple procedure described by Obadia et al. (2012). In brief, we computed the function over a range of possible time periods by determining the deviance r^2 statistic for each

iteration. Largest r^2 corresponds to the time window over which the ML model best described data.

To evaluate the time dependent reproduction number R_t we adopted the method developed by Wallinga and Teunis (2004). The transmission probability (p_{ij}) of individual i being infected by individual j at t_i, t_j onsets, respectively, can be described mathematically (Obadia et al. 2012) as:

$$p_{ij} = \frac{N_i \varphi(t_i - t_j)}{\sum\limits_{i \neq k} N_i \varphi(t_i - t_k)} \tag{2}$$

The net reproduction number R_j is then sum of all p_{ij} involving j as the infector $R_j = \sum\limits_j p_{ij}$ and it can be averaged over all cases with same date of onset as $R_t = \frac{1}{N_t} \sum\limits_{\{tj=t\}} R_J$.

Finally, since R_t are computed by averaging over all transmission networks compatible with observed incidence data, no assumption is made about the time dependence of the epidemic — unlike, for example the exponential growth in the well-known Bayesian approach (Obadia et al. 2012; Wallinga and Teunis 2004). Hence, we considered this model particularly suitable to estimate the reproduction number in the post-peak period where the transmission is expected to decrease.

All the above data analyses were performed using the R0 package (Obadia et al. 2012) as implemented in statistical software R, 3.6.2 version.

Results

Demographic and epidemiological data

Figure 1 shows COVID-19 incidence in Italy in the period of our observation, together with the dates in which the Italian government imposed restrictions, i.e., social distancing and school closure on March 4, lockdown of Lombardia region and of 15 provinces in northern Italy on March 8, national lockdown of Italy on March 11. Overall, the incidence of COVID-19 infection in Italy shows that the exponential growth period may take place during the first 15–20 days from the national epidemic onset (February 24, 2020).

Table S1-S2 (Supporting Information) show demographic and epidemiological data, respectively. As is widely known, Table S2 shows that the COVID-19 epidemic affected (and is affecting) the northern Italian regions harder, with $N = 16,859$ and $NA = 89,384$ on April 24, 2020, i.e., more than 80% of the cases of the country (with 54.7% of the Italian resident population), if we aggregate epidemiological and demographic data of the northern regions (Lombardia, Piemonte, Veneto, Emilia–Romagna, Liguria, Valle D'Aosta, Trentino–Alto Adige) plus the Marche and Toscana regions. Furthermore, in the Lombardia region the epidemic had a huge spread, with

$N = 71,256$ and $NA = 34,368$ on April 24, 2020, i.e., more than one third of the cases in the country (with 16.7% of the Italian resident population).

Estimation of the reproduction number R_0 and estimation of time-dependent reproduction number R_t

At the top of the panels in Figs. S1–S3 (Supporting Information) we have reported the incidence data for all the regions plus Italy. Initial inspection of the datasets shows again that the exponential growth period may take place during the first 15–20 days from the relative epidemic onset. It should be noted that for the evaluation of R_0 in the initial outbreak stage, we considered data from February 24 up to March 18, 2020, since data in a wider range could be affected by the national lockdown on March 11, 2020.

In Fig. 2a we show R_0 values obtained for SARS-COV-2 in all the regions and in Italy. Table 1 reports the same data represented in Fig. 2a, compared with those obtained by Guzzetta et al. (2020), D'Arienzo and Coniglio (2020), and Distante et al. (2020). According to our ML estimation, the northern Italy regions Friuli–Venezia Giulia ($R_0 = 3.61$), Liguria ($R_0 = 3.68$), Veneto ($R_0 = 3.73$), and Lombardia ($R_0 = 3.88$), presented the highest R_0, suggesting that one infected person will infect up to four other people. The same tendency is shown in the central region Lazio, with $R_0 = 3.62$. Apart from the Southern regions Basilicata ($R_0 = 2.73$), Molise ($R_0 = 2.52$), and Umbria ($R_0 = 2.44$) showing the lowest values, the reproduction number R_0 ranges from 3.00 to 3.49 in the rest of the country. The observed distribution of R_0 related to all the country is reported in Fig. 2b with the mean ($\pm$ SD) value of $\overline{R_0} = 3.29 \pm 0.38$.

The time evolution of reproduction number R_t for each region plus Italy is reported in the lower part of the panels of Figs. S1–S3. All the R_t profiles exhibited similar qualitative patterns, characterized by large initial R_t values followed by the decreasing over time of the parameter reaching $R_t \lesssim 1$ close to the peak of daily incidence ($\sim$ 20–30 days). Below this value, R_t further decreases, featuring a sort of plateau in the $R_t < 1$ "control regime" as detected for the country and for most of the regions. Uncertainty decreases over time because of the increased number of cases, but at the end of the data range considered (50–60 days) the estimates exhibited wide confidence intervals, reflecting a stochastic pattern for those regions presenting small number of cases. In Fig. 2c we presented the distribution of the first day (from the onset February 24, 2020) when the time evolution of R_t converges to $\lesssim 1$ in each region. In Table 2, the same results are listed with the corresponding calendar date together with the median R_t values determined in the last 7-day time window (18–24 April, 2020).
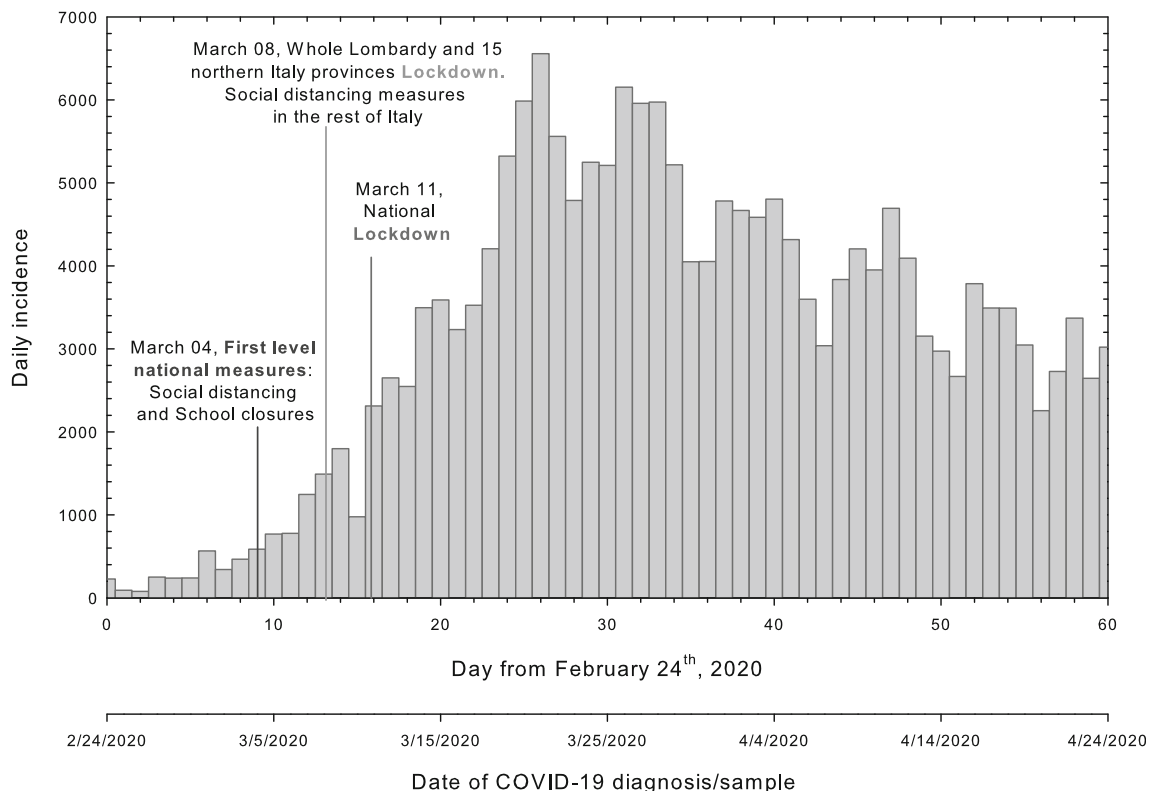
Fig. 1 Daily number of new confirmed cases in Italy of the first 2 months of COVID-19 epidemic. *Bars* are incidence data of COVID-19 in Italy from February 24 to April 24, 2020

Finally, in Fig. 3, median R_t values in the last 7-day time window (Table 2) are plotted in a three-dimensional array as a function of NA and DS as recorded on April 24, 2020.

Discussion

In this work, we analyzed the time evolution of incidence of the SARS-CoV-2 epidemic for 2 months from onset, February 24 to April 24, 2020 in all the Italian regions. We estimated the basic reproduction number (R_0), by using the ML method in the early stage of the epidemic. In addition, we determined time evolution of this parameter across the 2 months of the observational period. Finally, we linked R_t, with two indices, NA and DS, the first representing the number of contagious people and the latter the density of 'susceptible to infection' people in a region as recorded on April 24, 2020.

Firstly, we point out that these data can be considered only an approximation of the actual epidemic dynamics. Indeed, the reported number of cases strictly depends on the number of swabs that are used for Covid-19 testing and can be biased by several factors such as underreporting, delays in recording as well as errors in classification of cases (Micheli et al. 2020). Therefore, large data noise is generally observed, especially at the regional level, which requires a careful inspection of the

epidemic curve as well as data smoothing in order to avoid unrealistic reproduction number estimation. As described in the methods and results, for the evaluation of R_0 in the initial outbreak stage, we considered data from February 24th, up to March 18th. Data in a wider range can be affected by the national lockdown on March 11th. This period agrees well with a previous investigation where the same time window has been assumed as the infection period to determine R_0 for the whole Italy (Remuzzi and Remuzzi 2020).

Taking these preliminary considerations into account, our result of $R_0 = 3.22$ for Italy is highly consistent with values obtained by fitting the exponential growth rate of the infection across a 1-month period (Distante et al. 2020). A similar conclusion has been drawn for transmission dynamics in the Northern regions, and the same results were found for the Southern regions (Distante et al. 2020). In another work, Guzzetta et al. (2020) reported R_0 ranging from 2.50 to 3.00 for six selected Italian regions (Lombardia, Veneto, Emilia–Romagna, Toscana, Lazio, Puglia). While these values are lower than R_0 obtained here, a variability of ~0.5 for most of the regions is thus confirmed independently of geographical location. Again, Gatto et al. (2020), while including additional parameters like mobility and the spatial distribution of communities, determined a comparable initial generalized reproduction number $R_0 = 3.60$. Overall, these data support the idea

Table 1 Basic reproduction number (R_0) values for SARS-CoV-2 estimated in our study and three other works

Region	R_0 [Confidence Interval]			
	This work	Technical report, Istituto Superiore di Sanità, Rome, april 2020	D'Arienzo and Coniglio, 2020	Distante et al., 2020
Abruzzo	3.25 [95% CI: 2.82 to 3.73]	–	–	3.03
Basilicata	2.73 [95% CI: 2.16 to 3.38]	–	–	2.66
Calabria	3.27 [95% CI: 2.66 to 3.95]	–	–	2.84
Campania	3.40 [95% CI: 3.01 to 3.82]	–	–	3.14
Emilia–Romagna	3.49 [95% CI: 3.30 to 3.70]	2.84 [95% CI: 2.57 to 3.13]	–	3.38
Friuli–Venezia Giulia	3.61 [95% CI: 3.08 to 4.21]	–	–	3.04
Lazio	3.62 [95% CI: 3.15 to 4.12]	3.00 [95% CI: 2.68 to 3.33]	–	3.11
Liguria	3.68 [95% CI: 3.29 to 4.09]	–	–	3.10
Lombardia	3.88 [95% CI: 3.75 to 4.02]	2.96 [95% CI: 2.73 to 3.17]	–	3.60
Marche	3.09 [95% CI: 2.85 to 3.36]	–	–	3.12
Molise	2.52 [95% CI: 1.68 to 3.60]	–	–	2.60
Piemonte	3.19 [95% CI: 3.04 to 3.35]	–	–	3.40
Puglia	3.34 [95% CI: 2.97 to 3.74]	2.61 [95% CI:2.13–3.13]	–	3.11
Sardegna	3.00 [95% CI: 2.33 to 3.78]	–	–	3.00
Sicilia	3.35 [95% CI: 2.95 to 3.78]	–	–	2.99
Toscana	3.46 [95% CI: 3.12 to 3.83]	2.50 [95% CI: 2.18–2.83]	–	3.25
Trentino–Alto Adige (*)	3.21 [95% CI: 2.88 to 3.57]	–	–	Bolzano:2.90 Trento: 3.23
Umbria	2.44 [95% CI: 2.16 to 2.74]	–	–	2.79
Valle D'Aosta	3.43 [95% CI: 2.92 to 3.97]	–	–	2.95
Veneto	3.73 [95% CI: 3.46 to 4.04]	2.51 [95% CI: 2.18 to 2.86]	–	3.32
Italy	3.22 [95% CI: 3.14 to 3.29]	–	3.10	–

(*): The original incidence data related to Trento and Bolzano were merged into a single region called Trentino–Alto Adige, resulting in a geographical disaggregation of Italy into 20 regions

that epidemiological figures of the SARS-CoV-2 epidemic in Italy are slightly higher than those observed at the early stage of outbreak in Wuhan, China (Zhao et al. 2020).

The initial large values observed resulted from a sudden increase of independent first-reported infections which in many cases can be related to the so-called "super-spreading" events. Indeed, as observed for the SARS outbreak (Wallinga and Teunis 2004), in the early stage of the epidemic the time dependence of R_t shows a fluctuating pattern characterized by a wide confidence interval raised by the initial low number of cases used in the calculations. In this context, the super-spreading events cannot be necessarily triggered by a single infector, but it can be related to a few people who are perpetuating an epidemic in the susceptible population.

Here we observed that most of the regions have faced "super-spreading" events in the early stage of the epidemic, and the observation of such events is significant in the Southern regions of Basilicata, Calabria, Campania, Molise, Puglia, and Sicilia. Indeed, in these regions the overshoot of the R_t observed in the first 10 days can probably be correlated with the uncontrolled movement of people leaving the most affected northern regions to go to the south of Italy at the beginning of March. Furthermore, it should be noted that in some regions like Valle D'Aosta and Veneto, R_t suddenly increases around the second week of March 2020. Such modulations can be associated to the changes in the testing practices which promptly affected the ratio between the number of new confirmed symptomatic cases and the number of swabs owing to the step-like increase of daily incidence.

After the early stages, the R_t showed a decreasing trend which is likely to be affected by the temporal depletion of susceptible individuals (intrinsic factors) and by the implementation of control measures (extrinsic factors). Both these factors slow down the growth rate of incidence and deeply affect the shape and time-scaling of the epidemic peak driving R_t to fall below 1 (Nishiura and Chowell 2009).

We found that the mean value of time to reach the control regime is about 31 days from February 24 and about 14 days from the first day of nationwide lockdown (March 12, 2020). This mean value is in fair agreement with the value of 29 days detected for the whole of Italy, proving the self-consistency of our analysis. More precisely, we noted that Marche is the first
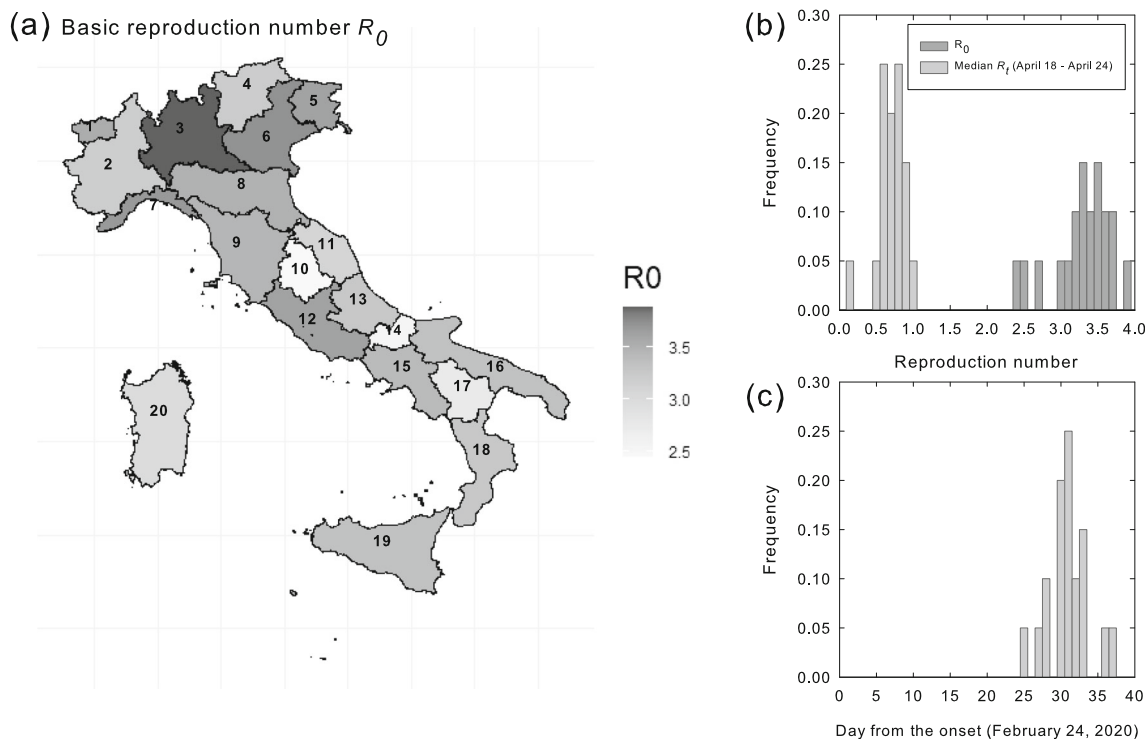
Fig. 2 The basic reproduction number (R_0) at the early stage of epidemic. **a** The map of Italy shows the basic reproduction number (R_0) in all the regions as determined in the early stage of the epidemic. **b** Observed distribution of R_0 and the R_t sorted by median values in the last 7-day time window (18–24 April, 2020) determined in each region. **c** Observed distribution of the first day from the onset (24 February, 2020) when the

time evolution of R_t converges to $\lesssim 1$ in each region. In **a** the different regions are numbered as follows: 1 - Valle D'Aosta; 2 - Piemonte; 3 - Lombardia; 4 - Trentino–Alto Adige; 5 - Friuli–Venezia Giulia; 6 - Veneto; 7 - Liguria; 8 - Emilia–Romagna; 9 - Toscana; 10 - Umbria; 11 - Marche; 12 - Lazio; 13 - Abruzzo; 14 - Molise; 15 - Campania; 16 - Puglia; 17 - Basilicata; 18 - Calabria; 19 - Sicilia; 20 - Sardegna

region reaching $R_t < 1$ on March 20, followed by Lombardia on March 22, Emilia–Romagna and Umbria on March 23, Veneto, Friuli–Venezia Giulia, Liguria and Silicia on March 25, and Piemonte on March 26. The last Italian region reached the control regime is the southern region Molise on April 1, 2020 due to delay of the epidemic onset.

Based on the timing of the targeted (March 9) and nationwide (March 12) lockdowns, this provides a direct evidence of the burden of social distancing measures introduced to control the epidemic. Indeed, the time gap between the introduction of the government measures and $R_t \leq 1$ ranges between 13 and 15 days for the most affected regions, namely Lombardia, Veneto, Emilia–Romagna, and Piemonte. Overall, after 20 days from the national lockdown all the regions displayed reproduction number below the unity. Recently, also Sebastiani et al. (2020), analyzing data in the period from March 1 to March 31, showed that the cumulative incidence growth rate peaked in the most Italian provinces on average 13.6 days after the restriction imposed by the Government.

In the control regime, we observed for about 30 days (up to April 24, 2020) that most of the regions experience a plateau in which R_t fluctuates just below 1, in

agreement with the smooth decreasing of the daily incidence. Exceptions are represented by Molise and Umbria, where R_t drops down to $\sim \frac{1}{2}$ according to the small number of positive cases updated. Compared to R_0, the median R_t of the last 7-day time window $\widetilde{R}_t$ presents a quite narrow distribution with a mean value of 0.71. Considering the decrease from the $\overline{R_0} = 3.29$ to median $\widetilde{R}_t$, we estimated that after 45 days the nationwide lockdown prevents about 78% of potential secondary infections on average.

Although the Italian Government's restrictive measures have proven to be of considerable utility in modifying the trend of the epidemic (Sebastiani et al. 2020), thus preventing even more devastating effects from the epidemic, the challenge of "Phase 2" appears even more demanding. In this respect, obtaining simple and effective indices to evaluate the state of activity of the epidemic seems mandatory: if the R_t index remains essential for understanding the trend in a given area; however, it is not the only parameter to account for. Briefly, if we consider two areas with the same R_t, the one which has a higher PD and/or NA must be considered more at risk, be monitored more carefully, and be potentially the target of more timely restrictive measures.

Table 2 Day from the epidemic onset which R_t reaches the $R = 1$ condition and the median reproduction number in the last 7-day time window

Region	First day$^{(+)}$ from the onset of epidemic at $R_t \lesssim 1$ (corresponding calendar date)	Median R_t values in the last 7-day time window (18–24 April, 2020l)
Abruzzo	31 (03/26/2020)	0.60
Basilicata	31 (03/26/2020)	0.57
Calabria	31 (03/26/2020)	0.71
Campania	36 (03/31/2020)	0.63
Emilia–Romagna	28 (03/23/2020)	0.76
Friuli–Venezia Giulia	30 (03/25/2020)	0.64
Lazio	32 (03/27/2020)	0.76
Liguria	30 (03/25/2020)	0.89
Lombardia	27 (03/22/2020)	0.91
Marche	25 (03/20/2020)	0.68
Molise	37 (04/01/2020)	0.72
Piemonte	31 (03/26/2020)	0.99
Puglia	33 (03/28/2020)	0.80
Sardegna	31 (03/26/2020)	0.58
Sicilia	30 (03/25/2020)	0.93
Toscana	33(03/28/2020)	0.72
Trentino–Alto Adige (*)	32 (03/27/2020)	0.77
Umbria	28 (03/23/2020)	0.54
Valle D'Aosta	33 (03/28/2020)	0.14
Veneto	30 (03/25/2020)	0.83
Italy	29 (03/24/2020)	0.84

($^{+}$) Date of epidemic onset February 24, 2020; (*) the original incidence data related to Trento and Bolzano were merged into a single region called Trentino–Alto Adige, resulting in a geographical disaggregation of Italy into 20 regions

The PD of a given area is clearly an R_t-independent risk factor for the development of an epidemic that spreads through human infection, although the PD of the different Italian regions may not be truly representative of the distribution of the population. Urban areas, and in particular metropolitan areas (Rome, Milan, Naples), have a population density higher than the regional ones. Furthermore, due to the peculiar Italian orography, some regions (for example Liguria, Valle D'Aosta, Trentino–Alto Adige) concentrate the population in a "habitable" area much less large than the total surface. Closely related to PD is DS, which expresses the number of people susceptible to contagion per km^2. In this phase of the epidemic, with a relatively limited number of cases, the use of DS instead of PD could seem superfluous, since the values of DS and PD appear similar. However, in some regions where the impact of the epidemic is still significant, such as Lombardia, we expect DS to be more appropriate than PD. In this work smaller areas have not been considered, but in some areas of Lombardy probably PD is no longer representative, i.e., the provinces of Bergamo, Brescia, and Lodi, where the infected represented more than 10% of the whole population on April 24, 2020. Taking into account this consideration, the higher DS (PD) is, the higher the probability of an elevated number of infections results if $R > 1$. Hence, DS is an important parameter to link to R.

After an initial phase of the epidemic in which the number of infections appears a reliable indicator; in a second phase other reliable indicators to monitor the evolution of the epidemic could be the incidence of intensive care units (ICU)-hospitalizations for COVID-19 and/or mortality rates for COVID-19 (García-Basteiro et al. 2020). However, both the incidence of ICU-hospitalizations and mortality rates suffer from a time delay of at least 7–15 days compared to the trend of the epidemic. Although the long-term mortality rate from COVID-19 will probably end up being the most reliable indicator (García-Basteiro et al. 2020), the delay of 15 days makes this parameter virtually unusable for making decisions in time. Although not completely independent of R and, like it, susceptible to estimation errors related to the number of swabs performed, NA is in our opinion an important parameter because it is representative of the activity of the epidemic in the period immediately preceding the day considered; again, NA is correlated with the healthcare burden related to the epidemic. Moreover, the higher the NA is, the more R will be difficult to reverse in that short time period.

Furthermore, looking at the national incidence data, the curve resulted in being asymmetrical with the descent part which scales slower than the ascent part. In this context, it is difficult to predict exactly the day-to-day trend and mostly the end of the epidemic, basing the forecasts exclusively on the data of R_t. As example, in a recent work, Zhang et al. 2020 predicted exactly the date of epidemic peak in Italy on the basis of the trend of R; however he did not have the same success in predicting the number of cumulative cases (172,451 expected), as they were already $N = 192,994$ on April 24, 2020. Moreover, he did not correctly estimate the epidemic end date, expected to be for June 1, 2020.

As a final remark, our analysis suggests that risk assessment based only on R_t, could underestimate, in some cases, the actual local transmission dynamics. This depends on demographics and social factors which are not included in the basic calculation of reproduction number with conventional methods (Obadia et al. 2012). For this purpose, while admitting their arbitrariness, we suggest the combined use of NA, DS, and R_t seems promising, because these three parameters can well describe the course of the epidemic and are relatively "easy to handle".
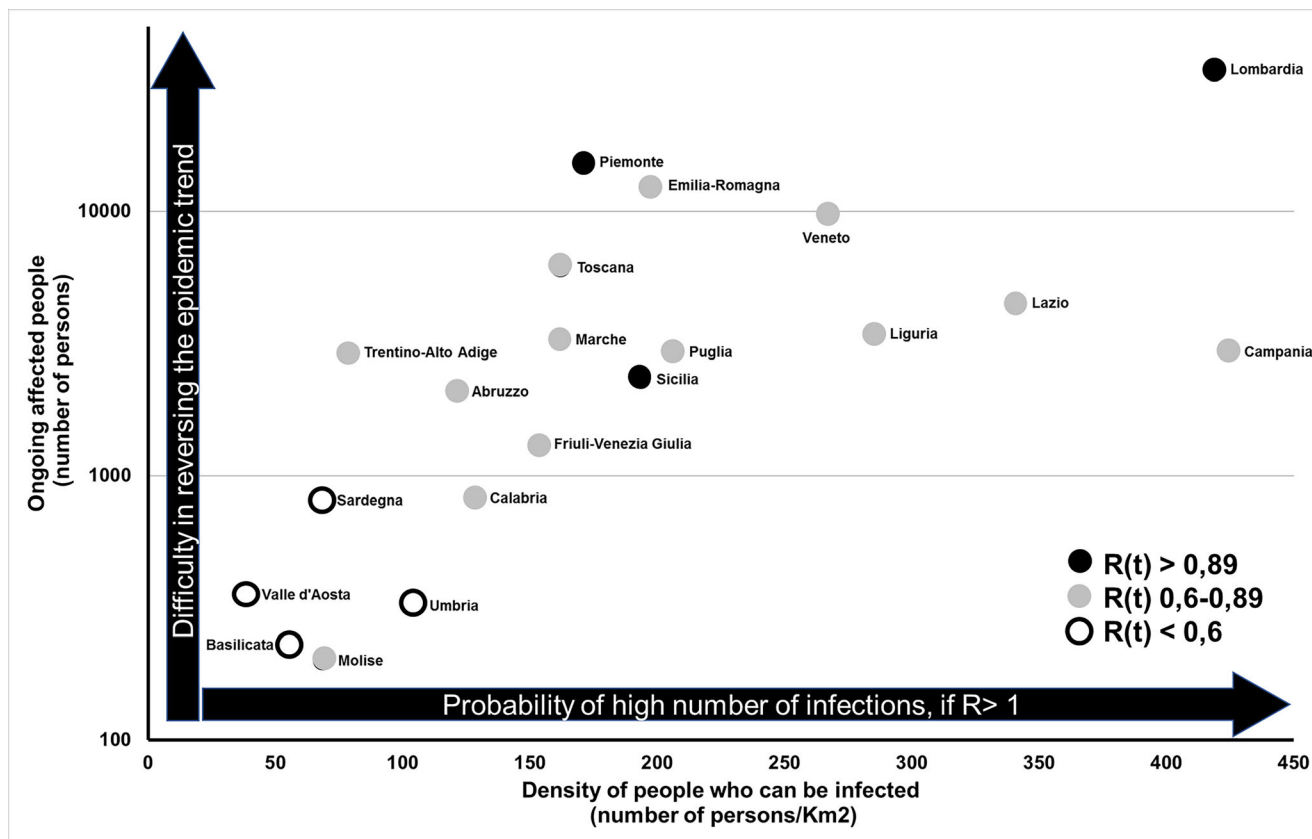
Fig. 3 The correspondence of R, density of people susceptible to infection (DS) and number of ongoing affected people (NA) on the date 24 April, 2020. Y-axis is plotted in logarithmic scale

Conclusions

Reproduction number is a fundamental parameter to analyze the epidemiological figures of COVID-19 in Italy. In this work we calculated R_0 and the evolution of R_t in the period of 2 months from the national onset (from February 24 to April 24, 2020). Our data confirmed the progressive decrease of R in all Italian regions and in the whole country starting from the early stages in which the R_0 values ranging between 2.44 (Umbria) and 3.88 (Lombardia), with $R_0 = 3.22$ (Italy). Moreover, we underlined the positive role of a national lockdown in preventing a larger spreading of the epidemic. Finally, we planned for our next work to combine use of two parameters, DS and NA, with R_t, in order to monitor the evolution of the epidemic in the regions and in all the country throughout 1 year of the COVID-19 outbreak (from March, 2020 to March, 2021) to highlight the impact of the restrictive measures taken by the Italian government and verify their effectiveness in containing the pandemic.

Declarations

References

Britton T, Scalia Tomba G (2019) Estimation in emerging epidemics: biases and remedies. J R Soc Interface 16:20180670. https://doi.org/10.1098/rsif.2018.0670

Carinci F (2020) Covid-19: preparedness, decentralisation, and the hunt for patient zero. BMJ 368:bmj.m799. https://doi.org/10.1136/bmj.m799

D'Arienzo M, Coniglio A (2020) Assessment of the SARS-CoV-2 basic reproduction number, R_0, based on the early phase of COVID-19 outbreak in Italy. Biosaf Health 2:57–59. https://doi.org/10.1016/j.bsheal.2020.03.004

Distante C, Piscitelli P, Miani A (2020) Covid-19 outbreak progression in Italian regions: approaching the peak by the end of March in Northern Italy and first week of April in Southern Italy. Int J Environ Res Public Health 27(17):3025. https://doi.org/10.3390/ijerph17093025

García-Basteiro AL, Chaccour C, Guinovart C, Llupià A, Brew J, Trilla A, Plasencia A (2020) Monitoring the COVID-19 epidemic in the context of widespread local transmission. Lancet Respir Med 8:440–442. https://doi.org/10.1016/S2213-2600(20)30162-4

Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A (2020) Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proc Natl Acad Sci U S A 117:10484–10491. https://doi.org/10.1073/pnas.2004978117

Guzzetta G, Riccardo F, Marziano V, Poletti P, Trentini F, Bella A, Andrianou X, Del Manso M, Fabiani M, Bellino S, Boros S, Urdiales AM, Vescio MF, Piccioli A; COVID-19 Working Group, Brusaferro S, Rezza G, Pezzotti P, Ajelli M, Merler S (2020) Impact of a nationwide lockdown on SARS-CoV-2 transmissibility, Italy. Emerg Infect Dis 27(1). https://doi.org/10.3201/eid2701.202114

Micheli V, Rimoldi SG, Romeri F, Comandatore F, Mancon A, Gigantiello A, Perini M, Mileto D, Pagani C, Lombardi A, Gismondo MR (2020) Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase. J Med Virol 93(3):1752–1757. https://doi.org/10.1002/jmv.26447

Nishiura H, Chowell G (2009) The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Castillo-Chavez C, Chowell G, Hayman JM, Bettencourt LMA (eds) Mathematical and statistical estimation approaches in epidemiology. Springer, Netherlands, pp 103–121

Obadia T, Haneef R, Boëlle PY (2012) The R_0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak 12:147. https://doi.org/10.1186/1472-6947-12-147

Remuzzi A, Remuzzi G (2020) COVID-19 and Italy: what next? Lancet 395:1225–1228. https://doi.org/10.1016/S0140-6736(20)30627-9

Sebastiani G, Massa M, Riboli E (2020) Covid-19 epidemic in Italy: evolution, projections and impact of government measures. Eur J Epidemiol 35:341–345. https://doi.org/10.1007/s10654-020-00631-6

Sjödin H, Wilder-Smith A, Osman S, Farooq Z, Rocklöv J (2020) Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Euro Surveill 25:2000280. https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000280

Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 160:509–516. https://doi.org/10.1093/aje/kwh255

Zhang X, Ma R, Wang L (2020) Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. Chaos Solitons Fractals 135:109829. https://doi.org/10.1016/j.chaos.2020.109829

Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang MH (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. Int J Infect Dis 92:214–217. https://doi.org/10.1016/j.ijid.2020.01.050