The COVID-19 infection in Italy: a statistical study of an abnormally severe disease.

De Natale Giuseppe[1,2,*], Valerio Ricciardi[3], Gabriele De Luca[4], Dario De Natale[5], Giovanni Di Meglio[5], Antonio Ferragamo[5], Vito Marchitelli[6], Andrea Piccolo[5], Antonio Scala[7], Renato Somma[1,8], Emanuele Spina[9], Claudia Troise[1,2]

[1] INGV, Osservatorio Vesuviano, Naples, Italy
[2] CNR-INO, Pozzuoli (NA), Italy
[3] Università della Campania 'L.Vanvitelli', Dept. Medicina Sperimentale, Naples, Italy
[4] University of Zurich, Dept. Physics, Zurich, Switzerland
[5] CoronaVerus, Naples, Italy
[6] Puglia Region Government, Dept. Mobility, Pub. Works, Ecology, Env., Bari, Italy
[7] Università degli Studi di Napoli "Federico II", Dept. Physics, Naples, Italy
[8] CNR-IRISS, Naples, Italy
[9] Università degli Studi di Napoli "Federico II", Dept. Neurosciences, Reproductive and Odontostomatology, Naples, Italy
*Correspondence: giuseppe.denatale@ingv.it; Tel.: +39 0816108510

Abstract

We statistically investigate the COVID-19 epidemics, which is particularly invasive in Italy. We show that the high apparent mortality (or Case Fatality Ratio, CFR) observed in Italy, as compared with other countries, is likely biased by a strong underestimation of infected cases. To give a more realistic estimate of the mortality of Covid-19, we use the most recent estimates of the IFR (Infection Fatality Ratio) of epidemic, based on CFR for Germany, and furthermore analyse data obtained from the ship 'Diamond Princess', a good representation of a 'laboratory' case-study from an isolated system in which all the people have been tested. From such analyses we try to derive more realistic estimates of the real extension of the infection, as well as more accurate indicators of how fast the infection propagates. We then try to point out, from the various explanations proposed, the dominant factors causing such an abnormal seriousness of the disease in Italy. Finally, we use the deceased data, the only ones estimated to be reliable enough, to predict the total number of infected people and the interval of time when the infection in Italy could stop.

Introduction

The infection of COVID-19, recently declared pandemic by WHO, represents perhaps one of the most serious Worldwide emergencies, potentially able to destroy social order and economies and to deeply change our lifestyle in the near future. The epidemics was firstly detected in China, in the city of Wuhan, at the end of December 2019; about 45 days after (mid-February 2020), it started to seriously affect several other countries (South Korea among the firsts). Since the end of February, it flared up in Italy and, since mid-March 2020, the epidemics was spread all over Europe, in the USA, Iran and many other countries (Worldometers, 2020).

Since the first appearance of this new Coronavirus (Worldometers, 2020), the COVID-19 infection has been treated with very mixed feelings: from just a disease a little more serious than a seasonal flu, to a very severe and troubling affliction. After some interlocutory days, the government of China showed serious concern and implemented very severe measures in the Hubei province, the centre of the epidemics, to contain the epidemic spreading of the infection. Around mid-February, among the most affected countries, besides China, there were South Korea, a rather natural candidate as bordering China, and then Italy and Iran, for less clear reasons. After one further month the epidemic appeared to have spread worldwide.

Here we want to focus on the COVID-19 epidemic in Italy that shows some peculiar features, distinguishing its evolution from the one observed in other countries. The epidemic appears very aggressive, both in terms of spread rate and mortality, which are however very uncertain parameters for this new virus.

In Italy, the infection is mainly focused in the Lombardia Region and the area around the Po river. The most affected regions are Lombardia, Emilia -Romagna and Veneto, which also represent the richest and more productive part of Italy. In Italy, the infection grew very fast, overcoming South Korea in the number of infected people as early as in the beginning of March 2020, today (March 25th, 2020) reaching 74.386 total infections. Moreover, it showed an average Case Fatality Ratio (CFR) over 9%, well above any other country and more than double with respect to the Hubei Region in China, where the new virus first appeared, and where CFR was significantly higher than in other parts of the China and higher than many other countries in the World.

In this paper, we will show statistical analyses of data associated to the Italian Covid-19 epidemics. The aim is to estimate a possible slowdown of the infection and also to understand our statistical predictions in relation with the severe containment measures taken by the Italian Government. We discuss alternative explanations for the very high CFR observed. To confront with the important problem of determining the true mortality (Infection Fatality Ratio, IFR) of the epidemics, we use the study of an isolated, well calibrated test case, represented by the infection spread on the 'Diamond Princess' cruise ship.

Mortality (IFR) estimates obtained within the 'Diamond Princess' are an 'unbiased' value, not affected by the underestimation of the number of infected people; also IFR values computed by the University of Oxford (Oke and Henegan, 2020) from Germany data were checked. Using various IFR estimations, we predict a much larger number of infected people in Italy with respect to the official one, suggesting that the CFR highly overestimates true mortality. We also discuss the likelihood of alternative hypotheses, often claimed to explain the high impact of Covid-19 in Italy, based on the influence of old average age of population, high antibiotic resistance, high number of smokers, pollution in the Po plain (e.g. Oke and Henegan, 2020).

Finally, we identify the best data set to analyse the statistical evolution of the epidemics in Italy, also comparing it with its prototypical behaviour obtained from the China dataset, to try to forecast the time of saturation of the infection.

The COVID-19 in Italy

The COVID-19 epidemic in Italy presents some peculiarities that make it very intriguing to analyse and understand. Initially it was thought that two independent focuses started in Codogno (15.962 inhabitants, Lodi province, Lombardia region) and Vò Euganeo (3416 inhabitants, Padua province, Veneto region) towns, but it is now generally understood that the virus started circulating earlier in the whole north of Italy. The epidemic rapidly blew up all over Italy, but particularly around the Po Valley, in Lombardia, Veneto and Emilia Romagna. These Regions are the richest ones in Italy, for their industries, agriculture in the Po Valley and international commerce. The Lombardia and the Po Valley, the most hit by the infection, are also the most polluted areas in Italy and probably also the most polluted in the whole Europe by the fine air particulate matter (PM10, PM2.5) and Ozone (Martuzzi et al., 2006:; Stafoggia et al., 2009).

The beginning and evolution of COVID-19 infection in Italy, and more specifically in Lombardia region, is from many points of view anomalous and highly lethal, with respect to all other countries worldwide, including China where the infection was born.

Although all the media (and also many specialists interviewed by media) highlight the velocity of infection in Italy as 'exponential', the number of infected people has never followed an exponential distribution, except in the very first few days. Figure 1a shows the number of recorded infections as a function of time (in days since February 24th 2020) in a semi-logarithmic scale. It is evident that the distribution is markedly different from a straight line, typical of an exponential distribution in such a scale ,and it's rather well fitted by a cubic polynomial (much slower than exponential), or by a logistic function, also present in the figure. Figure 1b shows the same quantity in a linear scale, together with the three mentioned fitting functions..

Although the number of infected people is the main parameter taken into account by the authorities, and the most highlighted by media, its real value is largely uncertain, and surely underestimated. In fact, it critically depends from the number of laboratory tests made on people to ascertain the infection, which is anyway limited and very small as compared with the number of inhabitants. Furthermore, the procedure to test people are highly variable within the different Regions of Italy and have changed over time in the last weeks; because of this inconsistency, this number is statistically very inhomogeneous and unfit to to interpret the actual evolution of the infection.

The number of tests in Italy, highly fluctuating but generally increasing in time except in the last days ranged from about 2.427 (February 27th) to 26.336 (March 21st), and decreased again to 25180 (March 22nd) and 17.066 (March23rd) (Il Sole 24 Ore, 2020). A very important quantity, namely the CFR, defined as the ratio of number of people deceased divided by number of total recorded infections, is extremely high in Italy (about 9%). Such a high value is dominated by the mortality in Lombardia, where about 50% of all the Italian infections have been recorded, with a CFR of about 11%. CFR is a generally overestimated value of the true mortality (IFR), given the likely underestimation of the real number of infection cases (i.e. including asymptomatic and pauci-symptomatic cases, which are easily overlooked by the small number of tests). IFR is the parameter which measure the percentage of deceases over the total population infected (including the generally unknown number of non recorded cases). Table 1(Oke and Henegan, 2020) reports the number of infections and CFR observed in several countries in the world. It is possible to note that the mortality rate for different countries is very variable, going from a minimum of 0.4% (i.e. Germany, Australia, Austria) to a maximum of 9% in Italy. Several observations are however around 1%-2% (Denmark, USA, Portugal, Belgium, etc.). It appears evident, therefore, that the CFR of Italy, and even more of Lombardia region, is absolutely extreme when compared to any other country. This CFR is more than double of the CFR obtained in China, where the epidemic first appeared. Therefore, besides CFR, it is important to obtain a reasonable estimation of the IFR already during the epidemic spread, to understand the real hazard of the disease and/or to estimate the real amount of infected people.

The IFR for COVID-19 and the mortality in Italy

The problem of unbiased estimation of the IFR for COVID-19 is complicated by the complex laboratory procedures that are necessary to identify the infected cases, ultimately limiting the total number of tested individuals. Since the IFR is the ratio between the number of deceased due to the disease and the total number of infected, the CFR, which is computed on the number of 'known' infected is an upper limit of IFR. Normally, CFR (and hence IFR) should be determined at the end of epidemic, because decease occurs at the end of the epidemic cycle, delayed with respect to the infection. At the very beginning of epidemic, the CFR could be underestimated just because the outcome of the disease (recovered or deceased) has not yet been reached. For COVID-19, however, it was noted in China (and in South Korea too) that CFR was generally higher in the first phase of the infection spreading. A further problem is that the amount of underestimated infected cases is generally variable, and could become progressively more critical as the real number of infected rapidly rises with time, while the number of tests remains stationary. From Table 1, it is clear that each country has a very variable CFR associated to the COVID-19 infection. Since the IFR is intrinsically overestimated by the CFR, assuming the virus strain is the same in every country, we can assume the minimum value of the observed CFR as the minimum upper limit for IFR. Following this procedure, the University of Oxford (Oke and Henegan, 2020) used one of the minimum CFR values, namely from Germany (CFR=0.4%) obtained in a rather initial phase, and halved it, then computing IFR=0.2%.

For the COVID-19 epidemics, there is an independent way to obtain a rather unbiased estimate of IFR using the only 'laboratory-like' case study: the 'Diamond Princess', the cruise ship anchored in the port of Yokohama from February 4th to March 2nd. Among the 3711 passengers, 705 were found infected, and 7 of them died. This is the only perfectly isolated case in which all the people have been tested, so that the number of infected ones is perfectly known. The IFR=CFR computed from the 'Diamond Princess' is then slightly less than 1%. Although statistically affected by a rather large uncertainty due to the small number of deceased, we can then assume that the IFR of COVID-19 is around 1%; the observation of a CFR significantly smaller than 1% (i.e. Germany, Norway, Australia, Austria, Ireland) indicate that, within the unknown uncertainty of the 'Diamond Princess' IFR, lower values are more probable than larger ones. In the light of such mortality range (less than 1%), it appears very abnormal the observed CFR in Italy (9%), and even more the CFR in the Lombardia Region (11%).

In the following, we will try to point out the possible explanations for such an extremely anomalous outcome. We will, again, assume that the virus strain is the same in every country, since there are not contrary evidences till now. The first, more obvious reason to explain such a high mortality, is to hypothesize that the number of infected cases is substantially underestimated. A clear sign of this is the fact that mortality increases during time, from around 2% at the beginning (February 20th) to 9% on March 20th. During this period the number of detected cases rose from few units to 47.000, whereas the daily number of laboratory tests changed from few hundreds to about 26.000.

Besides the CFR, which is only linked to the recorded infection cases, what is really interesting to understand is a 'local' IFR estimation for Italy, in order to derive the real total number of infected people. Lacking evidence for the existence of a different virus strain, more aggressive and lethal in Italy, we firstly assume that IFR is the same as in other places, and a distinctive trait of this epidemic. We could then choose between the University of Oxford estimate of IFR=0.2%, or the 'Diamond Princess' laboratory-like estimation of IFR=1%.

If the extreme mortality observed would be only due to the underestimation of the number of infected people, assuming a 'true' IFR ranging between 0.2% and 1%, to correct for the Italian 9% CFR, we should multiply, respectively, by 45 or by 9 the official number of infected cases (74.386 on March 25th); the resulting number then ranges between about 670.000 and about 3.3 million infected people.

We cannot exclude, however, the true IFR in Italy to be significantly higher than in other countries because of several concurrent reasons. For instance, the higher average age of Italian people has often been indicated as a possible explanation for the high COVID-19 CFR. Among all the countries of the World, Italy is actually in the second position for higher average age; however, the first position (oldest population) is occupied by Japan, which showed a very low number of infections and CFR (see Table 1) so that this possibility appears unlikely.

Another possible cause of comorbidity could be found in the high level of pollution in Lombardia region that is likely the most polluted region in Europe by fine particulate (PM10, PM2.5) and Ozone. As shown in several papers (e.g. Chen et al., 2010; Ye et al., 2016; Chen et al., 2017; Setti et al., 2020) there is a correlation between the diffusion of viruses and the pollution by fine particulate. Furthermore, exposition to fine powders contributes to enhance the severity for respiratory viral infections (Dominici et al., 2006; Ciencewicki and Jaspers, 2007). The incidence of fine particulate pollution could hence, in principle, be one of the reasons for the high mortality rate observed in Lombardia (and partially in Emilia Romagna, around the Po Plain). Although an effect of this kind of pollution in amplifying the mortality observed for a severe pulmonary disease seems reasonable, it is actually very difficult to quantify its incidence, and is difficult to believe it can be so strong with respect to other highly industrialized areas of Europe (e.g. Germany, France, UK, Netherlands, etc.). Moreover, a recent document issued by the Italian Aerosol Society, signed by about 60 scientists of various disciplines, rejects the hypotheses (Setti et al., 2020), pointing out that there is no clear evidence for a correlation between fine particulate and COVID-19 disease amplification (Contini et al., 2020).

Other tentative explanations for the extreme mortality were the high number of smokers in Italy and the antibiotic-resistance of Italian people (Oke and Henegan, 2020). However, regarding the percentage of smoking people, Italy's 23% is lower than the European average, 29% (WHO, 2016); regarding the antibiotic-resistance, on the contrary, Italy has actually the most critical position in Europe. Among about 33,000 yearly deceases in EU due to antibiotic-resistant bacteria, about 10,000 occur in Italy alone (ISS, 2019). Since generally therapies used against the Cov-SARS-2 involve one or more antibiotics, this issue could, in principle, result in higher mortality. Also in this case, however, it is not easy to quantify this effect, which however appears marginal in seasonal flu, since in this case CFR for Italy does not differ too much with respect to other European countries (ISS, 2019). Moreover, during Covid-19 antibiotics administration aims to avoid bacterial superinfection, whose mortality is negligible; so even if there was a more pronounced antibiotic resistance and this mechanism would be implied in higher mortality, this would probably account for only a very little part of the deaths. The remaining possibility is that the healthcare was unprepared for such an emergency due to a respiratory syndrome; actually, we just note that in Italy (60 million people), there were, before COVID-19 epidemic, about 5090 places in intensive care units (ICU); for comparison in Germany (82 million people) such places were 28.000. Actually, in terms of number of ICU divided by population, Italy occupies the 19th position among 23 European Countries. We observe indeed high pressure on ICU by severe/critical cases, mainly in Lombardia, where the number of ICU before the crisis was 900; this number increased to over 1000, but, at the moment (March 25th), the patients hosted in ICU are more than 1200, and, since March 14th, the Government of Lombardia already declared that available ICU were almost over. Another indication that something went wrong during the first phase of management of the infection by the Lombardia Hospitals is the very high number (6.205) of infected medical staff (ISS daily Info, March 25th). So, the Hospitals could have been the most effective carriers for the epidemic in the first phase in Lombardia, when a very fast exponential increase was observed.

Possible forecast of future behaviour of infection

In order to forecast the evolution and end of the COVID-19 epidemic in Italy, we could in principle use three kinds of data. The most obvious would be the daily infection data. However, such data are particularly unreliable, because too much dependent from the daily number of tests. They are generally very variable and inhomogeneous, both from a Region to another, and also in time. Since, as we noted, the real number of infected people is likely much higher (orders of magnitude) with respect to the sampled one, the inhomogeneous sampling can strongly condition the number of infections, making them not useful for a statistical study. Another possible indicator of the epidemic evolution is the number of people in Intensive Care Units (ICU). Contrarily to the number of infected, the number of people in ICU should be objective, because who has serious breathing problems must necessarily be hospitalized in ICU. However, in this phase of the epidemic crisis, this number has two problems: the first one is that, at least in Lombardia which however dominates the statistics, ICU places are mostly full, so not all people requiring them can be allocated. The second problem is that the daily numbers given by Civil Protection just mention the total number of people hosted in ICUs in that day, and not the daily incremental number. Hence, it is not possible to know the real cumulative number of people hosted in ICUs, because we don't know how many people each day came out from them, due to recovering or death.

The only quantity which has a rather rigorous statistical meaning, then, is the daily cumulative number of deceases. We then choose to use this number in order to statistically analyse the evolution of epidemic and to predict its end. Obviously, since we are particularly interested in determining the time at which the epidemic ends and the total number of infected people cumulated at the end of the epidemic, we have to correctly consider the relation of the daily cumulative number of deceases with the cumulative number of infected people. The number of deceased people is linked to the number of infected one by the IFR=D/I (D= deceased, I=infected); then, correcting the number of deceases for the constant factor represented by the inverse of IFR (I/D) gives the number of infected people. However, we must also consider that infection and decease are the two temporal limits of the disease: it starts with infection, proceeds with the symptoms, and then ends with one of the two possibilities: recovering or death. For doing so, we must consider the shift in time between the infection and the decease. For COVID-19, it can be estimated that the average time from infection to death is 16 days ((Jung et al., 2020). With such relations in mind between infected and deceased people, we fit the decease data to a logistic function of the general form:$y(t) = K\frac{1+me^{-t/\tau}}{1+qe^{-t/\tau}}$

Fig.2 shows the fit we made, using a logistic function, to the epidemic curve of the cumulative daily number of cases (infections) reported in the whole China, Hubei province and China without Hubei province. It is clear how, in the case of an epidemic already ended, so describing the whole epidemic behaviour from the start to the end, the logistic function gives a good fit to the process. In our case, as we explained, we are using the cumulative daily number of deceases to fit the logistic function, simply because is a much more reliable quantity when compared to the total reported cases..

In order to convert from cumulative number of deceases to cumulative number of infections, we use several tentative values of IFR: the one computed by University of Oxford, IFR=0.2%; the one computed from the mortality on the Diamond Princess, IFR=1%; and one much larger, in case the IFR for Italy would be much larger than in other countries, IFR=5%. Such a large range of IFR for conversion is useful to check the range of times of saturation, and obviously of the final amount of total infections.

Fig.3 shows the fits of the deceased data (black data points) reported in Italy with a logistic function. It is clear the logistic model gives a good fit to these data.

Fig.4 shows the fit of the estimated amount of total cases, by multiplying the daily number of deceases by the assumed IFR. The different logistic curves (blue, red and green), present in the figure, are the estimated number of contagious people for the different IFR values of 0.2%, 1% and 5%, respectively. From these values we also subtract the number of cases reported by Italian authorities to estimate the number of unreported contagious cases (cyan, pink, light green). It is evident that the number of reported cases is only a minor fraction of the estimated number of contagious people. In particular, for an IFR of 0.2%, 1% and 5% we estimate a total number of cases at March 25th (30 days from the beginning of the epidemics) of about 3.3 million, 670,000 and 134,000 depending on the value of the IFR. With a total number of reported cases in the order of 75000, Italy might be strongly underestimating the total number of infected people (including asymptomatic and pauci-symptomatic persons) of 98.6%, 93.2% or 66.2% depending on the IFR.

From the time dependence of the logistic function, we observe that in all cases the inflections of the respective curves are now exceeded by the data points, whatever the conversion factor IFR is used. This probably means that Italy as a whole have already overcome the maximum number of new daily infections. It is also important to note that, since the derivative of the logistic function is symmetric, data which overcome the inflexion point are well constrained, also if they do not cover the other part of the curve. In general, different growth functions like the Gompertz (1832), Janoschek (1957) or Richards (1959) sigmoids could be more suited to fit a dataset representing the curve of the daily number of infections of an epidemic because they allow their derivative to be non-symmetric and therefore resulting in a more generalized integral function. However, in our case, besides the very good fit given to our data, the test on the infection curve of China already shown in Figure 3 demonstrates that the logistic curve is a good approximation of the real process.

The total, true number of infections, at the end of epidemic, is obviously very variable, depending from the value assumed for IFR. It ranges (on March 25th)from a minimum of 134,000 (for

IFR=5%) to a maximum of 3.3 x 10^6 (for IFR=0.2%). We note, from best fitting curves, the peak value of the infections should have been reached around March 8th, four days after the closing of schools and just before the lockdown of Lombardia and then (two days after) of Italy. Moreover, our results predict that the point of 95% of the maximum value of the best fitting logistic curves will be reached on the last week of March, and that, within the first week of April, the real number of contagious people will be already well within the saturation, so that it practically will not increase anymore, assuming that lockdown will be maintained until that date.

Discussion

The COVID-19 epidemic in Italy shows very abnormal features, in terms of severity of disease and, particularly, mortality. The CFR (defined as the ratio between the number of deceased and the number of recorded infected) reaches here very high values, disproportioned with respect to any other Country: about 9%, rising to about 11% in the Lombardia Region, which alone represents half of the total number of infected people in Italy. Estimating a reliable value of the IFR (the ratio between the number of deceases and the total number of infected people), mainly for a virus hard to diagnose and requiring complex laboratory procedures, is very difficult. In this case, we considered two approaches: the first one, used by University of Oxford (Oke and Henegan, 2020) is based on the minimum statistical reliable CFR value, observed in Germany; the other one benefits of the possibility to study a very peculiar laboratory-like case study, namely the case of Diamond Princess, the cruise ship remaining at anchor in Japan (in the port of Yokohama) in which all the passengers and crew (3.771 people) were tested for COVID-19. Using the CFR of Germany, CFR=0.4%, University of Oxford assumed a 50% of non recorded infected; in this way, they computed a value of IFR=0.2%. In the alternative approach, we considered that in the Diamond Princess cruise ship 705 infected were detected, and 7 of them died. So, the mostly unbiased value of IFR=CFR=1% can be computed. We have then considered these values as the limiting values defining the range of IFR. If we assume that the infection number underestimation is the only reason for the large overestimation of the mortality rate, in order to determine the real number of infected cases able to correct the apparent mortality to the true value, we should multiply the 'official' number of recorded infections by the ratio between the CFR=9% and the IFR (=0.2% or 1%). In this way, we obtain a real number of infected people in Italy (today, March 25th) ranging between about 670,000 and about 3,300,000 people.

Even if this number appears very high (mainly with respect to the total number of reported cases of less than 90.000 in China, where however, using the same assumption, this amount should be multiplied by 4), it is not unreasonable for several reasons. Firstly, besides the high number of tests made in Italy (about 240.000), it is however only a small fraction of the total population (60 million people), and does not take into account the number of asymptomatic and pauci-symptomatic cases. The large number of potential daily infections in Italy during mobility for work or study can be also independently estimated by the statistics of the use of public transportation (the most likely to cause infection due to the high number of people assembled in small space). It results that 30 million people moves in Italy each day for work or study, with peak value in Lombardia (ISTAT report, 2017): about 56% of these people use public transportation; two thirds of the total travels for more than 15 minutes. So, considering only such a number of assembled people, and neglecting the likely high incidence of student assemblage in schools, of people in work places, and the more or less occasional frequentations of restaurants, clubs, pubs, sport matches, theatres, supermarkets etc., we get a minimum estimate of about 11 million people staying in close contact each day. A final indication that such number of effective infected cases is not unrealistic comes from the fit of the first days of infection, when the curve of increasing cases fitted well an exponential function (Fig. 1); by extrapolating that exponential curve till March 22nd it would predict about 6 million infected cases; i.e. well beyond the maximum number computed three days after March 25nd by our model. Obviously, such an hypothesis would assume that the exponential increase lasted till March 25nd (or however till few days before) and that the real exponential curve was missed because the limited number of tests progressively sampled a smaller and smaller percentage of the true daily number. From any point of view, in conclusion, a real number of 660-3.300 x103 infected people today does not appear unrealistic.

Other reasons could however equally affect the very high mortality observed (as CFR): discarding the existence in Italy of a virus strain significantly different and more aggressive (hypothesis actually impossible to verify because there is not yet the Italian genome available), a possible contribution could come from the high air pollution (from particulate matter PM10, PM2.5 and Ozone) of the zone around the Po Plain, where more than 50% of counted infections are clustered. Among the other hypothesized causes, the high average age (however lower than Japan, which showed a much lower CFR) is not likely to play a fundamental role; nor it is the number of smoking people (lower than the EU average). The observed high antibiotic-resistance, the highest one in Europe, has been also indicated as a possible contributing factor. However, although such problem causes in Italy about one third (10.000 cases; the double of France, and fourfold of Germany) of the total deceases for antibiotic resistance in the whole EU (33.000 cases), it does not appear to significantly affect the mortality (CFR) associated to seasonal flu (the most reliable comparison we can do, to infer the possible effect on COVID-19 mortality), which seems not significantly higher with respect to other European countries. We can compare, for instance, the number of deaths directly or indirectly associated to seasonal flu (the number of indirect deaths is much more significant in this case, because it is mostly linked to bacteria super-infections) in Italy and in Germany. The average yearly number of such deaths in Italy is around 8.000 (ISS, 2020); in Germany, the deaths associated with the 2017-2018 flu, although very severe that year, totalled 25.000 (Koch Institute, 2019)

A significant contribution to increase of the “local” IFR could, on the contrary, have come from the saturation of the public Hospitals, and in particular from the limited number (as compared for instance to Germany) of ICU. Italy, before the COVID-19 epidemic, had in fact 8.4 ICU per 10^5 citizens (5,090 ICU total), whereas Germany had 34 of them (28,000 ICU total). Also, some wrong actions taken by the public Hospitals in managing the first days of infection (testified by many media and by a dramatic percentage of infected medical staff), could have significantly enhanced the infection. This 19 doctors died till now (March 25th), and 6205 people from the medical staff have been infected (ISS daily Info, March 25th, 2020).

In this paper, we also use an indirect procedure, based on the analysis of the cumulative number of deceases which is the only reliable datum, in order to forecast the short-term evolution of the epidemic in Italy. As we have shown, the number of recorded infections is not statistically reliable for such analyses, since inhomogeneous in time and in space (among different Regions). Considering correctly the average time shift of the death with respect to the infection (16 days used here) we show that decease data, converted to infection data using a large range of possible IFR, are well fitted by logistic functions, all of them indicating the time of 95% of total infection (i.e. the starting of the flat part of the function, which represents the end of infection), is expected around the last days of March. The point of inflection of the logistic best fitting functions, which corresponds to the peak of the infections in Italy, occurred on March 8th. Looking at Fig.3, this date occurs just before the lockdown of Lombardia and, after 2 days, of the whole Italy. It then demonstrates the high mitigating impact of the lockdowns; however, since it also occurs only 4 days after the closing of the schools, it probably indicates it also has been an effective measure to contain the infection; then suddenly improved by the lockdowns. The high effectiveness of the closure of schools for containing epidemic spread has been specifically assessed by a number of epidemiologic studies (i.e. Adda, 2016). These results further confirm that the infection data we record today are only dependent on the variable and inhomogeneous sampling and have nothing to do with the true statistical evolution of the epidemic.

Looking at these very inhomogeneous data, it is on the other hands very clear that, since several days, the increase or decrease of new daily infections just depend on the number of tests. This is an obvious consequence of the fact, demonstrated here, that the true number of infections is much larger than the small sample tested: so, more tests you do, more cases you record. The information that the epidemic is reaching the saturation (and hopefully end, as forecasted, within the first half of April), although not apparent from the tested number of new infections, should be seriously considered in order to decide how to tune and eventually release the next measures for the containment of the infection

Conclusions

We analysed the COVID-19 epidemic in Italy, which showed to be the largest and most lethal in the World. We discussed the causes of such anomalous behaviour of the disease in Italy, where mortality appears much higher than in other Countries. Among the various hypotheses made till now to explain such an anomalous behaviour, it appears that neither the old age of population, nor the observed antibiotic-resistance of population (by far the highest one in Europe), nor even the smoking level, should have significant effects on the observed mortality. The most reasonable effect would involve a strong underestimation of the extension of the infection (implying a factor 10 to 45 more cases than the tested ones). A possible, further contribution could be given by the very high fine powder (and Ozone) pollution (one of the highest ones in Europe), which on one hand could facilitate the virus transmission, on the other hand could make more vulnerable and stressed the lung, causing heavier damages upon impact with virus. Another factor, emerging from our analyses (as well as from media), that likely amplified mortality, could have been the scarse preparedness and possible initial faults of the sanitary system, mainly in Lombardia where the epidemic first blew up in few days.

The problem of underestimation of the number of infections, coupled with a non homogeneous sampling and testing of the positive cases, makes these data unreliable to use for statistics aimed to forecast the epidemic evolution. We then use here the cumulative daily number of deceases, corrected for an appropriate IFR, to simulate the evolution of the epidemic by a logistic function. When corrected for IFR and time shift between infection and death, these data can be well fitted by a logistic function, and show that actually the peak of infection has been just overcome, and the saturation of the curve (end of epidemic) is expected within the first week of April or few days later. The information here obtained about the possible actual evolution of the epidemic and its likely end, which is not at all evident from the very inhomogeneous sampling of infections actually performed, should be seriously considered in order to decide what measures to undertake or relax in the next future. Moreover, considering the likely effectiveness of the schools closure, which is probably the minimally invasive measure in a social and economic sense, it should be considered as a primary measure when the lockdown will end. The study of this unprecedented medical catastrophe will hopefully give robust indications to avoid, in the future, the same errors and understatements. It will also help to seriously consider the importance to improve the sanitary systems, whose health and capability should be considered invaluable, also using a purely economic criterion.

References

Adda, J., 2016. Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data , The Quarterly Journal of Economics, 131, 2, 2016, 891-941, https://doi.org/10.1093/qje/qjw005

Chen, R., Chu, C., Tan, J., Cao, J., Song , W., Xu, X., Jing, C., Ma, W., Yang, C., Chen, B., Gui, Y., Kan, H. 2010. Ambient air pollution and hospital admission in Shangai, China. Journal of Hazardous Material, Volume 181, Issues 1-3, 15 September 2010, pp. 234-240 https://doi.org/10.1016/j.jhazmat.2010.05.002

Chen, K., Horton R.M., Bader D.A., Lesk C., Jiang L., Jones B., Zhou L., Chen X., Bi J., and Kinney P.L., 2017. Impact of climate change on heat-related mortality in Jiangsu Province, China. Environ. Pollut., 224, no. 317, doi:10.1016/j.envpol.2017.02.011.

Ciencewicki, J. & Jaspers,I. 2007 Air Pollution and Respiratory Viral Infection, Inhalation Toxicology, 19:14, 1135-1146, doi: 10.1080/08958370701665434

Contini et al., 2020. Società italiana di aerosol Informativa sulla relazione tra inquinamento atmosferico e diffusione del COVID-19, pp. 4 (in Italian) http://www.iasaerosol.it/attachments/article/96/Nota_Informativa_IAS.pdf

Dominici F, Peng R.D., Bell M.L., et al., 2006. Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. JAMA. 2006;295(10):1127–1134. doi:10.1001/jama.295.10.1127

Il Sole 24 Ore, 2020. (in Italian) https://lab24.ilsole24ore.com/coronavirus/#box_6

ISS, Infografica March 25th, 2020. https://www.epicentro.iss.it/coronavirus/bollettino/Infografica_25marzo%20ITA.pdf

ISTAT, 2018. Spostamenti quotidiani e nuove forme di mobilita' pp. 20 (in Italian) https://www.istat.it/it/files//2018/11/Report-mobilit%C3%A0-sostenibile.pdf

ISS report, 2020 Epidemia COVID-19 pp. 12 (in Italian) https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_23-marzo%202020.pdf

Janoschek, A., 1957. Das reaktionskinetische Grundgesetz und seine Beziehungen zum Wachstums- und Ertragsgesetz. Stat. Vjschr. 10, 25-37.

Jung, S.M., Andrei R. Akhmetzhanov, A.R., Hayashi, K., Linton, N.M., Yang,Y., Yuan, B., Kobayashi, T., Kinoshita, R., Nishiura, H., 2020. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. medRxiv, doi: 10.1101/2020.01.29.20019547

Gompertz, B., 1832 On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Phil. Trans. Roy. Soc. London 123, 513-585, 1832.

Koch Institute, 2019. Report 2018-2019 https://www.rki.de/EN/Content/infections/epidemiology/inf_dis_Germany/influenza/summary_2018-19.html.

Martuzzi M., Mitis F., Iavarone I., Serinelli M., 2006. Health impact of PM10 and Ozone in 13 Italian cities. World Health Organization Regional Office for Europe ISBN 92 890 2293 0 WHOLIS number E88700 World. Pp. 147

Oke, J,, Heneghan, C., 2020. Global COVID-19 Case Fatality Rates by CEBM and University of Oxford https://www.cebm.net/global-covid-19-case-fatality-rates/ (25 March 2020)

Richards, F. J., 1959. A Flexible Growth Function for Empirical Use, Journal of Experimental Botany, Volume 10, Issue 2, June 1959, Pages 290–301, doi: 10.1093/jxb/10.2.290

Setti, L. et al., 2020. Relazione circa l'effetto dell'inquinamento da particolato atmosferico e la diffusione di virus nella popolazione (in italian): http://www.simaonlus.it/wpsima/wp-content/uploads/2020/03/COVID19_Position-Paper_Relazione-circa-l%E2%80%99effetto-dell%E2%80%99inquinamento-da-particolato-atmosferico-e-la-diffusione-di-virus-nella-popolazione.pdf

Stafoggia M., Faustin A., Rognin M., Tessari R., Cadum E., Pacelli B., Pandolfi P., Miglio R., Mallone S., Vigotti M.A., Serinelli M., Accetta G., Dessi M.P., Cernigliaro A., Galassi C., Berti G., Forastier F., Gruppo collaborativo EpiAri. 2009 Epidemiol Prev. 2009 Nov-Dec;33(6 Suppl 1):65-76.

Ye X. , Peng L., Kan H., Wang W., Geng F., Mu Z., Zhou J., Yang D.. 2016 Acute Effects of Particulate Air Pollution on the Incidence of Coronary Heart Disease in Shanghai, China. PLoS ONE 11(3): e0151119. doi:10.1371/journal.pone.0151119

Wang D., Hu B., Hu C., et al., 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA, 323, 11, 1061–1069. doi:10.1001/jama.2020.1585.

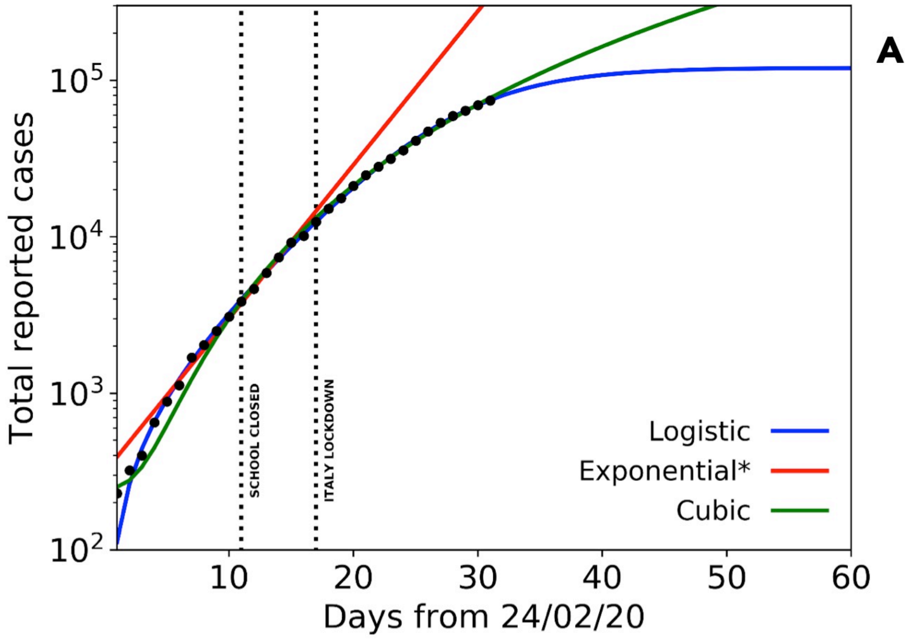
WHO (World Health Observatory), 2016. World Health Statistics tabacco smoking https://apps.who.int/gho/data/node.sdg.3-a-viz?lang=en

Worldometer 2020 https://www.worldometers.info/coronavirus/#countries

Country	***Cases***	***Deaths***	***CFR (%)***
Lombardia (Italy)	*28761*	*3174*	*11.0*
Italy	63927	6077	9.50
Iran	23049	1812	7.86
Spain	29470	2311	7.84
UK	6650	335	5.03

Netherlands	4749	213	4.48
France	19856	860	4.33
China	81171	3277	4.03
Japan	1193	43	3.60
Turkey	1529	37	2.41
Belgium	3743	88	2.35
Denmark	1460	24	1.64
Switzerland	8795	120	1.36
Sweden	2046	27	1.31
USA	43781	555	1.26
South Korea	8961	111	1.23
Canada	2091	24	1.14
Portugal	2060	23	1.11
Ireland	1125	6	0.53
Austria	4474	21	0.46
Germany	28480	123	0.43
Norway	2625	10	0.38
Australia	1887	7	0.37
Israel	1442	1	0.06

Table 1. Numbers of recorded infected people and deaths in several countries, on March 23rd 2020. Also indicated is the CFR (Case Fatality Ratio) defined as the ratio between the number of deaths and the number of recorded cases.



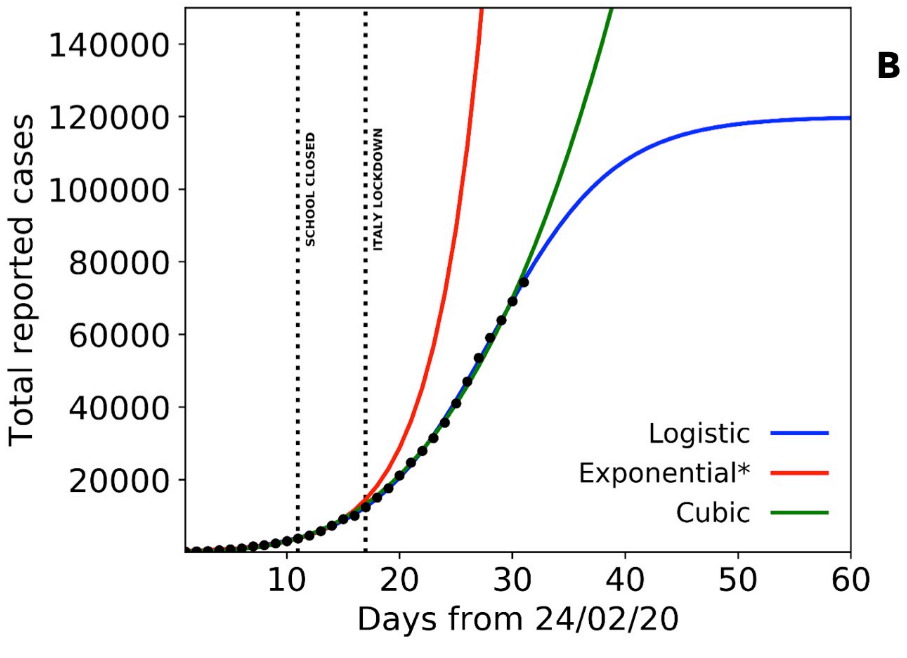


Figure 1. Total COVID-19 reported cases in Italy from February, 24th to March, 25th according to Protezione Civile (black dots) with the logistic (blue solid line), exponential (red solid line) and cubic (green solid line) Dotted black vertical lines mark the dates of Italian school lockdown and Italy total lockdown; the asterisk indicates that the exponential fit is based on data until March, 9th. A) fits obtained from the data in semi-logarithmic scale; B) fits obtained by data in linear scale.

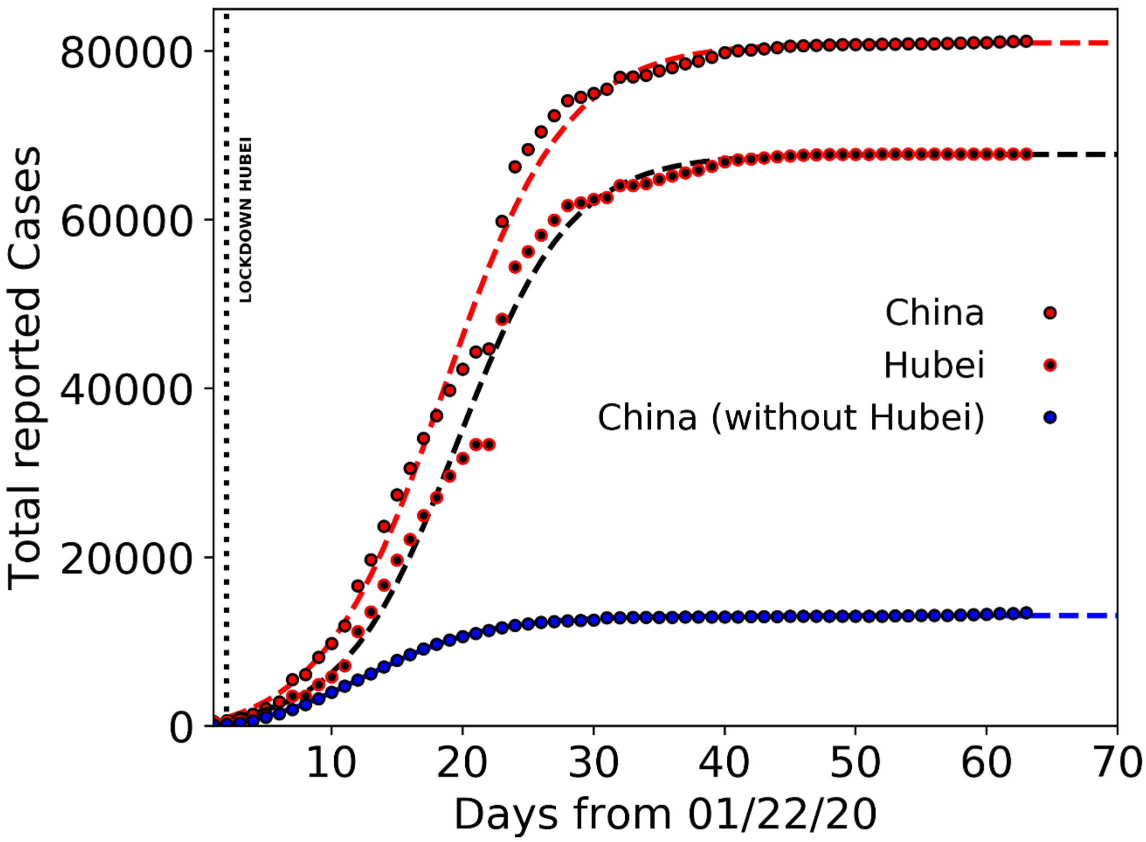


Figure 2. Total Covid-19 cases reported in China from January, 22nd to February, 25th according to Johns Hopkins University repository. Black-edged red dots, red-edged black dots and blue-edged black dots represent total Covid-19 cases registered in China, region of Hubei and China without Hubei, respectively while red, black and blue dashed lines are the logistic fits of the Chinese, Hubei and China without Hubei data, respectively.

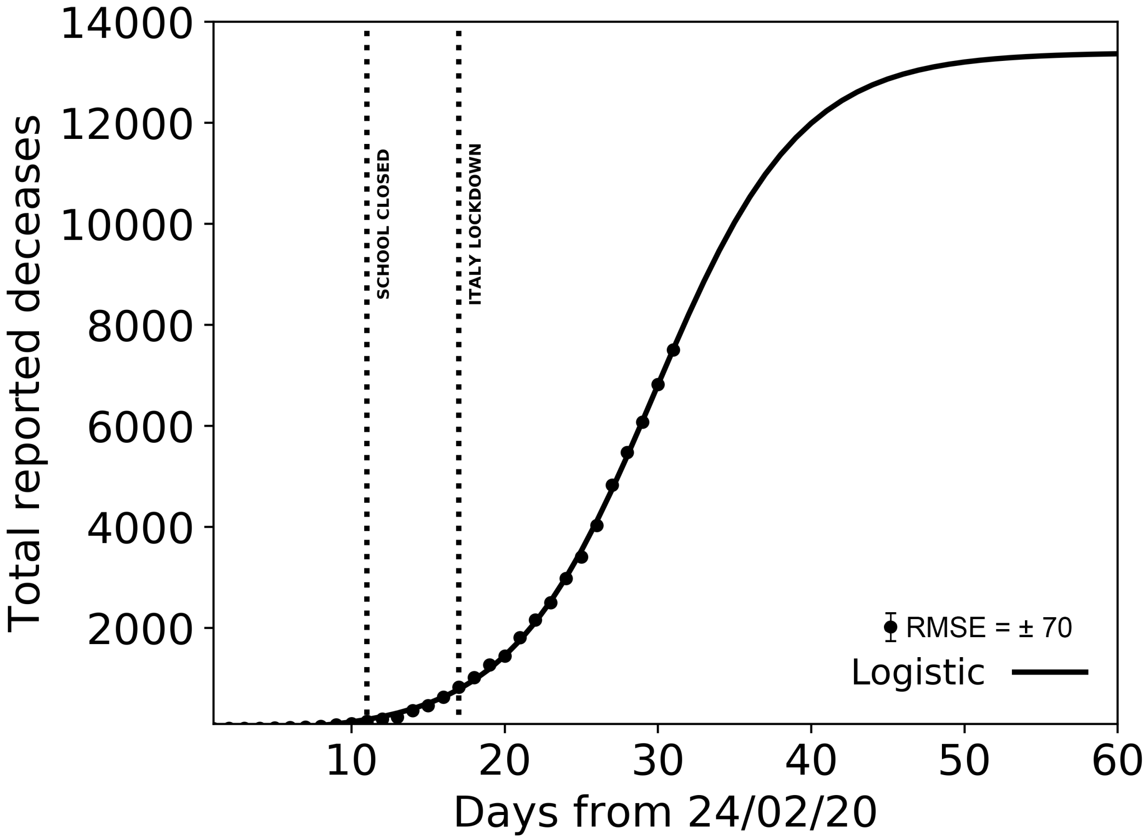


Figure 3. Total deceases reported in Italy from February, 24th to March, 25th according to Italian Civil Protection (black dots) and the logistic fit (black solid line) obtained from the data. Dotted vertical lines mark the dates of Italian school lockdown and Italy total lockdown. Also shown is the Root Mean Square Error (RMSE) computed from the data misfit.

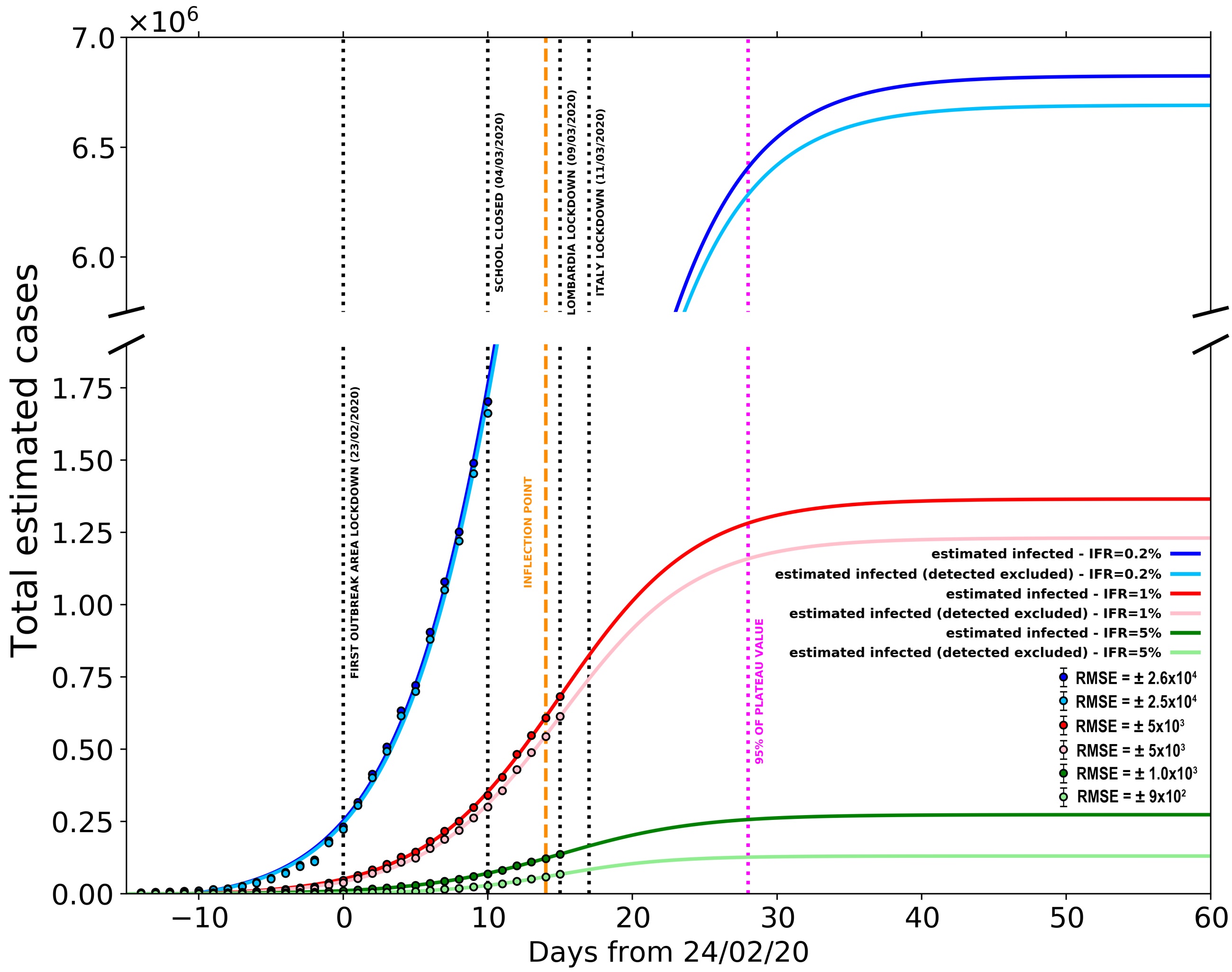


Figure 4. Estimated (total and undetected) COviD-19 cases in Italy based on three different IFR hypotheses: 0,2% (blue and sky blue dots), 1% (red and pink dots) and 5% (green and light green dots). Blue and sky blue solid lines represent logistic fits of total and undetected estimated cases with IFR=0,2%, respectively. Red and pink solid lines represent logistic fits of total and undetected estimated cases with IFR=1%, respectively. Green and light green solid lines represent logistic fits of total and undetected estimated cases with IFR=5%, respectively. Black dotted vertical lines mark the dates of Codogno area lockdown, Italian schools lockdown, Lombardia lockdown and Italy lockdown. Dark orange dashed vertical line marks the inflection points of the three curves, representing the total infected estimates; magenta dotted vertical line marks the 95% of the plateau of the three curves. Also shown are the Root Mean Square Errors of each curve (corresponding color), computed from the data misfit.