Research Article

Statistical investigation on covid 19 epidemics

Abstract

This paper describes a statistical investigation on the virus infection in Italy. Analysis of epidemic data and simulation tests, demonstrate the usefulness of statistical approach to foresee in advance the contagion widespread and organize preventive containment measures. The approach has general character, it is valid and applicable also to other Countries.

Keywords: covid, epidemic, analysis, statistics, simulation tests

Sebastiano Tosto
ENEA Casaccia Research Center, Italy

Correspondence: Sebastiano Tosto, ENEA Casaccia Research Center, via Anguillarese 301 00060 Roma, Italy, Tel +39 069968709, Email stosto@inwind.it

Introduction

The persisting alert status required by the recent epidemic Covid-19 virus all over the world, has stimulated the necessity of investigating new medical therapies to face the epidemic and evidenced the usefulness of predicting somehow the widespread of the infection process in the territories. The latter aspect, although having prevalent non-medical character, has relevant social interest to program in advance and organize financial interventions by the local health Institutions. For this reason, has been carried out the present statistical study on the number $n_i = n_i(t)$ of infected individuals as a function of time and related consequences in Italy. The basic questions that motivated this study were: did the progressive widespread of infection be somehow predictable since the early days of its evolution? Could the number of deaths be reasonably estimated and the probability of recovery rationally assessed? Despite the core problem of these questions has epidemiological character, there is a reasonable chance that the statistical investigation of data released by caregiver Institutions proves useful, whatever the key medical factors that govern the actual evolution of the infection might be. The ramp up stage of n_i is intuitively understandable. In fact n_i are acknowledged after a natural incubation time of the virus and successive medical tests to evidence its presence in the patients; all of this implies a latency time lapse during which number n_i of new infected persons, sometime asymptomatic, increments faster than its conversion to the numbers of recoveries $n_h = n_h(t)$ and deaths $n_d = n_d(t)$. The ramp down of n_i is also understandable because of several reasons, e.g. containment measures of the contagion promptly organized, enhanced efficacy of therapies meanwhile fine-tuned, statutory quarantine of infected individuals and eventually herd protection during the advanced stage of the epidemic. Clearly between these actions there is a transient stage where both effects balance, whence the occurrence of a peak of n_i as a function of time. This is in brief the statistical core of the endemic infection, which consists of numbers and mathematical functions that mirror and codify the activities in progress of Hospitals and Research Institutes. The ability of the statistical methods to summarize the behavior of great numbers of events of any kind, suggests the chance of elaborating a calculation model that accounts for the numerical aspects of the collective medical problem.

The present paper aims to investigate what kind of information can be obtained examining uniquely the infection data available in Italy along with that of recoveries n_h and deaths n_d. As the results appear encouraging, nothing prevents thinking that an analogous study can be carried out even elsewhere.

The statistical model

The Figure 1 shows the source of data released by "Protezione Civile": the ordinate reports the pertinent values of the aforesaid three

numbers, the abscissa expresses the time in days in the year 2020; the initial time is therefore February 24-th, the final time is May 10-th corresponding to the last reading useful to implement the statistical calculations presented here. These time boundaries are assumed enough to delineate the phenomenon in agreement with the purposes of the present paper. The data implemented for the calculations are cumulative at national level and gender inclusive (male + female); also, by definition, they are irrespective of the age of subjects, pre-existing state of health of the individuals and different realities of the regional health care organizations. The basic hypothesis is that n_i is anyway representable by a time function $F = F(t)$ characterized by a maximum at $t = t_{\max}$ to which corresponds a peak number $n_i = n_i^{\max} = n_i(t_{\max})$ of infected, as it is shown in figure 1. The following calculations are carried out as a function of time expressed in days starting from February 24-th, which becomes day 1.

As any statistical analysis requires homogeneous and comparable samples, hold the following basic assumptions about the mathematical processing of data:

i. The plots of fig 1 implement numbers sufficiently homogeneous to allow the self-consistency of a statistical model throughout the data sampling period. This implies for example that the initial measures of containment of virus infection do not change as a function of time, because this would perturb the natural course of recoveries and deaths with respect to the initial conditions.

ii. The change of environmental conditions to which the virus seems sensitive, e.g. typically the temperature or the degree of humidity, are not explicitly considered. A deep knowledge of the vitality of the virus and its importance from an epidemiological point of view are certainly essential factors for predicting behaviour and impact on the population. Yet, in lack of pertinent information, these factors and even the ability of the virus to modify itself are neglected.

The data in Figure 1 is the reference standpoint for any subsequent reasoning. In all of the following plots, the graphic symbols signify the respective experimental data, the curves characterise the trend of the mathematical functions aimed to represent them. The overlapping of plots and data is systematically shown throughout this paper to verify the chance of describing correctly the medical assessment via appropriate mathematical functions. This point is not trivial, because in fact the statistical homogeneity of data implemented for calculations is not controllable "a priori". The statistical parameters of the curves calculated by the statistical software are numbers with several decimal places, omitted or rounded for brevity; it is essential to demonstrate that best fit analytical functions consistent with the observed data effectively exist, whatever the actual computer numerical outputs might be. In other words, the matching between experimental symbols

and representative curves rather than the inherent mathematical details is the simplest proof about the reliability of the present calculation model.

The paper consists of two parts

The first one concerns all infection data available up to the time considered in the present paper, with the aim of describing the whole infection event as comprehensively as possible.

The second part repeats some selected calculations considering a few early data only: the aim is to show that actually these initial data are enough to extrapolate reasonably the whole event, thus evidencing that crucial features like peak time and related maximum numbers of infected, recoveries and deaths could have been decently estimated since the beginning of the infection. In fact, the calculations able to reproduce the observed data also provide extrapolated information.

Just that's the main purpose of the model. Three reference books on the standard statistical techniques are reported.[1–4] The present calculations have been carried out implementing the "nonlinear fit" statistical packages of MAPLE software. An ample literature has been produced recently to activate medical resources and face urgently the pandemic phenomenon; as expected, the prevalent character of these papers concerns the medical/clinical aspects of the virus problem. See[5] as concerns the infection in U.K.[6] as concerns virus diagnostics,[7] for prevention,[8] for implications of infection,[9] for transmission from animals to humans and for asymptomatic contacts,[10] for infection in Germany.

Among all medical aspects of the Covid 19 infection, the present paper focuses the attention of the reader on the impact of the virus aggressiveness on the population, strictly related to the statistical parameters emerging from the calculations.

Results of the statistical analysis of the trends depicted in the Figure 1

The search for the test function F suitable to represent the observed data of virus spread controlling $n_i = n_i^{obs}$ among the population must account for some mathematical requirements:
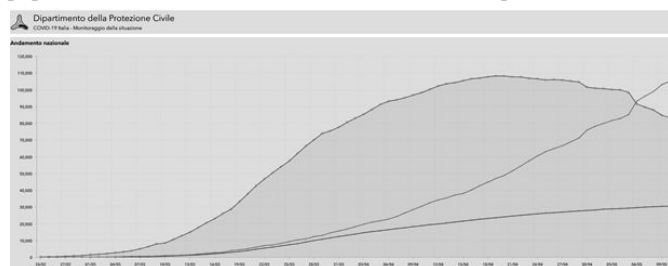

Figure 1 Source of data released by Protezione civile.

It must have a maximum as a function of time;

i. It must tend to zero for time tending to infinite, otherwise the number of new infections would persist indefinitely, contrarily to what emerges from epidemiological data historically occurred in the past;

The calculated peak must be later than the first detection date, otherwise F would automatically be incorrect;

Subsequently, a similar analysis will be also extended to the statistics of deaths and recoveries.

The results of a preliminary data analysis using a Gaussian "bell" type test function, typically used to represent statistical phenomena

that fulfil in principle the aforementioned boundary conditions, did not succeed; the initial trend of the Gaussian curve before the maximum increases too slowly to justify the number of n_i of (Figure 1), which instead have a more drastic exponential growth. For brevity this result is not reported here. A better representative function was instead the following modified Gaussian

$$F_i = A(t + t_1)^x \exp\left(-k(t - t_0)^2\right), \qquad (1)$$

With A, t_0, t_1, k, x best fit numerical parameters to be found. The problem is just to calculate these constants in order to reproduce and then extrapolate the few initial data available as reliably as possible. The graph in Figure 2 shows the result of nonlinear best fit regression of the representative function (1). Omitting here the mathematical details on the choice of the analytical form of F, shortly sketched below, we emphasize that the time factor $(t + t_1)^4$ is essential to modify the Gaussian according to the high increasing rate of the early n_i; this factor is thus representative of the clinical aggressiveness and infectiveness of the virus.
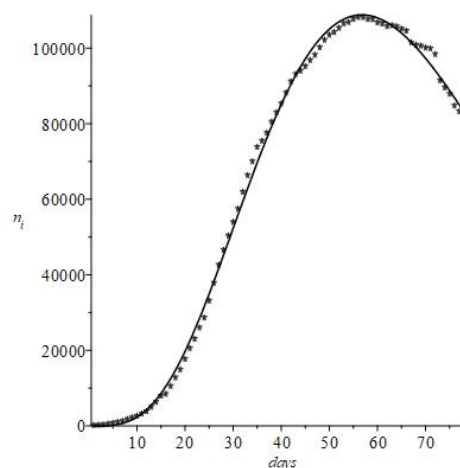

Figure 2 Result of nonlinear best fit regression of the representative function.

It is noted that with this representative function, the agreement with the experimental contagion data is reasonable, while the negative exponential in (1) is actually predictive of t_{max} and subsequent descent of the number of contagions as a function of time previously postulated "a priori".

With this function it is simple to calculate data of interest, summarized in the following eqs (2), which quantitatively account for the calculated peak number of infected persons n_i^{max} and the time $t(n_i^{max})$ to reach this peak corresponding to $\partial F / \partial t = 0$, i.e.:

$$t(n_i^{max}) = 56.8 \quad n_i^{max} = 108897;$$

the comparison of n_i^{max} with the observed number of cases n_i at the day of peak yields

$$n_i^{obs} = 108014 \quad deviation = 0.82\%. \qquad (2)$$

The fractional time could mean in principle the reaching of peak time between the 56-th and 57-th day after February 24-th; actually, this result must be intended as a numerical output with mere statistical meaning. What is important is that the deviation between calculated and observed number of infects is satisfactory. The peak number n_i^{max} of infected people is, as of today's date, about 1.4 per thousand of the Italian population with the present data used for the calculations.

Apparently, this calculation seems redundant, once having already visible all experimental values of n_i before and after the

peak. Actually, these results are significant for the comparison with the corresponding extrapolated values of the next section; it will be possible to assess quantitatively the predicting ability of analogous calculations obtained examining a few early data only.

The next plot of (Figure 3) represents in an analogous way the numbers of recoveries and deaths to complete the description of the global infection. The figure is obtained simply with the data corresponding to the days of (Figure 1) via the respective best fit functions, which are respectively

$$F_h = A \frac{(t+t_0)^3}{1 + B(t+t_1)^4}$$

$$F_d = A' \frac{C(t+t_0') + D(t+t_0')^2 + (t+t_0')^3}{1 + B'(t+t_1')^3}, \qquad (3)$$

Where the capital letters and $t_{0,1}$ are the statistical parameters to calculate explicitly the time trends of n_h and n_d. The (Figures 2&3) concern raw observed data, to show the actual chance of representing them via mathematical functions. A further important point to support the validity of the present approach is just the statistical nature of the infection, which appears in fact confirmed defining the function

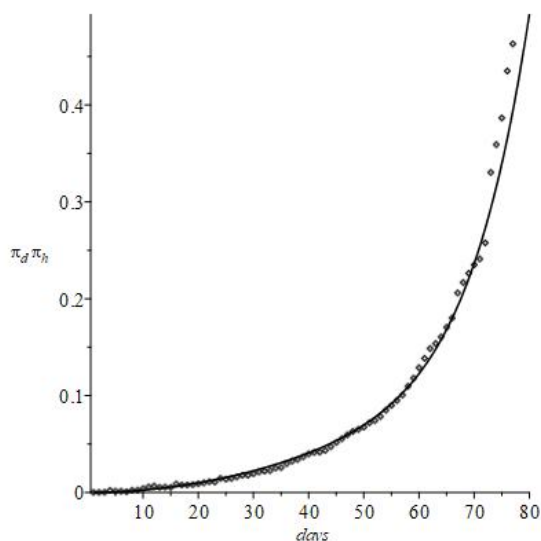$$\pi_h \pi_d = \frac{n_h}{n_i} \frac{n_d}{n_i} \qquad (4)$$

Figure 3 Numbers of recoveries and deaths to complete the description of the global infection.

Plotted in the Figure 4. Assuming that both n_h and n_d are consequences of a unique cluster n_i of infected individuals, π_h and π_d are definable as probabilities. The ordinate of the plot reports the joint probability that among n_i cases of infection there can be both recoveries and deaths, whose respective probabilities are combined according to the law of independent events. Indeed it is reasonable to think that π_h and π_d concern in fact independent classes of healthy and unhealthy, smoker and non-smoker individuals of different gender and age, whose immune defences imply different reactions to the infection.

On the one hand it is reasonable to think that π_h and π_d depend specifically upon the current clinical anamnesis of individuals with their own past stories, upon sex and all the other variables that contribute to cumulative (Figure 1). So, the collective trends of π_h

and π_d concern in fact independent classes of healthy and unhealthy people.
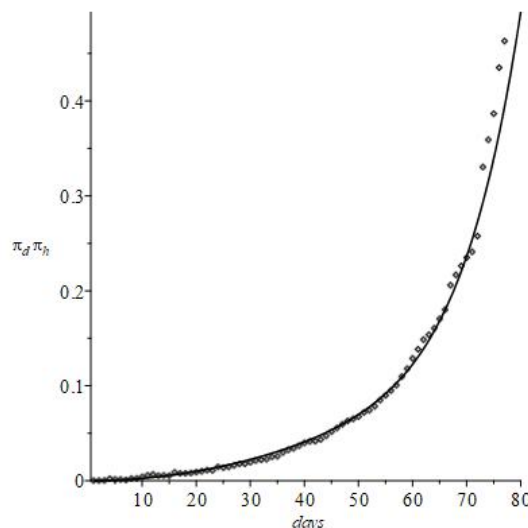

Figure 4 Statistical nature of the infection.

On the other hand, from a statistical point of view, in a sample of many subjects the differences from one individual to another appear summarized in a single collective behaviour whose global result is uniquely definable, the one summarized in the figure 1 that is the target of the present paper.

Just the regular trend of this figure encourages thinking that, despite the wide variety of biological parameters triggered by the complex virus/human body interaction, it is anyway possible to identify a well-defined collective behaviour hidden in the apparent randomness of personal situations. The identification of this joint probability is interesting as it concerns the chance of combining the available data to infer more information and deserves a tentative explanation.

Let $\pi_1 = p_h P_d$ indicate an unhealthy patient with small probability p_h of surviving and thus more likely to die with high probability P_d, whereas instead $\pi_2 = P_h p_d$ concerns the opposite chance for a healthy patient; P_h and P_d add further information to the left hand side of eq (4) because now the probabilities π_1 and π_2 emphasize explicitly the health states of the respective patients. Let the numbers n_h be large enough to gather individuals that anyway recover at the end of their therapeutic cycle, regardless of whether their initial health conditions are in principle classifiable with both low or high recovering probabilities; a typical example of this idea is that some individuals recover without need of intensive therapy, other ones instead need unfortunately invasive and prolonged hospital treatments but anyway recover. So, the initial p_h and P_h are not crucial themselves in determining the final recovery probability n_h / n_i; in other words, different patients classified as p_h and P_h merge into a global n_h defining a unique n_h / n_i. An analogous consideration holds for n_d / n_i regardless of p_d and P_d: in fact it is not true that die only individuals in intensive therapy, i.e. the ones classified as P_d. So when calculating $\pi_1 \pi_2$ the respective health conditions represented by the probabilities $p_h P_h$ and $p_d P_d$ are statistically equivalent to and simply summarized by n_h and n_d of eq (4). As a matter of fact, the respective curve of fig 4 is well defined and thus explainable with $P_h p_h$ and $P_d p_d$ described by a unique time behaviour.

Although this conclusion was not strictly evident "a priori", such a correlation is directly confirmed in Figure 5. The plot shows that the

number of recoveries n_h and deaths n_d are both functions of time mutually correlated through their common origin from the number of infected n_i only and thus are in fact uniquely definable: the implicit functions $n_h = H(n_i)$ and $n_d = D(n_i)$ merge together into the function

$$\frac{\text{recoveries}}{\text{deaths}} = At + Bt^4. \tag{5}$$

The regression coefficients at the right of the equation, determined as described above, provide quantitative estimates of the progress of the contagion containment as a function of time.
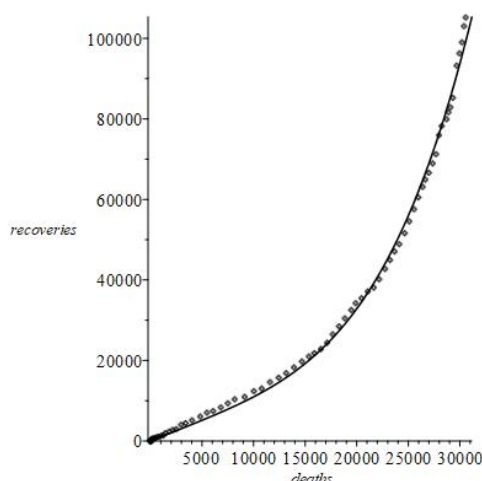

Figure 5 Number of recoveries and deaths.

The trend of n_h vs n_d evidences that the numbers of deaths and recoveries are well correlated in the population infected by the virus despite the diversity of the individuals. It is reasonable to guess that during the initial incubation period, soon after the virus attack, the n_i patients are in a transient state with suspended probabilities π_h and π_d before that they turn into either observed state defined by n_h or n_d; also here the solid line indicates the existence of a single ideal cumulative probability only, without discriminating serious patients recovered after intensive care or mild patients treated with simple drug therapy. The left-hand side represents the therapeutic efficacy; the different powers of t at the right-hand side concern short term and long-term ratios, when the second addend becomes expectedly predominant. So (5) measures the success of the full therapeutic cycle.

Herd immunity

The statistical model so far described has considered the raw n_i of (Figure 1) to describe in more detail in Figure 2 the ramp up law of the total number of infected people as a function of time towards to and around the peak. Presumably in this time range no further considerations are necessary besides the obvious requirement of data homogeneous and statistically representative. Yet there is no reason to exclude long term effects, like for example the so called "herd immunity". While the mere numerical analysis of data does not contain itself epidemiological information, this latter should be hopefully inferred from the former: no statistical model can be considered exhaustive without the ability of extracting information about this protective effect possibly hidden in the raw data. Since the herd immunity is reasonably ineffective in the case of the first initial data but becomes more important subsequently, the second half of the curve in the next (Figure 6) has been drawn cautiously by dashed line; without accounting for herd immunity, in principle this part of the plot estimates only the upper limit of the number of infections. Although

the satisfactory matching of raw n_i observed and calculated curve is good enough to neglect the necessity of corrections, at least with the data presently available, in this section however it is shown how to account for and assess the expected deviation from the long-term data.

In principle realistic values $n_i^{herd} \neq n_i$ accounting for the herd effect should be recalculated from the curve of (Figure 6) introducing a corrective factor $f(n_i) < 1$ at $t \gg t_{peak}$ through iterative calculations until convergence; however it requires hypothesizing an average number of contacts among a selected group of infected and recovered individuals to verify quantitatively the probability of a successful herd protection. In practice the approach to find an expression linking n_i^{herd} and n_i is more simply carried out by introducing a controllable correction factor that modifies reasonably the available short-term results already obtained.
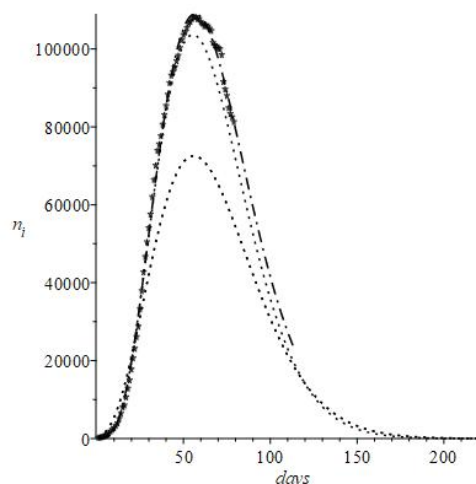

Figure 6 n_i observed and calculated curve.

An appropriate strategy to represent mathematically the herd effect is like this

$$n_i^{herd} \to n_i f(n_i) = \frac{n_i}{1 + Cn_i n_h} \tag{6}$$

Being C a proportionality constant; i.e. n_i is reduced by a variable factor which becomes progressively more important with the increase of n_i itself and with the number n_h of recoveries.

For $C = 0$ then $f(n_i) \equiv 1$, i.e. the herd effect is zero.

If $C \neq 0$, then the function (6) is proportional for $n_h \to \infty$ to n_h^{-1}. That is, if the number of recoveries increases then the number of infected individuals tends to zero, which is precisely what the protective effect of herd protection requires: from a probabilistic point of view, a healthy person surrounded only by recovered people or non-infected people has little chance of becoming infected.

Eq (6) is calculable with n_i and n_h extrapolated from the short-term data, where the herd effect is ineffective. Although a heuristic "ab initio" calculation model would be preferable, this equation yields the sought information once more by best fit statistical analysis of long-term infection data. In fact, the approach based on (6) is sensible only if the virus does not change; the early trend summarized by short term data is a fixed fingerprint of a given virus, whereas instead changes merely the long-term probability of infections by contact with recovered neighbours. Yet, if the virus evolves as a function of time, then a unique n_i curve is no longer representative; it should be replaced by two curves pertinent to the virus before and after modification, i.e. the herd effect is no longer definable via correction

of a unique short-term curve. However, the analysis of the long-term curve could in principle provide updated information about the virus modification occurred in the meantime; in this respect it is still sensible to check how the long term curve is modified by the constant coefficient C to infer information about the different short term trend.

An example of simulated results is shown in the Figure 6 with two test values of the constant C purposely chosen, i.e. $C = 10^{-10}$ and $C = 10^{-11}$.

This figure still reports by comparison the data of figure 1 with and without the correction introduced by eq (6); indeed, the main plot is just that of (Figure 2) plus an extrapolated part to be compared with two herd dotted curves. The results with $C = 10^{-10}$ show t_{max} equal to about 35 days, while the maximum peak of infected people n_{max} decreases to about 70,000 people infected. It is clear that this value of C foresees a too optimistic herd effect, which should have been already observed experimentally 50 days after February 24; since this is not realistic, the simulation was repeated with $C = 10^{-11}$, which in fact foresees the peak at 60 days, with a number of infected equal to about 102.000. As expected, anyway the alteration of the raw data of n_i is practically zero at $t < t_{max}$; the differences with respect to Figure 2 begin to become relevant much later only, reasonably at increasing n_i after the peak. The dot line that represents the long-term extrapolation of the best fit curve of n_i suggests its tendency to approach that of n_i^{herd}. Just this tendency to merge the early curves supports the validity of this reasonable value of C.

It can be concluded that this constant can be either (i) predetermined in order to simulate a herd effect practically negligible initially but progressively increasing with the increase of the infected and of the recoveries or (ii) determined uniquely by the statistical treatment of long term "tail data".

Early data analysis

This section concerns the results obtainable analysing the first 20 initial data only, $n_i^{red} < 20$ of the fig 1; the notation stands for "reduced" number of early observed data. The Figure 7 shows the result of best fit analysis of these few data only, which of course are insufficient to match decently the true trend of all observed data also reported by comparison. Nonetheless it appears that the peak time is still quite correct, it differs from the true peak time by about one day, whereas the related number of infected is underestimated by about 10% only. These results show however that the main epidemiological features of virus infection could have been realistically anticipated by a prompt analysis of statistical data. This is the true meaning of the present model.

Discussion

The overall indication that emerges from the present statistical model is that even a minimal alteration of environmental data causes a drastic variation in the spread of the virus and its ability of contagion, which is quantified by the best fit parameters of the exponential eq (1). On the one hand, owing to the exponential form, even small changes of these parameters affect strongly the dependence of n_i upon time. On the other hand, just this mathematical feature describing the contagion suggests that even elementary protection measures preventively actuated, e.g. safety distance between individuals or use of protection masks and so on, are effective in controlling the subsequent epidemic data. This conclusion appears evident in the Figure 6: even small changes of C, purposely introduced in the eq (6) to simulate a typical environmental effect such as herd protection, are enough to modify significantly the long-term evolution of the infection in progress.

Also, the fig 2 evidences an anomalous dispersion of data just after the peak, which implies an initial decreasing rate of n_i slower than predicted by the exponential of (1).
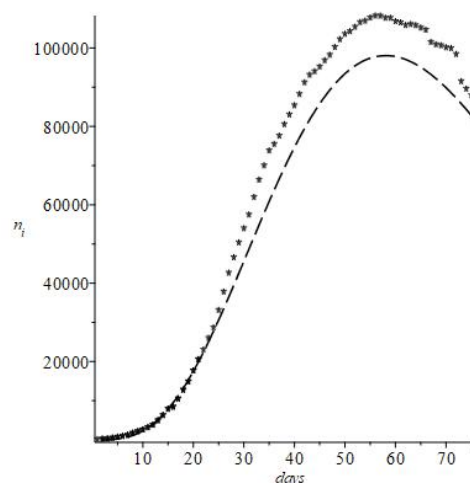

Figure 7 Result of best fit analysis.

This is not surprising. It is known in general that when a strong effect governs predominantly the kinetic of any process, an external perturbation simply adds a minor spread of data around the leading trend obtainable via statistical considerations only. Instead an equilibrium state in nature is by definition the resultant of many counterbalancing effects, thus particularly sensitive to any external perturbation; specifically, appear around the peak time the possible consequences due to the previous health conditions of the patients and their lifestyle.

Another problem is that the strict measures of restrain initially put in place by the Government authorities have been partially loosened, probably misinterpreting too optimistically the stage after the peak of infected patients as a symptom of reduced danger of contagion. After this short stage, however, the further n_i data have resumed the behaviour predicted by the eq (1).

It is commonly believed that the turning point of the infection advancement is identifiable with the peak time t_{max} of n_i; undoubtedly this is correct as concerns the effectiveness of the measures to prevent new contagions, but it does not highlight adequately the current state of the epidemic evolution.

Actually, the analysis so far carried out suggests the significance of the relationship between the incoming number n_i of hospitalized patients and outcoming numbers n_h and n_d representing either possible evolution of n_i after a successful or vain therapy cycle. Yet the experimental data of Figure 1 show the existence of a time such that $n_i - n_h < 0$, which also implies $n_i - (n_h + n_d) < 0$; the time evolution of n_i and n_h observed at increasing times suggests therefore the relevance of the number

$$N = n_i - (n_h + n_d)$$

that involves explicitly the effectiveness of the therapies on n_h patients. Clearly $N > 0$ implies overcrowding of hospitalized patients exceeding the chance of their efficient management; $N < 0$ implies the decreasing rate of new infected patients with respect to their conversion rate to either n_h or n_d. If however the growing rate of n_i decreases until becoming null at the peak time, while in the meantime the sum $n_h + n_d$ continues to increase, then one expects

first $N = 0$ at a given t^* and next even $N < 0$ at $t > t^*$. Hence, assuming that n_i still evolves at $t > t_{max}$ into $n_h(t)$ and $n_d(t)$, it happens that: (i) n_i tends to zero while contributing to the growth of n_h, (ii) n_d tends towards its final value of total deaths, (iii) N becomes more and more negative.

The (Figure 8) confirms these ideas: it appears that $t^* \approx 65$ days after which $N < 0$ indicates a number of new infected $n_i < n_h + n_d$ progressively decreasing.
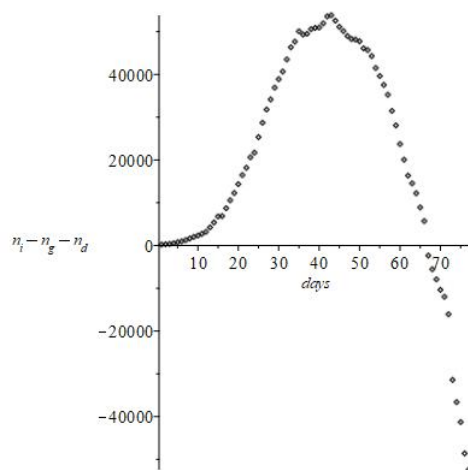

Figure 8 Number of new infected $n_i < n_h + n_d$ progressively decreasing.

On the one hand it is not surprising that the time t^* at which $N = 0$ does not necessarily coincide with $t_{max} \approx 57$ days calculated considering n_i only. On the other hand it is clear why N and not n_i is the crucial indicator of the turning-point between the aggressive stage of the virus and its progressive attenuation; when comparing n_i and $n_h + n_d$, the sign of N depends upon the relative rates at which the addends prevail each other.

All of this is explained in the Figure 9, showing that in effect N consists of three statistical distributions that end down with negative values; the three contributions of representative best fit functions resolved by the peak analysis are reasonably related to the three respective basic stages of contagion with $N \lessgtr 0$. The peak splitting equation has the analytical form

$$A(t-t_1)^4 \exp\left(-a(t-34)^4\right) + B(t-t_2)^6 \exp\left(-b(t-42)^2\right) + C(t-t_3)^4 \exp\left(-c(t-45)^2\right) \quad (7)$$
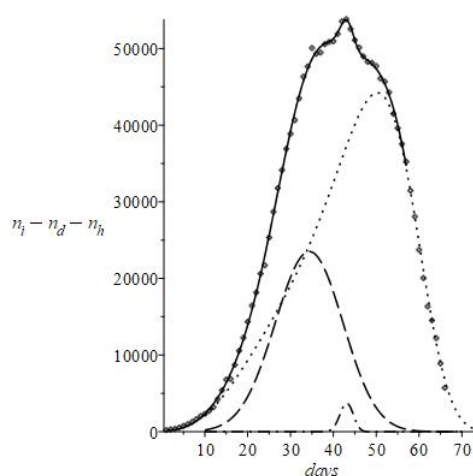

Figure 9 Effect N consists of three statistical distributions.

Conclusion

The statistical investigation provides reliable and significant indications on the epidemiological evolution of COVID-19, even extensible to extrapolated times. Although further data are still expected in the next days beyond the considered interval, the most critical aspects of the infection seem adequately represented even at the current end time.

The present information methodology has general character; it is valid and applicable without need of specific hypotheses besides the basic assumptions of any statistical model.

-The results show that the evolution of virus infection could have been predictable with reasonable approximation since the early days, which means good chance of organizing care structures and medical materials (e.g. face masks) useful in the most acute stage of the infection.

-The plots allow estimating quantitatively the efficacy of therapeutic protocols through the condition $n_i - n_h - n_d = 0$, which holds twice: at the peak time of n_i and before the onset of negative values $n_i - n_h - n_d < 0$ that defines the final stage of predominant contagion.

The approach is useful to forecast and plan future containment and coordination activities to support sufferance of sick.

Acknowledgements

None.

Conflicts of interest

The author declares that there is no conflict of interest.

References

1. Freedman DA. Statistical Models: Theory and Practice. Cambridge University Press. 2009.

2. Chiang CL. *Statistical methods of analysis*. World Scientific. 2003.

3. Schittkowski K. Data Fitting in Dynamical Systems. Boston: Kluwer; 2002.

4. Moss P, Barlow G, Easom N, et al. Lessons for managing high-consequence infections from first COVID-19 cases in the UK. *The Lancet*. 2020;395(10227):e46.

5. Cormac S. Coronavirus and the race to distribute reliable diagnostics. *Nature Biotechnology*. 2020;38(4):382–384.

6. Cohen J, Normile D. New SARS-like virus in China triggers alarm. *Science*. 2020;367(17):234–235.

7. Al-Omari A, Rabaan A, Samer S, et al. MERS coronavirus outbreak: Implications for emerging viral infections. *Diagnostic Microbiology and Infectious Disease*. 2019;93(3)265–285.

8. KW To, IFN Hung, JF W Chan, et al. From SARS coronavirus to novel animal and human corona viruses. *Journal of Thoracic Disease*. 2013;5(2):103–108.

9. C Huang, Y Wang, X Li, et al. Clinical features of patients infected with 2019 novel corona virus in Wuhan (China). *The Lancet*. 2020;395(10223):497–506.

10. Camilla Rothe, Mirjam Schunk, Peter Sothmann, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine*. 2020;382(10):970–971.