A statistical theory of the strength of epidemics.
The Italian Covid-19 case

Gabriele PISANO\textsuperscript{1} and Gianni ROYER-CARFAGNI \textsuperscript{1,2}

\textsuperscript{2}Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, I-43100 Parma, Italy

\textsuperscript{1}Construction Technologies Institute - Italian National Research Council (ITC-CNR), Via Lombardia 49, I-20098 San Giuliano Milanese, Milano, Italy

Abstract

We propose a theory of the lethality of epidemics which conforms to the classical Weibull model of the statistical strength of brittle structures which contain a population of crack-like defects. A relation between the probability of developing a critical pathology and the risk of death is adopted in the same manner that the mechanical strength model correlates the statistical spatial distribution of crack size to the risk of fracture at a prescribed stress level. The fracture toughness criterion that determines the onset of catastrophic crack propagation in the structure suggests a death criterion that considers the health system’s capability of treating the disease. Our theory relies upon a renormalization of the life-time that assigns the elderly a higher probability of developing a critical pathology. The theory predicts scaling laws for the age of death which are in agreement with the data collected in the Italian regions and provinces before and after the spread of COVID-19. A ”strength index” of the epidemic is defined and calculated by categorizing the mortality data, according to the victims’ age, from comparable periods of observation in both pre-epidemic and epidemic conditions.

KEYWORDS: Mathematical epidemiology; Covid-19; probabilistic mechanics; Weibull statistics; fracture mechanics.

\footnote{Corresponding author}

Email addresses: pisano@itc.cnr.it (Gabriele Pisano), gianni.royer@unipr.it (Gianni Royer-Carfagni)
1 Introduction

The purpose of this article is to show how consolidated models in mechanics, in particular in the stochastic mechanics of brittle materials, may give ideas for research in a very different area of speculation in order to quantitatively define, on a statistical basis, the effects of epidemics. The mechanical model, governed by similar equations, will allow to establish an intuitive correspondence between concepts well known to engineers, such as the stress state in a body, with other quantities, like the lethality of the epidemic. The comparison with the experimental data, which in this case are the current statistics of mortality in Italy in this period of Covid-19 infection, indicates the reliability of the proposed approach, with its potentialities and limitations. Indeed, since in the western world the epidemic from Covid-19 has spread in Italy first, the Italian case represents a paradigm for many other countries.

Mathematical models have been used for almost a hundred years in epidemiology. In their celebrated seminal work Kermack and McKendrick [1] proposed the Susceptible-Infectious-Recovered (SIR) model, detailing how a virus is transmitted from an infected person to a healthy one [2, 3]. Their ingenious ideas have been developed, modified and widely adopted [4, 5, 6, 7, 8, 9, 10], also within a probabilistic framework [11]. Our approach is complementary since we propose a new way to quantify the strength of an epidemic starting from a particular statistical treatment of measurable mortality data in a certain territory, within a given period of observation, which necessarily must consider the age of the victims. The paramount role of the demographic structure in the prediction of the expected number of deaths has been confirmed by recent studies [12], as the division by age is more important than the structure of families [13] in the analysis of the social contacts that influence the spread of a virus [14]; social distancing and other policies to slow transmission should always consider the age composition of local and national contexts.

In 1939 Weibull [15] proposed a theory of the strength of a structure made of a brittle material based on the idea that it contains inherent flaws of random spatial distribution, shape and size. The risk of fracture at a prescribed stress level is therefore determined by the probability that a large-enough defect exists which will propagate catastrophically at that level of stress. The flaw shape is typically assumed to be a flat slit (crack) of size \( \delta \). If \( \sigma \) denotes the component of stress normal to the crack plane, the toughness criterion of Linear Elastic Fracture Mechanics (LEFM) dictates that the crack will grow catastrophically (in mode I) when the so-called Stress Intensity Factor \( K_I = Y \sigma \delta^{1/2} \), where \( Y \) is a constant determined by the geometries of structure and crack, reaches the characteristic material’s fracture toughness \( K_{Ic} \) [16]. The Weibull model conceptualizes the structure as a chain comprised of equally stressed links, and whose resistance to fracture is limited by the strength of the weakest one. The expected strength is therefore a decreasing function of structural size (size-effect [17]), because as the length of chain increases there is a concomitant increased probability of the presence of a weak link.
The fight against epidemics requires a multi-disciplinary approach involving integrated cycles of prevention, response and recovery to support epidemiology [18]. The Weibull theory of structural strength suggests the possibility of formulating a statistical theory that quantifies the “strength” of an epidemic, where strength now refers to a precise quantitative measure of the lethality that results from the pathologies developed in individuals. Towards this end, it is first necessary to define what is the counterpart, in the infected human body, of the size of the stressed structure (the length of the chain), now associated with the probability of developing pathologies that are critical for the level of epidemic. This correlation, which is informed by existing epidemiological models, will be made with the age of the individuals which, however, needs be renormalized in order to take into account that the old have a higher probability than the young to develop severe pathologies. Such a renormalization, which depends upon genetics, quality of life and efficiency of the health system, is calibrated starting from mortality data in pre-epidemic conditions.

An index of epidemic $I_e$ is thus defined to measure its strength, starting from the theoretical modelling of the probability of death as a function of the severity of the epidemic and the renormalized age of individuals. The index is calculated from the excess mortality rates with respect to pre-epidemic conditions of previous years, which are considered [19, 20] the most reliable (and least-assumption laden) records of the effect of the epidemic, rather than the number of infected that is difficult to determine. The accurate estimate of $I_e$ over time and place is useful to follow the diffusion of the virus, to recognize the emergence of autonomous outbreaks but, even more so, to timely define appropriate countermeasures such as lockdown (prevention of spread), and to decide when, where and how to loosen the countermeasures (management of the epidemic). The theory is applied to the case of Italy.

The index $I_e$ has been calculated by comparing weekly and monthly mortality tables for February-April 2020 with the same periods of previous years. Considering territories of different size (regions and provinces), we estimate the epidemic flow over time, the emergence of autonomous outbreaks and the effect of the imposed lockdown. Monitoring the evolution of the index may be useful to manage the phase of recovery in the next future.

Are there other ways to measure the strength of the epidemic? Our experience in Italy indicates the use of the “crude” ratio between the number of deaths in epidemic and pre-epidemic conditions [21, 22, 23], but if we let ourselves be guided by a theory, we are led much further than where qualitative reasoning and empirical laws, based on data manipulation, could bring us, thanks to the formulation of precise equations and the definition of a single index, which accounts for the specific effects on the elderly and our capability of treating the disease. Our findings do not contradict the results of the observations; indeed they seem in perfect agreement with the statistical data relating to the deaths occurred in various territories of Italy as a consequence of the development of Covid-19, territories that are very different in terms of expectancy of life and diffusion of the virus. The simple comparison of the number of dead does not consider the demographic structures and consequently, as it will be verified, weights the deaths in an undifferentiated way, without being able to attribute a specific correlation with the effects of the epidemic.
In order to consider the problem mathematically, we start from hypotheses that, although moving away from the complex reality, should provide a sufficiently approximate view of the phenomenon. The basic results from the theory are presented in Section 2(a), whereas the mathematical calculations are detailed in Section 3(a). The representation will certainly be coarse but, if it is simple, it will be possible to conveniently apply the calculation and to check quantitatively, or at least qualitatively, if the obtained results correspond to the measured data and, therefore, verify the validity of the starting hypotheses. The comparison of the theoretical predictions with the mortality tables in Italy is illustrated in Section 2(b), while the precise way in which the data have been manipulated is indicated in Section 3(b). The agreement so far obtained is very promising. Although further studies are certainly needed to demonstrate whether the knowledge in materials science can be conveniently borrowed in the broad field of epidemiology, and viceversa, we are confident that new results can only be achieved with mutual interdisciplinary progress.

2 Results

(a) The theoretical framework

We start from the assumption that the life of a person is as a chain composed of life-segments that represent the rings in the chain analogy. The number of rings is successively increased to reflect the aging process, as rings represent potentially critical flaws. One may consider that a life-segment is a solar year of age, or submultiple, but this view is too simplistic. “It is mathematically demonstrable that the concept of time is closely related to age: time passes faster for old people”. This quote, by the American writer Alvin Toffler (1928-2016), introduces the idea that the usual life-time unit, e.g. the solar year, cannot indicate the reference scale to measure the aging process throughout the whole of human life. If a life-segment represents the nominal unit associated with one spot where potential damage may develop with equal probability, the number of life-segments contained in one solar year of age should be higher for an older individual than for a younger one.

Let $\Delta A_n$ denote the reference nominal life-segment. The real age $A_r$, expressed in number of solar (real) time-segments $\Delta A_r$, is rescaled to the nominal age $A_n$, equal to the number of $\Delta A_n$, through a renormalization group, expressed by $A_n = F(A_r)$. Elaborating the data on mortality recorded in Italy, we are led to consider a function $F(A_r)$ of the form

$$A_n = F(A_r) = A_r + (A_r - 45)^{\gamma_1} + (A_r - 70)^{\gamma_2}, \quad \gamma_1, \gamma_2 > 0,$$

where $A_r$ and $A_n$ are expressed in number of solar and nominal years, respectively, and $(\cdot)^+$ denotes the positive part function$^1$. Hence, there are two step-changes in life, the first at 45 and the second one at 70 solar years of age, beyond which, roughly speaking, “each

\hspace{1cm}$^1$This takes as input any real number and outputs the same number if this is non-negative, 0 otherwise.
year counts more”. The renormalization \( A_n = F(A_r) \) may depend, among other factors, on race, genetic heritage, environmental factors, quality of life, efficiency of the health system, and gender (male or female); it is assumed to be valid, on average, within a homogeneous ensemble.

The probability of death, with reference to a given period of observation \( \Delta t \) corresponding to the time interval in which mortality is monitored, is represented by the probability of developing in \( \Delta t \) a pathology that is critical for the level of epidemic. Such probability shall be higher for the older, according to the nominal age. The approach presumes the definition of variables measuring the level of pathology and epidemics, for which balance equations will be established on a statistical basis. Referring to Section 3(a) for the detailed calculation, here the basic results are synthetically reported.

We suppose that the severity of the pathology is synthetically measured by one variable \( \delta \). The probability of developing pathologies is defined for each nominal life-segment (i.e., in each potential damage spot), statistically treated as independent of other life-segments. Since serious pathologies (high \( \delta \)) are less probable than mild pathologies (low \( \delta \)), we consider a law \( \text{à la Pareto as per (3.1) in Section 3(a), expressed by the probability density function} \)

\[
p_{\Delta A}(\delta) \propto \delta^{-\alpha}, \quad \text{with} \quad \alpha > 1, \quad \text{of developing} \quad \delta \quad \text{for each} \quad \text{nominal life-segment.} \]

The statistical calculations provide the probability of developing \( \delta \) in the whole nominal life \( A_n \), composed of \( A_n/\Delta A_n \) life-time segments, in the form

\[
P_{A_n}^{\geq} (\delta) = 1 - \exp \left[ -A_n \left( \frac{\delta}{\eta} \right)^{1-\alpha} \right], \quad (2.2)
\]

where \( \eta > 0 \) depends upon a physiological minimal level of pathology in each \( \Delta A_n \), which has to be assumed to enforce a finite lifetime. Notice that probability increases with \( A_n \) because the higher is the number of life-segments of the individual, the higher the risk of developing pathologies.

We further postulate that the severity of epidemics is measured by another variable \( \sigma \): the higher is \( \sigma \), the lower is the pathology \( \delta \) necessary to cause death. In the simplest case, we assume the LEFM-like [16] death criterion

\[
Y \sigma \delta^{1/\beta} - K_{Ic} = 0 \quad \Rightarrow \quad \delta = \left( \frac{K_{Ic}}{Y \sigma} \right)^{\beta}, \quad (2.3)
\]

where \( Y \) is a dimensional constant and \( K_{Ic} \) depends upon human physiology. The parameter \( \beta > 0 \) models our capability of treating the disease with effective treatments: all other parameters being equal, the higher \( \beta \), the higher the deadly pathology.

Substituting in (2.2), the probability of death as a function of \( \sigma \) and the real age \( A_r \) reads

\[
P^D(A_r, \sigma) = 1 - \exp \left[ -F(A_r) \left( \frac{\sigma}{\eta_0} \right)^m \right]. \quad (2.4)
\]
This is a two parameter Weibull distribution with shape-parameter $m = (\alpha - 1)\beta$ and scale-parameter $\eta_0 = K_{IC}/(Y\eta^{1/\beta})$. Note that both are independent of $\sigma$, which represents the only variable describing the severity of the epidemics. The nominal age $A_n = F(A_r)$ determines the “size-effect” consequent to the fact that the higher is the number of life-segments of the individual, the higher is the risk of death.

In the mechanics of brittle materials [24], the counterparts of $\delta$ and $\sigma$ are the size of the cracks and the stress level [25, 26], respectively. The death criterion recalls the toughness criterion in LEFM [16], for which $\beta = 2$. The nominal life represents the size of the material body, which affects the probability of failure since the higher it is, the higher is probability of finding a crack of critical size with respect to the applied stress, according to classical LEFM.

In order to measure the strength of the epidemic, observe that (2.4) may be written as

$$\ln \ln \left[ \frac{1}{1 - P^D(A_r, \sigma)} \right] = \ln[F(A_r)] + m \ln \left[ \frac{\sigma}{\eta_0} \right],$$

which represents a straight line in the $\ln \ln \{[1 - P^D(A_r, \sigma)]^{-1} \} - \ln[F(A_r)]$ plane. Its intercept with the ordinate axis provides the value of $m \ln[\sigma] - m \ln[\eta_0]$, where both $m$ and $\eta_0$ do not depend upon $\sigma$. This plane somehow represents the counterpart, in this theory, of the Weibull plane [15], commonly used to interpret the strength of brittle material on a statistical basis and, because of this, it will be referred to as the epidemic Weibull plane. This representation is very useful because it allows a direct control of the reliability of the theory: one can collect statistical data on mortality and directly verify if they fall aligned in this plane.

Our aim is thus to compare a statistically-representative sample under “ordinary” (pre-epidemic) conditions and after the infection from Covid-19 (epidemic condition). No one can say what is the value of $\sigma$ in pre-epidemic condition; certainly, it is not associated with Covid-19 but rather to other kinds of mild epidemics (for example, seasonal flu). However, it cannot be zero, otherwise the probability of death would be null. For a comparative evaluation, without losing generality, we set $\sigma = 1$ in the pre-epidemic state.

Hence, from the analysis of the mortality data in pre-epidemic condition one can find the best-fit line in the epidemic Weibull plane and, setting $\sigma = 1$, from the intercept, one finds $-m \ln[\eta_0]$. Repeating the same argument for epidemic conditions, one obtains $m \ln[\sigma]$ since $-m \ln[\eta_0]$ is now known.

The “index of epidemic” should be a quantity directly measurable from mortality data that increases with $\sigma$. A parameter with such a property is

$$I_e = 100 \cdot \frac{\ln[\sigma]}{\ln[\eta_0]},$$

where both $\ln[\sigma]$ and $\ln[\eta_0]$ are positive and the factor 100 is introduced to render the numbers easier to read. In pre-epidemic (ante Covid-19) condition $I_e = 0$ because we have
set $\sigma = 1$; we expect $\sigma > 1$ and $I_e > 0$ in acute states of epidemics. Such index, independent of $m$, also accounts for our capacity of curing the illness because it decreases with increasing $\beta$, the parameter introduced in (2.3), which enters into $\eta_0$ defined in (2.4). It should be observed that $\sigma$ could be less than one when the mortality in the period under analysis is milder than in the corresponding period in pre-epidemic conditions. In this particular case $I_e$ may attain negative values, which should not be surprising since this is an index of comparison between two configurations, and does not have an absolute significance from the way it is calculated.

(b) Application to the Covid-19 epidemic in Italy

In Italy, positivity to the Covid-19 of two Chinese tourists was recorded on January $30^{th}$, while contagion officially started in February $21^{st}$-$22^{nd}$. According to the Italian Ministry of Health, deaths increased from 29 on February $29^{th}$, to 12428 at the end of March. Using published data, we start by analyzing 16 Italian regions, considered as homogenous boxes, geographically located as in Figure 2(a). Northern Italy is reputed the most infected zone; in southern Italy the infection is mild; central Italy is in-between. The remaining regions (Basilicata, Molise, Calabria, Sardegna) are not analyzed because their mortality from Covid-19 is almost null, at least so far. The period of observation $\Delta t$ is an interval of time that will be chosen within a two-month period of March and April: our results will in fact be obtained by processing the data of the deaths occurred in a time interval, variable according to the type of analysis, included in the period March-April 2020 and in an analogous interval in previous years.

We can then calculate the index of epidemic $I_e$ on a regional basis. Setting March as the period of observation, the renormalization $A_n = F(A_r)$ in pre-epidemic conditions is calibrated from the statistical analysis of the number of dead, sorted by age, in March of the years from 2015 to 2019, and doing the average. These data are published online by the Italian National Institute of Statistics (ISTAT) at https://www.istat.it/it/archivio for many Italian municipalities, which have been grouped by region of origin. The total population at the beginning of March is available online at http://demo.istat.it, together with the demographic structure for each solar year. Population and number of dead were categorized by age by considering nine sets for $A_r = \{0 - 10, 10 - 20, \ldots, 80 - 90, > 90\}$ solar years of age. Within the $i$–th set, the probability of death is $n_i^D$, i.e., the ratio between the number of dead and peer population. Then, the average probability of death in the period 2015-2019 is calculated for each $i$–th set. To pass from the histogram to points on a graph, we consider the age at the center of each interval (5, 15, $\ldots$, 85) and 100 for ages higher than 90 solar years. The distribution of (2.4) provides the theoretical scaling between the probabilities of death at any two real ages $A_r,1$ and $A_r,2$, to be compared with measured data. In all the considered regions, the assumed form of renormalization of (2.1) proved accurate. Equating the measured and expected probabilities at one point $A_r,2$, chosen to be $A_r,2 = 45$ because, here, $A_r,2 = A_n,2$, we calibrate $\gamma_1$ and $\gamma_2$ of (2.1). The results of the
Figure 1: Region Lombardia. Graphical regressions in the epidemic Weibull plane $\ln\ln\{1/[1 - P^D(A_r, \sigma)]\} - \ln[F(A_r)]$ of the probability of death $P^D(A_r, \sigma)$ as a function of the renormalized age $A_n = F(A_r)$. Comparison of the measured points (dots) and their linear interpolation in the epidemic Weibull plane according to the theory. (a) Conditions ante Covid-19 ($\sigma = 1$) in March (average of years 2015-2019); (b) Epidemic period of March 2020. The numbers above the dots indicate the solar years of age they refer to. Points corresponding to 5, 15, 25, 35 solar years of age are strongly sensitive to small variations in the number of deaths, which may derive from non-epidemic causes not considered by the theory.

Passing to March 2020, at https://www.istat.it/it/archivio one finds the number of deaths sorted by age. Since only population, but not the demographic structure, is recorded at http://demo.istat.it, we assume the this is equal to that for the year 2019. The theoretical model predicts that the parameters $\gamma_1$ and $\gamma_2$ do not vary, since all the effects of the epidemic are gathered in the parameter $\sigma$ in (2.4). The fitting with the experimental points confirms this finding, with very good approximation.

In order to calculate the index $I_e$ we use the graphical construction in the epidemic Weibull plane. Figures 1(a) and 1(b) show the measured points and the best-fit line in Lombardia in March 2015-2019 (pre-epidemic) and March 2020 (epidemic), respectively. The comparison demonstrates that the model excellently predicts the deaths in the middle and old ages. The approximation of the left-hand-side tail can be attributed to the fact that the number of dead is quite limited for the younger, so that even a small variation can provide a noteworthy deviation from the linear trend. It should be recalled that the deaths at a young age often occur from non-epidemic causes, such as road accidents, drug, homicides, which cannot be covered by our theory.

For the 16 considered regions, the values of the measured probability of death $P^D(45, 1)$, the calibrated coefficients $\gamma_1$ and $\gamma_2$, and the calculated quantities $m\ln[\eta_0]$ and $m\ln[\sigma]$ are recorded in Table 3 of Section 3(b). The index of epidemic has been calculated from (2.6) and the results for the month of March are shown in the histogram of Figure 2(b). Unfortunately, the index for the month of April could not be calculated with this procedure that relies upon raw mortality data, because, at the time of this writing, such data are not
available for periods after April 15th 2020.

It may be of interest to make a further analysis by considering that the Italian authorities have also released the number of deaths attributed to Covid-19 only. It is difficult, in our opinion, to unambiguously sort out such data: for example, the press reported that many elderly people died in retirement homes without having been tested for the virus. In any case, we tentatively assume that the total number of deaths in the period of observation in epidemic condition is the sum of the number of dead in pre-epidemic condition and those officially attributed to Covid-19 by the Italian institutions. However, such data are reported for the whole region and not for each municipality. In pre-epidemic condition, the number of deaths in each region are sorted by age on a yearly basis (see https://www.istat.it/it/archivio). Hence, the number of deaths for each $i$–th age set is now obtained by multiplying the percentage recorded in the mortality tables referring to the whole solar year by the total number of dead in March, both available at http://demo.istat.it. Moreover, since the mortality tables for the 2019 are not yet available, reference is made to the average of the years 2015-2018. Passing to March 2020, the deaths from Covid-19 are sorted out by age using the percentages attributed to each age class for the whole nation, published by the Italian National Institute of Health (ISS) on May 4th 2020 (https://www.epicentro.iss.it/coronavirus), as shown in Figure 7(a) of Section 3(b). The same procedure can be repeated also for the month of April, for which similar data are available. The resulting $I_e$ for the 16 considered regions is again represented in Figure 2(b).

The actual data that shall be used for the model calibration are those corresponding to March 2020 and labelled in Figure 2(b) as “raw mortality”. The most infected region is Lombardia followed, at a distance, by Emilia Romagna. The epidemic is milder in the central Italy, except Marche, where $I_e$ is much higher than in Toscana and Umbria. South of Lazio the infection is very limited. This is probably because here the epidemic developed later than in the North, so that the lockdown countermeasures, simultaneously imposed on the whole national territory, were more effective. Remarkably in Lazio, Umbria, Campania, Sicilia the Index even reaches a negative value: this means that the severity of epidemic, quantified by the parameter $\sigma$, is even less than in previous years. Indeed, we anticipate that in such regions the number of dead in March 2020 has been less than in previous years.

It should be observed that the values reached by $I_e$ by considering raw mortality data are, apart from Valle d’Aosta, much higher than those calculated by summing up the number of dead officially attributed to Covid-19: for example, in Piemonte it is almost double. According to this theory, this may be an indication that the number of deaths from Covid-19 could have been significantly underestimated. We may mention that the mayors of many municipalities in Italy have repeatedly complained in newspapers that the official death toll for Covid-19 was much lower than the difference between the death toll in March 2020 and in past years. On the other hand, there are regions where $I_e$ calculated from raw mortality data is negative, but it becomes positive when considering the number of dead officially attributed to Covid-19: obviously, if one sums up this number to the number of dead in previous years, the mortality cannot but increase.
Figure 2: (a) Geographic location of the various Italian regions under investigation; (b) values of the index of epidemic $I_e$ in March 2020 and April 2020, under the acute phase of Covid-19. Comparison between the values of the index obtained by analyzing the raw mortality data provided by ISTAT for March 2020 and by manipulating the official data of deaths attributed to Covid-19 by the Italian institution (official Covid-19), available also for the month of April.
Nevertheless, waiting for further data, one may estimate the evolution of the epidemic by comparing the values of $I_e$, calculated on the basis of the official number of Covid-19 deaths, in March and April 2020. The index for Lombardia remains almost unchanged in April, while it sensibly grows in Piemonte, Trentino Alto Adige and Valle d’Aosta, where it becomes higher than in Emilia Romagna. In general, $I_e$ is the highest in Northern Italy, with Veneto and Friuli Venezia Giulia the less infected. In Lazio, Umbria, Campania, Sicilia the Index slightly grows, achieving positive values.

Are there other ways to estimate the strength of the epidemic? To the best of our knowledge, it is customary to consider that a representative indicator is represented by the ratio between the number of dead under epidemic and in previous years. Table 1 reports the values of $I_e$ in March 2020, calculated from raw mortality data, and the values of $R = N_{d,20}/N_{d,15-19}$, where $N_{d,20}$ and $N_{d,15-19}$ are the number of dead in March 2020 and the corresponding average in the years 2015-2019, respectively. Of course $I_e$ and $R$ cannot be compared one another within the same region, because they are associated with different in type quantities; however, they can be compared in relation to different regions. Observe, in any case, that $R < 1$ means that the number of deaths in March 2020 has been less than in the same month of previous years; correspondingly, the index $I_e$ becomes negative.

The indicator $R$ can certainly provide significant information about the effect of an epidemic, but it is not accurate, in our opinion. In fact, the distribution of dead by age and the demographic structure of the population play a role of paramount importance in the evaluation of the diffusion of the epidemic in a certain territory. Compare, for example the results for Trentino Alto Adige and Liguria: the difference in $R$ is mild but the index $I_e$ in Trentino is almost double than in Liguria. This is principally due to the different demographic structure between the two regions: Trentino Alto Adige is a relatively “young” region, where about 73% of people are younger than 60, while this percentage drops to 65% Liguria. The present theory makes a distinction between the deaths among the elderly and the young people, associated with the greater attitude of the former ones to develop potentially lethal pathologies under epidemic. In other words, a variation in the number of deaths in the elderly is recognized as a more tangible sign of the effects of an epidemic than an equal variation in the young people. Our theory somehow recognizes that the increase of deaths in Trentino have occurred in a restricted part of the population, i.e., the oldest part, while this datum is smeared on the whole population if one considers the indicator $R$. The same trend can be appreciated also for other regions while analyzing the numbers in Table 1.

It should be mentioned, however, that the epidemic focused on restricted outbreaks within each region. Hence, we now calculate $I_e$ in smaller representative territories, included in the 9 Italian provinces shown in Figure 3(a). The calculations are made on a weekly basis from February 23th until April 11st, using the same raw mortality data released by ISTAT and available at https://www.istat.it/it/archivio and http://demo.istat.it for the years 2015-2020. The procedure is the same used for the regions. The extension of the territories investigated and how data have been processed are presented in detailed
Table 1: Values of the index of epidemic \(I_e\) in March 2020 and comparison with the indicator \(R = N_{d,20}/N_{d,15-19}\), where \(N_{d,20}\) and \(N_{d,15-19}\) are, respectively, the number of dead in March 2020 and the mean number of dead in the same month of the years 2015-2019. Raw mortality data released by ISTAT.

<table>
<thead>
<tr>
<th>Region</th>
<th>Abruzzo</th>
<th>Campania</th>
<th>Emilia R.</th>
<th>Friuli V.G.</th>
<th>Lazio</th>
<th>Liguria</th>
<th>Lombardia</th>
<th>Marche</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_e)</td>
<td>0.495</td>
<td>-0.545</td>
<td>3.452</td>
<td>0.377</td>
<td>-1.093</td>
<td>1.821</td>
<td>6.515</td>
<td>2.662</td>
</tr>
<tr>
<td>(R)</td>
<td>1.087</td>
<td>0.981</td>
<td>1.677</td>
<td>1.098</td>
<td>0.919</td>
<td>1.502</td>
<td>2.880</td>
<td>1.511</td>
</tr>
<tr>
<td>Piemonte</td>
<td>2.177</td>
<td>0.282</td>
<td>-0.651</td>
<td>0.696</td>
<td>2.844</td>
<td>-0.037</td>
<td>1.865</td>
<td>0.949</td>
</tr>
<tr>
<td>(I_e)</td>
<td>1.470</td>
<td>1.086</td>
<td>0.973</td>
<td>1.138</td>
<td>1.643</td>
<td>1.069</td>
<td>1.585</td>
<td>1.245</td>
</tr>
<tr>
<td>(R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Section 3(b). Values of \(I_e\) are plotted against time in Figure 3(b).

Although first official contagions were recorded on February 21\(^{st}\) – 22\(^{nd}\) and taking into account that the incubation time (about two weeks) and the course of the disease, the index in Lodi, the first infected province, was already high (\(I_e \simeq 3.8\)) in the week 23-29/02. It increased one week after (\(I_e \simeq 6.79\)), also affecting the neighboring Cremona, Piacenza and Brescia. However, it grew even more in Bergamo (\(I_e \simeq 4.62\)), although Bergamo does not border Lodi. Even in Pesaro-Urbino, which is 360 km far away from Lodi, \(I_e\) reached quickly very high values. Presumably, independent outbreaks developed here and in Bergamo.

The Italian government imposed the national lockdown on 9\(^{th}\) March, but, on February 24\(^{th}\), ten municipalities in the province of Lodi and one in Padova had already been preventively quarantined. This justifies why \(I_e\) decreased first in Lodi, after the peak in the third week of March (\(I_e \simeq 11.44\)), but it continued to increase in Bergamo, Brescia, Cremona. In Bergamo \(I_e\) exceeds the Lodi peak, achieving the dramatic value \(I_e \simeq 13.6\). Padova in Veneto was officially considered one of the first heavily-infected provinces, but this is not confirmed here, probably because the localized outbreak was promptly quarantined. Indeed Veneto is one of the less infected regions in the North, probably because of very effective countermeasure organized by the local authorities. Milano city has been officially considered highly critical, but the virus developed later than in the other provinces of Lombardia, and this is why national lockdown prevented the index from reaching very high values.

3 Methods

(a) The statistical theory

The starting point of the theory is the renormalization of the real age \(A_r\) to the nominal age \(A_n\) through \(A_n = F(A_r)\), to be considered valid on average within a territory reputed
homogeneous. A particular type of renormalization law is that in (2.1).

The severity of the pathology that an individual can develop is measured by one variable $\delta$. The model could be readily generalized by introducing additional parameters and following the same rationale. The analysis is referred to a given time interval of observation $\Delta t$, representing the time in which mortality is monitored. The statistics to be used refer to this observation time.

We then introduce a statistical law à la Pareto, of the type

$$p_{\Delta A_n}(\delta) = C_{\Delta t} \delta^{-\alpha}, \text{ with } \alpha > 1. \tag{3.1}$$

This represents the probability density function for the development of the pathology $\delta$ in the observation time $\Delta t$, associated with the renormalization length-scale for human aging represented by the single life-segment $\Delta A_n$. This law should universally hold for every person, irrespective of race, genetics, quality of environment and life, gender, because any distinction is accounted for through the renormalization $A_n = F(A_r)$.

Since the power law function diverges for $\delta \to 0$, it is customary to prescribe a normalization criterion. Hence, we assume that a physiological pathology level $\delta_{\text{min, } \Delta t}$, observable in $\Delta t$, is always present in each nominal life-segment. It represents the physiological minimal degradation due to aging of the human body: a person of nominal life $A_n$ will develop at least a pathology equal to $A_n/\Delta A_n \cdot \delta_{\text{min, } \Delta t}$, otherwise it would be theoretically possible
to live for ever. We can then write

\[
\int_{\delta_{\min,\Delta t}}^{\infty} p_{\Delta A_n}(\delta) d\delta = \int_{\delta_{\min,\Delta t}}^{\infty} C_{\Delta t} \delta^{-\alpha} d\delta = 1
\]

\[\Rightarrow C_{\Delta t} = \frac{\alpha - 1}{(\delta_{\min,\Delta t})^{-\alpha+1}},\]  

(3.2)

so that, for \( \delta \geq \delta_{\min,\Delta t} \), (3.1) becomes

\[
p_{\Delta A_n}(\delta) = \frac{\alpha - 1}{\delta_{\min,\Delta t}} \left( \frac{\delta}{\delta_{\min,\Delta t}} \right)^{-\alpha}.
\]

(3.3)

The probability of developing a pathology more severe or equal to \( \delta \geq \delta_{\min,\Delta t} \) in \( \Delta A_n \) reads

\[
P_{\Delta A_n}^{\geq}(\delta) = \int_{\delta}^{\infty} \int_{\delta_{\min,\Delta t}}^{\infty} p_{\Delta A_n}(\delta) d\delta = \left( \frac{\delta}{\delta_{\min,\Delta t}} \right)^{1-\alpha},
\]

(3.4)

which can be rearranged in the form

\[
P_{\Delta A_n}^{\geq}(\delta) = \left( \frac{\delta}{\delta_{\min,\Delta t}} \right)^{1-\alpha} = \Delta A_n \left( \frac{\delta}{\eta} \right)^{1-\alpha},
\]

(3.5)

where

\[
\eta = \frac{\delta_{\min,\Delta t}}{(\Delta A_n)^{1/(\alpha-1)}}.
\]

(3.6)

Therefore the probability of finding a pathology less than \( \delta \) is \( P_{\Delta A_n}^{\leq}(\delta) = 1 - P_{\Delta A_n}^{\geq}(\delta) \).

We treat the “development of a pathology in one nominal life-segment” as an event independent of what happens in other segments. In the whole nominal life \( A_n \), composed of \( A_n/\Delta A_n \) nominal life-segments, a pathology \( \geq \delta \) does not develop if this is true in all the segments. Hence, the probability \( P_{\Delta A_n}^{\leq}(\delta) \) of finding a pathology less than \( \delta \) in \( A_n \) coincides with the product of the probabilities of having the same property in all the \( \Delta A_n \), i.e.,

\[
P_{\Delta A_n}^{\leq}(\delta) = \left[ 1 - \Delta A_n \left( \frac{\delta}{\eta} \right)^{1-\alpha} \right]^{A_n/\Delta A_n}.
\]

(3.7)

This expression can be simplified under the hypothesis that \( A_n/\Delta A_n \ll 1 \). Taking the limit as \( A_n/\Delta A_n \to 0 \) and observing that \( \lim_{\epsilon \to 0}[1 + a \epsilon]^{1/\epsilon} = \exp[a] \), one obtains

\[
P_{\Delta A_n}^{\leq}(\delta) = \exp\left[ -A_n \left( \frac{\delta}{\eta} \right)^{1-\alpha} \right].
\]

(3.8)

Therefore, the probability \( P_{\Delta A_n}^{\geq}(\delta) = 1 - P_{\Delta A_n}^{\leq}(\delta) \) of finding in \( A_n \) a pathology \( \geq \delta \) results to be that of (2.2).
The epidemic is supposed homogenous in the considered region and is measured by one parameter $\sigma$. This again is a simplification, but a more detailed characterization can be obtained by considering $\sigma$ as a function of space (different values in diverse locations).

The event “death” is determined according to a death criterion represented by a function $G(\cdot, \cdot, \cdot)$: the event occurs when

$$G(\delta, \sigma, A_n) - K_{Ie} = 0.$$  \hfill (3.9)

Likewise, the “safe” domain is defined by $G(\delta, \sigma, A_n) - K_{Ie} < 0$. This is a general criterion, where a dependence upon $A_n$ takes into account that the epidemic can be, e.g., more aggressive on elderly than on young people. Given $A_n$, the critical level of pathology $\delta$, beyond which death occurs, is associated with the level of epidemic $\sigma$ through a function $\delta = g(\sigma, A_n)$, implicitly defined by (3.9), such that

$$G(g(\sigma, A_n), \sigma, A_n) - K_{Ie} = 0.$$  \hfill (3.10)

It is reasonable to require that $g(\cdot, A_n)$ is monotone decreasing because, for any $A_n$, the higher the level of epidemic, the lower the critical pathology. Moreover, one expects that when $\sigma \to 0$ it takes a pathology $\delta \to \infty$ to cause death, hence the further condition $\lim_{\sigma \to 0} g(\sigma, A_n) = +\infty$.

Because of these properties, the probability of death $P_D(A_n, \sigma)$ in the epidemic $\sigma$ at the nominal age $A_n$ coincides with probability of developing a pathology $\geq \delta = g(\sigma, A_n)$. Substituting in (2.2), one obtains

$$P_D(A_n, \sigma) = P_{A_n}^{\geq}(g(\sigma, A_n))$$

$$= 1 - \exp \left[ -A_n \left( \frac{\eta}{g(\sigma, A_n)} \right)^{\alpha-1} \right].$$ \hfill (3.11)

In the simplest case, dropping the dependence of $\delta = g(\sigma, A_n)$ upon $A_n$, one can consider the power law of (2.3), which satisfies the required monotonicity and asymptotic properties. Here, the parameter $\beta$ represents our capability of curing the disease caused by the epidemic.

Therefore, (3.11) simplifies into

$$P_D(A_n, \sigma) = 1 - \exp \left[ -A_n \left( \frac{\sigma}{\eta_0} \right)^m \right],$$ \hfill (3.12)

which is a two-parameter Weibull distribution with shape parameter $m$ and scale parameter $\eta_0$, independent of $\sigma$, given by

$$m = (\alpha - 1)\beta, \quad \eta_0 = K_{Ie}/(Y^{1/\beta}).$$ \hfill (3.13)

In terms of real life $A_r$, the renormalization $A_n = F(A_r)$ provides the expression of (2.4).
There are affinities with the statistical micro-mechanically-motivated models describing the population of macroscopic strengths of brittle materials. As pursued in [24], the starting point is a statistical characterization à la Pareto of the size $\delta$ of microcracks in a specimen [26]. Linear Elastic Fracture Mechanics (LEFM) [16] prescribes that cracks propagate when $Y \sigma \delta^{1/2} - K_{Ic} = 0$, where the first term is the Stress Intensity Factor (SIF), $\sigma$ is the nominal applied stress, $Y = 2.24/\sqrt{\pi}$ for semicircular thumbnail cracks of radius $\delta$ and $K_{Ic}$ the critical SIF, characteristic of the material. This is why the expression of (2.3) is a “LEFM-like” death criterion, even if the coefficient $\beta$ differs from $\beta = 2$. In this analogy, the level of the epidemic $\sigma$ represents the applied stress and the nominal life $A_n$ the size of the material specimen [17].

(b) Analysis of mortality data in Italy

The theory enables the quantitative analysis of the Covid-19 epidemic experienced by in Italy in March 2020. We analyze 16 regions, considered boxes with homogeneous mortality, distributions of the population and of deaths by age. This assumption provides a smeared view, within the same region, of the differences in genetic heritage, quality of environment and life, efficiency of health system. No differentiations are made between men and women, although a diverse mortality by gender has been recorded. We start by evaluating the pre-epidemic condition, chosen to be March in the period 2015-2019. The ISTAT has published the mortality data for many Italian municipalities, available online at https://www.istat.it/it/archivio, sorted out by age on a daily basis. The considered municipalities cover most of the population in each region, as indicated in Table 2.

Table 2: Percentage of the total regional population covered by the mortality data in the period 2015-2020 made available by ISTAT at https://www.istat.it/it/archivio.

<table>
<thead>
<tr>
<th>Abruzzo</th>
<th>Campania</th>
<th>Emilia R.</th>
<th>Friuli V.G.</th>
<th>Lazio</th>
<th>Liguria</th>
<th>Lombardia</th>
<th>Marche</th>
</tr>
</thead>
<tbody>
<tr>
<td>84.48%</td>
<td>78.95%</td>
<td>93.37%</td>
<td>73.16%</td>
<td>80.38%</td>
<td>92.02%</td>
<td>72.42%</td>
<td>80.08%</td>
</tr>
<tr>
<td>Piemonte</td>
<td>Puglia</td>
<td>Sicilia</td>
<td>Toscana</td>
<td>Trentino A.A.</td>
<td>Umbria</td>
<td>Valle d'Aosta</td>
<td>Veneto</td>
</tr>
<tr>
<td>93.77%</td>
<td>85.94%</td>
<td>72.75%</td>
<td>88.91%</td>
<td>92.87%</td>
<td>93.57%</td>
<td>92.73%</td>
<td>85.99%</td>
</tr>
</tbody>
</table>

For any two solar ages $A_{r,1}$ and $A_{r,2}$, (2.4) provides the scaling law

$$\frac{F(A_{r,1})}{F(A_{r,2})} = \frac{\ln[1 - P^D(A_{r,1}, \sigma)]}{\ln[1 - P^D(A_{r,2}, \sigma)]}. \quad (3.14)$$

We start by calibrating the renormalization law $A_n = F(A_r)$ in (2.1) by fitting the experimental points in all 16 considered regions in pre-epidemic condition. Since $A_n = A_r$ for $A_r \leq 45$ solar years, one can fix a point in this range (we fixed $A_{r,2} = 45$ solar years) and calculate the coefficients $\gamma_1$ and $\gamma_2$ by best fitting of the rescaling law of (3.14) with
experimental points. The results for the 16 regions are summarized in Table 3. Notice that, among the various regions, the difference between the values reached by the parameters $\gamma_1$ and $\gamma_2$ is small. The agreement is in general excellent, as demonstrated by the graphs for Lombardia reported in Figure 4. In general, the step changes at 45 and 70 solar years of age are confirmed in all the considered regions. We note that the probabilities of death associated with ages lower than 40 years are not optimally fitted by the theoretical curve, but this is due to the fact that they are strongly sensitive to small variations in the number of dead and that deaths at young ages often occur from non-epidemic causes (road accidents, drug, homicides). In general, the probability of death in the range 0-10 years is always higher than the theoretical prediction, but this is certainly influenced by the number of deaths at birth, which represents a special category. In general, our model does not account for deaths due to factors that are not associated with the natural degradation of the human body and the consequent potential ability to develop a pathology.

Figure 4: Region Lombardia in the pre-epidemic ($\sigma = 1$) condition of March (average of the years 2015-2019). Points represent the observed probability of death $P^D(A_r, \sigma)$ in the month, as a function of real age $A_r$ measured in number of solar years, while the curve indicates the theoretical prediction. (a) Left-hand-side portion for $0 \leq A_r \leq 65$ solar years; (b) right-hand-side portion for $A_r \geq 65$ solar years. The same correlation has been constructed for all the other considered regions, provinces and for Milano city, by using the demographic statistics published by ISTAT (http://demo.istat.it/).

We here assume that the renormalization function (2.1) is not affected by the level of epidemic $\sigma$, i.e., the function $G$ in equation (3.9) is not dependent upon $A_n$, and hence the parameters $\gamma_1$ and $\gamma_2$ are the same in pre-epidemic and epidemic conditions. Observe that (2.5) represents a straight line $y = x + c$ in the epidemic Weibull plane $y = \ln \ln \{1 - P^D(A_r, \sigma)\}^{-1}$ vs. $x = \ln F(A_r)$, with

$$c = m \ln[\sigma] - m \ln[\eta_0].$$

(3.15)

Setting $\sigma = 1$ in pre-epidemic condition, equation (3.15) provides $c = -m \ln[\eta_0]$. From the plot in the bi-logarithmic plane one finds the line $y = x + c$ that best fits with the measured points, and graphically derive $c = -m \ln[\eta_0]$ as the intercept on the $y$ axis. The results for Lombardia, plotted in Figure 1(a), show a very good agreement with the experimental points and provide $c|_{\sigma=1} = -m \ln[\eta_0] = -13.936$. In epidemic conditions (March 2020), with the
aforementioned graphical construction one finds again \( c \) of (3.15). The results for Lombardia are shown in Figure 1, from which one finds \( c = m \ln[\sigma] - m \ln[\eta_0] = -13.028 \). But recalling that \( m, \eta_0 \) and consequently \( m \ln[\eta_0] \) do not depend upon \( \sigma \), one can use the value calculated in pre-epidemic condition and hence determine \( m \ln[\sigma] = -13.028 + 13.936 = 0.908 \). The optimal fit of the model parameters to the experimental points shown in Figure 1(b) proves that the renormalization found for pre-epidemic condition, i.e., the parameters \( \gamma_1 \) and \( \gamma_2 \), is still valid for epidemic condition. The parameters that outcome from the graphical procedure are summarized in Table 3.

Table 3: Pre-epidemic condition (average of March 2015-2019), with \( \sigma = 1 \), and epidemic condition (March 2020), found by elaborating raw mortality data released by ISTAT in 16 Italian regions. Measured probability of death \( P^D(A_{r,2}, \sigma) \) at \( A_{r,2} = 45 \) solar years of age and \( \sigma = 1 \). Parameters \( \gamma_1 \) and \( \gamma_2 \) of the renormalization \( A_n = F(A_r) \), as per equation (2.1). Values for \( m \ln[\eta_0] \) and \( m \ln[\sigma] \) derived from interpolation of the measured points with the linear trend in the epidemic Weibull plane predicted by the theory.

<table>
<thead>
<tr>
<th></th>
<th>Abruzzo</th>
<th>Campania</th>
<th>Emilia R.</th>
<th>Friuli V.G.</th>
<th>Lazio</th>
<th>Liguria</th>
<th>Lombardia</th>
<th>Marche</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P^D(45, 1) )</td>
<td>( 6.1 \times 10^{-5} )</td>
<td>( 7.0 \times 10^{-5} )</td>
<td>( 5.9 \times 10^{-5} )</td>
<td>( 5.9 \times 10^{-5} )</td>
<td>( 5.4 \times 10^{-5} )</td>
<td>( 4.0 \times 10^{-5} )</td>
<td>( 5.4 \times 10^{-5} )</td>
<td></td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.89</td>
<td>1.95</td>
<td>1.92</td>
<td>1.87</td>
<td>1.95</td>
<td>1.98</td>
<td>2.02</td>
<td>1.90</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>2.95</td>
<td>3.00</td>
<td>3.00</td>
<td>2.95</td>
<td>3.02</td>
<td>3.02</td>
<td>3.10</td>
<td>3.00</td>
</tr>
<tr>
<td>( m \ln[\sigma] )</td>
<td>0.067</td>
<td>-0.073</td>
<td>0.472</td>
<td>0.051</td>
<td>-0.149</td>
<td>0.250</td>
<td>0.908</td>
<td>0.364</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Piemonte</th>
<th>Puglia</th>
<th>Sicilia</th>
<th>Toscana</th>
<th>Trentino A.A.</th>
<th>Umbria</th>
<th>Valle d’Aosta</th>
<th>Veneto</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P^D(45, 1) )</td>
<td>( 4.5 \times 10^{-5} )</td>
<td>( 6.4 \times 10^{-5} )</td>
<td>( 7.0 \times 10^{-5} )</td>
<td>( 4.6 \times 10^{-5} )</td>
<td>( 4.7 \times 10^{-5} )</td>
<td>( 5.3 \times 10^{-5} )</td>
<td>( 4.1 \times 10^{-5} )</td>
<td>( 5.1 \times 10^{-5} )</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>2.03</td>
<td>1.85</td>
<td>1.90</td>
<td>1.95</td>
<td>1.94</td>
<td>1.90</td>
<td>2.16</td>
<td>1.94</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>3.07</td>
<td>2.98</td>
<td>3.01</td>
<td>3.05</td>
<td>3.03</td>
<td>3.00</td>
<td>3.10</td>
<td>3.02</td>
</tr>
<tr>
<td>( m \ln[\sigma] )</td>
<td>0.301</td>
<td>0.038</td>
<td>-0.087</td>
<td>0.096</td>
<td>0.392</td>
<td>-0.005</td>
<td>0.261</td>
<td>0.130</td>
</tr>
</tbody>
</table>

The “index of epidemic”, defined in (2.6), reads \( I_e = 100 \cdot \ln[\sigma]/\ln[\eta_0] \) and it is therefore a measurable quantity. For the case of Lombardia, one obtains \( I_e = 100 \cdot 0.908/13.936 = 6.515 \). The procedure to determine this index is summarized in the flowchart of Figure 5. The values of the index for the regions under investigation are shown in Figure 2(b) and recorded in Table 1.

The values reached by \( I_e \) on basis of the raw number of dead are then compared with the ones that can be calculated by assuming the number of dead in the epidemic period as the sum of the number of dead in pre-epidemic conditions and the one officially attributed to Covid-19 by Italian institutions. In this case the total population for each region must be considered, since such data are associated with the whole region and not with the single municipalities. However, the data about mortality made available by ISTAT at https://www.istat.it/it/archivio do not cover all the municipalities in the region, as indicated in Table 2. Hence, in order to define the pre-epidemic conditions, the number of deaths for each \( i \)-th age set is obtained by multiplying the percentage given by the mortality tables referring to the whole solar year by the total number of dead in March, both available at http://demo.istat.it. Here, pre-epidemic conditions refer to the average in the
Figure 5: Schematic flowchart of the procedure used to calculate the index of epidemic. Mortality data in epidemic (period in 2020) and pre-epidemic (comparable period in previous years) conditions, sorted by age, are first elaborated to determine the renormalization \( A_n = F(A_r) \), where \( A_n \) and \( A_r \) are the nominal and real ages, respectively. The best-fit interpolation in the plane \( \ln \ln \left\{ \frac{1}{1-P^D(A_r, \sigma)} \right\} \) vs. \( \ln[F(A_r)] \) of the measured points with the line predicted by the theory provides the values \( m \ln[\sigma] \) and \( m \ln[\eta_0] \) (pre-epidemic with \( \sigma = 1 \)), whose ratio, multiplied by 100, represents the index of epidemic \( I_e \).

For epidemic conditions, what is needed is the deaths from Covid-19 sorted out by age for each region, but deaths attributed to Covid-19 furnished by the Italian National Institute of Health (ISS) are sorted by age only at the national scale. Hence, we are forced to assume that the total number of deaths is the sum of the mean value of the dead in March 2015-2018 and those attributed to Covid-19, sorted out by age using the percentages published for the whole nation, available online on https://www.epicentro.iss.it/ and shown in Figure 7(a). The total number of dead in March and April 2020, sorted out by region, are shown in Figure 7(b).

The resulting \( I_e \) in the 16 regions are represented in Figure 2(b) and compared with those obtained by analyzing the raw mortality data, when available. The parameters obtained in
the calibration process are summarized in Table 4. Observe that the parameters $\gamma_1$ and $\gamma_2$ and the product $m \ln[\eta_0]$ only slightly vary region by region. A moderate variation of such parameters is observed for each region by passing from March to April since mortality is not uniform during the solar year. The values of the index $I_e$ for March 2020 obtained from raw mortality data are much higher than those resulting from the analysis of the deaths officially attributed to Covid-19 from national institutions, which are probably underestimated. By comparing Figures 6(b) and 7(a), another inconsistency in the data published by ISS can be noticed in the very low percentage of deaths for people older than 90: probably their positivity to Covid-19 was not tested considering their age.

Passing to analyze the provinces indicated in Figure 3(a), their population has been obtained by grouping the municipalities according to their location, considering the mortality data week by week. The samples cover more than 70% of the provincial population, as indicated in Table 5. The city of Milano, densely populated, has been considered independently from the other municipalities in its province. For each considered week, the probability of death sorted by age in pre-epidemic condition has been calculated with reference to the average of the deaths occurred in the corresponding weeks in previous years. The number of deaths per day, sorted out by age, are available at www.istat.it/it/archivio, while the total population at the beginning of each month at www.demo.istat.it, together with its categorization by age on a yearly basis. Since ISTAT provides the demographic structures of the provinces only until the year 2019, we have assumed that, in the epidemic period, population by age is the same as in the 2019, and the total population has been estimated according to the demographic trend.

Following the same procedure used for the regions, one finds the results summarized in Table 6 in pre-epidemic and epidemic conditions, respectively. Consistently with the results of the regions, the coefficients $\gamma_1$, $\gamma_2$, and $m \ln[\eta_0]$ only slightly varies during the weeks and among the provinces. With respect to the region, there is a difference in the data analysis in the fact that, in pre-epidemic condition, we have assumed that the probability of death $P_D(A_{r,2}, \sigma = 1)$ at $A_{r,2} = 45$ solar years of age is the same for all the provinces and equal.
to probability of death in the whole national territory corresponding to the age range 40-50 years in the third week of March in the period 2015-2019. The reason for this is that the probabilities of death in the classes corresponding to the younger ages are strongly affected by small deviations in mortality, especially if the number of inhabitants is small, as in a province with respect to the regions. If we had considered as $P^D(A_{r,2}, \sigma = 1)$ the observed value in the province, there would have been a significant uncertainty in the evaluations of the index, especially for the less populated territories. The national datum appears more reliable for what concerns the deaths for natural diseases, since it is less affected by accidental deaths which, on the other hand, may determine a great variation in a small territory.

The values of $m \ln[\eta_0]$ and $m \ln[\sigma]$, reported in the same tables, have been graphically
Table 4: Pre-epidemic condition (March and April average in 2015-2018 and σ = 1) and epidemic condition (March and April 2020) found by summing up the average number of deaths in 2015-2018 with the number attributed to Covid-19 by the Italian institutions in the considered regions. Measured probability of death $P^D(A_{r,2}, \sigma)$ at $A_{r,2}$ = 45 solar years of age and $\sigma = 1$. Parameters $\gamma_1$ and $\gamma_2$ of the renormalization $A_n = F(A_r)$ as per equation (2.1). Values for $m \ln[\eta_0]$ and $m \ln[\sigma]$ derived from interpolation of the measured points with the linear trend predicted by the theory in the Weibull plane. Resulting index of epidemic $I_e$.

<table>
<thead>
<tr>
<th>Abruzzo</th>
<th>Campania</th>
<th>Emilia R.</th>
<th>Friuli V.G.</th>
<th>Lazio</th>
<th>Liguria</th>
<th>Lombardia</th>
<th>Marche</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^D(45, 1)$</td>
<td>5.3 $\cdot$ 10^{-5}</td>
<td>8.8 $\cdot$ 10^{-5}</td>
<td>6.9 $\cdot$ 10^{-5}</td>
<td>8.2 $\cdot$ 10^{-5}</td>
<td>6.7 $\cdot$ 10^{-5}</td>
<td>1.02 $\cdot$ 10^{-4}</td>
<td>5.7 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.81</td>
<td>1.90</td>
<td>1.95</td>
<td>1.91</td>
<td>1.95</td>
<td>1.96</td>
<td>1.86</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>2.73</td>
<td>2.82</td>
<td>2.78</td>
<td>2.87</td>
<td>2.87</td>
<td>2.67</td>
<td>2.91</td>
</tr>
<tr>
<td>$m \ln[\sigma]$</td>
<td>0.057</td>
<td>0.000</td>
<td>0.300</td>
<td>0.067</td>
<td>0.000</td>
<td>0.195</td>
<td>0.573</td>
</tr>
<tr>
<td>$I_e$</td>
<td>0.437</td>
<td>0.000</td>
<td>2.248</td>
<td>0.508</td>
<td>0.000</td>
<td>1.503</td>
<td>4.240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>March 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^D(45, 1)$</td>
</tr>
<tr>
<td>$\gamma_1$</td>
</tr>
<tr>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>$m \ln[\sigma]$</td>
</tr>
<tr>
<td>$I_e$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Piemonte</th>
<th>Puglia</th>
<th>Sicilia</th>
<th>Toscana</th>
<th>Trentino A.A.</th>
<th>Umbria</th>
<th>Valle d’Aosta</th>
<th>Veneto</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^D(45, 1)$</td>
<td>7.4 $\cdot$ 10^{-5}</td>
<td>7.6 $\cdot$ 10^{-5}</td>
<td>5.8 $\cdot$ 10^{-5}</td>
<td>6.7 $\cdot$ 10^{-5}</td>
<td>5.9 $\cdot$ 10^{-5}</td>
<td>8.9 $\cdot$ 10^{-5}</td>
<td>4.9 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.81</td>
<td>1.90</td>
<td>1.95</td>
<td>1.91</td>
<td>1.95</td>
<td>1.95</td>
<td>1.96</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>2.73</td>
<td>2.82</td>
<td>2.80</td>
<td>2.87</td>
<td>2.87</td>
<td>2.67</td>
<td>2.91</td>
</tr>
<tr>
<td>$m \ln[\sigma]$</td>
<td>0.150</td>
<td>0.023</td>
<td>0.395</td>
<td>0.134</td>
<td>0.034</td>
<td>0.359</td>
<td>0.600</td>
</tr>
<tr>
<td>$I_e$</td>
<td>1.134</td>
<td>0.174</td>
<td>2.924</td>
<td>1.001</td>
<td>0.252</td>
<td>2.737</td>
<td>4.391</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>April 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^D(45, 1)$</td>
</tr>
<tr>
<td>$\gamma_1$</td>
</tr>
<tr>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>$m \ln[\sigma]$</td>
</tr>
<tr>
<td>$I_e$</td>
</tr>
</tbody>
</table>

Table 5: Percentage of the total population of the provinces covered by the mortality data in the period 2015-2020 published by ISTAT.

<table>
<thead>
<tr>
<th>Bergamo</th>
<th>Brescia</th>
<th>Cremona</th>
<th>Lod</th>
<th>Padova</th>
<th>Parma</th>
<th>Pesaro U.</th>
<th>Piacenza</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.95%</td>
<td>76.46%</td>
<td>73.54%</td>
<td>89.44%</td>
<td>74.43%</td>
<td>95.80%</td>
<td>79.56%</td>
<td>81.55%</td>
</tr>
</tbody>
</table>

estimated and the resulting $I_e$ is plotted against the time in Figure 3(b).
A statistical theory of the strength of epidemics

We have proposed a multi-scale statistical theory addressed towards the definition of an index of epidemic to quantify its strength with respect to potential lethality. Basic assumptions of the theory are the renormalization of human age to determine the attitude of individuals to develop severe pathology and the introduction of variables representative of the levels of pathology and of epidemic, which are correlated through a death criterion that takes into account also our capacity to treat the disease. The model is calibrated with reference to the Covid-19 epidemic in Italy, through the processing and the consequent

Table 6: Pre-epidemic (average in March 2015-2019 and $\sigma = 1$) and epidemic condition (March 2020) in the Italian provinces under analysis and Milano city. Weekly variation from February 23rd to April 11th 2020, derived by elaborating data released by ISTAT. Assumed probability of death $P^D(A_{r-}, \sigma) = 6.67 \cdot 10^{-6}$ at $A_{r-} = 45$ solar years. Renormalization parameters $\gamma_1$ and $\gamma_2$ as per equation (2.1). Graphically-estimated values of $m \ln|\eta|$ and $m \ln|\sigma|$ as per equation (2.2). Index of epidemic $I_e$.
comparison of raw mortality data, irrespective of the cause of death, sorted by age in epidemic and pre-epidemic conditions. The theory is inspired by principles that are proper of the probabilistic mechanics of materials. It can well interpret the Italian statistical data of mortality, for various periods of observation and extension of the territory under analysis, thus allowing to follow the diffusion of the infection. The proposed index of epidemic, taking into account the demographic structure of the population, weights the number of deaths according to an age-based risk; therefore, it seems more reliable than other commonly used indicators which, considering only the excess mortality rates, are unable to make any kind of distinction in the deaths from accidental causes.

In order to evaluate more in detail the capacity of the health system, the proposed theory could also be extended and calibrated according to the number of people needing Intensive Care Units, instead of the number of dead. Moreover, the variables used in the theory could also be considered continuous functions of time and location, to model the kinetics of the epidemic and predict the number of dead and sick, as well as the expected effect of adopted, or adoptable, remedies, such as quarantine, lockdown, medical treatments and vaccines. To this respect our theory, through the definition of a reliable index of epidemic, may complement well-established mathematical models in epidemiology, among which the SIR model and its derivations are certainly the most used. A comprehensive multidisciplinary approach will be of paramount importance for the management of the epidemic, in particular to make decisions about when, where and to what extent to loosen restrictive countermeasures, especially the lockdown which, if prolonged beyond necessity, may cause irreparable damage to the economy of a country.

Data accessibility. Data used in this article are available on line at the following official websites of Italian national institutions: https://www.istat.it/it/archivio, http://demo.istat.it/, https://www.epicentro.iss.it/, http://www.protezionecivile.gov.it/home.

Authors’ contribution. GP and GRC contributed equally to the development of this work.

Competing interests. There are no competing interests to declare.

Funding. This research received no funding.

Acknowledgements. We would like to express our most grateful appreciation to Prof. Roberto Ballarini (University of Houston, Texas, USA), who kindly reviewed a preliminary draft of this article, suggesting significant changes to the original text. Prof. Ivo Iori (University of Parma, Italy) is gratefully acknowledged for helpful discussion.

References

A statistical theory of the strength of epidemics


