doi.org/10.26398/IJAS.0033-001

COVID-19 AND POSSIBLE BIAS IN STATISTICAL INFORMATION: PANDEMIC INDICATORS AND THE RISK OF INFODEMIC

Fabrizio Antolini[1]

Department of Business Communication, University of Teramo, Campus Aurelio Saliceti, Via Renato Balzarini, 1- Teramo, Italy

Samuele Cesarini

Research Fellow, University of Teramo, Campus Aurelio Saliceti Via Renato Balzarini, 1- Teramo, Italy

Abstract. *A widely discussed theme at this time in Italy concerns the statistical data used to evaluate the Covid-19 trend, given that some indicators have had considerable influence in determining policy choices. The statistical collection can produce an infodemic process that we must counteract and avoid. In some cases, information on the effects of Covid-19 is difficult to obtain, and data are not always scientifically collected. Pandemic data are often analysed without being standardised, or do not have an established definition. Some indicators are difficult to measure, while the use of a single recognised measure would help to better understand the effects of the pandemic. Moreover, the way in which statistical information on Covid-19 is disseminated also contributes to create a framing that can affect the use of the variables (or indicators) analysed. The correct use and interpretation of indicators becomes relevant when an increasing amount of information contains measurement errors due to non-structured data or interpretative framing. The aim of the article is to identify the potential bias that can be generated by reading pandemic data. The entire process of statistical collection, including its communication, should be monitored because good public choices should depend on the correct use of statistics.*

Keywords*: Data quality, epidemiological data, Covid-19, infodemic.*

1. INTRODUCTION

The pandemic crisis has changed the perception of the facts (Halpern et al., 2020), and with that, the quantitative information through which facts are described. This situation of general change (both objective and subjective) due to lifestyles creates a risk of incurring bias in the use, communication and interpretation of data (Antolini et al., 2006), and this could become a pandemic phenomenon itself. The

[1] Corresponding author: Fabrizio Antolini, email: fantolini@unite.it.

difficulty in interpreting the indicators chosen to describe the reality as it manifests itself is compounded by the difficulty of assessing the impact of Covid-19, which, first of all, requires a theoretical scheme of reference that inevitably influences the choice of the indicators themselves. The impact of Covid-19, in fact, has important economic influences in addition to its social and public health effects. In Italy, for example, a fall of 5.8% in household disposable incomes occurred in the second half of 2020, while savings increased by 5.3% (ISTAT, 2020). The increase in savings, which is due to the fall in consumption, can be interpreted differently. In fact, it can be configured as a proxy for the welfare system, which has not allowed such an incisive fall in household disposable income, or as a proxy for the existence of the unobserved economy - the underground and criminal economy estimated at 211 billion euros in Italy or equal to 11.9 per cent of Italy's GDP. In this latter case, the savings would be only partially connected to the trend in disposable income.

Therefore, evaluation of the effects of the pandemic requires, first of all, a theoretical framework of reference that can consider the different aspects (social, economic and psychological) that the pandemic impacts and attribute to each of them a 'specific weight'. Fortuitously, the enormous amount of data available is processed by following so-called heuristics to formulate the necessary evaluations. These are mental shortcuts that simplify the interpretation of reality and available data, with preference for a cognitive process, either by intuition or based on experience (Kahneman et al., 1982). In the pandemic context, which is inevitably characterised by a strong emotional component that pervades all sectors of society, evaluation requires homogeneous and good quality data in the dimensions of the analysis, and these data become the fundamental prerequisite (Franchet, 1991). Conversely, heuristics seem to work well in many cases, although good statistical background information (again, good in terms of accessibility and timeliness) remains a prerequisite.

The quality of statistical information is greatly influenced by the way data are collected (Eurostat, 2014), and this aspect influences the comparability of estimates of many phenomena, both social causes of death or immigration (Doody et al., 2001; Kelly, 1982) and economic (Diewert, 1995). The mode of data collection can determine an overestimation/sub-estimation of large amounts of data subsequently used to construct the indicators.

By contrast, the way in which the phenomenon is statistically represented depends on the data selection. All the above, together with the way statistical data are communicated, can produce informative bias:

> 'The errors arise from: sampling, frame coverage, measurement, non-response, data processing, and model assumptions' (Eurostat, ibidem p.3).

This is true both for the measures adopted and for the meaning that our "mental accounting" attributes to the data.

The aim of the paper is to represent a critical discussion of some pandemic indicators, with an in-depth diary of the events that took place in Italy. More analytically, the paper wants to highlight how the main pandemic indicators (i.e. the death count for Covid-19) have been constructed and managed also considering the correct perception of their meaning (i.e. the number of infected).

This inappropriate interpretation of the indicators, also carried out by the media, can influence both the choices of policy makers and the community.

The paper is structured as follows:

In Part 2, we address the methodological problem of the quality of measurement to represent reality. This section emphasizes the importance of choosing correct indicators and understanding their meaning; if both aspects are not taken into account, we risk generating a process of "infodemic", with a correlated loss of information.

In part 3, we analyse the importance of data collection, in this case the cause of death, to build homogeneous and therefore comparable indicators.

In part 4, we examine the measures taken by the Italian Government, while also considering the trends in the most important indicators used for the evaluation of the pandemic state.

In part 5, we explore the main statistical indicators, highlighting how the different calculation methodologies can produce different measures of the pandemic phenomenon.

Finally, in part 6, we analyse the potential influence of the communicative model used for cognitive mechanisms inside our minds on the perception of the Covid-19 phenomenon.

2. UNDERSTANDING THE INFODEMIC TRAP AS A LOSS OF INFORMATION. A METHODOLOGICAL OVERVIEW

In order to be used, the indicators must have direct operational definitions and a close semantic relationship with the property to be measured (Marradi, 2007). In addition, only concepts that refer to the specific properties present in the analysis or collection units, on which the phenomenon is analysed, can be considered as indicators. Conversely, the choice of indicators depends on the individual researcher, whose only, but unavoidable, limit is to keep in mind the relationship between the indicator and the indicated concept, without establishing whether it is a cause or an effect (Stevens, 1946; Stevens, 1951; Blalock, 1961; Sullivan, 1974):

> ‘The first principle is that an indicator should identify the essence of the problem and have a clear accepted normative interpretation. The second principle is that an indicator should be robust and statistically validated; the third principle is that an indicator should be responsive to effective policy interventions but not subject to manipulation; the fourth principle is that an indicator should be measurable in a sufficiently comparable way across member states. The fifth principle is that an indicator should be timely and susceptible to revision; the sixth principle is that the measurement of an indicator should not impose a too large burden on member states, on enterprise, or on the Union’s citizens’ (Atkinson et al., 2002, p. 21).

The validity of the indicators, therefore, is first conceptual, even if they must meet, as much as possible, certain requirements in order to be used. In fact, they should be *exclusive* (not replaceable with other indicators) and univocal (interpretable in an unambiguous way with respect to the direction of the variations) (Curatolo, 1974); sensitive and exhaustive (able to express the entire phenomenon observed by recording its variations in time or space); and finally, they should be faithful (able to attribute the same variations to changes in reality), regardless of the unit of measurement adopted. The operativity of the indicators (Maggino et Mola, 2007) (i.e. the ability to translate the concept into an indicator) must be considered, respecting as many of the requirements listed above as possible. In the case of Covid-19, for example, among the characteristics listed, faithfulness is the requirement that cannot always be met.

In addition, the ‘operational capacity’ of an indicator depends on the availability of the basic statistical information needed to translate the concept into measurement and use it. The concept of measurement must therefore be understood as ‘empirical measurement’, whose quality in observational data depends on the effectiveness of the instruments used for their collection (Hubbard, 2010, p.126):

> ‘Your problem is not as unique as you think; you have more data than you think; you need less data than you think; an adequate amount of new data is more accessible than you think’. (Hubbard, ibidem, p. 47).

During the pandemic crisis, infodemic (i.e. the circulation of an excessive amount of information, sometimes not accurately sifted) is determined sometimes as different interpretations of the pandemic phenomenon, to the point of producing a different semantic perception of the indicator, as well as generating, in some cases, problems of attribution of meaning (Antolini, 2009):

> ‘We’re not just fighting an epidemic; we’re fighting an infodemic’. (Ghebreyesus, 2020).

We must bear in mind that information has value both for citizens and government agencies:

> 'If you don't compute the value of measurements, you are probably measuring the wrong things in the wrong way'. (Hubbard, ibidem, p. 120).

Pandemic periods are inevitably characterised by high uncertainty and, therefore, by emotional involvement of all social agents, so a first change of variables concerns the amount of information available (Q_{inf}). During this pandemic, the failure to use scientifically valid statistical data, as well as the huge volume of circulating pandemic data, led to an infodemic situation. To help avoid misinformation and disinformation problems, the World Health Organisation (WHO) has provided important recommendations on how statistical information should be managed (WHO, 2020). The pandemic has taught us that increasing the volume of information available does not necessarily produce an increase in knowledge. On the contrary, when the information is not statistically well-founded, this creates a risk of providing numbers without improving the statistical information (Martini, 1998, p. 177). The end result is then the generation of a damage function (D) for society as a whole.

In fact, statistical information changes equation (1) in its volume (Vol_{inf}) and in the velocity (Vel_{inf}) by which information is exchanged in the same unit of time.

$$Q_{in:f} = Vel_{inf} \text{ x } Vol_{inf} \quad (1)$$

In the globalised world of information technology, data communication is also increasingly characterised by the simplification of the concepts underlying the indicator (Ioannidis, 2017; WHO 2020). The simplification of the concept (C_s) is directly proportional (k) to the speed at which information is shared by operators; therefore, we have equation (2):

$$C_s = kVel_{inf} \quad (2)$$

Inevitably, the simplification of the concept leads to the belief that reality is simple, thus producing a loss of information (especially in the presence of infodemic) that is equal to the damage function (D), and can be represented as equation (3):

$$D = Q_{inf} \text{ x } C_s \quad (3)$$

The logic-mathematical formalisation just presented cannot yet be measured, since we have no way to detect perfectly the variable velocity (Vel_{inf}). The volume of information, on the other hand, can be identified in the indicators/variables used by the government bodies in charge of drawing up daily bulletins (in Italy, this is the Civil Protection), although it remains difficult to measure the exact volume in the mass media because of the large use made of them, which cannot be quantified in precise terms.

The damage function (3) can have amplified effects on society, and this depends, as the WHO points out, on the communication tools used to disseminate information on pandemic data (see par.6).

3. THE COLLECTION OF STATISTICAL INFORMATION ON THE DEATHS FOR COVID-19: THE NEED TO USE STANDARDIZED DATA

As regards data concerning Covid-19, in Italy, the numbers are disseminated by the Ministry of Health and the Civil Protection, which have set up a repository on GitHub[2] that contains all data collected since the beginning of the pandemic. In this pandemic phase, in Italy, the decision was made to delegate to the Istituto Superiore di Sanità (ISS) the task of coordinating the COVID-19[3] Integrated National Surveillance, thereby integrating data from already existing surveys (e.g. the 'causes of death' survey) with those coming directly from the bodies that perform precise functions in the pandemic area (i.e. primary source data). The survey on the 'causes of death' uses ISTAT/D 4 and D 4 bis forms to record the annual causes of deaths that occurred in Italy, reporting the information on the death provided by the attending physician or by necroscopy (Part A of the death certificate), as well as demographic and social information (e.g. Part B of the death certificate) by the civil registrar. This statistical survey is subject to EU regulation (Reg CE N.1338/2008; 328/2011) and uses the ICD-10 classification, which defines cause of death as 'the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury' (ISTAT, 2003 p.7). The concept of the initial cause of death is now well assimilated and commonly used throughout the world to measure mortality. Very often, however, identifying the initial cause of death is difficult because multiple disease states may have contributed to a death.

The identification of the initial cause of death, in addition to requiring specific expertise in the medical field, requires knowledge of the selection rules laid down

---

2 https://github.com/pcm-dpc/COVID-19 (last access: November 2020).

3 See Ordinance 640 of the Presidency of the Council of Ministers - Department of Civil Protection and Circular 5889 of the Ministry of Health.

by society's World Health Organisation (WHO) to consider all demographic and health information on the death card and the possible links between the various causes (ISTAT, 2003 p. 8). The certifying doctor is required to indicate in Question 1 of the ISTAT form the most responsible cause of death, even if 'in many cases the form is not filled in correctly and is inconsistent with the recommendations provided by the WHO' (ISTAT, 2003 p. 8). In the case of Covid-19, the link between the pathology and death is affected by the existence of previous pathologies, some of which are chronic, since many affected people are, as we shall see, quite old. Establishing that Covid-19 directly caused the death of the person is therefore a complex operation that requires specific examinations that are not currently carried out systematically. In addition, grey or subjective areas can be established by the doctor, who could be induced to say that the person died because of Covid-19, for example, because he has Covid-19 or because he was in the medical department specifically dedicated to Covid-19. On the death certificate, although the other pathologies that may be present are also certified, it remains impossible to establish how much each of them affected the final event.

To find out more analytically about the causes of death from Covid-19, in the presence of comorbidity, it would have been very useful to analyse a sample of people who died in order to understand the risk associated with Covid-19 in determining the event of death in the presence of other diseases. Moreover, we know nothing about the environmental context or the person's lifestyle.

Behavioural and environmental variables, on the other hand, would be extremely useful in building a comprehensive epidemiological overview.

The data released by the Civil Protection (Figure 1) indicates an increase in the number of daily recorded deaths between March 2020 and November 2020 (the Regions and Autonomous Provinces transmit the death data to the Ministry of Health on a daily basis).

The figure clearly shows an increase in the curve, coinciding with the last weeks of October 2020 and the first week of November 2020. Moreover, the possible 'seasonality' of deaths in the cold months partly confirms the possibility of comorbidity with other diseases (for example, seasonal flu). Another possible consideration on the trend of the curve in question could be a renewed difficulty in a hospital system unable to manage the high demand for hospitalisation, as well as the fact that the Italian population has a very old age structure and patients are therefore unable to stay at home due to the lack of home care. Other possible causes can be found in the reopening of schools and public transport, as already warned in June 2020 by the WHO, situations that led to an increase in the spread of the virus (WHO, 2020), but at the moment we have no analytical record of these aspects.

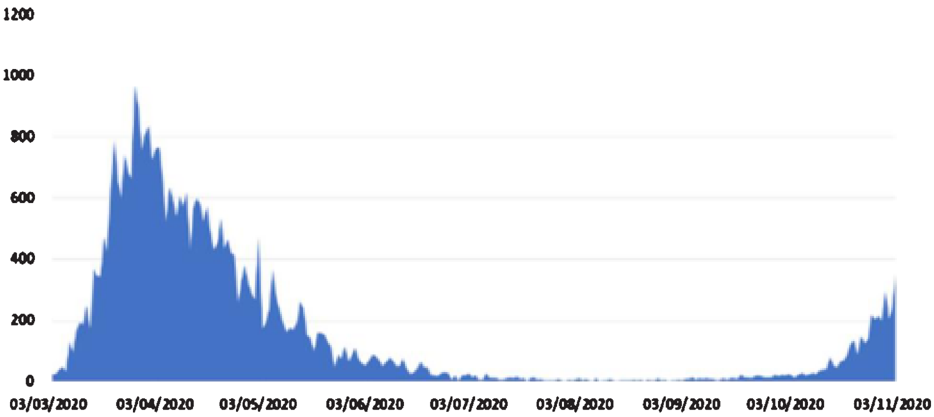

Fig. 1: Daily deaths in Italy (3 March 2020–3 November 2020)

Source: Authors' processing of Civil Protection data

The analysis carried out in Italy by the National Institute of Statistics (ISTAT) has tried to provide clarifications on the aspects previously illustrated by reporting a description of the medical records compiled up to June 2020 (ISTAT, 2020). This report indicates that 71.8% of the deaths of people who tested positive for SARS-CoV-2 had at least one contributing cause: 31.3% had one, 26.8% had two and 13.7% had three or more contributing causes. Of those, only 28.2% who died had no contributing cause of death, with a similar percentage in both sexes and different age groups. In the age group 0 - 49 years the proportion of deaths without contributing causes is lower (18%). The same survey states that Covid-19 is the directly responsible cause of death in 89% of the deaths of people who tested positive for SARS-CoV-2, while for the remaining 11% the causes of death are cardiovascular diseases (4.6%), cancer (2.4%), respiratory system diseases (1%), diabetes (0.6%), dementia and digestive diseases (0.6% and 0.5% respectively). The report on the characteristics of patients who died when positive for SARS-CoV-2 infection in Italy, as drawn up by the Istituto Superiore della Sanità on 4 November 2020, presents the data previously illustrated that refer to the latest available update to the fourth of November and also highlights other demographic data relating to Covid-19 deaths, which are aspects of fundamental importance when carrying out a mortality analysis (ISS, 2020).

According to the report, overall, almost the totality of Covid-19 positive deaths (96,6%) had at least one pre-existing disease: 13.1% had one, 19.1% had two and 64.4% had three or more previous diseases. Respiratory insufficiency was the most commonly observed complication (94.0% of cases), followed by acute renal damage (23.3%), hyper infection (19.2%) and acute myocardial damage (11.0%). The average age of a SARS-CoV-2 positive deceased person is 80 years; the modal classes of patients are those between ages 70–79 years, 80–89 years and finally 90 years and over. In addition, the report shows a very interesting fact that the median age of SARS-CoV-2 positive deceased patients (82 years) is 30 years older than the median age of those who contracted the virus (49 years). A breakdown by gender shows that 57.4% of deaths are male (42.6% female). The median age of women who died after becoming infected with SARS-CoV-2, at 85 years, is slightly higher than that of men, who are 79 years old. Another aspect that deserves highlighting is linked to the temporal trend (on a weekly basis) of the average age of the deceased, which shows a constant increase from the beginning of the epidemic until 85 years of age (1st week of July 2020) before decreasing slightly to lower values (80.50 years of age in the 1st week of November 2020). Here, too, it should be noted that not all the data mentioned are directly accessible and are therefore analysed using the reports of the Istituto Superiore di Sanità (ISS, 2020). This prevents the elaboration of other indicators by age group or territorial units that our access to micro-data, anonymously (to protect privacy regulations), would allow.

Therefore, it seems a difficult choice to understand. If we analyse the data of the main European countries from official sources (Eurostat), we do not have specific references to Covid-19, probably because of problems of homogeneity in data collection measures (Jougla et al., 1998) and precisely because of the difficulty of producing homogeneous data by number and age of the population, whereas data for cause of death are available. The analyses of deaths due to flu are always disseminated, especially in territorial analyses, by using a standardised death rate (Table 1), where deaths are weighted by age structure and with consideration of a standard reference population (Eurostat, 2020).

The analysis of standardised data allows a reasonable comparison to be made at a territorial level. Therefore, observing the disseminated flu mortality rate (Table 1) is possible to declare that the mortality associated with flu is significantly higher for the age group of over 65 years. Furthermore, the variability in the standardised mortality rate – which is much greater in some countries – deserves further investigation.

For instance, it could be useful considering the different organisation of national health systems to provide a possible explanation for the variability of this indicator. It includes, for example, the different level of accessibility to treatment, or the prevention policies adopted in each country through vaccination policies, or, finally, the different accuracy with which this cause of death (flu) is counted (Jougla et al.,1998).

Tab. 1: Causes of death (ICD 10): standardised flu mortality rate per region of residence (NUTS 2), year 2017 (per 100,000 inhabitants)

Countries	Total	Less than 65 years	65 years and over
Denmark	1.3	0.09	6.33
Germany	1.29	0.13	6.07
Greece	0.73	0.15	3.11
Spain	2.29	0.17	11.03
France	N/A	N/A	N/A
Italy	0.88	0.07	4.25
Netherlands	3.38	0.16	16.67
Austria	2.68	0.15	13.11
Portugal	1.01	0.08	4.87
Romania	0.12	0.02	0.53
Finland	4.63	0.23	22.79
Sweden	4.35	0.16	21.65
UK	1.08	0.18	4.84
Norway	5.17	0.32	25.19
Switzerland	3.54	0.15	17.55

Source: Authors' processing of Eurostat Causes of deaths data

Instead, in the case of Covid-19, the death count has been tallied by considering the absolute data at the regional and provincial levels, without standardizing this indicator. For this reason, many of these analyses cannot be carried out. Finally, to gain a better understanding of how much the ageing phenomenon characterizes our country, the Liguria region has 262.2 elderly people for every 100 young people, while that number in Campania is 135.1 (ISTAT, 2020).

4. MEASURES TAKEN BY THE ITALIAN GOVERNMENT AND THE EPIDEMIC SITUATION: A READING OF THE MAIN INDICATORS

Covid-19, identified as infection by the Sars-CoV-2 coronavirus, entered the international scenario on 30 January 2020, following the report by China (31 December 2020) of a cluster of pneumonia cases of unknown aetiology in the city of Wuhan. The WHO declared a state of public health emergency of global concern due to the coronavirus epidemic in China. Italy acknowledged the WHO's warning and from the following day – 31 January 2020 – proclaimed a state of emergency, for a period of six months, as provided for by the current legislation (D.L.n.59/2012).

The state of spread of the virus was unknown. From the very first stages of the epidemic, Italy was inspired by the precautionary principle (Penna, 2020).

On the same day of the state of emergency, a Task Force was also set up at the Ministry of Health to coordinate, in liaison with the relevant international institutions, all control actions to be taken to limit the spread of the virus and verify its compliance with WHO recommendations.

The situation began to take on the features we all know when, on 23 February 2020, following the outbreaks in Lombardy and Veneto, the Council of Ministers, following meetings with the Scientific Technical Committee (CTS), approved Decree Law 6/2020, made operational by the Prime Minister's Decree (DPCM from now on) on 23 February 2020. This decree introduced urgent measures for the containment and management of the epidemiological emergency of Covid-19.

On 1 March 2020, with 1577 positive cases registered and 140 intensive care posts occupied, a new DPCM came into force, implementing Decree Law no. 6 of 23 February 2020. This extended some of the measures already adopted for the containment and management of the epidemiological emergency by Covid-19 and introduced further measures, aimed at regulating the framework of interventions in a unified manner and ensuring uniformity throughout the country in the implementation of prophylaxis programmes.

On 8 March 2020, with by now 6387 total positives, 650 intensive care cases, 366 deaths (+ 359 since the first available count on 24 February 2020) and only 9% recovered (622) from any form of symptom on the recorded cases, the multiplicative phenomenon of the virus was now clear, as was the impossibility of sustaining such a high number of intensive care units. A new DPCM was therefore issued which contained further measures for the containment and management of the epidemiological emergency of Covid-19 throughout the country. More specifically, art. 1 of the new Prime Minister's Decree provided for the creation of a single area (the now famous red zone), including the territory of the Lombardy Region (53% of the national total of positive cases observed at the time) (Figure 2) and 14 other Provinces (five from Emilia-Romagna, five from Piedmont, three from Veneto and one from Marche). The indicator considered was the incidence of positives on the total national positives.

On 9 March 2020, a new DPCM extended the measures set out in Art. 1 of the DPCM of 8 March 2020 to the entire country (history will be left with the scenes of people going to trains to return to their families on the evening of 8 March 2020). This was the beginning of the so-called lockdown, which would be further extended on 11 March 2020 due to a situation considered alarming (Figure 3). The variables most considered were those present in the Civil Protection bulletin, i.e. inpatients with symptoms, intensive therapies, total hospitalised, isolated at home, total positives (current), total positives variation, new positives, discharged/recovered, deaths, total cases and swabs (RT-PCR).

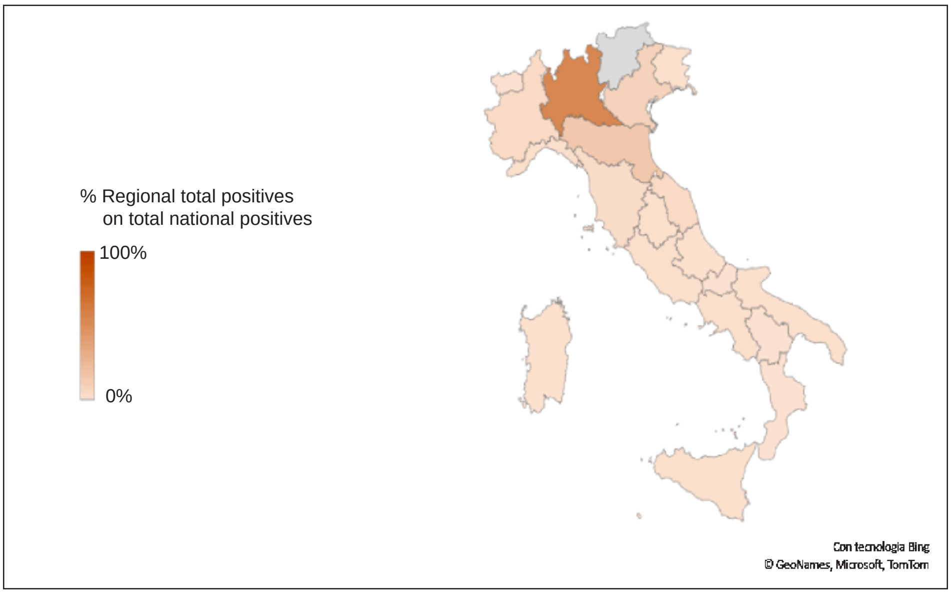

Fig. 2: Regional positives compared to total national positives - 8 March 2020

Source: Authors' processing of Civil Protection data

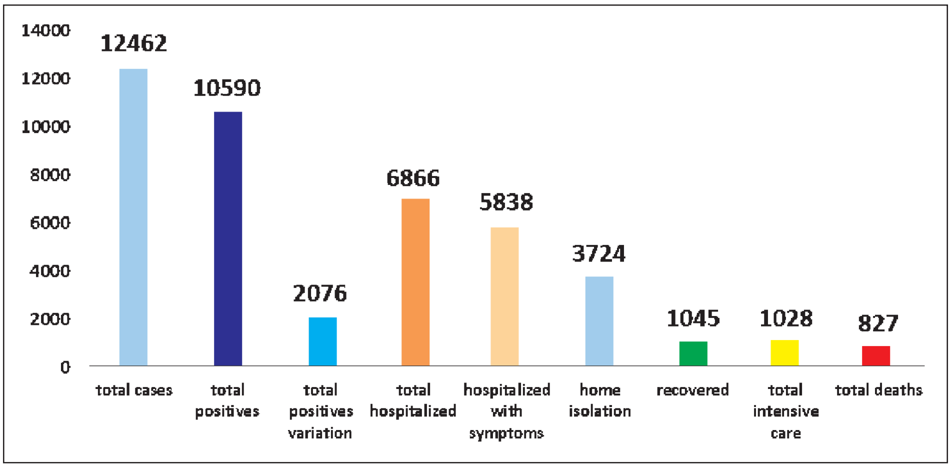

Fig. 3: Overview of Civil Protection Bulletin variables (DPCM - 11 March 2020)

Source: Authors' processing of Civil Protection data

Currently, some aspects with a risk of confusion and which may create misunderstandings should be highlighted. A substantial difference exists between ‘total positive’ and ‘total case’. Total positives refer to the people (stock measures) who are positive for the virus on that given day and exclude those who are no longer positive, which are instead included in the total cases. ‘Total case’ should not be confused with ‘total positive’, as these are two distinct indicators.

The number of people who are ‘positive’ are the number of people alive currently positive, while the number of ‘cases’ also includes deaths and people who have recovered. The total number of positives also includes those who are isolated at home and hospitalised and, as such, represents an important indicator for assessing the demand for care, whether at home or in healthcare facilities. The values of the total positives, total deaths and, finally, of total intensive care should be observed, as they are considered the most important indicators during the pandemic. Their relevance was also determined by the fact that they were the indicators most commented on by the mass media. Instead, the dates have been selected, considering the relevance of the legislative measures taken.

Table 2 shows the absolute values of the indicators mentioned, respectively considering as reference dates the following: 1) the DPCM of 11 March with the eloquent title ‘Io resto a casa’, which constitutes the day Italy was placed on lockdown; 2) the start of Phase 2 on 3 June; 3) the DPCM of 7 September; 4) the decreed law on 7 October comprising urgent measures sanctioning the extension of the state of epidemiological emergency due to COVID-19; and 5) the DPCM of 3 November sanctioning the return of curfew and the establishment of the system of ‘colours’, with three risk zones, namely, yellow, orange and red.

Tab. 2: Total positives, deaths and intensive care recorded on the reference days

DATE (2020)	MAIN INDICATORS (absolute values)		
	TOTAL POSITIVES	TOTAL DEATHS	TOTAL INTENSIVE CARE
11 March	10590	827	1028
03 June	39297	33601	353
07 September	32993	35553	142
07 October	62576	36061	337
03 November	418142	39412	2225

Source: Authors’ processing of Civil Protection data

Table 3 shows the absolute and relative variations of the indicators on the dates shown in Table 2 using as reference bases the previous day (n-1) and the seventh previous day (n-7), wherein n represents the dates in Table 2.

Tab. 3: Absolute and relative variations in total positives, deaths and intensive care on the reference days

DATE (n)	Δ(n-1) and Δ[n/(n-1)]						Δ (n-7) and Δ[n/(n-7)]					
	TOTAL POSITIVES		TOTAL DEATHS		TOTAL INTENSIVE CARE		TOTAL POSITIVES		TOTAL DEATHS		TOTAL INTENSIVE CARE	
	absolute	relative	absolute	relative	absolute	relative	absolute	relative	absolute	relative	absolute	relative
11 March	2076	24.38%	196	31,06%	151	17.22%	7884	291.35%	720	672.90%	733	248.47%
03 June	-596	-1.49%	71	0.21%	-55	-13.48%	-11669	-22.90%	529	1.60%	-152	-30.10%
7 September	915	2.85%	12	0.03%	9	6.77%	6239	23.32%	62	0.17%	35	32.71%
7 October	2442	4.06%	31	0.09%	18	5.64%	9929	18.86%	143	0.40%	46	15.81%
3 November	21630	5.17%	353	0.90%	203	9.12%	141685	33.88%	1507	3.82%	689	30.97%

Source: Authors' processing of Civil Protection data

While the total positives and intensive care occupants are subject to greater variability, since the positives will inevitably either recover or die, the data on the numbers of deceased will show a more stable trend, although increasing over time. On 11 March 2020, the circulation of Covid-19 showed significant variations both in absolute and relative terms. In particular, the week that led to the announcement of the generalised lockdown throughout Italy recorded an exponential trend in the main information indicators with dizzying increases of 291% in the total number of positives (+7884 cases), of 673% observed in confirmed deaths (+720) and of 248% in the intensive care units occupied (+733); this foreshadowed the upcoming collapse of the health system. This trend was also supported by an appreciable daily increase of 24%, 31% and 17%, respectively, which in absolute terms became 2076, 196 and 151, respectively, for the variables considered.

With the Ministerial Decree of 30 April 2020 'Covid-19 Emergency, health risk monitoring activities related to the transition from phase 1 to phase 2A', the criteria relating to health risk monitoring activities (as per Annex 10 of the Prime Ministerial Decree of 26 April 2020) for the evolution of the epidemiological situation were defined. The health risk connected to the transition from phase 1 to phase 2 would be monitored by identifying a number (21) of indicators with threshold and alert values at national, regional and local levels: process indicators on monitoring capacity; process indicators and capacity for diagnostic assessment, investigation and management of contacts; indicators relating to transmission stability and health service maintenance. One immediately evident issue was that the Regions, or at least a large part of them, did not have an integrated information system able to elaborate the 21 indicators provided. In the same way, the algorithm used to measure risk might not be able to express a qualitatively acceptable measure if the indicators could not be processed.

The collection of information and classification is carried out by the Ministry of Health, with the support of a direction cabinet involving the Istituto Superiore di Sanità and the Regions/Autonomous Provinces. However, the collection process appears fragmented and unable to control the quality of the statistical information provided by the offices of the Regions and the Autonomous Provinces.

The respect of the principle of parsimony in the selection of indicators should be the first objective. This would result in fewer indicators but would be consistent with the possibility that they may be further developed, providing quality information for subsequent risk calculations. For example, one unknown is the number of people who enter and leave intensive care units and, more importantly, the number of people who die and who remain alive. The availability of intensive care units, which is an important variable for constructing a territorial risk indicator, is not homogeneous in its definition. In some regions, the definition of the number of intensive care units has been modified by considering not only the intensive care units available, but also those that can be activated. The availability of these data to the scientific community would have encouraged greater collaboration in important analytical activities, such as the preparation of a map of the risk of virus transmission associated with the different classes of daily actions (Lettera 150, 2020).

With a not significant daily variation in recorded deaths (+0.21% between 2 June 2020 and 3 June 2020), even the weekly variation (27 May 2020 – 3 June 2020) observed was very small (+1.60%, for 529 units more). With the decrease in intensive care (-30% and -13.5%, respectively, weekly and daily), even the currently positive cases in June recorded substantial negative variations, with 11669 positives, or a reduction of 23% for the week and 1.50% for the day (596 units).

Italy was thus beginning to emerge from its most serious emergency phase, under the banner of a decreasing situation which authorised interregional movements, and the resumption of almost all activities, leading to a general opening that lasted until the summer months (Figure 4).

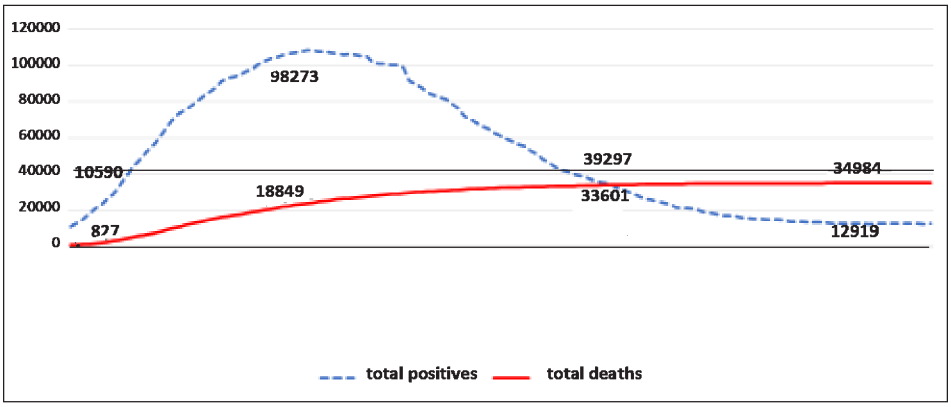

Fig. 4: Trend in total positives and deaths recorded in Italy (11 March 2020 – 31 July 2020)

Source: Authors' processing of Civil Protection data

The situation starts to change from mid-August onwards, with trends showing a clear worsening in Italy. Therefore, a DPCM was adopted – on 7 September 2020 – extending the state of emergency (contained in the Decree Law of 30 July 2020) to 7 October 2020 (Figure 5).

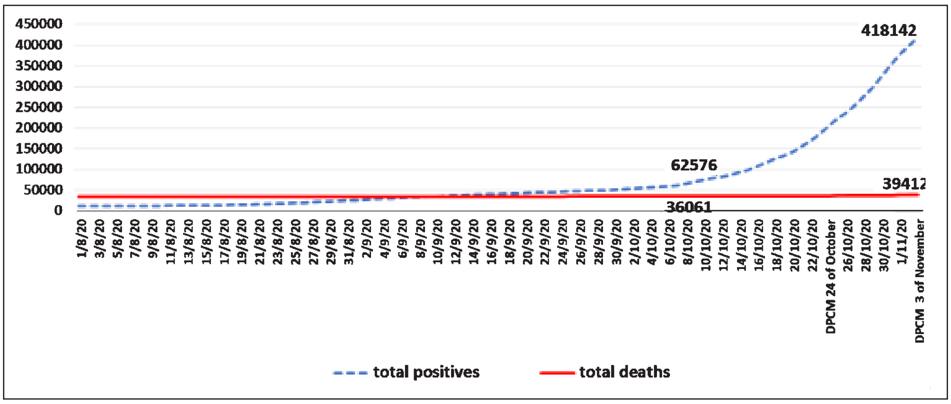

Fig. 5: Trend in total positives and deaths recorded in Italy (01 August – 3 November)

Source: Authors' processing of Civil Protection data

The incremental weekly and daily changes occurred in the current positives (+23% and 3%) was also reflected in the trend of occupied intensive care beds, which was interpreted as a clear warning sign. The 142 beds occupied from 1 to 7 September, combined with a daily increase of 7% (9 beds between 6 and 7

September 2020), support this interpretation, although the deaths still recorded incremental rates of essentially zero (0.17% and 0.03%) and probably contributed to the general perception of tranquility that characterised this period.

In this way, a policy dilemma was revealed: Is Covid-19 dangerous itself or because the national health system cannot manage it? This is a question we must try to answer because, depending on the information provided, the policy choices must necessarily be different.

On 7 October 2020, the Council of Ministers, having regarded a note from the Minister of Health and the opinion of the CTS, decided to extend the state of emergency, declared as a result of the declaration of ‘public health emergency of international importance’ by the World Health Organization (WHO), until 31 January 2021. It also approved a decree law (implementation of EU Directive 2020/739) containing urgent measures related to the extension of the declaration of the state of emergency due to Covid-19 and for the operational continuity of the Covid-19 alert system. The reasons for these decisions need framing in an appropriate manner. In October 2020, Italy remained the only member country still having a state of emergency in force. The situation in the last month showed a significant increase in the positive totals recorded; these were not yet exponential (62576 cases) and the trend in recorded deaths was still very stable and not particularly significant. In addition to the picture outlined so far, however, the number of intensive care beds in use showed a substantial increase in the last period.

The changes observed in the total positives and in the total of intensive care cases now assumed an increasingly stable trend. No longer were only those cases hospitalised with mild symptoms and home isolates growing, but also the intensive care component was expanding.

In addition to the usual stable trend in deaths, the beds occupied in intensive care units were perceived as a risk in the healthcare system. The daily increase of 18 units (+5.64% between 6 October 2020 and 7 October 2020) was an indicator of the impossibility of managing a situation that could soon change, given the growing and constant increase in new positives (+9929 cases in the week), which could have worse consequences.

On 3 November 2020, given the trend of the variables considered, a new DPCM was signed, containing the new measures to deal with the epidemiological emergency from Covid-19 and in force from 6 November 2020 to 3 December 2020. The new Prime Minister’s Decree identifies three areas – yellow, orange and red – corresponding to the different levels of criticality in the country’s regions and for which specific measures were envisaged. The areas are identified through the 21 indicators listed in the DM 30 April 2020.

This decision was taken in light of a national picture that showed a doubling of the positives recorded in the week (418142 from 28 October 2020 to 3 November 2020). The trend and daily changes were relatively minor (+33.88% and 5.17%) but netted much higher absolute values (+141685 and +21630). The trend in deaths also aimed at a concrete increase, well represented at +3.82% in the week (1507 deaths) and the daily trend was now close to 1% (+203 deaths between 2 November 2020 and 3 November 2020). The most critical aspect, however, was represented by the consolidation of the increase in intensive care, which presented very dynamic absolute and relative values in both the weekly and daily calculations.

The decisions taken so far have always considered indicators that give a negative meaning to the phenomenon, and this can be partly shared. However, decisions should also be made by analysing indicators that have a positive meaning (e.g. those relating to recovered people, the concentration in the relevant age groups, or people who have saved themselves from intensive care), as well as taking into account the transition matrix of those who go to intensive care from the hospital ward, or who enter intensive care directly from home, or who die or recover from intensive care. In some cases, hospitalisation is also due to dependent elderly people who do not have adequate home healthcare (Cogis, 2005). The structural or environmental variables, on the other hand, should be distinguished from the epidemiological ones, which are typically individual.

The considerations illustrated so far are in no way intended to affirm that the pandemic situation is not serious. In fact, it is very serious. However, statistical information is required that is adequate for the complex picture and that supports public and private decision making.

5. STATISTICAL INDICATORS AND BIAS OF REPRESENTATIVENESS

One of the difficulties in the correct use of pandemic data has concerned the accessibility of data and the use of indicators or variables that are not always homogeneous. Moreover, also the analyses of aggregate data can produce bias, in their representativeness. In the case of Covid-19 this situation – known as the Simpson paradox (Yule, 1903; Simpson, 1951) – occurs when deaths are counted in aggregate form without specifying the health condition about the existence of previous or chronic pathologies. It may generate an incorrect measure of the level of the lethality or mortality rate which may be underestimated or overestimated.

As illustrated below, an exhaustive overview of the pandemic phenomenon can be obtained by using other indicators as well, such as ‘excess deaths’ (see p. 31).

Several problems arise in the construction of the lethality rate, given that different measures can be obtained depending on the aggregate used as a denominator (Table 4).

Tab. 4: Lethality Rate Indicators

EQUATION	DESCRIPTION	PROS	CONS
L(x) = Nxt/Pxt*100	This measure is the ratio between the number of death and the "total cases" at time t.	It is often used to make comparisons between countries.	The total cases are uncertain.
L_2(x) = Nxt/(Nxt+Gxt)*100	This measure expresses the ratio between the number of deaths and total number of deaths and recovered from Covid - 19 at time t.	In this case the denominator is "certainly measurable".	This formula estimates a higher lethality rate in the early stages of the epidemic than the raw formula (Lx).
L_3(x) = N(x,t) /(Px-t) *100	This measure expresses the ratio between the number of deaths on day X and the "total cases" diagnosed a certain number of days earlier (12 days in the formula).	This measure reduces the systematic error due to the difficulty in recording the number of deaths and sick people at the same time.	This formula estimates a higher lethality rate in the early stages of the epidemic than the raw formula (Lx). Instead, the total cases remain uncertain

Source: Authors' elaboration

With reference to deceased persons, the survival index is the complement to 1 of the lethality rate (L_x)[4] or equation (4), the proportion of deaths *(N)* in relation to the total number of infected persons *(P_x)* in a given time span *(x)* and in the same population (WHO, 2020).

$$L(x) = Nx_t/Px_t*100 \tag{4}$$

[4] Survival Index is equal to 1 minus lethality rate.

This indicator is evidently rather variable and depends on the duration of the observation and, above all, on the uncertainty regarding the considerable difference between the recorded cases and the actual Covid-19 cases (Last, 2001; Battegay et al., 2020). On the other hand, the raw lethality rate, referred to in equation (4), does not allow assessment of the change in the severity of the infection (Figure 6).

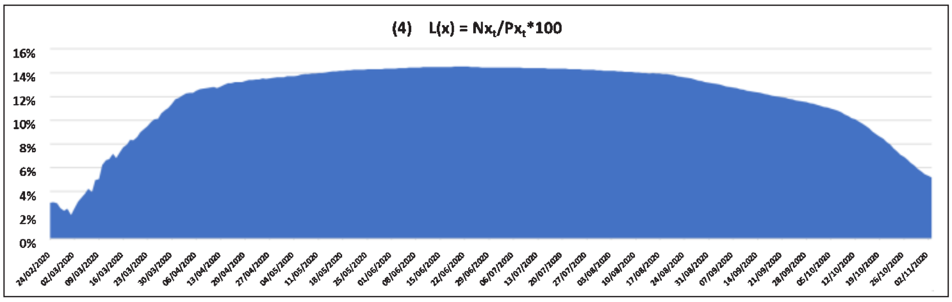

Fig. 6: Lethality rate trend from 24 February 2020 to 3 November 2020

Source: Authors' processing of Civil Protection data

The lethality rate using equation (4) for 3 November 2020 was 5.19%, compared to 39412 confirmed deaths and 759829 total cases (Civil Protection). This lethality rate is the Case Fatality Rate (CFR), i.e. the apparent lethality rate. These data are incomplete because it is impossible (or at least very difficult) to know 100% of the population who contracted the virus at a given time. However, the CFR is often used to make comparisons between countries, while ignoring its sensitivity to certain fundamental differences. As the data shown above confirm, the severity of the infection is influenced by the age and previous diseases of the infected patient. In this case, the comparison of lethality carried out on countries that reveal very different age group structures and very heterogeneous citizen medical records risks neglecting this important information in the judgments made on it. Furthermore, a significant diversity across territories exists in the National Health Systems and the absorption capacity of hospital systems, and this has had a considerable impact in the treatment of the infection, sometimes failing to prevent the deaths of many patients due to the difficulty in caring for them. In this regard, the operational difficulty still inherent in the way in which the patient is declared deceased for Covid-19 should be reiterated. This leads to the recording of non-homogeneous data that are inevitably reflected in a distortion in the comparison between lethality rates with numerators calculated in heterogeneous ways. Another issue of analysis is the inequality in swab-tracing policies in different countries. Compared to the early stages of contagion, all countries, including Italy, have improved their ability to perform tests and therefore to broaden the base of infected people used to perform counting. However, not all of them have done so at the same time, thereby inevitably leading to comparisons between calculated indicators with different operational capabilities.

One method to adjust the estimation of the lethality rate while the epidemic is ongoing may be to use equations (5) or (6) (Worldometer, 2020):

$$L_2(x) = Nx_t/(Nx_t+Gx_t)*100 \quad (5)$$

where:

G = number of recovered people

Equation (5) has the advantage of not needing an estimate of the total case variable and thus using the data of the recovered. This operation gives significant results when the average duration from hospitalisation to death is similar to the average duration from hospitalisation to discharge, as reported by the American Journal of Epidemiology (Ghani et al., 2005).

Using data available from the Civil Protection emergency repository for Covid-19, and using equation (5), the lethality rate on *3 November 2020* was 11.53%:

$$L_2 = \{39412 / [39412 + 302275]\} *100 = 11.53\%$$

However, a systematic error lurks in this methodology because the number of deaths and recovered is considered at the same time, ignoring the course of the infection and failing to meet the criteria of fidelity of the indicators mentioned above. The lethality rate should therefore be corrected by considering the time that elapses in a subject between the diagnosis of the disease and his or her death (Battegay, ibidem). This was estimated to be 12 days, since the median value calculated on the deceased persons was used (ISS, pt.8):

$$L_3(x) = N(x,t) /(Px\text{-}t) *100 \quad (6)$$

Using equation (6), for the same date, the lethality rate was 8.13%:

$$L_3 = [39412 /484869] *100 = 8.13\%.$$

Where 484.869 is the number of total cases on 23 October 2020.

The trends in lethality calculated using (5) and (6) are shown in Figures 7 and 8. Since equation (6) estimates lethality by comparing total deaths on a given day to the total number of cases recorded on the twelfth previous day, the first possible detection is on 6 March 2020 and no longer on 24 February 2020. This is because, until that date, the data on total cases (from the previous 12 days) are unknown.

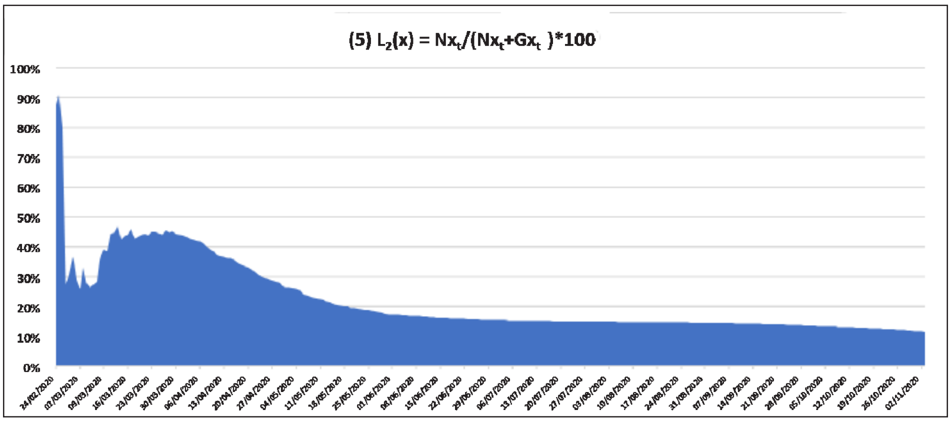

Fig. 7: Lethality rate trend from 24 February 2020 to 3 November 2020

Source: Authors' processing of Civil Protection data

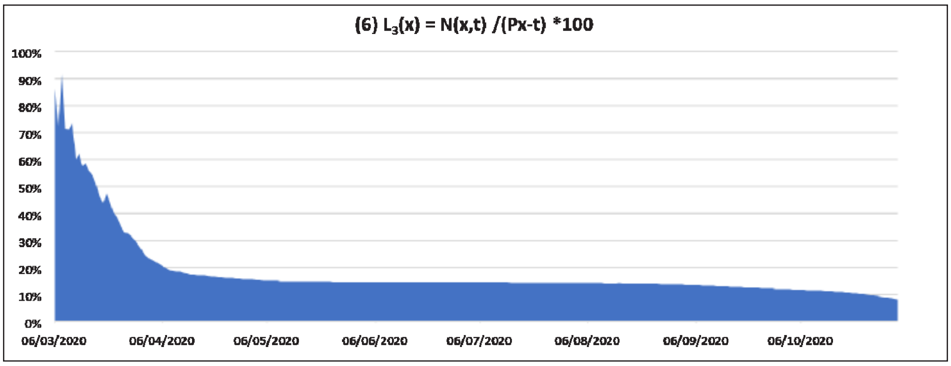

Fig. 8: Lethality rate trend from 24 February 2020 to 3 November 2020

Source: Authors' processing of Civil Protection data

The trend analysis, however, is more representative of the punctual measures. This is because the data, although affected by possible measurement errors, retains an intertemporal coherence, since the nature of the error would be systematic in the same direction in both indicators. The analysis confirms a situation of substantial stability for both indicators, starting from the second half of June 2020, when the equation (5) lethality rate begins to stabilise. Clearly, until mid-May, the rate reported in graph 7 is very different from the one calculated in equation (6) and reported in graph 8. This difference in trends is determined by the number of recovered people who, in the first months, were fewer in number but also had longer recovery times. The lethality rate of equation (6) also appears to be affected by bias if we consider that P_x has no asymptomatic patients who are not tested with swabs. This raises at least two concerns: on the one hand, the P_{xt} population of the denominator of equation (5) be much higher, and, on the other hand, counting the infected without considering the swabs does not allow for a homogeneous and comparable measurement.

To address the first concern, we could consider that the survey conducted by the National Statistics Institute (ISTAT, 2020) on the seroprevalence of SARS-CoV-2[5], with results published on 27 July 2020, allowed the identification of the proportion of people in the general population who have developed an antibody response to the virus and, therefore, would also be undeclared asymptomatic individuals. The results from a probabilistic sample (and therefore extendable to the entire population) of 64,660 people who had blood samples measured from 25 May 2020 to 15 July 2020, made it possible to state that 1,482,000 people (2.5% of the resident population in the household) tested positive for the virus IgG. They developed antibodies to SARS-CoV-2, and therefore had encountered the virus. This number was 6 times greater than the total number of cases detected by swabs and registered (243506) on the same date.

Using these results, the plausible infection fatality rate (IFR) can be calculated. This parameter is calculated by dividing the number of deaths by an estimate of the infections that considers both people with symptoms and weakly symptomatic or asymptomatic individuals. While the numerator calculates the number immediately (net of doubts about the data collection and publication systems), the latter requires further estimates identified by serological tests and work carried out by ISTAT, as mentioned above.

On 15 July 2020, the value of lethality rate of (5) and (6) was 15.15% and 14.50%, respectively. Assuming that the number of people who came into contact with Covid-19 is actually 1,482,000, using equation (4), and assuming that the catchment area is certain and known and that the Covid-19 death count is correct, a lethality rate of 2.36% would be obtained.

It is worth pointing out that from 23 March 2020 onwards, using the data provided by the Civil Protection, the lethality rates calculated from (4), (5) and (6) (i.e. the CFR and its adjustments), present precise values for which formula (4) produces results that are always lower than (6), which is respectively always lower

---

[5] According to the provisions of Decree Law no. 30 of 10 May 2020 'Urgent measures concerning epidemiological and statistical studies on SARS-CoV-2', converted into law on 2 July 2020. The survey was conducted from 25 May 2020 to 15 July 2020.

than those related to (5). The same comparison is obviously not possible with the data provided by the survey on the seroprevalence of SARS-CoV-2, which makes it possible to calculate the IFR, but not the punctual values over time because data don't available in a daily basis.

As mentioned above, equation (5) uses the values of the recovered to provide the lethality data. The values calculated in this way are understandably much higher than those calculated with equation (4) and (6) between March 2020 and May 2020, i.e. the period in which the recovered persons did not greatly exceed the total number of deaths. The differences between (4) and (6) are instead due to the choice to use the total cases on the same day as the deaths (4) versus the total cases 12 days earlier (6) and, inevitably, a smaller number when the trend of total cases is increasing (Figure 9).

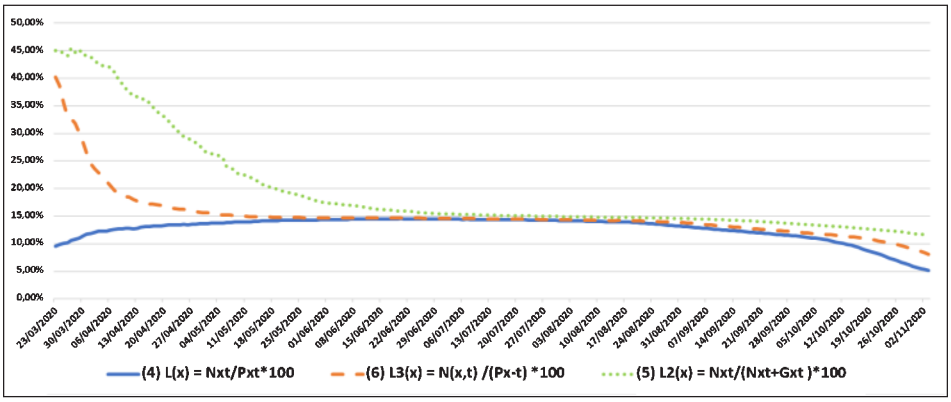

Fig. 9: Comparison of lethality rate trends (4) (5) and (6) from 23 March 2020 to 3 November 2020

Source: Authors' processing of Civil Protection data

As we have seen, all three lethality rate indicators present elements of serious difficulty regarding their validity and deciding which one to use as the main benchmark does not seem an easy exercise. However, one indicator that can provide a more immediate reading of the Covid-19 trend is the ratio between the deaths and recovered (Figure 10), since it is less volatile than the lethality rate and easier to interpret.

This ratio traces a curve that shows a peak in the early months of the epidemic (March 2020 - May 2020), as was the case for the lethality rate in equation (5) and is shown in Figure 6.

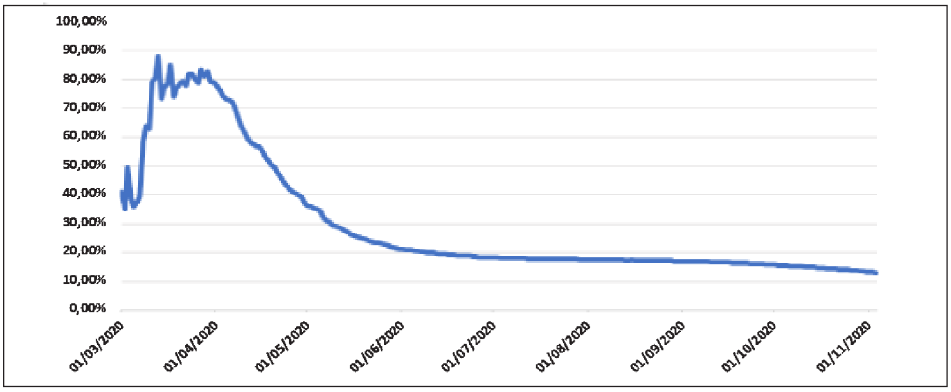

Fig. 10: Covid-19 deaths/recovered ratio (%) in Italy (March 2020–November 2020)

Source: Authors' processing of Civil Protection data

A measure considered more faithful to the lethality rate is the excess mortality, which measures the variation (percentage or absolute) in two different periods of the overall mortality rate, i.e. considering all causes, thus highlighting the deviation from normal conditions. In this regard, the Human Mortality Database, managed by the UK's National Statistics Office, offers the possibility to investigate the mortality that occurred in countries around the world during the Covid-19[6] pandemic. Excess mortality can be measured in several ways. The fastest method is to subtract the average number of deaths recorded in a given period ($D_{t,x}$) (e.g. one week) from the raw number of deaths observed in the same week ($D_{t,x}$) and compare that value to the values in previous years (i.e. the average number of deaths over the same period), calculated over a number of years (Average Deaths $\text{week}_{t,x\text{-}n}$), as in equation (7):

Excess Deaths week (t,x) = Deaths week (t,x) – Average Deaths week (t,x-n) (7)

Figure 11 shows how the raw number of weekly deaths (from 5 January 2020 to 28 June 2020) in 2020 differs significantly from the average number of deaths in the same weeks of the previous five years (2015-2019) using equation (7).

In the period between March 2020 and April 2020, the mortality curve is significantly higher than that obtained by calculating the average number of deaths in the period 2015–2019. This indicator has the undoubted advantage of evading the problem of the exact determination of the causes that determine the event of death. However, if used to compare countries or regions, this indicator produces bias, as it does not consider the different age structure of the population.

---

6 https://ourworldindata.org/excess-mortality-covid# (last access: October 2020).

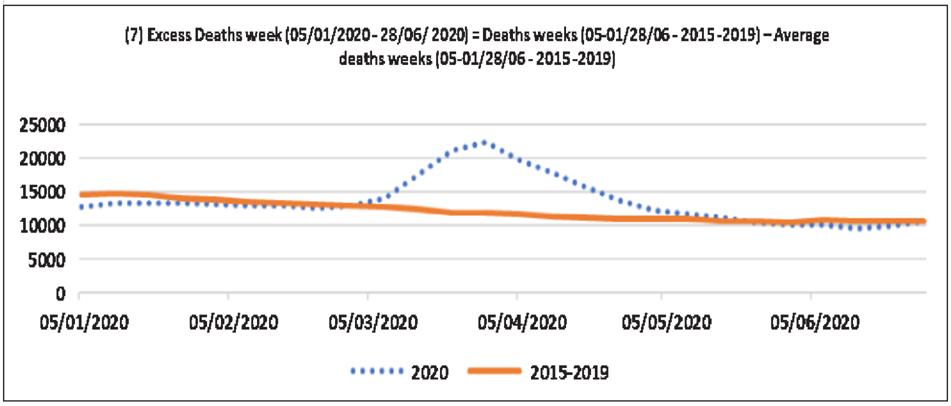

Fig. 11: Excessive weekly mortality (January 2020–June 2020) raw death count recorded in Italy

Source: Authors' processing of Civil Protection data

A more appropriate indicator is the P-score, which instead considers excess mortality in relative terms (i.e. as the percentage difference between the number of deaths per week and the average number of deaths in the same week in the last five years), as in equation (8).

P_{score} = {[Deaths week (t,x) – Average Deaths week (t,x-n)] / Average Deaths week (t,x-n)} *100 (8)

Figure 12 shows the P-scores, broken down by age group in this case. Clearly, the excessive mortality involves older age groups (74 years and older) in the period of greatest lethality associated with the virus (i.e. the months from March 2020 to May 2020). This graphical evidence shows the same trend, but different values are observed in terms of the intensity over the period when the peak of registered deaths due to Covid-19 occurs on 27 March 2020 (969). In percentage terms, the death count between 22 March and 29 March, as recorded in the age group 74 and older, was 101% higher (i.e., two times higher) than the average death count in the same week in the previous five years. This suggests a contribution of Covid-19, both directly and indirectly, to the total number of deaths registered in 2020 compared to the average number of deaths registered in the same period in the previous 5 years.

The trend of the excess mortality indicator is believed to show that the higher number of deaths recorded during the pandemic period is directly or indirectly attributable to Covid-19.

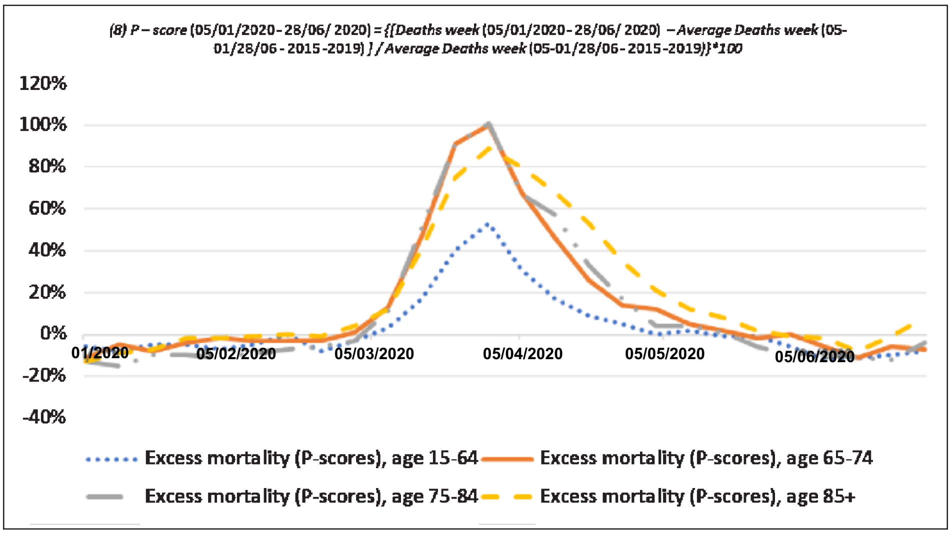

Fig. 12: (P - score) excessive mortality in Italy (January 2020–June 2020) compared to the 2015-2019 average in the same period

Source: Authors' processing of Civil Protection data

Excess mortality, however, is affected by the organisational and structural inefficiencies of the healthcare system. If put under stress, that system may no longer be able to adequately manage patients not affected by Covid-19, with even tragic outcomes (Antolini et al., 2020). The feeling is that our healthcare system has been hit by a stress test to which it could not respond, due to a deficiency in the number of intensive care units and, above all, due to the lack of a turnover that, over the years, has not guaranteed an adequate turnover of retired staff. Austerity policies have produced a linear cut in healthcare, and the territorial reorganisation of healthcare structures itself has, in many cases, not considered important variables, such as the morphology of the territory. Closing a hospital in the mountains is not the same as closing it in a flat area, since the spatial distances, in terms of time, can be very different (Antolini, 2015).

The trend of the new positives is conditioned by the number of swabs carried out and, therefore, the new positive indicator on swabs (or rate of positivity) is considered more reliable (Figure 13).

However, as each person can have more than one swab, one could consider the tested cases instead of swabs. The difference that the value of the indicator assumes is that putting the swabs, or the cases tested, as the denominator, as well represented in Figure 14, gives a more marked difference starting from the first ten days of October 2020. The figure shows the trend of the two curves, starting from 20 April

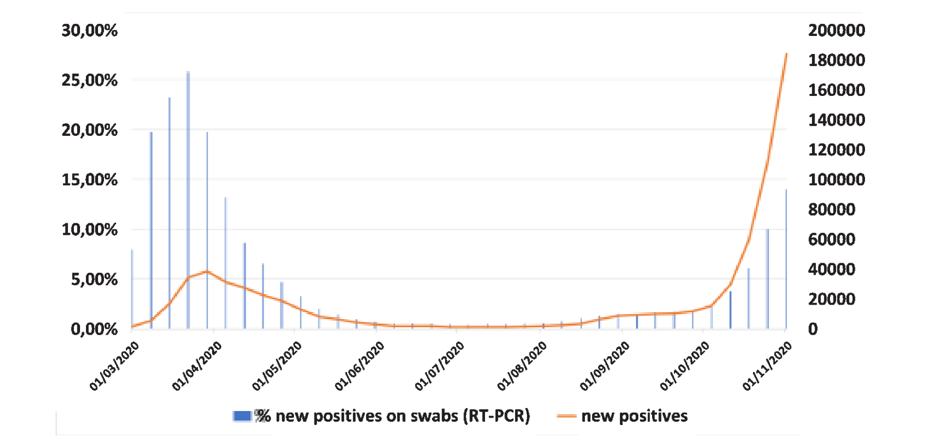

Fig. 13: Weekly trend of new positives and new positives/swabs in Italy March 2020 - November 2020)

Source: Authors' processing of Civil Protection data

2020, the date following the introduction of the variable cases tested at the Civil Protection repository. The divergence that opens in the final stretch of the two curves poses serious problems with regard to the identification of the real positivity rate, which is difficult to identify because the swab data are difficult to trace back to the many cases that produced it (many times the same person is/was tested) and can therefore lead to underestimating the relative positivity of the population.

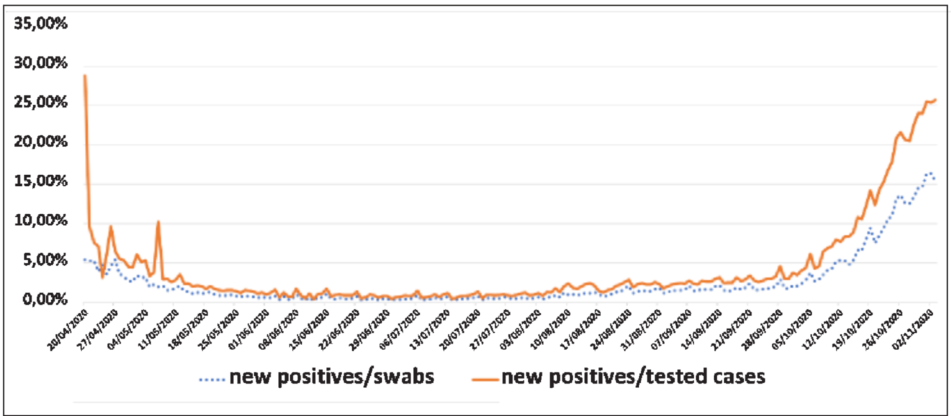

Fig. 14: Trend of new positives/swabs ratio and positive ratio with cases tested in Italy (April 2020–November 2020)

Source: Authors' processing of Civil Protection data

Finally, we need to take a position so that the present study is not used for non-scientific purposes: the pandemic exists and is serious. However, what is more serious is chasing the pandemic instead of managing it, and management can only be achieved using good quality statistics in an integrated form. However, this requires investments in human and technological capital.

6. MENTAL ACCOUNTING AND PERCEPTION BIAS

The need to have accurate information on the epidemiological status of Covid-19 and the capacity of the health system to cope with this emergency have sparked a debate on which indicators to use, as well as their reliability in describing the evolution of the pandemic state. First, however, we should highlight and deepen the way our minds perceive the facts: how we assess them as risky and determine the consequent human actions (Tverky et Kahneman, 1992). The first aspect of this reasoning concerns the perception of the mortality event, since our minds are more conditioned by the aspects to which we attribute a negative value compared to those we judge positive. This is particularly true under conditions where individuals are inclined to make extreme predictions based on inconsistent data (Thaler, 2015; Kahneman, 2017; Kahneman et Tverky, 1979). These are the mental biases responsible for a distorted perception of risk due to framing mechanisms that adapt the available information to an *a priori* constructed partial interpretation scheme (Tversky et Kahnenam, 1981; Thaler, 2015). The criteria that guide the choice and construction of indicators are obviously a function of the study's objective, but they are also affected by the same mental framing. In this case, the negativity of the events that are perceived as risky during the Covid-19 pandemic are the deaths and the need for intensive care hospitalisation.

Over the course of time, however, even the infected – or rather, the increase in new positives – have often been communicated as appearing to match the word 'dead', thereby ending up with a markedly negative connotation. Mental heuristics therefore leads to a circular-associative mechanism of information that acts as a shortcut (Thaler, 2015). The tendency is to belittle, and not deliberately, some of the new positives as asymptomatic or slightly symptomatic. The associative mechanism also influences the interpretation of the data, both at a statistical level and at the level of our mental accounting. As already pointed out by the philosopher David Hume in his essay on the human intellect, published in 1748, the associative mechanism can be traced back to three principles: similarity, contiguity in time and, finally, causality. These conditions have been enriched by the fact that cognitive psychologists are almost all inclined to consider ideas as nodes in a network – associative memory – where connections occur because causes are linked to their effects and things to their property, or because they are associated with the category to which they belong. The strength of association is determined by the way communication takes place. If the word Covid-19 is often expressed together with words like 'infected' and 'death' (Slovic 2000; Slovic, 2004; Stevens, 1975; Strack et al., 1988; Todorov, 2008), then mental heuristics (Kahneman, 2017; Slovic, Finucane et al., 2007) will reproduce the same associative mechanism as well in the future. Moreover, since the event of death is perceived as riskier (Kahneman et Tversky, 1973; Tversky et Kahneman, 1983; Dolan et Kahneman, 2008), the human mind attributes to Covid-19 the meaning of negativity attributed to death. In Italy, many people remember 18 March 2020, the day on which the Italian News showed the army trucks leaving the city of Bergamo with the coffins of the people who died of Covid-19. From that moment on, the associative and cognitive link between Covid-19 and the death event inevitably became stronger.

We do not have a measure of what has just been illustrated, but we can use Google Trends data as a proxy for this phenomenon. They show us how, in our minds, the death event is associated with Covid-19. To this end, we have observed the number of times the query 'deaths and Covid-19' and 'recovered and Covid-19' has been typed into the relevant search engine (Antolini et Grassini, 2019).

The results shown in Figure 15 confirm the greater attention paid to deaths compared to recovered for Covid-19, in the period of greatest criticality (March 2020 and April 2020) in Italy. Research levels show a sharp decrease coinciding with the summer months when the perception of risk was much lower.

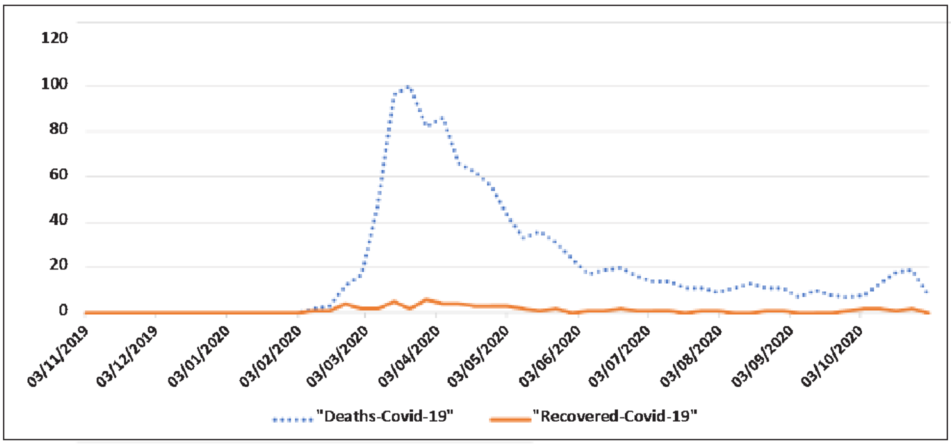

Fig. 15: Google Trends searches with 'Deaths-Covid-19' and 'Recovered-Covid-19' queries in Italy (3 November 2019–3 November 2020)

Source: Google Trends, authors' processing done on 5 November 2020, 23.32 hrs.

Therefore, having institutional communication can be an important asset in emergency situations (Hermes, 2020); however, for this reason, we must first select adequate basic statistical information.

CONCLUSIONS

In the light of the picture outlined here, in addition to an uncertain context in which it is difficult to know what will happen in the next period, the difficulty in interpreting the indicators relating to Covid-19 is evident, starting with the data on causes of death. The risk of an infodemic event is equal to risk of a pandemic one, with damage from the loss of information overlapping the serious one of the deceased (directly or indirectly) from Covid-19. In constructing the indicators, attention must be paid to the availability and quality of basic statistical information as well as to their format (i.e the use of standardized data when making spatial comparisons).

At present, the construction of measures suffers from various problems, both in terms of detection and methodology. These include assessment of the risk in an aggregate form, when an analysis by age group of the population is required.

For example, we do not have transition matrices that can describe how many people die in intensive care and how many return homes. This simple information would make the word ‘risk’ more understandable and, above all, operational. The deaths themselves, or rather their causes, show the difficulty of arriving at a homogeneous and comparable accurate estimate, especially at a territorial level.

The interpretation of the data, if it occurs in aggregate form, can also lead to ‘paradoxical’ interpretations if we do not consider the existence of any previous pathologies or the age of the deceased. The Covid-19 emergency represents a challenge on all fronts of the country’s life, especially from the statistical point of view. Statistics, or rather its quality, is an intangible asset which must be able to contribute to the well-being of the country and the implementation of good practices.

The same indicators may have different estimates depending on the methodologies and data used. The lethality rate, for example, can range from 15% to 2.36%. Lethality is measured by the number of deaths and the number of positive cases that have been confirmed by a medical test. However, among the positive cases, there should also be asymptomatic people who have not been tested or who have recovered without knowing they have been infected. In fact, for Covid-19, the main tracing instrument remains the swab, which is not used on the entire population, for a political choice (Crisanti, 2020). However, for this reason, the number of positive cases tends to be strongly underestimated because those who had the virus in asymptomatic form or with only slight health problems were not subject to swabbing and therefore escaped the total count. This also leads to an overestimation of lethality among the infected. The lethality rate, especially if stratified by age, should also consider the number of the population in different age groups, especially if the country, as is the case in Italy, is geriatric.

Even the indicator relating to excess mortality, which considers deaths for all causes and would be free of the problem of correctly identifying the exact cause of death, is affected by the problem connected to the effectiveness of the healthcare organisational structure. Another possibility is that an increase in deaths has occurred so that an excess mortality has arisen due to the impossibility of treating other pathologies because human and logistic resources have been directed towards limiting the damage of Covid-19. Moreover, analysing the new positives without counting the numbers of swabs carried out (better than the cases tested), can feed a vicious cycle where the greatest risk is to give numbers, when it would be more appropriate to use only statistics to improve the decision-making process of citizens and policy makers.

Nevertheless, the way in which statistical information is disseminated through modern communication channels can contribute to a framing that tends to associate only negative events (the deaths) with Covid-19, while positive events (the recovered) should also be commented on to provide a more balanced judgement. All these factors influence the way our ‘mental accounting’ acts in selecting and interpreting statistical data. In the communicative process, associating infected people with dead people is equivalent to attributing to the first noun the meaning of the second one. This process of mental heuristics is inevitable, and it also happens in financial markets. However, in conditions of information asymmetry, the decision-making process ends up reaching extreme positions. Google Trends’ data (big data) unequivocally demonstrate how, for citizens, Covid-19 is associated with the‘‘deaths’ and only to a small extent with the recovered. It follows that the positives are then themselves considered sick, if not future deceased, while the reality is much more analytical.

The pandemic has taught us that good-quality statistics are crucial for policy making, and even more so when the measurement environment is characterized by a high degree of uncertainty. Setting up an integrated information system is not sufficient if the sources that feed it are not checked for quality. This is why it is necessary to intervene by creating a unified administrative archive that can ensure the transparency of the data entries from which the basic information is taken. This task clearly cannot ignore the role of official statistics, in particular ISTAT (Biggeri, 2020), given that the permanent censuses were built on the construction of unified archives. Moreover, the delegation of the collection of epidemiological data to the regions does not seem to be appropriate, since not all regions have statistics offices capable of collecting and processing data in an adequate manner. The regional activities with regard to statistics require targeted investment to allow the regions to exploit the many administrative data available for the institutional duties they are called on to perform. Similarly, the indicators should be constructed by providing precise definitions (not as was the case for intensive care) to ensure their functionality in mapping the risk to be determined.

In establishing a set of indicators, their level of operability should be considered. Therefore, the difficulties in finding basic statistical information are also important to ensure the accessibility of data by the scientific community.

REFERENCES

Antolini, F., Del Colle, E., Bagnara, S., Castrucci, R., Barrile, R., (2006). *Comunicazione e diffusione dei dati statistici ed aspetti distorsivi. Rapporto di indagine Cogis*. https://trapanimarco.files.wordpress.com/2009/11/comm-garanzia-comunic-statistica-distorsioni-della-comunicazione.pdf. Last access: 11/11/2020.

Antolini, F., (2009). *La Statistica e le Statistiche. Dal dato amministrativo al dato statistico. Fonti e Strumenti di Analisi*. Led edizioni, Milano.

Antolini F., (2015). *Proceedings of the workshop "L'integrazione Socio-sanitaria. Un Obiettivo prioritario per poter rispondere ai bisogni del cittadino utente."* Giulianova.

Antolini, F., Grassini, L. (2019). Foreign arrivals nowcasting in Italy with Google Trends data. *Quality & Quantity*. 53.5: 2385-2401.

Atkinson, A., Cantillon, B., Marlier, E., Nolan, (2002). *Social Indicators: The EU and Social Inclusion.* Oxford University Press. 10.1093/0199253498.001.0001.

Battegay, M., Kuehl, R., Tschudin-Sutter, S., Hirsch, H., Widmer, A., Neher, R., (2020). 2019-Novel Coronavirus (2019-NCOV): Estimating the case fatality rate – A word of caution. *Swiss Medical Weekly*. 150: 5-6.

Biggeri L., *Caos Numeri Covid: mancano criteri su tutto, perché non si vuole usare l'Istat?*. https://www.ilsussidiario.net. Last access: 27/11/2020.

Blalock, H.M, (1961). Social Statistics. Evaluating the Relative Importance of variables. *American Sociological Review*. 26: 866-874.

COGIS, (2005). *L'informazione statistica sull'assistenza agli anziani in Italia.* https://www.camera.it/temiap/temi17/informazione_statistica_anziani_luglio_2005.pdf. Last access: 17/04/2021.

Crisanti A. (2020). Il mio Piano inascoltato per contenere il contagio. *Lettera 150*. 1: 37-38.

Curatolo, R., (1974). Indicatori sociali e qualità della vita. *Rivista Italiana di Economia, Demografia e statistica*. n. 3-4.

Diewert, W. Erwin, (1995). Price and volume measures in the system of national accounts. *Working Papers in National Bureau of Economic Research*, No. 510. University of British Columbia, Department of Economics.

Dolan, P. and Kahneman, D., (2008). Interpretations of utility and their implications for the valuation of health. *The economic journal.* 118.525: 215-234.

Doody, MM, Hayes, HM, Bilgrad, R., (2001). Comparability of national death index plus and standard procedures for determining causes of death in epidemiologic studies. *Annals of Epidemiology*. 11(1):46-50.

Eurostat, (2014). Quality of Statistics. Handbook on Methodology of Modern Business Statistics. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en. Last access: 15/10/2020.

Eurostat, (2020). European Statistical System handbook for quality and metadata reports. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-19-006. Last access: 10/01/2021.

Franchet, Y., (1991). International Comparability of Statistics: Background, Harmonization Principles and Present Issues. */Journal of the Royal Statistical Society.* Series A (Statistics in Society). 154 (1): 19–22.

Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin, J. T., Fraser, C., Lam, T. H, Ho, L. M., Chan, W. S., Anderson, R. M., Hedley, A. J., Leung, G. M., (2005). Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease. *American Journal of Epidemiology*. 162 (5): 479 – 486.

Ghebreyesus T., A., *Munich Security Conference*. https://www.who.int/director-general/speeches/detail/munich-security-conference. Last access: 15/01/2021).

Halpern D., F., Robert, J., Sternberg Scott, O. Lilienfeld, Candice, Basterfield, Shauna M. Bowes, Thomas, H. Costello, Jonathan Baron, D., Alan, Bensley, Douglas, A., Bernstein, Jason L., Braasch, G., Arthur, C., Graesser, Heather, A., Butler, Stephen, J., Ceci, Wendy, M. Williams, Gerd, Gigerenzer, Jane, S., Halonen, Dana, S. Dunn, Henry L. Roediger, III, Jeremy K. Yamashiro, John Ruscio, Kevin Brady, Anthony R. Pratkanis, (2020). *Critical Thinking in Psychology*. Cambridge University Press, 2 edition 2020.

Hermes, (2020). Niente sarà più come prima. *Journal of Communication*. 16: 99-122. ISSN 2284-0753, DOI 10.1285/i22840753n16p99.

Hubbard, D. W., (2010). *How to measure anything. Finding the value of "intangible" in business.* Wiley& Son, second edition, New Jersey.

Ioannidis, J. P. A., (2017). *Statistical Biases in Science Communication: What We Know About Them and How They Can Be Addressed*. The Oxford Handbook of the Science of Science Communication, Scholarly Research Reviews.

ISS, (2020). *Report Caratteristiche dei pazienti deceduti positivi all'infezione da SARS-CoV-2 in Italia*. https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia. Last access: 24/11/2020.

ISTAT, (2003). *La codifica automatica delle cause di morte in Italia: aspetti metodologici e implementazione della ICD X*. Istat, Roma. http://www.asl9.marche.it/mmg/DOCUMENTI/Codifica%20automatica%20delle%20cause%20di%20morte.pdf. Last access: 18/01/2021.

ISTAT, (2020a). *Conto trimestrale delle Amministrazioni Pubbliche, reddito e risparmio delle famiglie e profitti delle società.* https://www.istat.it/it/files//2020/10/Conto-trimestrale-AP-2020Q2.pdf . Last access: 27/11/2020.

ISTAT, (2020b). *Primi risultati dell'indagine di sieroprevalenza sul sars-cov-2.*

https://www.istat.it/it/files/2020/08/ReportPrimiRisultatiIndagineSiero.pdf. Last access: 27/11/2020.

ISTAT, (2020c). Impatto dell'epidemia Covid-19 sulla mortalità: cause di morte nei deceduti positivi a Sars-Cov-2. https://www.istat.it/it/files/2020/07/Report_ISS_Istat_Cause-di-morte Covid.pdf. Last access: 27/11/2020.

ISTAT, (2020d) *Bilancio demografico nazionale, Anno 2019*. https://www.istat.it/it/files/2020/07/Report_BILANCIO_DEMOGRAFICO_NAZIONALE_2019.pdf. Last access: 27/11/2020.

Jougla, E., Pavillon, G., Rossollin, F., De Smedt, M., Bonte, J., (1998). Improvement of the quality and comparability of causes-of-death statistics inside the European Community. EUROSTAT Task Force on "causes of death statistics". *Revue D'epidemiologie et de Sante Publique*. 46(6):447-456.

Kahneman, D., Tversky, A. (1973). On the Psychology of Prediction. *Psychological Review*. 80: 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. 47: 313-327.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press. doi: 10.1017/CBO9780511809477.

Kahneman, D., (2017). *Pensieri lenti e veloci.* (Italian Edition). Mondadori, Milano.

Kelly J., (1982). Possible Approaches to Improving the International Comparability of Immigration and Emigration Statistics - An Illustration of the Canadian Situation.'*International Migration Review*. 16 (3): 619-634.

Last, J. (2001). *A dictionary of epidemiology* (4th ed). Oxford University Press, Oxford. ISBN 0-19-514168-7, OCLC 43810820.

Lettera 150 (2020) Dati e Covid. La battaglia di lettera 150 per la trasparenza, a cura di Antolini, F., Branchini, P., Brasini, S., Contucci, P., Crisanti, A., Davide, F., Luisiani, A., Petrucci, A., Ravagna, G., Roccetti, M., Valditara, G., Zucchelli, C., (2020).

Lettera 150. 1 (2): 15-18. https://www.lettera150.it/wp-content/uploads/2020/12/Lettera150-002-interattivo.pdf.

Maggino, F., Mola, T., (2007). *Definizione e costruzione di indicatori soggettivi. Aspetti metodologici e presentazione di una applicazione allo studio dei reticoli, ASTRIS - Applicazioni Statistiche nella Ricerca Sociale*. Firenze University Press, Archivio E-Prints, Firenze.

Marradi, A., (2007). *Metodologia delle scienze sociali,* Il Mulino, Bologna.

Martini, M., (1998). Servizi di Pubblica utilità non profit: problemi di definizione, classificazione e registrazione statistica. In G. Vittadini, *Il non profit dimezzato*. Etas Libri, Milano, 177-191.

Slovic, P., (2000). *The Perception of Risk*. EarthScan, London, Sterling (VA).

Slovic, P., (2004). Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality. *Risk Analysis.* 24 (2): 311-312.

Slovic, P., Finucane, M., Peters E., MacGregor, D. G., (2007). The Affect Heuristic. *European Journal of Operational Research.* 177 (2007): 1333-1352.

Penna, T., (2020). Il principio di precauzione tra diritto e scienza nell'emergenza da CoViD-19. *BioLaw Journal - Rivista di BioDiritto.* Special Issue1/2020: 563-570.

https://iris.unito.it/retrieve/handle/2318/1739872/612893Il%20principio%20di%20precauzione%20tra%20diritto%20e%20scienza%20nell%e2%80%99emergenza%20da%20CoViD-19%20_TPenna_BioDiritto2020.pdf. Last access: 17/02/2021.

Simpson, E. H., (1951). The interpretation of interaction in contingency tables. *Journal of the /Royal Statistical Society B.* 13 (2): 238-241.

Stevens, S., (1946). On the theory of scales of Measurement. *Sciences*. 103 (2684): 677-680.

Stevens, S., (1951). Mathematics, Measurement and Psychophysics. In S. Stevens, *Handbook of Experimental Psychology*, Wiley, New York.

Stevens, S., (1975). *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects.* John Wiley & Sons, New York.

Sullivan, J.L, (1974). Multiple indicators. Some criteria selection. In M. Blalock, *Measurement in the Social Sciences*, 243-269. Aldine, Chicago.

Thaler, R. H., (2015). *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company, New York. ISBN 978-0-393-08094-0.

Todorov, A., (2008). Evaluating Face Trustworthiness: A Model-Based Approach. *Social Cognitive and Affective Neuroscience*. 3: 119-127.

Tversky, A. e Kahneman, D., (1981). Causal Schemes in Judgments Under Uncertainty. In M. Fishbein (Ed.), *Progress in Social Psychology*. 1: 49-72. Erlbaum, Hillsdale (NJ).

Tversky, A. e Kahneman, D., (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*. 90: 293-315.

Tversky, A., Kahneman, D., (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*. *5*(4): 297-323.

WHO, (2020a). *Estimating mortality from Covid-19.*

https://apps.who.int/iris/bitstream/handle/10665/333642/WHO-2019-nCoV-Sci_Brief-Mortality-2020.1-eng.pdf?sequence=1&isAllowed=y. Last access: 27/11/ 2020.

WHO, (2020b). *Coronavirus desease 2019 (Covid-2019). Situation report - 86.*

https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf?sfvrsn=c615ea20_6.

WHO, (2020c). *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation.* https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation. Last access: 21/02/2021.

WHO, (2020d). *Statement – Preparing for the autumn is a priority now at the WHO Regional Office for Europe.* https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-019/statements/statement-preparing-for-the-autumn-is-a-priority-now-at-the-who-regional-office-for-europe. Last access: 21/02/2021.

Worldometer, (2020). *Coronavirus Mortality Rate (Covid-19).*

https://www.worldometers.info/coronavirus/coronavirus-death-rate/.

Last access: 27/11/2020.

Yule G.U. (1903*).* Notes on the theory of association of attributes in Statistics. *Biometrika*. 2 (2). 121–134.