A novel approach to evaluate the asymptomatic COVID-19 contagions

Behrouz Pirouz [1], and Galileo Violini [2,*]

[1] Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende 87036, Italy; behrouz.pirouz@unical.it

[2] Pontificia Universidad Católica Madre y Maestra, Santo Domingo, Dominican Republic;

[*]Correspondence: leoviolini@yahoo.it

The original aim of this paper was to analyze a characteristics exhibited by certain data of the COVID-19 pandemic, that has been object of many comments and speculations, namely, the violent disagreement in lethality observed in Italy and Germany.

The results obtained put in evidence that, when comparing the time development of the deaths with that of contagions, one should not ignore that there is a time difference between the moment the illness is observed and the death of the patient. If one introduces a rough estimate of this effect, it arises quite naturally a method to answer one of the most important questions underlying many analysis of this pandemic. How large is the fraction of the population which is infected, and therefore contagious, in an asymptomatic way?.

The lethality is measured by the ratio R between fatalities and confirmed contagions. At the date of 31 March 2020, the ratio of Italian to German lethality was about ten, against a ratio of about 1.4 in the number of the contagions.

It is not surprising that both in Italy and Germany this fact and its possible reasons have been object of much attention and investigation. The most popular tentative of explanation had as starting point the remark that in Italy a relatively high rate of deaths occurred in the population segment above 70. This, accompanied by the remark that Italian average age is higher than the German one, provided a possible clue to explain the fact.

Another possible popular explanation focused on possible differences in the assessment of the cause of the death. An Italian analysis had indeed shown that in about fifty per cent of the deaths, the person had at least three, out of a list of about ten, serious illnesses, in some case being terminal. For the remaining 50%, there was a fifty-fifty distribution between people with one or two of those sicknesses, and only the 0.8 % died because of only COVID-19.[1] The implicit question was whether perhaps in other countries, and particularly in Germany, such multiple-cause deaths could have been ascribed to the previous illness.

A review of these arguments may be found in three papers that have been published between 30 March 2020 and 31 March 2020, two of them in newspapers of major circulation in Italy.

On 31 March 2020, Paolo Magliocco observed that scientists pointed to the difference in age population, to the lowest intergeneration relationship, and to the difference in health services quality[2].

The day before, Stefano Luna, had considered several hypotheses[3]. First of all, he quoted that German physicians excluded the possibility of differences in the diagnosis of the death cause*[1]. This forces to look for other possible explanations, such as the delay of the outbreak, considered nevertheless unlikely, the location of the epicenter of the outbreak,(in Italy it was localized in the North, in Germany it was spreaded in different Länder), the delay in the mitigation measures, the policies of hospitalization, (more crowding in Lombardy and selected hospitalization in Germany, the former associated also with a high number of fatalities among Health personnel*[2]), more efficient policy of tests, reserved in Germany to more symptomatic people, and again the age, but in a different sense, suggesting that there is a difference of the age distribution of the infected people.

The same day, Tonia Mastrobuoni in La Repubblica[5], provided some interesting data and informations that deserve to be quoted. The first is about tests, whose number in Germany seems to be now 500000 per week, against 160000 in Italy. She pointed out that the number of tests is a data on which the information is scarce and not easy to obtain. However, for what concerns Germany, she mentioned that three weeks ago they were 50000 to increase some time later to 160000. This fact should be remembered if one wants to analyze the connection between the number of the tests and the ratio between deaths and confirmed contagions. In fact the difference between Italy and Germany for that indicator has been a constant feature all over the time elapsed since the outbreak, and adds a mistery to a mistery when one considers that in the beginning Italy was the second country, if one does not consider Bahrein, for number of tests[6], and as reference 5 underscores, the increase of number of tests should contribute to reduce the deaths to contagions ratio. Moreover, she obtained confirmation that there are no univocous criteria to assess the cause of the death, and provides arguments which might support the hypothesis of an age bias in the cases which made the epidemic burst.

These press articles contributed to diffuse the results of a number of academic papers that had been published in the previous days. A very clear one is that of Matteo Villa[7], quoted in reference 5, which distinguishes between the concept of apparent lethality and

---

*[1] However, a posible contradiction of this statement will be mentioned in the following paragraph

*[2] On April 3, the number of deaths of Italian physicians has reached the value of 69, that of nurses 10, and that of infected Health personnel 10000.[4] This may cast some doubt on the estimate of Italy in Fig.3.

plausible lethality. With an analysis which applies to Italy the approach of the Imperial College group[8], Villa's paper estimates (although with a broad uncertainty) that the number of infected people is at least in the order of five times the confirmed contagion cases. If confirmed this result would reduce the significance of the R difference between Italy and Germany.

In his paper, Villa notices that the apparent lethality varies also in the different Italian regions, even more than at world level or between Italy and Germany, since in the region with the highest deaths to confirmed cases ratio, Lombardy, R=13.6%, and in that with the lowest ratio, Basilicata, R=1.1%.

Today a CNBC paper presents an interview to a German scholar and politician, Karl Lauterbach, which summarizes the German interpretation of their performance[9]. Two more arguments were introduced, the slow path of the initial outbreak, that allowed to treat the patients in the best hospitals of the country with a reduced pressure on them, and this indeed is very different from what happened for example in Milan or Bergamo, and in general the very high level of the Health system, for which the politician prevails on the academic, stressing the importance of the coalition to which his party belongs for the German Health policy.

Among the possible explanation of different developments of the pandemic, one can mention the idea of different effects of the virus, because of genetic mutations[10]. However, a much more economical approach seems to be to assume a uniform lethality irrespective of the country whose data are considered.

In this paper, we propose ourselves to study the possible correlation of what Villa calls apparent lethality with a number of indicators that are related to some of the mentioned suggestions. As we shall see, no convincing correlation is found, but our analysis might suggest a modification in the way of evaluating the apparent lethality.

We shall analyze the data for the ratio R, considering the 17 countries that, at the date of 31 March 2020, had the largest number of confirmed contagions. The data are taken from the usual daily updating of the pandemic[11].

The first indicator we will consider is the average age of the country[12]. The results are presented in Fig.1. On the x-axis are the average ages and on the y-axis the R values. The points represent the data for a given country. Similar explanation will hold for Figures 1 through 6, and for benefit of the reader, after Fig. 6, Table 1 will summarize the numerical values entered in Figures 1 through 6.

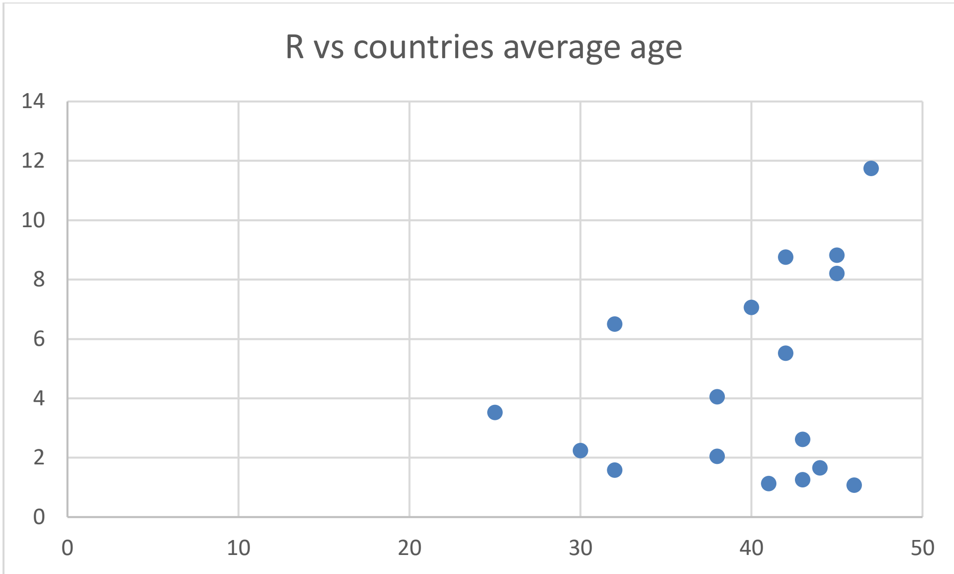

Figure 1

This figure seems to exclude any correlation between the ratio R and the average age of the country's population.

It has been suggested that the structure of the age and intergeneration relationship might explain the difference between Italy and Germany. For what concerns intergeneration relationships, attention was called to the difference in family relations between Italy and Germany, enhanced by the consideration that the outbreak started in different ways, largely due to hospital overcrowding by old people in Northern Italy and to diffusion by young people who had participated in age-characterised events in Germany. Despite an interesting qualitative research on this topics[13], we were unable to identify numerical indicators that could be put in relation with the data. Instead, for what concerns the discussion on age structure, we shall postpone it for consideration in Figures 5 and 6.

We shall study first the possible relation with an economical indicator of the countries. This consideration has a connection with another suggestion about possible differences of diffusion of the pandemia in different countries, that of home and/or office temperature. This in the winter is related to the heating cost which on turn is related to PPP (purchasing power parity)[14] . Therefore we explored a possible relation between R and countries PPP [15]. Figure 2 exhibits our results.

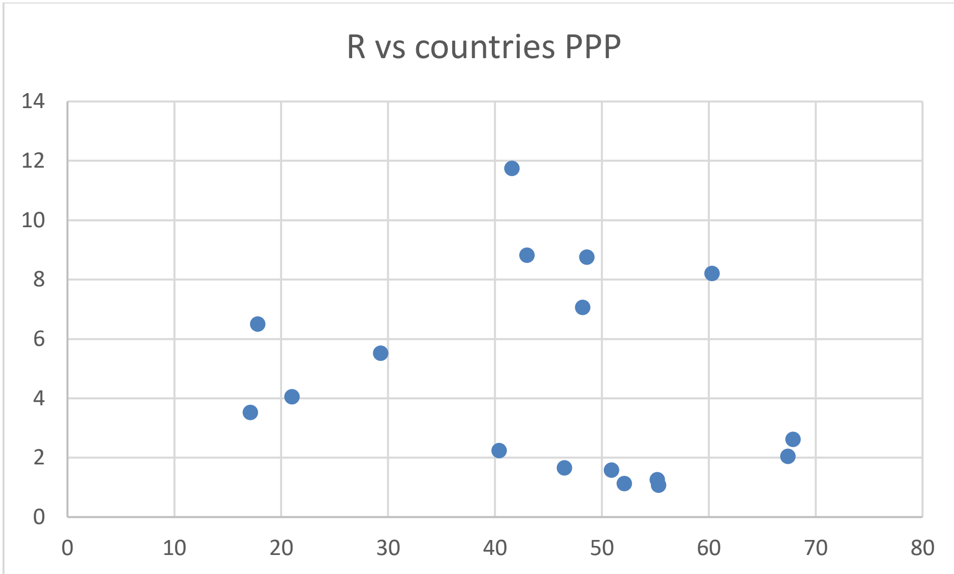

Figure 2

One may notice some regularity in this figure. There is a group of four countries with similar PPP per capita (Turkey, China, Iran, Brazil) grouped around a ratio of about 5, whereas wealthiest countries are divided in two clusters, one, to which Germany belongs, around 1-2%, and the other, to which Italy belongs, with a much larger value (7 to 11%).

The second sector we considered was the Health sector. For the sanitary system, we considered two indicators, the efficiency of the Health care[16] and the number of beds per person [17] available in the country. The corresponding results are presented in Figures 3-4.

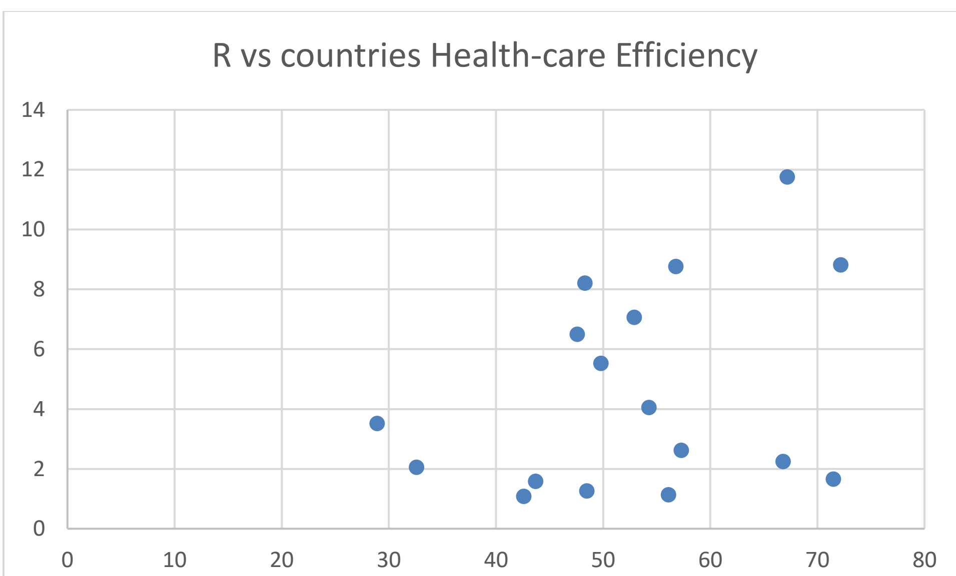

Figure 3

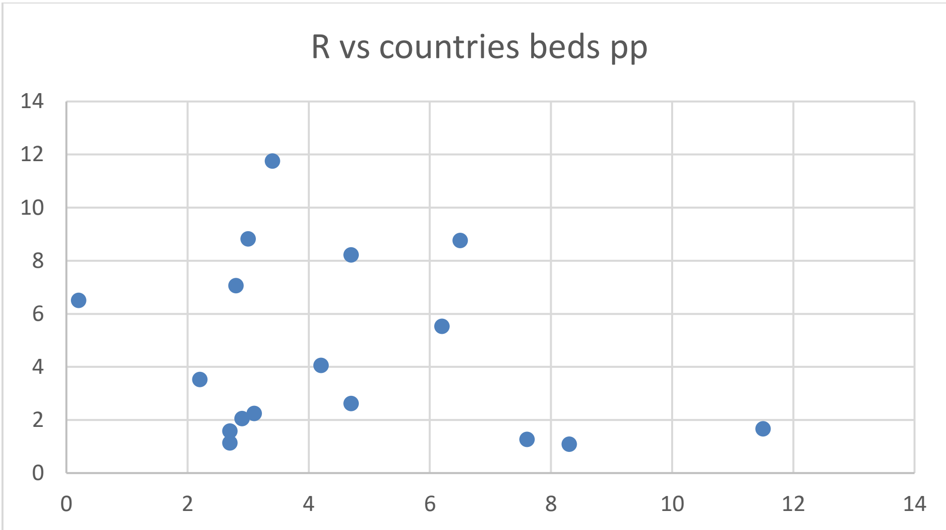

Figure 4

The results of these figures do not provide hints to understand the difference between Italy and Germany. Indeed, the first graph shows that Italy occupies a high ranking, and the same happens both, for Spain (with an R value of about 9) and for Israel and South Korea, with R values of about 2.

The number of beds per person could appear to be a good quantifier of the difference between Italy and Germany, but it is clear that is validity as an indicator explaining why deaths are more or less frequent worldwide is not confirmed when one considers the other countries, insofar one finds a similar performance in other four countries having one quarter of the beds per person of South Korea.

In a previous version of this paper, the column in Table 2 referring to this graph was wrong and repeated the second column of that Tabke. We apologize for this mistake now corrected, and thank Dr. Carlos Rodríguez for calling our attention to it.

Let us now return to demographic considerations. Instead of looking at the average age, we shall consider the ratio between the population older than 65 years and that between 25 and 64[12]. The results are presented in Figure 5.

This figure seems to tell us something. There is a regularity involving the majoriy of the countries whose representative points are lying along a line, but again there is a cluster of five countries which exhibit a different characteristics, those with R larger than 7%.

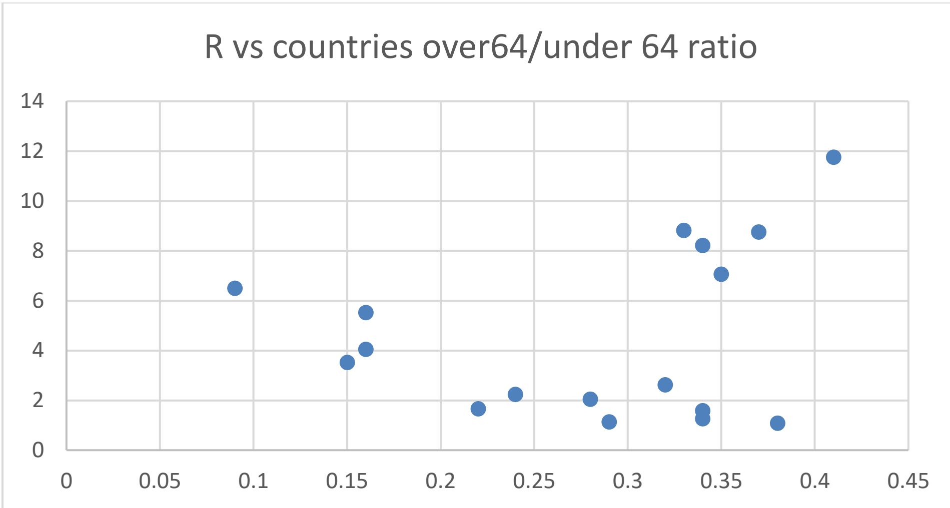

Figure 5

Do we learn anything if we look at another demographic quantity, that of the deaths per million? Probably yes, because the line of reasoning of, for example, the paper of ref.7 assumes that the behavior of the pandemic should be the same everywhere.

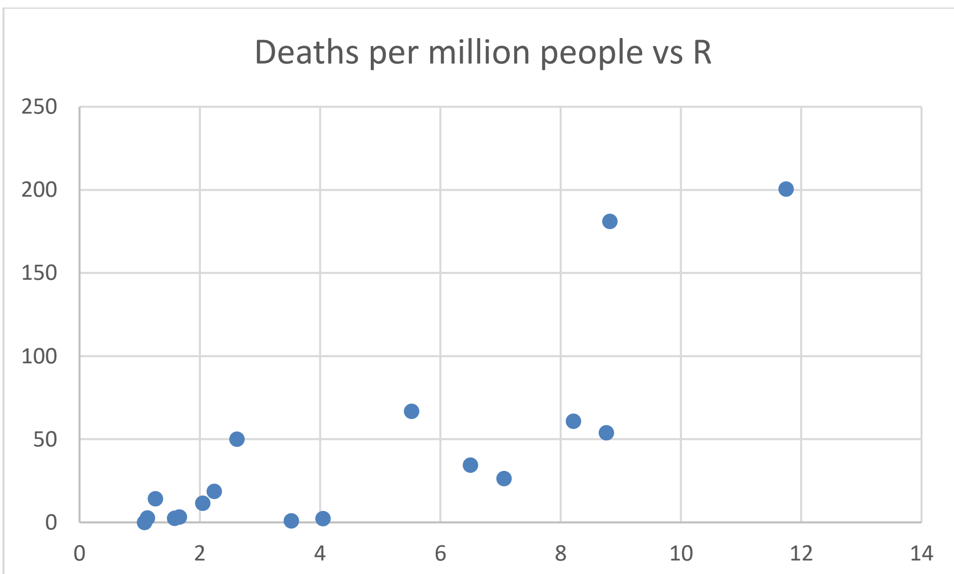

Figure 6

As we could have easily imagined, Italy and Spain are very far from the rest of the countries. However these data might suggest something about the problem of unobserved contagions.

If we buy the idea that one would expect that the effect of the virus should not depend on the country, one would expect that the number of the deaths per million should be the same.

Actually in Figure 6 most points lie on a strip on its lowest part. One can evaluate the average of this data, disregarding the differences between the populations. The result is 43. However, in Italy and Spain this number is not 43, it is about five times larger.

This is the fact that needs to be understood, probably more important than the differences between what percentage of confirmed contagions are the deaths. Unfortunately, even if one may arrive to reduce the 11% to about 1 %, through an increase of the contagion case which calls for a larger number of asymptomatic or semisymptomatic cases, the absolute number of deaths would remain to be explained.

For this, epidemic models can tell us that Italy is ahead of several European countries a couple of weeks, but it is unclear whether this explanation will be enough. Most probably one should try to clarify what was written in reference 5. “When we requested details to Koch Institute, the email answer was that there is no univocous criterium for the tests made after the death”.

The opportunity of uniform criteria is a something to consider, if not for this, for other possible epidemic, hopefully less violent than the COVID-19 one.

At level of understanding our data, it might deserve attention the consideration of the role played by several cases mentioned in the press where there have been situations of collective diffusion of the infections to vulnerable groups. This happened in some convent, or in some retirement house or in some crowded hospital. The study of these cases is perhaps more a job for the historians who will reconstruct these situations, but its effect on R is evident. Large numbers in a correlated way were added simultaneously to the numerator and denominator of R, with the effect of making it increase.

These considerations show that the assumption that the development rate of the ratio between deaths and contagions is country independent cannot be excluded and that deviation from this should not necessarily lead to infer that the evaluation of the number of contagions escaping counting is inadequate.

Table 1

Country	R	Fig 1	Fig 2	Fig 3	Fig 4	Fig5	Fig 6
Italy	11.75	47	41.6	67.2	3.4	0.41	200.6
Spain	8.82	45	43	72.2	3	0.33	181.1
France	8.76	42	48.6	56.8	6.5	0.37	54
Netherlands	8.21	45	60.3	48.3	4.7	0.34	60.8
UK	7.06	40	48.2	52.9	2.8	0.35	26.4
Iran	6.5	32	17.8	47.6	0.2	0.09	34.5
Turkey	5.52	42	29.3	49.8	6.2	0.16	66.8
China	4.05	38	21	54.3	4.2	0.16	2.3
Brazil	3.52	25	17.1	28.9	2.2	0.15	1
Switzerland	2.62	43	67.9	4.73	4.7	0.32	50.1
Israel	2.24	30	40.4	66.8	3.1	0.24	18.6
USA	2.05	38	67.4	32.6	2.9	0.28	11.6
South Korea	1.66	44	46.5	71.5	11.5	0.22	3.2
Belgium	1.58	32	50.9	43.7	2.7	0.34	2.5
Austria	1.26	43	55.2	48.5	7.6	0.34	14.2
Canada	1.13	41	52.1	56.1	2.7	0.29	2.6
Germany	1.08	46	55.3	42.6	8.3	0.38	0.01

Analogous consideration might perhaps help to explain the difference observed between different Italian regions. In this case one might recall that in Italy mitigation measurements were imposed more or less at the same time (around 10 March 2020) all over the country. At that moment the level of contagions was quite different and a South-North delay has indeed been observed on the regional outbreak. Region-independent measures might have produced effect on the development rate of the contagions, but that of the deaths would have been affected only at a later point.

This idea suggested us to explore how the results of the country analysis would be modified if one introduces by hand in a very simple way some delay between the contagions and the deaths.

The purpose of the following calculations is only to explore what might be the effect of this approach, without any pretention of rigor and of best fitting the data.

The basic point is to compare data at different days, with an amount of shift in date based on a rough estimate of the period between the contagion and the death.

We carried on two comparisons of what happened in the seventeen countries. In the first one we calculated the lethality on 31 March 2020 referring to the total number of contagions observed on 24 March 2020. The result is shown in Fig. 7. No special ordering has been assumed for the countries. The only relevant comment is that the ratio has an average of 0.099 which corresponds also to the median of our points. The mean quadratic error is 0.287. The two points above 0.2 correspond to UK and Belgium.

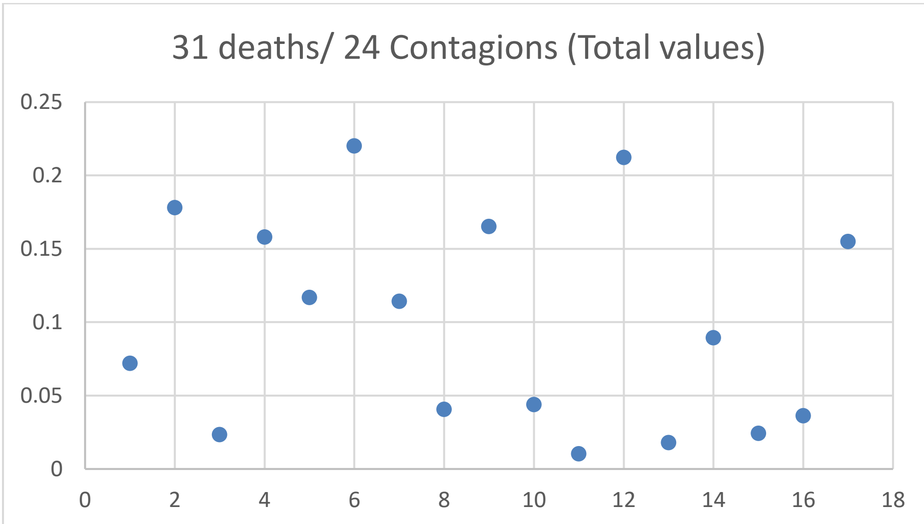

Figure 7

In Fig. 8 we compare the number of deaths on 31 March 2020 and that of the accumulated contagions between 17 March 2020 and 24 March 2020. The actual values entered in the figure, renormalized by dividing them for the South Korean value are presented in Table 2.

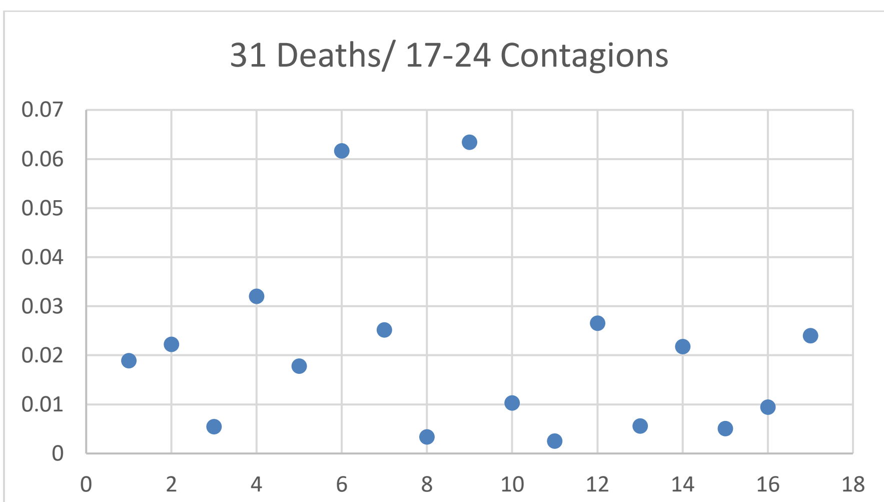

Figure 8

The average is 0.021with a mean quadratic error of 0.073, but this time one finds that practically all countries, again but Belgium and UK, have a similar behaviour. Italy is on the average, whereas Germany is not.

These results with the corresponding quadratic errors may be compared to those of the lethality calculated with data from the same day, in which case the average was 4.6 with a quadratic error of 13.43.

Within the statistically similarity of these results, one might suspect that the approach corresponding to Figure 8 is more promising in order to provide a unified description of this outbreak, perhaps with a modeling in some way more sophisticated than ours of the delay between the contagion and the death.

However, we can try to use our results to make a rough estimate of the number of asymptomatic infected people. We mentioned that the criteria which lead to test a patient have been different in different countries. This means that we do not really know to which extent the number of affected people includes persons with minor manifestations, if any, of the symptoms.

One can make two hypothesis of ratios that be country invariant. One is the ratio of confirmed cases to confirmed plus asymptomatic cases. A way of measuring this ratio would be a broad test in a site affected by the pandemic. A first example of this is what has been done in Vo', where it was found that the number of infected people was larger by 50-75% than that of confirmed infected people. However, the number of cases was limited (Vo´s population is slightly larger than 3000). An extension to a broader sample in Veneto has also been proposed. Of course, if the ratio is country invariant, the research could be carried out any where. One could think to a rather isolated region like Faroe Islands, where the pandemic was carefully tested already mid March, and is expanding at a relatively low path. Another possibility could be to consider a town which is a local focus of the pandemic with a limited population. For example, in Dominican Republic this would be the case of San Francisco de Macoris, town of 132000 inhabitants. One could select randomly (for example from the electoral lists) a statistically significant sample to determine experimentally the ratio between symptomatic and asymptomatic infected people.

Experimental statistics can be useful also in order to detect possible dependence on ethnical characters. Such a dependence is less likely to be relevant if one assumes that delayed lethality is an invariant characteristics of the pandemic*[3].

In order to develop this idea let us consider a particularly significant country ratio, that of South Korea. As it is well known, this country has carried on an extensive testing with a

---

*[3] Implicitly we are assuming a similar response capacity of the health system. This may be not true in situations like that of Northern Italy, and implies that the result we shall mention in Table 2 for Italy is possibly overestimated. One of us (G. V.) is indebted to Italian cultural attaché in Mexico, Professor Emilia Giorgetti, for an enlightening discussion of this aspect

careful follow-up of the contacts of the infected people. If we compare the values in Fig. 8 with the South Korean one we find that almost all of them are larger.

How can we interpret the deviations from 1? There should be no need that we remind that our calculation was oversimplified by the consideration of a period of 7 to 14 days to compare the delay between contagions and deaths, so that our results must be considered as indicative.

This allows to consider the value for Germany and Austria in agreement with that of South Korea. We recall that Germany too is among the leader countries for testing. The other two countries with a ratio smaller than 1 are China and Israel. The model might be inadequate for China because the data refer to the last period of the outbreak. For Israel, the number of deaths on 31 March 2020 was one of the smallest ones (20). This makes statistically less significant the case, and its fast growth in the following four days confirms it.

Therefore, if we renormalize the number of contagions to the South Korean case, and compare the value with that of confirmed cases we can estimate the lost cases. Of course, the specific value of this estimate depends on the choice of 7-14 days as delay between the data, but nevertheless might provide an order of magnitude estimation of the number of asymptomatic cases. The results are presented in Table 2.

Table 2

Country	R_C/R_{SK}	March 31st contagions	Estimated asymptomatic contagions ,	Country	R_C/R_{SK}	March 31st contagions	Estimated asymptomatic contagions ,
Italy	3.98	105792	315000	Switzerland	1.85	16605	14000
Spain	4.76	95923	360000	USA	3.39	188172	450000
France	5.74	52128	247000	Belgium	11.37	12775	132000
Netherlands	4.31	12667	42000	Canada	1.69	8527	6000
UK	11.06	25412	256000	Turkey	5.52	13531	61000
Iran	3.19	44605	98000	Brazil	3.91	5717	17000

We are grateful to Dr. Paolo Violini for his critical reading of the manuscript.

POSTSCRIPTUM

The results of our analysis for Italy are in very good agreement with those of two papers that follow a similar line of reasoning to estimate the number of infected people who should be added to that of the confirmed cases in order to have an estimate of the total infected people.

These papers came to our attention only after the completion of ours, but the coincidence of the three estimates, following similar though not identical lines of comparison, is remarkable.

The first paper has been published by Simone Ferro on 24 March 2020[A]. It follows an approach very similar too ours. The data it uses for comparison are South Korea number of deaths and age distribution. The limitations of the underlying assumptions are described very carefully and it is stressed that these limitations do not allow to consider the results presented in that paper more than an order of magnitude estimate. Nevertheless two results deserve to be mentioned here. The number of undiscovered cases, using 22 March 2020 data would be projected on 31 March 2020 to a value of 360000 asymptomatic cases, in remarkable agreement with our estimate. The second result refers to the age disaggregation of the cases, which is peaked in the 40-49 years sector.

It could be interesting to merge the two criteria and perform an analysis that makes use of both the age and the contagion-death delay. Given the coincidence of the results it is easy to predict that the basic result will remain unchanged, but most probably its statistical significance would be enhanced.

The second paper was actually published earlier, on 20 March 2020, but because of the greater similarity of Ferro's paper with our approach we preferred to discuss first Ferro's paper[B]. It is a paper of the LaSapienza Group. In this case the reference sample is China, and the comparison is made on the basis of the death number per age (with a significant correction due to different average age between the two countries) and on the assumption that the Chinese and Italian time evolution of the symptomatic cases be the same (with a shift of 39 days). The careful discussion of possible reasons of uncertainty of the results makes perhaps this order of magnitude calculation less stringent that that of Ref. A. In particular it is underscored that the Chinese data are extraordinary accurate in Guandong and less in Hubei, where the data could also be affected by the famous 40000 people banquet of 18 January 2020. However, in conclusion this analysis also points to a factor 4 reduction of the Italian lethality, similar to our result, and that, if corrected for the different average age, could be further reduced by a factor 2.

Let us return to the agreement with the paper of ref. A. Probably the Italian case has several peculiar reasons that make difficult to analyze it. The close agreement with ref.

A, using two different criteria of comparison (age and delay of the death) gives confidence on the validity of the method and, indirectly supports the prediction of the order of magnitude of the undected cases in the other countries analyzed in our paper.

References

1- Istituto Superiore di Sanita, Report sulle caratteristiche dei pazienti deceduti positivi a COVID-19 in Italia, 17/3/2020

2- https://www.lastampa.it/topnews/primo-piano/2020/03/30/news/che-cosa-sta-succedendo-in-germania-1.38657062

3- https://www.nonsprecare.it/perche-germania-pochi-morti-coronavirus

4- https://tg24.sky.it/cronaca/2020/04/02/coronavirus-medici-morti.html

5- https://rep.repubblica.it/pwa/generale/2020/03/30/news/germania_coronavirus_decessi_mistero_numero-252719991/

6- https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-COVID-19-ita-scheda-regioni-20200310.pdf

7- M. Villa, Coronavirus:la letalitâ in Italia, tra apparenza e realtà, ISPI, March 20,2020

8- N.M. Ferguson et al., Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, March 16, 2020

9- H. Ellyatt https://www.cnbc.com/2020/04/03/germany-has-a-low-coronavirus-mortality-rate-heres-why.html

10-L. Cuppini, “È vero che in Lombardia si muore di più perché il virus è mutato?”, *Corriere della Sera*, 21 marzo 2020.

11-https://coronavirus.politologue.com
https://coronavirus.jhu.edu/map.html

12-https://ourworldindata.org/age-structure

13-P. A. Dykstra, Cross-national Differences in Intergenerational Family Relations: The Influence of Public Policy Arrangements, Innovation in Aging, Vol. 2, January 2018, igx032

14-https://en.wikipedia.org/wiki/List_of_countries_by_electricity_consumption

15-"World Economic Outlook - GDP per capita". International Monetary Fund. October 2019. Retrieved 30 March 2020.

16-https://knoema.com/infographics/opambfb/health-care-efficiency-around-the-world

17-https://www.indexmundi.com/g/r.aspx?v=2227&l=it

Postscriptum references

A- S. Ferro, https://www.lavoce.info/archives/64469/quanti-sono-davvero-i-contagiati-in-italia March 24, 2020

B- S. Franchini, E. Molinari, G. Parisi, F. Ricci Tersenghi, https://www.scienzainrete.it/articolo/ma-quanti-sono-gli-ammalati-di-covid-19-italia/simone-franchini-enzo-marinari-giorgio, March 20, 2020