Does death from Covid-19 arise from a multi-step process?

Neil Pearce[1], Giovenale Moirano[2], Milena Maule[2], Manolis Kogevinas[3,4], Xavier Rodo[3,5], Deborah A Lawlor[6,7], Jan Vandenbroucke[1,8,9], Christina Vandenbroucke-Grauls[10], Fernando P Polack[11,12], Adnan Custovic[13]

(1) London School of Hygiene and Tropical Medicine, London, UK
(2) Unit of Cancer Epidemiology, Department of Medical Sciences, University of Turin
(3) ISGlobal, Barcelona, Spain
(4) IMIM (Hospital del Mar Research Institute), Barcelona, Spain
(5) ICREA, Barcelona, Spain
(6) MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
(7) Population Health, Bristol Medical School, University of Bristol, Bristol, UK
(8) Leiden University Medical Center, The Netherlands
(9) Aarhus University Denmark
(10) Department of Medical Microbiology and Infection Prevention, Amsterdam UMC, The Netherlands
(11) Fundacion INFANT, Buenos Aires, Argentina
(12) Department of Pediatrics, Vanderbilt University, Nashville, TN
(13) National Heart & Lung Institute, Imperial College London, United Kingdom

ABSTRACT

The Covid-19 death rate increases exponentially with age, and the main risk factors are having underlying conditions such as hypertension, diabetes, cardiovascular disease, severe chronic respiratory disease and cancer. These characteristics are consistent with the multi-step model of disease. We applied this model to Covid-19 case fatality rates (CFRs) from China, South Korea, Italy, Spain and Japan. In all countries we found that a plot of ln (CFR) against ln (age) was approximately linear with a slope of about 5. We also conducted similar analyses for selected other respiratory diseases. SARS showed a similar log-log age-pattern to that of Covid-19, albeit with a lower slope, whereas seasonal and pandemic influenza showed quite different age-patterns. Thus, death from Covid-19 and SARS appears to follow a distinct age-pattern, consistent with a multi-step model of disease that in the case of Covid-19 is probably defined by comorbidities and age producing immune-related susceptibility.

INTRODUCTION

Respiratory viral infections are a leading cause of mortality globally (1), and their global impact has been highlighted by the huge public health threat of coronavirus disease 2019 (Covid-19) (2) Primary infection with SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) and a relatively mild disease followed by successful clearance of the virus seems to represent the norm. However, in a minority of patients, there is systematic inflammation and an exaggerated immune response (cytokine storm) resulting in respiratory and multi-organ failure and potentially death (3).

SARS-CoV2 is particularly aggressive for the elderly, who represent between 73 and 90% of fatal Covid-19 cases (4, 5). Mortality, need for intensive care, mechanical ventilation, and hospitalizations increase with advancing age (6-8). Median age among 20,133 patients hospitalized with Covid-19 in the UK was 73 years, and more men were admitted than women (60% vs. 40%) (9). In hospitalized patients older than 75 years, mortality is 29.4%, and in intensive care units (ICU) can reach 43.5% (10). It is currently unclear whether children have similar infection rates to adults (11, 12), or are equally infectious; however, when infected they have fewer symptoms, and deaths in children are very rare (13).

Overall, death from Covid-19 increases monotonically and steeply with age. This is in contrast with other respiratory pathogens, such as respiratory syncytial virus (14) and seasonal influenza (15, 16) which cause substantial numbers of deaths in children and have a J-shaped pattern with age. In contrast, pandemic influenza viruses often cause severe disease in middle-aged adults (17-19), even without pre-existing co-morbidities (20), suggesting different pathogenesis from that for seasonal influenza. One exception was mortality from pandemic influenza in 2009, which had different age-distribution to other influenza pandemics (15); severe disease was relatively high in young adults aged 20 to 40 years, with comparatively few cases among those >65 years of age, and appeared to be driven by a non-protective anamnestic response to the new virus by previously acquired antibodies. Thus, there are some notable features of Covid-19 death which suggest that it has a different relationship with age than most other severe respiratory pathogens.

The mechanisms explaining excess Covid-19 mortality in elderly are unknown (21). Having an underlying condition such as hypertension, diabetes, cardiovascular disease, severe chronic respiratory disease and cancer increases the risk. However, a strong predictor of mortality after adjusting for major comorbidities is increasing age itself (9). Furthermore, the

extreme rarity of fatal outcome in children suggests that one must have gone through a number of changes, either as a result of unknown aspects of aging, or as a result of chronic disease (effectively speeding up the aging process), which increase susceptibility, leading to severe disease and death. These characteristics of Covid-19 death are consistent with the multi-step model of disease (22), a model which has primarily been used for cancer (23, 24), and more recently for amyotrophic lateral sclerosis (ALS) (25). If one assumes that a disease (or an outcome such as Covid-19 death) is caused in a single step process, then the incidence in a particular year will be proportional to the risk of having undergone that specific step in that year; this will depend on the level of exposure to the relevant causal factor. If it requires more than one step, then the chance of undergoing the first step by a particular age will be proportional to that age (e.g., a 60 year old is twice as likely to have gone through a particular step as a 30 year old); if there are n steps involved, then the probability that the first n-1 steps have occurred by a particular age will be proportional to age to the power of n-1:

$$i = u_1 u_2 u_3 ... u_{n-1} u_n t^{n-1}$$

Where i=incidence, t=age, n=number of steps, and the u_i are a series of constants. It then follows that:

$$\log(i) = (n-1)\log(t) + c \quad \text{(where c=}\log(u_1 u_2 u_3 ... u_{n-1} u_n))$$

Thus, a plot of log (incidence) against log (age) will be a straight line, and the slope (n-1) will be one less than the number of steps (n). For epithelial tumours, the average number of steps appears to be about 6, and in some instances (such as colorectal cancer), the relevant steps have been identified (24). Similar findings have been obtained for ALS, whereas some other chronic neurological conditions (e.g. multiple sclerosis) do not follow a multi-step pattern (25).

Given what is known about the age-distribution of Covid-19 death, and its clinical characteristics, we hypothesised that Covid-19 death follows an age-pattern consistent with the multi-step model. One might question whether such a model, which assesses risk across all ages, would apply to Covid-19 death, since Covid-19 is a new disease. However, evidence to date suggests that usually a person will die from Covid-19 because they have been infected and they have one or more chronic conditions. Thus, one can conceive of the process involving the acquisition of these chronic conditions and infection with SARS-CoV-2. Briefly, if someone has already accumulated a number of the necessary steps, then they are

'primed' to have severe Covid-19 if infection occurs. If they are not 'primed', then infection can occur, but disease is usually mild. This is most obvious in children, who are unlikely to accumulate the steps, and in whom infection is almost always mild. Thus, it is plausible that the multi-step model could apply to death from Covid-19, with the early steps occurring through life-course and 'priming' an individual to have severe disease if and when they experience the final step, i.e. SARS-CoV-2 infection. Since males have a higher death rate in each age-group than females (9, 26), we also aimed to compared the findings between sexes.

METHODS

Issues with available data for testing the multi-step model for Covid-19 deaths

As we have discussed elsewhere (27), there are a number of potential problems with reported Covid-19 infection and mortality data. These include non-random population sampling for testing (mostly symptomatic people get tested) which means that diagnosed cases are likely to be more severe, misclassification of infection and/or misclassification of cause of death. Moreover, there is little valid information on the underlying population infection rates. This means that estimates are available for the case fatality rate (CFR=deaths/diagnosed cases), but few estimates are available for the infection fatality rate (IFR=deaths/total infected). For the current analyses, using IFRs may have been preferable, but in almost all cases only CFRs were available in our data sources. Even for the CFR, there are significant problems in estimating it accurately (28, 29), as is the case for pandemic influenza (30). Under-ascertainment of infections is most likely to be a problem in younger age-groups where there is likely to be less testing, because there are fewer symptomatic infections (13).

Data sources

We needed data on Covid-19 CFRs or IFRs by age-group, ideally separately for males and females. We sought potential data sources through literature searches, examination of national statistics online, and through enquiries with colleagues in different countries. In some instances, population death rates (i.e. with total population denominators) were available, but these are not appropriate for the multi-step model in the current analysis which focusses on case fatality. Therefore, we restricted the analysis to data with CFRs (cases as denominators) or IFRs (infections as denominators).

China: The data for estimates of the CFR and IFR for China were taken from the publication by Verity et al (31). Briefly, these involved 44,672 PCR-confirmed cases in Wuhan and elsewhere in China during the period 1st January to 11th February 2020, extracted from the

WHO-China Joint Mission Report (32); during the same period 1,023 Covid-related deaths were reported across China. The main analyses in the publication were adjusted for censoring, demography and under-ascertainment. We report the findings using the unadjusted estimates, since these are more comparable to the data from other countries. However, only the adjusted IFR is available in the report and therefore to compare the findings for the CFR with those for the IFR in the China data we used the adjusted estimates for both.

South Korea: Data were retrieved from the online website of the Korean Centre for Disease Control (KCDC). KCDC included all confirmed Covid-19 cases collected by the surveillance system from 20th January to 17th of March 2020. During the identified timespan a total of 8,320 cases and 81 Covid-19 related deaths had been confirmed in South Korea (33).

Japan: The data for Japan were obtained from the website of the Ministry of Health, Welfare and Labour, 6th May 2020; up to that date there were 15,300 cases and 537 deaths reported.

Italy: The data for estimates of the CFR for Italy were obtained from the scientific report released by the National Health Institute (ISS) on April 17th (34). The data refers to all laboratory-confirmed cases collected by the national surveillance system from 21st February to 16th April 2020. During the identified period 159,107 Covid-19 cases and 19,996 Covid-related deaths were reported in Italy.

Spain: The data was obtained from the official reports of the Spanish government (Ministry of Health and the Instituto de Salud Carlos III). It includes the reported cases, recovered and deaths by age category (35). During the reporting period up till 18th May 2020, there were 239,125 Covid-19 cases and 19,186 deaths.

As a comparison, we also obtained similar data for three other respiratory diseases (seasonal influenza, pandemic influenza and Severe acute respiratory syndrome-SARS). Finally, we conducted an analysis of all-cause mortality to explore whether our results for Covid-19 reflected a more general aging process. The details of the data sources, and findings for these, are shown in the Supplementary Appendix.

Data analysis

Under-ascertainment of cases is mostly likely to be a problem in younger age-groups, and deaths are rare in this age-group and the rates are therefore unstable (ln (CFR) cannot be calculated when there are zero deaths). We therefore included all ages (0-99 years) in the descriptive analyses, but restricted the multi-step analyses to ages 30 years and over (an approach that has been used in some other analyses of Covid-19 mortality (36)). We analysed

age-specific (and where available, sex-specific) death rates (i.e. CFRs and IFRs) in ten-year age-groups (for publications which reported five-year age-groups we amalgamated these into ten-year groups in order to standardize the analysis). We used the mid-point of each age-group and regressed the natural log of the CFR or IFR against the natural log of age, using standard linear least-squares regression. As in similar previous analyses (25), we used unweighted regression so that the estimated slope was not biased towards the older age-groups where there were many more deaths. The multi-step model predicts a slight flattening of the curve in the youngest and oldest age-groups - this occurs because deaths in young age-groups often involve people who have inherited at least one step, and therefore have a high risk but a lower slope – in the oldest age-groups many people will already have accumulated 1-2 steps, so once again the risk is higher but the slope is lower (24). For the multi-step model, we are interested in the overall slope with equal weight on all of the age-groups (essentially it is the middle age-groups which are predicted to provide the slope which is one less than the number of steps). We did the regressions for each country, by sex (data available for Spain and Italy), and we also did combined analyses adjusting for country; these were weighted by the total number of cases in each country's dataset.

RESULTS

Multi-step analyses for Covid-19

For all analyses, we found a linear log-log relationship between the CFR and age, with a small flattening of the curve in the youngest and oldest age-groups (Table 1). All five countries had an overall slope of about 5. Figures 1 and 2 show the findings for the CFR rate by age, and the plot of ln (CFR) against ln (age), for all five countries combined. The country-specific findings are shown in Figures S1-S5 (Supplementary Appendix).

For the two countries where we had data separately for males and females (Italy and Spain), the male death rates were higher at each age-group, but the slope was lower - by 0.7 (Figures 3 and 4).

In one country, China, we had data for both CFR and IFR, and these yielded very similar findings (not shown in table); these used adjusted data (see above), and yielded slopes of 5.5 for the CFR and 5.4 for the IFR (Figures S6 and S7, Supplementary Appendix).

Multi-step analyses for other respiratory virus diseases and total mortality

7

The findings for selected other respiratory virus diseases (Seasonal influenza, Pandemic influenza and SARS) are shown in Figures S8-S10 (Supplementary Appendix). Seasonal influenza (Figure S8) and pandemic influenza (Figure S9) showed age-patterns that were different from each other, and markedly different from that for Covid-19 (Figure 2). Although seasonal influenza generally showed an exponential increase with age, there were also high death rates in the 0-4 age-group (i.e. the overall pattern is J-shaped). Data on mortality in hospitalised influenza cases in England from January 2000 to August 2007 (16), i.e. during non-pandemic years, did not clearly show a log-log pattern, and the fitted slope is only 1.4 (Figure S8). Pandemic influenza (H1N1) showed a very different age-pattern, with peak mortality in adulthood. Modelling data from England (37) on mortality due to 2009 H1N1 influenza pandemic shows a J-shaped pattern mortality in the youngest age-group, with no log-log pattern, and the fitted slope is close to zero (Figure S9). SARS showed a roughly linear log-log relationship (Figure S10), but with a lower slope (3.6) than was observed for Covid-19 (Figure 2). Thus, SARS showed a similar log-log age-pattern to that of Covid-19, albeit with a lower slope; in contrast, seasonal and pandemic influenza showed quite different age-patterns, with little evidence of a log-log relationship, and with substantially lower slopes. The log-log plot of total mortality in the Northwest Italy for the first six months of both 2019 and 2020 shown in Figure S11 indicates that Covid-19 mortality and all-cause mortality both increase markedly with age, but that there are differences in age-patterns (the pattern for total mortality appears to be non-linear – slightly J-shaped – whereas the pattern for Covid-19 mortality is linear with a slight S-shape which is predicted by the model), and in the slope of the log-log trend (total mortality has a slope of about 7, rather than the value of about 5 for Covid-19 mortality).

DISCUSSION

As hypothesized, we found a linear log-log relationship between the Covid-19 CFR and age, with a small flattening of the curve in the youngest and oldest age-groups (as predicted by the multi-step model (24)). All five countries gave similar values for the overall slope (Table 1), i.e. about 5, indicating a multi-step process of about 6 steps. As with other applications of the multi-step model (e.g. cancer), this is a population average, and it may well be that some pathways to the outcome may involve a different number of steps (24). We have also shown a similar log-log age-pattern for SARS, albeit with a lower slope (indicating a smaller number of steps); in contrast, seasonal and pandemic influenza showed markedly different age-

patterns. Thus, death from Covid-19 and SARS appears to follow a distinct age-pattern, consistent with a multi-step model of disease.

A number of limitations of the data should be recognized. Firstly, the appropriate denominator depends on what the hypothesis is. For cancer, where the multi-step model has been most commonly used, the hypothesis is that there are steps involved in developing cancer, so the outcome/numerator is incidence and the denominator is total population; most risk factors for cancer that follow the multi-step model (e.g. a genetic mutation) affect incidence and usually not survival (24). In the current context, the outcome is Covid-19 death, and the hypothesis relates to mortality in those who become infected. However, in most cases, only case fatality rates (CFRs) were available, and the denominator was diagnosed cases rather than all infections. A related problem is selective case identification - most cases are identified by testing symptomatic people, and non-symptomatic cases are largely missed, and even in those tested there will be false positive and false negative results (this is also a problem with other respiratory diseases such as pandemic influenza (30)). Given these uncertainties, the available data that we have used for these analyses is likely to be subject to inaccuracy and misclassification (of outcome, denominator, or both). However, these uncertainties will not bias the analysis unless they operate with different strengths at different age-groups. For example, if the problem is just that the CRF is double the IFR at each age-group greater than 30 years (because at each age-group, only one-half of the infections are tested and diagnosed), this will affect the absolute value of the CFR but will not affect the slope of the log-log linear trend. Furthermore, it is noteworthy that we have obtained very similar findings for the CFR in several different populations and contexts, with differing methods for ascertaining cases.

A key difference with the application of the multi-step model in cancer and in Covid-19 should also be noted. The model of cancer involves changes taking place in a single cell. Clearly, this does not apply to Covid-19. Rather, we think of individuals in a population and the different 'steps' that make them more susceptible for a severe course of Covid-19 (much like the 'promotor' steps in the cancer model). In the cancer model, there are some steps which apply to all cells, e.g. when one is born with a genetic mutation. The same may apply to Covid-19, i.e. the steps involved are likely to be systemic for an individual and can be genetic. It is also possible that comorbidities, or other changes associated with the aging, may lead to increased systemic inflammation, and an exaggerated immune response. Once again, these would be systemic changes in an individual rather than changes in a single cell.

If we can assume that the findings presented here are valid, given all the uncertainties of the data, what do they mean? The multi-step model has been used for cancer since 1954, but with a few exceptions, the relevant stages have not yet been identified. Those that have been identified typically involve mutations in DNA or other cellular changes which lead to cell proliferation. The model has produced a number of testable hypotheses, e.g. relating to the stage (step) at which various carcinogens act (38), the dose-response relationships for factors such as smoking, and the changes in the age-incidence patterns after smoking cessation (24). The application of the multi-step model to ALS (25) is beginning to yield similar benefits, including clarifying the role of genetic susceptibility which increases the disease rate at each age-group but produces a lower slope compared with persons without such susceptibility (39). Thus, recognizing diseases that show a multi-step pattern provides the foundations for identifying the steps and being able to intervene across the life-course; this includes not only attempting to prevent specific steps from happening, but also taking early therapeutic action pointed at an existing step in those who are infected. Importantly, for Covid-19, it illustrates that identification of final steps in (elderly) patients could prevent severe disease and death.

Given that SARS-Cov-2 is a newly discovered virus, it is not surprising that the stages involved in Covid-19 mortality are not currently readily identifiable. Given the rapidity of Covid-19 death following infection, it is likely that SARS-Cov-2 itself acts at a late stage of the multi-step process, with the previous stages having occurred before infection, throughout the life-course. Although it is not yet clear what these steps may be, they clearly involve changes associated with chronic conditions such as hypertension, diabetes, cardiovascular disease, severe chronic respiratory disease and cancer. They also involve other cumulative changes associated with aging (e.g. one change may be the nasal expression of ACE-2 receptors with aging (40)), since age itself is a risk factor for Covid-19 death. Collectively, these comorbidities, and age itself, appear to be markers of immune-related susceptibility, which makes the difference between experiencing a mild illness, or a severe illness possibly resulting in death. However, we and others would argue that age is not a process in and by itself, but rather that age is a surrogate variable for the accumulation of several 'steps' that leave permanent damage. What is involved in the specific steps in the model, and how they effectively 'speed up' the aging process, remains unclear. For example, why does a 60-year-old with diabetes have a similar Covid-19 mortality risk (if infected) as a 63 year old without diabetes (41)? Perhaps having obesity involves 'insults' such as systemic inflammation, and may prime someone to be more susceptible to the systemic inflammation and cytokine storms

that can occur with Covid-19. If one is old enough, then this same 'insult' may occur for other 'unknown' reasons, which is why such risks increase with age.

Does having diabetes or a high BMI, for example, involve a specific biological change (analogous to those observed for cancer and ALS) which is an intrinsic part of the process leading to Covid-19 mortality? If this is the case, are these conditions clearly-defined 'steps' which can potentially be identified? If so, one would expect that people who have one of these conditions (e.g. diabetes), would have a higher CFR for Covid-19 death at each age-group, but a lower slope – this is a hypothesis that can be tested in future as more data with sufficient numbers of the underlying conditions emerge. On the other hand, are these conditions, like age, just markers of more general increases in immune-related susceptibility which accumulate across the life-course? In which case there must be other steps that 'prime' someone for severe disease when they are infected.

The differences between males and females are intriguing, and mechanisms explaining excess Covid-19 mortality in males are unknown (21). It has been suggested that this may be from an inherent characteristic of being male (42). In multi-step model terminology, this would mean that males were born with one step already in place, analogous to being born with a particular genetic mutation which accounts for one step of a multi-step cancer process. In this scenario, males would have higher death rates at each age-group, but their overall age slope would be one lower, since one step had already been acquired at birth. We found that the male CFR was higher than that in females at each age-group, but that the slope of the regression was 0.7 lower in males. This is close to the difference of 1.0, which was the a priori hypothesis. We propose that this step may be related to a relative deficiency in innate immune responses to viruses in males compared to females. Recent analysis in a population-based birth cohort has shown that interferon (IFN)-α, -β and -γ responses to three common respiratory viruses and two viral mimics are deficient in males compared to females, indicating that the excess Covid-19 mortality in males is likely at least in part explained by impaired innate anti-viral immune responses in males compared to females (Prof S Johnston, personal communication). Endosomal-expressed Toll-like receptors (TLRs) 7/8/9 are engaged by positive strand RNA viruses such SARS-CoV-2, which require endosomal processing as part of viral entry into cells (43). Their activation results in production of type I IFNs, which is an important step for the induction of antiviral immunity. TLRs 7 and 8 are encoded by loci on X chromosome locus, and biallelic expression of X-linked genes could enhance TLR7-8 expression in female immune cells (44), thereby providing a mechanistic

explanation for our observation. We propose that the notion that biallelic expression of X-linked TLR7-8 is unlikely to be strictly binary, and will not occur in every female, could explain 0.7 (rather than 1) difference.

Our findings should be regarded as preliminary, and require further replication in other populations. Moreover, their etiological significance is not yet clear, though specific hypotheses described above, such as whether diabetes is likely one of the steps, can be tested as more data emerges. Nevertheless, the patterns are strikingly consistent across the countries studied. These findings are consistent with a multi-step model of disease involving a six-step process that in the case of SARS-COV-2 is probably defined by comorbidities and age producing immune-related susceptibility. It is also striking that we see a similar pattern with SARS (another Coronavirus), perhaps indicating a similar mechanism for mortality, whereas we see different age-patterns for other severe respiratory infections, perhaps indicating that different mechanisms are operating. The clear log-log patterns of the CFR rate with age that we have identified indicate that there is a potential to identify population subgroups which are 'primed' for severe disease if they experience a SARS-CoV-2 infection. Identification of these steps would be potentially important for prevention, therapy, and priority for vaccination for SARS-COV-2 infection.

Conflicts of interest

D.A.L reports support from Medtronic Ltd and Roche Diagnostics for biomarker research unrelated to this publication.

AC reports personal fees from Novartis, personal fees from Thermo Fisher Scientific, personal fees from Philips, personal fees from Sanofi, personal fees from Stallergenes Greer, outside the submitted work.

Availability of data and material (data transparency)

This is publicly available data, and was analysed with Excel, so there is no code; original data is available from the authors.

References

1.   Mizgerd JP. Lung infection--a public health priority. PLoS Med. 2006;3(2):e76.
2.   Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-9.
3.   Coperchini F, Chiovato L, Croce L, Magri F, Rotondi M. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. Cytokine Growth Factor Rev. 2020.
4.   Kunitz SJ. The health of populations: general theories and particular realities. New York: Oxford University press; 2007.
5.   Pearce N, Davey Smith G. Is social capital the key to inequalities in health? American Journal of Public Health. 2003;93(1):122-9.
6.   Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382(18):1708-20.
7.   Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med. 2020;26(4):506-10.
8.   Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507-13.
9.   Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ. 2020;369:m1985.
10.   Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(15):458-64.
11.   Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv. 2020:doi.org/10.1101/2020.03.03.20028423.
12.   ONS. Coronavirus (COVID-19) Infection Survey pilot: England, 14 May 2020. London: Office of National Statistics; 2020 14 May 2020.
13.   Castagnoli R, Votto M, Licari A, Brambilla I, Bruno R, Perlini S. Severe acute respiratory syndrome Coronavirus 2 (SARS-COV-2) infection in children and adolescents: a systematic review. JAMA Pediatrics. 2020:in press.
14.   Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390:946-58.
15.   Chowell G, Bertozzi SM, Colchero MA, Lopez-Gatell H, Alpuche-Aranda C, Hernandez M, et al. Severe respiratory disease concurrent with the circulation of H1N1 influenza. New England Journal of Medicine. 2009;361:674-97.
16.   Cromer D, van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: a statistical analysis to inform vaccine policy. Journal of Infection. 2014;68:363-71.
17.   Gagnon A, Miller MS, Hallman SA, Bourbeau R, Herring DA, Earn DJD, et al. Age-specific mortality during the 1918 influenza pandemic: unreavelling the mystery of high young adult mortaloty. PLoS One. 2013;8:e69586.
18.   Reichert T, Chowell G, McCullers JA. The age distribution of mortality due to influenza: pandemic and perr-pandemic. BMC Medicine. 2012;10:162.
19.   Woo G. Age-dependence of the 1918 pandemic. British Actuarial Journal. 2019;24:1-16.

20.   Monsalvo AC, Batalle JP, Lopez MF, Krause JC, Klemenc J, Zea J, et al. Severe pandemic 2009 H1N1 influenza disease due to pathogenic immune complexes. Nature Medicine. 2011;17:195-9.

21.   Wenham C, Smith J, Morgan R, Gender, Group C-W. COVID-19: the gendered impacts of the outbreak. Lancet. 2020;395(10227):846-8.

22.   Peto R. Epidemiology, multistage models, and short-term mutagenicity tests. Int J Epidemiol. 2016;45(3):621-37.

23.   Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. British journal of cancer. 1954;8(1):1-12.

24.   Frank SA. Dynamics of cancer. Princeton, NJ: Princetone University Press; 2007.

25.   Al-Chalabi A, Calvo A, Chia A, Colville S, Ellis CM, Hardiman O, et al. Analysis of amyotrophic lateral sclerosis as a multi-step procss: a population-based modelling study. Lancet Neurology. 2014;13:1108-13.

26.   Jin J-M, Bai P, He W, Wu F, Liu X-F, Han D-M, et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. Public Health 2020:doi.org/10.3389/fpubh.2020.00152.

27.   Pearce N, Vandenbroucke JP, VanderWeele T, Greenland S. Accurate statistics on Covid-19 are essential for policy guidance and decisions. American Journal of Public Health. 2020:in press.

28.   Battegay M, Kuehl R, Tschudin-Sutter S, Hirsch HH, Widmer A, Neher RA. 2019-Novel Coronavirus (2019-nCoV): estimating the case fatality rate - a word of caution. Swiss Medical Weekly. 2020;150:w20203.

29.   Ruan S. Likelihood of survival of coronavirus disease 2019. Lancet Infectious Diseases. 2020.

30.   Nishiura H. Case fatality rate of pandemic influenza. Lanet Infectious Diseases. 2010;10:443-4.

31.   Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infectious Diseases. 2020:in press.

32.   WHO. Report of the WHO-China Joint Mission on coronavirus disaese 2019 (COVID-19). Feb 28, 2020. 2020.

33.   Korean Centre for Disease Control. Updates on COVID-19 in Republic of Korea,17 March 2020. Seoul, South Korea: Korean Centre for Disease Control; 2020.

34.   ISS TfC-dDMIeSdI, Istituto Superiore di Sanità. Epidemia COVID-19, Aggiornamento nazionale. Dipartimento Malattie Infettive e Servizio di Informatica, Istituto Superiore di Sanità; 2020 16 April 2020.

35.   Sanitarias CdCdAyE. Actualización nº 109. Enfermedad por el coronavirus (COVID-19). 18.05.2020. Madrid: Centro de Coordinación de Alertas y Emergencias Sanitarias; 2020.

36.   Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L. Estimating excess 1- year mortality from COVID-19 according to underlying conditions and age in England: a rapid analysis using NHS health records in 3.8 million adult. MedRxiv. 2020:doi.org/10.1101/2020.03.22.20040287.

37.   Mytton OT, Rutter D, Make M, Mak M, Stanton EAI, Sachedina N, et al. Mortality due to pandemic (H1N1) 2009 influenza in England: a comparison of the first and second waves. 140. 2012:1533-41.

38.   Pearce N. Multistage modelling of lung cancer mortality in asbestos textile workers. International Journal of Epidemiology. 1988;17(4):747-52.

39.   Chio A, Mazzini L, D'Alfonso S, Corrado L, Canosa A, Moglia C, et al. The multi-step hypothesis of ALS revisited: the role of genetic mutations. Neurology. 2018;91:e635-e42.

40.   Bunyavanich S, Do A, Vicencio A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. Journal of the American Medical Association. 2020:doi:10.1001/jama.2020.8707/.

41.   [Available from: https://alama.org.uk/covid-19-medical-risk-assessment.

42.   Sharma G, Santos Volgman A, Michos ED. Sex Differences in Mortality from COVID-19 Pandemic: Are Men Vulnerable and Women Protected? JACC Case Reports. 2020:doi.org/10.1016/j.jaccas.2020.04.027.

43.   Yang N, Shen HM. Targeting the Endocytic Pathway and Autophagy Process as a Novel Therapeutic Strategy in COVID-19. Int J Biol Sci. 2020;16(10):1724-31.

44.   Souyris M, Cenac C, Azar P, Daviaud D, Canivet A, Grunenwald S, et al. TLR7 escapes X chromosome inactivation in immune cells. Sci Immunol. 2018;3(19).

Table 1: Slope of ln (death rate) versus ln (age) for the five countries, for CFR, for the total population and separately for females and males where available

Country	N of cases	Total	95% CI	Females	95% CI	Males	95% CI
China	44,672	4.88	3.96-5.80				
South Korea	8,320	5.66	3.01-8.31				
Japan	15,300	5.77	4.75-6.79				
Spain	239,125	4.94	3.95-5.93	5.37	4.25-6.50	4.63	3.80-5.46
Italy	159,107	4.93	3.70-6.15	5.43	4.00-6.86	4.76	3.90-5.61
Total	**466,524**	**4.94**	**4.51-5.37**	**5.40**	**4.67-6.12**	**4.68**	**4.19-5.17**

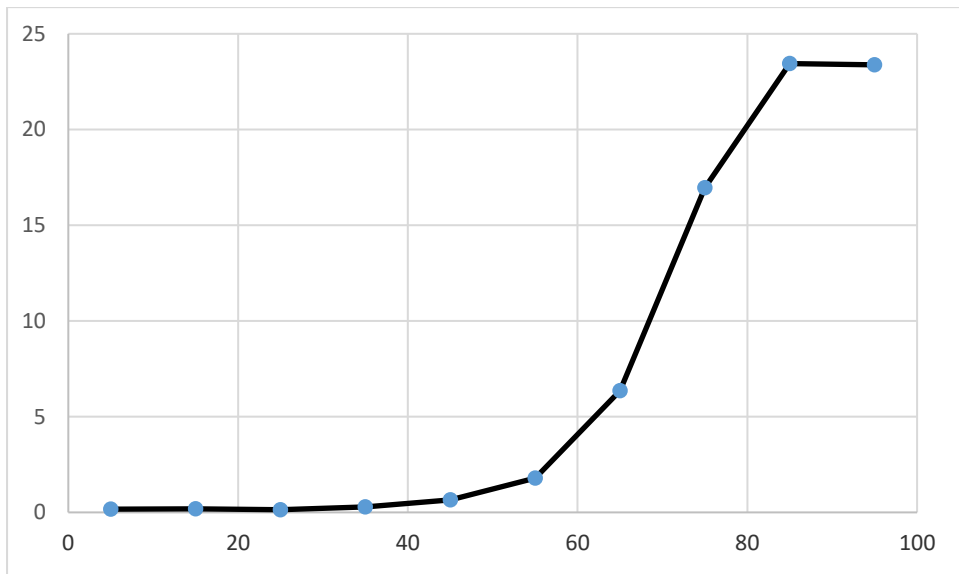**Figure 1: CFR by age for the four countries combined**

Figure 2: Ln (CFR) versus ln (age) for the four countries combined
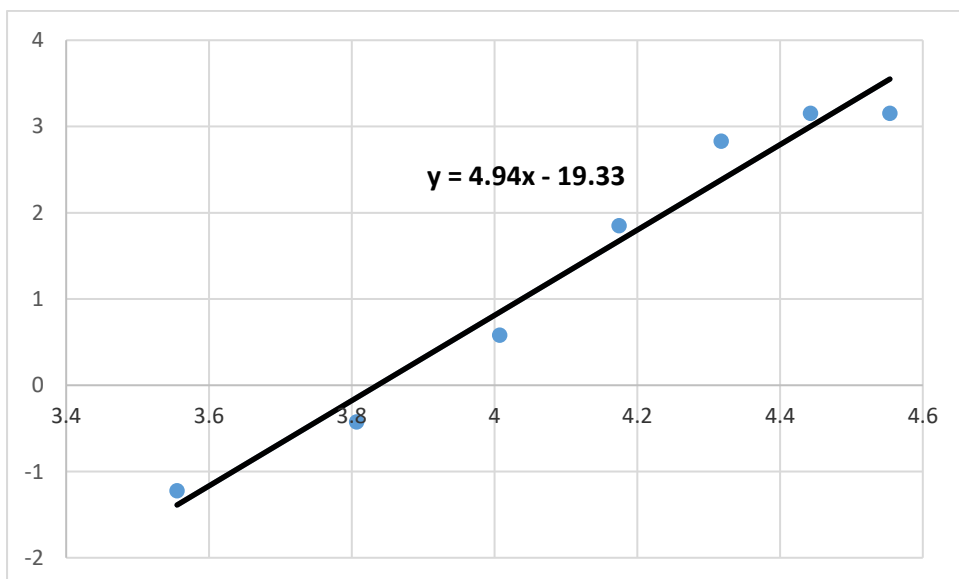

y = 4.94x - 19.33

Figure 3: CFR by age for two countries combined (Italy and Spain), by sex
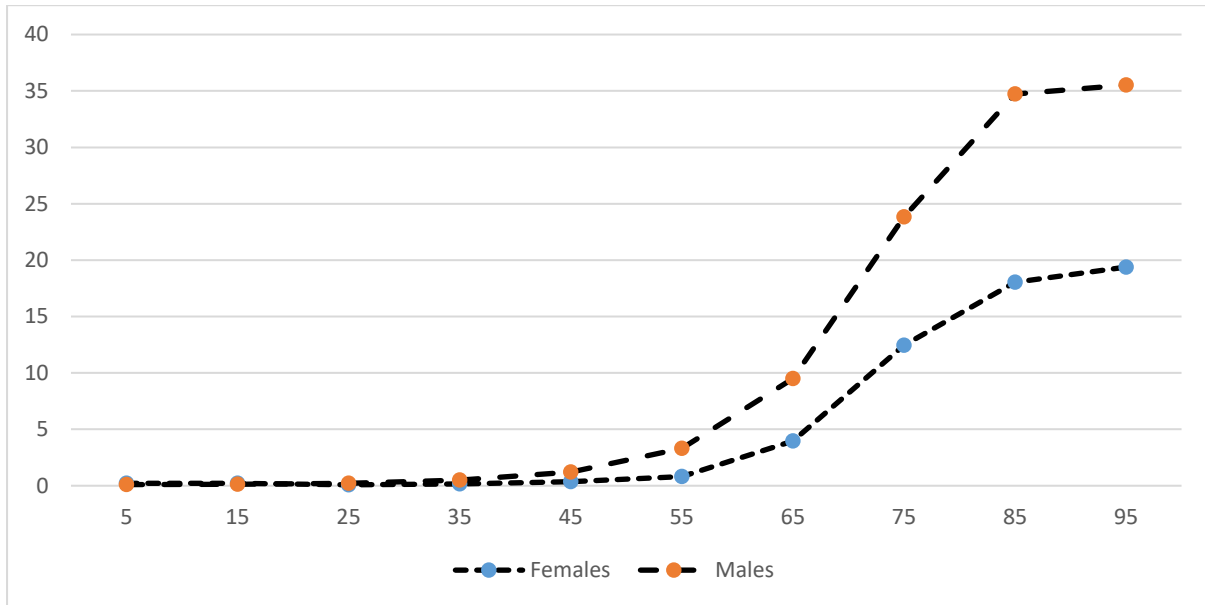

Figure 4. Ln (CFR) versus ln (age) for two countries combined (Italy and Spain), by sex

$y = 4.70x - 17.49$ (males)

$y = 5.40x - 21.31$ (females)