# Comparing successive COVID-19 waves within and between countries: A challenge when dealing with imperfect data

Catalina Torres[1,2], France Meslé[1], Magali Barbieri[1,3], Florian Bonnet[1], Carlo Giovanni Camarda[1], Emmanuelle Cambois[1], Arianna Caporali[1], Étienne Couppié[1], Jenny Garcia[1], Iris Hourani[1], Svitlana Poniakina[1], Jean-Marie Robine[1,4]

1.Institut national d'études démographiques (INED), France ; 2. Muséum National d'histoire naturelle, France ; 3. University of California, Berkeley, USA ; 4. INSERM, France.

**Abstract:** The COVID-19 pandemic is characterized by a succession of waves which developed differently over time and space. Comparing the demographic characteristics of people who died during these different waves across various countries could provide valuable insights into the efficiency of the health measures implemented by the governments. However, such comparisons are not so straightforward given the heterogeneity of available data between countries/data sources. In a recently published paper, we discussed a number of issues to take into consideration when carrying out international comparisons of COVID-19 mortality. These issues were illustrated with concrete examples, using data from the "Demography of COVID-19 deaths" database for the first phase of the epidemic. The purpose of the present paper is to extend those analyses to the second and third waves, especially in those countries with comparable death counts. Further analyses will answer the following questions: How did mortality change from one wave to the next? Were age patterns modified? Was there convergence between male and female mortality? For which ages? Can these changes be explained by health policies and, for the most recent, vaccination ? Some preliminary results are included in this extended abstract.

The COVID-19 pandemic is characterized by a succession of waves with different characteristics over time and space. Comparing the demographic profiles of people who died during these different waves across various countries provides valuable insights into the efficiency of health measures implemented by the governments (from basic prevention, such as regular hand washing, face masks and lockdown, to testing and immunization). However, such comparisons are not straightforward given the heterogeneity of the data available across countries. Although many countries have published series of death counts attributable to COVID-19 to monitor the pandemic, these data are far from being directly comparable across space and time; thus, comparative analyses must be conducted cautiously. The collection of data carried out through specific observation and registration systems (e.g. the civil registry and epidemiologic surveillance systems) potentially leads to differences in data coverage and representativeness between sources, even in the same country. Such variations introduce biases when carrying out undocumented international comparisons.

We collected COVID-19 death counts by age and sex for 19 countries in Europe, Asia and North America. Data are available in the "Demography of COVID-19 deaths" database (https://dc-covid.site.ined.fr/en/). The data were extracted from official statistics provided and updated periodically in each country (e.g. from national public health institutes or national civil registration systems). We endeavor to qualify these data, by highlighting the critical points to consider for international comparisons. In the metadata collected for each country, we focus on three main items:

data definitions (e.g., testing strategies, case confirmation mechanism, and the consideration of "probable cases" and their definition), data collection (e.g., type of system, coverage by place of death, verification, reporting time lag), and data publication (e.g., date of reference, periodicity). With concrete examples, we illustrate how differences in these three components can lead to biased conclusions regarding COVID-19 mortality, drawing attention to the data specificities, differences and shortcomings to be avoided.

In a recently published paper (Garcia et al., 2021)[1], critical points for international comparisons of COVID-19 mortality were listed and illustrated. These points will be revisited in the present study, taking into consideration the most recent data. The most important ones are presented here below.

First, the definition of deaths attributable to COVID-19 differs from one source to the other, even within the same country. The cause of death can be attributed to COVID-19 as a result of various biological tests, clinical diagnosis, imaging test or if the infection is mentioned on the death certificate. These differences are important because they define whether the COVID-19 deaths that are counted in the statistics correspond to individuals who died due to COVID-19 or to people who were infected but died from another cause (i.e. died with COVID-19). Statistics relying on death certificates – where the underlying cause of death is indicated – are often available with some delay, in comparison with data issued from surveillance systems. In the second half of 2021, however, more and more data from civil registration (death certificates) will become available as national statistics offices publish the totality of the death records for 2020 and for the first semester of 2021.

Second, data coverage is often difficult to assess. Some sources are explicitly limited to specific places of death (e.g. hospitals, nursing homes). In other cases, the published statistics cover, in principle, all places of death; however, the degree of reporting completeness by place of death is not specified. In particular, the distribution of deaths by age and sex may be available for a fraction of the total deaths only, but this fraction is often not representative of the total.

Third, depending on the country/source, deaths may be reported by date of occurrence, date of reporting (i.e., when a death is declared to the administrative or health authorities, locally or centrally) or date of publication. The different types of date to which the data can refer to are important to consider when making international comparisons, given the time lag between occurrence and publication.

Garcia et al. (2021) classified the countries/sources included in the "Demography of COVID-19 deaths" database in three groups, according to the definition of COVID-19 used as well as the degree of data coverage. The authors hold that comparisons should only be carried out for countries within the same group. In the aforementioned study, analyses on COVID-19 mortality were only performed for countries/sources belonging to the group with "comprehensive" death counts, i.e. "countries [and sources] whose data include statistics from the vital registration system, where COVID-19 is mentioned on the death certificate, or surveillance systems or health agencies that report both laboratory-confirmed and suspected COVID-19 deaths" (p. 49). These countries are England and Wales (Office for National Statistics, ONS), Scotland (National Records of Scotland, NRS), Belgium (Sciensano), France (CépiDc), and the United States (National Center for Health Statistics, NCHS). We showed in

[1] Garcia, J., Torres, C., Barbieri, M., Camarda, C., Cambois, E., Meslé, F., Poniakina, S. & Robine, J. (2021). Differences in COVID-19 mortality: Implications of imperfect and diverse data collection systems. *Population*, 2021/0, URL: https://www.cairn-int.info/journal-population-2021-0-page-35.htm

that paper the impact of the epidemic on the age-standardized death rates from COVID-19 in these countries (Figure 1) and identified differences in the age patterns (Figure 2). Those analyses focused on the first wave of the pandemic (until July 2020).

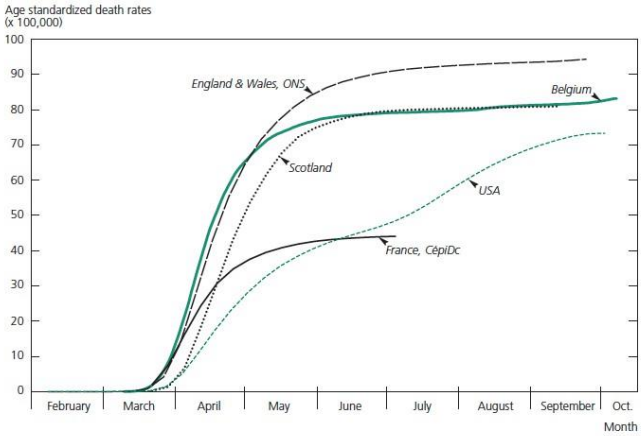


**Figure 1.** Cumulative age-standardized COVID-19 death rates by country (Source: Garcia et al., 2021)

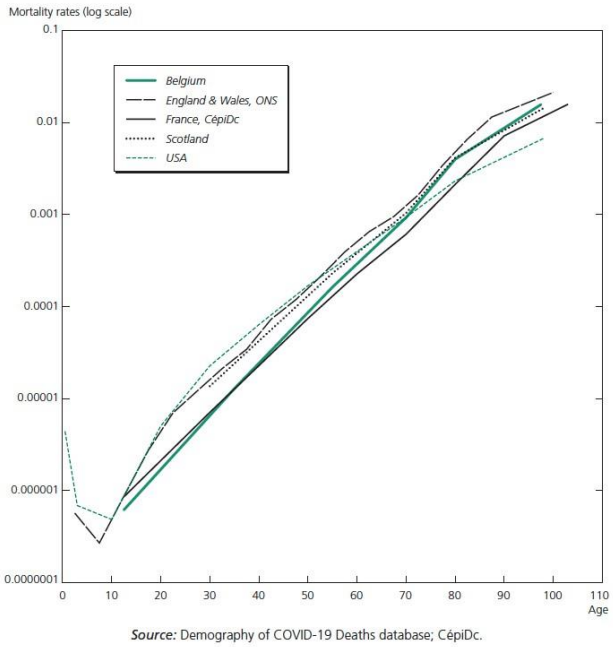


**Figure 2.** COVID-19 mortality age pattern by country on July 15, 2020 (Source: Garcia et al., 2021)

The purpose of the present paper is to extend the analysis to the second and third waves in the countries belonging to the comprehensive group. We will examine whether this group can be extended to include additional countries/sources where data coverage has improved. Next autumn, mortality data will become available for the period up to July 2021. These data will be used to update the "Demography of COVID-19 deaths" database, together with the extensive documentation (metadata) also available on the website. The documentation pays particular attention to changes in definitions and other critical aspects that may have an impact on the observed data. Further analyses extended to the second and third waves should answer questions such as : How did mortality change from one

wave to the next? Were age patterns modified? Was there convergence between male and female mortality? At which ages? Are these changes related to improvements in data collection and coverage (e.g. increased testing) or could they be explained by health policies, and for the most recent period by immunization policies?

Preliminary results comparing the first and the second waves show some changes in the age pattern of mortality (Figure 3). In particular, while the age structure in England and Wales was very close to the one observed in the USA up to age 70 during the first wave, mortality at these ages was less pronounced in England and Wales than in the USA during the second wave and much closer to Belgium. Furthermore, in the USA, mortality at older ages (much lower than in the other countries during the first semester of 2020), reached the same level as in other countries during the second semester. These trends must be further analyzed to identify their driving forces.

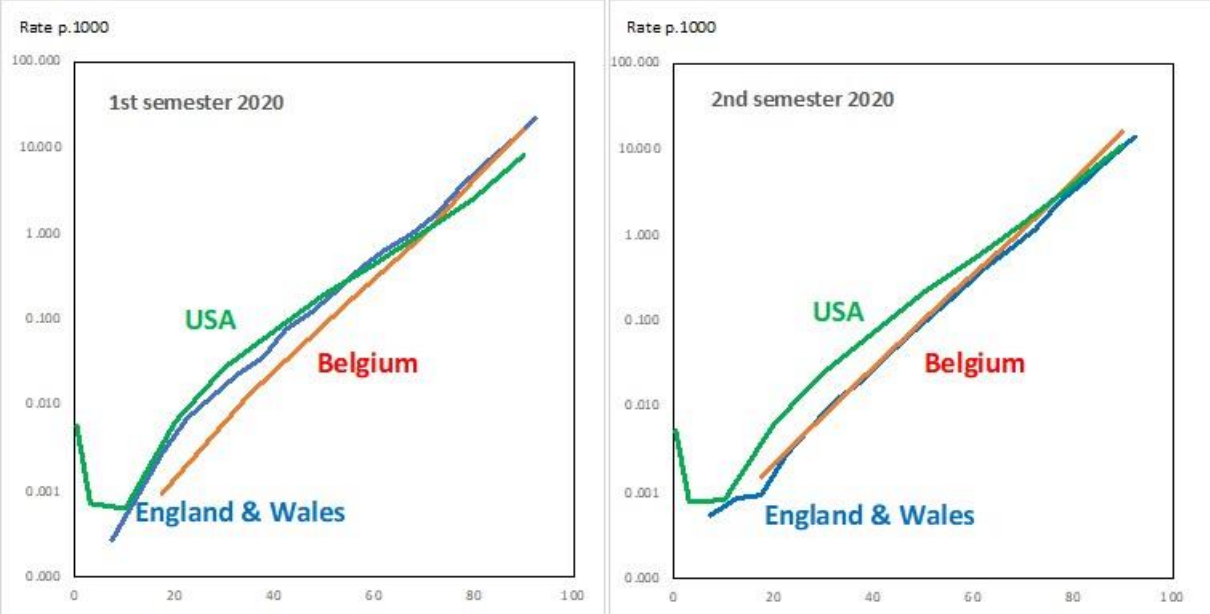


**Figure 3.** COVID-19 mortality patterns by age in the US, England and Wales and Belgium. First and second semester of 2020

Source : The Demography of COVID-19 Deaths (2021). National Institute for Demographic Studies (INED) (distributor). https://dc-covid.site.ined.fr/en/ [accessed 03/05/2021]

The age pattern of COVID-19 mortality also changed over time. Figure 4 displays the distribution of deaths from COVID-19 by age in three months (April 2020, November 2020, and March 2021), corresponding to the peak of the three epidemic waves in the observed countries. In April 2020, 84 % of deaths were occurring at ages 70 and over in England and Wales and 81% in France. One year later, this percentage has decreased to 72 % in the former but increased to 84 % in the latter (Figure 4). The difference between the two countries may be due to the timing of the immunization campaign which started earlier and developed much more rapidly in England and Wales than in France.

In the proposed paper, these very preliminary results will be extended in time and space. Comparative analyses on mortality by age and sex will be systematically conducted on groups of countries for which data comparability will have been ascertained based on the available metadata, paving the way to

substantive explanations for the differences found across countries and providing new evidence to improve policies to control the COVID-19 epidemic in particular and other contagious infections in general.

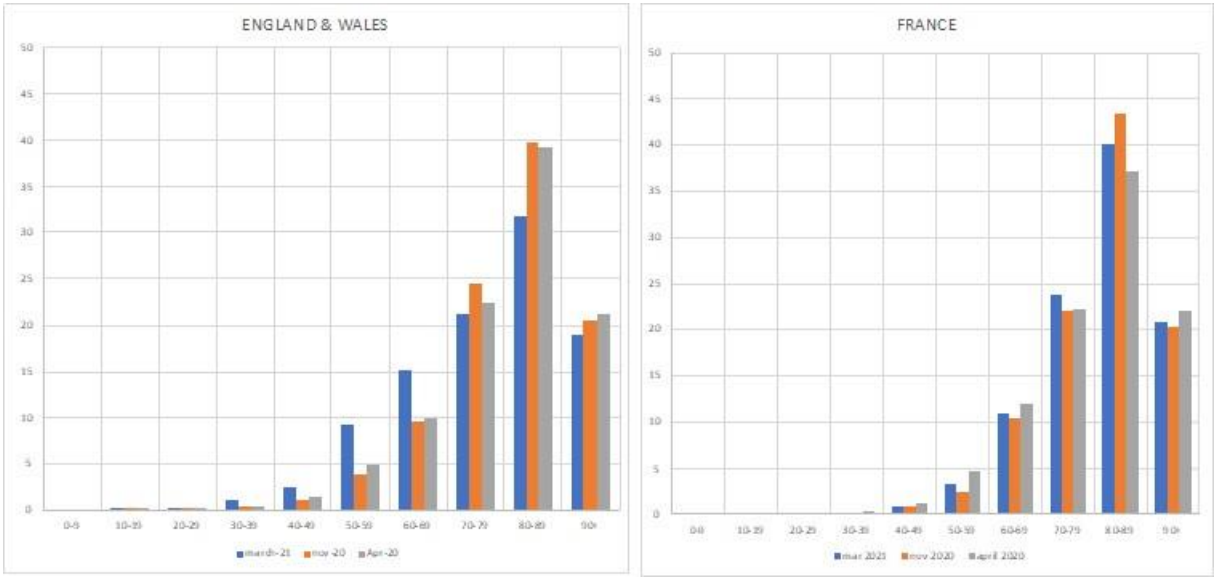


**Figure 4.** Age distribution of COVID-19 deaths in England and Wales and in France in April 2020, November 2020 and March 2021

Source : The Demography of COVID-19 Deaths (2021). National Institute for Demographic Studies (INED) (distributor). https://dc-covid.site.ined.fr/en/ [accessed 03/05/2021]