



HAL
open science

The Demography of COVID-19 Deaths. A gateway to well-documented international data

Arianna Caporali, Etienne Couppié, Jenny Garcia, Iris Hourani, Svitlana Poniakina, Magali Barbieri, Florian Bonnet, Carlo Giovanni Camarda, Emmanuelle Cambois, Daria Korotkova, et al.

► **To cite this version:**

Arianna Caporali, Etienne Couppié, Jenny Garcia, Iris Hourani, Svitlana Poniakina, et al.. The Demography of COVID-19 Deaths. A gateway to well-documented international data. XXIXe Congrès international de la population de l'UIESP 2021 (IPC 2021), Dec 2021, Hyderabad (Conférence virtuelle), India. <hal-04606399>

HAL Id: hal-04606399

<https://hal.science/hal-04606399v1>

Submitted on 10 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

The Demography of COVID-19 Deaths

A Gateway to Well-Documented International Data

Arianna Caporali¹, Étienne Couppié¹, Jenny Garcia¹, Iris Hourani¹, Svitlana Poniakina¹, Magali Barbieri^{1,2}, Florian Bonnet¹, Carlo Giovanni Camarda¹, Emmanuelle Cambois¹, Daria Korotkova³, France Meslé¹, Olga Penina⁴, Jean-Marie Robine^{1,5}, Markus Sauerberg⁶, Catalina Torres^{1,7}

1. Institut national d'études démographiques (INED), France ; 2. University of California-Berkeley, USA; 3. Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine; 4. Nicolae Testemitanu State University of Medicine and Pharmacy of the Republic of Moldova; 5. INSERM, France; 6. Vienna Institute of Demography, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, Univ. Vienna); 7. Muséum National d'histoire naturelle, France

Keywords: COVID-19, Mortality, Cross-country comparative analyses

Abstract short

Since the start of the pandemic, “The Demography of COVID-19 Deaths” database (dc-covid.site.ined.fr/en) aims at providing COVID-19 related death counts along with well-documented metadata. The targeted audience is the research community and beyond (e.g., policy-makers, journalists). As of mid-May 2021, the database contains daily cumulative death counts attributable to COVID-19 reported by 19 countries in Europe and beyond, and updated weekly. Data are broken down by sex and age and, when possible, by place and date of occurrence of the death. Data documentation includes detailed descriptions of the sources and points out issues of quality and coverage of the data. Through this information, it is possible to assess data inconsistencies hampering comparability over time and/or across countries. In particular, the database webpages point out seven data-related issues to take into account for accurate cross-country comparisons of the pandemic. Compared to other databases on the same topic, the strength of “The Demography of COVID-19 Deaths” database is to provide high-quality data and metadata through a rigorous process of data collection. The poster describes the approach and content of the database with the purpose of promoting its usage.

Abstract long

1. Objectives

Since the start of the pandemic, the importance of timely, high-quality, publicly accessible COVID-19 related data by age and sex has been essential in order to monitor its impact. Many national authorities and statistical bureaus have made available daily counts of COVID-19 deaths. These may be incomplete since data collection has often been carried out through health surveillance systems or by establishing *ad hoc* collection systems after the start of the epidemic [1]. COVID-19 death counts have been extensively re-circulated, collated, commented on, and compared¹; however, they are usually poorly sourced and documented.

“The Demography of COVID-19 Deaths” database (dc-covid.site.ined.fr/en) [2] focuses on collecting death counts attributable to COVID-19, providing details on the exact definition of COVID-19 death used in each country, the national data collection system, and the processing of the data. The database provides the *cumulative death counts by sex and age* and, when possible, by *places and dates of occurrence of the deaths*. The database specifically aims to offer *tools for assessing data coverage and comparability, over time and across countries* that can be used by the research community and others (policymakers, actuaries, journalists,

¹ The “Demography of COVID-19 Deaths” webpage lists some of the related databases here: [<https://dc-covid.site.ined.fr/en/useful-links/>]. One example is the COVERAGE-DB database [3] or “Our World in Data” [4].

etc.) for accurate trend analysis. Related documentation is collected and published in the database with the following information:

1. The exact source of data and its main features, such as the type of data collection system (e.g., health surveillance system, vital statistics, hospital records, etc.).
2. The degree of completeness of the statistical information, as well as the changes over time in each country (e.g. whether the reported deaths only include those occurring in hospital settings or also those occurring in nursing homes or private residences).
3. The reference date for which deaths are reported, which can be the date of occurrence, the date of registration, or the date when the figures are officially published.
4. The criteria considered in the attribution of a death to COVID-19, e.g., whether data sources only include confirmed cases or also suspected ones.

2. Database area and population coverage

As of mid-May 2021, 19 countries are included in the database, namely Austria, Belgium, Canada, England and Wales, Denmark, France, Germany, Italy, the Netherlands, Norway, Portugal, the Republic of Korea, the Republic of Moldova, Scotland, Spain, Sweden, Switzerland, Ukraine, and the United States of America (Figure 1). Additional countries are forthcoming (e.g., Romania) or being investigated (e.g., Japan).

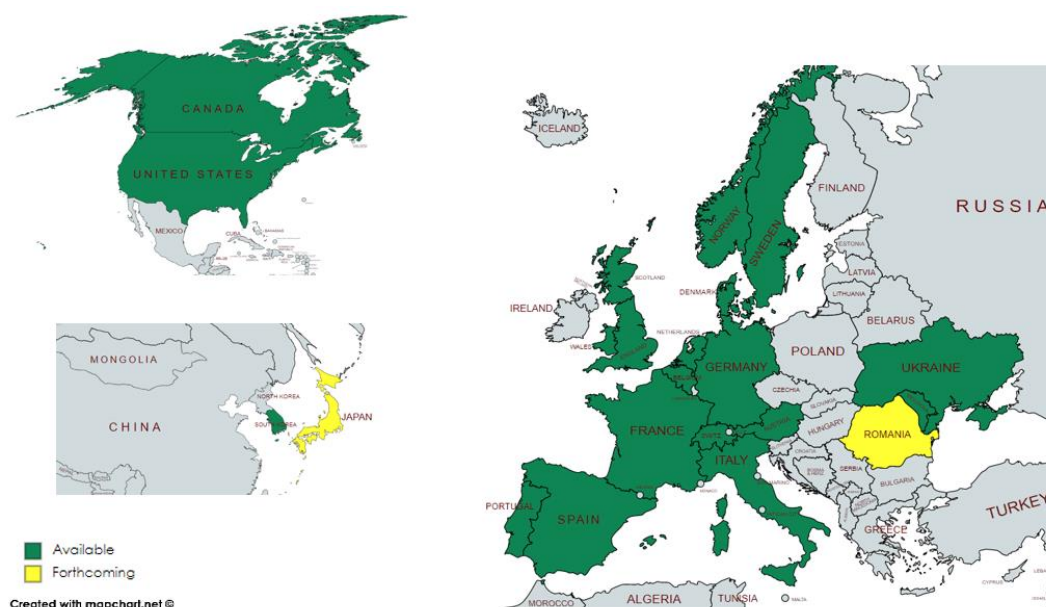


Figure 1: Countries for which data are available or forthcoming as of mid-May 2021

3. Database content and data availability

The “Demography of COVID-19 Deaths” database consists of a collection of webpages. The website is bilingual, available in both English and French. First there are *country-specific webpages* (dc-covid.site.ined.fr/en/data) that contain: 1) a spreadsheet file with the actual data, 2) explanatory notes, 3) all the original documentation (methodological documents and official reports) from which the data and the metadata have been extracted, and 4) the archive with all prior versions of the data files. The country-specific webpages are intended for non-specialist users. They provide data in a user-friendly format along with warnings about data idiosyncrasies that may hamper country comparison.

Second *pooled sets* (dc-covid.site.ined.fr/en/data/pooled-datafiles/) include cumulative death counts for all countries and are meant for expert users. There is a pooled data file with data by sex and age, another one with data by places of death; other pooled files contain data by registration date and by

occurrence date of the deaths. There is also a pooled data file with population estimates and one includes the metadata available for all countries. Users can merge data and metadata pooled files to keep track of changes in the data definitions over time.

Third *the description of seven data-related key issues* [dc-covid.site.ined.fr/en/presentation] highlight the potential uses that may be made of the data, while pointing out the specificities, differences and shortcomings of the data sets. The metadata provide detailed explanations for each data source about:

1. Data coverage (i.e. whether the data include only hospital deaths, all deaths, or deaths which occurred elsewhere, and, if available, a description of the cause-of-death certification process),
2. Data collection systems (i.e., the national protocols for reporting deaths),
3. The information originally available for each death (e.g. sex, age, date and place of occurrence),
4. The time at which the death count is stopped before publication by the organization in charge of disseminating the data.

Users can explore the data availability in the database according to their characteristics through an online application (<https://ineddemographiecovid19.shinyapps.io/DataViz/>, figure 2). This tool shows the heterogeneity of the data across data sources and over time, as to e.g. : width of age groups, data collection systems, data coverage, identification of COVID-19 (suspected or confirmed, cause of death), reference date (e.g., registration, publication, or occurrence date).

Data availability on 'The Demography of COVID-19 deaths' database

This application allows to explore the data available on the website: 'the Demography of COVID-19 deaths' according to the metadata characteristics

Data source: the Demography of COVID-19 deaths (2020). French Institute for Demographic Studies - INED (distributor). Extracted from: <https://dc-covid.site.ined.fr/en/data/pooled-datafiles/>
(last data update: 22/04/2021)



Figure 2: Screen shoot of the database online application allowing to browse the available data and metadata

4. Data preparation and updating

Each member of the database operational team is in charge of a maximum of three countries. They monitor data availability daily and collect all the statistical data and documentation made available by national statistics offices. The country specialists prepare the country data files and the documentation. In-country experts may be contacted to contextualize and interpret the data. In addition, three team members oversee the work carried out by the country specialists before the weekly publication of each country update. They check the accuracy of the daily data against the original data sources and monitor completeness of the documentation. This double-checking ensures correctness and accuracy of the data at the highest level. The

operational team completes the data update procedure by preparing and releasing the pooled data and metadata sets.

5. Database use

Using the data for the first wave of the pandemic available in the “Demography of COVID-19 deaths” database, the database team’s first analysis pointed out the need to take data heterogeneity into account when identifying national characteristics of COVID-19 mortality rates and trends, and especially when doing international analyses [5]. While users can harmonize age-group intervals at their convenience, they cannot do anything about the other sources of heterogeneity in the data (e.g., different definitions of COVID-19 related deaths between data sources). To correctly interpret the pattern of trends and international differences in COVID-19 pandemics, therefore, comparisons should only be carried out between sources that use comparable COVID-19 attribution and confirmation methods.

The data have also been used by international researchers and have already led to articles published in influential journals [e.g., 6]. The database is also monitored by the general public and by journalists, with a mean of 27300 pages viewed per week.

6. Conclusion

The main strengths of “The Demography of COVID-19 Deaths” database is the provision of high-quality data and metadata guaranteed by a rigorous process of data collection, preparation, documentation, and verification, and weekly data updates. The main weakness of the database results from the fact that the stringency of the data collection and metadata review processes limits the number of countries available in the database. Users are invited to contact the database team at eo-dc-covid@listes.ined.fr with their questions or suggestions as to how improve the database.

References

1. INED research group on COVID-19 related deaths (2020). The contribution of demography to the COVID-19 pandemics – 27 March 2020. Available online: https://www.ined.fr/fichier/rte/39/Demo_Covid19_pandemics_Eng.pdf.
2. The Demography of COVID-19 Deaths (2021). National Institute for Demographic Studies (INED) (distributor). Extract from: <https://dc-covid.site.ined.fr/en/> (11/05/2021).
3. Riffe, T., Acosta, E., and COVerAGE-DB project team. (2020). COVerAGE-DB: A database of age-structured COVID-19 cases and deaths”. DOI: <https://doi.org/10.1101/2020.09.18.20197228>. Accessed at : <https://www.medrxiv.org/content/10.1101/2020.09.18.20197228v2.article-info> (11/05/2021).
4. Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodés-Guirao, L., and Roser M. (2020). Coronavirus Pandemic (COVID-19). Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/coronavirus>.
5. García J., Torres C., Barbieri M., Cambois E., Camarda, C.G., Caporali A, Meslé F., Poniakina S., and Robine J.M. (2021). Differences in COVID-19 Mortality: the implications of imperfect and diverse data collection systems, *Population* 2021/0 (Prepublication), p. 35-71. <https://www.cairn-int.info/journal-population-2021-0-page-35.htm>.
6. Bhopal, S. S. and Bhopal, R. (2020). “Sex differential in COVID-19 mortality varies markedly by age”. *The Lancet*, Volume 396, Issue 10250, 532 - 533. 13 August 2020. DOI: [https://doi.org/10.1016/S0140-6736\(20\)31748-7](https://doi.org/10.1016/S0140-6736(20)31748-7).