

Data Management Plan

(see https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf)

General Information

Administrative information

Dr. Diego Alburez-Gutierrez, Principal Investigator (PI); Prof. Dr. Emilio Zagheni
Project: "Family bereavement due to Covid-19 and its consequences for society: a new approach combining macro-level demographic estimations and micro-level analysis"
Funding scheme: Volkswagen Stiftung's "Corona Crisis and Beyond – Perspectives for Science, Scholarship and Society"

1 Data Description And Collection Or Re-Use of Existing Data

How will new data be collected or produced and/or how 1a will existing data be re-used?

This project will use secondary data on excess mortality and will not collect any primary data. The only dataset to include weekly international age-specific excess mortality rates is the "Short-term Mortality Fluctuations" (www.mortality.org), which is curated by experts from the Human Mortality Database project. Other high-quality data sources include the INED Database on Covid-19 Excess Mortality (<https://dc-covid.site.ined.fr/en/>) and the ongoing COVerAGE-DB (https://github.com/timriffe/covid_age) of which the PI is a collaborator. Other potentially useful sources include the Our World in Data COVID-19 dataset (<https://ourworldindata.org/coronavirus>) and The Economist's tracker for Covid-19 excess deaths (<https://github.com/TheEconomist/covid-19-excess-deaths-tracker>).

Data quality is a major concern in the context of an ongoing pandemic. Excess mortality is widely acknowledged to be the best source for studying the mortality effects of the pandemic (at least for the countries that are able to collect and share this information). Known issues include a reporting lag (deaths may be reported days or weeks after their occurrence), under-reporting of deaths (especially in low-income settings), and retrospective adjustments (statistical offices revising their past estimates with new information). These issues are well known to the Principal Investigator, who is currently involved in a separate project to collect data on excess deaths from Covid-19 (https://github.com/timriffe/covid_age).

We will acknowledge the use of these publicly-accessible data sources in our documentation and publications.

What data (for example the kind, formats, and volumes), 1b will be collected or produced?

The project will produce a set of spreadsheets documenting the expected burden of family bereavement by country, age, sex, and type of relative lost.

Our estimates will be provided both in Excel (.xlsx) and Comma-separated-value (.csv) formats.

Providing the estimates in easy-to-navigate Excel format will encourage officials and other non-technical users to engage with the data. Providing the same data in open-access .csv files will be useful for researchers as it can be easily read by any statistical software. Csv file can also be stored in public repositories and are less likely to become obsolete in the future compared to proprietary formats such as .xlsx.

The size of the dataset will depend on the number of countries included in the analysis. For reference, a similar dataset produced by the PI weights 80MB when all world countries are included (https://research-app.shinyapps.io/child_death_paa/)

2 Documentation And Data Quality

What metadata and documentation (for example the methodology of data collection and way of organising 2a data) will accompany the data?

In addition to the metadata accompanying our dataset, we will write a methodological paper stating clearly how the estimates were produced using our novel methodology.

Disclaimer:

The VolkswagenStiftung does not assume any liability for the information published in this Data Management Plan.

| | |
|------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | The metadata describing the dataset of bereavement will follow the guidelines established by the Human Mortality Database protocols (https://www.mortality.org/Public/Docs/MethodsProtocol.pdf), which are a standard in the reporting of mortality data by the demographic community. The code to produce the data will also be shared and commented thoroughly following Google's R style guide recommendations (https://google.github.io/styleguide/Rguide.html). |
| | The project and data will be managed using version control software in the platform www.github.com , which is standard for managing and sharing open-source projects. The GitHub repository will include a Readme file providing a general introduction and sub-directories containing the input data, the code used to produce the output, the output data files, and metadata documents. The PI has extensive experience managing collaborative projects in GitHub (see, for example, this repository created by the PI: https://github.com/alburezg/EGM). |
| | The metadata will detail (1) how the data management was conducted, (2) how excess mortality rates were obtained, (3) how excess mortality rates were translated into probabilities of bereavement, and (4) how these probabilities were used to produce a range of estimates based on different scenarios. The metadata will also include the necessary information on software version, packages used (and their versions) and data sources to allow users to fully reproduce the analysis or update it in the future. |
| 2b What data quality control measures will be used? | We will conduct 'code-reviews' with members of the Lab of Digital and Computational Demography at the Max Planck Institute to ensure that the results are reproducible and minimize the chance of human error. These reviews will also be useful to make sure that the implementation of our novel method is properly documented and accessible to other experts in the field. |

Disclaimer:

The VolkswagenStiftung does not assume any liability for the information published in this Data Management Plan.

3 Storage And Backup During The Research Process

| | |
|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| How will data and metadata be stored and backed up during the research? 3a | All the data and code from this project will be stored locally on the team's computers, online in the version-control platform GitHub and in a centralized secure location. The Max Planck Institute for Demographic Research, where the project will be hosted, has a centralized backup system that performs regular backups of the users profiles and data and stores this information in a secure Cloud, which is physically stored within the institute (in Germany). |
| How will data security and protection of sensitive data be taken care of during the research? 3b | This project will not use or produce any sensitive data. All input data are publicly accessible and all output data will be aggregated and impossible to use to uniquely identify individuals. All team members will have access to the data at all times. This does not pose a risk since all the secondary data that we will use is publicly available. The Max Planck Institute for Demographic Research, a German institution, has established protocols for data protection, which have been updated after the introduction of the General Data Protection Regulation. The project will be discussed the the Data Protection Officer at the institute. |

4 Legal And Ethical Requirements, Codes of Conduct

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| If personal data are processed, how will compliance with legislation on personal data and on security be ensured? 4a | No personal or sensitive data will be collected, used, or produced as part of this study. |
| How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable? 4b | The data and code used to produce the data will be in the public domain under a CC-BY-4.0 license, which permits permits almost any use subject to providing credit and license notice. All participants in the project have agreed to making the data available without restrictions. |
| What ethical issues and codes of conduct are there, and how will they be taken into account? 4c | We will only provide country-level estimates which cannot be used to identify individuals. We will strive for clarity in the resulting literature and in the metadata to reduce the likelihood of the data being misinterpreted or misrepresented. Project members will be subject to the Max Planck Society's "Code of Good Conduct" and "Rules of Good Scientific Practice". The Society has clear protocols to deal with academic misconduct: "Rules of Procedure in Cases of Suspected Scientific Misconduct" (https://www.mpg.de/about_us/procedures) |

Disclaimer:

The VolkswagenStiftung does not assume any liability for the information published in this Data Management Plan.

5 Data Sharing And Long-Term Preservation

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?</p> <p>5a</p> | <p>We project three main mechanisms for sharing the data resulting from this study. We expect most users to engage the data through an easy-to-use online dashboard (see, for example, this dashboard created by the PI for another project: https://research-app.shinyapps.io/child_death_paa/) which will be updated in real-time. The same data will be available through the Center for Open Science (see below). Finally, the latest version of the data and code will be available in a public GitHub repository for the more technically-inclined users. The Max Planck Institute for Demographic Research has a well-staffed media department, which can help distribute the data among the international network of demographers and sociologists, health scientists, and practitioners. The same is true for Population Europe, a trans-national network of research institutions currently being hosted by the Max Planck Institute (https://twitter.com/PopulationEU).</p> |
| | <p>The data and metadata will be stores in a project from the Center for Open Science (https://osf.io/) for long-term storage. A physical copy will be stored in the servers at the Max Planck Institute for Demogrpahic Research.</p> |
| | <p>We anticipate a final version of the data to be made available within one year of the start of the project. We envision a short embargo period for preliminary versions of the data if required by the peer-reviewed journal to which we submit a description of the data. We consider it prudent to wait until the end of the peer-review process before making the data publicly available. This will allow us to include reviewers comments and possibly improve our estimation methodology to make sure of the robustness of our results.</p> |
| | <p>After the end of a potential embargo period (see previous point), the data will freely available without restrictions to all interested parties.</p> |
| <p>How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?</p> <p>5b</p> | <p>No data needs to be destroyed given its personal or sensitive nature as this project will rely on publicly available secondary data exclusively.</p> |
| | <p>Indicate how it will be decided what data to keep. Describe the data to be preserved long-term.</p> |
| | <p>Explain the foreseeable research uses (and/or users) for the data.</p> |
| <p>What methods or software tools are needed to access and use data?</p> <p>5c</p> | <p>The database in csv format can be accessed using any standard data management software such as MS Excel, SPSS, Stata, SAS, or R. The analysis will be performed and shared in the R language for statistical programming, which is a widely used (and open-source) software. R offers sophisticated and robust analysis at no cost and can be integrated with the Docker software to create 'containers' that ensure the computational reproducibility of the results in the future (even if the source code of the R language or any dependencies used for the analysis change in the future).</p> |
| | <p>Data will be stored in a special repository from the Center for Open Science, a widely used and free-to-use repository in the social sciences (https://osf.io/).</p> |
| <p>How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?</p> <p>5d</p> | <p>The Center for Open Science offers unique and persistent object identifiers (DOI) for data projects. The PI has experience managing data in this platform (https://osf.io/mpwjq/).</p> |

6 Data Management Responsibilities And Resources

| | |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>6a</p> | <p>Diego Albrez-Gutierrez (PI) will be responsbile for overseeing the data management and production of metadata.</p> <p>Ivan Williams will be responsible for mantaining the metadata documenting the production of the estimates.</p> <p>Data storage, backup, and archiving will be the responsibility of a Research Assistant hired for this proeject. This reserach assistant will also be responsible for consucting tests of data quality.</p> |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Disclaimer:

The VolkswagenStiftung does not assume any liability for the information published in this Data Management Plan.

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?</p> | <p>We do not anticipate costs related to data archiving and preservation, as described in section 5. Permanent access to a server to host the dashboard where the results will be shared is available free-of-charge from the Max Planck Institute for Demographic Research. As such, staff time will be the main expenditure related to data management. In order to account for this, we will clearly identify tasks related to data management in the job description of the Research Assistant who will be directly involved with producing the empirical estimates and the final data products that will be shared. Apart from traditional metadata documentation, Diego Alburez, Ivan Williams and Emilio Zagheni will be involved in preparing top-quality academic publications to be submitted to high-impact peer-review journals describing the estimation procedures.</p> |
| | <p>We do not anticipate any additional charges from data repositories as such. Permanent access to a server to host the dashboard where the results will be shared is available free-of-charge from the Max Planck Institute for Demographic Research.</p> |

Disclaimer:

The VolkswagenStiftung does not assume any liability for the information published in this Data Management Plan.