Short-Term Mortality Fluctuations Data series (STMF)

Dmitri Jdanov, Vladimir M. Shkolnikov, Ainhoa Alustiza Galarza with assistance of Carl Boe and Magali Barbieri

This is a preliminary version of the STMF note. The methodology, data quality, and data presentation descriptions will be further elaborated.

Please let us know if you have any suggestions or comments by writing to: hmd@mortality.org.

Contents

Background

The Short-term Mortality Fluctuations (STMF) data series are a new component of the Human Mortality Database (HMD). These series are established to provide data for scientific analysis of all-cause mortality fluctuations by week within each calendar year. The decision to add this new resource to the HMD was triggered by the COVID-19 pandemic of 2019-20. An additional motivation for this HMD extension was increasing importance of short-term or seasonal mortality fluctuations that are driven by temporary hazards such as influenza epidemics, temperature extremes, as well as man-made or natural disasters. The relative importance of short-term excess mortality increases in the context of a general mortality decline. It is important also that these particular problems tend to disproportionally affect vulnerable population groups such as the elderly. The STMF series enable monitoring the fluctuations and aim at facilitating research on avoidable life losses, their driving forces, and related health policies in a comparative perspective by presenting a standardized set of weekly counts of deaths across a range of countries.

Objective and internationally comparable data are crucial to evaluate the political strategies used to address epidemics and other public health crises. Indicators based on disease incidence and fatality as well as on cause-specific mortality are valuable but have important shortcomings that make comparisons across countries and time problematic. In contrast, being able to look at short term fluctuations in all-cause mortality (such as captured by weekly or monthly excess deaths) comprise an important complement to other types of data. Weekly death counts constitute a solid data basis for the most objective and comparable way of assessing the scale of short-term excess mortality across countries and over time.

STMF is an important data resource complimentary to national and international monitoring systems. In the context of the COVID-19 pandemic, an increasing number of countries have published weekly death counts in a timely manner. The EUROSTAT has recently launched a collection of weekly death counts covering many EU countries and NUTS2 regions within these countries. In addition, a research group led by MPIDR research scientists developed the COVerAGE-DB, a database that includes information on COVID-19 cases, deaths, and tests by age and sex.

Data

The STMF provides open and user-friendly access to detailed data on mortality by week, sex, and aggregated age group. The data series contain death counts, death rates, and original input data used to produce these output indicators. The data is supplemented by detailed country-specific documentation including precise references to the original data sources. The data series formats, structure, and methods are described in the subsequent sections of this document. STMF is mainly focused on (but not limited) to the countries included in the main HMD database.

Data quality issues

In general, the STMF is based on the data that are collected for the core HMD. The HMD follows a number of criteria for inclusion of individual countries. In particular, only countries demonstrating high-quality statistical systems are included in the database, i.e. those where the census and vital registration system cover close to 100 percent of the population. Nevertheless, the weekly statistics have their specific features that should be considered.

1) Deaths by date of occurrence are preferred for analysis since artificial fluctuations typically affect death counts by date of registration (because vital registration offices are closed on week-ends and holidays in particular). However, depending on the country, weekly death counts are sometimes provided by date of registration only. In most cases, there is no way to convert the date of registration into the date of occurrence. For each country, the exact type of input data and other important details are indicated in the metadata file.

2) The data for the last weeks of a year as well as data for all weeks of the most recent year might be incomplete due to delayed registration of deaths. Statistical offices may still revise these data backward in the course of the next updates.
3) Deaths and death rates are provided by calendar week starting from Monday, Saturday or Sunday (depending on country-specific standards). Please check the country-specific metadata file for details.
4) Each year in the STMF output files include 52 weeks, and each week contain 7 days, including the first and the last of the year. The week arrangement follows the ISO 8601-2004 guidelines. According to this rule, days of the previous years and, respectively, deaths, are included in this first week (whenever January 1st was a Tuesday, Wednesday or Thursday) or in the last calendar week (whenever December 31 was Thursday, Friday or Saturday). Thus, weeks arranged according to ISO-8601 differ slightly from the respective calendar year and cover a slightly different number of deaths compared to the core HMD or the national statistics. Exceptionally, some years include 53 weeks. That is the case for 1992, 1998, 2004, 2009, 2015 and 2020.

The STMF data are provided without any adjustments for potential data quality problems, such as undercounts. The series is not smoothed either. All known country-specific quality issues are documented in country-specific metadata files.

Updates

The monitoring of weekly data is an important task and should ideally be performed regularly. We are planning to update the STMF on a monthly or bimonthly basis. However, more frequent updates (weekly updates) will be carried out during the COVID-19 pandemic. As data become available from more countries over the next few months, additional data series will be included in the STMF.

Output data

The STMF output includes three files: STMF.xlsx, STMF.csv, and STMFmetadata.pdf. The first two files contain all available data series and provide the same data in two different formats. The third file provides country-specific information on the nature of the initial statistics, the exact sources for the data, and other important details.

Each line (record) in the STMF.xlsx and STMF.csv files have an identical structure. Each record is unique (i.e., there are no duplicated records) and contains all necessary information for a particular week of each year. The order of records is usually chronological. Each line includes the following fields:

1) The first four columns identify the country, year, week, and sex (m – males, f – females, b – both sexes combined)
2) The next six columns (5 to 10) provide weekly death counts by age group (0-14, 15-64, 65-74, 75-84, 85+) and the total (all ages combined)
3) Columns 11 to 15 contain the corresponding age-specific death rates,
4) Column 16 provides the crude death rates for all ages combined.
5) The last three columns (17 to 19) include explanatory indicators:

a. “Split” indicates whether the data were split from aggregated age groups (1/0 for yes/no, respectively). For example, Split equals 1 if the original age scale was 0-4, 5-29, 30-65, 65+, and Split=0 if the original age scale was 0, 1-4, 5-9, …, 90+.
b. “SplitSex” indicates if the original data are available by sex (0) or data was split by sex (1)
c. “Forecast” is equal to 1 for all years where forecasted population exposures were used to calculate the weekly death rates.

The first worksheet of the Excel file (STMF.xlsx) provides a summary of information relative to the data. The following worksheets present country-specific data. Each worksheet contains all of one country-series.

The CSV data file (STMF.csv) is optimal for a use in statistical packages. This is a conventional comma-separated file with all countries’ data provided in one data set. The file has the following heading (with each field matching the description above):
CountryCode,Year,Week,Sex,D0_14,D15_64,D65_74,D75_84,D85p,DTotal,R0_14,R15_64,R65_74,R75_84,R85p,RTotal,Split,SplitSex,Forecast.

Input data

The STMF Input files are country-specific comma-separated files (available for download as a single zip file). The files are named in the following way: XXXstmf.csv, where XXX is the 3-digit ISO country code with additional characters for sub-populations (including 3-7-digits, all uppercase letters, see https://www.mortality.org/Data/DataAvailability for details). Each country-file contains standard headings (first line), which represent the field identifiers. The CRLF ("\r\n") combination of characters is used as a record delimiter and a comma (",") as the field delimiter. Missing values are coded as a single dot ("."). Optionally, input data files may contain spaces to improve text file readability. The spaces serve no other purpose.

Headings:

PopCode, Area, Year, Week, Sex, Age, AgeInterval, Deaths, Type, Access

Each record refers to the death count for one particular week. Each death count (field “*Deaths*”) corresponds to a particular population/country name (field "*PopName*"), geographical area (field "*Area*"), calendar year (field "*Year*"), calendar week (“Week”), sex (field "*Sex*"), age/lower limit of age interval (field "*Age*"), the length of the age interval (field "*AgeInterval*"), the type of the data (field "*Type*"), and the type of data access (field "*Access*").

The formats of the fields are as follows:

1.1) Population name – country code.

1.2) Area (2-digit). This field serves to reflect territorial (or country) coverage.

1.3) Calendar year (4-digit). Year in which the deaths occurred or were registered (depending on field Type).

1.4) Week (2-digit). Week in which the deaths occurred or were registered (depending on field Type).

1.5) Sex (1-char). The character 'm' denotes males, 'f' denotes females, and 'b' indicates both sexes combined.

1.6) Age (on 1-, 2- or 3-digit or 3-char). For age groups, the value is always equal to the lower age limit, and TOT[1] and UNK stand for total (all ages combined) and unknown ages, respectively.

1.7) The length of the age interval (1- or 2-digit or '+' for the open age interval). Typically: 1, 5, 10, or + for the open-ended age interval.

1.8) Deaths (numerical field, no fixed length). The number of weekly deaths.

1.9) Type (1-char). R for deaths by day/week of registration, O for deaths by day/week of occurrence.

1.10) Access (1-char). This field indicates the confidentiality/accessibility of the data ("C" - confidential, "O" – publicly accessible, “U” – unknown).

Methods[2]

The data processing begins from retrieving yearly population exposures and death counts. The standard HMD estimates of population and mortality measures are usually not available for the most recent year(s) and months. Thus, the available HMD estimates had to be complemented by the forecasted projections covering the most recent year(s). We apply the Lee-Carter model[3] to extrapolate annual death rates. This model is fitted using the available HMD data series starting from 2005. These rates are used then to estimate age-specific population exposures and death counts under the assumption of zero migration.

Next, we collect weekly death counts by sex and age. In some cases, these counts are produced from data sets containing individual death records. The original input data are to be re-classified according to the standard aggregate age groups of STMF (0-14, 15-64, 65-74, 75-84, 85+). If the original weekly deaths are available for detailed age groups (e.g. 5-year intervals), the weekly counts for the STMF age groups are computed by summing the corresponding detailed death counts. If the original weekly deaths are classified into age groups broader than in the STMF, we apply proportional splitting. This procedure is applied for each year independently. The annual proportions of narrower age groups within broader age groups for the whole year are used to redistribute weekly death counts by narrower age groups narrower age groups as follows:

$$\hat{D}_y^w(x, x+a) = D_y^w(x, x+b) \cdot \frac{D_y(x, x+a)}{D_y(x, x+b)},$$

1 The total is always taken from the original data. It may be different from the total yielded by summing up age specific counts.

2 This section will be extended.

3 We use the R package “demography” by R.Hyndman (https://cran.r-project.org/web/packages/demography/demography.pdf)

Where *a* and *b* are the upper limits of age intervals (*a*<*b*); $\widehat{D}_y^w(x, x+a)$ denotes the estimated number of deaths in the age interval $(x, x+a]$ for week w of year *y*; $D_y^w(x, x+b)$ is the number of death in (broader) age interval for week w in year *y*; $D_y(x, x+a)$ and $D_y(x, x+b)$ are the corresponding death numbers for the whole year *y.*

A similar splitting procedure is applied when the original data is not split by sex. Sex-specific weekly death counts in each age group are estimated using the observed sex-ratio in the annual total death counts in the same age group:

$$\hat{D}_y^{w,males}(x, x+a) = D_y^{w,total}(x, x+a) \cdot \frac{D_y^{males}(x, x+a)}{D_y^{total}(x, x+a)}.$$

The weekly age-specific death rates are calculated as follows:

$$m_y^w(x, x+a) = \frac{D_y^w(x, x+a)}{E_y\,(x, x+a)/52},$$

where $E_y(x, x+a)$ denotes annual population exposures in age interval $(x, x+a]$ in year *y*.

History of the project and research team

This work began in mid-March 2020, in a unique and uncertain context caused by the Coronavirus disease 2019 (COVID-19) epidemic. The standpoint of Vladimir M. Shkolnikov and Dmitri A. Jdanov to reflect the excess in all-cause mortality related to this virus in different countries around the world lead us to focus on weekly mortality from all causes (short-term mortality) data as the most appropriate data basis to analyze the effects of this kind of events health shocks. Therefore, since March 2020 different members of the Demographic Data laboratory at the Max Planck Institute of Demographic Research are working on collecting international weekly death counts by age and sex, checking and processing the data and developing the database named the Short-term Mortality Fluctuations (STMF).

The STMF was launched in May 2020. After the first phase of development we realized that the STMF ought to be integrated in the HMD: It follows HMD guiding principles (comparability, flexibility, accessibility, reproducibility), it heavily relies on published HMD data series, and might be useful for many HMD users. Thus, the STMF was published on the HMD website.

The STMF, as part of the HMD project, is a joint project of two teams based in the Laboratory of Demographic Data at the Max Planck Institute for Demographic Research (MPIDR) and at the Department of Demography of the University of California, Berkeley (UCB).

Vladimir M. Shkolnikov, Dmitri Jdanov, and Ainhoa Alustiza Galarza (all MPIDR) conceptualized the idea of the STMF, Dmitri Jdanov designed the database. Ainhoa Alustiza Galarza is primarily responsible for data collection and contacts with data providers and STMF users. Magali Barbieri and Carl Boe (both UCB) gave continuous and valuable advice and feedback on data presentation and methods. Carl Boe also designed the web presentation of the STMF. In July 2020, Laszlo Nemeth together with Vladimir

Shkolnikov developed the Visualization Toolkit enabling instant calculation and visual inspection of excess mortality across countries.

Acknowledgment

The publication of the STMF data series has been only possible thanks to the enthusiasm, effort, and the goodwill of a large number of people and institutions. In March 2020 there were, to our knowledge, a few countries publishing freely weekly death counts by age and sex: England and Wales, France, the Netherlands, Portugal, and the USA. Nevertheless, due to the urgency of the circumstances and the demand of different research institutions, most of the European National Statistical Institutes (NSIs), put a lot of effort to develop and publish in record time weekly death series by age groups and sex.

Our gratitude to Eurostat that encouraged and coordinated the European NSIs to develop weekly death statistics, and publishing the special collection of weekly deaths *demomwk*.

Our appreciation to the NIS and their analysts that, despite of the enormous workload, they have been working so hard and diligent and always responding us so kindly to our data requests and questions:

- Australia: The Australian Bureau of Statistics and its analysts Heidi Rummery and Angelina McRae
- Austria: Statistics Austria
- Belgium: STATBEL, the Belgian Statistical Office
- Bulgaria: The Bulgarian National Statistical Institute
- Canada: Statistics Canada
- Chile: *Departamento de Estadísticas de Información De Salud* (DEIS) of the Health Ministry of teh Chilean Government.
- Croatia: Croatian Bureaus of Statistics
- Czech Republic: The Czech Statistical Office, and Anna Altová from the Faculty of Science, Charles University, Prague
- Denmark: Statistics Denmark and Dorthe Larsen
- England and Wales: The Office for National Statistics and Dave Leon, from the London School of Hygiene and Tropical Medicine
- Estonia: Statistics Estonia
- Finland: Statistics Finland and Markus Rapo
- France: INSEE, Institut National de la Statistique et des Études Économiques
- Germany: Destatis, Statistisches Bundesamt and Felix zur Nieden
- Greece: ELSTAT, the Hellenic Statistical Authority and Christou Christos
- Hungary: the Hungarian Central Statistical Office
- Iceland: Statistics Iceland, Brynjólfur Sigurjónsson and Heiðrún Sigurðardóttir
- Israel: Israel Central Bureau of Statistics and David Landau
- Italy: ISTAT, Istituto Nazionale di Statistica
- Latvia: Central Statistical Bureau of Latvia

- Lithuania: Statistics Lithuania and Alina Norgèlaitè
- Luxembourg: STATEC, National Institute of Statistics and Economic Studies of the Grand Duchy of Luxembourg
- Netherlands: Statistics Netherlands
- New Zealand: Statistics New Zealand and the Data Service of the Ministry of Health
- Northern Ireland: The Northern Ireland Statistics and Research Agency (NISRA)
- Norway: Statistics Norway and Anders Sønstebø
- Poland: Statistics Poland
- Portugal: Statistics Portugal (INE) and the eVM or Vigilância eletrónica de Mortalidade em tempo real (Real Time Electronic Mortality Surveillance)
- Republic of Korea: Statistics Korea (KOSTAT) and ByeongJe Park from the Vital Statics Division.
- Russia: International Laboratory for Population and Health (ILPH) at the National Research University Higher School of Economics (HSE University), Evgeny M. Andreev and Sergey Timonin.
- Scotland: National Records of Scotland (NRS)
- Slovenia: Statistical Office of the Republic of Slovenia
- Slovakia: National Statistical Institute of Slovakia and Jan Meszaros of the Demographic Research Centre
- Spain: Instituto Nacional de Estadística
- Switzerland: Swiss Federal Statistical Office, Celine Schmid and Christoph Junker
- Sweden: Statistics Sweden, Tomas Johansson and Ann-Marie Persson
- Taiwan: The Division of Survey, Department of Statistics, Ministry of the Interior, Roc (Taiwan)
- USA: Centers for Disease Control (CDC) and Prevention and the National Center for Health Statistics (NCHS)

We are grateful to Karolin Kubisch and our student assistant Ellen Jahr for their hard work on data collection. Special thanks go also to Timothy Riffe for his direct and indirect support.

We are also very grateful to all the STMF users who give us feedback and comments on this project, especially on the early stages of the database. Thanks to your enthusiasm and contributions we can improve this data series.