Short-term Mortality Fluctuations Dataseries

This is a preliminary version of the STMF. We are still working on the methodology, data quality checks, and data presentation. Please let us know if you have any suggestions or comments.

Contents

Background

Short-term Mortality Fluctuations (STMF) data series is a new part of the Human Mortality Database. These series are established to provide data for scientific analysis of all-cause mortality fluctuations by week within each calendar years. The decision to establish this new data resource under the HMD was primarily in response to the COVID-19 pandemic of 2019-20. An additional motivation for this HMD extension was an increasing importance of short-term or seasonal mortality fluctuations that are driven by temporal risk factors such as seasonal patterns of influenza or temperature extremes. The relative importance of short-term mortality outbreaks increases in the context of low annual levels of mortality. It is also evident that these particular problems affect particular population groups such as elderly, in particular. The STMF series enable interrogation of these fluctuations and aims at facilitating research on human life losses, driving forces, and related health policies.

Objective and internationally comparable data are crucial to determine the effectiveness of different strategies used to address epidemics and other abrupt challenges. Metrics based on disease incidence and fatality as well as on cause-specific mortality 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 provide the most objective and comparable way of assessing the scale of short-term disasters across countries and time.

Data

The STMF provides open and user-friendly access to detailed data on mortality by week, sex, and aggregated age group. The dataset contains death counts, death rates, and original input data used to produce output estimates. In addition, an exhaustive country-specific documentation is provided along with precise references to the original data. The dataset structure methods are described in the subsequent sections of this document.The data collection 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 certain criteria for inclusion of individual countries such as requirements for maintaining high quality statistical systems and having the completeness of registration of vital events close to 100 percent. Nevertheless, the weekly statistics has own specific which should be taken into account.

1) The data on deaths aree provided by the date of registration or date of occurrence. There is no way to convert the date of registration into the date of occurrence. The exact type of the data and further details are indicated in the metadata file.
2) If the data are provided by date of registration, there could be artificial fluctuations in weekly death figures related to special events (e.g. end of the statistical period, public holidays).
3) The data for last weeks of a year as well as data for all weeks of the most recent year might be incomplete due to a delayed registration. Statistical offices may still revise these data in the course of the next update.
4) 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 documentationfor details.
5) Each year in the STMF refers to 52 weeks, each week has 7 days. In some cases, the first week of a year may include several days from the previous year or the last week of a year may include days (and, respectively, deaths) of the next year. In particular, it means that a statistical year in the STMF is equal to the statistical year in annual country-specific statistics.

The data in the STMF are presented without any corrections. The original data are notadjusted for death undercounts or smoothed. All known country-specific quality issues are documented in the country-specific metadata file.

Updates

The monitoring of weekly data is an important task and should be ideally performed on a regular basis. Generally, the update the STMF will be performed every one-two months. However, more frequent updates will be implemented in critical situations such as COVID-19 pandemic. The STMF should be considered only as a complimentary data resource in addition to national and international monitoring systems. Since April 2020, a number of countries publishing weekly death counts hasnotably increased. The EUROSTAT launched a collection of weekly death counts which may eventually include all EU countries in the future. In order to monitor the most recent trends.

Output data

The STMF output includes three files: STMF.xlsx, STMF.scv, and STMF.pdf. The first two files contain all available data series and provide identical data in two different formats. The last file provides country-specific metadata including information about data sources and and shape of the original data.

Each line (record) in the STMF.xlsx and STMF.scv files has an identical structure. Each record is unique (i.e., there are no duplicate records) and contains all necessary information for a particular week of each year. The sequence of records has no significance. Following data fields are included:

1) First four columns define a country, year, week, and sex (m – males, f – females, b – both sexes combined)
2) Next six columns (5 to 10) provides weekly death counts by age group (0-14, 15-64, 65-74, 75-84, 85+) and total (all ages combined)
3) Column 11 to 15 contains age-specific death rates, column 16 is reserved for crude death rates
4) Last three columns are data data-specific indicators:
    a. "Split" shows if data were split from aggregated age groups (0 if the original data has necessary detailed age scale). For example, if the original age scale was 0-4, 5-29, 30-65, 65+ the Split indicator will be equal 1.
    b. "SplitSex" indicates if the original data are available by sex (0) or data are interpolated (1)
    c. "Forecast" is equal 1 for all years where forecasted population exposures were used to calculate weekly death rates.

The Excel file (STMF.xlsx) is structured by country, i.e. country-specific data series are represented on separate worksheets. In addition, there is a worksheet with a data summary. The CSV file (STMF.csv) is provided to use in statistical packages. This is a standard comma-separated file with all data combined into one table. The file has the following heading (fields match 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 in one zip file). The filename has the following structure: XXXstmf.csv, where XXX is a population code (3-7-digits, all character values must be uppercase, see https://www.mortality.org/cgi-bin/hmd/DataAvailability.php for details). Each country-specific file contains standard headings (first line), which represent field identifiers. Each record contain all necessary information and is unique (i.e., there are no duplicate records). 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, the data files may contain a number of spaces to improve text file readability. The spaces have no other function.

Heading:

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

Each weekly death count refers to one record. Each death count (field "*Deaths*") is determined by population name (field "*PopName*"), geographical coverage (field "*Area*"), calendar year (field "*Year*"), calendar week ("Week"), sex (field "*Sex*"), age (field "*Age*"), the length of age interval (field "*AgeInterval*"), type of the data (field "*Type*"), and type of access (field "*Access*").

The format of the fields is as follows:

1.1) Population name – country code (see definition of country codes).

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 registered (depending on field Type).

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

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

1.6) Age (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 and unknown ages, respectively.

1.7) The length of the age interval (1- or 2-digit or '+' for open age interval). For example: 1, 2, 3, 5, 10, +.

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

The data processing starts from the collecting annual population exposures and death counts. The standard HMD estimates are not available for the most recent year(s) and months. Thus, the available HMD estimates are complemented by the forecasted projections covering the most recent period. . Lee-Carter model[2] to extrapolate annual death rates and using forecasted rates are used to estimate age-specific population exposures and death counts under the condition of zero migration. The Lee-Carter model is fitted using the available HMD data series starting from 2005.

At the next step, the original data are splitted into the standard age groups. This procedure is applied for each year independently using proportions from annual age-specific death counts (observed or forecasted):

[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] We use R packacge "demography" by R.Hyndman (https://cran.r-project.org/web/packages/demography/demography.pdf)

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

where $D_y^w(x, x+b)$ denotes the number of death in age interval $[x, x+b)$ in week *w* of year *y*; $D_y^w(x, x+b)$ is the number of death in age interval $[x, x+b)$ in year *y*.

A similar procedure is applied when sex-specific data are not available. 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:

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

Finally, the age-specific death rates 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*.