ABOUT MORTALITY DATA FOR ENGLAND AND WALES, CIVILIAN POPULATION

by Dmitri Jdanov 1 September 2005
Revised by Domantas Jasilionis Last updated: 05 May 2022

For the period of the two World Wars, population estimates and death counts provided by the General Register Office cover only the civilian population. The data series for the total population of England and Wales (including the military) uses estimates from Jdanov et al. (2005) for population (1912-1920 and 1932-1950) and for death counts (1914-1920 and 1939-1950) during war-time periods. Nevertheless, for comparison purposes, the Human Mortality Database (HMD) also presents this series based solely on the official data. During the two war-time periods (1912-1920 and 1939-1950), this series comprises only the civilian population. During other periods, the mortality estimates from this series are identical to those for the total population. Although cohort life tables are available for this civilian series, users should be aware that they are of questionable quality for cohorts that experienced significant military losses during the wars.

DATA QUALITY ISSUES

During war-time, the population of males shows a sudden decline because soldiers are not included. Population estimates during war-time are available by five-year age groups. Unfortunately, the standard HMD method for splitting such data into single years of age does not work well due to the irregular "migration" pattern. Therefore, these population estimates were split using a modified spline method (see Appendix 2).

Single-year death counts show age heaping effects during World War I and between the two World Wars. Signs of age heaping can be observed at ages ending with zero, such as 60 and 70 years (Appendix 3). Appendix 3 shows that mortality at ages 60 and 70 for some years in the 1920s and 1930s seems to be similar or higher than at age 71.

For other details, see the *Background and Documentation* files for the Total Population of England and Wales.

Revision NOTES

Changes with the December 2017 revision:

Life tables: All life tables have been recalculated using a modified methods protocol. The revised protocol (Version 6) includes two changes: 1) a more precise way to calculate a0, the mean age at death for children dying during the first year of life and 2) the use of birth-by-month data (where and when available) to more accurately estimate population exposures. These changes have been implemented simultaneously for ALL HMD series/countries. For more details about these changes, see the revised Methods Protocol (at http://www.mortality.org/Public/Docs/MethodsProtocol.pdf), particularly section 7.1 on Period life tables and section 6 and Appendix E, on death rates. The life tables calculated under the prior methods (Version 5) remain available at v5.mortality.org but they have not been, and will not be, updated.

APPENDIX 1:

Description of the original data used for HMD calculations

DEATHS

Period	Type of Data	Age grouping	RefCode
1841-1909	Annual number of deaths to the *de facto* population, by sex and age groups (nx1 rectangles)	0, 1, 2, 3, 4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+	2
1910-2001	Annual number of deaths to the *de facto* population, by sex and age groups (1x1 rectangles)	0, 1, 2, …maximum age attained	3, 4, 5, 20
2002-2014	Annual number of deaths to the *de facto* population, by sex and age groups (1x1 rectangles)	0, 1, 2, …110+	13, 14, 18, 19, 21, 22, 28, 34, 40
2015-2020	Annual number of deaths to the *de facto* population, by sex and age groups (1x1 rectangles)	0, 1, 2, …105+	45, 49, 53, 58

POPULATION

Period	Type of Data	Age grouping	Comments	RefCode
1841, 1851, 1861, 1871, 1881, 1901, 1951, 1961	Census counts	0,5,10,..,85+	*de facto* population	7, 8
1911, 1921, 1931	Census counts	0, 1, 2, 3, …, 100+	*de facto* population	7
1911-1921, 1931-1951	Annual mid-year population estimates (of permanent residents)	0, 1, 5, 10, …., 85+	These estimates were split into single years of age using a non-standard method (see Appendix 2 for details)	6
1962-2020	Annual mid-year population estimates (of permanent residents)	0,1,2,3,….,90+	The raw data for 1981-1990 shown in the HMD are rounded to hundreds (RefCode= 11). The original population estimates (RefCode= 10) were used for calculations but cannot be released publicly.	9, 10, 41, 46, 50, 54, 57

APPENDIX 1 (CONTINUED):

DESCRIPTION OF DATA USED FOR LEXIS DATABASE

BIRTHS

Period	Type of Data	RefCode
1841-2020	Annual live birth counts, by sex	12, 17, 24, 25, 29, 30, 33, 39, 43, 47, 51, 55

BIRTHS BY MONTH

Type of data: Annual live birth counts by month
Period covered: 1938-2020.
RefCode(s): 31, 32, 42, 44, 48, 52, 56.

APPENDIX 2:

Splitting aggregated civilian population estimates during war-time

By Dmitri Jdanov

Civilian data for England and Wales for the period of World War I and World War II are available only by five-year age groups. The HMD Methods Protocol (MP) requires splitting these data into one-year age groups. The method described in the MP (see section titled “Intercensal survival with data in n-year age groups”) starts with a distribution of the population in more recent years, and uses reverse survival to estimate the age distribution for earlier years. This approach gives acceptable results in most cases. The main assumption of this method is that migration is (nearly) uniformly distributed within age groups. That assumption does not hold for males aged 15-19 during war-time because conscription begins at age 18. Thus, in this age group migration is heavy at ages 18-19 (during war-time), but negligible at ages 15-16. Furthermore, at the first post-war census, these conscripts had been demobilized, and therefore are included in the census counts. The standard HMD methods would have us redistribute out-migration (conscripts) uniformly within the whole age group 15-19, thus producing implausible results. As shown on Figure A2.1, the estimates show an unexpected spike in the population at age 15-16. In addition, the estimates for ages 20-40 exhibit a strange “sawtooth” pattern, which is especially noticeable in 1914-15 (red and green lines) and 1939 (shown in blue).

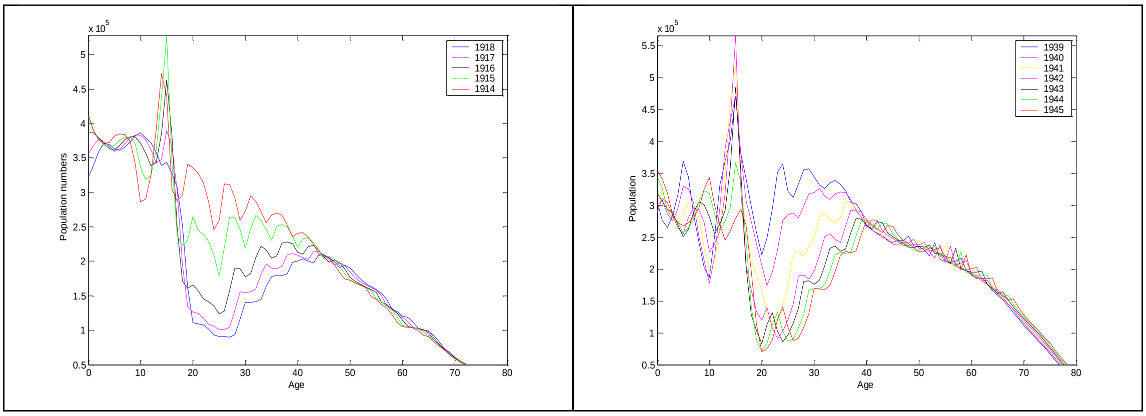


Figure A2.1. Population estimates for 1914-1918 (World War I) and 1939-1945 (World War II) by single year of age, England and Wales, males. The standard method for splitting population data was used. Note peaks at ages 15-16 and “sawtooth” pattern in 1914-1915 (red and green lines) and 1939 (blue).

Thus, the standard method cannot be applied for splitting population data in 1914-1918 and 1939-1945 in England and Wales. The simplest alternative in this case would be to apply the same approach as for death counts and split cumulative population numbers using a cubic spline. Unfortunately, this method assumes smooth changes (without large fluctuations) in the population structure. Therefore, it cannot be applied in the specific case of war. So, to split these data into one-year age groups, we used the method described below. This strategy is based on the distribution of cumulative net migration.

Suppose that war began in year t_{begin} and ended in t_{end}. Population estimates $P(x,t)$ are available, aggregated by five-year age groups within the time interval $(t_0;T)$, where $t_0 < t_{begin}$ and $T > t_{end}$ are the years of the last pre-war and first post-war censuses, respectively. Census data are available by one-year age groups. Let $\hat{P}(x,t)$ represent population estimates in year *t* ($t_0 < t < T$) at age *x* calculated as follows:

$$\hat{P}(x,t) = \begin{cases} P(x-(t-t_0),t_0) - \sum_{i=t_0}^{t-1} [D_U(x-(t-i),i) + D_L(x-(t-i)+1,i)], \\ \qquad \text{for } t-t_0 \le x < 20 \text{ and } (x \ge 20 \text{ and } t \le t_{begin}) \\ B(t-x) - \sum_{i=t-x}^{t-1} [D_U(x-(t-i),i) + D_L(x-(t-i)+1,i)], \\ \qquad \text{for } x < t-t_0,\ t \le t_{begin} \text{ and } x < \min(t-t_0,20), t > t_{begin} \\ P(x+(T-t),T) + \sum_{i=t}^{T-1} [D_U(x-(t-i),i) + D_L(x-(t-i)+1,i)], \\ \qquad \text{for } t_{begin} < t,\ x \ge 20 \end{cases} \tag{A2.1}$$

where $B(t)$ is the number of live births in year *t*, $D_L(x,t)$ and $D_U(x,t)$ are the numbers of lower-triangle and upper-triangle deaths at age *x* in year *t* (see p.9 of the Methods Protocol for details), and $P(x,t_0)$ and $P(x,T)$ are pre-war and post-war census counts, respectively. In other words, $\hat{P}(x,t)$ is a post-censal estimate (based on the pre-war census) for all ages in pre-war years and for ages less than 20 after the beginning of the war. For older ages (20+) after the war begins, we use pre-censal estimates (based on the post-war census). These estimates provide the population distribution assuming zero migration[1].

Let ${}_5m_x(t)$ denote the difference between official population estimates by five-year age groups, ${}_5P_x(t)$, and our estimates (A2.1) for that same age group $[x,x+5)$:

---

[1] Note that for the period after the war begins, these estimates may not be consistent across the whole age range within a given calendar year. For example, during World War I, $P(19,1915)$ is calculated by subtracting deaths since the last pre-war census, whereas $P(20,1915)$ is calculated by adding deaths that occurred before the first post-war census. Thus, we may observe a discontinuity at age 20. Nevertheless, the estimates are consistent within each five-year age group because pre- and post-censal estimates are not mixed within five-year age intervals.

$$ {}_5m_x(t) = {}_5P_x(t) - \sum_{i=0}^{4} \hat{P}(x+i,t) \quad . \qquad \text{(A2.2)} $$

This difference can be interpreted as the implied migration (accumulated over the period from t_0 to t) for the respective cohorts aged x to $x+5$ at the beginning of year t. Now it is possible to split ${}_5m_x(t)$, instead of ${}_5P_x(t)$, into single years of age, which simplifies our task, because more flexible assumptions can be formulated for migration than for the population.

Let $M(x,t) = \sum_{i=0}^{x} {}_5m_i$ be the cumulative (across age) net migration up to age x. We assume that $M(x,t)$ is known for ages 0,1,5,10,15,20,…,85. This collection of ages is sufficient during peacetime when there is no mobilization of troops. For war-time, it is necessary to estimate one more data point at age 18 (because we expect a big increase in migration between ages 18 and 20). To define $M(18,t)$, we fit a linear regression model to the values of cumulative migration $M(x,t)$:

$$ M(x,t) = \alpha + \beta x \text{ , for } x = 1,5,10,15. \qquad \text{(A2.3)} $$

Then, using the coefficients from the regression model, we estimate that $M(18,t) = \hat{\alpha} + \hat{\beta}(18)$. That is, we assume that migration follows a linear trend from ages 1 thru 17.

Next, we fit a standard cubic spline function (e.g., see McNeil et al., 1977 and Appendix B of the MP) to $M(x,t)$ for $x = 0,1,5,10,15,18,20,25,..85$. Using the coefficients from the cubic spline function, we calculate fitted values $\tilde{M}(x,t)$ by single years of age. Then, we derive estimates of net migration $\tilde{m}(x,t)$ by single years of age as follows:

$$ \tilde{m}(x,t) = \tilde{M}(x+1,t) - \tilde{M}(x,t) \text{ , for } x = 0,1,2,...85. \qquad \text{(A2.3)} $$

Finally, we use these estimates to calculate population numbers (adjusted for migration) by single year of age:

$$ \tilde{P}(x,t) = \hat{P}(x,t) + \tilde{m}(x,t), \quad x = 0,1,2,....85. \qquad \text{(A2.4)} $$

The results of such and interpolation of population data are presented in Figure A2.2

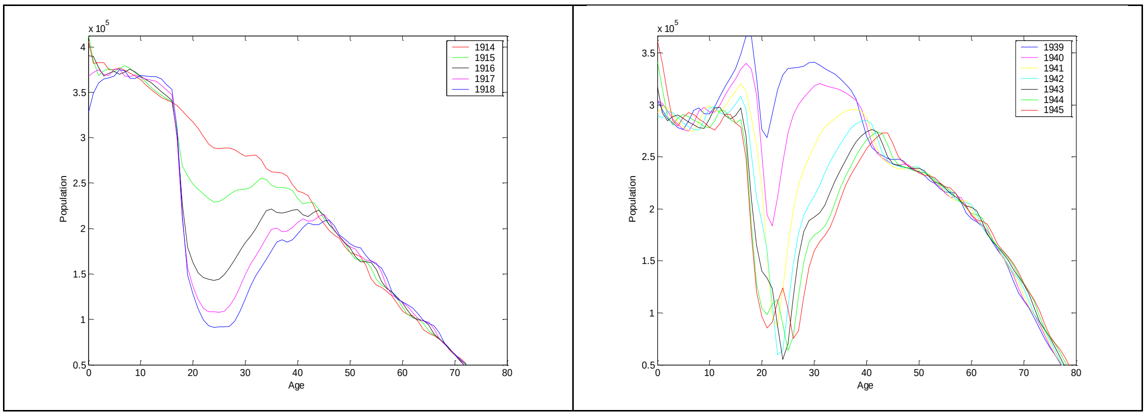
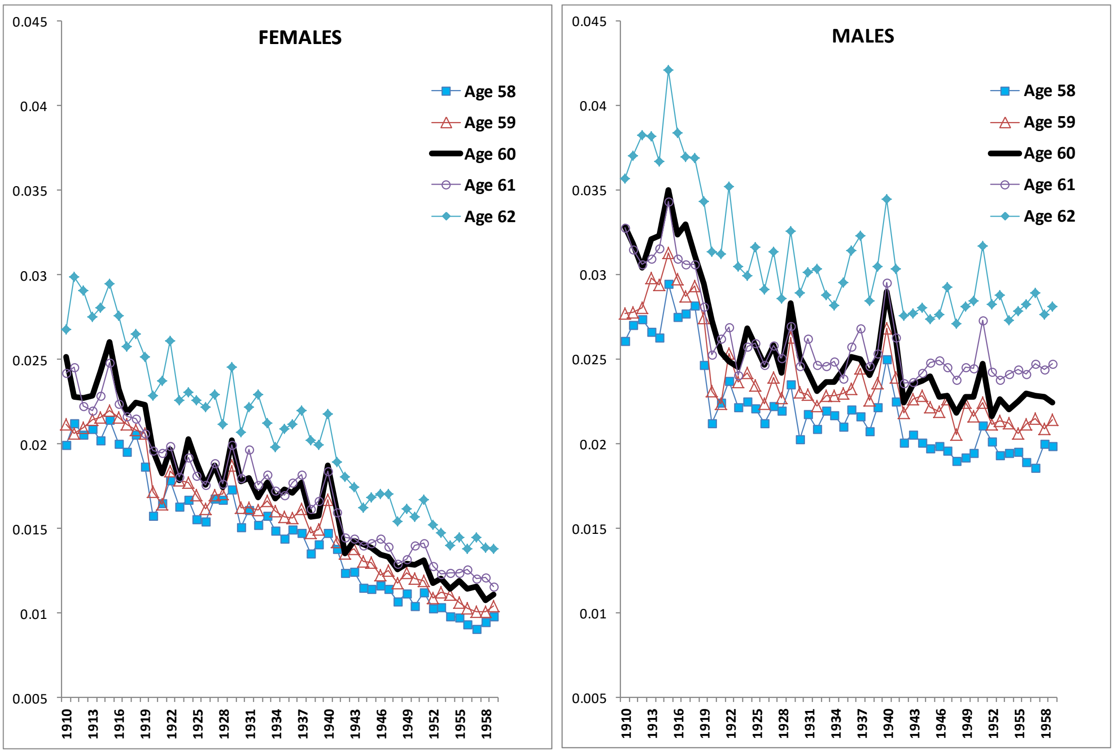


Figure A2.2. Population estimates for 1914-1918 (World War I) and 1939-1945 (World War II) by single year of age, England and Wales, males. The method used is based on splitting cumulative migration. Note that peaks at 18 in 1939-1944 are explained by irregularities in the population structure due to World War I (decrease in birth rate in 1914-1918 and rebound in 1918).

APPENDIX 3:

Mortality rates for selected ages, England & Wales Civilian population, males, 1910-1959.

A) Ages 58-62:



B) Ages 68-72: