ABOUT MORTALITY DATA FOR GERMANY

By Rembrandt Scholz, Dmitri Jdanov, Eva Kibele, Pavel Grigoriev, Sebastian Klüsener

Last updated by Isabella Marinetti on May 10, 2022

GENERAL

From 1949 to 1990, Germany was split into two parts, which were referred to as the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR). For the period since the German reunification, this report distinguishes between “East Germany”, which corresponds to the area of the former GDR, and “West Germany”, which corresponds to the area of the former FRG.

Germany is a federal republic consisting of 16 states (referred to as *Länder* or *Bundesländer*). In the Human Mortality Database (HMD), the following *Länder* combine into West Germany: Baden-Württemberg, Bayern, Bremen, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, and Schleswig-Holstein, as well as West Berlin. East Germany consists of the following *Länder*: Mecklenburg-Vorpommern, Brandenburg, Sachsen-Anhalt, Sachsen, Thüringen, and East Berlin. This division was maintained in the HMD until 2015. Since 2016, West Berlin is included in East Germany and consequently excluded from the territory of West Germany.

The official statistical system in Germany

There is close co-operation between the Federal Statistical Office and the statistical offices of the *Länder*. In Germany, the Federal Government and the *Länder* share the responsibility for the collection and production of official statistics, following the tradition of the Federal Republic of Germany. The Federal Statistical Office compiles federal results from the *Länder* data and publishes them; the statistical offices of the *Länder* publish their results separately. There is a clear division in the dissemination of statistical data between the Federal and the *Länder* offices. The Federal Statistical Office publishes data for Germany as a whole with a geographical breakdown at the level of the *Länder* only. The publication of regional data at or below the *Länder* level, even down to the level of local administrations, is the task of the statistical offices of the *Länder*.

Between 1876 and 1945, vital registration statistics in all parts of Germany was based on the same law (the *Personenstandsgesetz*). Census taking has been harmonized across the member states of the German Empire since 1925. Previously, most of the states had counted the *de facto* population, while the Saxony and the Hanseatic city states of Hamburg, Lübeck, and Bremen had reported the *de jure* population (Lee and Schneider, 2005). The “*Personenstandgesetz*” was slightly modified between 1945 and 1990, with separate practices in the FRG and the GDR. Since October 3rd, 1990, the official day of reunification, population statistics in both parts of Germany have been collected following the same laws.

Differences in population statistics between the two parts of Germany prior to 1990 existed in the definitions used for live births and for the resident population. In the GDR, the term "live birth" was used when two signs of life occurred, while in the West only one sign was required. The GDR only included in the resident population people who had been living in the country for at least six months, whereas in FRG the resident population included all persons in residence.

DATA SOURCES

All of the raw data used for the Human Mortality Database (HMD) originate from statistical data provided by the Federal Statistical Office in Wiesbaden and the Statistical Offices of the German states. The HMD Input Database (Input DB) includes death, birth and population counts for all years since 1990.

TERRITORIAL COVERAGE

There has been no territorial change since 1990.

DEATH COUNT DATA

Coverage and Completeness

The annual statistics include all registered deaths of residents of Germany (source: laws for the registration of births, marriages, and deaths). We are not aware of any substantial misreporting in the death statistics. The quality of age reporting is high as since 1876 birth and death registration has been conducted by the registry office in all parts of Germany, and death registration has been requiring a birth certificate from the registry office.

For the years 2018-2020, the German Statistical Office (Destatis) introduced a new procedure to deal with small numbers of deaths. When the number of cases is less than 3 or when there are single values identical to the marginal totals, the non-zero number of death counts in that cell are suppressed for privacy protection. In addition, the relative additional non-risky cells are suppressed as well (secondary suppression or complementary suppression) to avoid re-identification of the deceased individuals, following German rules (see Rothe, 2015). The suppressed cells issue arises when splitting Germany into West and East Germany. In particular, Eastern Germany, with a smaller population than the Western part, is more likely to experience a small number of deaths, especially at young ages (5-25 years old), where death rates are very low, and at very high ages (105 years and more), where the number of survivors is very small. Hence, whenever in the East German dataset a cell with a number of deaths less than 3 and/or the total of that specific age is also less than 3, that row death count has been suppressed by the National Statistics Office in both the East and West German datasets. The average proportion of suppressed cells over the three calendar years 2018-2020 is 0.09% for East Germany and 0.12% for West Germany.

We used the Iterative Proportional Fitting (IPF) method to estimate suppressed values. The IPF algorithm was initially developed by Fienberg (1970) and later revised by Bishop et al. (1975). The IPF iteratively recalculates the cell values in a contingency table so that the row and column sums of the cells become equal to the marginal totals (see further details in Appendix 2).

The algorithm is applied independently to data in each year (2018-2020) and sex and results in adjusted age-specific data cells that correspond to the marginal totals for East Germany.

Suppressed cells in the West German dataset were calculated by difference between the Total and the East German counts.

BIRTH COUNT DATA

Coverage and Completeness

The definition of a "live birth" has changed over time:

- In 1991-1993: Newborn child whose breathing *or* heartbeat started after complete separation from the mother independent of cutting the umbilical cord and delivery of the placenta, and whose body weight was no more than 1,000 grams.
- Since 1994: Newborn child whose breathing *or* heartbeat started after complete separation from the mother independent of cutting the umbilical cord and delivery of the placenta, and whose body weight was not more than 500 grams.

Prior to 1994, stillbirths were under-estimated compared with the following period because the minimum birth weight for classification as a stillbirth rather than as a spontaneous abortion (i.e., miscarriage) changed on March 31st, 1994, from 1,000 grams to 500 grams.

POPULATION COUNT DATA

Coverage and completeness

Official population estimates are based on census counts. Between census years, the birth, death, and migration counts are used by the Federal Statistical Office in Wiesbaden and the Statistical Offices of the German states to calculate the population on the last day of each year (December 31st).

In the FRG, the last census prior to reunification (in 1990) took place on May 25th, 1987. In the former GDR, it was conducted on December 31st, 1981. The population register was already established in East Germany in the 1970s. The results of GDR census of 1981 revealed high quality of population register as the difference between these data sources turned out to be fairly small (about 20 thousand people). This population register was used as the basis for an enumeration conducted in East Germany on October 3rd, 1990, the official day when West and East Germany were reunited. The population register in the East was closed in 1992 because it was not compatible with the rules on data protection in the FRG. In the time period after 1988, population movement between West and East Germany was very intense, particularly in the years around the re-unification. However, the first census after unification was carried out in 2011 only. This was the first official register-based census, and it is disputed whether such a procedure can deliver results of the same quality as a traditional census (Coleman, 2013). For an overview on the potential deficiencies of the 2011 census see Scholz and Kreyenfeld (2016). That it was quite challenging to implement a register-based census in Germany is also underlined by the fact that the Federal Statistical Office decided not to use the officially published census results by age and sex as the basis for post-censal population estimates. Instead, a modified version of the estimates was published, which used additional statistical information (see Kaus and Mundil-Schwarz, 2015). Like the Federal Statistical Office, for the HMD calculations we decided to rely on the modified official census data.

The difference between the old population estimates produced earlier by the Federal Statistical Office in Wiesbaden and the Statistical Offices of the German states and the population estimates based on the new census of 2011, as obtained for the 1st of January 2012, is quite substantial. There were actually 977,000 fewer males and 539,000 fewer females compared to the previously published official population estimates.

DATA QUALITY ISSUES

Quality of data on population counts

As mentioned above, the differences between the population estimates constructed from the most recent census of 2011 and those constructed from the former censuses of 1981 (East Germany) and 1987 (West Germany) are quite substantial. Unfortunately, the Federal Statistical Office decided not to recalculate back the population estimates by age and sex for the complete inter-censal period. Some of the discrepancies between the two series of estimates arise in part from the massive internal and international migration flows experienced by Germany in the late 1980s and in the 1990s due to the fall of the Iron Curtain and German reunification. In consequence, the standard HMD inter-censal method, which assumes that migration is uniformly distributed across the inter-censal period is not applicable in this case.

Another problem is related to substantial efforts by the statistical office to improve its estimation of the population by cleaning its registers in the years prior to the 2011 census. This was done with the help of local (regional) statistical offices which were instructed to make sure that all out-migrants had been de-registered and that no one was registered in two different areas simultaneously, using alternative sources of information such as tax records. Without these prior cleaning activities, the differences between the population estimates based on the former censuses and those based on the 2011 census would have been even larger. Unfortunately, these adjustments were only taken into account in the year when the error was detected while the undocumented outmigration event or a multiple registrations might have occurred years or decades before. Data cleaning operations were implemented by the State Statistical Offices in various degrees throughout the inter-censal period, but two periods witnessed substantial corrections: 2004 and 2008–2010. To our knowledge, information about the exact number of individuals removed from the registers and the exact dates of the removals from the resident population counts is unknown.

The Syrian refugee crisis and the massive migration flow into Germany during the most recent years have had a notable impact on the size and age-structure of the population. According to the 2014 official estimates, net migration in Germany reached 550,000 and it increased again in 2015, reaching one million (with two million arrivals and one million departures (Statistishes Bundesamt, 2017)). Most of the migrants were young males with a sex ratio particularly skewed in their favor in 2015.

Intensive in-migration eventually affected the quality of population estimates at younger ages because of problems related to the registration of refugees and asylum seekers. Though migrants are required to be registered, official estimates of the number of refugees are affected by an under-coverage which is difficult to quantify, though issues of multiple counts have also been raised as migrants could have registered each time they moved to a new area within the country[1].

[1] The Statistical Office acknowledges the problems related to refugee registration, stating that: “It must be assumed that the 2015 figures on people seeking refuge are affected by undercoverage which cannot be quantified because the registration authorities were unable to register all people seeking refuge in a timely manner. In addition, double counting

To resolve all these issues, we used research estimates by Klüsener et al. (2018) for the intercensal period from 1987 to 2012. The method used in this paper is a modified version of the HMD inter-censal method that takes into account migration flows.

REVISION HISTORY

Changes with the May 2022 Revision

The German Statistical Office (Destatis) introduced a new procedure to deal with small numbers of deaths for the years 2018-2020. When the number of cases is less than 3 or there are single values identical to the marginal total, the non-zero number of death counts in that cell is suppressed for privacy protection, and also the relative additional non-risky cells are suppressed (secondary suppression or complementary suppression) to avoid re-identification of the deceased individuals. As mentioned above, we implemented a special method based on the IPF algorithm to reconstruct the data (see further details in Appendix 2).

Changes with the December 2017 Revision

Since January 2016, the territory of East Germany includes the city of Berlin.

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://v6.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 will not be further updated in the future.

Changes with April 2017 Revision

We carried out some further refinement to the adjustment implemented with the previous revision to construct population estimates for the inter-censal period. These estimates were originally published in Klüsener et al. (2018). This modification resulted in very minor changes in population size (predominantly at ages above 80 years), and had a negligible impact on mortality rates.

Changes with April 2016 Revision

The inter-censal population estimates for the period 1987–2011 have been revised to account for the results of the last census. For this, we used a country-specific method. The

may have occurred because of false register entries in connection with the distribution of people seeking refuge within Germany" (https://www.destatis.de/EN/FactsFigures/_CrossSection/Refugees/Refugees.html) adjustments to the population estimate numbers were implemented in two main steps. In the first step, we accounted for the corrections made by the statistical offices of Germany to their population estimates prior to the 2011 census. This adjustment accounted for errors that occurred due to multiple registrations and undocumented out-migration events. The adjustments assumed that when they were implemented (in 2004 and in 2008–2010), all outmigration rates above an ‘expected‘ level of outmigration could be attributed to the adjustments made by the statistical offices. Using a simple spline interpolation we estimated the ‘expected’ migration rates by cohort. The difference between ‘expected’ and ‘observed’ events in years 2004 and 2008–2010 was assumed to reflect artificial migration counts, which were then added to the accumulated error. In the second step, this error was redistributed by cohort and by sex assuming that it had accumulated in a linear manner over the inter-censal period. For more details see Data Quality Issues section and Klüsener et al., 2018*).*

In the previous revisions we applied an adjustment factor to the population older than 90 years because of the low quality of population estimates at advanced ages (Scholz and Jdanov, 2007). The new official estimates for the population at ages 90+ years (based on the 2011 census) are extremely close to the alternative HMD estimates derived previously from the 1981 and 1987 census and vital statistics using the aforementioned correction factor. As indicated in Figure 1, the difference between the new estimates and the previous HMD estimates is close to zero, while the difference between the previous HMD estimates and the previous official estimates used to be large, particularly among males in Western Germany (see more details about the latter comparison in Jdanov, Scholz, Shkolnikov, 2005).

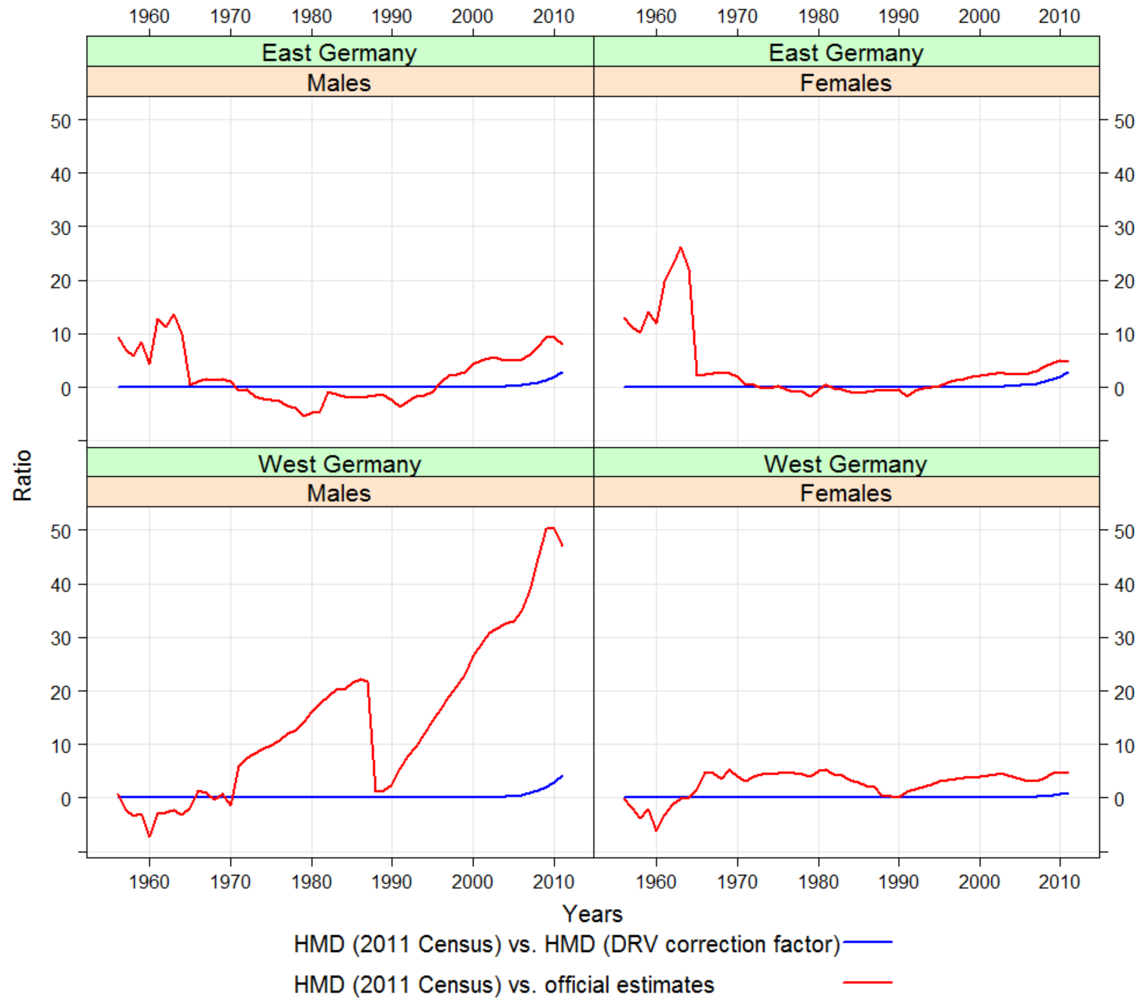

Figure 1. Relative difference (per cent): HMD (2011 Census) vs. HMD (DRV correction factor) and HMD (2011 Census) vs. old official population estimates; ages 90+

ACKNOWLEDGEMENTS

We would like to thank: Walter Becker, Jutta Gebhardt, Hans-Peter Bosse, Dieter Emmerling, Gisela Papaja-Yadegari, and Anja Conradi-Freundschuh at the Federal Statistical Office Wiesbaden; and Anita Depner, Beate Koehn, Viola Kraemer, Rolf Paletta, R. Hopf-Engelk, and Monika Zornik at the Statistical Office of Berlin for their kind help and assistance in compiling the German data. We express our special gratitude to Bettina Sommer (Federal Statistical Office) and Juergen Paffhausen (Statistical Office of Berlin) for very useful comments. We thank the staff at the research data center of the Deutsche Rentenversicherung (DRV) for the provision of pension statistics.

REFERENCES

Coleman, D. (2013). The twilight of the census. *Population and Development Review*, *38*(s1), 334-351.

Jdanov, D.A.; Scholz, R.D.; Shkolnikov, V.M. (2005).Official population statistics and the Human Mortality Database estimates of populations aged 80+ in Germany and nine other European countries. *Demographic Research*, 13:14, 335-362.

Kaus, W., & Mundil-Schwarz, R. (2015). Die Ermittlung der Einwohnerzahlen und der demografischen Strukturen nach dem Zensus 2011. *Wirtschaft und Statistik* 2015(47), 18-38.

Klüsener S, Grigoriev P, Scholz RD, Jdanov DA (2018). Adjusting inter-censal population estimates for Germany 1987-2011: approaches and impact on demographic indicators. *Comparative Population Studies* 43, 31-64.

Lee W. R., Schneider M. S., 2005, "Amtliche Statistik zwischen Staat und Wissenschaft, 1872-1939", in MAckensen R., Reulecke J. (eds.), *Das Konstrukt „Bevölkerung" vor, im und nach dem "Dritten Reich"*, Wiesbaden, VS Verlag für Sozialwissenschaften, pp. 50-91.

Scholz, R. and D. Jdanov. (2007). Verfahren zur Korrektur der Bevölkerungsbestände der amtlichen Statistik im hohen Alter. MPIDR Working Paper WP-2007-002. Rostock: MPIDR. (http://www.demogr.mpg.de/Papers/Working/wp-2007-002.pdf)

Scholz R. and Kreyenfeld M. (2016) The register-based census in Germany: historical context and relevance for population research. *Comparative Population Studies* 41(2): 175-204.

Statistishes Bundesamt (2017). Official website of the German Statistical Office [DESTATIS]. Available at https://www.destatis.de

Rothe, P. (2015). Statistische Geheimhaltung–der Schutz vertraulicher Daten in der amtlichen Statistik. Bayern in Zahlen, 5, 294-303.

Bishop, Y., S. Fienberg, and P. Holland (1975). Discrete Multivariate Analysis: Theory and Practice. MIT University Press.

Fienberg, S. (1970): An iterative procedure for estimation in contingency tables, The Annals of Mathematical Statistics, Vol. 41, N. 3, 907-917.

Jasilioniene, A., Jdanov, D. A., Sobotka, T., Andreev, E. M., Zeman, K., Shkolnikov, V. M., ... & Rodriguez, G. (2015). Methods protocol for the human fertility database. Rostock: Max Planck Institute for Demographic Research.

APPENDIX 1:

DESCRIPTION OF THE ORIGINAL DATA USED FOR HMD CALCULATIONS

DEATHS

Period	Type of Data	Age Grouping	Comments	RefCode(s)†
1991-2009	Annual number of deaths to residents by sex, age and birth cohort (Lexis triangles)	0, 1, 2, 3, … maximum age attained		5, 6, 7, 14, 15, 16, 24, 25, 28
2010-2017	Annual number of deaths to residents by sex, age and birth cohort (Lexis triangles)	0, 1, 2, 3, …110+		28, 31, 35, 38, 41
2018-2020	Annual number of deaths to residents by sex, age and birth cohort (Lexis triangles)	0, 1, 2, 3, …106+		48, 54, 55

POPULATION

Period	Type of Data	Age Grouping	Comments	RefCode(s)†
1990-2011	Annual population estimates (as of December 31st) by sex and age	0, 1, 2, 3, ...95+	Reconstructed data	33
2012-2020	Annual population estimates (as of December 31st) by sex and age	0, 1, 2, 3, …100+		34, 37, 40, 45, 49, 50

BIRTHS

Period	Type of Data	Comments	RefCode(s)†
1990-2020	Annual live birth counts by sex		12, 13, 20, 21, 22, 23, 29,32, 36, 39, 42, 46, 51

BIRTHS BY MONTH

Type of data: Annual live birth counts by month

Period covered: 1946–2020

RefCodes: 43, 44, 47, 52, 53

† The reference code is used to link 'Input Data' with the primary data source

APPENDIX 2:

DESCRIPTION OF THE IPF ALGORITHM

As explained in the main text, the German Statistical Office (Destatis) has introduced a new procedure for the year 2018 to 2020 to deal with small numbers of deaths in order to protect confidentiality: in the death tabulations by calendar year, sex, age, cohort for East and West Germany, cell counts equal to 1 or 2 are suppressed. For any suppressed cell in either East Germany (the most common occurrence given the relatively small population) or West Germany, both cells are suppressed so that the missing number cannot be reconstructed by subtraction. There are a number of methods available to impute a value in such situations. In the HMD, we decided to use the Iterative Proportional Fitting (IPF) method.

The Iterative Proportional Fitting (IPF) algorithm is a commonly-used technique appropriate to adjust raw data in age-specific data cells that correspond to the marginal totals, hence ensuring consistency in a contingency table.

We applied the IPF procedure separately to each year and subpopulation. We started with East Germany as its population is lower than in West Germany. When needed, we could also apply this procedure to West Germany but, since all values are documented for the country as a whole, it was simple to estimate the values missing for West Germany as the difference between total Germany and East Germany.

The IPF algorithm is applied to the matrix of suppressed cells, where we have only the rows and columns totals. Before implementing the algorithm, we have to assign an initial (starting) value for each missing cell, which is done separately for each single age group. When an age-specific cell is suppressed for both sexes combined, the missing values are replaced by the ratio of the difference of the reported total deaths and the actual sum over all ages, to the number of missing cells in that year.

The rationale used to fill the suppressed sex-specific cells when the total is not suppressed is that at least one of the two cells of the male and female death counts should be equal to 1 or 2, or it would not have been suppressed by the statistics office.

The procedure can be formally expressed as follows.

If d_a^{tot} (for a given year) is unknown:

$$\mathrm{d}_{\mathrm{a}}^{\mathrm{tot}} = \frac{\mathrm{d}_{\mathrm{D}}^{\mathrm{tot}} - \mathrm{d}_{\mathrm{OBS}}^{\mathrm{tot}}}{\mathrm{n}^{\mathrm{miss}}} \; for\ any\ age\ a$$

If d_a^{tot} (for a given year) is known:

$$d_a^f = 1; \quad d_a^m = d_a^{tot} - d_a^f \; for\ a < 100$$
$$d_a^f = d_a^{tot} - d_a^m; \quad d_a^m = 1 \; for\ a \geq 100$$

Where d_a^{tot} represents the death count for both sexes at age *a*, d_D^{tot} represents the reported death counts by the statistics office in a given year for both sexes, d_{OBS}^{tot} is the actual sum of death counts in a given year for both sexes, and d_a^m and d_a^f are age- and sex-specific death counts. In this formula, the fact that the probability of death is higher for men than for women is taken into account, as well as the fact that there are many more women than men surviving to old ages.

At this point, the starting value for the IPF is defined and the next steps was implemented as follows. The IPF technique was applied to data for each year, on the contingency table *n x 3*, in which each of the three columns represent data for men, women and the two sexes combined, and *n* is the number of rows with missing values in that year. The row sums (i.e. the total number of deaths at each age over both sexes) are known (at least after the previous step has been carried out). The column sums (the sex-specific totals across all ages with suppressed death counts) can be calculated by difference between the sex-specific total number of deaths at all ages and the actual number of deaths at all ages with a suppressed value.

The process consists of the following steps.
Let index k be an iteration number.
Firstly ($k = 0$), we define as described before the starting values of the suppressed data matrix.

Knowing that:

$D_i^k(x)$ is the estimated (after k steps) number of deaths of gender i ($i = 1, 2$) at age x and in a specific cohort;
$D(x)$ is a fixed number of deaths at age x and in a specific cohort, for both genders;
D_i is a fixed sum (over ages) of all deaths of gender i of a specific cohort.

we can apply the IPF algorithm. The marginal total is distributed proportionally between the cells, first by row (following equation 1):

$$D_i^{k+1}(x) = \frac{D_i^k(x)}{\sum_i D_i^k(x)} \cdot D(x) \qquad (1)$$

and then by column (following equation 2):

$$D_i^{k+2}(x) = \frac{D_i^{k+1}(x)}{\sum_i D_i^{k+1}(x)} \cdot D_{\mathrm{i}} \qquad (2)$$

The distance function Λ is calculated at each step by difference between the cell values and the marginal totals:

$$\Lambda^k = \sum_x \left(\sum_i D_i^k(x) - D(x)\right)^2 + \sum_{\mathrm{i}} \left(\sum_{\mathrm{x}} D_i^k(x) - D_I\right)^2 \qquad (3)$$

The iterative process is interrupted when the value of the distance function (equation 3) is less than 10^{-8}.