

About Spain Data on Causes of Death

By: Rosa Gómez Redondo, Aina Faus-Bertomeu, Svitlana Poniakina and Noelia Cámara Izquierdo

Last updated: 16/04/2023

General

The collection of official demographic statistics for Spain began in 1856 with the founding of the General Statistics Commission of the Kingdom (*Comisión de Estadística General del Reino*). In 1857, the commission became the officially designated General Board of Statistics (*Junta General de Estadística*), and the first modern population census was conducted on May 21st of that same year. In 1873 the agency was renamed the Geographic and Statistical Institute (*Instituto Geográfico y Estadístico*), and in 1931 it was further renamed the Cadastral and Statistical Geographic Institute (*Instituto Geográfico Catastral y Estadístico*). Finally, in 1945, the organization was split into the National Geographic Institute (*Instituto Geográfico Nacional*) and the National Statistics Institute (*Instituto Nacional de Estadística*, INE). The latter is Spain's current official statistical office and is in charge of conducting censuses as well as collecting and publishing other official demographic, social and economic statistics. Since 1989, INE has been responsible for coordination with the Comunidades Autonomas (regions in Spain with autonomous governments) and has supervised all of the state statistical services within the General State Administration (INE, 2003).

Since 1860, censuses have been approximately every 10 years. They took place in the years 1857, 1860, 1877, 1887, 1897 and, starting in 1900, every 10 years until 1970. Further censuses took place in 1981, 1991, 2001 and 2011. Prior to 1970, censuses were generally conducted on December 31st, but in 1981 this was changed to March 1st. This is because many Spaniards are away from home during the Christmas and New Year holidays (Reher & Valero Lobo, 1995:23). The most recent census was conducted on November 1st 2011.

Starting in 1564, parish registries were required to collect vital statistics data (i.e., births, deaths and marriages), although baptism and marriage registries had often existed much earlier. After 1600, death registries became more common but were often fraught with data quality problems. The most important problem was the under-registration of infant deaths. In 1870, the law established a Civil Status Registry, which made available a continuous series of vital statistics data. Nonetheless, problems with data reliability persisted until the beginning of the 20th century. After 1900, data from the Civil Status Registry were almost always of better quality than the parish registries. Furthermore, data since 1870 can be considered of comparable quality to other countries in the region (Reher & Valero Lobo, 1995:85-86).

In 1878, Spain began publishing vital statistics data (based on the Civil Status Registry) in a series called the Natural Movement of the Population (*Movimiento Natural de la Población*, MNP). Since 1900, this series has been published annually.

In 1975, a number of changes were introduced that substantially improved data quality (Del Campo & Navarro López, 1987). The definition of a live birth was changed to conform to the international standard. In addition, the population covered by vital statistics was changed from the *de facto* population (i.e., of fact (*de hecho*), based on the place of birth/death) to the *de jure* population (i.e., of law (*de derecho*), based on the permanent legal residence). Deaths were first classified by both the completed age and birth cohort. Furthermore, the World Health Organization's classification began in this year to be applied for causes of death.

It was not possible to study continuous series of causes of death in Spain until the beginning of the 20th century. Since 1871, when the Civil Registry (*Registro Civil*) was established, cause of death data appear to have been collected continuously. However, these data must be treated cautiously until 1902, when the statistic on Deaths by Cause of Death was founded. In this same year, the statistics were standardized and a plan was implemented for publishing on a yearly basis the population's movements over the preceding year. Additionally, the title *Movement of the Spanish Population* was changed to *Natural Movement of the Spanish Population (Movimiento Natural de la Población, MNP)*, thus excluding migrations. From that moment on, the data seemed to be better than in previous years, although there were still problems, especially at the beginning of the century, during the Civil War and in the following years (Mirón et al., 2002).

Since 1901, Spain has used the international classification of diseases for cause-of-death statistics. The first revision (of the Bertillon list) was in use between 1901 and 1909. The second revision had a long implementation period, from 1910 to 1930, because the third revision was never implemented. The fourth revision also had a long implementation period, from 1931 to 1950, as the fifth revision that was adopted by WHO in 1938 was never applied. The sixth revision was adopted in 1948 and represented the greatest impact on the classification of deaths since the beginning of the century. It was used in Spain from 1951 to 1965. The seventh revision was applied for a very short time (between 1966 and 1968). The eighth revision was adopted more quickly and was in effect for the period 1969 to 1979. The ninth revision ran from 1980 to 1998 and involved changes in the presentation of results and publication of new data. The 10th revision incorporated new data processing, including the online publication of the data since 1999.

Death and birth counts presented in the database come from the official statistics that are published in *Natural Movement of the Spanish Population*. Population counts are based on published census counts. For more recent years, these data are available in computerized files that are published on the INE website (http://www.ine.es/inebmenu/mnu_dinamicapob.htm). The format of the data varies over time. For more information, see the 'Background and Documentation' section of the Spanish data at the Human Mortality Database, written by Gleit et al. (2014).

Specific episodes in Spain's demographic history

Nadal (1976) noted that the mortality decline associated with the demographic transition began later in Spain than in other European countries. During the late 19th century, there were several historical events that had an important effect on the demographic statistics. These include a cholera epidemic in 1884 and the Spanish-American War (*La Guerra de Cuba*) in 1898. Events that affected the demographic trends in the 20th century include the 1918 influenza epidemic and the Spanish Civil War (1936-39 and post-war period 1941-42).

Spain experienced a baby boom from 1957 to 1967, and after 1975 the total fertility rate (TFR) began to decline at a faster rate. By 1981, the TFR was below replacement level. In recent years, there appears to have been a slight increase in fertility linked to immigration (i.e., immigrants having higher birth rates than the native Spanish population).

During the 1960s and up to 1974, many people emigrated from Spain to other countries. This trend reversed in and around 1976-1980. Since then, immigrants have come largely from other parts of Europe. Starting in the mid-1990s, the flow of immigrants from outside the European community increased. Following this, net migration peaked around 2007 and turned negative sometime around 2010.

Territorial coverage

The national territory of Spain includes the Peninsula, the Balearic and Canary Islands, Ceuta, and Melilla (Argüeso, 2003). The death and birth counts for the period 1861 to the present cover this same territory. Prior to 1975, the North African territories of Ceuta and Melilla (which form a very small part of the total population of Spain) were included, respectively, in the Cadiz and Malaga (or Granada) provinces. Since 1975, they have been listed as separate regions. Currently, Spain is divided into 50 provinces in 17 autonomous regions (*Comunidades Autonomas*) and two autonomous towns (Ceuta and Melilla).

Typically, the published census counts cover this same territory. However, the data by sex and age in 1940, 1950 and 1960 do not include Ceuta and Melilla. Thus, the census counts have been adjusted so that they are consistent with the death counts. The population counts for all the other census years cover the same territory as the death counts (Argüeso, 2003). See Gleit et al. (2014) for details.

Part I – Vital statistics and population censuses

1. Death count data

Coverage and completeness

As noted above, death data prior to 1975 refer to the *de facto* population. After this, data refer to the *de jure* (usually resident) population. The death counts include military deaths that occurred within Spanish territory (e.g., during the Civil War). However, deaths of Spaniards residing outside of Spain are not included in the vital statistics.

The place of residence and nationality is an important aspect. The INE microdata include deaths of foreign visitors in Spanish territory while the HMD excludes them. Therefore, foreign visitors must be removed from the INE data. To do this, we must look at two rating variables, which are different in each revision of the ICD:

- For the years 1980-1993: the variable is 'province of residence' (*provincia de residencia*, PROVRES) and the foreign population is identified with the code 53;
- For the years 1994-2008: the variable is 'province of residence' (PROVRES) and the foreign population is identified with the code 66;
- For the years since 2009: the variable is 'place of residence' (*lugar de residencia*, LUGRES) and the foreign population is identified with the code 2.

The data appear to be of reasonably good quality, except during a few specific periods (e.g., the Civil War and immediately afterwards). In the first part of the 20th century, there may have been some under-registration of births and deaths (Reher & Valero Lobo, 1995:90). During the Civil War (1936-39), registration of some deaths was delayed until 1941. There were age-heaping problems prior to the 1960s, which were worse for females than males. By 1975, many of the problems were resolved and the MNP is now considered a complete and reliable source of vital statistics (Del Campo & Navarro López, 1987).

Specific details: infant mortality

Prior to 1975, infant deaths were under-registered because the legal definition of a live birth excluded newborns that had died within the first 24 hours of life (for more details, see the section on 'Birth Count Data') (Gómez Redondo, 1992:4). This problem was less important early in the century because infant mortality was high and these 'false stillbirths' formed a smaller proportion of the total number of infant deaths. Yet, as infant mortality declined, the omitted infant deaths began to represent a substantial proportion of all infant deaths. Since 1930, the natural movement of the population has provided data on infant deaths in the first 24 hours of life, making it possible to correct for these false stillbirths. Based on these data, the under-registration of deaths in the first year of life is estimated to be about 5% in 1930 but reached 30% by 1974 (Gómez Redondo, 1992:9). The infant death counts for 1930-74 have been corrected in the HMD to include the 'false stillbirths'. See Gleit et al. (2014) for details.

2. Population count data

Coverage and completeness

Prior to 1960, most of the published census data refer to the *de facto* population. After this, most census data refer to the *de jure* (usually resident) population. Death counts refer to the *de facto* population until 1975, and until 1970 the *de facto* census counts are used. The *de jure* counts start with the 1981 census.

Prior to 1900, the census data were affected by under-counting, particularly for children of a young age and, in some cases, for women aged 50 and older. There may have also been under-counts for men of military age and problems of age heaping (the tendency to report ages ending in '0' or '5'). From 1900 to 1920, the census data progressively improved and the 1930 census counts are considered to be of good quality. However, in 1940, there was a tendency to over-count due to duplications, which continued until the 1981 census. For the 1981 and 1991 censuses, there is some suspicion of under-counting. The degree of underestimation has diminished over time (Reher and Valero Lobo, 1995: pp. 27-28). Despite these problems, Reher and Valero Lobo (1995: p.29) argue that Spanish census data are of good quality and as reliable as those of other countries in the region.

With the goal of providing international coherence to the Spanish data, the Human Mortality Database (HMD) data are used for the period 1980-2020. (www.mortality.org. Accessed: April 2023)

3. Birth count data

Coverage and completeness

As already noted, birth data covered the *de facto* population until 1975, when they began to cover the *de jure* population thereafter.

The definition of a live birth also changed in 1975. Prior to this, the law counted a live birth only if the newborn survived at least 24 hours after birth. In 1975, the statistical definition of a live birth was changed to conform to the international standard (Gomez Redondo, 1985, 1995, 1992).

There may have been some under-registration of births in the early 20th century (Leasure, 1963). However, data quality improved as the century progressed (Reher and Valero Lobo, 1995:90).

Specific details

As noted in the section on 'Death count data', data are available on infant deaths in the first 24 hours of life for the period 1930-1974.

These births comprise a very small proportion ($\approx 1\%$) of all live births, so this adjustment has little substantive effect. As data on false stillbirths are not available prior to 1930, the birth counts during that early period (1908-29) have not been corrected (i.e., newborns born alive but dying within the first 24 hours were excluded).

Part II – Information on CoD collection

Coinciding with the establishment of the Civil Registry law in 1957, INE proposed a new collection system to address the growing demand for demographic and sanitary information. New statistical death and stillbirth bulletins came into force. These were independent of those used by the Civil Registries. With the introduction of these bulletins, the information being gathered increased significantly. However, not all of the introduced variables could be exploited due to the fact that many informants struggled to fill them in. The death bulletins that are currently used correspond to this reform, by which not only was the content modified but also the format, in order to make the presentation more attractive and logical.

In 1980, the statistical bulletins on births and stillbirths were joined into one single bulletin, from which the data can be retrieved for live births and those who died before 24 hours. This revision involved modifying the presentation of results and publication of new data. For instance, the results at a regional (autonomous regions) level have been collected in a single volume since 1987, and the mortality rates have been published since 1985.

As a result of modernizing the Statistics on the Natural Movement of the Population, a process that began in 2007, a series of reforms to the Death Statistics was implemented in 2009. The changes aimed at improving the quality of the cause-of-death information by adapting its reporting and collection to WHO and Eurostat recommendations. With this in mind in 2009, the two documents gathering the causes of death (the Medical Death Certificate and the Statistical Death Bulletin) were unified into a single document: CMD/BED (García Ferruelo, 2007). Furthermore, information and communication technology was introduced into the process of collecting information: CMD/BED were adapted to a format by means of optical character recognition, which speeds up the process of capturing information.

4. Death certificate

The information about causes of death is collected on two statistical forms: the Medical Death Certificate/Statistical Death Bulletin (*Certificado Médico de Defunción/Boletín Estadístico de Defunción* – CMD/BED); and, for deaths that occur within the first 24 hours of life, the Statistical Birth Bulletin (*Boletín Estadístico de Parto* – BEP). Both forms collect the data by following the criteria established by WHO.

Regarding the BED, a single model was used (1975-1988) during the implementation of ICD-9. With the objective of improving data collection in the implementation of ICD-10, several models were used. ICD-10 started with a model that had a short validity period (1989- 1992) due to difficulties in its implementation. Then, the format was improved in a new model that was used from 1993 to 2008. The BED and CMD were unified with the modernization of the Statistics on the Natural Movement of the Population, which began in 2007. This model was implemented in 2009 and is currently used.

The BEP also had various modifications. The first model was used from 1975 to 1981. In 1980, the form started to distinguish between live births and stillbirths. This model was used until 1994. Since 1994, the cause of death has been reported on the BEP. In 1995, the model underwent a format change, which was used until 2006. Finally, the current model was established and came into force in 2007.

The bulletin sections for reporting causes of death are different in both models. In CMD/BED, four diseases that comprise the chain of events leading to death are described. In BEP, only two causes of death are stated – one related to the mother and another to the foetus.

The variables that are the subject of study include the underlying cause of death (UCD), selected according to the International Classification of Diseases (ICD), and the sociodemographic variables: age (or weeks of gestational age), sex, nationality and place of residence. The variables on the BED are defined and classified into the following groups (see also Table 1):

1. Time variables

- Primary variables: birth date and death date. The maximum detail for this information is the exact day.
- Derived variables: age and cohort. This is obtained by processing the primary time variables.

2. Geographic variables:

- Place of death (municipality)
- The deceased's place of usual residence

3. Personal variables

- Sex: male/female
- Marital status: single, married, widowed and divorced/separated.
- Job: coded and classified into 12 broad groups of the National Job Classification (*Clasificación Nacional de Ocupaciones - CNO*).

4. Underlying cause of death

UCD is currently classified according to the 10th ICD-10 revision up to the level of the 4-digit codes.

- Death: Any death of a person born alive, no matter how many hours he/she lived, is considered a death.

- Cause of death: Cause of death is defined as a set of diseases, pathological states or injuries that led to or contributed to the death, and the circumstances of the accident or violence that produced those injuries.
- Underlying cause of death: According to the WHO definition, the underlying cause of death refers to the disease or injury that initiated the chain of morbid events leading directly to death or the circumstances of the accident or violence that produced the fatal injury.

Table 1: Format of the Statistical Death Bulletin (BED), common variables from 1980 to date

Variables	Categories	Characteristics
Time		
Death date	Death date	Exact day
Birth date	Birth date	Exact day
Derived variables	Age and cohort	Completed age, 5-year age groups, reduces groups
Geographic		
Death place	Province and municipality	Municipality of death occurrence
Residence place	Province and municipality	Usual residence place
Personal		
Sex	Male, female	
Marital status	Single, married, widowed, divorced/separated	
Profesion	12 broad groups	Hi non-response rate
Cause of death		
Underlying cause of death	4-digit ICD codes	

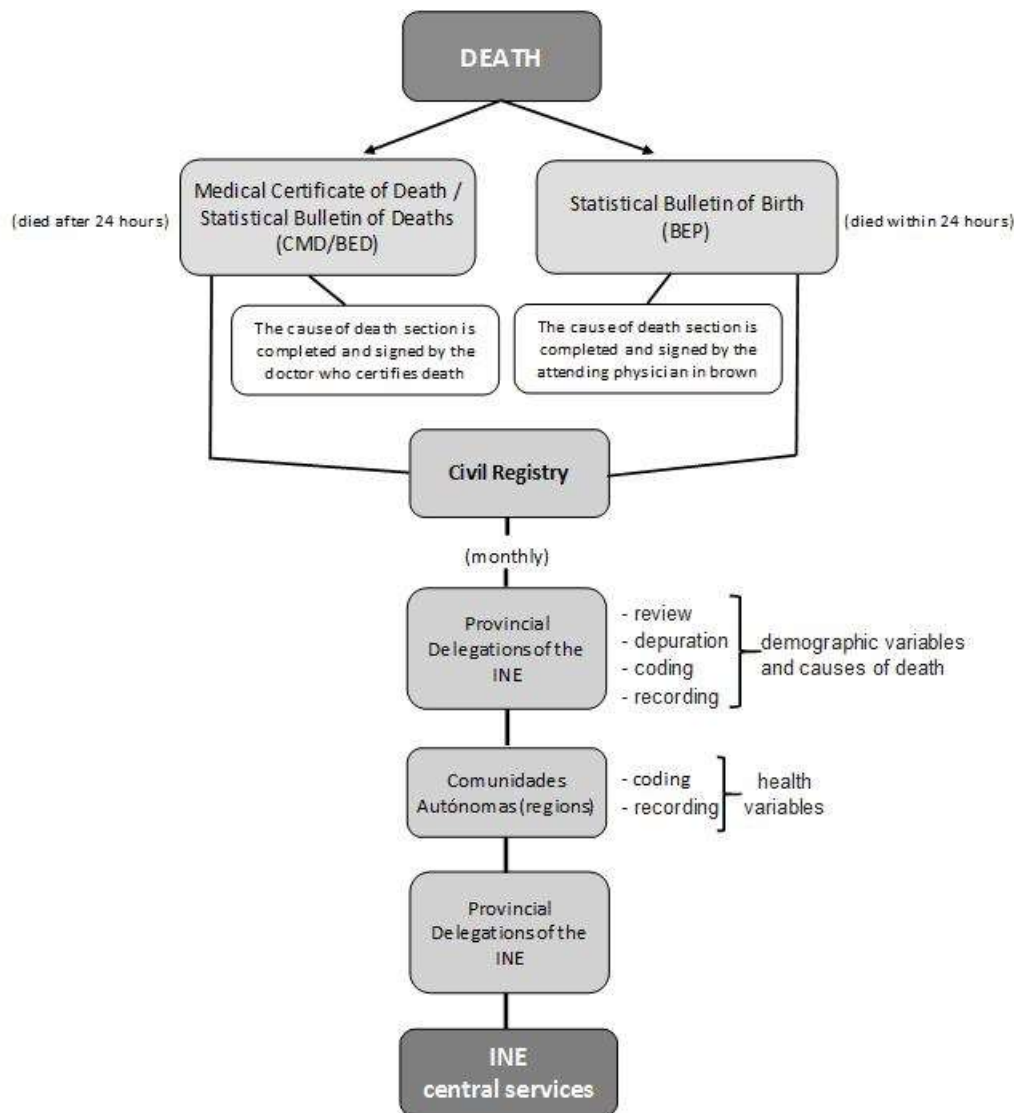
Source: Institute of Statistics of Andalucía.

The *Medical Death Certificate* (CMD) should be filled out by the medical doctor who certifies the death as well as the cause of death. The remaining (personal and geographic) data are recorded by relatives and people close to them. To facilitate the process in large municipalities, people can obtain the CMD and BED in pharmacies. Thus, the doctor can simultaneously complete the certificate and the BED section on the causes of death. In other municipalities, funeral homes provide the bulletins.

5. Cause-of-death coding

The Civil Registries send the filled-in BED and BEP forms on a monthly basis to the INE Provincial Delegations, where the demographic variables are revised, cleaned up, coded and registered. This information is used to immediately update other statistical files. Autonomous regions are in charge of the process of coding and registering the health variables of the deaths that have occurred in their territory. The INE Provincial Delegations proceed with the selection and coding of the underlying cause of death (UCD). The Provincial Delegations provide the autonomous communities with monthly information in electronic format, depending on whether they have a collaboration agreement for both statistics. Then, the BED and BEP are sent back, with the health variables included, to the INE Provincial Delegations. The statistical bulletins are then sent to the INE Central Services together with the corresponding computer files transmitted by Electronic Data Interchange (Figure 1).

Figure 1. The registration process of the Spanish death certificate



The INE coordinates the coding work of every autonomous community through assessment and the provision of technical support, as well as by continuously following up in order to guarantee the homogeneity of data collection. Additionally, the registering process followed by the autonomous communities must respect the format and norms established by the INE.

6. Classifications in use and collected data

Table 2 provides an overview of classifications used in Spain from the beginning of the 20th century. The **first revision** of the so-called **Bertillon list** was formally applied to the period 1901 to 1909. This period is characterized by the modernization and generation of the MNP data. In the framework of this update, several variables were standardized: age, sex, province of registration¹ and, from 1903,

¹ From 1900, the death statistics by cause and sex were published at the same level of detail as those corresponding to all Spain and for each province for the periods 1900 to 1907, and 1919 to 1939. It is noteworthy that there are abbreviated

classifications by cause of death, all of which were combined with age groups (Coll and Fortea, 1995). Joint-frequency tables were published for 1900-1902. For instance, for deaths in reference to the state, province and provincial capitals, the joint frequencies of deaths were given by seasons and age group or by season and causes of death. However, the joint frequencies of mortality by age group and cause of death were not published until a few years later. One of the weakest points during this period is the high number of cases classified as 'other diseases' or 'unknown and ill defined', something that makes a detailed analysis for this period difficult (Sanz, 1995).

Table 2: Revisions of the ICD and its implementations in Spain

Revisions	Adopted by WHO	Used in Spain	Number of items	Categories		Total
				External causes	Reasons for health assistance	
First	1900	1900	157	22		179
Second	1909	1910	157	32		189
Third	1920	-	166	39		205
Fourth	1929	1931	164	36		200
Fifth	1938	-	164	36		200
Sixth	1948	1951	769	153	88	1010
Seventh	1955	1966	800	153	88	1041
Eight	1965	1969	858	182	48	1088
Ninth	1975	1980	909	192	77	1178
Tenth	1989	1999	1575	373	82	2032

Source: Rodríguez Jústiz et al. (2000).

The **second revision** had a long usage period, from 1910 to 1930, due to the fact that the third revision was never implemented. Spain standardized death rates in 1930, thus allowing more precise comparisons among the different capitals and provinces. The data were presented in two lists: one detailed by gender with 189 causes of death, and the other detailed by gender and age with 38 causes of death (Blanes, 2007).

The **fourth revision** also had a very extensive duration, from 1931 to 1950, because the fifth revision (approved internationally in 1938) was never implemented due to the Civil War and the difficulties of the post-war period. This also led to a lack of precision in the census and limited accuracy of the causes

classifications for province capitals for the period 1908 to 1918, but not for the whole provincial group. Furthermore, the absence of available information for deaths by exact age for the year 1906 is especially notable.

of death (Gómez Redondo, 1992; Mirón et al., 2002). The numbers of deaths corresponding to 1936-1939 MNP are the most doubtful of the 20th century, as violent deaths are not properly registered.

The **sixth revision** was approved by WHO in 1948 and it had a considerable impact on the continuity of existing death statistics by cause. The 6th revision was implemented in Spain for the period 1951-1965. This classification changed in name and size: from *Classification of causes of death* (with approximately 200 items) to *Classification of diseases and causes of death*, (with more than 1,000 categories).

The **seventh revision** was implemented in Spain over a short period, between 1966 and 1968.

In 1975, Spain underwent major constitutional and administrative reforms. Autonomous statistical offices were founded and gradually took over several administrative competencies, including the compilation of mortality statistics. Around the same time, the **eighth revision** (approved by WHO in 1965) was implemented. The 8th revision was in use during the period 1969 to 1979.

Of the major conceptual changes from the 1970s, the most important change relates to the statistical definition of stillbirths (INE, methodology; Gómez, 1992). Regarding causes of death after 1975, the extended use of informatics for treating and exploiting vital statistics allowed for detailed publication of the full 4-digit ICD-8 cause-specific mortality². Another important change was the classification of deaths by place of residence instead of place of death. From the same year, cause-specific statistics were published in an independent monograph.

The **ninth revision** was adopted by WHO in 1975 and implemented in Spain from 1980 to 1998. During the 9th ICD revision, the statistical bulletins of births and stillbirths were joined into one single form, from which the data for live births and deaths before 24 hours can be retrieved. Changes were also introduced to the publication system; for example, mortality rates have been published along with the data³ since 1985, and the results at the autonomous regional level have been compiled in a single volume since 1987.

Finally, the **tenth revision** was adopted by WHO in 1989 and implemented in Spain as of 1999. Compared to the ICD-9, the ICD-10 is an alpha-numeric system with double the number of items. It also introduced changes to the rules for selecting the underlying cause of death (Robles, 2009:574). As of 1999, the cause-specific data has been available online. The data are published at four regional levels: national, autonomous, provincial and capitals. In addition, further distinction is made for mortality type: general mortality, infant mortality, perinatal mortality and late foetal deaths.

Several changes also occurred in the provided age-group detail throughout the period covered by mortality data. Before 1933, the last open-age interval was 60 years (Mirón et al., 2002). From 1934, the threshold was extended to over 70, and the last age group of over 85 years was established in 1950 (Blanes, 2007).

² As mentioned in the 1969 edition of 1966 MNP statistics, the processing of the information on deaths was one of the first computerized statistical operations.

³ Until 1985, only data about proportional mortality (cause-specific deaths per 1,000) or crude death rates (per 100,000 inhabitants) were published. Since then, age-standardized mortality rates are computed and published at the regional level using both direct and indirect standardization methods.

6.2 Collected data

The overview of the collected data is provided in Appendix 1 at the end of this document. Regarding technical aspects of the data collection, the data for the period 1980-1998 were provided in .txt format, while the data for 1999-2008 were coded into an internal INE format (converted later to text). The format of the ICD items slightly differed between ICD-9 and the ICD10:

- For ICD-9 (1980 - 1998): The cause-of-death code is divided into two parts. The variable CODCAU1 provides the first three digits of the cause of death (numeric format). The fourth digit is provided in the variable CODCAU4.
- For ICD-10 (1999 - present): The cause of death is divided into three parts. The first one (CODCAU1) contains the alphabetic character, the second one (CODCAU23) lists the last two numeric digits and the third one (CODCAU4) adds the fourth numeric digit.

6.3 Data sources

We use computerized data for the period between 1980 and 2012. The data come from the INE series, 'Deaths by Cause of Death'. These are anonymized microdata files containing a four-digit code for cause of death. They are not granted public access and were formally requested from INE.

6.4 Specific treatment of the raw data

Before the reconstruction itself, the data were carefully checked for mistakes in coding or tabulation and for improbable combinations of age, sex and cause of death (which typically occur in senility, sex-specific diseases or infant mortality codes). Find the table of detected cases in the Appendix 2.

Codes that are not to be used as underlying causes of death (the so-called asterisk codes) represent a specific issue in the quality of mortality statistics. We checked both ICD-9 and ICD-10 data for the presence of asterisk codes and, although these codes were not observed in ICD-10, there were many cases in ICD-9. Asterisk codes were replaced by corresponding dagger codes and, in cases where this was not possible, they were recoded to the ill-defined group.

We also had to deal with the updates that are implemented regularly in ICD-10. In the framework of the ICD updates, some ICD-10 three-digit items are further sub-divided at the 4-digit level (such as A09 or R95). In some items, the 4-digit level is re-arranged (such as in the case of C82 or N18) or a brand new 3-digit item is introduced (such as C86 or J09). These changes were made to different items in different years. To obtain comparable time series for the ICD-10, we attributed the new (updated) items to their hypothetical 'target' items. Appendix 3 summarizes the solutions adopted for individual cases.

7. Specific transition documents

To establish correspondence between ICD-9 and ICD-10, we used a correspondence table developed by the Madrid Autonomous Region.

Part III – Reconstruction information

8. Reconstruction information

In order to reconstruct the series of causes of death in Spain, we used the methodology that Jacques Vallin and France Meslé proposed for the reconstruction of causes of death in France.

The first step in the method is to build a correspondence table between all elements of the two revisions. We based our correspondence table on the Madrid Community Autonomous correspondence table (mentioned above).

The correspondence table was created based on the mortality data of four years of transition (from ICD-9 in 1997-1998 to ICD-10 in 1999-2000). This means that only those causes from which a death actually happened were considered. To solve this problem, the lists of items used for Spain in 1997-98 (ICD-9) and 1999-2000 (ICD-10) were matched with WHO lists and all of the missing items were integrated into the Spain-specific lists.⁴ For these additional items, correspondences were established using the WHO ICD-9/ICD-10 translator.

After adjusting the standard lists for ICD-9 and ICD-10 to the WHO lists, and after the treatments mentioned in section 6.4, we worked with a total of 5,248 items in the ICD-9 list and a total of 10,218 items in the ICD-10 list.

In the second step, we built the fundamental associations of items, which gather the same medical content over the two revisions and ensure statistical continuity. This step began in the spring of 2011 and was finished in November 2013. We worked with the data covering the four years around the transition: 1997 and 1998 for ICD-9 and 1999 and 2000 for ICD-10.

While building the fundamental associations, we verified that time series for groups of items united under one association have a smooth trend without disruption in the year of transition. This was done with an R-program called 'Check Disruption', designed by Giancarlo Camarda. If the program detected statistically significant disruption, the associations were reviewed again and modified so as to eliminate the tendency toward disruption. This step was repeated each time a modification to the correspondence table was introduced.

However, there were several problems regarding the use of the correspondence table built by the Madrid Community Autonomous Region and the WHO Translator:

1) Some associations were too big and bound up items from different ICD chapters. To solve the first problem, all of the associations were re-examined by Aina Faus Bertomeu with the use of a specially designed Java application that employed data from the adjacent transition years (from 1997 to 2000) and ensured medical as well as statistical coherence. In the end, many unnecessary links were cut, the number of associations increased and the most complex associations were simplified.

2) Errors were observed in associations of type 1:1 of the original correspondence table. In principle, such associations should have statistical and semantic consistency. However, this did not happen in many cases for two reasons. First, the correspondence was misspelled, i.e., the ICD-9 and ICD-10 codes did not have similar names and there was therefore no semantic coherence. To resolve these errors, we used the ICD manuals in order to find the correct headings. Second, even after matching the codes

⁴ In the end, we obtained 5,365 different codes for ICD-9 and 10,636 different codes for ICD-10.

and names of both ICDs, there was still statistical inconsistency. In these cases, the items were associated to more relevant items in order to achieve coherence.

3) For some ICD-9 and ICD-10 codes, the correspondences were not defined by the original correspondence tables (marked as 'undef'). This is mainly because these causes were introduced in ICD-10 and were not taken into account in ICD-9. The modus operandi in these cases was to find the relevant correspondences in the ICD-9 and ICD-10 manuals.

4) We also found errors in complex associations (1:N; N:1; N:N), typically by not achieving the expected statistical or semantic consistency. In these cases, a protocol to follow was developed.

5) Some of the ICD-10 codes present in the Madrid Community Autonomous correspondence table reflected the ICD-10 modifications that were made in 2003 (i.e., M723, M725, Q314, Q350, Q352, Q354, Q356, Q358, R500, R501, X342, X343, X344, X345, X346, X347). These correspondences were neglected, as they appeared after the transition period.

6) A specific issue was observed for the category of secondary malignant neoplasms, which did not exist in the ICD-9 data. The problem lay in finding corresponding deaths for ensuring statistical and medical continuity. We decided to group them with primary tumours, but the result does not always have statistical consistency.

7) Another problem (observed even at the level of the three-digit items) involved transferring between mental and nervous diseases in the ICD-10 and ICD-9.

In order to solve these problems, we consulted researchers from the MODICOD project and health professionals from the Carlos III Institute of Health in Madrid.

After the modifications, the final correspondence table resulted in 3,512 fundamental associations. Table 3 summarizes the distribution of the associations by type.

Table 3: Distribution of fundamental associations of items by type and death counts between the ICD-9 revision of 1980 and ICD-10 revision of 1999.

Association type	Associations	
	Number	Proportion, %
type 1:1	2161	62
type 1:n	110	3
type n:1	772	22
type n:n	469	13
Total	3512	100

In the next step, we computed the transition coefficients. This was done using a Visual Basic macro developed by Olga Penina. If the program detected inconsistencies in the calculation, the links in the association needed to be changed.

The next step was the reconstruction itself, which was done by multiplying the coefficients and the data by items in the old revision for the years 1980 to 1998. The result was checked by testing statistically significant disruptions in the year of transition. If the result was unsatisfactory, either the correspondence in the association or a transition coefficient at the given age was changed. After this change, all of the stages were repeated.

A posteriori corrections

The totals for all causes were compared with the death counts from the Human Mortality Database (HMD). The deviations were very small and appeared mainly because of rounding. In order to adjust the death counts to the HMD, we redistributed differences proportionally among all of the causes separately for each age group.

9. Redistribution of ill-defined causes of death

The redistribution of ill-defined and unknown causes (except cause R95_ 'Sudden Infant Death Syndrome') was done proportionally among all other causes.

References

- ARGÜESO, A. (2003): *Personal Communications*, via e-mail, June 6th and 23rd.
- BLANES LLORENS, Amand (2007): *La mortalidad en la España Del Siglo XX. Análisis demográfico y territorial*, Universitat Autònoma de Barcelona, PhD Thesis.
- COLL, Sebastián y FORTEA, José Ignacio (1995): *Historia Económica de España. Recursos y Sectores Productivos*, Vol. 1, Banco de España - Servicio de Estudios de Historia Económica, nº 32, Madrid
- DEL CAMPO, S. NAVARRO LÓPEZ, M. (1987): 'La estadísticas de población'. Chapter 6 in: *Nuevo Análisis de la Población Española*. Barcelona: Editorial Ariel, S.A.
- GARCÍA FERRUERO, Margarita (2007): 'Causas de Muerte, una Estadística Básica en Continua Evolución', *Revista Índice*, 20:16-18
- GLEI, Dana; GÓMEZ REDONDO, Rosa; ARGÜESO, Antonio & CANUDAS-ROMO, Vladimir (2014): 'About Mortality Data for Spain', *Human Mortality Database*, Background and Documentation, [online] <http://www.mortality.org/cgi-bin/hmd/country.php?cntr=ESP&level=1> (Last Update: Tim RIFFE, 01/04/2014)
- GÓMEZ REDONDO, Rosa & BOE, Carl (2005). 'Decomposition Analysis if Spanish Life Expectancy at Birth: Evolution and Changes in the Components by Sex and Age'. *Demographic Research*, 13-20: 521-546
- GÓMEZ REDONDO, Rosa (1985). 'El descenso de la mortalidad infantil en Madrid, 1900-1970'. *REIS* 32: 101-139
- GÓMEZ REDONDO, Rosa (1992): *La mortalidad infantil española en el siglo XXI*, Centro de Investigaciones Sociológicas, Co-Editorial Siglo XXI, pp.451
- GÓMEZ REDONDO, Rosa (1995). 'Vejez Prolongada y Juventud menguada'. *REIS* 71-72:79-108.
- GÓMEZ REDONDO, Rosa (1997). 'La mortalidad de la España actual'. *Política y Sociedad* 26:41-62
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE): *Metodología de la recogida y tratamiento de las causas de muerte* [online] http://www.ine.es/daco/daco42/sanitarias/metodologia_00.pdf
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE) (2003): *Short History of INE*. Retrieved 24th July. (<http://www.ine.es/info/historia.htm>).
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE) (2005): *Population Projections Calculated from the 2001 Population Census, Hypothesis on the Evolution of Demographic Phenomena According to the Different Situations Considered*. Retrieved 29th November 2005 (http://www.ine.es/en/metodologia/t20/t2030251h_en.htm).
- LEASURE, W. (1963): 'Factors Involved in the Decline of Fertility in Spain, 1900-1950'. *Population Studies*, 3: 217-285.
- MIRÓN J.A; ALONSO, M.; MÉNDEZ, M.; LÓPEZ, I.; SÁENZ M. A. (2002): 'Mortalidad Prematura por Enfermedades Infecciosas en España 1908- 1995', *Rev. Panamericana Salud Pública*, 12(4):262- 268
- NADAL, J. (1976): *La Población Española (Siglos XVI a XX)*. Barcelona: Ariel.
- RODRÍGUEZ JUSTÍZ, Félix; FERNÁNDEZ NÚÑEZ, Andrés & BALY GIL, Alberto (2000): 'Apreciaciones Sobre la Clasificación Internacional de Enfermedades', *Revista Cubana de Higiene y Epidemiología*, 38(3):215-9.
- REHER, D-S., VALERO LOBO, A. & García Sestafe, J.V. (1995). 'Fuentes de Información Demográfica en España'. *Cuadernos Metodológicos*, No. 13. Madrid: Centro de Investigaciones Sociológicas.
- ROBLES GONZALEZ, Elena (2009): '¿De qué se mueren los ancianos en España?', *Estudios Geográficos*, 267(LXX):567- 598.
- SANZ GIMENO, Alberto (1995): 'Infancia, mortalidad y causas de muerte en España en el primer tercio del siglo XX (1906- 1932)', *REIS* 1:129- 154.

List of acronyms

INE – National Statistics Institute (*Instituto Nacional de Estadística*)
MNP - Natural Movement of the Population (*Movimiento Natural de la Población*)
CMD - Medical Death Certificate (Certificado Médico de Defunción)
BED - Statistical Death Bulletin (Boletín Estadístico de Defunción)
BEP - Statistical Birth Bulletin (Boletín Estadístico de Parto)
CNO - National Job Classification (Clasificación Nacional de Ocupaciones)
WHO – World Health Organization
ICD – International Classification of Diseases
HMD – Human Mortality Database

Appendixes

Appendix 1. Overview of the collected data for Spain.

Revisions	Totals	Data	Sex	Age	Marital Status	Causes	Profession	Season	Nationality	Territorial Coverage
CIE-1 (1900- 1909)	√		√	√ (1) (2)		√ (3)	√ (4)	√ (5)		State, province and capitals (6)
CIE- 2 (1910-1930)	√		√	√		√ (7)	√	√		State, province and capitals (8)
CIE- 3	–	–	–	–	–	–	–	–	–	–
CIE-4 (1931- 1950)	√		√	√ (9) (10)		√ (11)	√	√		State, province and capitals
CIE-5	–	–	–	–	–	–	–	–	–	–
CIE-6 (1951- 1965)	√		√	√		√	√	√		Province and capitals (12)
CIE-7 (1966- 1968)	√		√	√		√	√	√		Province and capitals
CIE- 8 (1968- 1980)	√ (13) (16)	√	√	√ (14) (16)	√	√ (15) (16)	√ (17)	√		State, province and capitals (16) (18)
CIE-9 (1980- 1998)	√ (19)	√	√ (19)	√ (19)	√	√ (19)	√	√		State, province, autonomous community (20), capitals
CIE-10 (1999- to date)	√	√	√	√	√	√	√ (21)	√ (22)	√	State, province, autonomous community (20), capitals

1. Lack of information available on the exact age for deaths from 1906.

2. After 1903, the cause of death classifications combined with age groups are available.
3. A high number of cases were classified as 'other diseases' or 'unknown and ill-defined'. For the period from 1902 to 1907, the crosstabs focused on the cause of death by sex, age and place.
4. After the early 20th century, data was given for deaths by occupation, although at first without crossing this information with ages.
5. From 1900 to 1902, crosstabs (double-entry table) were included. For example, frequency stations and mortality by age group or by stations and causes of death were given.
6. Only the period 1908-1918 gives the causes of death for the provincial capitals, but not for each province. The periods 1900-1907 and 1919-1939 are only provincial.
7. Introduction of crosstabs (double-entry table) from frequencies of mortality by age and cause of death.
8. From 1930, standardized mortality rates were applied to allow tighter comparisons between capitals and provinces.
9. Available data of causes of death combined with age groups for each province for the periods 1903-1907 and 1930-1939.
10. From 1933, the group 65 and over were unified (before, they were limited to 60 or older).
11. The Civil War and the post-war period entailed a lack of rigor and precision regarding the causes of death. Still, for the period 1931 to 1940, there are crosstabs that focus on the causes of death by sex, age and place.
12. Since 1960, a high proportion of women have started having children in hospitals in the capital, where these births are registered. This causes an under-registration of births and deaths in rural populations.
13. Since 1975, those born alive but die before 24 hours are considered as deaths and not stillbirths, as before.
14. Variables derived from age: age, five-year groups, small groups and large groups.
15. Since 1975, the basic cause of death is classified into four digits.
16. Since 1975, the crosstabs focus on the cause of death by sex, age and place.
17. A large percentage of non-response.
18. After 1975, the place of death registration and habitual residence differs.
19. Until 1985, only published data on proportional mortality (mortality by cause of death by 1000) or average rate (deaths by cause per 100,000 population). After this, average rates are based on age and sex structure (direct method) or a common mortality (standardized mortality ratio), giving standardized rates by age and sex.
20. Since 1987, the regional results (autonomous community) are collected in a single volume, before they were edited as appendices.
21. Encoded by a reduced list of the National Classification of Occupations (*Clasificación Nacional de Ocupaciones*, CNO). This database is characterized as incomplete.
22. Includes all deaths that occur in the state territory, independently of the place of origin. Deaths of Spanish nationals or residents in Spain that occurred outside of Spain are not collected

Appendix 2. Recoded age – specific causes.

Year	Sex	Origin cause	Target cause	Age groups							
				0-5	5-10	15-20	30-35	35-40	40-45	45-50	60-65
1980	male	R54_	R99_				1	1			
1980	female	R54_	R99_					1		1	
1981	female	R54_	R99_						1		
1985	male	G20_	G219			1					
1988	male	G20_	G219		1						
1990	female	F531	F59_								1
2000	male	Q000	Q899			1					
2015	male	G20_	G219			1					

Appendix 3 List of modifications to the ICD-10 data regarding classification updates

Original cause	Target cause	Type
A09_	A099	obsolete
B485	B488	dragger
B970	B340	non-UCD
B972	B342	non-UCD
B973	B343	non-UCD
B974	B348	non-UCD
B975	B348	non-UCD
B976	B343	non-UCD
B977	B344	non-UCD
B978	B348	non-UCD
C80_	C809	obsolete
C832	C839	obsolete
C834	C839	obsolete
C836	C839	obsolete
C842	C849	obsolete
C843	C849	obsolete
C850	C859	obsolete
C912	C919	obsolete
C932	C939	obsolete
C941	C947	obsolete
C945	C947	obsolete
C952	C859	obsolete
C961	C969	obsolete
C963	C969	obsolete
C97_	C969	non-UCD
D463	D469	obsolete
D752	D759	obsolete
D760	D763	obsolete
F100	X45_	non-UCD
F110	X42_	non-UCD
F120	X42_	non-UCD

F130	X41_	non-UCD
F140	X42_	non-UCD
F190	X40_	non-UCD
G903	G909	obsolete
I150	I139	non-UCD
I151	N289	non-UCD
I152	E349	non-UCD
I158	I139	non-UCD
I159	I139	non-UCD
I220	I212	non-UCD
I221	I212	non-UCD
I228	I212	non-UCD
I229	I212	non-UCD
I252	I258	non-UCD
I48_	I489	obsolete
I848	K649	obsolete
I849	K649	obsolete
J987	J989	error
K350	K358	obsolete
K351	K358	obsolete
K359	K358	obsolete
K511	K519	obsolete
K553	K559	error
K581	K580	error
K588	K580	error
K662	K669	error
K85_	K859	obsolete
L412	L419	obsolete
L89_	L899	obsolete
M725	M729	obsolete
N180	N189	obsolete
N188	N189	obsolete
O96_	O969	obsolete
O97_	O979	obsolete
P703	P969	non-UCD
P704	P969	non-UCD
P708	P969	non-UCD
P709	P969	non-UCD
P720	P969	non-UCD
P722	P969	non-UCD
P729	P969	non-UCD
Q314	Q319	obsolete
R003	R008	error
R170	R17_	error
R500	R509	obsolete
R501	R509	obsolete
R95_	R959	obsolete

