

OFFICIAL STATISTICS

# The Statistics on Causes of Death: characteristics and improvements

Margarita García Ferruelo<sup>1</sup>, M. Rosario González García<sup>2</sup>

<sup>1</sup> Instituto Nacional de Estadística, margarita.garcia.ferruelo@ine.es

<sup>2</sup> Instituto Nacional de Estadística, rosario.gonzalez.garcia@ine.es

Received: November 3, 2022. Accepted: December 26, 2022.

---

**Abstract:** The Statistics on Causes of Death is a key tool for the Public Health. This article describes the complex process of the statistics, the advances achieved in recent years, such as the implementation of an international automatic system for coding multiple causes of death and for the selection of the underlying cause (IRIS) or the improvement in obtaining the external causes of death, as well as its usefulness for the epidemiological studies and health research. It is also discussed some of the lessons learned during the worst pandemic period, that, without any doubt, have highlighted the need of a more efficient method to get information through the implementation of an Electronic Death Certificate. And, finally, it is proposed to collect other variables of interest for the analysis of the causes of death using available administrative sources.

**Keywords:** underlying cause of death, multiple cause of death, International Classification of Diseases, coding, IRIS System, IMLweb, administrative data

**MSC:** 62P10, 62P25, 91B82

---

## 1 Introduction

This work is part of the review process within the strategy of the European Statistical System (ESS) to implement one of the recommendations of the Code of Good Practices (CoP) through the evaluation of key statistics by external experts. For this purpose, a session has been included within the framework of the XXXIX Spanish conferences on Statistics and Operation Research and of the XIII Conference of Public Statistics to present the Statistics on causes of death methodology, with two speakers from Complutense University of Madrid and the National School of Health of the Carlos III Health Institute.

The Statistics on Cause of Death is an annual operation legally supported by Regulation (EU)

328/2011 of the Commission, of April 5, 2011, which develops Regulation (EC) 1338/2008 of the European Parliament and of the Council regarding community statistics on public health and safety and health at work in the field of statistics on causes of death, which establishes the commitments acquired by the Member States and Eurostat in relation to the statistics on Causes of Death, Commission Regulation (EC) 328/2011 specifies the scope, the definitions, the list of variables, the reference period, the deadline to send the data and the set of variables and metadata to be provided to Eurostat.

The main objective of the statistics is to know the pattern of mortality. Information on causes of death is the traditional study that has allowed us to know better the health situation since the 19th century and continues to have great potential as a tool for decision-making in public health. This statistic began based on a list of 5 diseases. Later, the cause of death was classified according to a list of 99 diagnoses, origin of the International Classification of Diseases (ICD) of the World Health Organization (WHO) and nowadays there is the tenth revision of this classification (ICD- 10), containing more than 12,000 diagnostic codes, which allows classifying the mortality causes with a high degree of clinical specification (OPS 1995).

The purpose of the causes of death data is to know the pattern of mortality associated with each sex, group of age and geographical areas, as well as its evolution over the time. The key variable is the underlying cause of death, which is selected following the ICD-10 criteria from the diseases informed by the physician in the Medical Death Certificate. (INE 2020)

The study of mortality based on the underlying cause of death constitutes a great value tool; in fact, it has allowed to detect changes in the trend of certain diseases. Four decades ago, cardiovascular diseases were the main cause of almost half of deaths, but their relative weight has been decreasing since then (26% in 2021) and, although they remain the leading cause of death, they have been displaced for other causes such as tumours, dementia, or Alzheimer's disease. These last two diseases mentioned have been the main cause of more than 30,000 deaths in 2021, 66% more than in 2005. And with the outbreak of the pandemic, this pattern has undergone some new changes.

Another important contribution of the underlying cause of death, with relevant social and economic impact, is to identify the "guilty" causes of premature deaths and quantify the years of life lost (INE 2021) due to these causes. Violent deaths—that is, traffic accidents, poisonings, falls or other accidents, suicides, or homicides—are the causes that deprive of more years of life. People who die of a violent cause, live, on average, 27 years less than the years they would theoretically have to live. And if we refer to tumours, we would be talking about 12 years of life lost. However, if Statistics on causes of death were limited to the study of the underlying cause, the opportunity to identify the associations of the most frequent pathologies and to know the true dimension of mortality would be lost. An aging population, like the case of Spain, have a significant increase of chronic diseases in which several pathologies converge, although they do not lead to death, they can contribute to hastening it. This is the case, for example, of diabetes or hypertension.

When one of these diseases is informed on the certificate, its selection as underlying cause will depend on the rest of the diseases that have also been mentioned. For example, if hypertension and COVID-19 appear together, depending on the order in which both diseases are informed, the ICD-10 selection rules may penalize hypertension by displacing it in favour of COVID-19. Thus, the opportunity to know and analyse the true dimension of hypertension in mortality would be lost, apart from to identify the most frequent associations with other pathologies and, therefore, to adopt more effective prevention action measures. Another recent example of the importance of the multiple causes of death availability (INE 2021) has been their effect on respiratory diseases with the irruption of the COVID-19. Respiratory system diseases have been displaced as underlying cause by the COVID-19 when both were informed in the same certificate. Thanks to the multiple causes, the

impact of respiratory diseases on mortality was not lost and it was possible to obtain information on the complications derived from COVID-19 and the comorbidities presented by the people who died. Apart from that and without any doubt, the usefulness of the causes of death data is also linked to the continuous quality improvement to be able to respond to new demands.

In recent years, great efforts have been made to improve the quality of these statistics. An international automatic system for coding (IRIS) has been implemented and, the Institutes of Legal Medicine and Forensic Sciences have been incorporated as a source of information of the causes of deaths in case of deaths with judicial intervention.

In terms of responding new demands, the Causes of Death Statistics offer very precise and high-quality information, however it does not provide information in the short term as evidenced during the COVID-19 crisis. This fact is due to the administrative steps required by the Civil Registers before sending the Death Certificate (paper document with the needed information) to the NSI and its complex process of mass scanning and OCR review. The solution is found in the Electronic Death Certificate, a project framed in the digitization of the Civil Register as established by Law 20/2011 and in which the NSI collaborates with the Civil Register. Nevertheless, Causes of Death Statistics is the source of numerous epidemiological studies and health research. The pandemic, that has given visibility to epidemiology, has also made this Statistics more visible, proof of this is the huge number of accesses to its results. In 2019 there were around 260,000 queries and since 2020 they are approaching two million.

## 2 Statistical processes in the pandemic context

When the COVID-19 pandemic emerged, the information on mortality that was becoming known referred to estimated mortality data (Daily Mortality Monitoring System - MoMo - and Experimental Statistics on weekly estimations of deaths) and mortality declared to the National Epidemiological Surveillance Network (RENAVE). On the one hand, the mortality declared by COVID-19, provided by the Ministry of Health from the registers of the Autonomous Communities, met the criterion of deaths with a positive COVID test without distinguishing the direct cause of these deaths and, on the other hand, the estimated mortality made it possible to know the excess mortality, but without having information on the causes of death.

This daily information was essential for the epidemiological surveillance of the pandemic; however it was not enough to know the real impact of COVID-19 on mortality. The source that would provide the best estimation of mortality attributable to the pandemic would be the Statistics on Causes of Death.

The Death Certificate, document in which the physician informs the sequence of diseases that finally lead to death, is the source of information for the Causes of Death Statistics.

The Civil Registers send the death certificates to the Provincial Delegations of NSI to be scanned. These certificates are designed for optical recognition. Taking into account that the terms to be recognized are diseases, their recognition is quite complicated, for that reason it is necessary to develop a diseases dictionary from the information provided by physicians in the death certificates over the years. Currently, this dictionary has around 170,000 different terms and the recognition success level is around 85-90%. Some difficulties had to be faced during the health crisis, such as the collapse in the Civil Registers during the first wave of the pandemic, a greater volume of deaths

**OMC** ORGANIZACIÓN MÉDICA COLEGIADA DE ESPAÑA CONSEJO GENERAL DE COLEGIOS OFICIALES DE MÉDICOS **CERTIFICADO MÉDICO DE DEFUNCIÓN**

 <https://certificados.ogcom.es> **Id. Electrónico** **Colegio de** **Nº Certificado**

**CLASE 3ª SERIE A**

D. / Día: \_\_\_\_\_ en Medicina y Cirugía, colegiado/a en \_\_\_\_\_, con el número \_\_\_\_\_ y con ejercicio profesional en \_\_\_\_\_

**CERTIFICO la defunción de**

Nombre del fallecido/a: \_\_\_\_\_  
 1º Apellido del fallecido/a: \_\_\_\_\_  
 2º Apellido del fallecido/a: \_\_\_\_\_

Fecha de nacimiento: Día \_\_\_\_ Mes \_\_\_\_ Año \_\_\_\_ Sexo: Varón  Mujer

Documento de identidad:  D.N.I. Número: \_\_\_\_\_  Pasaporte Número: \_\_\_\_\_  N.I.E. (Tarjeta de Residencia) Número: \_\_\_\_\_

Hora y fecha de la defunción: Hora: minutos \_\_\_\_:\_\_\_\_ Día \_\_\_\_ Mes \_\_\_\_ Año \_\_\_\_  
 ¿En qué municipio ocurrió la defunción?: \_\_\_\_\_  
 Domicilio particular  Centro hospitalario  Residencia socio-sanitaria  Lugar de trabajo  Otro lugar

**Causas de defunción (ver instrucciones en página 2)** Intervalo de tiempo aproximado<sup>1</sup>

**I. Causa inmediata<sup>2</sup>**

(a) \_\_\_\_\_ Debido a \_\_\_\_\_ Horas \_\_\_\_ Días \_\_\_\_ Meses \_\_\_\_ Años \_\_\_\_  
**Causas intermedias<sup>3</sup>**  
 (b) \_\_\_\_\_ Debido a \_\_\_\_\_ Horas \_\_\_\_ Días \_\_\_\_ Meses \_\_\_\_ Años \_\_\_\_  
 (c) \_\_\_\_\_ Debido a \_\_\_\_\_ Horas \_\_\_\_ Días \_\_\_\_ Meses \_\_\_\_ Años \_\_\_\_  
**Causa inicial o fundamental<sup>4</sup>**  
 (d) \_\_\_\_\_ Debido a \_\_\_\_\_ Horas \_\_\_\_ Días \_\_\_\_ Meses \_\_\_\_ Años \_\_\_\_

**II. Otros procesos<sup>5</sup>** \_\_\_\_\_ Horas \_\_\_\_ Días \_\_\_\_ Meses \_\_\_\_ Años \_\_\_\_

¿Ha habido indicios de muerte violenta? Si  No  ¿Se practicó autopsia clínica? Si  No

¿La defunción ha ocurrido como consecuencia directa o indirecta de?:  
 Accidente de tráfico No  Si  Accidente laboral No  Si  Fecha del mismo: Día \_\_\_\_ Mes \_\_\_\_ Año \_\_\_\_

Figure 1: Medical death certificate

and, the most important, dealing with a new disease in the process of codification. In March 2020, when the first deaths due to covid occur, the Optical Recognition System did not recognize the terms referring to this new disease because of not being included in the dictionary. For that reason, coders from the Autonomous Communities, within the collaboration agreements with the NSI for the work of the Statistics, had to compare exhaustively the recognition result with the certificate images and correct any errors. As soon as possible, the terms used in the first certificates referring to COVID were included in the dictionary (around 145 different expressions).

Once the information on the certificates has been scanned and reviewed by coders, the next step is to identify the initial cause of death, applying the ICD-10 rules in the sequence of diseases informed by the physician in the Death Certificate, and finally determine the underlying cause of death. The Volume 2 of ICD-10 describes these rules according to medical criteria and applying medical logical relations between the diseases informed in the certificate. This is undoubtedly the most complicated part of the process and requires continuous medical supervision and advice.

Until 2013 the coding process was manual, and the emerging doubts were solved through a forum created for this purpose between the NSI and the Regional Mortality Registers. However, the manual

coding of causes of death, apart from being affected by the same problems as other manual coding, that is, it requires time, requires numerous human and financial resources, and is very sensitive to systematic errors by coders, presents another specific problem: the selection of the underlying cause of death must be based on the guidelines described in the ICD and these are characterized by their complexity and, above all, their numerous exceptions. In the context of improving the quality of this statistics, the IRIS automatic system was implemented in Spain with the 2014 data, as a consequence, a fundamental advance has been achieved in terms of punctuality and comparability of the information, both at national and international level (Carrillo and González, 2016). IRIS is an intelligent system prepared to work with the WHO medical death certificate model, which is in force in Spain. It works using algorithms based on codes assigned to medical terms, on causal medical relations and on the application of selection rules in accordance with the guidelines of Volume 2 of the ICD-10. To get an idea of how valuable this tool, it is enough to mention that the number of relations between diseases programmed exceeds 29 million.

Although Iris is linked to the names of its two co-founders, Lars Age Johansson and Gérard Pavillon, as well as the rest of the Core Group members, it is important to highlight that the automatic system success is also a consequence of the involvement of all countries that, to a greater or lesser extent, are part of the project. The development of a tool that guarantees the comparability of causes of death statistics at a world level could not be understood without a coordinated international cooperation. IRIS is a language-independent international software because it works with codes and this implies the development of a dictionary in the national language that associates an ICD-10 code with each disease. Currently, the Spanish dictionary consist of 167,000 standardized terms associated with their ICD-10 code. The COVID-19 pandemic brought new challenges for IRIS since their decision tables had to be modified to include the new disease.

The inclusion of COVID-19 in the statistical process began from the moment that WHO incorporated the new disease into the ICD-10, assigning two different codes to distinguish between COVID-19 virus identified (with positive test) and COVID-19 virus not identified (suspected), thereby making it possible to determine the mortality directly caused "by" confirmed COVID and "by" suspected COVID.

Apart from that, as mentioned before, in addition to the underlying cause, the Statistics also provides information on the diseases that have contributed to the death and have been reported by the physician in the Death Certificate (multiple causes). In this way, deaths with the presence of COVID-19 without being this disease the underlying cause, that is, mortality "with" confirmed COVID and "with" suspected COVID could be identified.

In the same way, multiple causes gave information about complications due to COVID-19 and the comorbidities of people who died of this disease. Respiratory failure and pneumonia were the most frequent complications reported on the Death Certificates of people who died of COVID-19, both identified and not identified virus. In terms of comorbidities, hypertensive disease and renal failure were the main comorbidities in COVID-19 virus identified mortality and dementia in the case of COVID-19 virus not identified mortality. The results showed that there were 60,358 deaths due to COVID-19 virus identified in 2020 and another 14,481 deaths due to suspected COVID-19 due to having symptoms compatible with the disease (COVID-19 virus not identified). In addition, physician certified 8,275 deaths due to other causes, but having COVID-19 as comorbidity contributing to the death. In 3,770 cases the physicians identified the virus and in 4,505 cases they suspected its presence due to having symptoms compatible with the disease.

The other action to improve quality that deserves to be highlighted is the incorporation of the Institutes of Legal Medicine and Forensic Sciences as a source of information on deaths with judicial intervention. Although quantitatively judicial deaths have little relative weight (between 5-6% of

mortality), they are very relevant qualitatively because most of them are premature and avoidable deaths.

Up to 2019, the main source of this information was the judicial authorities and around 40% of reported judicial deaths had incomplete information on causes of death. This implied that the Mortality Register of Autonomous Communities, depending on their available human resources, had to improve this information by contacting the coroners, creating comparability problems between regions and time series.

The NSI has developed a software in order the coroner to provide the information on deaths with judicial intervention, the tool is in line with the statistical needs, that is, focused on obtaining the ICD-10 code and guarantees coverage and comparability. The incorporation of the Institutes of Legal Medicine and Forensic Sciences in the circuit of deaths with judicial intervention information, as recommended by numerous studies, has led to a very significant improvement in quality. In addition, the collaboration of the Institutes of Legal Medicine and Forensic Sciences positions them as a key source of information in the circuit of official statistics on causes of death, reinforcing their social and health projection. The software also provides reports with the same format and common international health language that facilitates comparability between Institutes of Legal Medicine and Forensic Sciences and the preparation of its annual reports.

### 3 Some examples of epidemiological studies and health research based on the Statistics on Causes of Death

As mentioned in the introduction, the Statistics on Causes of Death is the source of information for numerous epidemiological studies and health research. In order to show the potential offered by the Statistics, it is presented below some of the numerous research projects based on monitoring the mortality of specific cohorts' members:

- Monitoring of the mortality of the cohort of patients affected by the Toxic Oil Syndrome. The Carlos III Health Institute has been monitoring for 37 years the Toxic Oil Syndrome epidemiological cohort with 20,643 affected since the beginning of the outbreak. The NSI annually provides the vital status of those affected and the cause of death in case of decease to identify a mortality pattern of this epidemic. Up to now, more than 5,000 deaths have been identified.
- EPI-CT project on potential health effects of exposure of children and adolescents to ionizing radiation during TAC scans carried out by the Centre for Research in Environmental Epidemiology (CREAL) of Catalonia. This project came to light through a collaboration agreement between the NSI and the Centre for Research in Environmental Epidemiology (CREAL) of Catalonia. The objective was to determine the potential effects of the ionizing radiation doses applied during the TAC scans, to see the possible effects and, according to the results, to reduce and optimize these doses. There was first a pilot study with 10,000 patients and later with more than 200,000 patients. The cohort was crossed with 27 mortality data files.
- BIFAP Project: Database for Pharmacoepidemiologic Research in Primary Care. The Spanish Agency for Medicines and Health Products requires information on mortality by cause to evaluate the safety and effectiveness of new medicines.
- Prospective study of the Health Research Institute of the Hospital La Paz on the follow-up during 12 months of a 2,000 individual cohort with a suicide attempt. This study aims at assessing the incidence of re-attempted suicide and identify risk factors. For this purpose, this cohort has been crossed with the mortality data files.
- Mortality study among medical professionals. The Council

of Official Colleges of Physicians has prepared a mortality study among the medical professionals. This initiative analyses, for the first time in Spain, the expectancy life and causes of death of Spanish physicians, based on the data from its register and mortality data from the NSI. The study analyses the evolution of the number of medical professional deaths in the period 2005-2014 and the main causes of death in this group.

## 4 Other user demands

The availability of additional variables of interest for the mortality analysis is a reiterated demand from researchers.

Based on 2016 data, variables such as educational level and activity status have been obtained from administrative sources and incorporated into the microdata file. Occupation, at the one-digit of the National Classification of Occupations (CNO 11) level, has been assigned in case the deceased over 16 years of age was working at the time of death.

For this purpose, the information from the pre-census population files that were prepared for the elaboration of the 2021 Population and Housing Census has been used (INE 2014).

The sources of information have been the Mutual insurance companies of civil servants (MUFACE, MUGEJU and ISFAS), the file of current contracts from the Public Employment Service (SEPE) and the 2011 and 2001 Censuses.

Also from 2010 data, the geographical coordinates of the census section corresponding to the residence of each deceased are made available to the researchers in the microdata file.

For the future, other information of interest, such as the average income of the census section, it will be incorporated into the microdata files for researchers as soon it is available.

## References

- Carrillo, Jesús and María del Rosario González (2016). Iris: Codificador automático internacional de causas de muerte. Technical report, Instituto Nacional de Estadística.
- Instituto Nacional de Estadística (INE) (2014). Método de asignación de nivel educativo, relación con la actividad laboral y ocupación. Technical report, Instituto Nacional de Estadística (INE).
- Instituto Nacional de Estadística (INE) (2020). Certificado médico de defunción/ boletín estadístico de defunción. Technical report, Instituto Nacional de Estadística (INE).
- Instituto Nacional de Estadística (INE) (2021). Metodología de los resultados detallados. Technical report, Instituto Nacional de Estadística (INE).
- Organización Panamericana de la Salud (OPS) (1995). Clasificación estadística internacional de enfermedades y problemas relacionados con la salud – 10 revisión. Technical report, Washington, D.C.