Title

Beyond the underlying cause of death: an algorithm to study multi-morbidity at death.

Journal

Population Health Metrics, 22(1)

Authors

Grippo, Francesco

Frova, Luisa

Pappagallo, Marilena

et al.

Publication Date

2024-12-18

DOI

10.1186/s12963-024-00356-8

Peer reviewed

Population Health Metrics

RESEARCH

Open Access

Beyond the underlying cause of death: an algorithm to study multi-morbidity at death

Francesco Grippo[1*], Luisa Frova[1], Marilena Pappagallo[1], Magali Barbieri[2,3], Sergi Trias-Llimós[4], Viviana Egidi[5], France Meslé[2] and Aline Désesquelles[2]

Abstract

Background In countries with high life expectancy, a growing share of the population is living with several diseases, a situation referred to as multi-morbidity. In addition to health data, cause-of-death data, based on the information reported on death certificates, can help monitor and characterize this situation. This requires going beyond the underlying cause of death and accounting for all causes on the death certificates which may have played various roles in the morbid process, depending on how they relate to each other.

Methods Apart from the underlying cause, the cause-of death data available in vital registration systems do not differentiate all other causes. We developed an algorithm based on the WHO rules that assigns a "role" to each entry on the death certificate. We distinguish between the following roles: originating (o), when the condition has initiated a sequence of events leading directly to death; precipitating (p), when it was caused by an originating condition or one of its consequences; associated (a), when it contributed to death but was not part of the direct sequence leading to death; ill-defined (i), i.e., conditions such as symptoms or signs or poorly informative causes. We applied this algorithm to all death records in four countries (Italy, France, Spain and the US) in 2017.

Results The average number of originating causes is similar in the four countries. The proportion of death certificates with more than one originating cause—a situation typical of multi-morbidity—ranges from 10% in the US to 18% in Spain. All ages combined, the proportion of deaths with at least one associated cause is higher in Italy (41%) and in the US (42%) than in France (29%) and in Spain (27%). It is especially high in the US at all adult ages. Variations in the average number of causes between the four countries are mainly due to precipitating and ill-defined causes.

Conclusions The output of our algorithm sheds light on cross-country differences in the average number of causes on death certificates. It also opens the door for improvements in the methods used for multiple cause-of-death analysis.

Keywords Mortality, Causes of death, Aging, Multi-morbidity

*Correspondence:
Francesco Grippo
frgrippo@istat.it
Full list of author information is available at the end of the article

Background

To a large extent, increase in life expectancy since the 1970s is due to a decrease in the incidence of, but also to better survival from, chronic diseases and more specifically from cardiovascular conditions [1, 2]. In low mortality countries, an ever-growing share of the population is living with several diseases—a situation referred to as multi-morbidity [3–5]. In this context, monitoring and characterizing multimorbidity is a legitimate objective. It relies on data from health surveys and health data systems but, as multi-morbid patients are at higher risk of dying [6, 7], cause-of-death statistics based on the information reported on death certificates, usefully complement the picture. Analysis relying on a single cause, typically the underlying cause of death (UC)—i.e., according to the World Health Organization (WHO) terminology, "*the cause that initiated the sequence of events leading to death*" –, cannot shed light on that trend. Identifying patterns of multi-morbidity at death requires looking beyond the underlying cause of death to account for all the causes reported on the death certificate, herein labeled multiple causes (MC). In the last decade, the so-called "multiple cause-of-death (MCOD) approach" has emerged as a new and promising field of research. Methods and indicators have been developed in order to take full advantage of the information reported by certifying physicians. Besides indicators that aim at recalculating mortality levels attributed to a given condition by considering all entries on the certificate, indicators that measure the strength and the specificity of associations of causes have been proposed [8, 9]. Methods derived from network analysis have also been successfully applied to describe all links between causes mentioned on death certificates [10]. Another methodological path has been taken recently, which consists of summarizing the information on the death certificate according to different types of morbid processes (simple, ill-defined and multi-morbid) [11].

The death certificate recommended by WHO comprises two parts (see Appendix). In Part I, the certifying physician is asked to report all conditions involved in the morbid process that directly led to death, from the immediate cause of the death to the cause that initiated the sequence. Part II is for "*any other significant condition that unfavorably influenced the course of the morbid process but is not related to the condition directly causing death*" [12]. Causes reported on the death certificates may have played a different role in the morbid process, depending on how they relate to each other. But apart from the underlying cause, the cause-of death data available from vital statistics systems list all other causes without differentiating between them, except for their location on the death certificates (line, order and part), which provides information on their interrelationships.[1] Sometimes death certificates include more than one sequence in Part I, each with a different "originating" cause and a different train of complications leading to death. This situation is of particular interest when monitoring multi-morbidity at death. Causes mentioned on Part II of the death certificates that have played an indirect role in the process leading to death also signal multi-morbidity. Manton and Stallard [15] referred to them as «background factors for other causes»: when they are combined with another serious disease, the risk of dying increases, reflecting either a «synergistic» or «additive» morbid process [16]. This is typically the case for hypertension, diabetes, as well as for frailty symptoms [17]. In terms of public health, it is also of interest to characterize the consequences of underlying causes.

In this paper, we present an algorithm that assigns a "role" to each entry on the death certificate. We distinguish between the four following roles: originating (o), when the condition has initiated a sequence of events leading directly to death[2]; precipitating (p), when it was caused by an originating condition or one of its consequences; associated (a), when it contributed to death but was not part of the direct sequence leading to death. In addition, there are ill-defined causes of death (i), i.e., conditions such as symptoms or signs or poorly informative causes. We illustrate the results of our approach with data from four high-income high-data-quality countries: France, Italy, Spain and the United States. More precisely, we present basic statistics about the frequency of these different types of entries on the death certificates in those countries. This categorization of the causes reported on the death certificates opens the door for improvements in the methods used for MCOD analysis.

Data and methods

We use individual-level multiple cause-of-death data for year 2017[3] from France, Italy, Spain, and the United States. All four countries implement the medical certification form recommended by the World Health Organization to report causes of death with slight variations in terms of

[1] In the 1980's, the National Center for Health Statistics in the United States (NCHS) developed the software TRANSAX (for Translation of Axis), with the aim of producing more meaningful multiple cause data [13]. The software applies the rules and provisions of the ICD in order to create "*the data necessary for person-based tabulations by translating the axis of classification from an entity basis to a record basis*", i.e. by aggregating and arranging the codes in order to better represent the morbid processes responsible for the death [14]. The TRANSAX system may be considered a pioneering system in accounting for relationships among the conditions reported on death certificates.

[2] In principle, the underlying cause is one of them but as mentioned before, there may be several originating causes on a given death certificate.

[3] 2017 is the last year before the COVID-19 pandemics for which data are available in all four study countries.

the order of the reporting and the space available to report the causes on the hard copy version of the certificate. Data are provided by the French National Institute for Health and Medical Research (INSERM) in France, by the Italian National Institute of Statistics (ISTAT) in Italy, by the Spanish National Institute of Statistics (INE) in Spain, and by the National Center for Health Statistics (NCHS) within the Centers for Disease Control (CDC) in the United States. The analysis includes all death certificates recorded in 2017,[4] amounting to 591,535 deaths in France, 650,590 in Italy, 424,523 in Spain, and 2,813,503 in the US. All four countries use automated coding systems. Causes of death are automatically coded under the 10th Revision of the International Classification of Diseases (ICD-10). The United States uses the MICAR-ACME system (2009 version) [18], while France, Italy and Spain use the IRIS system (2016 version) [19]. Both systems are highly consistent and strictly follow all WHO rules for the coding of causes and the selection of the underlying cause of death.

We developed an algorithm that labels each disease or condition on the death certificate, as represented by an ICD-10 code, according to their role in the morbid process. The algorithm uses the location of the cause on the certificate (Part I or Part II of the death certificate, line number and position on each line). In brief, first, the algorithm identifies 'ill-defined' codes. Second, the program uses the decision tables embedded in IRIS for the UC selection among the remaining codes, as described in the ICD-10 WHO Manual [12]. These decision tables are implemented to identify one or more 'originating' causes for any given sequence(s). Originating causes are supposed to be reported alone on the lowest line of Part I, but sometimes they are reported elsewhere (e.g. on Part II). Finally, the algorithm classifies the remaining causes in Part I as 'precipitating' and those in Part II as 'associated'.

The steps followed by the algorithm are described below in more detail and illustrated on Fig. 1. Table 1 provides some examples of death certificates with coded causes and the labels assigned to each cause during the processing steps.

1. Initial step

   A. All ill-defined codes are labelled as such (label = 'i'). Ill-defined codes (uninformative causes of deaths, symptoms and signs) are mainly included under chapter 18 ("Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified") of the ICD. This list is presented in the Appendix.
   B. All remaining causes in Part I are initially labelled as originating ('o').
   C. All remaining causes in Part II are initially labelled as associated ('a').

2. Identifying the originating causes

   A. Checking sequences in Part I. This section of the algorithm identifies the starting point of the sequences leading to death described in Part I. The ICD-10 WHO Manual includes a list of the conditions that can be "**due to"** another given condition. If a condition can be due to another condition located on a lower line, the program labels the condition in the upper line as precipitating ('p'). This process is performed iteratively for each pair of codes. Example 1 in Table 1 shows a simple case: excluding all ill-defined codes, the only specific code in Part I (I489, unspecified atrial fibrillation) is due to the single code reported on the lowest line (I219, unspecified acute myocardial infarction), so I489 is flagged as precipitating (p). Example 2 is more complex: there are two originating causes (G309, unspecified Alzheimer disease, and J449, unspecified chronic obstructive pulmonary disease): G309 cannot be due to J449, so G309 remains as originating (o).
   B. Checking for possible originating causes in Part II. This step derives from an ICD-10 rule to identify an originating cause even when reported in an improper location on the death certificate, for instance in Part II. This rule explicitly identifies conditions that are "obvious consequences" of other conditions. The concept of "obvious consequence" refers to a disease or condition that is typically a complication from another cause reported on the death certificate and that can unarguably be considered as the result of this other condition even if the sequence of conditions is not reported in the expected order. Considering only codes labelled as originating in Part I or associated in Part II, the algorithm evaluates if a condition is an "obvious consequence" of another. If that is the case, the condition which is the"obvious consequence" is labelled as 'precipitating'. In example 3 of Table 1, because the ICD considers that E86 (volume depletion) is an obvious consequence of F03 (unspecified dementia), E86, previously labelled as 'originating', is converted into 'precipitating' in step 2B and F03 becomes the 'originating' cause.

[4] For Italy and Spain all deaths that occurred in the country are considered. For France deaths of people residing in overseas departments or abroad are not included. For the US, only deaths of national and legal residents are included.

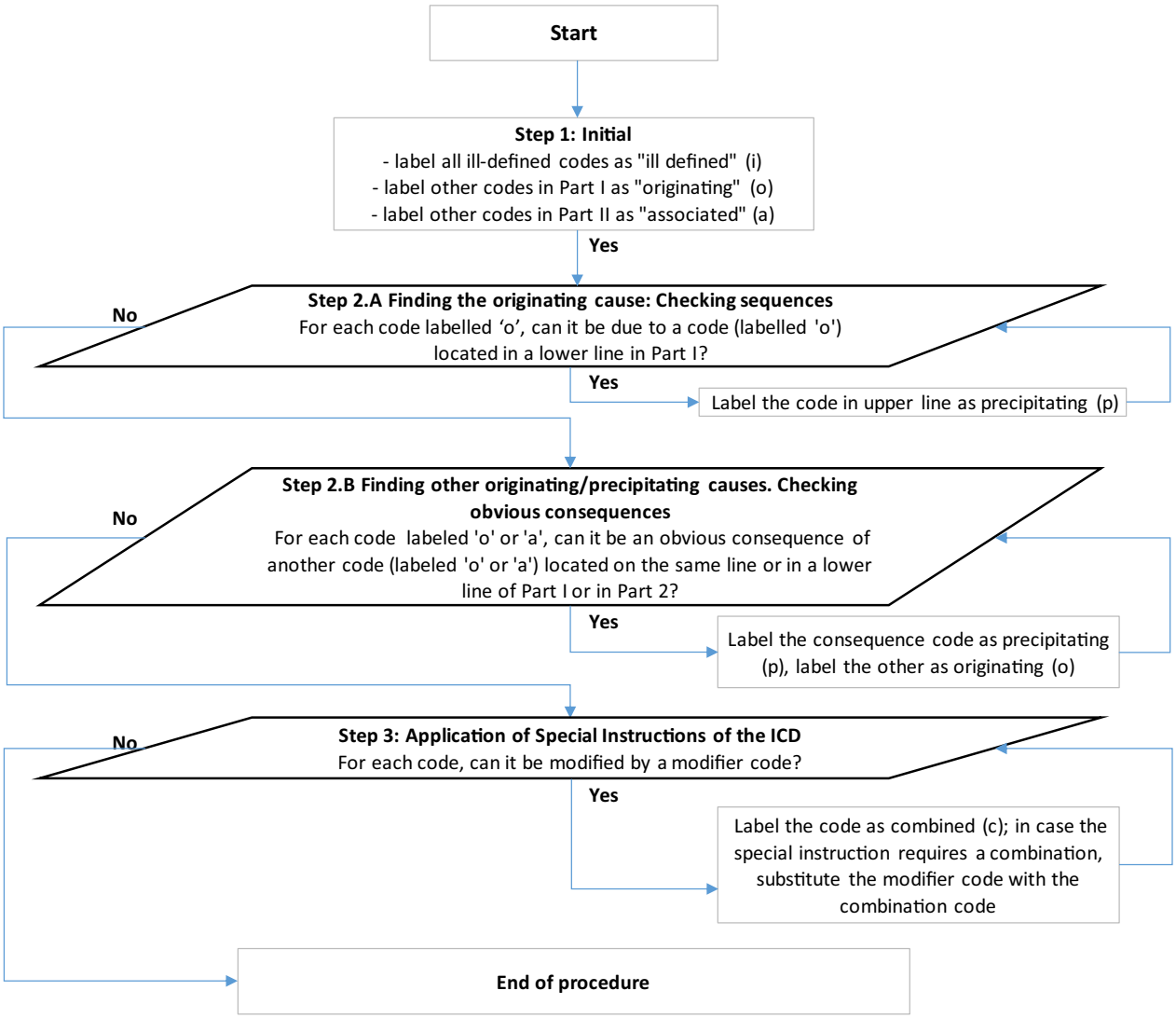

Fig. 1 Summary of the steps followed by the algorithm

3. Application of the ICD Special Instructions. In some instances, the ICD-10 Manual identifies specific pairs of codes from which to identify the underlying cause. These instructions require either to give preference to one of the two codes or to combine the two codes into a third one. We decided to include these instructions into our algorithm for consistency with WHO rules regarding the selection of the underlying cause. The instructions are described in detail in the ICD-10 WHO Manual [12].

For each 'originating' or 'associated' cause, our program thus checks if there is any cause reported on the death certificate for which a special instruction applies. Example 2 in Table 1 indicates that the originating cause J449 (Chronic obstructive pulmonary disease, unspecified) is modified by the presence of J189 (pneumonia) and the resulting code is J440 (Chronic obstructive pulmonary disease with acute lower respiratory infection). Code J449 is subsequently dropped and replaced by J440. J189 is then labeled as combined ('c').[5] In Example 4, B485 (pneumocystosis), labelled as the originating cause, is reported as a consequence of R75 (HIV positive status). According to WHO rules, B485 is then modified and becomes B206 (HIV disease resulting in *Pneumocystis jirovecii* pneumonia).

In the end, the output of the algorithm is a file where all causes of the multiple cause-of-death file have been flagged according to the role they played in the process leading to death. The initial version of the program was

[5] Combined codes represent a small share of all entries in the four countries: 6.3% in France, 3.8% in Italy, 4.7% in Spain, 6.2% in the US.

Table 1 Output of the algorithm on four death certificates

Step	Example 1				Example 2				Example 3				Example 4			
1 Initial step	PI.a	I469	Cardiac arrest, unspecified	i	PI.a	J960	Acute respiratory failure	i	PI.a	I469	Cardiac arrest, unspecified	i	PI.a	B485	Pneumocystosis	o
	PI.b	I489	Atrial fibrillation and atrial flutter, unspecified	o	PI.b	J189	Pneumonia, unspecified	o	PI.b	R64	Cachexia	i	PI.b	R75	Laboratory evidence of human immunodeficiency virus [HIV]	i
	PI.c	I219	Acute myocardial infarction, unspecified	o	PI.c	G309	Alzheimer disease, unspecified	o	PI.b	I489	Atrial fibrillation and atrial flutter, unspecified	o				
	PII	E149	Unspecified diabetes mellitus: Without complications	a	PI.d	J449	Chronic obstructive pulmonary disease, unspecified	o	PI.c	E86	Volume depletion	o				
									PII	F03	Unspecified dementia	a				
2A Checking sequences (due to rules)	*Rules found in Iris decision tables:*				*Rules found in Iris decision tables:*				*Rules found in Iris decision tables:*				No change			
	I489 *is due to* I219				J189 *is due to* G309				I489 *is due to* E86							
	PI.a	I469	Cardiac arrest, unspecified	i	PI.a	J960	Acute respiratory failure	i	PI.a	I469	Cardiac arrest, unspecified	i				
	PI.b	I489	Atrial fibrillation and atrial flutter, unspecified	p	PI.b	J189	Pneumonia, unspecified	p	PI.b	R64	Cachexia	i				
	PI.c	I219	Acute myocardial infarction, unspecified	o	PI.c	G309	Alzheimer disease, unspecified	o	PI.b	I489	Atrial fibrillation and atrial flutter, unspecified	p				
	PII	E149	Unspecified diabetes mellitus: Without complications	a	PI.d	J449	Chronic obstructive pulmonary disease, unspecified	o	PI.c	E86	Volume depletion	o				
									PII	F03	Unspecified dementia	a				
2B Checking obvious consequences	No change				No change				*Rules found in Iris decision tables:*				No change			
									E86 *is obvious consequence of* F03							
									PI.a	I469	Cardiac arrest, unspecified	i				
									PI.b	R64	Cachexia	i				
									PI.b	I489	Atrial fibrillation and atrial flutter, unspecified	p				
									PI.c	E86	Volume depletion	p				
									PII	F03	Unspecified dementia	o				

Table 1 (continued)

Step	Example 1	Example 2	Example 3	Example 4
3 Application of special instructions of the ICD	No change	*Rules found in Iris decision tables:* J449 *is modified by* J189 *and transormed into* J440 PI.a J960 Acute respiratory failure i PI.b J189 Pneumonia, unspecified c PI.c G309 Alzheimer disease, unspecified o PI.d J440 Chronic obstructive pulmonary disease with acute lower respiratory infection o	No change	*Rules found in Iris decision tables:* R75 *modifies* B485 *which is transformed into* B206 PI.a B206 HIV disease resulting in Pneumocystis jirovecii pneumonia o PI.b R75 Laboratory evidence of human immunodeficiency virus [HIV] i

developed in the C computer language.[6] For a wider use, it is currently being converted into the R language.

Results

We first looked into the average number of mentions on the death certificates for each of the four study countries. All ages combined, the average number of entries per death certificate is highest in Italy (4.4) and lowest in the US (3.2). France and Spain are in an intermediate position with 3.5 entries and 3.7 entries per death certificate, respectively.

The average number of entries as well as the country ranking vary with age (Fig. 2). However, at all ages, Italy is the country with the highest number of entries. The extreme positions of Italy and the US become the rule over the age of 60 years. At those ages, France and Spain are in-between Italy and the US, with a relatively similar number of entries. At the youngest ages, the average number of entries is also highest in Italy, while the lowest numbers are found in Spain and the US. In the US, it increases with age to reach a maximum among young adults of 3.5 entries per death certificate. Spain exhibits the lowest average number of entries per death certificate for age group 15 to 55 years.

Figure 3 shows that the average number of originating causes is very similar in the four countries (ranging from 1.00 in France to 1.17 in Spain). Differences are also relatively small for ill-defined causes, with the exception of the US where, on average, the number of ill-defined entries is low. Variations between the four countries are mainly due to precipitating and to associated causes that, however, represent a small share of all entries. Italy has the highest number of precipitating causes (1.38), followed by Spain (1.04), France, and the US (0.75 in both countries). As far as associated causes are concerned, the ranking is quite different from that for all entries together, with the highest values in Italy and the US, and the lowest in France and Italy.

Figure 4a shows that the average number of originating causes is not only very similar in all four countries but also quite stable with age. In all age groups, the average numbers are close to or slightly over/under one. Over the age of 75, Spain slightly overpasses the other countries. France has the lowest average number of originating cause, especially at very old ages (over the age of 95). It may happen that there is no originating cause or that there is more than one (Fig. 4b). The proportion of death certificates with more than one originating cause ranges from 10% in the US to 18% in Spain.

[6] The source code and the compiled software can be obtained upon motivated request to the authors of the paper.

The average numbers of both precipitating and associated causes by age group are very heterogeneous across the four countries. At all ages, Italy has the highest number of precipitating causes, increasing from a low of 0.95 at age 0 to a high around 1.5 among young adults and adults until the age of 80 years and then decreasing to 1.0 over the age of 95 years (Fig. 5). The age pattern in Spain has a particularly striking shape: after a first peak for children, the average number of precipitating causes declines to a minimum around ages 30–34 years and the second peak at only 1.1 is reached at ages 85–89 years. The pattern for France is a softened version of the Spanish one, with a higher average among children and for deaths between the ages of 65 and 89 years, but the average number of precipitating causes never exceeds 0.85. By contrast with its three peers, the US curve peaks at young adult ages, reaching 1.3 at ages 25–29 years. It then declines to below Spain at ages 50–54 years, and below France at ages 60–64 years. The ranking of the four countries for all ages combined is the same as that over the age of 60 years, where the largest share of deaths occurs. Among young adults (15–39 years of age), the average number of precipitating causes in the US is exceeded only by that in Italy, while France and Spain exhibit the lowest numbers.

Associated causes represent the smallest share of all entries on the death certificates (from 11% in Spain to 22% in the US). In France and Spain, the average number of associated causes is the highest for the age group 85–89 years at 0.53 and 0.46, respectively (Fig. 6a). In Italy, the highest value is for deaths at age 80–84 years when it reaches 0.77. This is also the highest value reached in the US but the age profile is again very different: while in France, Spain and Italy, the average number of associated causes progressively increases up to ages 80 to 90 years, and then decreases—more or less sharply—at older ages, the US pattern is characterized by relatively high numbers of associated causes even at young ages, and a less pronounced increase with age. As an example, at age 15–19 years, there are 0.39 associated causes per death certificate in the US, a level reached only at age 50–54 years in Italy.

Figure 6b shows the proportion of death certificates with at least one associated cause. All ages combined, this proportion is much higher in Italy (41%) and in the US (42%) than in Spain (27%) and France (29%). In the US, from age 20 to age 90 years, 40 to 45% of all death certificates include at least one associated cause. Such a level is reached in Italy only, over the age of 70. In France and Spain, the proportion of death certificates with at least one associated cause reaches its maximum value between the ages of 85 and 94 years, at around 32%. Compared

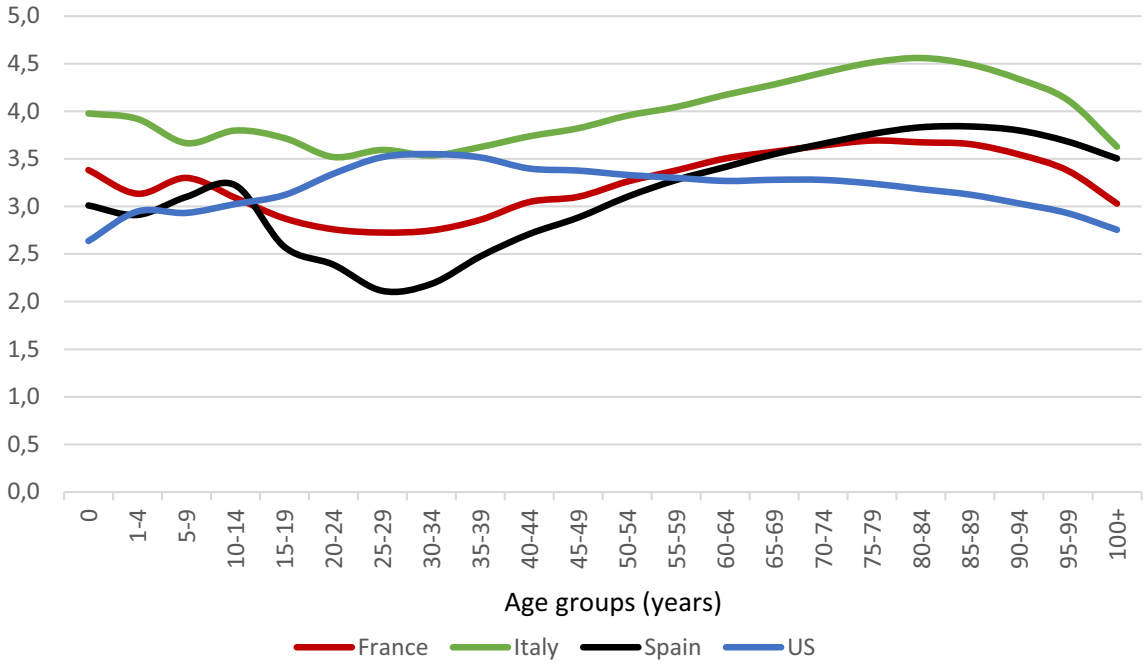

Fig. 2 Average number of entries per death certificate (all roles) by age group—All deaths in 2017 in France, Italy, Spain, and the US

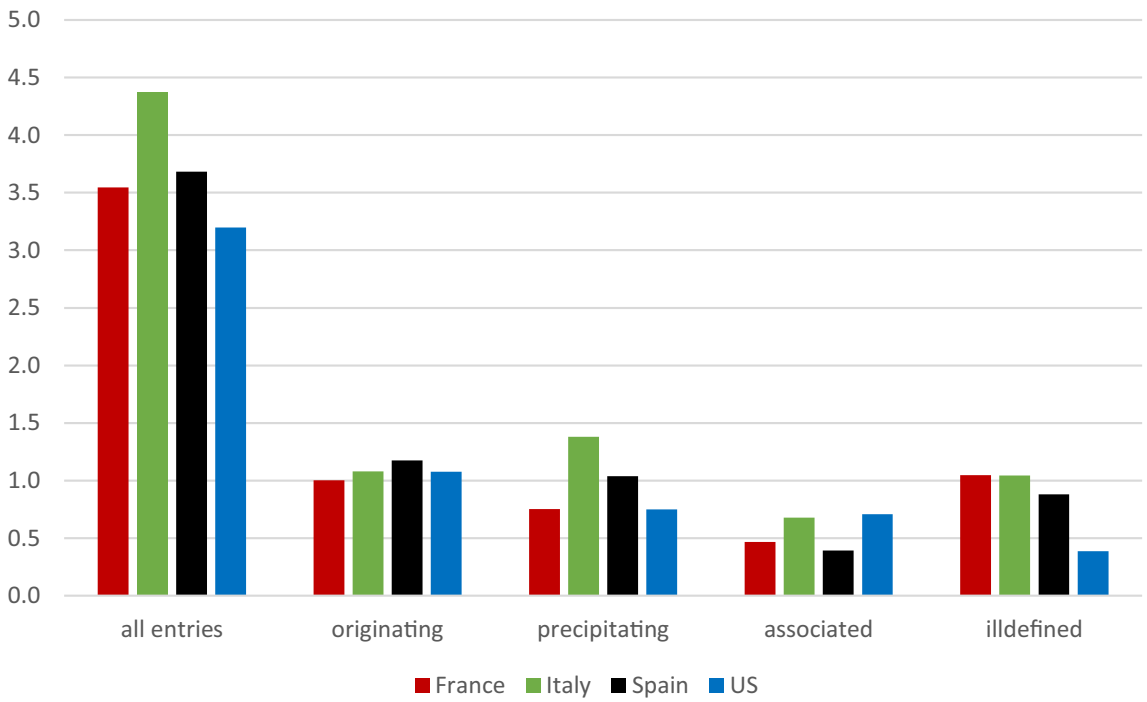

Fig. 3 Average number of entries per death certificate according to their role in the morbid process—All deaths in 2017 in France, Italy, Spain, and the US

with the US, the three European countries show a clearer upward trend with age followed by a decline at older ages.

The average number of ill-defined mentions, which tends to increase with age, is lowest in the US at all ages, with a minimum of 0.13 at ages 20–24 years and a

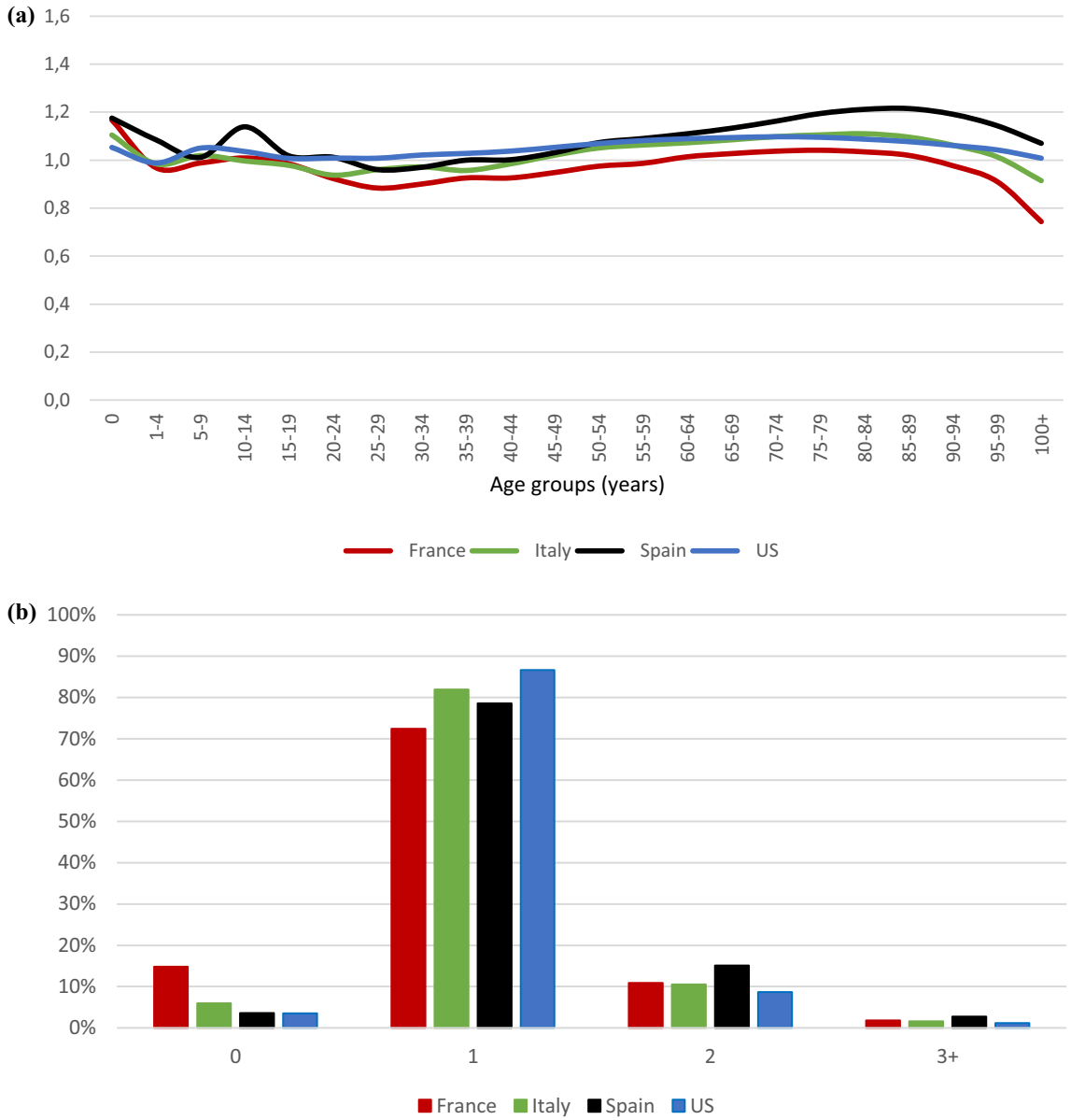

Fig. 4 **a** Average number of originating causes per death certificate—All deaths in 2017 in France, Italy, Spain, and the US **b** Distribution of the death certificates according to the number of originating causes—All deaths in 2017 in France, Italy, Spain, and the US

maximum of 0.5 at ages 100 and above (Fig. 7). The number of ill-defined codes (around one per death certificate on average) is similar in France and Italy and slightly lower in Spain.

Discussion

In this paper, we described an algorithm based on the rules established by WHO to select the underlying cause of death, which labels diseases or conditions reported on the death certificate according to their role in the morbid process ('originating', 'precipitating', 'associated' and 'ill-defined'). We applied this algorithm to all 2017 death records in Italy, France, Spain, and the US. We computed the average number of entries per death certificate in all four countries and by age groups, for these different types of entries. The proposed classification of causes of death provides information on the differences between countries in the average number of entries on the death certificate. As such, it offers some insights into the quality of the cause-of-death data. For example, the percentage of death certificates with more than one originating cause of death could be taken as an indicator of inaccurate

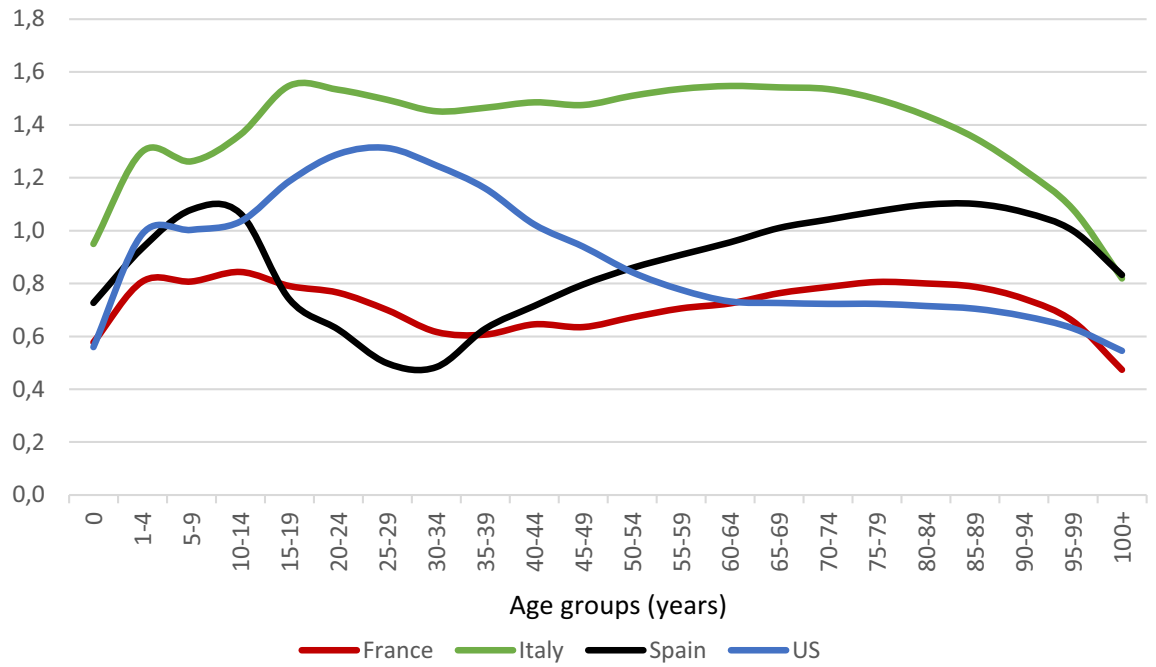

Fig. 5 Average number of precipitating causes per death certificate– All deaths in 2017 in France, Italy, Spain, and the US

certification. As will be shown later, it can also be used to identify the types of causes that contribute most to cross-country differences. Beyond this methodological interest, the purpose of the tool is to refine and extend the use of multiple cause-of-death data by identifying the role that each condition has in the fatal process. In particular, the tool allows identifying cases where death is due to multiple independent processes (multi-morbidity) rather than to a single cause of death. So far, a proxy indicator for identifying multiple cause at death was the total number of codes on certificates, which was clearly not satisfying and provided no information on the relationships between causes.

Our results confirm that the average number of causes on death certificates varies greatly from one country to the other, as well as by age within each country. These differences may reflect real differences in the epidemiological profiles of the countries under study. However, we cannot exclude that differences in certification practices contribute to these variations, since more conditions listed on the death certificate could result from differences in the training and recommendations provided to certifiers, with encouragements to describe the morbid process in more details in some settings than in others. Similarly, it is possible that slight differences in the format of the death certificate could induce such variations, for instance by providing more space to list causes of death. Yet, assessing the extent to which these two aspects could have affected the results was not possible in our settings.

The classification of the cause-of-death mentions enables us to identify which types of causes vary most from one country to the other. Additional analyses based on this classification suggest that the variability of originating causes around its mean value of 1.1 is extremely low, accounting for less than 1% of the variance of the total number of entries. The same is true for associated causes with little variability (8% of the total), indicating very similar values around the average of 0.56 associated cause per certificate. In contrast, the highest variability is observed for ill-defined (32%) and precipitating causes (32%). The role of covariance is also very important (28%), which expresses the fact that the numbers of causes of each type are correlated with each other. This could result ofromdifferences in both underlying cause-of-death profiles and certification practices.

The age-related pattern of variability differs for the four different types of mentions (Fig. 8). For originating causes, variability remains consistently low across all age groups; that of precipitating causes exhibits two peaks, one at ages 25–29 years and another at ages 60–64 years; associated causes show higher variability during the central years of life, up to the age of 65 years. Notably, the number of ill-defined causes varies widely across countries under age 30 years and above age 85 years. The covariance between the different roles has its highest positive values under the age of 5 and over 75 years,

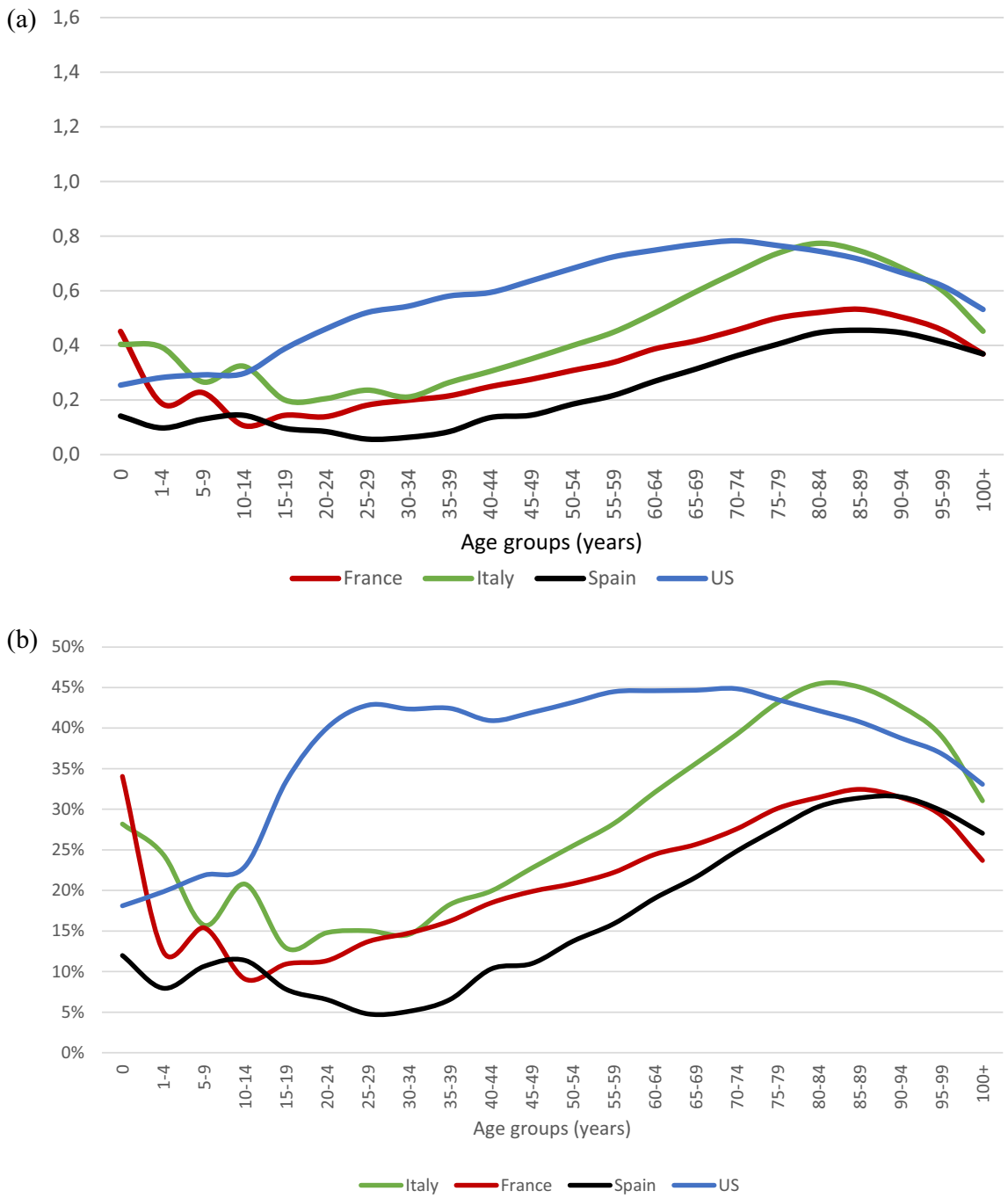

Fig. 6 **a** Average number of associated causes per death certificate– All deaths in 2017 in France, Italy, Spain, and the US **b** Percentage of the death certificates with at least one associated cause by age group—All deaths in 2017 in France, Italy, Spain, and the US

peaking at the age of 80–84 years, suggesting that at those ages, relationships between the different types of entries contributes to increase the variability across countries. For 20–24 years old, as well as for 35–64, on the contrary, the covariance is negative and it contributes to lower the variability among countries.

There generally is one originating cause per death certificate. Though it is not the aim of our algorithm, this cause is expected to be selected as the underlying cause. We verified that overall, in 89.3% of the cases, the underlying cause corresponds, at the three-digit level, to one of the originating causes identified by the algorithm

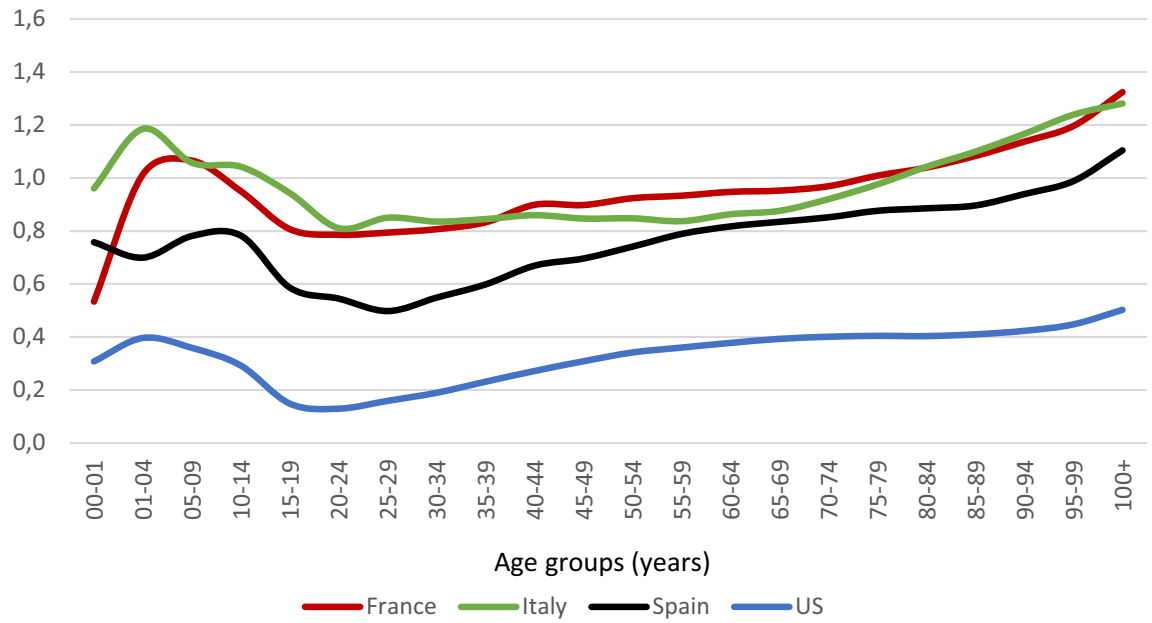

Fig. 7 Average number of ill-defined entries per death certificate—All deaths in 2017 in France, Italy, Spain, and the US

(ranging from 79.7% in France to 91.2% in the US). There are several reasons for the gap in the identification of the underlying cause between our algorithm output and the national statistics office determination. First, our algorithm does not treat ill-defined causes in the same way as the WHO selection process. Per WHO rules, for certificates with only ill-defined causes, an ill-defined code is assigned to the UC. By contrast, our algorithm never labels any of the ill-defined causes as originating: ill-defined causes are flagged as ill-defined at the beginning of the process, i.e., a separate category from originating causes. This allows identifying cases without a well described sequence of diseases/injuries leading to death. In addition, our list of ill-defined codes differs slightly from that of WHO (see Appendix), which could result in an inconsistent application of coding rules on some certificates. This applies especially to certificates with acute heart failure reported as an originating condition in Part I, since such a condition is considered ill-defined by WHO but not by our algorithm. Nevertheless, these cases are uncommon.

By contrast, a condition reported in Part II can be flagged as originating with our algorithm only when a well-defined cause in Part I is an "obvious consequence" of this condition. Otherwise, it is considered as associated. Another reason for the difference with the WHO recommendations is that we do not take into account the time interval sometimes reported on the death certificates between the onset of a condition or disease and death. Next, the decision tables embedded in our algorithm are the 2020 decision tables while the underlying cause in the countries under study was determined using a less recent version of these tables. Last, certificates that are manually coded in country production (in Italy they account for 21% of the total) might be lacking the complete reporting of multiple causes. These certificates are mainly associated with external causes, mentions of surgery, or they correspond to complex cases not considered by the decision tables. Indeed, when we exclude records with ill-defined underlying causes or external causes (including surgeries), the percentage of overall agreement rises to 93.3% (ranging from 89.9% in France to 94.0 in the US).

It may happen that there is no originating cause on the death certificate or that there are several originating causes. Though this latter case can be considered as the result of errors in certification, it interestingly can be interpreted as a situation of multi-morbidity, with two (or more) independent morbid processes being reported by the certifying physician due to their inability to identify the true originating cause. Similarly, death certificates with at least one associated cause—which we find to be especially prevalent in the US at all adult ages—also correspond to situations of multi-morbidity: theses associated causes have contributed to the morbid process as "background factor" without causing the death [15].

By identifying multi-morbidity at death, the algorithm follows the method developed by Grippo et al. [11],

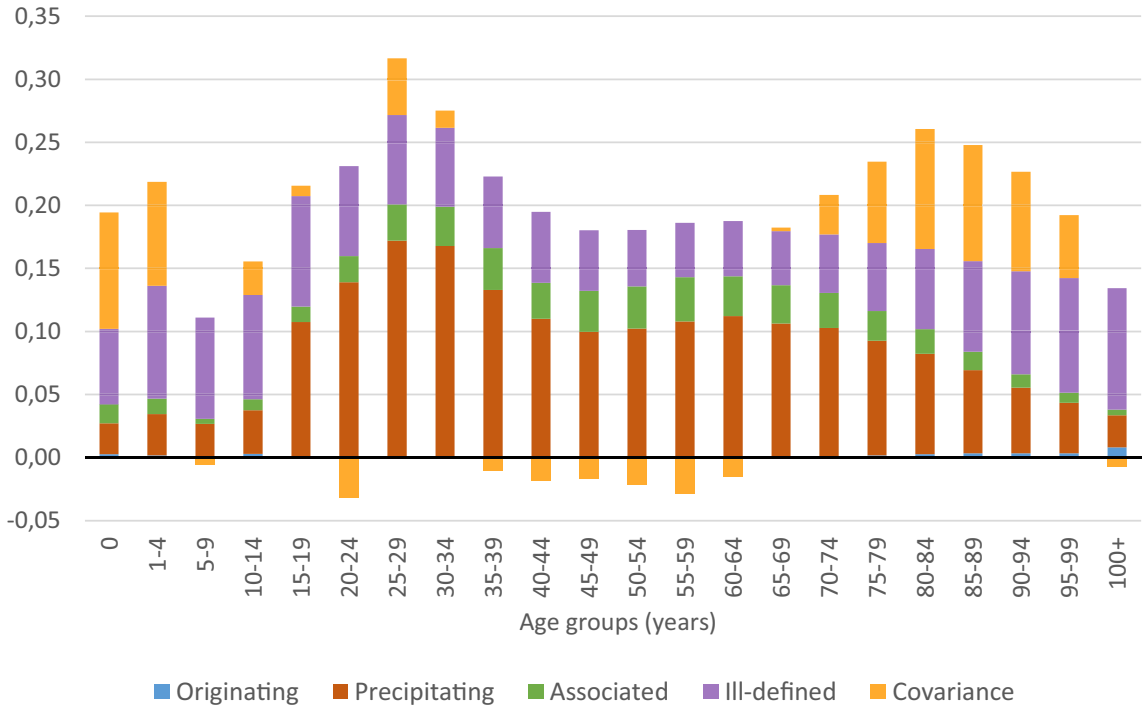

Fig. 8 Variability in the number of entries according to their role in the morbid process and the age group—All deaths in 2017 in France, Italy, Spain, and the US

which aimed to classify deaths according to three main types of processes (simple, multi-morbid and ill-defined). However, it overcomes several limitations of this previous work. First, the computer program developed by Grippo et al. did not account for the fact that causes in Parts I and II of the death certificates can be causally related. Though it does not directly apply to our analysis, which relies on data pre-dating the pandemic, our approach is likely to encounter many such situations in the context of COVID-19, when mentions of pneumonia on Part I as a result of the infection with the SARS-CoV2 virus mentioned on Part II have become common [20, 21]. Second, the output of the program developed by Grippo et al. was a simple count of the number of processes on the death certificate but, apart from the underlying cause of death, the algorithm implemented in this early paper did not explicitly label the various entries on the death certificates. As a consequence, it was not possible to characterize which precise causes on the death certificates contributed to the morbid process as precipitating causes and which led to the identification of the process as multi-morbid. A preliminary version of this algorithm has been used to describe the complications of COVID-19 as well as the most frequent preexisting conditions [22].

Conclusion

The output of our algorithm suggests possible improvements in the methods based on multiple cause-of-death data. Some analyses require using the death (rather than the mentions) as the unit of observation. Weighing strategies for the entries have been developed to estimate mortality rates such that the sum of the weights for a given death equal to one [23–26]. This approach raises important issues due to the arbitrary nature of the weights. The ability to identify originating, precipitating and associated causes from processed death data greatly enhances the utility as the "role" assigned to each cause can potentially inform further development on non-arbitrary weights.

More generally, the method we have developed may be used to "clean" the data in an appropriate way given the objectives of the study. The output from our method offers new opportunities for monitoring mortality profiles by computing MCOD indicators (e.g., rates, standardized ratio of multiple to underlying cause and cause-of-death association indicators) separately for the different types of causes on the death certificates (originating, associated and precipitating). Such distinction is meaningful in a public health perspective. Precipitating causes that are complications or consequences of an originating cause could be monitored as targets for tertiary prevention. Examining associated causes can shed light

on causes that, though not lethal, increase the vulnerability and the risk of dying of persons with other more severe conditions (e.g., cancers, cardio-vascular diseases, or infectious diseases).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12963-024-00356-8.

Supplementary file 1.

Acknowledgements
Not applicable.

Author contributions
Francesco Grippo, Luisa Frova and Marilena Pappagallo developed the C algorithm, ran it on the data for Italy and wrote the Method section of the paper. Francesco Grippo provided advice on the application of the WHO rules for the conceptualization of the algorithm. Magali Barbieri, Aline Désesquelles and Sergi Trias-Llimós ran the program on (respectively) the data for the US, for France and for Spain. France Meslé (MD) provided medical advice in the conceptualisation of the algorithm. Viviana Egidi and Luisa Frova performed the variability analysis. Aline Désesquelles coordinated the project and wrote the first draft of the paper (except the Method section). All the authors contributed to writing the successive versions of the paper.

Funding
Magali Barbieri received funding from the University of California Berkeley Center for the Economics and Demography of Aging (CEDA) (NIH Grant #P30AG012839) for this research though the views expressed in this manuscript do not necessarily reflect the official policies of NIH. Sergi Trias-Llimós acknowledges funding from the Spanish State Research Agency (Grant RYC2021-033123-I).

Availability of data and materials
For France: access to the cause-of-death data is restricted to authorized users via the SNDS (Système national des données de santé) platform. For Italy: Data can be requested to the Istat contact centre (https://contact.istat.it/s/?language=it). For Spain: Multiple cause-of-death data are not publicly available but interested researchers have the possibility to obtain data access by contacting INE (www.ine.es). For the US: the multiple cause-of-death data used in this study are publicly available from the NCHS at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/DVS/mortality/mort2017us.zip. The source code of the C application can be requested to frgrippo@istat.it.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Istituto Nazionale di Statistica (ISTAT), Viale Liegi, 13, 00198 Rome, Italy. [2]Institut National d'Etudes Demographiques (INED) 9 Cours Des Humanités, CS 50004, 93322 Aubervilliers Cedex, Paris, France. [3]Department of Demography, University of California, 310 Social Sciences Building, Berkeley, CA 94720-2120, USA. [4]Center for Demographic Studies (CED-CERCA), Carrer de Ca n'Altayó, Edifici E2, 08193 Bellaterra, Spain. [5]Dipartimento di Scienze Statistiche, Sapienza University of Rome, Viale Regina Elena, 295, 00161 Rome, Italy.

Received: 27 November 2023 Accepted: 15 November 2024
Published online: 18 December 2024

References

1. Leon DA. Trends in European life expectancy: a salutary view. Int J Epidemiol. 2011;40(2):271–7.
2. Mensah GA, Wei GS, Sorlie PD, Fine LJ, Rosenberg Y, Kaufmann PG, Mussolino ME, Hsu LL, Addou E, Engelgau MM, Gordon D. Decline in cardiovascular mortality: possible causes and implications. Circ Res. 2017;120(2):366–80.
3. Palladino R, Pennino F, Finbarr M, Millett C, Triassi M. Multimorbidity and health outcomes in older adults in ten European health systems, 2006–15. Health Aff. 2019;38(4):613–23.
4. Head A, Fleming K, Kypridemos C, Schofield P, Pearson-Stuttard J, O'Flaherty M. Inequalities in incident and prevalent multimorbidity in England, 2004–19: a population-based, descriptive study. Lancet Healthy Longev. 2021;2(8):e489–97.
5. Souza DL, Oliveras-Fabregas A, Minobes-Molina E, de Camargo CM, Galbany-Estragués P, Jerez-Roig J. Trends of multimorbidity in 15 European countries: a population-based study in community-dwelling adults aged 50 and over. BMC Public Health. 2021;21(1):1–10.
6. Salive ME. Multimorbidity in older adults. Epidemiol Rev. 2013;35(1):75–83.
7. Jani BD, Hanlon P, Nicholl BI, McQueenie R, Gallacher KI, Lee D, Mair FS. Relationship between multimorbidity, demographic factors and mortality: findings from the UK Biobank cohort. BMC Med. 2019;17(1):1–13.
8. Désesquelles A, Salvatore MA, Pappagallo M, Frova L, Pace M, Meslé F, Egidi V. Analysing multiple causes of death: which methods for which data? An application to the cancer-related mortality in France and Italy/Analyse des causes multiples de décès: quelles méthodes pour quelles données? L'exemple de la mortalité par cancer en France et en Italie. Eur J Population. 2012;28(4):467–98.
9. Bishop K, Balogun S, Eynstone-Hinkins J, Moran L, Martin M, Banks E, Joshy G. Analysis of multiple causes of death: a review of methods and practices. Epidemiology. 2023;34(3):333.
10. Egidi V, Salvatore MA, Rivellini G, D'Angelo S. A network approach to studying cause-of-death interrelations. Demogr Res. 2018;38:373–400.
11. Grippo F, Désesquelles A, Pappagallo M, Frova L, Egidi V, Meslé F. Multimorbidity and frailty at death: a new classification of death records for an ageing world. Popul Stud. 2020;74(3):437–49.
12. WHO.International Statistical Classification of Diseases and Related Health Problems (The) ICD-10. Fifth Edition. 2016.
13. https://www.cdc.gov/nchs/nvss/mmds/acme_transax.htm
14. Chamblee RF, Evans MC. TRANSAX, the NCHS system for producing multiple cause-of-death statistics, 1968–78. Vital and Health Statistics. 1986;l(20).
15. Manton KG, Stallard E. Temporal trends in US multiple cause of death mortality data: 1968 to 1977. Demography. 1982;19(4):527–47.
16. Speizer FE, Trey C, Parker P. The uses of multiple causes of death data to clarify changing patterns of cirrhosis mortality in Massachusetts. Am J Public Health. 1977;67(4):333–6.
17. Trias-Llimós S, Barbieri M, Egidi V, Frova L, Grippo F, Meslé F, Désesquelles A. Frailty at death: an examination of multiple causes of death in four low mortality countries in 2017. Demogr Res. 2023;49(2):13.
18. https://www.cdc.gov/nchs/nvss/mmds/about_mmds.htm
19. http://www.iris-institute.org
20. Fedeli U, Schievano E, Avossa F, Barbiellini Amidei C, Grande E, Grippo F. Different approaches to the analysis of causes of death during the COVID-19 epidemic. Eur Rev Med Pharmacol Sci. 2024;25(9):3610–3.
21. Gundlapalli AV, Lavery AM, Boehmer TK, Beach MJ, Walke HT, Sutton PD, Anderson RN. Death certificate–based ICD-10 diagnosis codes for COVID-19 mortality surveillance—United States, January–December 2020. Morb Mortal Wkly Rep. 2021;70(14):523.
22. Grippo F, Navarra S, Orsi C, Manno V, Grande E, Crialesi R, Frova L, Marchetti S, Pappagallo M, Simeoni S, Di Pasquale L, Carinci A, Donfrancesco C, Lo Noce C, Palmieri L, Onder G, Minelli G, Italian National Institute of Health COVID-19 Mortality Group. The role of COVID-19 in the death of

SARS-CoV-2–positive patients: a study based on death certificates. J Clin Med. 2020;9(11):3459.
23. Piffaretti C, Moreno-Betancur M, Lamarche-Vade A, Rey G. Quantifying cause-related mortality by weighting multiple causes of death. Bull World Health Organ. 2016;94(12):870.
24. Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death. Epidemiology. 2017;28(1):12–9.
25. Bishop K, Moreno-Betancur M, Balogun S, Eynstone-Hinkins J, Moran L, Rao C, Joshy G. Quantifying cause-related mortality in Australia, incorporating multiple causes: observed patterns, trends and practical considerations. Int J Epidemiol. 2022;52(1):284–94.
26. Trias-Llimós S, Permanyer I. Cause-of-death diversity from a multiple-cause perspective in the United States. Demography. 2023;60(1):73–98.

Publisher's Note