Reconstructing Long-Term Coherent Cause-of-Death Series, a Necessary Step for Analyzing Trends

Markéta Pechholdová^{1,2} · Carlo-Giovanni Camarda¹ · France Meslé¹ · Jacques Vallin¹

Published online: 19 December 2017

© Springer Science+Business Media B.V., part of Springer Nature 2017

Abstract Every time the classification of causes of death is changed, time series of deaths by cause are disrupted in more or less profound ways. When changes involve only the merging of several items or splitting a single item into several new categories, the problems caused by these ruptures are not too difficult to solve. A more or less severe imbroglio occurs, however, each time a new item results from recombining portions of different split items. Sometimes, but very rarely, some countries proceed to a bridge coding during the year of transition, which can help reconstruct coherent time series. This article first summarizes the general principles of the method developed for France by Meslé and Vallin to reconstruct complete series for France from 1925 to 1999 in the detailed list of the 9th WHO International Classification of Diseases (ICD), doing so by successively bridging a posteriori the five versions of the ICD that were in use during that period. Second, it reports on several methodological improvements that have been developed with the aim to reconstruct and analyze mortality trends by cause in sixteen industrialized countries.

Keywords Mortality · Causes of death · Time series reconstruction · Automation · ICD · Bridge coding

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10680-017-9453-1>) contains supplementary material, which is available to authorized users.

✉ Markéta Pechholdová
pechholdova@gmail.com

¹ INED, Paris, France

² University of Economics, Prague, Prague, Czech Republic

1 Introduction

Understanding mortality differences and changes greatly depends on the availability of good data on medical causes of death. From infection to cardiovascular diseases and violent deaths, the most prominent ways to die are strongly and specifically related to various aspects in societal, economic, and political contexts. Consequently, medical cause-of-death statistics are a precious tool for studying the epidemiological conditions of survival and for devising ways to fight premature death—either by creating the appropriate therapeutic tools and making them universally available or by implementing well-targeted prevention services and campaigns. This is especially true for analyses of East–West differences and of observed trends toward divergence or convergence. In every developed country, there currently exist cause-of-death data that are of rather good quality. However, their comparability over space and time is quite problematic, because today’s data rely on a long historical process that developed according to national specificities, which inevitably results in substantial discontinuities in time and differences in space.

Physicians through the ages have sought to uncover the etiology of different diseases and the paths to treatment and recovery. In contrast, the idea of collecting individual diagnoses to produce health and mortality statistics is quite recent. The first difficulty that arose in establishing counts was how to most efficiently classify diseases and injuries. The idea of classifying diseases was already present in Ancient Greece, in Arab medicine and in the European Middle Ages (through Albert de Souabe, for example), but the first modern attempts occurred in the eighteenth century with the work of Boissier de la Croix (1730). It was not until the mid-nineteenth century that a consensus was reached regarding the need for an international classification of diseases that met the requirements of statistical data collection. This took place at the first International Statistical Congress (Brussels 1853), and nearly another half-century would pass before the adoption of the International Classification of Diseases in 1893 (at the ISI World Statistics Congress in Chicago). The next ISI Congress (Christiana 1899) decided to revise this classification every 10 years, with the First Revision of the International Classification of Diseases (ICD-1) adopted in 1900 in Paris. Until the present, the international classification has been revised ten times, progressively taking a very different shape from those of the late nineteenth and early twentieth centuries. Discussions regarding ICD-11 are currently in progress. These revisions present one of the greatest problems in analyzing historical trends in mortality by cause.

Although the general structure of ICD has not changed much, the number of items has been multiplied by 50, from 203 to more than 10,000. Moreover, the relationships between the various items are very complex, making it impossible to easily track the diseases through various revisions. Finally, statistical offices do not always use the same classification detail. All together, the consequence of the ICD revisions and various levels of classification detail is that raw data cannot reflect actual long-term trends in mortality by cause. Even if, in the framework of the present EJP special issue, only the most recent decades are of actual interest, it is

necessary to deal with the two or three most recent ICD revisions, which were not necessarily the least disruptive.

Available raw cause-of-death data cannot be used to draw any conclusions about patterns in the prevalence of any major pathological process (Vallin and Meslé 1988). ICD revisions resulted not only in the very frequent subdivision of former items into several new items or the much less frequent merging of several items into a single one (developments that would not cause great damage), but they also resulted in splitting various items from older classifications and reorganizing them under completely different new classifications. As one would imagine, such complex rearrangements should be minimized when working on data that is based on the complete detailed ICD lists. However, in reality, this is not at all the case: each ICD revision has resulted in a large number of complex rearrangements of many items from the complete detailed list, and, more importantly, such complex redistributions usually involve very large numbers of deaths (for France, see: Vallin and Meslé 1988; Meslé and Vallin 1996).

Thirty years ago, a method was established for reconstructing coherent time series in order to study long-term mortality trends by cause in France, which resulted in a complete reclassification of deaths into the detailed ICD-8 list for the whole period 1925 to 1978¹ (Vallin and Meslé 1988). Various researchers then applied the method to more recent French data as well as to several other countries. While it allows relying on coherent time series to follow epidemiological changes that were driving mortality trends, it also presents the great inconvenience of being very laborious and time-consuming, which is why it is rarely applied. Other authors have preferred more rapid solutions of employing regression (Janssen and Kunst 2004) or time series techniques (Rey et al. 2011; van der Stegen et al. 2014). In these approaches, continuity is assessed and established based on the statistical estimation of trends before and after classification changes. Unfortunately, the approaches of Janssen and Kunst (2004) and Rey et al. (2011) do not assure obtaining the same number of deaths after the reconstruction as before. This inconvenience was corrected by van der Stegen et al. (2014) by adding a constraint which forces the sum of the corrections to be zero. Moreover, all of these alternative approaches worked only on a very abridged list of causes (ICD chapters in van der Stegen et al., 13 selected causes in Rey et al. 2011), on a selected subset of high-quality data countries (only Western countries), on selected age groups (only above 60 in Janssen and Kunst 2004), or even without distinguishing age and sex (van der Stegen et al. 2014). As such, these methods cannot be seen as alternatives to the initial one of Vallin and Meslé, because they never yielded continuous time series in sufficient detail and fully classified by age and sex. It is also questionable how well these methods (based on the statistical strength of the modeled trends) would perform if applied to individual ICD codes for which there are few deaths and high or irregular yearly fluctuations. The more rapid time series methods also fail to pay attention to the actual medical content of the ICD items, which can produce misleading results. According to van der Stegen et al. (2014), not taking into

¹ 1925 is the first year for which national statistics of death by sex, age, and cause are available; 1978 is the last year ruled by ICD-8 in France.

account the medical content in the applied corrections is an advantage, because very limited information is then required. This claim is, however, in conflict with the general purpose of the time series reconstruction, which aims to understand as precisely as possible the pathological background of long-term mortality trends. Finally, these methods are presented as universally applicable, but both Janssen and Kunst (2004) and Rey et al. (2011) acknowledge that a country-specific approach is needed due to persisting differences in certification and coding practices.

In the absence of any satisfying alternative, it was therefore pressing to refine the original French method in order to facilitate applying it to more and more countries. This paper proposes several ways to automate—if not the entire process—at least its most time-consuming steps.

The first section will thus present the original method succinctly, while the second will report on the novelty of automating several procedures. Throughout the text, we will also discuss the methodological advantages of the proposed method against other proposed reconstruction approaches, and we will additionally address the possibilities of using transition facilitating tools produced by the WHO and by national statistical authorities.

2 Reconstructing Coherent Time Series of Deaths by Cause: A General Approach

Typically, little is known about the statistical effects of implementing a new classification. In rare cases, the administrative body responsible for cause-related mortality statistics has conducted dual classification when a new classification was first implemented² (which is usually called “bridge coding”). In all other cases, it is necessary to undertake the painstaking task of comparing the contents of the items of the former classification with those of the new one and to identify the exchanges between items.

2.1 The Case of Bridge Coding

Bridge coding produces independent dual coding of a sample of individual death certificates according to both the former and the new classifications. It should ideally include all deaths in the given year, but in most cases the sample is more restrictive.

Bridge coding is primarily conducted with the aim of measuring the statistical effects of the classification change. Under ideal conditions, bridge coding also allows directly computing the *transition coefficients* which specify, for the corresponding year, the proportion of each item from the old classification that should be attributed to every related item in the new classification. Applying transition coefficients to annual deaths by cause observed before the classification

² For example, in England in 1979, the OPCS undertook a dual classification of a sample of deaths using the intermediate lists from both the eighth and ninth revisions of the ICD, which made it possible to reconstruct the corresponding statistical series with constant definitions (Meslé and Vallin 1993). In the USA, bridge coding studies have been conducted for the last 6 ICD-transitions at various levels of detail (NCHS 2001).

change then enables redistributing them according to the new classification under the hypothesis that such transition coefficients are stable throughout the period of interest. In other words, we assume that bridge coding applied to any year in the period would have produced the same result as the one actually organized at for the year of transition. This hypothesis is strong but we can generally assume that biases will be fairly negligible if the work is performed at a great enough level of detail.

Theoretically, when bridge coding is available, it can easily solve some issues related to ICD revisions. Figure 1 gives an example of annual death rates by ischemic heart disease for the USA across the four periods of ICD-7, ICD-8, ICD-9 and ICD-10. Clearly, the reconstructed series (solid line) perfectly corrects the discontinuities caused by classification changes (dotted line).

The main problem with bridge coding is related to the size of the sample. If it is applied to a small population or to an overly small sample from a greater population (to minimize the cost), it is impossible to work with a sufficiently detailed list of causes. Many relationships will be missing or uncertain, and many necessary transition coefficients will thus remain unknown. Even though a rather large sample was used in a very large country such as the USA, many items in the detailed list had to be grouped together to avoid the problem of having too many empty cells. Most samples are much smaller, and the reconstruction can thus be carried out only at the level of very large groups of causes, for which it is questionable to assume that transition coefficients are stable.

One may also hope that transition coefficients established for one country could be used to reconstruct series in countries where no bridge coding is available. However, this is in fact quite impossible. For example, looking at French data, we identified two typical statistical discontinuities caused by the switch to ICD-10 in

Fig. 1 Trends in mortality rates by ischemic heart disease in the USA after reconstructing annual death series on the basis of bridge coding over 3 revisions, from ICD-7 to ICD-10. Items 420 of ICD-7, 410-414 of ICD-8 and ICD-9, and I20-I25 of ICD-10. *Source:* Barbieri and Meslé 2008, p. 25

1999: the annual deaths by “pneumonia” (items 480–486 of ICD-9 and items J12–18 of ICD-10) dropped dramatically, while those by “septicemia” (item 038 of ICD-9 and items A40–41 of ICD-10) rose suddenly (Meslé and Vallin 2008). Can we reconstruct French series by applying US or British transition coefficients obtained from bridge coding? Figure 2 clearly shows that although the results might be acceptable for pneumonia, the results for septicemia are not. Obviously, the difficulty in transferring coefficients from one country to another results from differences in national coding practices that affect the use of either ICD-9 and/or ICD-10.

In the same direction, it was also found that applying coefficients computed from the England and Wales bridge coding between ICD-8 and ICD-9 to French data produced very disappointing estimates (Meslé and Vallin 1993).

2.2 Creating an a Posteriori Cross-Tabulation (the Meslé-Vallin Method)

Because bridge coding is largely unavailable and given its limitations listed above, the statistical coherence of the time series can be achieved a posteriori by methodologically matching the detailed medical content of two successive ICD revisions. In order to do this, a method was developed at INED in the 1980s that—when applied to the available French data—has allowed the reconstruction of 1925–1978 cause-specific mortality series classified according to the eighth revision of the ICD (Vallin and Meslé 1988). Later, taking into account the changes caused by the switch to ICD-9 (Meslé and Vallin 1996), the previous series were themselves converted in order to obtain the 1925–1999 series according to ICD-9, as the aim of the method is to reclassify the deaths into the most recent classification revision.

This achievement came at the cost of a very long and laborious process. Six transitions had to be dealt with, from the third to the ninth revision, and each time, the method had to be adjusted in response to specific problems, not only because the ICD revisions were of widely differing nature and scale,³ but also because of the need to face more specific difficulties linked to the peculiarities of the French data, which varied from period to period.

To summarize the general principles of that method, let us take as an example the transition from ICD-8 to ICD-9 (see Meslé and Vallin 1996 for more detail). To obtain a posteriori transition coefficients, three basic stages are followed:

1. The construction of two symmetrical *correspondence tables* (CT), linking items from the past classification to items from the new classification and vice versa;
2. The definition of *fundamental associations of items* (FAI), bringing together the same medical diagnoses from the two classifications and ensuring statistical continuity between them; and

³ The most elaborate exercise was without a doubt the transition from the fifth revision (adopted in 1938 and still under the influence of Bertillon’s first classifications) to the sixth revision (adopted in 1948 under the auspices of the WHO and greatly influenced by the American medical tradition).

Fig. 2 An attempt to reconstruct French series using US or British coefficients obtained from bridge coding for the transition to ICD-10. Items 480-486 of ICD-9 and items J12-J18 of ICD-10 for “pneumonia”; item 038 of ICD-9 and items A40-41 of ICD-10 for “septicemia”. *Source:* Meslé and Vallin 2008, pp. 353–354

3. The establishment of a *transition table* (TT), indicating how deaths are to be distributed so that those originally classified according to the past classification could be reclassified in line with the new one (or reciprocally).

To create the first correspondence table (CT-1), we take the first item of ICD-8 (000.0, Classical cholera) and systematically search in the alphabetical index of ICD-9 (WHO 1978) for any ICD-9 item that includes any diagnosis mentioned or suggested by the description given in the systematic detailed list from the ICD-8 *Manual* (WHO 1968). In this case it is very simple, since only item 001.0 of ICD-9 (cholera due to *Vibrio cholerae*) is involved. We then do the same for the next ICD-8 item. Naturally, in many cases, CT-1 will indicate several ICD-9 items. At the end of the process, all ICD-8 items will have been reviewed and, if the work was correctly performed, all ICD-9 items will have been mentioned at least once. The second correspondence table (CT-2) is then built in the same way: for each item in the detailed list given by the ICD-9 *Manual* (WHO 1977), every ICD-8 item that includes any diagnosis mentioned in the ICD-9 item is identified. Again, all ICD-9 items are reviewed, and all ICD-8 items should be mentioned at least once. It is important to note here that these tables are built on the basis of medical definitions and contents given by ICD volumes. They are theoretical links. In the subsequent steps, these medical links have to be confirmed against the actual statistical contents of ICD items.

To do so, FAIs are then built in two steps. First, correspondence tables from the first stage are used. We start with the first ICD-9 item of CT-2 by looking for the corresponding item(s) in ICD-8; then, for each ICD-8 item, other corresponding ICD-9 items are given by CT-1; then, for each new ICD-9 item, CT-2 gives other corresponding ICD-8 items, if there are any, and so on until all links in either direction have been taken into account. The first FAI is then complete, and the next one can be built starting with the first ICD-9 item that has not yet been taken into account, and so on until the last ICD-9 item. Of course, in the example above, the first FAI is very simple since only one ICD-8 item corresponds to the first ICD-9 item, and vice versa. Very often, however, there are multiple correspondences from at least one side, and sometimes links on both sides result in highly complex associations (Table 1). In order to reduce the complexity, non-useful links can be deleted. Such negligible links mostly relate to cases where very few or even no deaths are involved in items that should have been shared according to complex theoretical interconnections.

In the second step, the coherence of each FAI has to be confirmed statistically by inspecting the time series of total annual numbers of deaths for all ICD items included in the FAI over the whole period covered by ICD-8 and ICD-9. If a suspicious change appears in the year of the transition (and cannot be explained either by historical trends or by usual fluctuations), it is necessary to first check that the building process was performed correctly and, if so, to thus conclude that coding practices did not correctly follow the medical ICD definitions. After determining the precise cause, the FAI must then be corrected accordingly.

The third stage is to convert the FAIs into a transition table. There are four types of associations. The first does not require any transition table, since it consists of a

Table 1 An example of a complex fundamental association of items (ICD-9/ICD-8) (FAI #852, dealing with some heart diseases)

ICD-9 items			ICD-8			
Item nb	Title	Deaths in 1979	Deaths in 1978	Item nb	P/T	Title
411	Other acute and subacute forms of ischemic heart disease	2483	15	411.0	T	Other acute and subacute forms of ischemic heart disease with hypertensive disease
			111	411.9	T	Other acute and subacute forms of ischemic heart disease without mention of hypertension
			630	412.0	P	Chronic ischemic heart disease with hypertension
			9049	412.9	P	Chronic ischemic heart disease without mention of hypertension
412	Old myocardial infarction	30	“	412.0	P	
			“	412.9	P	
414.0	Coronary atherosclerosis	2986	“	412.0	P	
			“	412.9	P	
414.1	Aneurysm of heart	35	“	412.0	P	
			“	412.9	P	
414.8	Other specified forms of chronic ischemic heart disease	1314	“	412.0	P	
				412.9	P	
414.9	Chronic ischemic heart disease, unspecified	1938	“	412.0	P	
				412.9	P	
			1	414.0	P	Ischemic heart disease, asymptomatic with hypertension
			5	414.9	P	Ischemic heart disease without hypertension
429.2	Cerebrovascular disease, unspecified	1144	“	412.0	P	
			“	412.9	P	
Total FAI no 852	411, 412, 414, 429.2	9930	9811			411, 412, 414

The letter T indicates that the ICD-8 item is totally included in the ICD-9 item; P indicates that the ICD-8 item is partially included in the ICD-9 item

link between a single ICD-9 item and a single ICD-8 item (as in the example of cholera above). The second type consists of several ICD-8 items merged into a single ICD-9 item, and this does not require a transition table either, since one hundred percent of each ICD-8 item involved simply has to be attributed to the corresponding ICD-9 item. For the third type of FAI, which consists of splitting one ICD-8 item into several ICD-9 items, the transition table is trivial: deaths registered under the ICD-8 item are divided among ICD-9 items in proportion to death counts observed in the first year of ICD-9.

The only problem appears with the fourth type of FAI, called “complex associations” since they link several ICD-9 items to several ICD-8 items through a complex reorganization of sub-components of different items. In this case, it is necessary to build a table that cross-tabulates all of the ICD-9 items involved with all of the ICD-8 items involved (Table 2). To compute transition coefficients, it is first necessary to estimate the number of deaths that would have been attributed to each ICD-8 item in 1979 (the year of transition) by applying the 1978 distribution to all deaths observed for the association in 1979. Then, each cell for which a

Table 2 Cross-tabulation of 1979 deaths according to both ICD-8 and ICD-9 (FAI #852, dealing with certain heart diseases)

ICD -9 items	ICD-8 items						Total 1979
	411	411.9	412	412.9	414	414.9	
Cross-tabulation							
411	15	112	153	2203			2483
412			2	28			30
414			195	2791			2986
414.1			2	33			35
414.8			86	1228			1314
414.9			126	1806	1	5	1938
429.2			74	1070			1144
Total 1979	15	112	638	9159	1	5	9930
Total 1978	15	111	630	9049	1	5	9811
Transition coefficients from ICD-8 to ICD-9							
411	1000	1000	240	241			
412			3	3			
414			306	305			
414.1			3	4			
414.8			135	134			
414.9			197	197	1000	1000	
429.2			116	116			
Total	1000	1000	1000	1000	1000	1000	

Shaded areas are excluded since no link exists between the two items

correspondence was previously defined can be filled in and the transition coefficients can be computed. A further distinction can be drawn between two categories of complex FAIs: those for which only one redistribution solution exists and those for which an arbitrary choice must be made between different distribution hypotheses.

Finally, transition coefficients can be used to convert data initially coded according to ICD-8 into the ICD-9 system and therefore to produce continuous time series for the whole ICD-8 and ICD-9 periods (Fig. 3).

However, it is still necessary to conduct two important additional operations before such series are completed. The first is to check the statistical continuity of the series once again at the level of specific ICD items, looking for suspicious breaks in each series for the year of the transition. If such discontinuities are found, it will be necessary once again to either rethink the transition coefficients (especially where it has been necessary to make a choice between several possible solutions in a complex FAI) or to even modify the association itself. At the final level, statistical continuity is checked also by age group and sometimes sex.

The final adjustment is less directly linked to ICD changes, but instead depends on the way these changes were implemented by national authorities or on the fact

Fig. 3 Trends in annual deaths before and after reconstruction (FAI #852, involving certain heart diseases), example from France

that these authorities may at any time decide to make changes in their use of ICD items. Such unexpected events sometimes result in abrupt changes that can be detected by inspection of time trends and corrected if their causes are documented or sufficiently understood.

As with series reconstructed on the basis of bridge coding, transition coefficients obtained by this type of a posteriori cross-tabulation are specific to a given country and cannot be applied to another country satisfactorily. Even for countries under the same administrative system, such as the republics of the former USSR during the Soviet period, transition coefficients obtained in one are often not applicable to another (Meslé and Vallin 2012).

3 Automating Specific Steps in the Reconstruction of Time Series: Four Main Achievements

While the method of a posteriori cross-tabulation is consensually recognized as the most accurate, it is sometimes criticized for being laborious and subjective. In a recent research effort, a series of new tools described below were designed. In order to reduce the “manual work,” save time, and reduce subjectivity, automated or semiautomated algorithms have been developed, tested, and implemented for each of the three main steps of the Meslé-Vallin method (producing correspondence tables, building fundamental associations of items, estimating transition coefficients). The automation deals primarily with the implementation of the ICD-10, which on the one hand represents the most challenging change to mortality statistics in the second half of the twentieth century, while on the other hand it enables the use of new transition helpers such as the WHO ICD-9/ICD-10 Translator or the bridge coding studies.

3.1 Correspondence Table

Until recently, the possibilities for automatizing this task were quite limited. When ICD-10 was adopted, the WHO produced a full correspondence table for the first time: the so-called WHO ICD-9/ICD-10 Translator (WHO 1997). This was built in a partially automatic fashion by means of comparing computerized English and French indexes. The WHO Translator thus provides virtually all hypothetical correspondences and, as such, is a great tool for automatizing the ICD-9 to ICD-10 transition. However, it cannot be applied directly for several reasons. First, no correspondences could be found for some codes and they are left as “undefined.” Second, some codes are linked only to the secondary (“asterisk”) codes or to the injury (S or T) codes, which are not used as underlying causes of death. Third, it does not define correspondences for HIV codes, which were added to the ICD-9 by a 1986 addendum (WHO 1986). Fourth, it is based on the first version of the ICD-10 and, as such, it does not include ICD-10 updates, which are now being adopted almost on a yearly basis. And finally, as the aim was to search for all hypothetical correspondences, many of the codes form huge associations (the largest association based on the WHO Translator contains 4780 links, the second largest contains 1870

Fig. 4 Matrix of correspondences based on the WHO ICD-10/ICD-9 Translator

links). Figure 4 represents all of the WHO Translator's valid correspondences plotted as a matrix, where each dot represents a link. The non-proportionality of the number of rows and columns (5231 rows vs. 10,369 columns) reflects the increase in detail of ICD-10, while a quasi-diagonal pattern means that a majority of items have kept a similar position in the classification. Numerous dots outside of the diagonal, however, point to the complexity of the changes between ICD-9 and ICD-10.

At the same time, more countries produced bridge coding at more or less detailed level. For three countries, it was possible to access ICD-9 to ICD-10 bridge coding at the most detailed level: the USA, England and Wales, and France. Unlike for the WHO Translator, which is based on hypothetical medical correspondences, the bridge coding studies reflect all exchanges observed in real data. In contrast, extinct diseases (such as cholera, plague, or smallpox) are not represented in the bridge coding at all, and correspondences for rare diseases are not statistically significant. Similarly to the WHO Translator, applying the bridge coding to reconstruct the series results in connecting almost all ICD items into one large association (the largest association derived from the bridge coding contains 49,041 links). The results of the bridge coding studies thus even accentuate the complexity of the exchange between ICD-9 and ICD-10, as is well demonstrated in Fig. 5, which plots the matrix of all correspondences observed in any of the three mentioned bridge coding studies.

In Fig. 5, the total number of deaths for which the plotted links were observed is 2,923,174 (53,869 from France, 551,093 from England and Wales, and 2,318,212 from the USA). In spite of such considerable numbers of deaths, there is still an under-representation of numerous ICD items: 45% of the codes are missing from ICD-9 and 56% from ICD-10. This basically means that even full-scale bridge coding (such as that for the USA) cannot be used as the only source for reconstructing time series and that a large part of the correspondences has to be defined by other means. The information from the two sources—the WHO

Fig. 5 Matrix of correspondences based on available bridge coding studies

Translator and the bridge coding—can, however, be appropriately combined: the links proposed by the WHO Translator and observed in either of the three bridge coding studies can be considered as both medically justified and validated by the practice.

For reducing the complexity of the correspondences given by the WHO Translator, an approach based on social network analysis (Wasserman and Faust 1994) was developed and tested (Remund 2016). This method uses criteria based on the structure of each association by targeting, in order of priority, the WHO Translator links that are generating the most complexity (i.e., the links with high “betweenness”) while at the same time are not observed in practice (in the bridge coding). The more a link contributes to the complexity of its association and the less empirical support it relies on, the weaker it is.

In the developed approach that is currently being tested on the reconstruction of the French series between ICD-10 and ICD-9, we are thus using the following three criteria to simplify the complexity of the correspondence table: (1) the betweenness (as it is known from the social network theory that high betweenness vertices have the potential to disconnect the clusters if they are removed); (2) the validity (to delete, in priority order, the links that have never been observed empirically in the bridge coding); and (3) the frequency (the number of times each item appears in order to avoid generating orphans). Based on combining the WHO Translator, available detailed bridge codings, and the social network theory method, this approach yields a medically justified and reasonably complex correspondence table (with maximum number of links within one association equal to 138), which can then be used directly for building the FAIs in the subsequent work.

3.2 Building Fundamental Associations of Items

The construction of fundamental associations of items has been fully automatized. At first, an iterative algorithm (Fig. 6) that builds associations directly from the correspondence table was developed in R software (Pechholdová and Camarda 2014). The output table of the algorithm was then directly used as input in another

Fig. 6 Estimated transition coefficients over ages. FAI #852, involving certain heart diseases. Values in parentheses in each panel title represent the number of deaths involved in each exchange for all ages

automated tool, which was developed for visualizing the associations as they were designed by the Meslé-Vallin method. The correct visualization of the associations includes the names of the items, the observed death counts, and—most importantly—the order in which the items are placed in the association, all of which are crucial for the further work of balancing the statistical content of the associations.

A program for visualizing the associations was developed in Visual Basic for Applications (VBA), by Bâzgan and Penina (2016). Automation of the association formatting has several important advantages: a significant reduction in manual effort, standardization of the formatting, elimination of potential (human) mistakes, and flexibility in updating the associations any time the correspondence table is modified.

The statistical continuity of the FAIs is then assessed by inspecting the trends for the sum of deaths in each association according to the same process as described in section 3.4 for specific cause-of-death time series. Detecting a disruption indicates the need to modify the correspondence table (and the association). To assist in decisions about modifying the FAIs, a semiautomated tool was developed by Grigoriev (2016). The algorithm visualizes the trend in the sum of all deaths in the association before and after the modification to assess its potential effect. The final decision, however, relies on the country specialist and on the individual character of the issue being treated; as such, it cannot be further automatized.

3.3 Estimating Transition Coefficients

Relying on the basic assumption that proportions observed among new items after the revision remain equal in the years before the revision, we embed the estimation of transition coefficients in a least-squares problem with asymmetric constraints.⁴ We can express Tables 1 and 2 as series of linear equations where the unknowns are the transition coefficients. These equations satisfy both row and column sums of Table 2. A plain ordinary least-squares (OLS) solution would seem appropriate. However, two issues arise within this framework: (1) estimated coefficients need to be bounded between 0 and 1; (2) different correspondence tables can yield different conditions.

To overcome the former issue, we enforce the inequality constraints (coefficients must be between 0 and 1) by adding an asymmetric penalty to the OLS algorithm and iteratively solving its penalized version (Eilers 2005).

Within the system of equations coming from the row/column sums of the correspondence table and the boundary constraints, we can encounter three different conditions:

1. A unique solution, which is the case for all simple associations, splitting, merging, and most of the smaller complex associations. These cases pose no problem in the estimation of transition coefficients.

⁴ Further details on the underlying statistics as well as R routines for reproducing the results can be found in associated supplementary material.

2. An infinite number of solutions, which occurs in the complex associations for which an arbitrary choice must be made between different distribution hypotheses.
3. No solution, indicating that either the assumption of equal proportions among new items in the turning years is not correct, or the correspondence table needs to be modified by reconsidering exchanges between items from their medical content, as explained above.

In the most common case of infinite possible outcomes, we need additional constraints for selecting one of the equally optimal solutions. We assume that for items with unknown distribution, the proportions between death counts in each cell and marginal totals should be as close as possible. This approach will thus select, from among all infinite solutions, the transition coefficients which maintain an internal proportion over rows and columns similar to the proportions observed at the marginal level, thereby ensuring an internal coherence of the estimated correspondence table. When we apply this approach to FAI #852, we get results which are similar—though not equal—to those manually computed and presented in Table 2 (see Table 1 in online supplementary material).

Finally, we generalize the proposed approach by allowing age-specific exchanges between old and new items. We consider the complete matrix of deaths over old items and age groups and compute the associated matrix of expected deaths, assuming equal proportions by new cause of death in the 2 years of transition for each age group.

By introducing two-dimensional regression elements, which are simply an augmented version over the age groups of the unidimensional structures, we can then either estimate transition coefficients for each age group independently or enforce the smoothness of the series of age-specific coefficients by adding a roughness penalty (Eilers and Marx 1996).

Figure 6 presents transition coefficients over age for the three proposed assumptions: (1) constant coefficients for all ages; (2) independent coefficients for each age group; and (3) smoothly changed coefficients over age groups. In the presented example, we apply the last two solutions to data from age 30 onward, which aggregate 99.93 and 99.86% of total amount of deaths in association FAI #852 for 1978 and 1979, respectively. The first solution (blue horizontal lines) assumes constant transition coefficients over age groups, which differ considerably from the independently estimated coefficients (red dots), suggesting that the assumption of fixed proportionality over age is too strong.

However, independently estimated transition coefficients over age groups seem to wiggle with respect to the natural regular pattern of mortality over ages. Therefore, if we assume that volatility in transition coefficients is merely due to random fluctuations, enforcing the smoothness of the transition coefficients over age groups (red lines) seems to be an appropriate compromise between fixed proportionality and the randomness of the independent age-specific estimate.

3.4 Identifying Suspicious Discontinuities

The description of the method in Sect. 2 suggested that the time series require thorough checking at several steps in order to maintain statistical continuity. So far, mostly visual inspection has been used to detect suspicious disruptions.

A study by Rey et al. (2011) proposed an automated detection of jumps developed earlier by Zhang et al. (2009) under the name Polydetect. The method is based on local polynomial regression with linear kernel smoother and bootstrap estimation of the bandwidth. The validity of the jump is then evaluated based on comparing the signal-to-noise ratio at the given time point to a threshold value *alpha*. However, the Polydetect fitting and optimization method is computationally intensive, requires subjective choices about several sub-procedures (choice of kernel, choice of the level of *alpha*, and choice of the number of bootstraps for bandwidth estimation), and performs poorly on sparse and irregular data. Automatic detection of ICD-related disruptions based on regression prediction techniques was also attempted by Barbieri et al. (2008).

To fully suit the needs of the reconstruction method at different levels of detail, including the breakdown by age and sex, a method for disruption detection based on nonparametric P-spline smoothing was developed (Camarda and Pechholdová 2014). In this approach, cause-specific log-mortality trends before and after the classification change are considered in a Poisson distribution using the population size as offset. Working with a Poisson distribution allows taking into account the actual size of the process and adjusting the level of smoothing accordingly (the lower the death counts, the greater the amount of smoothing). The trends on both sides of the revision year are then forecasted to the middle of the revision year, assuming that without the revision they would smoothly follow the estimated trend for (at least) half a year. Based on these estimates, an expected difference is computed along with its confidence interval. If the confidence interval contains the value of 0, then the difference is considered as statistically nonsignificant (at the given confidence level).

Similarly to the Polydetect method, the choice of the confidence level is arbitrary and therefore subjective. However, to minimize the subjectivity, we have undertaken an unprecedented experiment: several experts involved in the reconstruction of mortality time series were asked to visually assess the disruptions in a blind test (based on a subset of 100 causes of death). We compared the results of this multiple visual inspection to different confidence levels of our method in order to obtain the setting that is as close as possible to the expert opinions. Based on the evaluation of this experiment, we have arrived at an “ideal” level of 80% confidence.

The method was implemented in R software and included some additional features (such as an option for GAM smoothing, a parameter for sizing the standard population and the confidence level). Most importantly, the program produces a set of graphs for all tested items (either FAIs, ICD codes, or age-specific death rates) with a flagged legend, which becomes red in cases where a significant disruption is detected.

Fig. 7 Automated detection of disruptions—examples of output. Bold lines represent observed trends, and dotted lines represent smoothed estimation of the trend with respect to the stochasticity

Fig. 8 Examples of associations in which at least 1, 2, 3, or 4 experts have identified a disruption and where the model with 80% confidence level did not identify disruption

An example of the output of the method is presented in Fig. 7 (the check for France was performed on the associations between ICD-9 and ICD-10). The right panel shows a disruption-free trend (FAI #232), while the left panel provides an example of a clear disruption detected by the method (FAI #269). In both cases, the 80% confidence level was used.

However, when automated detection is compared to expert opinion, some dubious cases arise. Figure 8 provides four examples of trends used in a blind test (experts vs. algorithm), in which a given number of experts (from at least one to at least four) considered the trends to be “disrupted” or “suspicious,” while they passed undetected by the method even at the 80% confidence level. These (rare) false negatives are inherently unavoidable in particularly irregular trends, especially when irregularities occur around transition years.

The method has recently become a part of the automatized production of continuous cause-specific time series. A positive detection result usually indicates a need to modify the correspondence table. However, every detected disruption is always re-evaluated in its specific medical and statistical context by the country specialist in charge.

4 Conclusion

The cause-of-death classification revisions seriously impact statistical counts and represent the greatest obstacle to understanding long-term changes in cause-of-death trends. In order to establish continuous time series, it is necessary to produce a complete analysis of all changes introduced in the classification and to estimate their impact on the statistical series while respecting both the theoretical medical contents of every cause-of-death item (before and after changes) and their actual statistical counts. This is an enormous task, especially if there is no information available about the expected exchanges (such as a bridge coding or a translator). Until recently, all the work was done manually and was thus limited to a very low number of countries. To be able to extend it to a greater number of cases, it was necessary to find a means to make this easier.

The proposed alternative approaches based on time series methods seem to be inapplicable if our aim is to work in great detail while respecting the medical definitions of items as much as possible. Nor can the direct use of bridge coding itself provide a quick and acceptable solution, as many ICD items are missing or involved in very complex exchanges. Finally, despite the existence of an official correspondence table produced by the WHO to facilitate the transition between ICD-9 and ICD-10 (the WHO Translator), a simple solution remains out of reach due to, again, very complex exchanges between items and to the persistence of country-specific coding practices, both of which exclude the possibility of adopting any universal correspondence table. It seems unrealistic to find any global solution that can respect the actual changes and avoid any smoothing that may be misleading.

Given these difficulties, we are proposing an automated and standardized extension of the detailed a posteriori cross-tabulation method by Meslé and Vallin.

To accomplish this, we have fully automatized several steps of the method, specifically: constructing and visualizing fundamental associations of items, computing transition coefficients, and detecting statistical disruptions. We have tested the possibilities of combining the existing information from the WHO Translator with information from bridge coding studies in order to build a realistic initial correspondence table between all ICD-9 and ICD-10 items. However, defining a correspondence table perfectly suited to the observed data in the given national context cannot be further automatized; thus, the table remains the main input required in terms of the amount of labor and expertise.

The availability of standard and automatic routines represents a great step forward in the field of reconstructing continuous time series. In several countries such as France, Japan, Spain, Germany, Poland, and Russia, the continuous series of causes of death classified according to ICD-10 have either been recently reconstructed or the work is at an advanced stage. Reconstructing several countries at once using the same methodology allows researchers to share existing resources (such as correspondence tables) and to collect information about repetitive issues and their solutions, which further accelerates the work. Using the automated options proposed here, the method was also used to produce most of the data series analyzed in this special issue.

The question of bridging two successive ICD revisions has become prominent with the adoption of ICD-10 and the numerous changes within it, including the unprecedented increase in the level of classification detail, the reorganization of the main chapters, and the changes to the coding rules. Currently, the WHO is preparing to adopt a new ICD: the 11th revision. In the near future, we can thus expect another major disruption in cause-specific trends and losses in historical continuity. It therefore remains highly desirable to establish an exact, functional, and reasonably demanding method for bridging statistical discontinuities caused by changes to the ICD.

Acknowledgements This research was supported by the French Institute for Demographic Studies (INED) and the Max Planck Institute for Demographic Research (MPIDR). This collaboration was supported by two research grants: Project ANR-12-FRAL-0003-01 “Diverging Trends in Mortality and Future Health Challenges” (DIMOCHA). AXA project “Mortality Divergence and Causes of Death” (MODICOD).

References

- Barbieri, M., & Meslé, F. (2008). Comparing long term trends in cardiovascular mortality in three high-income countries, paper at the *Second human mortality database symposium*, Rostock, Allemagne, 13–14 June 2008.
- Barbieri, M., Robert, C., & Boe, C. (2008). Automating the redistribution of deaths by cause over ICD changes. Paper at the Second Human Mortality Database Symposium, Max Planck Institute for Demographic Research, Rostock, Germany, 13–14 June 2008.
- Bâzgan, V., & Penina, O. (2016). VBA macro for building fundamental associations of items. Paper at *MODICOD closing seminar*, Rostock, MPIDR, Jan 25–26, 2016.
- Boissier de Lacroix (*alias* Sauvage) (1730). *Nouvelles classes de maladies qui dans un ordre semblable à celui des botanistes comprennent les genres et les espèces de toutes les maladies avec leurs signes et leurs indications*. (Ouvrages de la bibliothèque universitaire de Montpellier: EC 83 (8°)).

- Camarda, C. G., & Pechholdová, M. (2014). Assessing the presence of disruptions in cause-specific mortality series: A statistical approach. Paper at the *79th annual meeting of the population association of America*. Boston (USA), 2014.
- Eilers, P. H. C. (2005). Unimodal smoothing. *Journal of Chemometrics*, 19(5–7), 317–328.
- Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, 11(2), 89–102.
- Grigoriev, P. (2016). A function to assess the modifications of the associations. paper at the *MODICOD closing seminar, Rostock*, MPIDR, Jan 25–26, 2016.
- Janssen, F., & Kunst, A. E. (2004). ICD coding changes and discontinuities in trends in cause-specific mortality in six European countries, 1950–1999. *Bulletin of the World Health Organization*, 82(12), 904–913.
- Meslé, F., & Vallin, J. (1993). Causes de décès : de la 8^e à la 9^e révision, deux cas différents, la France et l'Angleterre, in Alain Blum & Jean-Louis Rallu (ed.), *Démographie européenne. Volume II. Dynamiques démographiques*, (pp. 421–445). Paris, John Libbey/INED. (Congresses and Colloquia, 9).
- Meslé, F., & Vallin, J. (1996). Reconstructing long-term series of causes of death: The case of France. *Historical Methods, a Journal of Quantitative and Interdisciplinary History*, 29(2), 72–87.
- Meslé, France, & Vallin, Jacques. (2008). The effect of ICD-10 on continuity in cause-of-death statistics. The example of France. *Population-E*, 63(2), 347–359.
- Meslé, F., & Vallin, J. (2012). Mortality and causes of death in 20th century Ukraine. Dordrecht, Heidelberg, London New York, Springer, 279 p. (Demographic Research Monograph).
- NCHS (National Center for Health Statistics). (2001). Comparability of Cause of Death between ICD-9 and ICD-10: Preliminary Estimates, *National Vital Statistics Reports* 49 (2), 1–32. (Department of Health and Human Service).
- Pechholdová, M., Camarda, C. G. (2014). Construction associations automatically, paper at the *Modicod/Dimocha workshop*, INED, Paris (France), October 2014.
- Remund, A. (2016). C Visualize, analyse and simplify associations using social networks tools. Paper at the *MODICOD closing seminar, Rostock*, MPIDR, January 25–26, 2016.
- Rey, Grégoire, Aouba, A., Pavillon, Gérard, Hoffmann, R., Plug, I., Westerling, R., et al. (2011). Cause-specific mortality time series analysis: A general method to detect and correct for abrupt data production changes. *Population Health Metrics*, 9, 52.
- Vallin, J., & Meslé, F. (1988). *Les causes de décès en France de 1925 à 1978*. Paris, INED/PUF, 608 p. (Travaux et Documents, Cahier 115).
- van der Stegen, R. H. M., Koren, L. P. H., Harteloh, P. P. M., Kardaun, W. P. F., & Janssen, F. (2014). A novel time series approach to bridge coding changes with a consistent solution across causes of death. *European Journal of Population*, 30(3), 317–335.
- Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. In: *Structural analysis in the social sciences* 8. Cambridge, New York: Cambridge University Press.
- WHO (World Health Organization). (1968). *Manual of the international statistical classification of diseases, injuries, and causes of death: Based on the recommendations of the Eighth Revision Conference, 1965, and adopted by the Nineteenth World Health Assembly* (Vol. 1, p. 501). Geneva, Switzerland: WHO.
- WHO (World Health Organization). (1977). *Manual of the international statistical classification of diseases, injuries, and causes of death: Based on the recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-ninth World Health Assembly. Volume 1*. Geneva, Switzerland, p. 773.
- WHO (World Health Organization). (1978). *Manual of the international statistical classification of diseases, injuries, and causes of death: Based on the recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-ninth World Health Assembly. Volume II. Alphabetical Index—Geneva* (Switzerland).
- WHO (World Health Organization). (1986). *International classification of diseases, ninth revision. Clinical modification*, vols. 1 and 2 update. (HTLV-III/LAV Infection codes. Official Authorized Addendum. Geneva: 20).
- WHO (World Health Organization). (1997). *International classification of diseases: Translator: Ninth and Tenth Revisions*. Geneva, World Health Organization.
- Zhang, B., Su, Z., & Qiu, P. (2009). On jump detection in regression curves using local polynomial kernel estimation. *Pakistan Journal of Statistics*, 25(4), 505–528.