Application of Multiple Cause of Death Information to Eliminate Garbage Codes

Agnieszka Fihel[1,2], Magdalena M. Muszyńska-Spielauer[3]
[1]Centre of Migration Research, University of Warsaw; [2]Institut des Migrations, Paris;
[3] Vienna Institute of Demography/Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna)
Corresponding author: Agnieszka Fihel, a.fihel@uw.edu.pl

Introduction

International studies of mortality rely on the comparability of cause-of-death data between countries. European countries follow the World Health Organization's (WHO) guidelines on registering deaths and adopt the same classification of causes of death as the WHO's International Statistical Classification of Diseases and Related Health Problems (ICD). Despite the implementation of concordant rules and the ICD, important differences in the quality of cause-of-death data hinder international comparisons; in Europe, the main problem stems from the assignment of so-called garbage codes (GCs), that is causes of death that are not useful in analyses of public health and mortality (Murray and Lopez, 1996).

Motivation and research objective

For several years now, WHO's international comparisons of mortality exclude Poland (Mathers et al., 2005; Naghavi et al., 2010) because of the high prevalence of GCs: almost 20% of deaths are annually assigned to ill-defined cardiovascular symptoms and another 6-7% to symptoms, signs and ill-defined conditions. This study aims to reclassify cardiovascular GCs in Poland into well-defined causes of death and investigate cardiovascular mortality structure by specific causes. The reclassification is based on a unique dataset including all original MCoD mentions in death certificates, which in Poland are usually destroyed for the sake of confidentiality as soon as a medical coder has defined and validated the code of the underlying cause of death. However, the scans of death certificates for 2013 were preserved and made available for this research.

Data and method

The dataset concerns 387,988 permanent residents of Poland deceased in 2013 in the territory of Poland. The contents of scanned death certificates, including the deceased's demographic characteristics, circumstances of death, all MCoD mentions made by medical doctors pronouncing the death and the ICD code of the underlying cause defined by medical coders, were digitalized. The MCoD mentions consist of both original descriptions and their ICD codes established for the purpose of quantitative analysis.

The study refers to all conditions recognized by the WHO (2013) as cardiovascular GCs (Table 1). The analysis consisted of two steps. First, all cardiovascular GC individual records were manually verified, where a detailed description of well-defined medical conditions was provided and constituted a logical chain of events leading to death, the code of the underlying cause was replaced with a well-defined one from the given death certificate. Second, coarsened exact matching was applied (Snyder et al., 2014; Stevens et al., 2010) to death counts that had not been reclassified in the previous step: each GC death record was matched with all death records that had been registered due to both a well-defined underlying cause and a contributing

mention including the same GC. Controlling for sex and 5-year age groups (up to 85 and over), cardiovascular GC deaths were redistributed across the well-defined causes they matched with proportionally to the number of established combinations. For instance, 29,167 deaths certified to heart failure (I50) were matched in 60,480 combinations with well-defined causes, including 2,070 combinations with cardiomyopathy (I42). Allowing for age and sex, 3% (2,070: 60,480) of deaths initially assigned to heart failure were reclassified as due to cardiomyopathy. Both redistribution steps were conducted at the accuracy level of four-digit ICD codes (i.e. I70.9 instead of I70).

Without additional medical documentation evidencing individual morbid histories, the possibility of verifying GC redistribution validity is limited. We, therefore, examined the accuracy of our redistribution by comparing the results to mortality patterns observed in three other Eastern European countries: the Czech Republic, Hungary and Slovakia. Death counts by cause and population estimates for these countries were both derived from the WHO Mortality Database (2020). Age-specific death rates and age-standardized death rates (ASDRs) were compared, the latter being based on the European standard population for 2013. To assure comparability, deaths due to symptoms, signs and ill-defined conditions (ICD-10 codes starting with R) and cardiovascular GCs for these three countries were redistributed proportionally across all other causes.

Results

In 2013, 86,856 deaths, which constituted 22% of all deaths in Poland, were assigned to fourteen cardiovascular GCs. The most prevalent GCs were heart failure (I50) and generalized and unspecified atherosclerosis (I70.9, Table 1). These two GCs were also most frequently corrected in the first step of the reclassification. Heart failure was frequently recorded as an underlying cause of death despite the coexistence of well-defined mentions such as chronic ischaemic heart disease (I25), atrial fibrillation and flutter (I48) and vascular dementia (F01). Generalized and unspecified atherosclerosis, in turn, was often registered despite well-defined contributing mentions of atherosclerotic heart disease (I25.1), atherosclerosis of well-defined arteries (I70.0-8) and cerebrovascular diseases (I60-I69).

The second step of the reclassification, coarsened exact matching, covered 61,249 GC deaths that were combined with 131,418 death records due to well-defined underlying causes; 9.8 million combinations were established. Redistributions were made predominantly to well-defined cardiovascular diseases (58%), in particular ischaemic heart diseases (chronic heart disease, I25, and acute myocardial infarction, I21), hypertensive heart disease (I11) and cerebrovascular diseases (cerebral infarction, I63, and stroke, not specified as haemorrhage or infarction, I64). Deaths originally registered as cardiovascular GCs were also redistributed to other groups of diseases, primarily cancers (17%), such as malignant neoplasm of bronchus and lung (C34), colon (C18), and breast (C50), and endocrine, nutritional and metabolic diseases (7%), in particular diabetes mellitus type 2 (E11). Altogether, in both manual correction and coarsened exact matching, death records were redistributed to well-defined cardiovascular diseases (59%), cancers (12%), endocrine, nutritional and metabolic diseases (6%), and respiratory diseases (5%).

As a result of the redistribution, mortality due to well-defined cardiovascular diseases increased considerably (Figure 1): the ASDR for ischaemic heart diseases rose from 224 to 308 per

100,000 for men and from 120 to 179 per 100,000 for women. For cerebrovascular diseases, the ASDR rose from to 148 to 178 per 100,000 for men and from 118 to 147 per 100,000 for women. In contrast, the unusually elevated mortality due to diseases of arteries, arterioles and capillaries, driven mostly by generalized and unspecified atherosclerosis, decreased almost fourfold for men and fivefold for women. Mortality due to other forms of heart disease, with cardiac arrest and heart failure as the most important causes, decreased threefold for both sexes. After redistribution, cardiovascular mortality structure by large groups of causes in Poland became similar to the structure observed in other countries, with ischaemic heart diseases and cerebrovascular diseases dominating. However, ASDR due to ischaemic heart diseases still remained significantly lower in Poland than in other countries studied.

As cardiovascular GCs were redistributed to all chapters of ICD classification, overall cardiovascular mortality decreased by 12% for men and 15% for women and became lower than in other countries. Concurrently, mortality levels due to other large groups of diseases, notably cancers and respiratory diseases, increased but remained similar to levels registered in other countries.

Conclusions and discussion

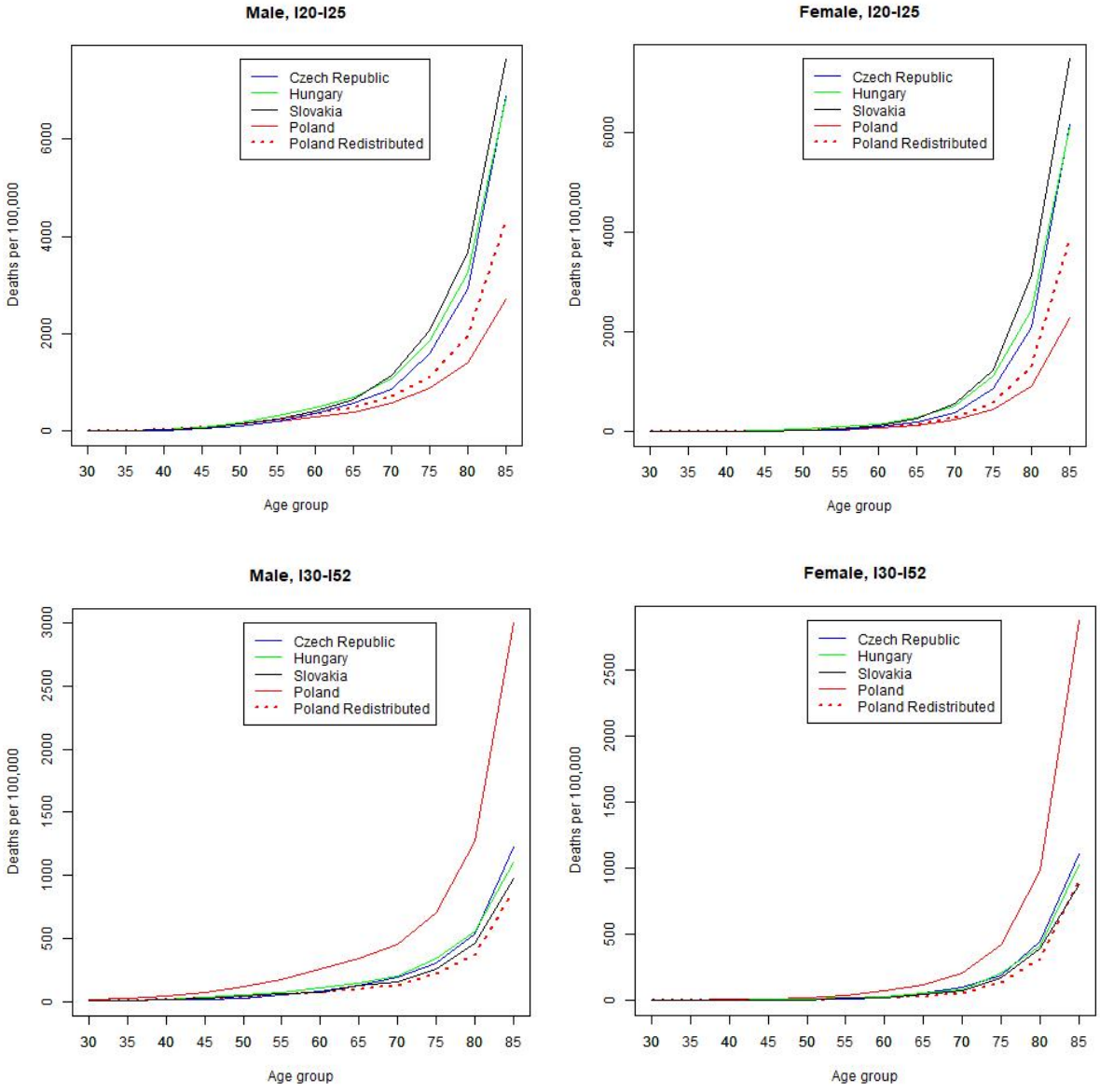
When the prevalence of GCs is considerable, the standard approach to mortality accounting only for well-defined causes with proportionally redistributed GCs yields questionable results. This applies particularly to countries without complete vital registration systems, but also concerns several European countries (Naghavi et al., 2010). To conduct international comparisons of cause-of-death mortality, more and more complex methodologies for GCs are being developed (Vos et al., 2020), most often based on regression models (Ahern et al., 2011; Fink et al., 2012; Foreman et al., 2016; Murray et al., 2008, 2006) and hence imposing *a priori* assumptions on the functional form of redistribution models (Stevens et al., 2010). Other studies employ expert judgement to establish possible medical associations between causes of deaths and verify these associations based on empirical data (Ahern et al., 2011; Naghavi et al., 2010). In contrast, the coarsened exact matching approach using MCoD is a relatively simple missing data replace method but its advantage over the abovementioned techniques lies in the fact that it is based directly on the observed relationships between causes of death. However, due to low availability and accessibility of MCoD data in most countries with a high prevalence of GCs, this method has previously been applied in only two studies: in both cases to heart failure in North or South American countries (Murray et al., 2008; Snyder et al., 2014).

In this study 22% of all deaths in Poland in 2013 – those originally registered due to cardiovascular GCs – were re-assigned to a well-known coexisting disease based on the MCoD data. As a result, the level of well-defined cardiovascular mortality increased, and the structure of cardiovascular mortality by groups of diseases became similar to those observed in countries with a low prevalence of GCs and at the same stage of the health transition. Hence, we demonstrated that the presented redistribution improved the comparability of cause-of-death data between countries with different data quality.

Cardiovascular GC and its ICD-10 code	Original death counts	Death counts after 1st step correction
Essential (primary) hypertension, I10	978	978
Pulmonary embolism without mention of acute cor pulmonale, I26.9	1,643	1,643
Cardiac arrest, I46	6,939	5,539
Ventricular tachycardia, I47.2	14	12
Ventricular fibrillation and flutter, I49.0	84	17
Heart failure, I50.0,1,9	38,372	29,167
Complications, ill-defined descriptions of heart disease, I51.4,5,6,8	4,342	3,309
Generalized and unspecified atherosclerosis, I70.9	34,407	20,507
Other and unspecified disorders of circulatory system, I99	77	77
Total	86,856	61,249

Table 1. Number of death counts due to cardiovascular Garbage Codes under study, before and after first-step correction, Poland 2013

Source: own elaboration based on the MCoD data from Statistics Poland.



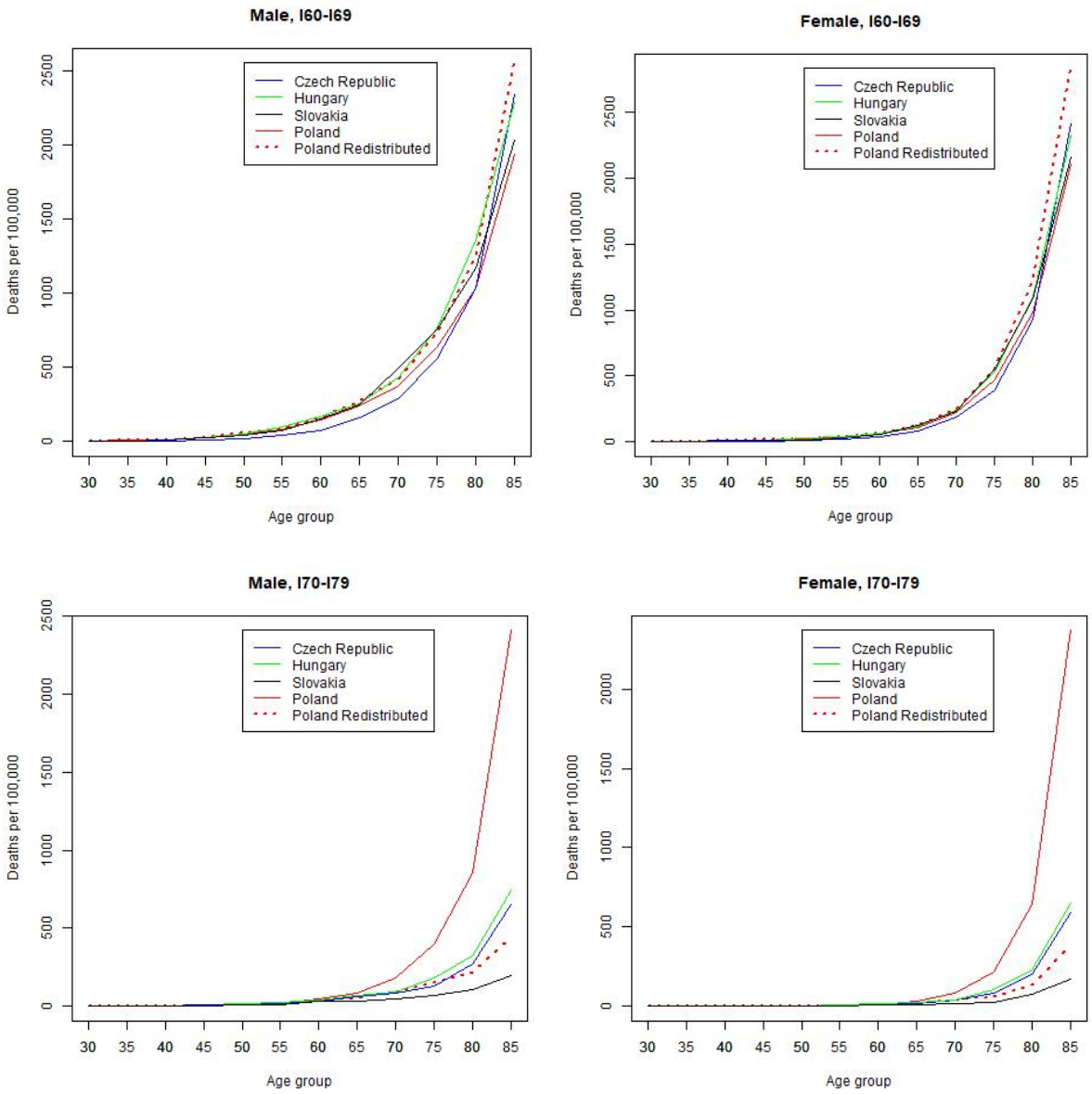


Figure 1. Age-specific death rates due to selected groups of cardiovascular diseases[1,2] in selected countries, 2013, by sex (per 100,000)

Notes: [1]Ischaemic heart diseases I20-I25, other forms of heart disease I30-I52, cerebrovascular diseases I60-I69, diseases of arteries, arterioles and capillaries I70-I79; [2]With symptoms, signs and ill-defined conditions (R codes) redistributed proportionally across all causes of death in all countries, with cardiovascular GCs redistributed proportionally across all causes of death in the Czech Republic, Hungary and Slovakia.

Source: own elaboration based on the MCoD data from Statistics Poland.