

Redistributing Ill-defined Causes of Death – A Case Study from the Burden 2020-Project in Germany

Annelene Wengler (✉ wenglera@rki.de)

Robert Koch Institute: Robert Koch Institut <https://orcid.org/0000-0001-8260-7794>

Heike Gruhl

German Environmental Agency

Dietrich Plaß

Germany Environmental Agency

Janko Leddin

Robert Koch Institute: Robert Koch Institut

Alexander Rommel

Robert Koch Institute: Robert Koch Institut

Elena von der Lippe

Robert Koch Institute: Robert Koch Institut

Research

Keywords: Burden of Disease, redistribution, cause of death, ill-defined death, mortality

Posted Date: December 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-117779/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on March 15th, 2021. See the published version at <https://doi.org/10.1186/s13690-021-00535-1>.

Abstract

Background

The cause of death statistics in Germany include a relatively high share (26% in 2017) of ill-defined deaths (IDD). To make use of the cause of death statistics for Burden of Disease calculations we redistribute those IDD to valid causes of death.

Methods

The process of proportional redistribution is described in detail. It makes use of the distribution of the valid ICD-codes in the data. We use examples of stroke, diabetes, and heart failure to illustrate how IDD are reallocated.

Results

The largest increases for both women and men can be found for *lower respiratory infections, diabetes mellitus, and stroke*. The numbers of deaths for these causes more than double after redistribution.

Conclusion

This is the first comprehensive redistribution of IDD within the German cause of death statistics. Performing a redistribution is necessary, otherwise there would be an underreporting of certain causes of death or large numbers of deaths coded to residual or unspecified codes.

Background

Globally, burden of disease (BoD) analyses are performed to assess the health state of populations [1]. Applying a standardized concept which covers all relevant health impairments allows comparing different diseases and injuries as well as related risk factors [2]. A key component of BoD studies is the summary measure of population health the *Disability-Adjusted Life Year* (DALY). DALY summarize the amount and severity of health problems experienced by a population and consider both fatal and non-fatal health outcomes [3]. The effect of the fatal health outcome is expressed by using the *Years of Life Lost*, the non-fatal health outcome is measured through *Years Lost due to Disability* (YLD).

The calculation of the YLL is usually based on the cause of death (CoD) statistics, which use information from death certificates. In most cases the death certificate includes more than one cause and often the full chain of events leading to death. In Germany the CoD are classified according to the principles of the International Statistical Classification of Diseases and Related Health Problems (ICD-10, WHO Version). Besides the physician certifying the death, specialized coders in the statistical and health offices, as well as supporting software may influence the decision on the main, underlying CoD, that is then transferred to the national CoD statistics [4]. The underlying CoD should be the starting point of the chain of events leading to death. Secondary preceding causes and comorbidities recorded on the death certificate are not included in the nationally reported CoD statistics in Germany [5]. Furthermore, in some federal states the coding of the underlying CoD is done electronically through the implementation of specific software (Iris/MUSE). However, many federal states are still coding manually [6–9]. This may result in imprecisions.

Due to various reasons some ICD-10-codes in the CoD statistics are considered to be not sufficiently informative or valid for BoD estimations – sometimes referred to as garbage codes [10]. In the following, we refer to them as ill-defined (causes of) death or IDD. These codes describe conditions which cannot or should not be considered as an underlying CoD [11, 12]. Reasons for this might be missing information regarding the death or lack of training of the person coding the death [13]. A more detailed description on the major types of IDD can be found in Appendix 1.

Though this phenomenon occurs in all CoD statistics worldwide the amount of the IDD varies largely across countries [14–16]. Comparing the shares of IDD in the CoD statistics for the years 2015 or 2016 from six countries with rather advanced

health systems, Mikkelsen et al. revealed, that the share of IDD in Germany (26%) ranges between the shares found in Canada (22%) and in Japan (36%) [15]. Depending on the amount of IDD, the CoD statistics may not accurately reflect a country's mortality, hampering comparisons or leading to biased priorities [17]. In consequence, as the underlying cause is not clearly identifiable large amounts of deaths with residual or unspecific codes may not be considered when deriving specific public health measures. Furthermore, the actual importance of certain CoD may be largely underestimated. IDD in this respect challenge BoD studies worldwide because a valid recording and reporting of CoD in a population, is the basis for calculating YLL and hence BoD estimations.

The *Institute for Health Metrics and Evaluation (IHME)*, responsible for the *Global Burden of Diseases, Injuries, and Risk Factors (GBD) Study*, has provided a thorough classification of the IDD which is updated with each cycle of the GBD study. Before calculating YLL these IDD are redistributed to valid CoD, using different statistical methods and employing profound and comprehensive algorithms [10, 16]. In general, the methods used for the redistribution in the GBD study include proportional reassignment, fixed proportions, regression models, experts' opinions, and fractional assignment of a death assigned to multiple causes [10]. The World Health Organization developed a different classification of IDD [18], still resulting in a relatively high share of deaths being classified as IDD in Germany [19].

Besides the GBD study, national BoD assessments are becoming increasingly available [20–22]. The project BURDEN 2020 – Burden of disease in Germany at national and regional level – is piloting a national BoD study for Germany [23]. The identification of IDD in this study follows the methodology provided by the GBD study [24, personal communications with M. Naghavi 2019]. Furthermore, part of the GBD study's redistribution methods are adopted and applied to the German context. In contrast to the GBD study's calculations for Germany, in BURDEN 2020 the YLL calculation and thus the IDD redistribution are performed not only on national but also on subnational level. The aim of the present article is to provide an in-depth description of the procedures for redistributing the IDD to valid codes in BURDEN 2020 and the impact this has on the case numbers for specific CoD in Germany.

Methods

The German CoD statistics provide only one CoD, the underlying CoD. In BURDEN 2020 we therefore chose a redistribution method which refers to this one underlying CoD and makes use of the proportional distribution of deaths across valid ICD-codes. In general, 4 steps need to be taken to adjust the CoD statistics for calculating BoD: **1)** defining and grouping IDD in IDD packages, **2)** defining valid target codes (reflecting the probable *true* underlying CoD) for each IDD package, **3)** deciding on and applying a redistribution methodology, **4)** structuring and grouping of ICD-codes to form suitable CoD. We follow the definition of IDD and the corresponding target codes (step 1 and 2) from the GBD study but chose our own approach for the redistribution (step 3). The grouping of ICD-10-codes to cause groups (step 4) has also been adapted from the GBD study. Hence, we shortly describe the four steps of utilizing CoD statistics for BoD estimations in the following, before we describe the complex redistribution processes (here step 3) in detail in the second part of the Methods section.

1) IDD definition and grouping in IDD packages

The GBD list of IDD contains more than 7,000 ICD-10-codes (3- and 4-digit codes) of which 859 actually occur in the German CoD statistics in 2017. Overall, 932,269 deaths were registered in 2017 and 26% of all deaths were coded as IDD [19]. The IDD are grouped in IDD packages. This grouping is based on the thematic association of certain ICD-codes and entails that all deaths belonging to one package are redistributed together following the same procedure. For each package a set of so-called target codes (see following section) is defined.

Depending on the type of IDD (impossible, intermediate, immediate, and unspecified cause; see Appendix 1), different objectives are pursued with the redistribution. Impossible IDD should be more generally allocated to various plausible CoD. The aim of handling sequela (intermediate and immediate causes) is to trace back to the underlying CoD. Unspecified

causes should be transferred to more specific ones. In the correction process age, sex, and regional (place of residency) assignments are not changed.

2) Definition of target codes for IDD

The IDD have to be reassigned to valid ICD-codes and consequently to valid CoD. The valid codes are called target codes and for each IDD package a set of target codes is defined. Target codes can be considered as the probable *true* underlying CoD in case that a specific IDD was coded. BURDEN 2020 follows the GBD methodology in this step and thus, uses the same sets of target codes [16]. In general, the definition of the target codes requires understanding the pathology and epidemiology of the IDD. The target codes were made available to the BURDEN 2020 project through personal communication with Mohsen Naghavi as part of a memorandum of understanding between the Robert Koch Institute and IHME.

3) Decision regarding redistribution method

In BURDEN 2020 we apply a proportional redistribution method for dealing with IDD. This means that the distribution of the IDD to target codes starts from the empirical proportion of valid codes as reported in the death register. In Fig. 1 the redistribution of the IDD is depicted by using an example. The target codes for the IDD *unspecified stroke* can be grouped in three specific stroke groups: *ischemic stroke*, *intracerebral hemorrhage*, and *subarachnoid hemorrhage*. In our example, ischemic stroke accounts for 59.9% of all valid codes. Accordingly, 59.9% of all IDD are redistributed to the group of ischemic strokes. Hence, the original proportion remains stable meaning that the distribution of valid codes is the same before (in blue) and after (in red) the redistribution. The main assumption in the proportional redistribution is, that the empirical distribution of target codes represents the actual distribution of CoD in the population. The redistribution with all necessary steps to be taken will be explained in detail in the following section (*Approach for the redistribution in BURDEN 2020*).

4) Structuring and Grouping of ICD-codes

The CoD statistics in Germany are available on a very detailed level of ICD-10 (four-digit codes). However, to get a comprehensive picture of CoD in Germany and to increase usability for public health concerns, this information needs to be aggregated and simplified. For this purpose, the hierarchical organization of the CoD at different levels, adopted from the GBD study, is implemented in BURDEN 2020 (Fig. 2) [16]. At the highest level (level 1), all valid ICD-10-codes are classed into three broad cause categories: 1) Communicable, maternal, neonatal, and nutritional diseases (CMNN), 2) Non-communicable diseases (NCD), and 3) Injuries. Level 2 disaggregates these level 1 causes into 21 cause groups. NCD for example are subdivided into cardiovascular diseases, neoplasms, chronic respiratory diseases etc. (see Table A1 in appendix). On level 3 cardiovascular diseases for example are further distinguished as among others in ischemic heart disease or stroke. For some ICD-10-codes level 3 is the most detailed cause level. Where more detailed data are available or specific policy requirements exist a further disaggregation at level 4 is possible [24; personal communications with M. Naghavi 2019]. Diabetes for example on level 4 is divided into diabetes mellitus Type 1 and Type 2. For stroke we can differentiate between ischemic stroke, intracerebral hemorrhage, and subarachnoid hemorrhage on level 4.

This cause hierarchy (see Fig. 2 and Table A1 in the appendix) also is the framework for presenting results on the disease burden related to cause specific mortality. It is comprehensive in the way that all valid ICD-codes are assigned to one exclusive cause on each level [16]. We adapted this GBD mapping of ICD-codes to CoD and refer to the different levels in the results section of this paper and other publications [25, 26].

Approach for the redistribution of IDD in BURDEN 2020

A step-by-step description of the redistribution procedure applied in BURDEN 2020 is presented in Fig. 3. The share of 26% IDD of all deaths in 2017 is grouped and assigned to target codes (see section before). As the IDD are grouped in 166 different packages (159 occurring in the German data), each package has a clearly assigned set of target codes whereas target codes are not exclusively assigned to only one package. Hence, specific codes can be assigned as target codes for several IDD packages.

The IDD packages are redistributed successively, shifting IDD to valid ICD-codes. In this way, the redistribution procedure is a stepwise reassignment. Accordingly, the number of valid ICD-codes increases with each redistributed package. Since more cases are defined as valid after each redistribution step, the number of cases forming the empirical distribution of valid target codes for the next step of the redistribution increases consecutively.

Beyond the rational of using distinct packages, the redistribution is carried out age and sex specific. For instance, when reallocating unspecified stroke which is considered an IDD in women aged 82 we use the valid distribution of target codes within the group of women aged 80 to 84. Furthermore, the redistribution is performed on a subnational level (here federal states) which assures that regional variations in the CoD statistics are considered (see below).

Example for redistribution: Stroke

For most IDD packages a set of general target codes is defined which are neither age nor sex specific. One example outlined here is a package that encompasses all IDD belonging to the category of *unspecified stroke*. Different target ICD-codes are assigned to this package which can be grouped to three different causes: *ischemic stroke*, *intracerebral hemorrhage*, and *subarachnoid hemorrhage* (see Fig. 4). All IDD belonging to the unspecified stroke package are redistributed to the same target codes. However, the proportions of those target codes vary by age and sex. Figure 4 shows the causes containing the specific target codes for unspecified stroke and their proportions in women and men aged 80 to 84. Among women 60.1% of all cases with unspecified stroke are reassigned to ischemic stroke, 33.0% are moved to intracerebral hemorrhage, and 6.9% belong to subarachnoid hemorrhage after redistribution. For men in the same age group the proportions are 58.5%, 34.4%, and 7.1%, respectively.

Example for redistribution: Gastrointestinal bleeding – with varying target codes

Some IDD packages have sets of target codes that vary by age or sex. *Gastrointestinal bleeding* is only reassigned to *other diarrheal diseases* and *other digestive disease* for deaths under 15 years of age (here illustrated for boys aged 1–4). However, for deaths at the age of 15 or older (here displayed for men aged 80–84) a larger set of target codes and resulting underlying CoD is assumed, including *colon and rectum cancer*, *stomach cancer*, and *cirrhosis and other chronic liver diseases* (see Fig. 5).

Subnational redistribution

Redistribution is not only carried out within each sex and age group but is also performed on a subnational level. In the project BURDEN 2020 [26] we aim to report BoD estimates for the 96 German spatial planning regions (SPR). However, in less populated SPR empirically only very few or no deaths may be assigned to the target codes, especially in younger age groups. This results in a distribution with missing target codes (e.g. 0% of cases), which makes a redistribution of IDD to those valid codes impossible. To overcome this problem, still taking regional variation into account, we chose the distribution of valid ICD-codes of the 16 federal states for subnational redistribution. Through the last place of residency of the deceased, we can identify the SPR as well as the federal state. Hence, for redistributing IDD age and sex specific empirical proportions of target codes on the federal state level are used (see example in Fig. 6). Additionally however, the results can still be presented for each SPR.

Estimation of the uncertainty intervals

In Fig. 3 *estimating uncertainty* is presented as a separate process. More specifically, we want to estimate uncertainty that evolves from redistributing IDD to valid target codes. We consider this by randomly reassigning new codes and repeating this

procedure 1,000 times. In practice, to each person with an IDD we generate 1,000 random numbers between 0 and 1. Thereafter, we use the specific distribution of target codes for reassigning ICD-codes (Fig. 7).

Referring again to stroke, the distribution for women (aged 80 to 84), which was 60.1% *ischemic stroke*, 33.0% *intracerebral hemorrhage*, and 6.9% *subarachnoid hemorrhage* is redefined in intervals. Proceeding this way the random numbers can be easily used for reassigning new ICD-codes. Random numbers between 0 and 0.601 are assigned to the cause ischemic stroke, values between 0.601 and 0.931 to intracerebral hemorrhage, and values between 0.931 and 1.000 to subarachnoid hemorrhage. This means that a higher proportion of a target code results in a higher probability for assignment to this code. The uncertainty intervals can be derived by observing the distribution across the 1,000 draws in terms of minimum, maximum and mean number of deaths with a specific target code. This results in uncertainty intervals for all valid CoD. Additionally, the 1,000 draws can be used to depict uncertainty when calculating YLL [25].

Results

In the first part of the results section, we want to illustrate how the redistributed cases are reallocated from IDD to valid causes, looking at specific examples. In the second part, we take a broader perspective and compare all cases before and after redistribution by cause level.

Examples: heart failure, stroke, and diabetes

Heart failure redistribution

One of the biggest groups of IDD in Germany is heart failure. Physicians often choose heart failure as underlying CoD which results in 39,300 cases in the German CoD statistics in 2017 [14]. Heart failure is defined as an IDD because it cannot be the underlying CoD but is rather the consequence of other underlying causes (e.g. ischemic heart disease). Nevertheless, this code is largely used as a main CoD, hampering the use of CoD statistics for BoD analyses.

The defined target codes for heart failure are spread across the whole ICD catalogue (Fig. 8). Of the 39,300 heart failure cases on average almost 22,800 cases are redistributed to *cardiovascular diseases*, 6,000 to *neoplasms*, and 4,300 to *diabetes and kidney diseases*. 2,700 cases are redistributed to *chronic respiratory diseases* and 2,000 to *respiratory infections and tuberculosis*. For 2017, the largest cause group on level 2 is cardiovascular diseases of which on level 3 the largest causes are *ischemic heart disease*, *stroke*, and *hypertensive heart disease*. All heart failure IDD reallocated to neoplasms on level 2, are mainly composed of *tracheal, bronchus, and lung cancer, colon and rectum cancer*, as well as *breast cancer* on level 3.

Stroke redistribution

Besides observing which are the biggest groups of target codes for a specific IDD, it is also interesting to analyze which IDD contribute most to a concrete valid CoD. Below is the example for stroke as a cause showing a significant increase by 34,200 deaths after redistribution (Fig. 9).

The data for 2017 show that the IDD group of *unspecified stroke* that includes 24,500 cases on average contributes most to the increase of stroke as a valid CoD. Thereafter, *heart failure* contributes around 4,400 cases, the *impossible CoD* 1,200, and *hypertension* 1,000. As a result, the number of deaths from stroke has increased from 30,975 cases to 65,218 (see Table 3).

Diabetes redistribution

Diabetes is another CoD where the number of deaths increases largely due to redistribution, especially since a large amount of unspecified diabetes types are recorded on death certificates. Figure 10 indicates that the increase of diabetes cases after

redistribution is around threefold. Whereas, the heart failure (Fig. 8) and stroke (Fig. 9) examples put emphasis on the redistributed cases, Fig. 10 explicitly includes valid cases before and after redistribution. Figure A1 in the Appendix additionally depicts the precise IDD groups contributing to diabetes type 1 and 2. Besides the already stated largest group of *unspecified diabetes*, IDD belonging to the groups of *chronic kidney disease due to unspecified type*, *heart failure*, and *impossible CoD* contribute most to diabetes.

Cases before and after redistribution

Cases before and after redistribution on level 1 and 2

On level 1 all CoD are divided into three groups. Before redistribution 10,091 deaths are assigned to *communicable, maternal, neonatal, and nutritional diseases* (CMNN: 1.5%), 649,658 to *non-communicable diseases* (NCD: 93.9%) and 31,718 to *injuries* (4.6%). After redistribution 36,930 cases are defined as CMNN (4.0%), 850,534 (91.2%) as NCD and 44,805 (4.8%) as injuries (Table 1). The number of cases with CMNN as CoD displays a threefold increase. This results in 4.0% of all deaths being assigned to this group instead of 1.5% before redistribution. The share of NCD on the other hand decreases in the process of redistribution from 93.9–91.2%. However, by numbers of deaths it remains by far the largest group.

Table 1
Cases by causes of death – Level 1, before and after redistribution of cases with ill-defined causes of death

	Before redistribution	Share	After redistribution	Share
Communicable, maternal, neonatal, and nutritional diseases	10,091	1.5%	36,930	4.0%
Non-communicable diseases	649,658	93.9%	850,534	91.2%
Injuries	31,718	4.6%	44,805	4.8%
Invalid ICD-codes	240,802			
Total	932,269		932,269	
Source: causes of death statistics, Germany, 2017, own calculations				

On level 2 the ranking of the top three CoD – *cardiovascular diseases*, *neoplasm*, and *neurological diseases* – remains unchanged. In contrast, after redistribution *diabetes and kidney diseases* move two ranks up to the fourth position. *Chronic respiratory diseases* and *digestive diseases* both lose one rank after redistribution. Apart from that, the effect of the redistribution can be evaluated looking at the percent of increase. *Cardiovascular diseases* and *neoplasms* cause the most deaths in Germany and account for 64.5% of all deaths. Irrespective of the unchanged ranking for these conditions with 42% and 25% respectively, we nevertheless observe a large increase in case numbers after redistribution. The highest increases, however, can be shown for CoD which are probably underreported: *respiratory infections* (1,135% increase) and *diabetes and kidney diseases* (79% increase). The increase in respiratory infections mostly results from unspecified pneumonias being reallocated. Also, the number of *HIV/AIDS and sexually transmitted infections* increases significantly (235% increase), though this in general is only causing a small number of deaths.

Table 2 Deaths by causes on level 2 before and after redistribution of IDD

Before redistribution			After redistribution				
Rank	Causes of death (level 2)	No. deaths	Rank	Causes of death (level 2)	No. deaths	UI	%-change
1	Cardiovascular diseases	234673	1	Cardiovascular diseases	333185	(332708 - 333672)	42%
2	Neoplasms	214720	2	Neoplasms	267717	(267330 - 268200)	25%
3	Neurological disorders	72407	3	Neurological disorders	76193	(76041 - 76395)	5%
4	Chronic respiratory diseases	39924	4	Diabetes and kidney diseases	53471	(53203 - 53745)	79%
5	Digestive diseases	33974	5	Chronic respiratory diseases	49522	(49251 - 49771)	24%
6	Diabetes and kidney diseases	29797	6	Digestive diseases	41921	(41705 - 42112)	23%
7	Unintentional injuries	19758	7	Unintentional injuries	28112	(27914 - 28335)	42%
8	Other non-communicable diseases	13984	8	Respiratory infections and tuberculosis	25060	(24870 - 25246)	1134%
9	Self-harm and interpersonal violence	9128	9	Other non-communicable diseases	15611	(15492 - 15719)	12%
10	Substance use disorders	5636	10	Self-harm and interpersonal violence	12240	(12126 - 12376)	34%
11	Enteric infections	4531	11	Substance use disorders	7973	(7879 - 8067)	41%
12	Transport injuries	2832	12	Enteric infections	6942	(6849 - 7054)	53%
13	Musculoskeletal disorders	2325	13	Transport injuries	4453	(4377 - 4564)	57%
14	Skin and subcutaneous diseases	2169	14	Skin and subcutaneous diseases	2456	(2401 - 2502)	13%
15	Respiratory infections and tuberculosis	2030	15	Musculoskeletal disorders	2436	(2400 - 2465)	5%
16	Maternal and neonatal disorders	1485	16	Maternal and neonatal disorders	1591	(1567 - 1614)	7%
17	Other infectious diseases	887	17	Other infectious diseases	1338	(1279 - 1390)	51%
18	Nutritional deficiencies	817	18	HIV/AIDS and sexually transmitted infec.	1061	(1031 - 1095)	235%
19	HIV/AIDS and sexually transmitted infec.	317	19	Nutritional deficiencies	904	(883 - 929)	11%
20	Mental disorders	48	20	Mental disorders	49	(49 - 52)	3%
21	Neglected tropical diseases and malaria	24	21	Neglected tropical diseases and malaria	33	(26 - 45)	36%

Source: causes of death statistics, Germany, 2017, own calculations; UI: uncertainty intervals

Top 20 causes of death (level 3) before and after redistribution

In Table 3 the 20 most frequent CoD on level 3 before and after the redistribution of the IDD are presented. For women the top three CoD after redistribution are *ischemic heart disease*, *stroke*, and *Alzheimer's disease and other dementias*. For men these are *ischemic heart disease*, *tracheal, bronchus and lung cancer*, and *stroke*.

As before, we observe quite large variation in the percent of increase in case numbers after redistribution. The largest increases for both women and men can be found for *lower respiratory infections* (+ 1450% and + 1239%, respectively), *diabetes mellitus* (+ 224% and + 217%), and *stroke* (+ 122% and + 96%). As indicated before lower respiratory infections increase to such extent due to the great amount of unspecified pneumonias in the data.

Table 3 The 20 most frequent causes of death, for women (A) and men (B) before and after redistribution (level 3)

A – women

Before redistribution			After redistribution				
Rank	Causes of death (level 3)	No. deaths	Rank	Causes of death (level 3)	No. deaths	UI	%-change
1	Ischemic heart disease	56820	1	Ischemic heart disease	80376	(79917 - 80807)	41%
2	Alzheimer's disease and other dementias	35618	2	Stroke	38144	(37919 - 38395)	122%
3	Breast cancer	18423	3	Alzheimer's disease and other dementias	37403	(37276 - 37526)	5%
4	Hypertensive heart disease	17428	4	Hypertensive heart disease	23429	(23161 - 23700)	34%
5	Stroke	17147	5	Breast cancer	22576	(22366 - 22778)	23%
6	Tracheal, bronchus, and lung cancer	16532	6	Chronic obstructive pulmonary disease	20409	(20226 - 20575)	30%
7	Chronic obstructive pulmonary disease	15730	7	Tracheal, bronchus, and lung cancer	19724	(19545 - 19895)	19%
8	Atrial fibrillation and flutter	13633	8	Chronic kidney disease	17532	(17360 - 17744)	30%
9	Chronic kidney disease	13524	9	Colon and rectum cancer	16419	(16210 - 16595)	44%
10	Colon and rectum cancer	11366	10	Atrial fibrillation and flutter	14453	(14354 - 14544)	6%
11	Pancreatic cancer	9061	11	Diabetes mellitus	14048	(13926 - 14170)	224%
12	Falls	7699	12	Lower respiratory infections	12698	(12580 - 12808)	1450%
13	Non-rheumatic valvular heart disease	7303	13	Falls	11280	(11138 - 11397)	47%
14	Ovarian cancer	5400	14	Pancreatic cancer	11187	(11048 - 11329)	23%
15	Cirrhosis and other chronic liver diseases	5328	15	Non-rheumatic valvular heart disease	8796	(8648 - 8915)	20%
16	Parkinson's disease	4444	16	Cirrhosis and other chronic liver diseases	6658	(6559 - 6756)	25%
17	Diabetes mellitus	4334	17	Ovarian cancer	6280	(6193 - 6381)	16%
18	Stomach cancer	3745	18	Stomach cancer	5013	(4903 - 5125)	34%
19	Leukemia	3515	19	Parkinson's disease	4491	(4473 - 4510)	1%
20	Urinary diseases and male infertility	3387	20	Diarrheal diseases	4229	(4161 - 4305)	53%
22	Diarrheal diseases	2757	21	Leukemia	4113	(4051 - 4180)	17%
56	Lower respiratory infections	819	22	Urinary diseases and male infertility	3790	(3728 - 3851)	12%

Source: causes of death statistics, Germany, 2017, own calculations

B – men

Before redistribution			After redistribution				
Rank	Causes of death (level 3)	No. deaths	Rank	Causes of death (level 3)	No. deaths	UI	%-change
1	Ischemic heart disease	68794	1	Ischemic heart disease	88973	(88611 - 89339)	29%
2	Tracheal, bronchus, and lung cancer	28903	2	Tracheal, bronchus, and lung cancer	35307	(35084 - 35494)	22%
3	Chronic obstructive pulmonary disease	18755	3	Stroke	27074	(26924 - 27255)	96%
4	Alzheimer's disease and other dementias	17607	4	Chronic obstructive pulmonary disease	23129	(22921 - 23309)	23%
5	Prostate cancer	14378	5	Alzheimer's disease and other dementias	18697	(18611 - 18801)	6%
6	Stroke	13828	6	Prostate cancer	17848	(17671 - 18007)	24%
7	Colon and rectum cancer	13207	7	Colon and rectum cancer	17737	(17530 - 17927)	34%
8	Cirrhosis and other chronic liver diseases	10059	8	Cirrhosis and other chronic liver diseases	12874	(12717 - 13038)	28%
9	Pancreatic cancer	8948	9	Lower respiratory infections	11914	(11817 - 12013)	1239%
10	Chronic kidney disease	8251	10	Diabetes mellitus	11672	(11571 - 11768)	217%
11	Atrial fibrillation and flutter	7349	11	Pancreatic cancer	10776	(10637 - 10902)	20%
12	Hypertensive heart disease	7125	12	Falls	10222	(10090 - 10373)	46%
13	Falls	6979	13	Chronic kidney disease	10209	(10097 - 10331)	24%
14	Self-harm	6698	14	Self-harm	8949	(8827 - 9058)	34%
15	Parkinson's disease	6606	15	Hypertensive heart disease	8797	(8651 - 8928)	23%
16	Stomach cancer	5310	16	Atrial fibrillation and flutter	7785	(7722 - 7851)	6%
17	Non-rheumatic valvular heart disease	4973	17	Stomach cancer	6892	(6766 - 6994)	30%
18	Liver cancer	4605	18	Parkinson's disease	6678	(6656 - 6702)	1%
19	Leukemia	4346	19	Bladder cancer	6412	(6298 - 6544)	59%
20	Alcohol use disorders	4332	20	Alcohol use disorders	6048	(5965 - 6143)	40%
22	Bladder cancer	4022	21	Non-rheumatic valvular heart disease	5699	(5614 - 5808)	15%
23	Diabetes mellitus	3679	22	Liver cancer	5393	(5298 - 5476)	17%
57	Lower respiratory infections	890	25	Leukemia	5043	(4969 - 5115)	16%

Source: causes of death statistics, Germany, 2017, own calculation

Discussion

In this paper the method of redistributing IDD to valid CoD, as carried out in the German BoD study BURDEN 2020, has been illustrated. Whereas some parts of the method like identification of IDD and target codes were adopted from the GBD study, the redistribution and the approach to calculate uncertainties were in large parts own developments. Our aim was to depict how CoD statistics can be made suitable for BoD estimations and beyond. Since 26% of all deaths in Germany in 2017 are defined as IDD, we observe considerable changes in cases numbers before and after redistribution. For the specific CoD stroke, diabetes, and respiratory infections the numbers more than double.

The method used for the redistribution is in part an adaptation of the GBD method, adjusted to the German data. The results show that there are some significant differences, when comparing the final CoD data. The GBD study (2017) was estimating about 15,000 more deaths in 2017 in Germany than the German national CoD statistics [16]. This difference is most likely due to the additional correction steps that IHME applies when assessing the data quality and population coverage. The difference in the total number of deaths is reflected also in the cause specific number of deaths. However, the proportions of the main groups of CoD are very much comparable between the GBD and the German BURDEN 2020 study.

The development and implementation of methodologies for redistributing IDD is also done in other countries [21, 22, 27, 28]. The methods applied mainly reflect the available country specific data. Accordingly, the handling of IDD in national BoD studies differs significantly. In Scotland the CoD statistics include information on multiple CoD and in some cases deaths can be linked to individual clinical records. Thus, it is possible to develop a more precise, country specific method of identifying and redistributing IDD [29]. Likewise, Australia has developed an own algorithm of redistributing IDD. It includes several methods such as data linkage for obtaining additional information, usage of multiple CoD statistics, and proportional redistribution [28]. Other countries that lack multiple CoD statistics are forced to rely on alternative methods. For instance, Brazil has performed further research based on information from different health service providers or verbal autopsy. In cases where the actual CoD was not possible to be defined a proportional redistribution method was applied [27]. In the Netherlands experts pursue a one-number-policy, they do not redistribute IDD but instead use the CoD statistics without adjustment [30].

The above described method of redistribution is applied on the federal state level. It must be considered that the subnational differences in mortality registration procedures influence the quality of the data and the regional amount of IDD [19, 31]. Thus, the uncertainty bands are of high importance additionally depicting the variation in quality of CoD registration between the regions.

Limitations and Strengths

The adopted method used the definition of IDD as developed by IHME as part of the GBD study. However, it must be considered that in many cases scientists and physicians may have different perspectives on which diseases should be defined as IDD. For instance, for some CoD there is no consensus whether the condition should be classified as underlying CoD or as a sequela of another disease, the second being an intermediate or secondary CoD. A critical example is septicemia, which is considered an IDD in the GBD study, with its own redistribution package. This assessment is controversially discussed, as some experts see septicemia, at least for a part of the reported deaths, as the underlying CoD [32].

Another limitation of the study is the lack of multiple CoD data. The redistribution methodology in Germany could be largely improved should this kind of data become available. Further research is underway to test possible redistribution methods using multiple CoD data for some regions in Germany. Related to this, a further limitation of the applied method is the assumption that the valid CoD present the *true* distribution of valid codes. To overcome this issue, in BURDEN 2020 uncertainty intervals supporting the interpretation of results are provided indicating the margin in which the actual death counts may vary.

Advantageously, the applied redistribution method is transparent and comprehensible. Another strength of the study is the high quality of the German mortality data, especially with regard to registering the correct number and the age and sex of the deceased. In Germany almost full coverage of all deaths can be assumed and hence no methods for correction of possible underreporting must be applied. For many other countries, where mortality data do not have the same quality, and consequently a lower coverage, the GBD study has developed methods for corrections [16]. Another strength of the study, is the redistribution of IDD on a subnational level. As shown before [19] the quality of the CoD statistics in Germany differs strongly between the federal states. Additionally, we generally expect and observe differing mortality patterns across the federal states [33, 34], e.g. due to differences in age structure and socio-economic status [35, 36].

Outlook

The method described here reflects the availability of data in Germany. It is the first comprehensive redistribution of IDD within the CoD statistics for Germany. Further methodological developments are possible. We have only analyzed data from one year (2017). Looking at trend data (3-year or 5-year period) might limit random variations in the CoD data. Other aspects of the improvement include the usage of multiple CoD data which will allow a better determination of the target codes and the redistribution proportions. Furthermore, the selection of the target codes needs a better documentation and possible a revision in the future. At the moment the selection of the target codes is based on current research and expert assessment, provided by IHME, with not always clear facts and description.

Performing a redistribution method on the CoD statistics is currently very important. Otherwise there would be an underreporting of certain CoD or large numbers of deaths coded to residual or unspecific codes. However, more efforts should be put into obtaining a better quality of death registries and hence CoD data. This encompasses defining the underlying CoD as well as providing information on the accompanying diseases [31]. The nationwide implementation of the Iris/MUSE software, which improves the electronic processing and correction of CoD data, is a step in that direction and will contribute to better registration of the underlying CoD [8, 9]. From a public health perspective these successive improvements are of large importance as CoD data are an important information base for the identification of needs, the prioritization of actions required, and the development of targeted interventions.

BURDEN 2020 study group

Alexander Rommel, Elena von der Lippe, Annelene Wengler, Michael Porst, Aline Anton, Janko Leddin, Thomas Ziese (Robert Koch Institute), Helmut Schröder, Kathrin Schüssel, Gabriela Brückner, Jan Breitzkreuz (WIdO – AOK Research Institute), Dietrich Plass, Heike Gruhl (German Environment Agency)

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

The datasets generated during the current study are not publicly available due to confidential information from IHME (Institute of Health Metrics and Evaluation). Data are however available from the authors upon reasonable request and with permission of IHME. The basis of our analysis are the German causes of death data. Without regional assignments and aggregated they are available as an excel file through the German Federal Statistical Office

(https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Todesursachen/_inhalt.html Excel: *Ergebnisse der Todesursachenstatistik für Deutschland - Ausführliche vierstellige ICD10-Klassifikation – 2017.xlsx*).

Competing interests:

The authors declare that they have no competing interests.

Funding:

The study 'BURDEN 2020 – The burden of disease in Germany at the national and regional level' is funded by the Innovation Fund at the Federal Joint Committee (funding code: 01VSF17007).

Authors' contributions:

AW performed the statistical analyzes and wrote relevant parts of the paper; HG prepared most of the graphs and was a major contributor in writing the manuscript; DP also wrote relevant parts of the paper; JL supported the analyzes with technical knowlegde and contributed to the structuring of the paper; AR largely contributed to introduction and discussion of the results; EvdL: accompanied the whole process of redistribution and wrote relevant parts of the paper

All authors were substantially involved in developing the method of redistribution. They all read and approved the final manuscript.

Acknowledgements:

We thank Ronny Kuhnert for statistical advice with regard to transforming random numbers to new ICD-codes.

References

1. Murray CJL, Lopez AD: **Measuring global health: motivation and evolution of the Global Burden of Disease Study.** *Lancet (London, England)* 2017, **390**(10100):1460-1464.
2. Murray CJ, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C, Naghavi M, Salomon JA, Shibuya K, Vos T *et al.* **GBD 2010: design, definitions, and metrics.** *Lancet (London, England)* 2012, **380**(9859):2063-2066.
3. McKee S, Leach-Kemon K, Redford S. **What we measure and why.** <http://www.healthdata.org/acting-data/what-we-measure-and-why>
4. Zack F, Kaden A, Riepenhausen S, Rentsch D, Kegler R, Büttner A: **Fehler bei der Ausstellung der Todesbescheinigung.** *Rechtsmedizin* 2017, **27**(6):516-527.
5. **Todesursachenstatistik. Todesursachen in Deutschland** [https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Gesundheit/todesursachen.pdf?__blob=publicationFile&v=3]
6. **Iris-Institut** [<https://www.dimdi.de/dynamic/de/klassifikationen/icd/iris-institut/>]
7. Eckert O: **Elektronische Kodierung von Todesbescheinigungen.** *Bundesgesundheitsblatt* 2019, **62**:1468-1475.
8. Eckert O: **Verbesserte Qualität der nationalen und internationalen Todesursachenstatistik durch den Kodierkern MUSE WISTA,** *Statistisches Bundesamt* 2017, **4**:118-132.
9. Eckert O, Vogel U: **Todesursachenstatistik und ICD, quo vadis?** *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2018, **61**(7):796-805.
10. Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R: **Algorithms for enhancing public health utility of national causes-of-death data.** *Population Health Metrics* 2010, **8**(1):9.
11. **The concept of underlying cause of death** [<https://crvsgateway.info/The-concept-of-underlying-cause-of-death~616>]

12. WHO (World Health Organization) (ed.): **International statistical classification of diseases and related health problems, 10th revision, Volume 2, Instruction manual, Fifth edition**; 2016.
13. Gleich S, Weber S, Kuhn J: **Das Unbehagen mit der Leichenschau und der Todesursachenstatistik – A never-ending story?** *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2019, **62**(12):1413-1414.
14. Stolpe S: **Regional differences in the use of ill-defined causes of death in cardiovascular mortality.** *European journal of public health* 2017, **27**(suppl_3).
15. Mikkelsen L, Iburg KM, Adair T, Fürst T, Hegnauer M, von der Lippe E, Moran L, Nomura S, Sakamoto H, Shibuya K *et al*: **Assessing the quality of cause of death data in six high-income countries: Australia, Canada, Denmark, Germany, Japan and Switzerland.** *International Journal of Public Health* 2020, **65**(1):17-28.
16. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A *et al*: **Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017.** *The Lancet* 2018, **392**(10159):1736-1788.
17. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A *et al*: **Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.** *The Lancet* 2018, **392**(10159):1859-1922.
18. WHO: **WHO methods and data sources for country-level causes of death 2000-2015.** In. Edited by Department of Information EaR. Genf 2017.
19. Wengler A, Rommel A, Plaß D, Gruhl H, Leddin J, Porst M, Anton A, von der Lippe E: **ICD-Codierung von Todesursachen: Herausforderungen bei der Berechnung der Krankheitslast in Deutschland.** *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2019, **62**(12):1485-1492.
20. National Institute for Public Health and the Environment (RIVM), Ministry of Health Welfare and Sport: **The Public Health Foresight Report, The Netherlands**; 2018.
21. ScotPHo: **The Scottish Burden of Disease Study, 2015 Overview report.** Edinburgh: NHS Scotland; 2017.
22. **Belgian National Burden of Disease Study (BeBOD)**, [<https://www.sciensano.be/en/projects/belgian-national-burden-disease-study>]
23. Rommel A, von der Lippe E, Plaß D, Wengler A, Anton A, Schmidt C, Schüssel K, Brückner G, Schröder H, Porst M *et al*: **BURDEN 2020—Burden of disease in Germany at the national and regional level.** *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2018, **61**(9):1159-1166.
24. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A *et al*: **Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017.** *The Lancet* 2018, **392**(10159):1736-1788.
25. Wengler A, Rommel A, Plaß D, Gruhl H, Leddin J, Ziese T, Lippe E, Group ftBS: **Durch Tod verlorene Lebensjahre (YLL) in Deutschland. Eine umfassende Analyse des Sterbegeschehens im Projekt BURDEN 2020.** *Deutsches Ärzteblatt International* submitted 2020.
26. **BURDEN 2020: Potenzial und Nutzen** [https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/Krankheitslast/Potenzial/burden_potenzial_node.html]
27. França E, Teixeira R, Ishitani L, Duncan BB, Cortez-Escalante JJ, Morais Neto OLd, Szwarcwald CL: **Ill-defined causes of death in Brazil: a redistribution method based on the investigation of such causes.** *Revista de Saúde Pública* 2014, **48**:671-681.
28. AIHW: **Australian Burden of Disease Study: impact and causes of illness and death in Australia 2015.** *Australian Burden of Disease series no 19* 2019, **Cat. no. BOD 22**(Canberra).
29. Ian Grant, Oscar Mesalles-Naranjo, Grant Wyper, Jade Kavanagh, Elaine Tod, Colin Fischbacher: **Burden of Disease in Scotland. Invited Chapter, "The burden of disease is generally greater in more deprived areas, and that burden is more**

likely to be fatal.”. 2018.

30. Hilderink HBM, Plasmans MHD, Poos MJJC, Eysink PED, Gijzen R: **Dutch DALYs, current and future burden of disease in the Netherlands.** *Archives of Public Health* 2020, **78**(1):85.
31. Schneider NF, Mueller U, Klüsener S: **Sterblichkeit in Deutschland muss besser und schneller erfasst werden - Lehren aus der COVID-19-Pandemie.** *RatSWD Working Paper* 2020, No. 272, Rat für Sozial- und Wirtschaftsdaten.
32. Pechholdová M: **Sepsis-related mortality in the Czech Republic: multiple causes of death analysis.** *Epidemiol Mikrobiol Imunol* 2017, **66**(2):73-79.
33. Rau R, Schmettmann CP: **District-Level Life Expectancy in Germany.** *Dtsch Arztebl International* 2020, **117**(29-30):493-499.
34. Doblhammer G, Kreft D, Dethloff A: **Gewonnene Lebensjahre.** *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2012, **55**(4):448-458.
35. Kroll LE, Schumann M, Hoebel J, Lampert T: **Regional health differences – developing a socioeconomic deprivation index for Germany.** In.: Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung; 2017.
36. Lampert T, Hoebel J, Kroll LE: **Social differences in mortality and life expectancy in Germany. Current situation and trends.** *Journal of Health Monitoring* 2019, **4**(1):3-14.

Figures

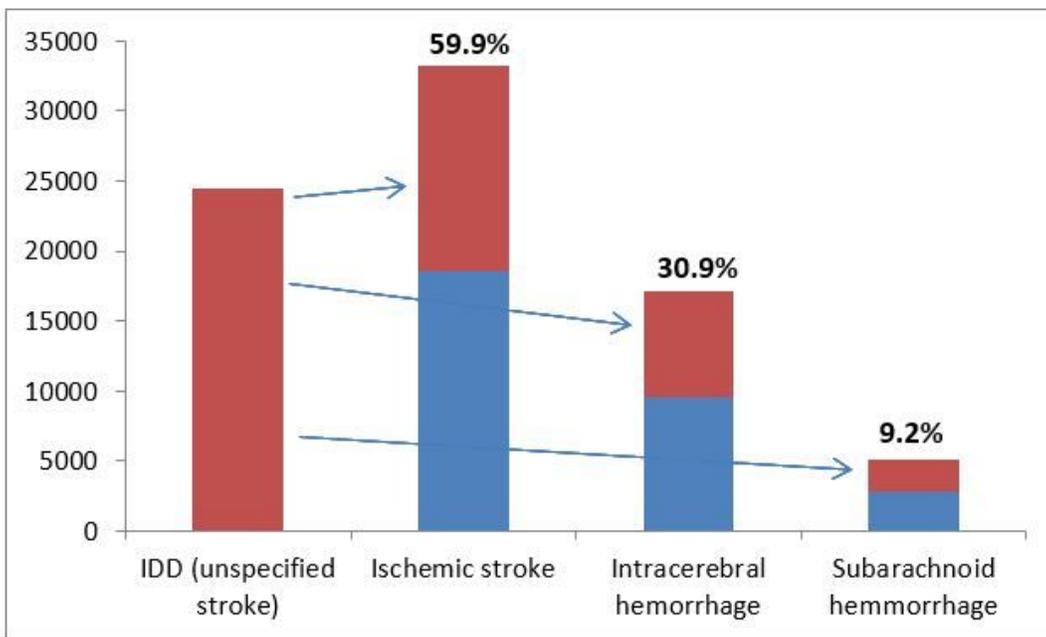


Figure 1

Example for the proportional redistribution of unspecified stroke Source: own depiction

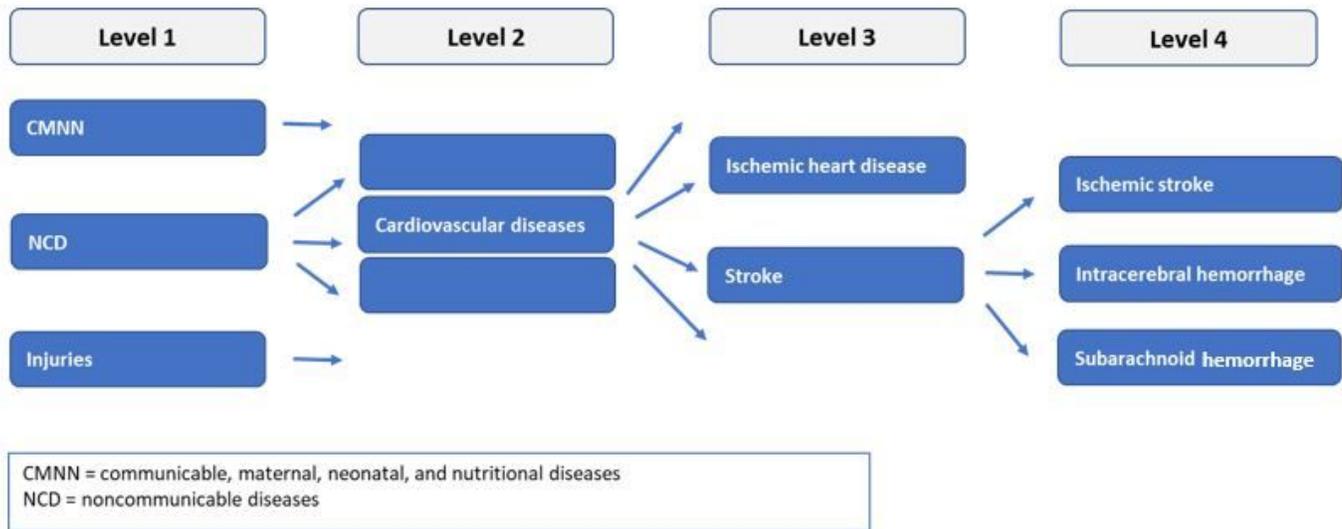


Figure 2

Causes of death hierarchy of the GBD study Source: [16, Appendix]

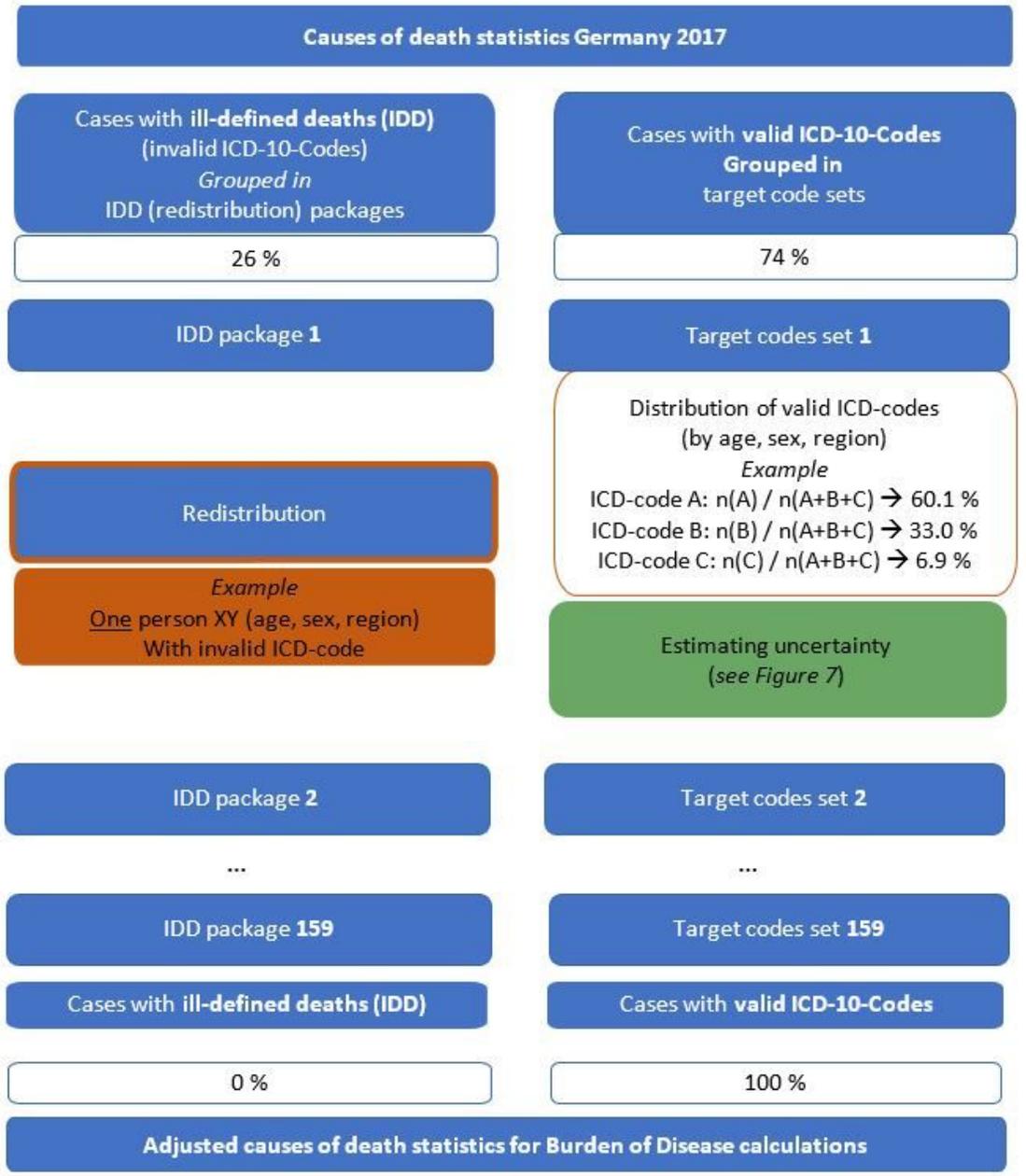


Figure 3

Consecutive process of redistributing IDD to valid ICD-codes Source: own depiction of redistribution of IDD to valid ICD-codes in the project BURDEN 2020 IDD: ill-defined deaths; n: number of deaths

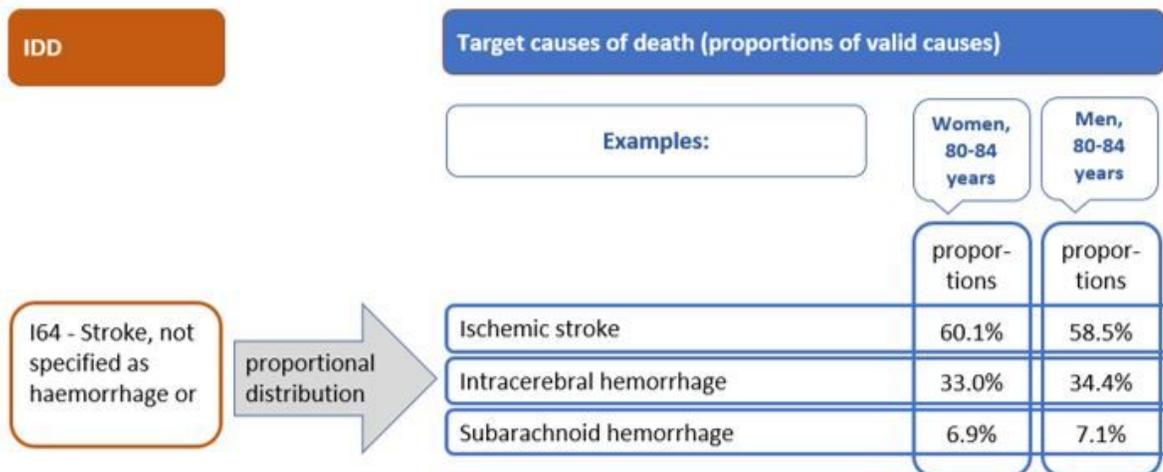


Figure 4

Example for the redistribution of unspecified stroke IDD in women and men aged 80 to 84 Source: own depiction

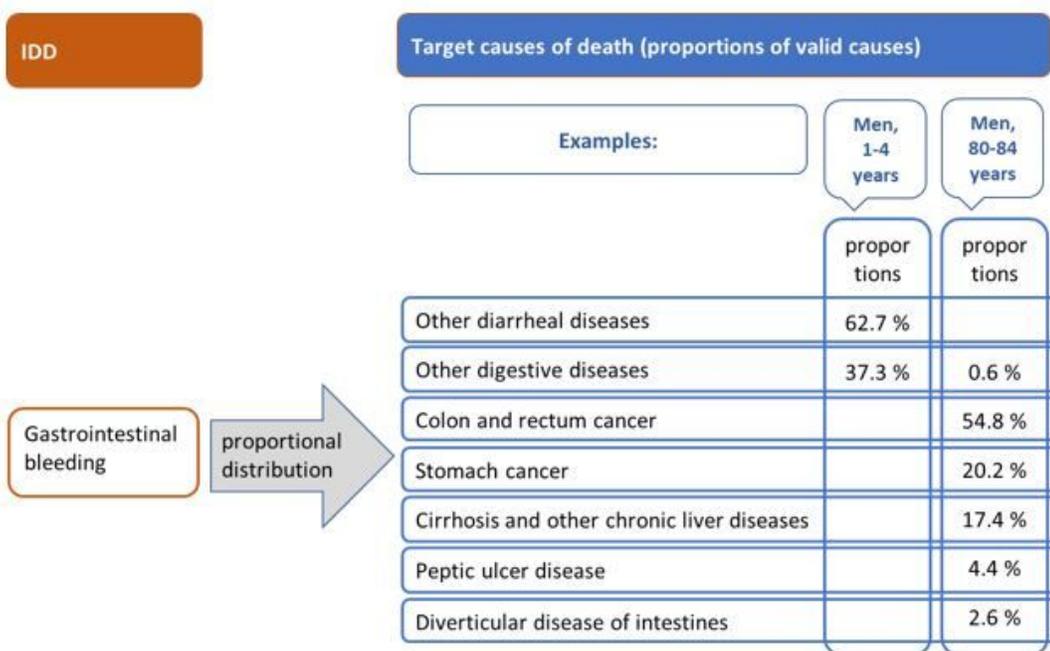


Figure 5

Example for the age specific redistribution of gastrointestinal bleeding Source: own depiction

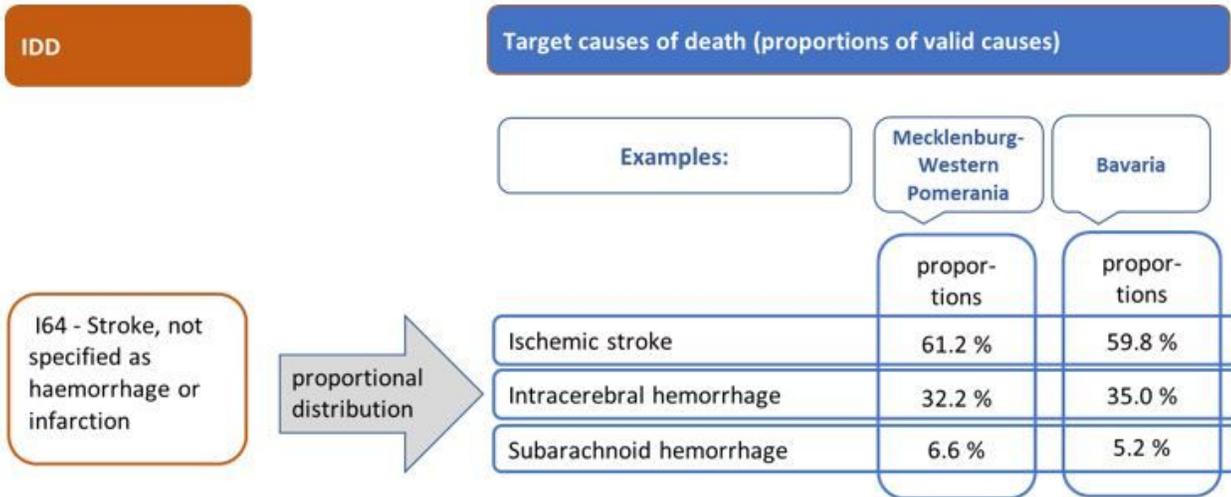


Figure 6

Example for the subnational redistribution of IDD in the federal states of Mecklenburg-Western Pomerania and Bavaria

Source: own depiction

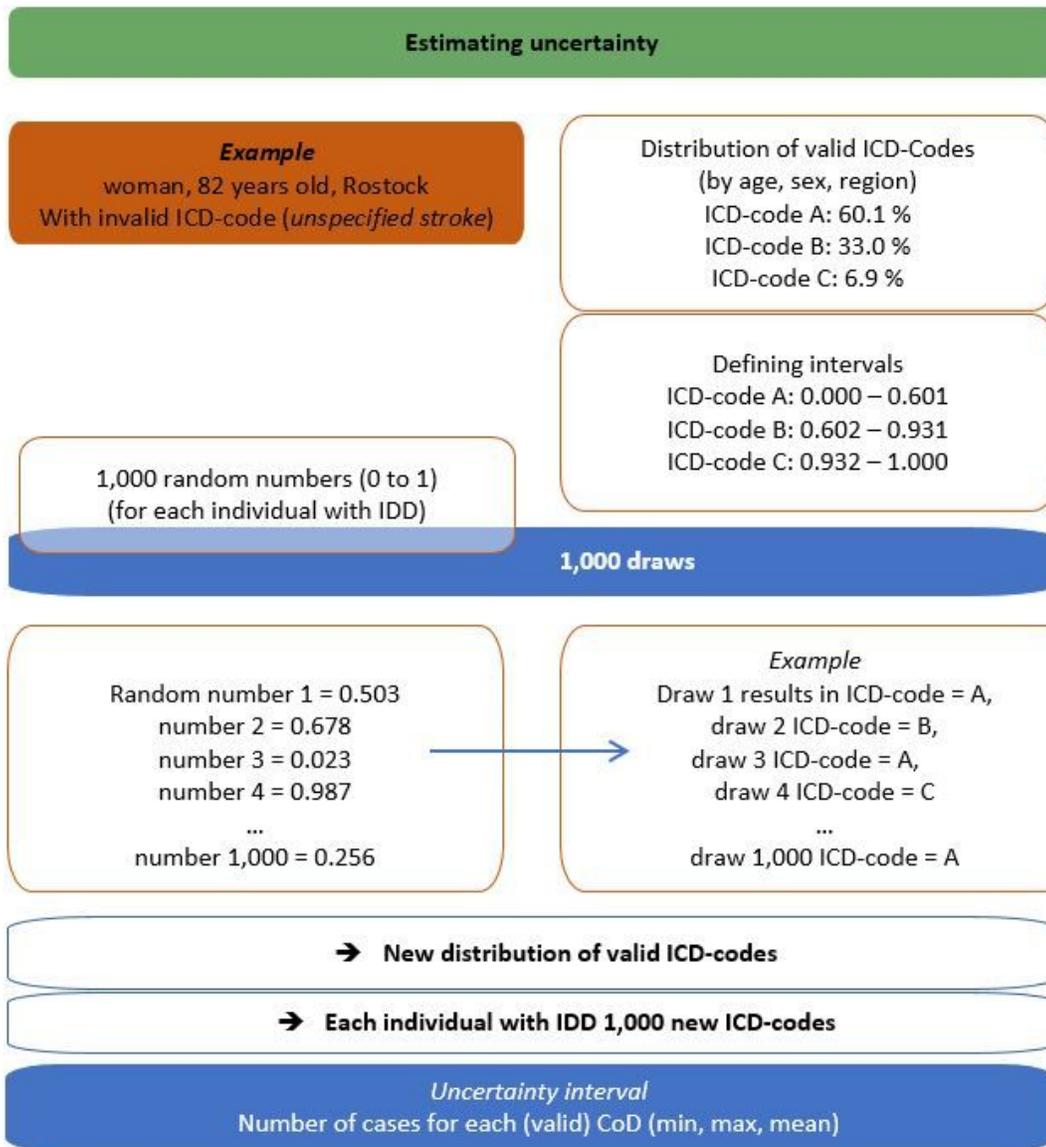


Figure 7

Example for estimating uncertainty when redistributing IDD to valid ICD-codes Source: own depiction of estimating uncertainty in the project BURDEN 2020 IDD: ill-defined deaths

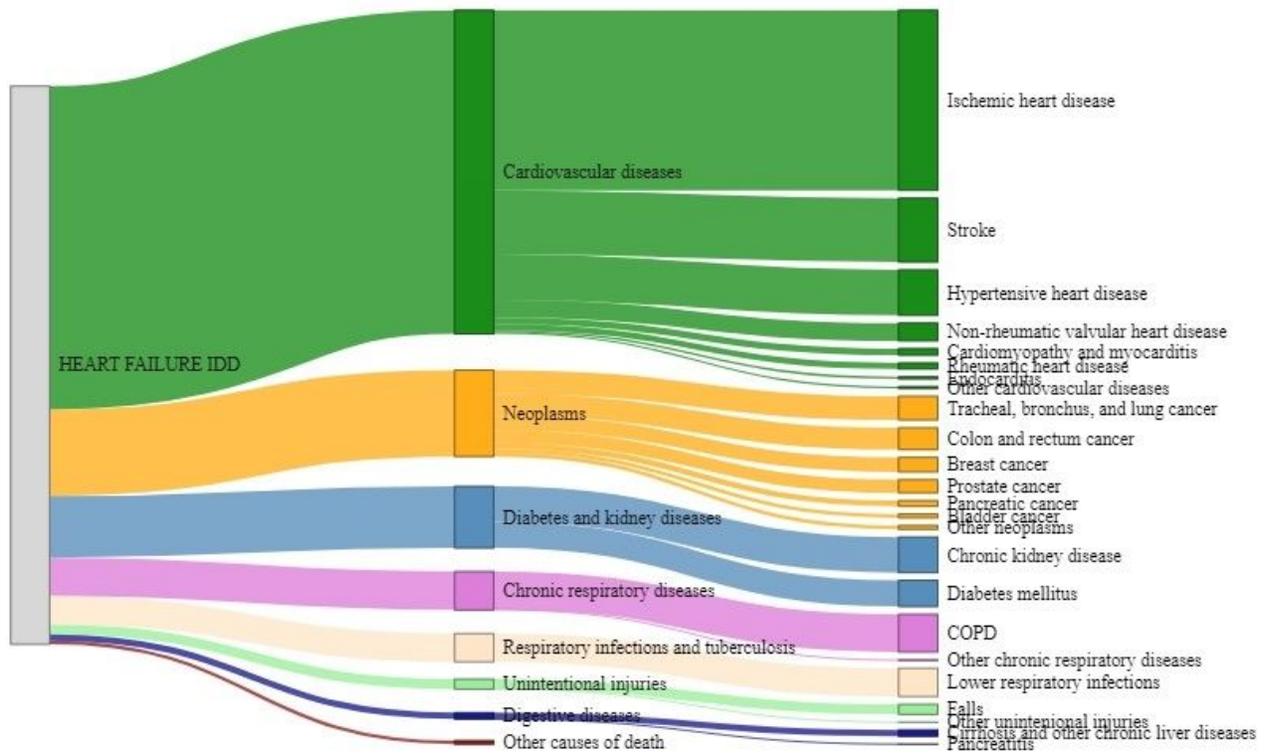


Figure 8

Redistribution of cases classified heart failure IDD to causes (level 2 and 3) Source: causes of death statistics, Germany, 2017, own calculations

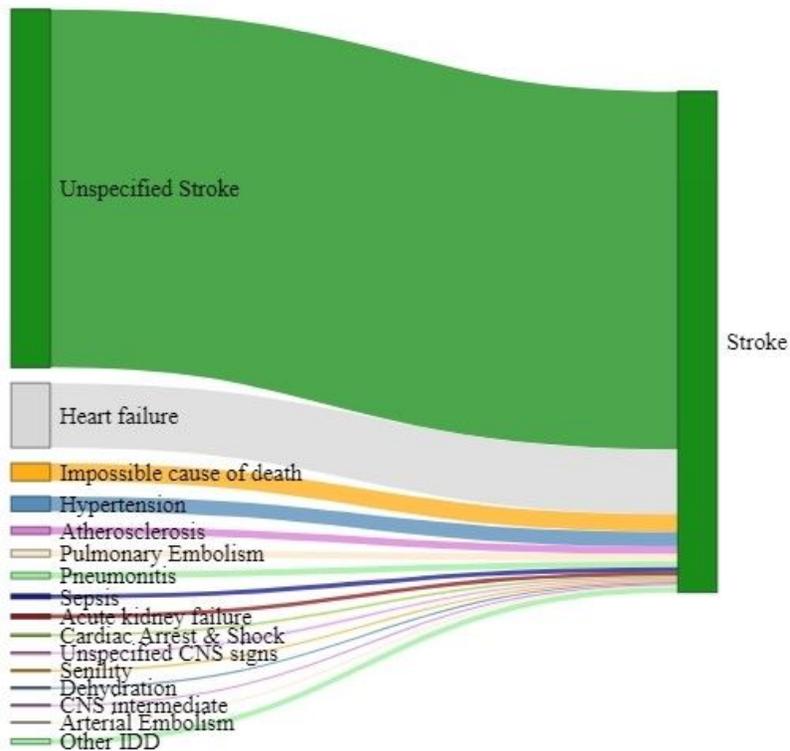


Figure 9

Redistribution of cases classified as IDD to stroke (level 3) Source: causes of death statistics, Germany, 2017, own calculations
IDD: ill-defined deaths, CNS: central nervous system

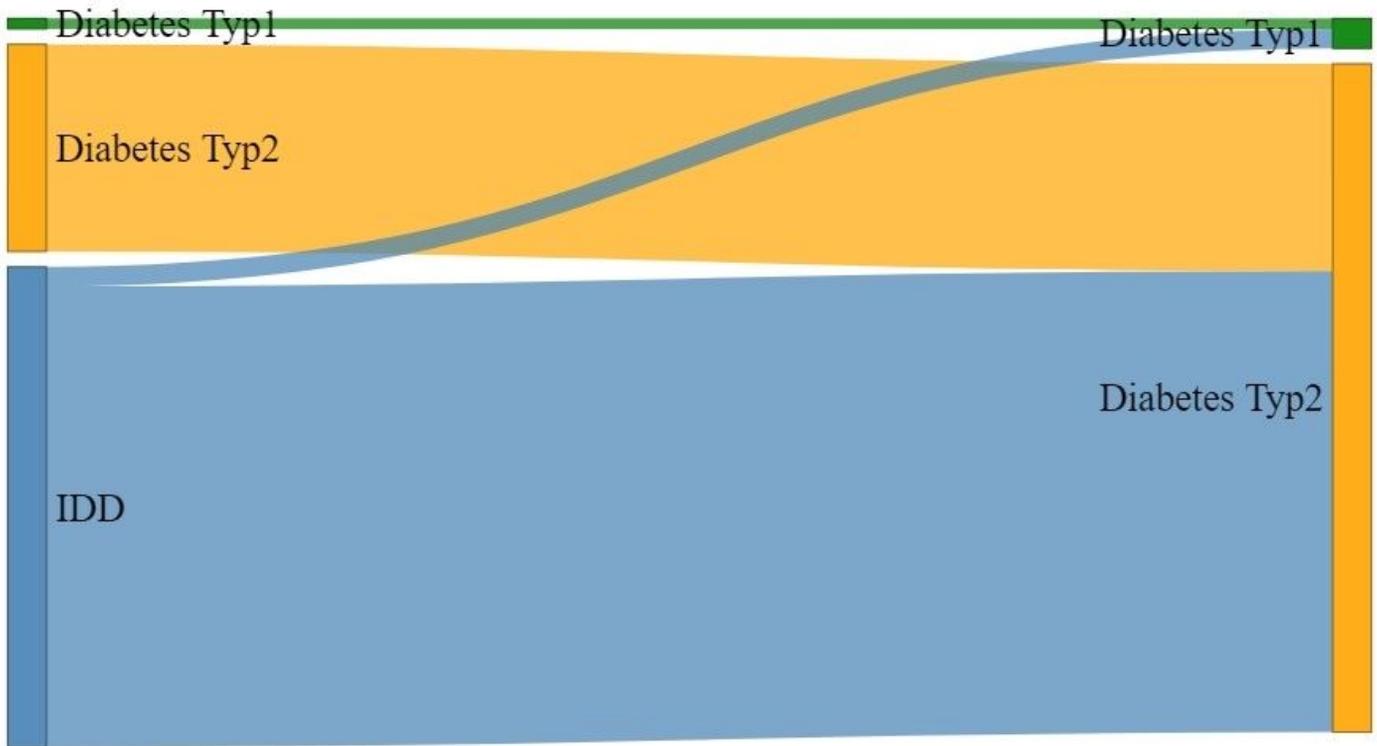


Figure 10

Redistribution of cases classified as IDD to Diabetes mellitus Type 1 and 2 on level 4 Source: causes of death statistics, Germany, 2017, own calculations

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [APPENDIX.docx](#)