

# Format of the data files in the Human Cause-of-Death Database

## Contents

General information .....	1
Data files .....	2
Death counts data files .....	2
Files with age-specific death rates .....	4
Files with CDR and SDR.....	4
Population exposures .....	4
Births data files .....	5
Metadata files .....	5
Lists of causes of death .....	5
Background and Documentation.....	6
References.....	6
Annex 1. ISO 3166-1 alpha-3 codes .....	7
Annex 2. Age group format .....	8
Annex 3. Shortlist .....	9
Annex 4. Intermediate list .....	10

## General information

Every country page in the Human Cause-of-Death Database (HCD) provides access to the following types of data: death counts, age-specific death rates, crude death rates (CDR), standardized death rates (SDR), population exposures, birth counts, and documentation. Death counts and death rates are provided by three cause-of-death lists (full, intermediate, and short) with redistributed ill-defined causes of death and death at unknown age. The intermediate and the short lists are standardized for all countries, and are provided as Annex 2 and 3, while the full list may differ.

Data classified according to the full list are missing for countries with short time series covering only the period since the implementation of ICD 10 in the country.

The complete set of input data for a country includes the following files:

Death counts:

- **XXX\_d\_full\_idr.csv** – death counts by causes of death according to the full list;
- **XXX\_d\_interm\_idr.csv** – death counts by causes of death according to the intermediate list;
- **XXX\_d\_short\_idr.csv** – death counts by causes of death according to the short list;
- **XXX\_d\_idinterm.csv** – ill-defined death counts according to the intermediate list;
- **XXX\_d\_idshort.csv** – ill-defined death counts according to the short list;

Age-specific death rates:

- **XXX\_m\_full\_idr.csv** – death rates by causes of death according to the full list;
- **XXX\_m\_interm\_idr.csv** – death rates by causes of death according to the intermediate list;
- **XXX\_m\_short\_idr.csv** – death rates by causes of death according to the short list;

Crude death rates:

- **XXX\_cdr\_full\_idr.csv** – crude death rates by causes of death according to the full list;
- **XXX\_cdr\_interm\_idr.csv** – crude death rates by causes of death according to the intermediate list;
- **XXX\_cdr\_short\_idr.csv** – crude death rates by causes of death according to the short list;

Death counts:

- **XXX\_sdr\_full\_idr.csv** – standardized death rates by causes of death according to the full list;
- **XXX\_sdr\_interm\_idr.csv** – standardized death rates by causes of death according to the intermediate list;
- **XXX\_sdr\_short\_idr.csv** – standardized death rates by causes of death according to the short list;

Population exposures:

- **XXX\_e.csv** – population exposures;

Birth counts:

- **XXX\_b.csv** – birth counts;

Metadata files:

- **XXX\_list.pdf** – specification of lists of causes of deaths;
- **XXX\_bd.pdf** – background and documentation;
- **XXX\_ref.pdf** – reference file.

Here XXX denotes the country code (3-character ISO code with no spaces, such as FRA, USA, etc., see Annex table 1). All data files are comma separated files with the extension “.csv”. A dot is used as decimal separator. Each record is unique; one line represents unique combination of country, year, sex, and cause, which cannot be repeated in the same file. This particularly means that alternative versions of data for one and the same country-year-sex-cause combination are not allowed. Only in the full list are zero death items not displayed. Missing values are coded as dots (“.”).

## Data files

### Death counts data files

The death counts are provided classified by causes of death according to the full, intermediate, or short lists. The intermediate and the short lists are standardized for all countries, and are provided as Annex 2 and 3.

The full list is country specific; it may be ICD10 4-digit, 3-digit or another country-specific abridged list. The full list may vary by year. While the zero death items are missing in data files, the full list in metadata file, the file XXX\_list.pdf, contains all the codes that have been reconstructed.

None of the three lists include ill-defined causes of death. Deaths counts originally classified as ill-defined are re-distributed. In addition we provide numbers of ill-defined death by age group as they were available before the re-distribution for short and intermediate lists. The list of ill-defined causes is provided in a separate file.

The structure of all *death counts data files* is identical. The first line is the header with variable names. Each file contains 32 variables. The first six columns refer to identifying variables: *country, year, sex, list, agf, cause*:

ID	Variable name	Description
1	<i>country</i>	3 character country identifier
2	<i>year</i>	Year of occurrence
3	<i>sex</i>	1-males, 2-females, 3-both sexes
4	<i>list</i>	List identifier (see table LISTS for full lists, otherwise <i>short</i> or <i>intermediate</i> depending on data )
5	<i>agf</i>	Age group format (see Annex 2)
6	<i>cause</i>	Cause identifier, depending on the list

Columns 7-32 refer to age-specific death counts. Column 7 contains deaths from all ages combined. Columns 8-32 contain deaths according to the given age group format, with up to 2 decimal places (if numbers are non-integer). Deaths at unknown ages must be redistributed (there is no column for deaths at unknown age). The variable names are as follows:

ID	Variable name	Age group
7	total	all ages combined
8	d0	0
9	d1	1-4
10	d5	5-9
11	d10	10-14
12	d15	15-19
13	d20	20-24
14	d25	25-29
15	d30	30-34
16	d35	35-39
17	d40	40-44
18	d45	45-49
19	d50	50-54
20	d55	55-59
21	d60	60-64
22	d65	65-69
23	d70	70-74
24	d75	75-79

25	d80	80-84
26	<b>d85p</b>	85+
27	d85	85-89
28	<b>d90p</b>	90+
29	d90	90-94
30	<b>d95p</b>	95+
31	d95	95-99
32	<b>d100p</b>	100+

Where “p” stands for “plus” (open-ended age interval). If, for example, the last interval for the whole period is 85+, variables 27-32 are generated as missing (“.”)

### Files with age-specific death rates

Files with age-specific death rates have same structure as death files. Death rates are calculated per 1,000,000 and published as integer numbers.

### Files with CDR and SDR

The first six columns of data files with CDRs and SDRs are identical to first six columns of death files. The following columns present values of CDR or SDR, respectively. While the CDR file contains only one data variable (‘cdr’), the SDR file provides SDRs calculated using two versions of the European Standard population<sup>1</sup> for total population and aggregated age groups:

ID	Variable name	Description
7	<i>SDR1976</i>	SDR according to 1976 standard
8	<i>SDR2013</i>	SDR according to 2013 standard
9	<i>SDR1976_0_14</i>	SDR according to 1976 standard for age group 0-14
10	<i>SDR1976_15_39</i>	SDR according to 1976 standard for age group 15-39
11	<i>SDR1976_40_64</i>	SDR according to 1976 standard for age group 40-64
12	<i>SDR1976_65p</i>	SDR according to 1976 standard for age group 65+
13	<i>SDR2013_0_14</i>	SDR according to 2013 standard for age group 0-14
14	<i>SDR2013_15_39</i>	SDR according to 2013 standard for age group 15-39
15	<i>SDR2013_40_64</i>	SDR according to 2013 standard for age group 40-64
16	<i>SDR2013_65p</i>	SDR according to 2013 standard for age group 65+

### Population exposures

Exposures files contain 30 variables with the following identifiers:

ID	Variable name	Description
1	<i>country</i>	3 character country identifier
2	<i>year</i>	Year of observation
3	<i>sex</i>	1-males, 2-females, 3-both sexes
4	<i>agf</i>	Age group format

And the following data, with 2 decimal places:

<sup>1</sup> We use European Standard Population as of 1976 and as of 2013

ID	Variable name	Age group
5	total	all ages combined
6	e0	0
7	e1	1-4
8	e5	5-9
9	e10	10-14
10	e15	15-19
11	e20	20-24
12	e25	25-29
13	e30	30-34
14	e35	35-39
15	e40	40-44
16	e45	45-49
17	e50	50-54
18	e55	55-59
19	e60	60-64
20	e65	65-69
21	e70	70-74
22	e75	75-79
23	e80	80-84
24	<b>e85p</b>	85+
25	e85	85-89
26	<b>e90p</b>	90+
27	e90	90-94
28	<b>e95p</b>	95+
29	e95	95-99
30	<b>e100p</b>	100+

### Births data files

Birth data are provided with 3 identifying variables and one variable with the actual number of live births:

ID	Variable name	Description
1	<i>country</i>	3 character country identifier
2	<i>year</i>	Year of occurrence
3	<i>sex</i>	1-males, 2-females, 3-both sexes
4	<i>b</i>	number of yearly live births

### Metadata files

#### Lists of causes of death

These lists are provided as PDF files. These files contain all lists of causes of death used in data except standard short and intermediate lists. The first list is the most detailed list of causes of death used in production of reconstructed (continuous) data series.

In case of ICD10 4-digit lists, the codes which are terminal 3-digit codes (i.e. have no 4-digit detail), are displayed with an underscore at the end (so that all codes have an equal length of 4, e.g. A09\_ instead of A09).

If possible, both English and national text titles are provided for the full list.

In the data files, the lists have a string identifier in column *list*. The following identifiers are used for known lists:

List identifier	Description
icd10_3	ICD10 3 digit
icd10_4	ICD10 4 digit
short	CoDD shortlist
interm	CoDD intermediate list
belarus	Belarus
ukraine	Ukraine
baltic	Baltic list
russia	Russian list

If, due to data availability, some items of the intermediate list cannot be constructed, they are recoded as missing (“.”) in the data files. Respective information is should be given at the beginning of the *list* file.

### **Background and Documentation**

This file contains descriptive information about the data series.

### **References**

This file is used to provide references to data source used in the data series

## Annex 1. ISO 3166-1 alpha-3 codes

BLR	Belarus
CZE	Czech Republic
DEU	Germany
GDR	East Germany
FRG	West Germany
ENW	England and Wales
ESP	Spain
EST	Estonia
FRA	France
JPN	Japan
LTU	Lithuania
LVA	Latvia
MDA	Moldova
POL	Poland
ROU	Romania
RUS	Russian Federation
UKR	Ukraine
USA	United States

## Annex 2. Age group format

The age group format describes standard 5-year age schedules. Age schedules used in the database and respective values of age group format are given in the table below:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
0	0	0	0
1-4	1-4	1-4	1-4
5-9	5-9	5-9	5-9
10-14	10-14	10-14	10-14
15-19	15-19	15-19	15-19
20-24	20-24	20-24	20-24
25-29	25-29	25-29	25-29
30-34	30-34	30-34	30-34
35-39	35-39	35-39	35-39
40-44	40-44	40-44	40-44
45-49	45-49	45-49	45-49
50-54	50-54	50-54	50-54
55-59	55-59	55-59	55-59
60-64	60-64	60-64	60-64
65-69	65-69	65-69	65-69
70-74	70-74	70-74	70-74
75-79	75-79	75-79	75-79
80-84	80-84	80-84	80-84
<b>85+</b>	<b>85+</b>	<b>85+</b>	<b>85+</b>
-	85-89	85-89	85-89
-	<b>90+</b>	<b>90+</b>	<b>90+</b>
-	-	90-94	90-94
-	-	<b>95+</b>	<b>95+</b>
-	-	-	95-99
-	-	-	<b>100+</b>

The age group format may vary between countries and years, depending on the open-ended interval. To assure consistency across different age group formats, “lower” open-ended intervals are computed for age group formats 2-4. In age group formats 1-3, the missing values (“.”) are generated where data is not available.



### Annex 3. Shortlist

No.	Title	Category codes according to ICD10
0	All causes	A00-Y98
1	Certain infectious diseases	A00-B99
2	Neoplasms	C00-D48
3	Diseases of the blood and blood-forming organs	D50-D89
4	Endocrine, nutritional and metabolic diseases	E00-E90
5	Mental and behavioral disorders	F00-F99
6	Diseases of the nervous system and the sense organs	G00-G44, G47-H95
7	Heart diseases	I00-I52
8	Cerebrovascular diseases	G45, I60-I69
9	Other and unspecified disorders of the circulatory system	I70-I99
10	Acute respiratory diseases	J00-J22, U04
11	Other respiratory diseases	J30-J98
12	Diseases of the digestive system	K00-K93
13	Diseases of the skin and subcutaneous tissue, musculoskeletal system and connective tissue	L00-M99
14	Diseases of the genitourinary system and complications of pregnancy, childbirth and puerperium	N00-O99
15	Certain conditions originating in the perinatal period and congenital malformations/anomalies	P00-Q99, R95
16	External causes	V01-Y98

#### Annex 4. Intermediate list

No.	Title	Category codes according to ICD10
0	All causes	A00-Y98
1	Other specified intestinal infections	A00-A08
2	Diarrhoea and gastroenteritis of presumed infectious origin	A09
3	TBC	A15-A19, B90
4	Septicaemia	A40-A41
5	Other bacterial diseases	A20-A28, A30-A39 A42-A44, A46, A48-A49
6	HIV disease	B20-B24
7	Viral hepatitis	B15-B19
8	Other viral diseases	A80-A89, B00-B09, B25-B34
9	Other and unspecified infectious and parasitic diseases	A50-A75, A77-A79, A90-A99, B35-B60, B64-B89, B91, B92, B94-B97, B99
10	Malignant neoplasms of lip, oral cavity and pharynx	C00-C14
11	Malignant neoplasm of esophagus	C15
12	Malignant neoplasm of stomach	C16
13	Malignant neoplasms of colon	C18
14	Malignant neoplasm of rectum and anus	C19-C21
15	Malignant neoplasms of liver and intrahepatic bile ducts	C22
16	Malignant neoplasm of pancreas	C25
17	Other malignant neoplasm of digestive system	C17, C23-C24, C26
18	Malignant neoplasm of larynx	C32
19	Malignant neoplasms of trachea, bronchus and lung	C33-C34
20	Malignant neoplasm of skin	C43, C44
21	Malignant neoplasm of breast	C50
22	Malignant neoplasm of cervix uteri	C53
23	Malignant neoplasms of uterus	C54-C55
24	Malignant neoplasm of ovary	C56
25	Malignant neoplasm of prostate	C61
26	Malignant neoplasm of other genital organs	C51, C52, C57, C58, C60, C62, C63
27	Malignant neoplasm of bladder	C67
28	Malignant neoplasms of kidney and other urinary organ	C64-C66, C68
29	Malignant neoplasms of meninges, brain and other parts of central nervous system	C70-C72
30	Leukemia	C91-C95
31	Other malignant neoplasms of lymphoid, hematopoietic and related tissue	C81-C90, C96
32	Malignant neoplasms of independent (primary) multiple sites	C97
33	Other cancer	C30-C31, C37-C41, C45-C49, C69, C73-C80
34	In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior	D00-D48
35	Diabetes mellitus	E10-E14

36	Malnutrition	E40-E46
37	Other endocrinologic and metabolic diseases	E00-E07, E15-E16, E20-E35, E50-E68, E70-E90
38	Blood diseases	D50-D89
39	Dementia, vascular, senile or unspecified	F01, F03
40	Alcohol abuse	F10
41	Drug abuse	F11-F19
42	Other mental disorders	F04-F09, F20-F99
43	Systemic atrophies and demyelinating diseases of the central nervous system	G10-G12, G35-G37
44	Parkinson's disease and other extrapyramidal and movement disorders	G20-G25
45	Alzheimer's disease and other degenerative diseases of the nervous system	G30, G31
46	Epilepsy	G40-G41
47	Other diseases of nervous system	G00-G09, G43-G44, G47-G83, G90-G99, H00-H95
48	Rheumatic heart diseases	I00-I09
49	Essential hypertension	I10
50	Hypertensive disease (heart, kidney and secondary)	I11-I15
51	Acute myocardial infarction	I21-I23
52	Atherosclerotic cardiovascular and heart diseases	I25.0, I25.1
53	Other IHD	I20, I24, I25.2 to .9
54	Pulmonary heart diseases	I26-I28
55	Non rheumatic valve disorders	I34-I38
56	Cardiac arrest	I46
57	Heart failure	I50
58	Other heart diseases	I30-I33, I40-I45, I47- I49, I51
59	Intracranial haemorrhage	I60-I62
60	Cerebral infarction, occlusion, and stenosis	I63, I65, I66
61	Other cerebrovascular diseases	G45, I64, I67
62	Sequelae of cerebrovascular disease	I69
63	Diseases of arteries, arterioles and capillaries	I70-I78
64	Other circulatory diseases	I80-I99
65	Influenza	J09-J11
66	Pneumonia	J12-J18
67	Other acute respiratory infections	J00-J06, J20-J22, U04
68	Asthma	J45-J46
69	Other chronic obstructive pulmonary disease	J40-J44, J47
70	Pneumonitis due to solids and liquids	J69
71	Pneumoconioses and chemical effects	J60-J68, J70
72	Other respiratory diseases, principally affecting the interstitium	J80-J84
73	Other diseases of the respiratory system	J30-J39, J85-J98
74	Gastric and duodenal ulcer	K25-K28
75	Hernia	K40-K46
76	Enteritis, colitis and other intestinal diseases	K35-K38, K50-K63
77	Alcoholic cirrhosis of liver	K70

78	Other cirrhososes of liver	K74
79	Other diseases of liver	K71-K73, K75, K76
80	Cholelithiasis and other disorders of biliary tracts	K80-K83
81	Diseases of pancreas	K85-K86
82	Other digestive diseases	K00-K22, K29-K31, K65-K66, K90-K92
83	Diseases of skin and subcutaneous tissue	L00-L98
84	Diseases of the musculoskeletal system and connective tissue	M00-M99
85	Renal tubulo-interstitial diseases	N00- N15
86	Renal failure	N17-N19
87	Other diseases of urinary system	N20-N36, N39
88	Diseases of genital organs	N40-N99
89	Complications of pregnancy, childbirth, and puerperium	O00-O99
90	Certain conditions originating in the perinatal period	P00-P96
91	Congenital malformations, deformations, and chromosomal abnormalities	Q00-Q99
92	Sudden infant death syndrome (SIDS)	R95
93	Transport accidents	V01-V99
94	Accidental falls	W00-W19
95	Accidental drowning and submersion	W65-W74
96	Accidental exposure to smoke, fire and flames	X00-X09
97	Accidental poisoning by alcohol	X45
98	Accidental poisoning by other substance	X40-X44, X46-X49
99	Other accidental threats to breathing	W75-W84
100	Suicide and self-inflicted injury	X60-X84
101	Assault	X85-Y09, Y35, Y36
102	Event of undetermined intent	Y10-Y34
103	Complications of medical and surgical care	Y40-Y84
104	Other accidents and late effects of accidents (remainder)	W20-W64, W85-W99, X10-X39, X50-X59, Y85-Y91, Y95-Y98