

Chapter 8

Data Collection, Data Quality and the History of Cause-of-Death Classification

Vladimir Shkolnikov, France Meslé, and Jacques Vallin

Until 1996, when INED published its work on trends in causes of death in Russia (Meslé et al. 1996), there had been no overall study of cause-specific mortality for the Soviet Union as a whole or for any of its constituent republics. Yet at least since the 1920s, all the republics had had a modern system for registering causes of death, and the information gathered had been subject to routine statistical use at least since the 1950s. The first reason for the gap in the literature was of course that, before perestroika, these data were not published systematically and, from 1974, had even been kept secret. A second reason was probably that researchers were often questioning the data quality; however, no serious study has ever proved this. On the contrary, it seems to us that all these data offer a very rich resource for anyone attempting to track and understand cause-specific mortality trends in the countries of the former USSR – in our case, in Ukraine. Even so, a great deal of effort was required to trace, collect and computerize the various archived data deposits.

This chapter will start with a brief description of the registration system and a quick summary of the difficulties we encountered and the data collection methods we used. We shall then review the results of some studies that enabled us to assess the quality of the data.

V. Shkolnikov (✉)

Max Planck Institute for Demographic Research, Konrad Zuse Strasse 1,
18057 Rostock, Germany
e-mail: shkolnikov@demogr.mpg.de

F. Meslé • J. Vallin

Institut National d'Études Démographiques, Bd. Davout 133,
75980 Paris Cedex 20, France
e-mail: mesle@ined.fr; vallin@ined.fr

8.1 The Registration and Coding System

The history of cause-of-death statistics in Ukraine is a relatively short one. It was only in 1925 that a system for registering cause of death on a regular basis was instituted throughout Ukraine (and across the USSR). The first Soviet Classification of Causes of Death, drawn up in 1922 and officially adopted in 1924 (Sadvokasova 1955), was quite close to the 1920 International Classification of Diseases (ICD).

However, full registration of causes of death was still not achieved, because doctors alone were empowered to write death certificates (Bystrova 1965), and some places – notably rural areas¹ – had very few doctors. It was only in towns that the system functioned properly, covering less than half the population of Ukraine. The situation improved from 1958, when the Ministry of Health and the TsSU (Central Statistical Directorate of the Soviet Union) decided to authorize another category of medical worker, the *feldsher*,² to make out a death certificate when there was no doctor. In 1959, 15% of death certificates in rural areas were completed by a *feldsher*. The proportion then decreased steadily throughout the 1960s and 1970s, falling to 5% or 6% in the mid-1980s.

Since most of the data we processed were compiled under the Soviet regime, we shall start by describing the system of registration and coding in force during that era. We shall go on to show that the change in political regime has not as yet called this system into question.

8.1.1 The Soviet System of Registering and Coding Causes of Death

After a death, the deceased's relatives must obtain a medical certificate of cause of death from the responsible institution (hospital, *poliklinika* [district clinic] or *sudebno-meditsinskaia ekspertiza* [Forensic Medical Examiner's Office]) and take it to the district register office (ZAGS). They exchange it for a civil death certificate that serves as both a burial permit and a legal document for inheritance purposes. ZAGS then sends the medical death certificate to the regional office of statistics (for the *Oblast* [province] or similar administrative area³).

As in most countries, the medical death certificate distinguishes three levels of causes: the underlying (principal or primary) cause, the immediate cause and contributory (associated or secondary) causes. Statistical tables are based on the

¹ The archives contain no record of deaths by cause for rural areas before the early 1950s.

² Within the Soviet health system, someone with an intermediate qualification, between a midwife and a doctor.

³ Such as Autonomous Republic or Autonomous Territory.

Table 8.1 Cause-of-death Classifications in the USSR since the Second World War

Period	Title	Number of items	Age groups used
1955–1964	1952 classification (Soviet classification of causes of death, 3rd revision)	116	0, 1, 2, 3–4, 5–6, 7–13, 14–15, 16–17, 18–19, 20–24, 25–29, 30–39, 40–49, 50–59, 60–69, 70 and over
1965–1969	1965 classification (based on ICD-7)	210+13 ^a	0, 1, 2, 3, 4, 5–9, 10–14, ..., 80–84, 85 and over
1970–1980	1970 Classification (based on ICD-8)	185+10 ^a	0, 1, 2, 3, 4, 5–9, 10–14, ..., 80–84, 85 and over
1981–1987	1981 classification (based on ICD-9)	185+10 ^a	0, 1, 2, 3, 4, 5–9, 10–14, ..., 80–84, 85 and over
Since 1988 ^b	1981 classification, adapted for deaths from injury and poisoning (based on ICD-9)	175+10 ^a	0, 1, 2, 3, 4, 5–9, 10–14, ..., 80–84, 85 and over

^aFor classifying deaths from injury and poisoning according to nature of injury

^bThe 1988 revision was modified in the 1990s in order to identify previously non-existent causes (like AIDS) or causes that had not been recorded singly (such as hunger)

underlying cause. In Ukraine, regional offices of statistics are responsible for coding the cause of death, and each one produces an annual table ('Form No. 5') giving the number of deaths by sex, age group and cause. These regional tables are then sent to be processed and totalled up centrally for the whole of Ukraine (as was the case for the other Soviet republics) by Goskomstat of Ukraine; until 1991, they were then processed at the level of the whole Soviet Union by Goskomstat of the USSR. These tables are the main source of the mortality data used in our study.

The USSR never adopted the World Health Organization's *International Classification of Diseases, Injuries and Causes of Death*, but used its own classification, with a detailed list that, from 1965, included about 200 items. From the foundation of the Soviet state, seven different versions of this classification were used in succession, five of them after the Second World War. Table 8.1 gives an overview of the latter.

The Ministry of Health and Goskomstat have had joint responsibility for modifications⁴ to the certificate of cause of death and to the rules on declaration and coding, as well as for successive revisions of the Classification. Goskomstat entirely computerized its processing of mortality data in 1988; until then, cross-tabulations had been produced manually.

Four of these successive versions were in use between 1965 and the Independence of Ukraine: the 1965, 1970, 1981 and 1988 Revisions. A detailed description of these is to be found in Annex III on the Website (<http://www.demogr.mpg.de/books/drm/009> or <http://extras.springer.com/>).

⁴To our knowledge, new directives were introduced at least eight times – in 1954, 1964, 1966, 1980, 1984, 1986, 1989 and 1992.

Three particular features of the Soviet system for declaring and recording cause of death are worth emphasizing.

1. The Soviet Classification has always differed significantly from the ICD definitions, although they were brought somewhat closer together from 1965 onwards. In particular, the number of Soviet items is far smaller than the number of ICD items: 210 items in 1965–1969, as against over 2,000 in ICD-7; from 1970, 185 instead of almost 3,000 in ICD-8 and then over 5,000 in ICD-9 (WHO 1977, 1978). Goskomstat produced a special document with a table showing correspondences between the 1981 Soviet Classification and ICD-9 (Goskomstat 1981).
2. Up to 1988, some items in the Soviet Classification did not appear in the ordinary annual tables of statistics ('Form No. 5'): these causes (cholera, plague, suicide, homicide and occupational accidents) were concealed for political reasons. In order to ensure consistent "All causes" totals, the deaths attributed to these 'hidden' causes were added to the "Ill-defined causes of death". In fact, the hidden causes were also accounted for separately in a special, top-secret table ('Form No. 5b'), which we were finally able to consult for the years 1963–1982 and 1984–1987 – though not for 1983. However, we were able to subject that year to an indirect estimate from the data available for the whole USSR and from the special category reserved for deaths from injury and poisoning.⁵ There is no mystery as to why the Soviet authorities kept these causes secret: in 1970, for example, the standardized male mortality rate from homicide in the Soviet Union was almost eight times the European average. The authorities were concerned to keep such information, which was viewed as politically dangerous, out of reach of both inside and outside observers.
3. The coding system is decentralized. Therefore, despite all the instructions, newsletters and directives issued by the Ministry of Health and Goskomstat, there is

⁵ At the end of our preliminary investigations in Moscow, we still had a problem with these hidden causes: for 1983 and 1984, data were available only for the whole USSR, not by individual republic. However, it was still possible to find the number of 'hidden deaths from injury and poisoning' for each republic – and therefore, here, for Ukraine – by subtracting officially acknowledged deaths from injury and poisoning from all deaths from injury and poisoning. The ordinary tables classify deaths from injury and poisoning twice, once according to the external cause of death and once according to nature of injury. All deaths from injury and poisoning, whatever their cause, are classified according to the nature of the injury, and the relevant table clearly shows the total number of deaths from injury and poisoning – including those relating to hidden causes. After using this approach to find the total number of hidden deaths from injury and poisoning, we distributed them for each republic in proportion to the structure of external causes of death for the whole USSR. In the end, the only remaining hidden deaths for these 2 years were deaths from plague and cholera, classified under 'Ill-defined causes of death'; but by that period these two causes had become insignificant. In fact, we managed to find Ukraine's 1984 statistics for hidden causes in Ukraine itself. Therefore, 1983 is the only year for which indirect estimates are used here.

some risk of regional differences in coding practices. Cause-of-death coding by regional offices of statistics is not checked or corrected at the national level. In fact, analysis of the data by republic suggests that some instructions were applied at different dates or using different methods. However, we can reasonably hope for less heterogeneity within each republic than between republics.

8.1.2 *Little Change Since Independence*

As far as we have been able to find out, independence for Ukraine does not seem to have led to any significant amendments in the system for collecting information on causes of death, still in force from the Soviet period until 2004. In particular, the last Soviet classification of causes of death remained in use for 15 years ahead. Apart from the very specific changes brought into effect in the 1990s, already mentioned in Table 8.1, the Soviet detailed list of causes of death remained in force throughout the whole period covered by this study. Ukraine was late adopting the International Classification of Diseases, contrary to the Baltic States and Russia. It has been done in 2005 only and only on the basis of a simplified list of 258 groups of ICD-10 items, that we shall call “2005 Ukrainian classification”. Results of that change are too recent to give us the means to deal with its statistical effect definitely in the framework of that book, but we shall refer to data as they are, as far as possible to complete series used in the following chapters until 2006.

8.2 The Available Data

Very little data on causes of death was published in the Soviet period. For the 1960s and early 1970s, some overall figures on mortality from cancer and diseases of the circulatory system⁶ for the whole Soviet Union were published in the journal *Vestnik statistiki*, in the *Narodnoe Khoziaistvo SSSR* statistical annuals and in the 1973 demographic yearbook *Naselenie SSSR* (TSSU 1975).

The situation became even worse between 1974 and 1987, after the Soviet Government decided to ban any publication on mortality and causes of death, since unfavourable trends in this area had become a taboo subject.

It was only in 1987–1988 that perestroika and glasnost opened up a completely new era in statistics and that cause-of-death data were published systematically for the first time. But at first these gave only age-specific mortality rates for very broad groupings of causes (infectious diseases, neoplasms, diseases of the circulatory system, respiratory diseases, injury and poisoning). Since then, current

⁶Mortality rates by sex and age for 1966–1967, 1968–1969 and 1969–1970; for 1973, crude rates only.

data have become more easily accessible and more widely used, though no retrospective picture has been available before the publication of the French version of that book, in 2003.

In order to be able to analyse long-term cause-of-death series in sufficient detail, we mainly had to use the original handwritten returns produced by Goskomstat, which became accessible in 1988 when the archives were opened.

However, for the years 1971, 1976 and 1980–1990, we obtained a copy of the computerized files from the Data-Processing Centre of the Ministry of Health of the Russian Federation, giving numbers of deaths by sex, age, cause and republic. For the post-1990 years, we obtained paper copies or computer files (depending on the year in question), up to and including 2006.

For the other years (1959–1970, 1972–1975 and 1977–1979), we gained access to the original tables, kept in the Russian State Archive of the Economy.⁷ It was no small task to determine whether certain tables existed (notably for ‘hidden’ causes), to locate them, to photograph thousands of original manuscript sheets (bound in heavy registers that could not be taken away) and, finally, to key in all the data.

8.3 Data Quality⁸

All countries face the issue of validity and comparability in recording causes of death. These depend very much on the quality of diagnoses, on the system for registering and coding causes, on the training given in medical schools, and on practitioners’ habits and priorities.

During the Soviet period, three major surveys looked at the quality of registration of causes of death. They were all conducted in similar ways. A certain number of medical death certificates was collected in each region chosen. Next, experienced doctors checked the quality of diagnosis and coding, by comparing the underlying cause of death declared on the certificate both to the true diagnosis, which could be established from the medical file and the post-mortem report, and to the item under which it was finally coded. None of these surveys related directly to Ukraine, but they allow us to lift a corner of the veil over the quality of observations made in the Soviet Union, and therefore in Ukraine.

⁷ Russian State Archive of the Economy (RGAE), fonds 1562, series 27 (files 833, 1023, 1187, 1328, 1464, 2632, 2638, 2655–2658, 5873, 5874, 5881, 9742, 9743, 9752, 9753), series 33 (files 980, 1361, 1700, 6627, 6984, 7320, 7652, 7933), series 34°C (files 174, 356, 529, 701, 882), series 44 (files 2625, 2655–2658), series 45 (files 2368, 2369, 5873, 5874, 5881, 9742, 9743, 9752, 9753), series 46 (files 1587, 1588, 1595, 1596), series 47 (files 1430, 1431, 1438, 1439), series 48 (files 1289, 1290, 1299, 1300), series 49 (files 1859–1860, 1869–1871), series 50 (files 1758, 1759, 1768–1770), series 55 (files 1908, 1909), series 56 (files 1936, 1937, 1947, 1948).

⁸ This section covers essentially the same ground as a text about the quality of cause-of-death data in Russia, already published by INED (Meslé et al., 1996).

The first survey related to death certificates was completed in central Russia (Tula, Novomoskovsk, Tambov, Michurinsk) in the early 1960s (Bystrova 1965). The second one dealt with a sample of 1979 certificates from different regions of Russia (Bednyi et al. 1980, 1981). The last one involved certificates completed in Belorussia (now Belarus) and in Turkmenistan in 1981–1982 (Ovcharov and Bystrova 1982). The results of these surveys taken together might have provided some very interesting data. Unfortunately, the description of the material used is very cursory: not even the total number of deaths studied in each survey is shown, and there are no data on age or sex. In Bednyi's survey, there is no mention of which regions provided the certificates used.

Despite their serious inadequacies, these surveys gave us important indications. They offer two types of results (Tables 8.2 and 8.3).

Table 8.2 gives the percentage of errors in medical diagnoses and in coding. The total percentage of diagnostic errors varies from 6.6% (Minsk, 1981–1982) to 12.7% (Russia, 1979), while the percentage of coding errors varies from 4.1% (Minsk, 1981–1982) to 17.7% (Tula, etc., 1960). For most causes of death, there is a higher proportion of errors in coding than in diagnosis. For cancer (neoplasms) and for injury and poisoning, the number of errors is small; it is much higher for hypertensive, cerebrovascular and ischaemic heart diseases, respiratory diseases and digestive diseases.

Table 8.3 gives the final differences observed between real cause and registered cause after the two types of errors quantified in Table 8.2 have been combined. In many cases, each type of error cancels out the other.

The table shows some over-registration of deaths from cerebrovascular disease, coronary atherosclerosis, atherosclerotic cardiosclerosis and respiratory diseases and, in contrast, under-registration of deaths from cancer, hypertensive diseases and myocardial infarction. In fact, without knowing the absolute numbers, we cannot check whether these opposed trends offset each other within the given sample. In order to test the validity of the observations, we applied the set of coefficients from Table 8.3 to cause-specific death statistics for Russia (1960 and 1979), Belorussia (1981–1982) and Turkmenistan (1981–1982). There was little difference between the total numbers of 'All causes' deaths before and after correction; so we were able to place some confidence in the results of these surveys.

These surveys do not confirm the widely-held view that mortality from diseases of the circulatory system was overestimated in the USSR. There are indeed fairly large errors for the various diseases that make up this group, but they offset one other, so that in the end the total percentage error is modest. Furthermore, only the first study indicates slight over-registration of mortality from diseases of the circulatory system; the other two show under-registration of 2–3%.

These findings contradict the point of view – defended by Iuri Belenkov et al. (1987), among others – that there was substantial over-registration of deaths from causes related to the circulatory system, especially among old people. The data available from these surveys did not allow a more precise analysis of differences by age. Given the overall results (Table 8.3), if circulatory diseases in the older age groups were really over-estimated, there would be an accompanying under-estimate

Table 8.2 Percentage of errors in diagnosing and coding causes of death, according to three Soviet surveys. Table taken from Meslé et al., 1996

Cause of death	(1)		(2)		(3)		(3)	
	Diagnosis	Coding	Diagnosis	Coding	Diagnosis	Coding	Diagnosis	Coding
<i>Infectious diseases</i>								
Tuberculosis	1.8	—	10.5	14.0	23.2	5.6	16.3	14.3
<i>Neoplasms</i>	3.8	6.1	4.6	0.8	3.9	3.0	4.9	18.1
Digestive organs	4.0	3.9	—	—	—	—	—	—
Respiratory organs	3.0	2.2	—	—	—	—	—	—
Female genital organs	0.0	5.2	—	—	—	—	—	—
Breast		0.0	3.9	—	—	—	—	—
Leukaemia	4.1	7.6	—	—	—	—	—	—
<i>Circulatory diseases</i>	14.7	25.1	17.0	26.4	3.3	7.6	20.5	8.8
Rheumatism	7.5	10.6	32.4	1.4	—	—	—	—
Hypertension	11.6	25.7	5.4	40.0	—	—	—	—
Ischaemic diseases	—	—	24.9	21.4	—	—	—	—
Cerebrovascular diseases	16.5	30.6	6.6	18.3	—	—	—	—
<i>Respiratory diseases</i>	13.0	22.8	24.5	24.6	11.8	7.8	10.2	15.8
<i>Digestive diseases</i>	12.8	21.1	29.7	5.9	12.8	3.7	22.1	13.0
<i>Genitourinary diseases</i>	14.6	14.6	—	—	7.3	11.8	8.0	37.5
<i>Congenital anomalies</i>	12.5	5.0	—	—	2.2	7.7	16.7	40.0
<i>Injury and poisoning</i>	3.3	8.9	0.0	3.7	1.4	3.1	4.9	23.3
<i>Total</i>	10.8	17.7	12.7	17.2	6.6	4.1	9.0	10.1

(1) Bystrova (1965)

(2) Bednyi et al. (1980, 1981)

(3) Ovcharov and Bystrova (1982)

Table 8.3 Over-estimated or under-estimated causes of death, according to three Soviet surveys. Number of deaths classified in a given category per 100 deaths actually falling into that category. Table taken from Meslé et al., 1996

Cause of death	Survey			
	(1) Tula, Novomoskovsk Tambov, Michurinsk	(2) Unspecified regions of Russia	(3) Minsk (Belorussia – now Belarus)	(3) Ashkhabad (now Ashgabat – Turkmenistan)
<i>Infectious diseases</i>	–	–	96.3	84.4
Tuberculosis	100.0	–	–	–
<i>Neoplasms</i>	96.3	–	95.5	96.4
Stomach	96.8	–	–	–
Oesophagus	100.0	–	–	–
Respiratory organs	99.2	–	–	–
Female genital organs	94.3	–	–	–
Breast	75.0	–	–	–
<i>Circulatory diseases</i>	102.3	97.2	98.8	96.9
Rheumatism	96.1	95.8	–	–
Hypertension	71.2	60.0	–	–
Ischaemic heart disease	–	96.0	–	–
Coronary atherosclerosis	106.6	103.8	–	–
Myocardial infarction	86.5	92.2	–	–
Cardiosclerosis	133.5	–	–	–
Cerebrovascular diseases	135.3	114.9	–	–
<i>Respiratory diseases</i>	111.3	–	117.2	111.7
<i>Digestive diseases</i>	88.5	–	95.4	114.7
<i>Genitourinary diseases</i>	100.0	–	101.2	88.0
<i>Congenital anomalies</i>	100.0	–	100.0	66.7
<i>Injury and poisoning</i>	99.5	–	98.6	94.2

(1) Bystrova (1965)

(2) Bednyi et al. (1980, 1981)

(3) Ovcharov and Bystrova (1982)

at other ages. In the absence of any other indications, we felt it prudent not to transfer any deaths from circulatory causes into the other major nosological categories. On the other hand, some redistribution of deaths between the different circulatory items might be helpful – for example, from the seemingly over-estimated cerebrovascular to the under-estimated hypertensive diseases. However, the unfavourable trends in mortality from circulatory diseases in Ukraine (like in Russia and other European republics of the former USSR) reflect a real deterioration much more than any increased over-estimation.

References

- Bednyi, M. S., Ivakina, V. N., Dmitriev, V. I., & Nikolskii, A. V. (1980). *Анализ медико-демографических показателей* [Analysis of medico-demographic indicators]. Moscow: Republican Scientific Research Laboratory for Medical Demography.
- Bednyi, M. S., Ivakina, V. N., Dmitriev, V. I., & Nikolskii, A. V. (1981). *Анализ степени точности диагностики заболеваний и причин смерти*. [Analysis of accuracy in diagnosing diseases and causes of death], *Информационное письмо Министерства здравоохранения РСФСР* [Newsletter of the Ministry of Health of the Russian Federation].
- Belenkov, I., Zhdanov, V., Pichugin, V., Marchenko, O., Yurova, I., Sidorenko, V. (1987). *Совершенствование диагностики причин смерти от основных болезней системы кровообращения*. [Improving diagnosis of causes of death from the main diseases of the circulatory system]. In E. I. Chazov (Ed.), *Болезни сердца. Медико-демографические исследования* [Heart disease. Medico-demographic research] (pp. 12–15). Moscow: Pan-Soviet Scientific Centre for Cardiology, Academy of Medical Sciences of the USSR.
- Bystrova, V. A. (1965). *Степень точности врачебной регистрации причин смерти*. [Accuracy in medical records of causes of death]. In V. A. Bystrova (Ed.), *Методологические вопросы медицинской и санитарной статистики* [Methodological issues in medical and health statistics] (pp. 60–80). Moscow: Nauka.
- Goskomstat (1981). *Краткая номенклатура причин смерти 1981года, основанная на международной статистической классификации болезней, травм и причин смерти IX пересмотра (1975 год)* [Abridged 1981 Classification of causes of death, based on the international classification of diseases, injuries and causes of death, ninth revision, 1975]. Moscow: Goskomstat of the USSR.
- Meslé, F., Shkolnikov, V. M., Hertrich, V., & Vallin, J. (1996). *Tendances récentes de la mortalité par cause en Russie 1965–1994*. Paris: INED and CDEH. 140 p. + 2 diskettes. (Bilingual French and Russian publication. Données statistiques, No. 2).
- Ovcharov, V. K., & Bystrova, V. A. (1982). *Медико-социальная характеристика воспроизводства населения СССР и союзных республик* [Medico-social characteristics of the reproduction of the population in the USSR and its Republics] (Scientific report no. 78032560). Moscow: Institute for Social Medicine and Health Care Organization.
- Sadvokasova, E. A. (1955). *Лекция X*. [Lecture X]. In E. A. Sadvokasova (Ed.), *Статистика здоровья населения* [Population Health Statistics] (pp. 83–93). Moscow: Medgiz.
- TSSU SSSR (Tsentral'noe Statisticheskoe Upravlenie SSSR) (1975). *Население СССР 1973. Численность, состав и движение* [Population of the USSR, 1973. Numbers, composition and change] (208 p). Moscow: Statistika.
- WHO (1977). *Manual of the international classification of diseases, injuries and causes of death, based on the recommendations of the Conference for the Ninth Revision, 1975 [...]* (Vol. 1, 781 p.). Geneva: World Health Organization.
- WHO (1978). *Manual of the international classification of diseases, injuries and causes of death, based on the recommendations of the Conference for the Ninth Revision, 1975 [...]* (Vol. 2, Alphabetical Index, 641 p.). Geneva: World Health Organization.