



# Comparison of Two Information Sources for Cause-of-Death Follow-up in the Russian Federation: The Asbest Chrysotile Cohort Study

J. Schüz<sup>1</sup> E. Kovalevskiy<sup>2,3</sup> M. Moissonnier<sup>1</sup> A. Olsson<sup>1</sup> D. Hashim<sup>1</sup> H. Kromhout<sup>4</sup> S. Kashanskiy<sup>5</sup>  
O. Chernov<sup>2</sup> I. Bukhtiyarov<sup>2,3</sup> E. Ostroumova<sup>1</sup>

<sup>1</sup>Section of Environment and Radiation, International Agency for Research on Cancer (IARC/WHO), Lyon, France

<sup>2</sup>Federal State Budgetary Scientific Institution “Izmerov Research Institute of Occupational Health,” Moscow, Russian Federation

<sup>3</sup>I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russian Federation

<sup>4</sup>Institute of Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

<sup>5</sup>Yekaterinburg Medical Research Center for Prophylaxis and Health Protection in Industrial Workers, Yekaterinburg, Russian Federation

**Address for correspondence** J. Schüz, PhD, Section of Environment and Radiation, International Agency for Research on Cancer (IARC/WHO), 150 cours Albert Thomas, 69372 Lyon, France (e-mail: schuzj@iarc.fr).

Methods Inf Med 2020;59:9–17.

## Abstract

**Background** The Asbest chrysotile cohort was set up in Asbest town, Sverdlovsk oblast, Russian Federation, among the current and former workforce of the world’s largest operating chrysotile mine and its processing mills, to investigate cancer risk in relation to occupational exposure to chrysotile.

**Objectives** The cohort of 35,837 people was followed-up for mortality using cause-of-death information from official death certificates issued by the Civil Act Registration Office (ZAGS) of Sverdlovsk oblast from 1976 to 2015. Data were also retrieved from the electronic cause-of-death registry of the Medical Information Analytical Centre (MIAC) of Sverdlovsk oblast, which was launched in 1990 and operates independently of ZAGS. The objectives were to compare the completeness of record linkage (RL) with ZAGS and with MIAC, and to compare the agreement of cause-of-death information obtained from ZAGS and from MIAC, with a focus on malignant neoplasms.

**Methods** RL completeness of identifying cohort members in ZAGS and in MIAC was compared for the period 1990 to 2015. In the next step, for the comparison of the retrieved cause-of-death information, 5,463 deaths (1,009 from cancer) were used that were registered in 2002 to 2015, when causes of death were coded using International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) nomenclature by MIAC. For ZAGS, original cause-of-death text from the death certificates was obtained and then coded according to ICD-10 by the International Agency for Research on Cancer/World Health Organization (IARC/WHO). Agreement was evaluated at various levels of detail, and reasons for any disagreements between the MIAC and the IARC/WHO ICD-10-coded cancer diagnosis were systematically explored.

## Keywords

- ▶ mortality register
- ▶ cause of death
- ▶ International Statistical Classification of Diseases and Related Health Problems
- ▶ asbestos
- ▶ Russian federation

received  
December 16, 2019  
accepted after revision  
March 26, 2020

DOI <https://doi.org/10.1055/s-0040-1710381>.  
ISSN 0026-1270.

© 2020 Georg Thieme Verlag KG  
Stuttgart · New York

License terms



**Results** A total of 10,886 deaths were obtained from all avenues of follow-up for the period 1990 to 2015 in the cohort; 10,816 (99.4%) of these were found in ZAGS. This percentage was 88.3% if only automated deterministic RL was used and 99.4% when deterministic RL was complemented with manual searches of cohort members. Comparison of the cause-of-death information showed agreement of 97.9% at the ICD-10 main group level between ZAGS (coded by IARC/WHO) and MIAC. Of 1,009 cancer deaths, 679 (67.3%) cases had identical coding, 258 (25.6%) cases corresponded at the three-character ICD-10 level, 36 (3.6%) had codes that were within the same anatomical or morphological cluster, and for only 36 (3.6%) cases were major discrepancies identified. Altogether, the agreement between IARC/WHO coding of cause-of-death information from ZAGS and MIAC coding of malignant neoplasms was therefore 96.4%.

**Conclusions** RL completeness and agreement of cause-of-death information obtained from ZAGS and from MIAC were both very high. This is reassuring for the quality of cancer mortality follow-up of the Asbest chrysotile cohort. For future epidemiological studies in the Russian Federation, ZAGS appears to be a reliable information source for mortality follow-up, if the automated RL is complemented with manual searches of cohort members. MIAC is a good resource for prospective studies.

## Introduction

Although all forms of asbestos are known to be carcinogenic to humans,<sup>1</sup> there are open questions related to the quantification of the exposure-response relationship with cancers known or suspected to be caused by chrysotile.<sup>2</sup> The Asbest chrysotile cohort was set up in Asbest town in Sverdlovsk oblast (province), Russian Federation, among the current and former workforce of the world's largest operating chrysotile mine and its processing mills; details are described elsewhere.<sup>3</sup> In brief, the cohort consisted of 35,837 people who worked for at least 1 year in the mine or processing mills in Asbest between 1975 and 2010. This cohort was followed-up for mortality from 1976 to 2015 with vital status ascertained from official records of Sverdlovsk oblast. For cohort members who died while resident in the oblast, their original death certificates were retrieved from the Sverdlovsk oblast Civil Act Registration Office (ZAGS; abbreviation based on the Russian name) to determine the cause of death. In addition to ZAGS archives data, we used information on death cases and causes from the Medical Information Analytical Centre (MIAC) of Sverdlovsk oblast. MIAC was established more recently by the Sverdlovsk oblast Ministry of Health to be the central link in organizing the collection and processing of information and indicators of medical statistics, medical demographic, financial, and personnel components of health care in Sverdlovsk oblast. The MIAC cause-of-death registry includes electronic records starting from 1990. MIAC receives information, including medical death certificates, directly from medical institutions and performs the cause-of-death information extraction and coding independently of ZAGS. The main outcomes for mortality analyses in our study are deaths from cancers of sites known or suspected to be associated with chrysotile.

The Asbest Chrysotile Cohort Study is the first large-scale epidemiological study with international participation in Sverdlovsk oblast. An "oblast" in the Russian Federation is a federal administrative division, similar to a province, and Sverdlovsk oblast is spread over the slopes of the North and Middle Urals and the Western Siberian plain, with approximately 4.3 million inhabitants. Because there is an overall scarcity of cancer cohort studies in the Russian Federation, all procedures for obtaining permissions for access to data from official registration offices, for record linkage of the cohort with the registry data, and for the compilation of the obtained raw data for epidemiological purposes had to be developed for our study. Notably, Russian legislation on access to personal data changed in the post-Soviet period when it became virtually impossible for health authorities and scientific organizations to get access to ZAGS data. Hence, it was important to have rigorous measures in place for data quality assurance and to obtain data from multiple sources and to explore their agreement. Therefore, a unique quality assurance measure to determine the quality of the outcome data of our cohort study was to use the MIAC data as an independent source of mortality data to check the quality of the ZAGS data, in terms of both completeness and accuracy. The MIAC data alone were not sufficient for our cohort study because they did not cover the entire follow-up period. ZAGS was the primary source of mortality data for the Asbest Chrysotile Cohort Study, with 40 years of mortality follow-up (1976–2015). For ZAGS, the original cause-of-death text information was obtained by the study team, and the coding of the underlying cause of death for all deceased cohort members was done according to the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) at the International Agency for Research on Cancer/World Health Organization (IARC/WHO). From MIAC, for the purpose of validation,

we received the underlying cause of death already coded according to ICD-10 for the deaths occurring in 2002 to 2015. Record linkage of cohort members with ZAGS was done using automated deterministic (perfect-match) linkage complemented with manual searches of cohort members. MIAC had its own stochastic record linkage procedure in place, from which we used only the perfect matches for the purpose of this validation study.

Here, we report the results on completeness and quality of the comparison of cause-of-death information obtained from ZAGS and from MIAC, with a focus on neoplasms because this is the major outcome of interest in our cohort study. These results are informative not only for our study but also for providing a benchmark for determining reliable mortality data sources for future epidemiological studies in the Russian Federation.

## Objectives

The first objective was to compare the completeness of mortality data obtained from independently performed record linkages of the study cohort with two computerized cause-of-death registries, ZAGS and MIAC. The completeness of ZAGS mortality data are an important indicator of the data quality of the Asbest Chrysotile Cohort Study for which ZAGS served as the primary source of mortality data, and this was explored as part one of the first objective. The completeness of MIAC mortality data are important for future epidemiological studies, because the MIAC cause-of-death registry was established to facilitate and standardize medical services and research, and this was explored as part two of the first objective.

The second objective was to compare the agreement of cause-of-death information obtained from the ZAGS and MIAC registries, where the processes of death certificate collection, cause-of-death extraction, data entry, and cause-of-death coding according to ICD-10 are done independently of each other. Specific areas of disagreement were assessed among cancer causes of death (malignant neoplasms, ICD-10 codes: C00–C97).

## Methods

### Cause-of-Death Information Sources and Record Linkage Procedures

In the Russian Federation, and earlier in the Soviet Union, a medical death certificate is an official standardized document to ascertain and to report on the fact, circumstances, and causes of death. The part of the medical certificate on cause of death is completed in accordance with rules and principles described in the various editions of ICD. The certificate includes part I for diseases related to the chain of events leading directly to death, and part II to report on unrelated but relevant contributing health conditions.

In our study, the primary source of information on deceased persons and their causes of death was the electronic cause-of-death registry and archives of the Sverdlovsk oblast ZAGS. Death registration is obligatory, and the registration procedure is explicitly defined by Russian Federation law. Death registra-

tion includes two steps. First, a medical death certificate is issued within 1 day after the cause of death was ascertained. Second, based on the medical death certificate and the passport of the deceased person, ZAGS registers the death event and issues an official death certificate. The registration of death in ZAGS is performed when a relative of the deceased person applies for death certification. ZAGS records all personal information of deceased persons, including all names (first name, patronymic name, and surname), complete date of birth, birthplace, date and place of death, and causes of death as indicated in the medical death certificate. The registration of death in ZAGS must be performed within 3 days after the medical death certificate was issued. For some death cases (e. g., forensic investigation), an original cause of death could be updated or modified if further postmortem pathology examination provided additional information, but it is up to the relatives of the deceased person whether they provide this updated cause-of-death information to ZAGS for recording. In 2011, the Sverdlovsk oblast ZAGS started computerizing paper archival records dating back to 1919, allowing record linkage of the mortality data. Hence, by the time we started the record linkage, all information was in electronic format for the study follow-up period of 1976 to 2015.

As the first step to identify deceased persons in the study cohort, deterministic record linkage using all names and date of birth was performed. If a cohort member could have several spellings of their name, all the spellings were used in the record linkage. As the second step, manual searches of cohort members were performed to reduce the number of missed matches due to errors in the personal identifiers.

In addition to linkage with ZAGS data, we also retrieved information from MIAC. MIAC periodically receives information from medical death certificates directly from all medical institutions in the oblast and enters all death information, including causes of death, independently of ZAGS. Unlike ZAGS cause-of-death information, MIAC can occasionally update cause-of-death information, because the information from medical death certificate is sent directly to MIAC by the medical institution that determined the cause of death. Therefore, MIAC receives the final information from all autopsy records, whereas in ZAGS, the outcome of an autopsy is registered only if the relatives of the deceased person provide this updated cause-of-death information. MIAC is not required to archive paper copies of medical death certificates; paper records are destroyed 1 year after they are issued by the medical institution, according to current legislation in the Russian Federation. MIAC has an electronic database of individuals' deaths with coded causes of death, with records starting from 1990. The database is supposed to have complete coverage of deaths that occurred in residents of Sverdlovsk oblast. However, death records for certain years were irretrievably (partly) lost or were not collected for organizational reasons; gaps were identified for 1993, 1994, and mainly 2009. Especially for 2009, when the database system was moved to new software, some records were not entered. For record linkage between the cohort and MIAC data, MIAC used a stochastic record linkage procedure that allowed for various combinations with a varied degree of

accuracy of the personal identifiers: use of two names instead of three (from among first name, patronymic name, and surname), use of the initials instead of the first name and patronymic name in full, and use of part of the date of birth (month and year, day and month, or just year) instead of the complete date of birth. For the purposes of this comparison, we selected records matched on all three names and complete date of birth (i.e., perfect matches).

### Cause-of-Death Coding

During the study follow-up period of 1976 to 2015, several editions of ICD, as well as domestic statistical disease nomenclatures were used to code causes of death in the Russian Federation, and earlier in the Soviet Union (→ [Supplemental Figure S1](#); available online only). The main limitation of domestic statistical disease nomenclatures was the combination of several specific diseases into one group with one aggregated code assigned. Those aggregated codes were used for statistical reporting but are too crude for epidemiological studies because they are lacking the required level of precision on an individual's cause of death. Moreover, ICD revisions before the 10th revision did not have an individual code for mesothelioma, which is one of the major outcomes of interest for the cohort study. Finally, ZAGS do not systematically have coded causes of death in their registries.

To overcome all the above-mentioned limitations, at IARC/WHO, we received the original text from the death certificates of each deceased study cohort member from ZAGS; the selection of underlying cause of death was performed by a Russian medical doctor, considering all the available information from each death certificate and using the ICD-10 rules for selection of the underlying cause of death, blindly to the MIAC data (→ [Supplemental Figure S1](#); available online only).

As a result of the cohort linkage with MIAC death records, we received information on individuals' causes of death either coded using a special non-ICD-based MIAC nomenclature or coded using ICD-10 from 2002 onward (→ [Supplemental Figure S1](#); available online only), but the original text of causes of death was not available (because paper records had to be destroyed after 1 year). In MIAC, statisticians select and code the underlying cause of death, and in case of injury or poisoning also the external cause leading to death; sometimes the medical practitioner who issued the death certificate coded the underlying cause of death that MIAC may use after checking for correctness.

For the validation study, we restricted the time period for comparing the cause-of-death information from ZAGS and MIAC from 2002 to 2015, for which MIAC consistently provided the underlying cause of death coded according to ICD-10 (→ [Supplemental Figure S1](#); available online only).

### Statistical Methods

For the comparison of the completeness of record linkage of the study cohort with ZAGS and with MIAC, we compared only cohort members who died in 1990 to 2015 (because the MIAC electronic cause-of-death registry was established in 1990). For ZAGS completeness is reported in two ways,

namely, by including only those who were identified in ZAGS through the automated deterministic record linkage described above and by including also those who were identified during the manual searches of cohort members. We report simple numbers and percentages of completeness.

From the sample described above, for comparison of the agreement of cause-of-death information, we used only deaths from 2002 onward, because MIAC started coding cause-of-death information according to ICD-10 in 2002. We compared agreement between the ICD-10 main groups for all deaths. For neoplasms overall and for malignancies in particular, we used the detailed ICD-10 code and made comparisons of diagnosis at different ICD-10 levels. All disagreements between the IARC/WHO codes using the original cause of death text from ZAGS and the MIAC codes were systematically and individually assessed to determine whether the disagreement was the result of a coding error in either IARC/WHO or MIAC codes or whether the original text on the death certificate was ambiguous.

## Results

### Overall Completeness of Record Linkage with ZAGS

From all available sources of vital status, described in detail elsewhere,<sup>3</sup> the total number of deaths in the cohort during the period from January 1990 to May 2015 was 10,886. Of those, 10,816 (99.4%) were found in ZAGS (9,613 through automated deterministic record linkage and 1,203 through manual searches of cohort members). This corresponds to a record linkage completeness of ZAGS of 88.3% if only automated deterministic record linkage is used and of very high 99.4% when the deterministic record linkage was complemented with manual searches; this addresses part one of the first objective. Only 70 deaths found in other sources were not found in ZAGS (10,886 – 10,816). Of those 70 deaths, 11 were found only in MIAC. The 59 deaths of cohort members not identified in ZAGS or in MIAC were found through record linkage with the National Pension Fund and the Federal Migration Service in Sverdlovsk oblast.

As noted above, manual searches of cohort members in ZAGS identified 1,203 deaths of cohort members in 1990 to 2015 that were not initially found through automated deterministic record linkage. This could be because for the manual searches we used additional information, such as last known address, and allowed for even broader spelling variations to account for potential typographical errors in personal identifiers. For example, sometimes the birth month was written in Roman numerals on the death certificate and then at the data entry stage it was converted into Arabic numerals incorrectly (e.g., February was written as “II” on the death certificate and entered as “11” instead of “2” in the database). Another typical data entry error was transposing the day and month of the date of birth.

### Completeness of Record Linkage with ZAGS Compared with MIAC

For the comparison of the completeness of record linkage between ZAGS and MIAC, we used as the denominator only the deaths found either in ZAGS through deterministic record

linkage or in MIAC. For the period from January 1990 to May 2015, this was a total of 9,979 deaths. Of those, 8,986 (90.0%) were identified in both ZAGS (through automated deterministic record linkage) and MIAC. Of the 993 discrepant cases, 627 (6.3% of the total deaths) were found in ZAGS but not in MIAC, and 366 (3.7% of the total deaths) were found in MIAC but not in ZAGS. Consequently, independently of each other, 96.3% of cases were found in ZAGS and 93.7% of cases were found in MIAC; this addresses part two of the first objective. For lung cancer deaths, 526 (96.7%) of the 544 deaths were identified in ZAGS through deterministic record linkage, and all of the remaining 18 were later found through manual searches (leading to 100% completeness of record linkage for lung cancer). For mesothelioma, all of the eight deaths that occurred within the search window were found in both ZAGS and MIAC.

For the 627 cases not found in MIAC, there were 11 to 34 missing deaths per year with no apparent time trend, with the exception of 145 missing deaths in 2009. This was known to be a problematic year, and we had been informed about possible underrecording in that year (see **Methods**). Excluding 2009 from the analysis shown above would improve the completeness of the record linkage of MIAC data from 93.7 to 95.0%.

For the 366 cases not found in ZAGS, there were 12 to 24 missing deaths per year during the period 1990 to 2008 (with the exception of 37 in 2002, for unknown reasons), and there were 7 or fewer missing deaths per year from 2009 to 2015. The resulting ZAGS record linkage completeness for 2009 to 2015 is 97.3%, which is slightly higher than the completeness of 96.3% for the total period of 1990 to 2015 as shown above. Of the 366 cases not found in ZAGS through deterministic record linkage, 361 were later found in ZAGS through manual searches of cohort members (**Table 1**).

### Comparison of Cause-of-Death Information

Comparison of cause-of-death information was based on 5,463 deaths in 2002 to 2015 with cause of death coded using ICD-10 both for ZAGS data (coded by IARC/WHO) and MIAC data (coded by MIAC), to address the second objective. Based on ZAGS data, about half (50.6%) of the coded deaths were due to circulatory system diseases (ICD-10 main group I), followed by neoplasms (18.4%). **Table 2** compares ZAGS-

based (IARC/WHO-coded) and MIAC-coded causes of death at the ICD-10 main group level, with overall agreement of 97.9% (5,349 out of 5,463 deaths). This demonstrates high quality of coding practice. It also shows that if there were any potential later revisions in the medical death certificates, which were more systematically available at MIAC, they did not usually result in major changes in the main group coding.

A detailed exploration of 1,009 cancer deaths (identified in either the ZAGS-based IARC/WHO-coded or MIAC-coded databases, or both, as ICD-10: C00–C97) showed that 679 (67.3%) cases had identical coding, corresponding at the four-character ICD-10 level, and 258 (25.6%) cases corresponded at the three-character ICD-10 level (**Table 2**). Of the remaining 72 cases, 36 (3.6% of the total cancer deaths) had cancer diagnosis codes that were within the same anatomical or morphological cluster of neighboring ICD-10 groups that will be combined in risk analyses such as cancers of digestive organs, the genital tract, lymphoma, or leukemia (called partial agreement in **Table 3**). Altogether, the agreement between IARC/WHO-coded ZAGS data and MIAC coding of malignant neoplasms was therefore 96.4%.

Major discrepancies were found for only 36 deaths (3.6% of the total cancer deaths) of which 19 were coded as cancer deaths by IARC/WHO but not by MIAC or vice versa. We found six deaths based on ZAGS data on underlying cause of death that were not coded as cancer by IARC/WHO, but in the MIAC database these deaths were coded as cancer, namely, two cancers of the stomach, two of the pancreas, and two of the lung. Manual checks of text information for these six deaths available in the ZAGS cause-of-death registry did not provide any evidence of a cancer, confirming the initial cause-of-death coding. Hence, it is likely that MIAC later received updated information through the revised medical death certificate. It is less likely but also possible that the information on cause of death in the ZAGS database was incomplete due to data entry or data transfer errors, because the noncancer conditions indicated as the cause of death in ZAGS (cachexia, posthemorrhagic anemia, and acute cardiovascular insufficiency) are possible consequences of a cancer progression.

In five deaths, there was a cancer diagnosis code based on ZAGS data but not in MIAC data. These included five deaths due to cancers of the colon, ovary, kidney, and bone (secondary cancer) and multiple myeloma, where the cancer diagnosis was clearly stated in the text of the death certificate, justifying the coding correctness of the IARC/WHO coding. These discrepancies between ZAGS and MIAC information could be due to data entry errors or incomplete information in the MIAC electronic database; they could also be due to coding errors in MIAC when a complication was coded instead of the initiating condition (e.g., intestinal obstruction by ovary cancer). Five cases were coded by MIAC as malignant brain tumors, whereas the cause of death in ZAGS was written as “brain tumor,” that is, it remained unclear whether the tumor was benign or malignant. Therefore, those five cases were coded by IARC/WHO as brain tumors of uncertain or unknown behavior. Similarly, a diagnosis of acute lymphoproliferative disease was coded as lymphoproliferative disease of unspecified behavior by IARC/WHO but as an acute

**Table 1** Record linkage completeness between ZAGS and MIAC based on the deaths found either in ZAGS through deterministic record linkage or in MIAC (total of 9,979 deaths), 1990–2015

n (%)	MIAC (found)	MIAC (not found)	Total
ZAGS (found)	8,986 (90.0)	627 (6.3)	9,613 (96.3)
ZAGS (not found) <sup>a</sup>	366 (3.7)	NA	366 (3.7)
Total	9,352 (93.7)	627 (6.3)	9,979 (100.0)

Abbreviations: MIAC, Medical Information Analytical Centre; NA, not available; ZAGS, Civil Act Registration Office.

<sup>a</sup>Of the 366 patients, 361 were later found through manual searches in the ZAGS registry.

**Table 2** Concordance between cause-of-death coding in ZAGS and MIAC data by ICD-10 main groups (disease chapters), 2002–2015

ZAGS cause-of-death code	MIAC cause-of-death code														Total		
	A00–B99	C00–C97	D00–D48	D50–D89	E00–E90	F00–F99	G00–G99	I00–I99	J00–J99	K00–K93	L00–L99	M00–M99	N00–N99	Q00–Q99		R00–R99	S00–T98
Infectious (A00–B99)	84	1						2	3								90
Malignant neoplasms (C00–C97)		990	1	1			1	1		1					2		996
Nonmalignant neoplasms (D00–D48)		6	3														9
Blood, blood forming organs (D50–D89)		1		4													5
Endocrine, nutritional, metabolic (E00–E90)					33			1									34
Mental (F00–F99)						6				1							8
Nervous system (G00–G99)						39		3							1		43
Circulatory system (I00–I99)	1	2				3		2,716	9	6			1		28		2,767
Respiratory system (J00–J99)	1	2				1		2	164						5		175
Digestive system (K00–K93)										351					3		354
Skin (L00–L99)	3										4				1		8
Musculoskeletal system (M00–M99)											5		1				6
Genitourinary system (N00–N99)								2	1				33				36
Congenital malformations (Q00–Q99)								1						1			2
Symptoms, not elsewhere classified (R00–R99)		1			1			7		2			1		195		207
External causes (S00–T98)								6	1						1	714	723
<b>Total</b>	<b>89</b>	<b>1,003</b>	<b>4</b>	<b>5</b>	<b>34</b>	<b>6</b>	<b>45</b>	<b>2,741</b>	<b>178</b>	<b>361</b>	<b>4</b>	<b>5</b>	<b>36</b>	<b>1</b>	<b>197</b>	<b>754</b>	<b>5,463</b>

Abbreviations: ICD, International Statistical Classification of Diseases and Related Health Problems; MIAC, Medical Information Analytical Centre; ZAGS, Civil Act Registration Office.

**Table 3** Concordance between cause-of-death coding<sup>a</sup> of all cancer deaths from ZAGS or MIAC (ICD-10 C00-C97), 2002–2015

Total (n)	Agreement		Partial agreement	No agreement	
	All 4 ICD-10 characters n (%)	First 3 ICD-10 characters n (%)	Neighboring groups combined in risk analyses n (%)	Within cancer group (C00–C97) n (%)	Outside cancer group (C00–C97) n (%)
1009	679 (67.3)	258 (25.6)	36 (3.6)	17 (1.7)	19 (1.9)

Abbreviations: ICD, International Statistical Classification of Diseases and Related Health Problems; MIAC, Medical Information Analytical Centre; ZAGS, Civil Act Registration Office.

<sup>a</sup>ZAGS: coded according to ICD-10 by International Agency for Research on Cancer/World Health Organization based on original text information on underlying cause of death as recorded in ZAGS; MIAC: Coded according to ICD-10 by MIAC (original text not archived).

unspecified leukemia in MIAC. One case with a diagnosis of cancer of the uterus in ZAGS was coded as carcinoma in situ of cervix uteri in MIAC. Finally, one case of metastasis with unknown primary in ZAGS was coded as noncancer disease (stroke sequelae) in MIAC. In both cases, either MIAC had new information through the revised medical death certificate or wrong information was entered in the ZAGS cause-of-death registry but no coding error was detected.

The remaining 17 major discrepancies between IARC/WHO and MIAC cancer coding were as follows. First, in the ZAGS registry four cases were referred to as metastatic cancers of a specific site (stomach, liver, and bone) with unknown site of primary cancer. These cases were coded by IARC/WHO as secondary malignant neoplasm of a specific organ, whereas in MIAC these cases were coded as primary cancer of that organ. Second, one death from stomach cancer and one from lung cancer, with metastasis in both cases, were coded as primary stomach and lung cancer, respectively, by IARC/WHO, whereas in MIAC, they were coded as secondary cancer of the liver and secondary cancer of the lung. Third, one case of mesothelioma of the pleura was erroneously coded as cancer of the pleura in MIAC, and one case of melanoma of cheek mucosa was erroneously coded as skin cancer of the lip in MIAC. Another discrepancy in three cancer codes between ZAGS and MIAC data also seems to be due to coding errors in the MIAC registry: cancer of the accessory sinus was coded as cancer of the nasopharynx, cancer of the thigh as cancer of connective and soft tissue unspecified, and cancer of the vulva as cancer of the breast. Another reason for the coding discrepancies was a mix-up of the codes for cancers of the larynx, pharynx, and hypopharynx in MIAC (four cases). Two cases referred to in death certificates as malignant neoplasms of lymphoid tissue (mantle-cell lymphoma, lymphoplasmacytic lymphoma) had leukemia codes in MIAC. Those altogether 17 discrepancies showed coding errors in the MIAC registry, although some may be explained by MIAC having revised information available; for the Asbest Chrysotile Cohort Study it was important to know that the coding performed by IARC/WHO correctly reflected the cause-of-death information available on the ZAGS death certificate.

## Discussion

The comparison of data from two sources of causes of death confirmed the value of the extensive efforts made to collect

follow-up data for the Asbest Chrysotile Cohort Study from all available sources. Obtaining permissions for access to data was a laborious but worthwhile bureaucratic effort, and obtaining the data became an iterative process because there was little previous experience of providing data from these administrative registries for epidemiological studies. Clearly, medical research benefits from the creation of the MIAC electronic cause-of-death registry in the oblast, because MIAC occasionally has better cause-of-death information than ZAGS (through receiving potentially updated medical death certificates directly from the medical institutions) and has advanced record linkage procedures. However, because MIAC was established more recently, with only the coded causes of deaths available and very short storage of paper copies of medical death certificates, the MIAC registry is important as a supplementary but not as an alternative primary mortality data source compared with ZAGS for historical cohort studies. Based on this assessment, MIAC is not as complete as ZAGS, and periodic linkage between the two registries will be of benefit for both.

For the Asbest Chrysotile Cohort Study, we found excellent record linkage completeness for ZAGS which is reassuring evidence for the mortality follow-up of the cohort. However, the high completeness of record linkage was only achieved by doing manual searches of cohort members as a second step, resulting in overall identification of more than 99% of deaths of oblast residents from among those combined from ZAGS, MIAC, or other vital status follow-up sources. Identification of deaths based solely on automated deterministic record linkage resulted in a completeness rate of 88.3% that would not be sufficient for epidemiological quality standards.

Comparison of ICD-10 codes in the IARC/WHO-coded ZAGS data and the MIAC-coded data showed an excellent agreement, especially for cancer deaths. This confirms both the high quality of ICD-10 coding, as well as it confirms there was only small proportion of inconsistent information on underlying causes of death of deceased persons in the two registries, albeit their independent processes in collecting and recording this information. In case of discrepancies, however, there was value in having access to the original death certificates rather than using codes, which could be influenced by different coding practices and changes in underlying medical classifications over time. We had anticipated more discrepancies because MIAC receives

information with a potentially updated diagnosis, but this advantage was less obvious than expected because numbers were small. Notably, because information collection, cause-of-death extraction, data entry, and coding of potentially not even the same underlying medical death certificate were totally independent between ZAGS and MIAC, this cross-check of ZAGS and MIAC cause-of-death information went beyond agreement of coding but was an external validation of how complete and accurate the cause-of-death information in ZAGS really was.

Quality checks of mortality registries have recently been performed for other historical cohort studies in the Russian Federation, on radiation-related health effects, including oblasts in the region of the Southern Urals like the present study but with much smaller numbers of subjects. A validation study of 246 death certificates registered in the mortality registry of Ozyorsk town showed 98% validity of the ICD-9 main group level, and among neoplasms 89% were in the same three-digit diagnostic category.<sup>4</sup> Another validation study of the mortality registry at the Urals Research Center for Radiation Medicine in Chelyabinsk, with coding of 500 death certificates by three different institutions, showed agreements on ICD-9 main group level of 80 to 86%.<sup>5</sup> Both studies used ZAGS death certificates as a primary source for cause-of-death ascertainment, but their follow-up periods started in the 1950s. Danilova et al investigated differences in coding practices across various oblasts of the Russian Federation using death counts and population estimates from the Russian Federal State Statistics from 2002 to 2012. They concluded that overall there was a high degree of variation caused by different coding practices, especially for certain cardiovascular diseases, nervous system diseases, and ill-defined causes of death, but a high level of consistency for cancer.<sup>6</sup> Nevertheless, this confirmed our choice to centrally code the causes of death according to ICD-10 for the long 40-year time span of our cohort study. In Switzerland, in a somewhat different approach by comparing causes of death with hospital diagnosis for the Swiss National Cohort, for 83% of individuals, the cause of death could be traced among the hospital diagnosis, showing that overall there are some limitations of cause-of-death information.<sup>7</sup>

### Limitations and Strengths

The main limitation of our study is that MIAC covers only the more recent part of the follow-up period of the Asbest Chrysotile Cohort Study, starting from 1990. We have no comparison data for 1976 to 1990. No discernible time trend in the completeness of record linkage was observed in the 1990s, that is, a period when the country transitioned through major political and economic challenges. Therefore, it is unlikely that reporting of mortality data was much worse before then. For the accuracy of cause-of-death coding, we had only comparison data starting from 2002. Because we had access to the original death certificates for the entire study follow-up period, we are confident that the coding quality for the earlier years is equally good. Another limitation is that with this approach alone it was not possible to precisely quantify

the overall follow-up completeness of the Asbest Chrysotile Cohort Study, because there may have been individuals who were missed in the record linkages with all sources for vital status information among those who were censored before the end of the study follow-up in 2015.<sup>3</sup> Although from the recent years only a small portion of patients was missed in ZAGS that were found in other sources, there may have been more patients missed in the earlier years of follow-up, especially before the data were immediately entered into an electronic registry (such as from typographical errors from retrospectively entering information from paper documents from the 1970s and 1980s, or possibly from the loss of documents). A limitation inherent in all studies around the world using death certificates is that there is no information of how well the underlying cause of death recorded on the death certificate resembles the true underlying cause of death.

Despite its limitations, this study has some specific strengths. This is the first study to compare independent cause-of-death information sources in the Russian Federation in a large-scale epidemiological study. The long-term study period has important implications for conducting historical cohort studies in the Russian Federation in general. Furthermore, this study has been strengthened by international collaboration, and the assessments of completeness and accuracy will serve as an important benchmark for the accuracy of mortality outcomes in future public health studies.

### Conclusion

In conclusion, the very high completeness of record linkage indicates a good-quality follow-up for cancer mortality in the Asbest Chrysotile Cohort Study based on the vital status and cause-of-death information obtained from the Sverdlovsk oblast ZAGS. This was, however, achieved only by complementing the automated deterministic record linkage between the cohort and ZAGS data with manual searches of cohort members. Comparison of cancer deaths using a second, independent source (MIAC) affirmed the accuracy of the coding of cause-of-death information in the Asbest Chrysotile Cohort Study. This was achieved because we obtained the original text information from ZAGS death certificates and were able to perform the cause-of-death coding with our own experienced staff at IARC/WHO, thus avoiding errors associated with changes in classifications and coding practices with time, when coding conversions need to be applied.

For future epidemiological studies in the Russian Federation, ZAGS appears to be a reliable information source for mortality follow-up, if the automated deterministic record linkage is complemented with manual searches or if ZAGS implements stochastic record linkage, as is done by MIAC. MIAC has the advantage of occasionally having better information on causes of death, although this advantage over ZAGS data was perhaps less than anticipated. For the time being, the MIAC cause-of-death registry has to further optimize data completeness before becoming a major information source for mortality follow-up in future epidemiological studies.



### Conflict of Interest

E.K. and S.K. reported receiving, on behalf of their institutes and personally through consulting firms, payments from companies to evaluate exposure to asbestos and risk of asbestos-related disease in those workplaces. All other authors have no competing interests to declare. For full transparency, E.K. reported participation as an occupational and environmental health expert as part of the delegation of the Russian Ministry of Health at multiple World Health Assembly meetings as well as at the Conference of the Parties to the Basel and Rotterdam Conventions. E.K. and S.K. reported attending meetings organized by the International Chrysotile Association and reported that all expenses for attendance were paid by their respective institutes.

### Acknowledgments

The work was supported by the Ministry of Health of the Russian Federation in the framework of the Federal target program “National System of Chemical and Biological Safety of the Russian Federation” of 2009–2014 and of 2015–2020 under a general framework of action between the Federal state budgetary scientific institution “Izmerov Research Institute of Occupational Health” (IRIOH) and the International Agency for Research on Cancer (IARC). The study is monitored by an independent Scientific Advisory Board (SAB), which oversees the progress of the study; the SAB members are Professor Franco Merletti (Chair), Professor Mads Melbye (until 2017), Professor Julian Peto, Professor Martin Rööslä (from 2017), and Dr Antti Tossavainen. The authors also like to thank the Data

Entry Team at Asbest and the Study Team members of the Asbest Chrysotile Cohort Study not involved in this particular task, for their contribution to the study. Where authors are identified as personnel of IARC/WHO, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of IARC/WHO.

### References

- 1 IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Arsenic, metals, fibres, and dusts. IARC Monogr Eval Carcinog Risks Hum 2012;100(Pt C):11–465
- 2 Schüz J, Schonfeld SJ, Kromhout H, et al. A retrospective cohort study of cancer mortality in employees of a Russian chrysotile asbestos mine and mills: study rationale and key features. *Cancer Epidemiol* 2013;37(04):440–445
- 3 Asbest Study: occupational exposure to chrysotile in workers in mines and processing facilities in Asbest, Russian Federation. Available at: <http://asbest-study.iarc.fr>. Accessed March 31, 2020
- 4 Azizova TV, Fedirko V, Tsareva Y, et al. Mayak workers study cohort. An inter-institutional comparison of causes of death in the cause-of-death register of Ozyorsk in the Russian Federation. *Methods Inf Med* 2012;51(02):144–149
- 5 Startsev N, Dimov P, Grosche B, Tretyakov F, Schüz J, Akleyev A. Methods for ensuring high quality of coding of cause of death. The mortality register to follow Southern Urals populations exposed to radiation. *Methods Inf Med* 2015;54(04):359–363
- 6 Danilova I, Shkolnikov VM, Jdanov DA, Meslé F, Vallin J. Identifying potential differences in cause-of-death coding practices across Russian regions. *Popul Health Metr* 2016;14:8
- 7 Zellweger U, Junker C, Bopp M; Swiss National Cohort Study Group. Cause of death coding in Switzerland: evaluation based on a nationwide individual linkage of mortality and hospital in-patient records. *Popul Health Metr* 2019;17(01):2