Mayak Workers Study Cohort*
An Inter-Institutional Comparison of Causes of Death in the Cause-of-Death Register of Ozyorsk in the Russian Federation

T. V. Azizova1; V. Fedirko2; Y. Tsareva1; F. Tret'yakov1; C. Funch Lassen3; S. Friis3; J. Schüz2
1Southern Urals Biophysics Institute (SUBI), Ozyorsk, Russia;
2International Agency for Research on Cancer (IARC), Lyon, France;
3Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark

Keywords
Mortality register, cause of death, inter-institutional comparison, radiation, Russia

Summary
Background: The cause-of-death register at the Southern Urals Biophysics Institute (SUBI), Ozyorsk, Russia, was established to document the number and causes of deaths in the Mayak workers cohort, which includes all persons (N = 22,377) employed at Mayak nuclear facility between 1948 and 1982. Most workers were occupationally exposed to high doses of ionizing radiation and have been shown to have increased risks of various chronic diseases including cancer.

Objectives: To investigate the quality of cause of death coding in the SUBI register.

Methods: A random sample of 246 deaths (~1% of the total) was coded independently at the SUBI and the Danish Cancer Society using the International Classification of Diseases 9 (ICD-9). Proportions of matching codes were calculated.

Results: Overall, 233 deaths (95%) were identically classified using the ICD-9 main category matching. Excluding mismatches that were considered to be incorrectly coded during validation, the validity of the register increased to 98%. Using the specific ICD-9 first three-digit matching, 182 deaths were identically coded (74%) and the respective validity of the register was 85%. There were also some non-resolvable discrepancies demonstrating limitations of assigning one code for each death or using language-adapted ICD-9 version.

Conclusions: This validation study was an important quality check of a register used for mortality follow-up in a highly influential epidemiological study on radiation-related health effects. The results of the inter-institutional comparison were generally favourable; however, since the comparison revealed individual mismatches and some systematically differing coding practices, it is essential to repeat it on a regular basis in order to maintain a high quality.

Correspondence to:
Joachim Schüz, PhD
International Agency for Research on Cancer (IARC)
Section of Environment and Radiation
150 cours Albert Thomas
69372 Lyon Cedex 08
France
E-mail: schuzj@iarc.fr

1. Introduction

The Mayak Production Association (Mayak PA) was the first nuclear facility at the Southern Urals in the former Soviet Union, situated 10 km from Ozyorsk, and 100 km from Chelyabinsk. The Mayak PA was commissioned on June 19, 1948 and consisted of facilities for production of the weapons-grade plutonium including reactors, radiochemical and plutonium production facilities, and auxiliary facilities. Due to the lack of experience in operating such facilities and short deadlines set by the Government for the production of plutonium, most workers were occupationally exposed to high doses from external and/or internal radiation during the period of Mayak PA commissioning and development (1948–1958) [1, 2].

The individual dosimetry monitoring of external exposure for all nuclear workers has been initiated since the Mayak PA commissioning. The continuous internal dosimetry monitoring of workers exposed to transuranium radionucleotides started in the 1960s. Medical follow-up of workers has been performed on the regular basis by a specially designed standard program, which includes obligatory pre-employment medical examination and routine medical examinations on an annual basis [3, 4].

The Mayak workers study cohort includes all workers first employed at one of the main facilities between the Mayak PA commissioning and the end of 1982, irrespective of sex, age, nationality, or occupation. The cohort study has several advantages including its large size (22,377 individuals), long-term period of follow-up (over 60 years), individually measured radiation doses, wide range of external and internal doses (0 up to 10 Gy), heterogeneity by sex, age, and ethnicity, full information on vital status and health effects (over 95% complete), and additional information on non-radiation risk factors assessed during the medical examinations [5, 6].

The epidemiological studies of the Mayak worker cohort have shown an association between exposure to plutonium and increased mortality from lung, liver, bone cancers [7], and leukaemia and solid cancers taken as a whole [8]. An increased
incidence of circulatory diseases (ischemic heart disease and cerebrovascular diseases) was also observed among the Mayak workers first employed in 1948–1958 associated with external gamma-ray dose and in some analyses with internal dose from plutonium [9, 10].

The basis for mortality follow-up is the cause-of-death register established at the Southern Urals Biophysics Institute (SUBI) in 1988. All deaths among Mayak worker cohort members and a more recently established Mayak worker offspring cohort are recorded in this register [11, 12].

2. Objectives

The aim of the present study was to conduct an inter-institutional comparison of cause of death coding in the SUBI register with the coding of the same cases by specialists from an external professional institution.

3. Methods

The cause-of-death SUBI register includes over 35,000 cases of death that occurred in the city of Ozyorsk, Chelyabinsk region, Russian Federation during the period between 1948 up to now. During the follow-up period of the Mayak worker cohort (1948–2008), 31,527 individuals have died in Ozyorsk. Cases of death among children were excluded, resulting in a total of 30,120 eligible death certificates for our validation.

The mortality data were collected from the medical death certificates (Supplemental document 1) obtained from the Office of Civilian Registration. Standard medical certificates of death were filled in by a physician for all cases of death occurring in the city of Ozyorsk after the year 1948. All causes of death were coded by the same technician in accordance with the Russian edition of the Ninth Revision of International Classification of Diseases (ICD-9) and entered into the computer database. The verification of data in the cause-of-death register is performed routinely through comparison of the data stored at the SUBI and at the Civilian Registration Office of Ozyorsk, the only institution in the city that registers the death and stores all death certificates; thus, the completeness of data in the register is 100%. Additionally, quality assurance checks of the death information have been performed in the framework of the previous epidemiological studies [9, 10], in which it was found that 80%, 25%, and 50% of death certificates were confirmed by autopsy before 1990, 1990–2000, and after year 2000, respectively. Malignant neoplasms were found to be histologically verified for 80% of all cases.

The inter-institutional comparison of cause of death coding was performed to validate the quality of data in the SUBI register. For validation, 251 death certificates were randomly sampled from the eligible 30,120 death certificates at the 120-pace from all cases of death registered between 1948 and 2008 in the cohort of Mayak workers first employed at the main facilities in 1948–1982, by doing so accounting for the mortality structure within this cohort. The cause-of-death coding was performed by specialists from SUBI and the external validation coding was performed by professional coders (medical doctors) at the Institute of Cancer Epidemiology of the Danish Cancer Society (DCS), Copenhagen, independently from each other. For the validation study, the original texts of the death certificates were entered in an Excel spread sheet and together with the original ICD-9 codes were delivered to the International Agency for Research on Cancer (IARC), where the original codes were removed and kept separately before forwarding the data to the DCS. A Russian speaking scientist from IARC was then assisting the Danish coders during the whole coding process. The original ICD-9 coding was linked to the validation coding only after all cases were double coded.

3.1 Statistical Methods

Death certificates were coded independently by professional coders at the SUBI and the DCS using the adapted ICD-9 in Russian and original ICD-9 in English, respectively. Data were merged in one file, and proportions of correctly classified causes of death were calculated after excluding two typing errors in the original codes when entering the data in the Excel spread sheet and three cases with ambiguous old terminology causing problems in proper translation (namely ‘haematosarcoma’ and ‘lympholeukosis’), yielding a final study population of 246 deaths. Proportions were calculated on two levels: 1) ICD-9 main classes 2) specific ICD-9 topics (first 3 digits of ICD-9 code). The main categories of diseases and other conditions used in this study and their corresponding ICD-9 codes are shown in Supplemental Table 1. The Russian and original English edition are often different in the fourth digit of ICD-9 (mainly no existing fourth digit in the Russian edition), so consequently this level of detail was not used in the comparisons. The SUBI register was used as the gold standard in the validation. In addition, we further computed Cohen’s Kappa coefficients to evaluate the inter-rater reliability [13].

After the linkage, the discrepancies were discussed at the round table involving all coders and classified as either i) non-preferable code in the cause-of-death register, ii) non-preferable code used during validation, iii) non-resolvable discrepancy, or iv) both codes valid. Also it was checked whether discrepancies revealed any systematically different coding patterns.

4. Results

The frequency distributions of the deaths selected for this study by age and sex are shown in Table 1. Since 2000, on average, about 1,000 deaths per year have occurred among individuals occupationally exposed at the Mayak PA. Thus, there were 1,142 deaths registered in 2008, including 37.3% of individuals who died at the age of < 65 years, 23.6% at 65–74 years, and 38.7% at ≥ 75 years.

The validation study represents approximately 1% of all death certificates from the study cohort of Mayak workers entered into the SUBI cause-of-death register. Of the 246 eligible death certificates, 233 were identically classified using the main category matching (94.7%) and 182 using the first three-digit matching (74.0%); Tables 2, 3 and 4). As stated above, the SUBI register was used as the gold standard for the interpretation of the results of
Methods Inf Med 2/2012 © Schattauer 2012

146 T. V. Azizova et al.: Mayak Workers Study Cohort

Table 3 and 4. The Kappa coefficient for the inter-rater reliability was 0.92 (95% confidence interval: 0.88–0.96; P-value < 0.0001) for the main categories, showing good agreement [13].

4.1 Description of Mismatches in the Main ICD-9 Categories

Overall thirteen mismatches within the main ICD-9 categories were identified. Based on a round-table discussion among the coders, six were identified as non-preferable coding in the register, three as non-preferable coding during the validation, one case who died in 1950 with death certificate terminology that was not classifiable without extra information, and three cases for which no agreement was reached as to what code to use since both classifications appeared to be valid (e.g., case with ‘Chronic glomerulonephritis’, ICD-9 code ‘582’ and ‘Embolism and thrombosis’, ICD-9 code ‘453.8’). Therefore only six mismatches were considered to be incorrectly coded in the SUBI cause-of-death register, which gives a validity of 97.6% on the ICD-9 main classes’ level.

4.2 Description of Mismatches by the Specific ICD-9 First Three Digits of the Code

4.2.1 Neoplasms

There were seven mismatches in the category ‘Neoplasms’ (ICD-9 ‘140–239’).

Among them, four cases with multiple possibly primary cancers which were coded as one non-systematically selected cancer in the SUBI register whereas in the validation study these cases were coded as ‘Multiple cancer’ (ICD-9 code ‘199.0’). It should be noted that there are no subtopics in ICD-9 code ‘199’ in the adapted version of International Classification of Diseases (ICD) published in Russian. There was one death certificate with old terminology used for cancer of lymphatic and hematopoietic tissues, ‘lymphomatos lymphoma’. In the register, this case was coded as unspecified lymphoid other malignant neoplasm (code ‘202’), whereas in the validation study as lymphoma of lymphoplasmacytoid type (code ‘200.8’). The latter coding was only available in the original ICD-9 version (volume 2 Alphabetical Index).

Another case of unspecified malignant neoplasms of haematopoietic tissue was coded as unspecified leukaemia, code ‘208.9’ in the register and as unspecified neoplasm ‘200.8’ in the validation study. The last discrepancy was in the coding of cancer in the peritoneum and retroperitoneum, for which in the register, a non-preferable code

Table 1

Distribution of the random sample of cases of deaths by sex and age

<table>
<thead>
<tr>
<th>Age groups (years)</th>
<th>Men (N, %)</th>
<th>Women (N, %)</th>
<th>Both sexes (N, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;65</td>
<td>51 (40.5)</td>
<td>29 (24.2)</td>
<td>80 (32.5)</td>
</tr>
<tr>
<td>65–74</td>
<td>45 (35.7)</td>
<td>39 (32.5)</td>
<td>84 (34.2)</td>
</tr>
<tr>
<td>≥75</td>
<td>30 (23.8)</td>
<td>52 (43.3)</td>
<td>82 (33.3)</td>
</tr>
<tr>
<td>Total</td>
<td>126</td>
<td>120</td>
<td>246</td>
</tr>
</tbody>
</table>

Table 2

Distribution of underlying causes of death in the SUBI register and validation study by main ICD-9 categories

<table>
<thead>
<tr>
<th>Categories in the SUBI register</th>
<th>Cases in the validation study</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>6 100 7 1 1 1 101 2 19 4 0 1 3</td>
<td>246</td>
</tr>
</tbody>
</table>

* Abbreviations of the categories of diseases and other conditions used in the study and the corresponding ICD-9 codes are presented in Supplemental Table 1.
'158' (malignant neoplasms of retroperitoneum and peritoneum) was used; while in the validation study a more preferable code '195.2' (malignant neoplasm of abdomen) was utilized.

4.2.2 Diseases of the Circulatory System

A total of 37 mismatches were found in this category. Among them, 15 were due to systematic use of special coding in the register when a more general ICD-9 topic should be used (ICD-9 '414.9'; "Other forms of chronic ischemic heart disease, unspecified") as proposed in the validation study. Seven mismatches in the sub-category of cerebrovascular diseases (ICD-9 '430 to 438') were also noted and attributed to systematic coding practices in the register. Out of the remaining fifteen mismatches, a round-table discussion assigned three as non-preferable coding in register, six as non-preferable coding during the validation, and five as non-resolvable discrepancies.

4.2.3 Diseases of the Digestive System

There were two mismatches in this category, for which a non-preferable coding was used in the register.

4.2.4 Other Mismatches in the Coding

For one case in the category 'Infections and parasitic diseases' (ICD-9 '001–139'), a non-preferable coding was used during validation due to difficulties in translation of a tick-borne viral encephalitis, one case of genitourinary disease (ICD-9 '580–629') with non-preferable coding during validation, and three cases of injuries (ICD-9 '800–999') that were coded as external causes by the register but based on death certificate information (preferable) during the validation study.

5. Discussion

Overall, we observed a good agreement in this inter-institutional comparison of ICD-9 coding of causes of death, with 95% agreement on the ICD-9 main classes level (Kappa coefficient of 0.92) and 74% agreement on the ICD-9 specific category level (first three digits). Discussion of discrepancies showed that they were due to different reasons, namely use of non-preferable codes in the SUBI cause-of-death register, use of non-preferable codes during the validation process and because of some cases of non-resolvable coding discrepancies as both codes could have been used. Calculation of the validity of the SUBI cause-of-death register after removing non-preferable codes during validation and non-resolvable cases yielded validities of 98% on the main class level and 85% on the category level, with the latter increasing to 94% if some systematic non-preferable coding practices would be replaced by more appropriate ones.

There were a number of challenges when performing this comparison. First,
the original texts were in Russian, with the validation done by experts from Denmark, with English as the bridging language. This was a potential source of error, but clarification of discrepancies showed there was only little effect introduced by translation.

Second, the death certificates stem from a long time period of more than 60 years, with medical terminology and the practices of determination of causes of death changing over time. However, only one (0.4%) discrepancy was attributed to use of terminology that is not used anymore and was therefore difficult to code during the validation process. Time therefore might have played a minor role as the register was long-term, over a period of more than 60 years.

The SUBI cause-of-death register is a special register created for a specific task and defined population, compared to a population-based mortality register, and therefore has high quality in completeness and level of detail expected. No cancer case was missed by the restriction.

In the context of this validation study, it was also discussed whether and, if yes, how a conversion to ICD-10 should be done. Given the amount of work and the long time span covered by the register, it was not recommended to convert already coded cases, but to switch to using a ICD-10 in the near future and to develop procedures to combine ICD-9 and ICD-10 for the same classes of disease before running the analyses, as it is for instance done in the US Surveillance, Epidemiology and End Results (SEER) program [14]. However, the future use of ICD-10 would not solve discrepancies detected during the validation and those require some re-coding and specific solutions. Another option for discussion is the use of semi-automatic coding systems [15], but their possible advantages for such a specialized cause-of-death registers need more detailed evaluation.

From this validation study, we are able to make some project-specific and general recommendations. With regard to the SUBI cause-of-death register, some systematic coding practices need to be re-considered, and adaption to international coding rules would reduce the number of overall discrepancies by 25 (39% of all discrepancies). The three actions to be taken in the existing cause-of-death register are as follows. Firstly, to use ICD-9 code ‘414.9’ (“Other forms of chronic ischemic heart disease, unspecified”) more often as a general term, as the use of the specific category in these cases was not supported by the text on the death certificates. Secondly, to switch to using a ICD-10 in the near future and to develop procedures to combine ICD-9 and ICD-10 for the same classes of disease before running the analyses, as it is for instance done in the US Surveillance, Epidemiology and End Results (SEER) program [14]. However, the future use of ICD-10 would not solve discrepancies detected during the validation and those require some re-coding and specific solutions. Another option for discussion is the use of semi-automatic coding systems [15], but their possible advantages for such a specialized cause-of-death registers need more detailed evaluation.

From this validation study, we are able to make some project-specific and general recommendations. With regard to the SUBI cause-of-death register, some systematic coding practices need to be re-considered, and adaption to international coding rules would reduce the number of overall discrepancies by 25 (39% of all discrepancies). The three actions to be taken in the existing cause-of-death register are as follows. Firstly, to use ICD-9 code ‘414.9’ (“Other forms of chronic ischemic heart disease, unspecified”) more often as a general term, as the use of the specific category in these cases was not supported by the text on the death certificates. Secondly, within the main class of circulatory disease, the systematic preference of ICD-9 code ‘436’ (“Acute but ill-defined cerebrovascu-

### Table 4 Agreement between the death register and validation study by the specific ICD-9 three digit subcategories

<table>
<thead>
<tr>
<th>Major ICD-9 category</th>
<th>ICD-9 Codes</th>
<th>N (validation)</th>
<th>N (register)</th>
<th>N matching by 3 digits</th>
<th>Validity*, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Infectious and Parasitic Diseases</td>
<td>001–139</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>62.5</td>
</tr>
<tr>
<td>II. Neoplasms</td>
<td>140–239</td>
<td>100</td>
<td>105</td>
<td>93</td>
<td>88.6</td>
</tr>
<tr>
<td>III. Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders</td>
<td>240–279</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>70.0</td>
</tr>
<tr>
<td>IV. Diseases of the Blood and Blood-forming Organs</td>
<td>280–289</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>V. Mental Disorders</td>
<td>290–319</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100.0</td>
</tr>
<tr>
<td>VI. Diseases of the Nervous System and Sense Organs</td>
<td>320–359</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>VII. Diseases of the Circulatory System</td>
<td>390–459</td>
<td>101</td>
<td>97</td>
<td>59</td>
<td>60.8</td>
</tr>
<tr>
<td>VIII. Diseases of the Respiratory System</td>
<td>460–519</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>100.0</td>
</tr>
<tr>
<td>IX. Diseases of the Digestive System</td>
<td>520–579</td>
<td>19</td>
<td>16</td>
<td>14</td>
<td>87.5</td>
</tr>
<tr>
<td>X. Diseases of the Genitourinary System</td>
<td>580–629</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>50.0</td>
</tr>
<tr>
<td>XIII. Diseases of the Musculoskeletal System and Connective Tissue</td>
<td>710–739</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>XVI. Symptoms, Signs, and Ill-defined Conditions</td>
<td>780–799</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>XVII. Injury and Poisoning</td>
<td>800–999</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>001–999</td>
<td>246</td>
<td>246</td>
<td>182</td>
<td>74.0</td>
</tr>
</tbody>
</table>

*Validity was calculated within each class as [No. matching codes by 3 digits]/[No. codes in register] * 100%
lar disease”) ought to be avoided. Thirdly, for injuries, the direct cause of death should be used for the coding and not the external cause of death. Another systematic difference between the two coding groups was detected in the sub-category of haematological malignancies. Here, specifically for lymphoma, a well described weakness of ICD-9 [16] was circumvented by each center with their own coding policy, but these policies were different. Given that haematological malignancies are one of the major endpoints in radiation risk studies, we recommend using ICD-10 in addition to the original coding, as ICD-10 provides an improved and more appropriate classification of these diseases. However, to our mind, the validation was reassuring as the observed minor differences in cause-of-death coding in the SUBI register would not have any major impact on the results of epidemiological studies investigating cause-specific mortality among individuals in the background population of the SUBI register.

As a general note, we emphasize that when comparing results across countries, in which country-specific ICD-9 versions are used, one has to be aware of the differences for a correct interpretation of the results; however, this will refer mainly to sub-category level analyses.

Another more general recommendation is the re-consideration of using only one code for the classification of the cause of death, i.e., entry of the full information from medical certificates of death (main disease causing death, disease or concomitant states directly causing death, and intermediate causes) into the SUBI register database. For solving the issue of coding primary multiple cancers (ICD-9 code ‘199.0’) we recommend to provide additional information on all cancers. The majority of the main class level discrepancies in the inter-institutional comparison were due to the fact that the coding teams tried to incorporate multiple diseases into a single code. The use of the full information on causes of death is very important for further reduction of coding difficulties, although it is still preferable to define one primary cause of death.

6. Conclusions

In conclusion, this validation study was an important quality check that clearly demonstrated a high quality of the SUBI cause-of-death register used for mortality studies in the highly influential epidemiological investigations on radiation-related health effects in the Southern Urals. This is not only in the sense of identifying discrepancies, but also to strengthen the confidence in the results of the epidemiological studies, for which it is crucial that the validity of the register is high. The results of our inter-institutional comparison were generally favourable; however, since our comparison did reveal individual mismatches and some systematically differing coding practices, it is essential to repeat it on a regular basis in order to maintain the high quality.

Acknowledgements

This work is part of the SOLO project (Epidemiological Studies of Exposed Southern Urals Populations) funded by the Seventh Framework Programme of the European Atomic Energy Community (grant agreement number 249675). The authors like to thank Dr Ausrele Kesminiene (IARC) for her help in translating the causes of death from Russian to English with medical competence.

References