

SHORT REPORT

Feasibility of cohort studies in Estonia

Kaire Innos, Mati Rahu, Kaja Rahu

Abstract

Objectives—To describe the methodology and feasibility of follow up for vital status in retrospective cohort studies in Estonia. **Methods**—A cohort of 7412 workers who had been employed at two factories in Tallinn between 1946 and 1988 was followed up for vital status from the date of first employment until death, emigration, or the end of the study, 31 December 1995, whichever occurred first. The cohort was electronically linked with the National Population Registry of Estonia that was created in 1992 and includes personal identification numbers of Estonian citizens and residents, and the Mortality Database that contains information from death certificates issued in 1983–95. A manual search was carried out on several non-computerised population data sources and archives.

Results—By 31 December 1995, the vital status of 6780 (91.5%) subjects could be traced (4495 (60.6%) subjects were alive, 1993 (26.9%) had died, and 292 (3.9%) had emigrated). Analysis by calendar period of leaving work showed that the proportion of subjects traced was lowest in the group of workers who had left work between 1946 and 1955 (58.4%), especially those whose age at leaving work was <30 (53.2%) or >60 years (42.3%). Among subjects who left work in 1956–65, 1966–75, and 1976–88, the follow up rate was 84.7%, 94.6%, and 98.2%, respectively.

Conclusions—The findings, which are especially important for occupational epidemiology, confirm the feasibility of conducting retrospective cohort studies in Estonia. Most of the issues discussed in the paper apply to other former Soviet countries.

(Occup Environ Med 1999;56:499–502)

Keywords: Estonia; cohort studies; population registration

From 1940 to 1991, Estonia (area 45 216 km², population 1.5 million in 1997¹) belonged to the Soviet Union (USSR). During that period, vital statistics and migration registration data were collected, but computerisation was very limited. Also, the only registry that gathered health related information was the Estonian

Cancer Registry, which holds information on cancer cases diagnosed in Estonia since 1968.² In August 1991, Estonia regained independence. Shortly after that, in June 1992, the National Population Registry was established and personal identification numbers were adopted. However, retrospective cohort studies, which require the tracing of individual people through the Soviet period, must still use non-computerised data. To illustrate the methodology and feasibility of cohort studies in Estonia today, we report the follow up procedures of a recently conducted retrospective cohort study of furniture workers.

Subjects and methods**STUDY SUBJECTS**

The study cohort consisted of all workers employed between 1946 and 1988 at two furniture plants in Tallinn, the capital of Estonia. The information for the study population had been gathered for a previous study conducted in 1984–8.³ Demographic and employment information had been collected from company records and the vital status of the subjects had been checked at the Address Bureau of Estonia (see later) during the previous project.

The cohort members were followed up for vital status from the date of first employment until death, emigration, or the end of study date (31 December 1995), whichever occurred first.

DATA SOURCES

The National Population Registry contains personal data and information on the place of residence of Estonian residents and citizens. The registry was created in June 1992 and includes data on people who have been Estonian citizens or permanently resided in Estonia at any time since the creation of the registry. The registry assigns a personal identification number to each registered person. The data are continuously updated through data exchanges with other institutions where relevant events are registered.

The Mortality Database contains data on death certificates for all deaths in Estonia in 1983–95. For the period 1986–91, the data are without personal identifiers (names). The database was created at the Department of Epidemiology and Biostatistics, Institute of Experimental and Clinical Medicine, as a joint project with the Statistical Office of Estonia,⁴

Department of
Epidemiology and
Biostatistics, Institute
of Experimental and
Clinical Medicine,
Tallinn, Estonia
K Innos
M Rahu
K Rahu

Department of Public
Health, Tartu
University, Tartu,
Estonia
K Innos

Correspondence to:
Dr Kaire Innos,
Department of Epidemiology
and Biostatistics, Institute of
Experimental and Clinical
Medicine, Hiiu 42, 11619
Tallinn, Estonia.
Telephone 00372 2 514342;
fax 00372 6 706814;
email kaire@
ekmi.online.ee

Accepted 6 February 1999

Table 1 Results of follow up for vital status in the cohort by period of leaving work and by age at leaving work, 1946–88

	Period of leaving work*				All periods
	1946–55	1956–65	1966–75	1976–88	
Subjects (n)	539	1591	1912	3370	7412
Traced (%)	58.4	84.7	94.6	98.2	91.5
Alive	22.4	41.6	60.0	76.1	60.6
Dead	35.4	41.3	30.8	16.5	26.9
Emigrated	0.6	1.8	3.8	5.6	3.9
Untraced (%)	41.6	15.3	5.4	1.8	8.5
Traced individuals in age groups (%):					
<30**	53.2	82.9	92.1	97.6	88.0
30–39	64.0	83.2	94.5	98.8	90.6
40–49	70.3	90.6	96.4	98.1	94.7
50–59	72.0	90.4	97.3	98.7	96.3
≥60	42.3	80.7	97.1	97.7	90.5

*Subjects who were employed more than once were classified by the latest period.

**Age at leaving work.

by compiling the annual files from the Statistical Office and recoding all necessary variables. In the source files, the underlying cause of death was coded with the Soviet modification of the ninth revision of the international classification of diseases (ICD-9) for 1983–93,⁵ and the conventional ICD-9 for 1994–5; in the database, all deaths were recoded to the ICD-9 codes. The death certificates from 1996 to the present, in which cause of death is coded with the ICD-10, are kept in separate files at the Statistical Office. The original death certificates issued in Estonia since 1964 are stored in the archives of the Statistical Office and are available for manual search.

The Address Bureau of Estonia contains paper records on Estonian residents aged ≥16 since 1966.⁶ It was the main registry of population and vital statistics during the Soviet time. Currently, the Bureau is a division of the Office of Citizenship and Migration of Estonia, and constitutes the primary source for emigration data. The records of dead and emigrated people are kept in the Bureau for about 10 years. As the work is done manually, this rule is not always strictly followed. We, for example, conducted our search in 1996 and obtained emigration information for people who had emigrated in 1983.

In each town and county there are Civil Registration Offices where births, deaths, marriages, and name changes are recorded. A duplicate of every record is sent to the State Civil Registration Office. The central archive contains documents since 1966; local archives go further back—for example, the archive of the office in Tallinn has records starting from 1926. In recent years, some of the offices have started to computerise their records.

Some parish and town administrations keep records of people who have lived in their territory.

METHODS

The cohort was electronically linked with the National Population Registry and the Mortality Database. The records were matched by surname, first name, date of birth, and when available, father's name, place of residence, and previous surname.

All subjects were manually searched at the Address Bureau. Subjects not found in the

previous sources were manually searched in the archives of the Civil Registration Offices. In some cases, parish and town administrations for the person's last known address were contacted by post. The same variables were used for searching subjects manually as for electronic linkage.

Results

By 31 December 1995, among the 7412 subjects, 4495 (60.6%) were alive, 1993 (26.9%) had died, and 292 (3.9%) had emigrated. The vital status of 632 (8.5%) subjects was unknown. Among male and female workers, 7.1% and 10.2%, respectively, could not be traced.

The results of follow up for vital status were examined by calendar period of leaving work, and by age at leaving work (table 1). The proportion of successfully traced subjects increased steeply over time and of the 632 subjects lost, 468 (74.1%) had left work before 1966. For the workers employed after 1965, 96.9% were successfully traced.

The age of people at leaving work was markedly related to the success of tracing for the first period (1946–55) with the lowest follow up rates occurring in both marginal age groups. During the whole follow up period, the proportion of successfully traced subjects was lowest among workers who had left work at an age of <30 years.

Personal identification numbers were obtained from both the National Population Registry and the Mortality Database for 1992–5. In total, personal identification numbers could be found for 4648 cohort members—that is, for 92.5% of those who had been alive at or since the time when the numbers began to be assigned. For Estonians and non-Estonians, the respective percentages were 94.4% and 90.6%. Identification numbers were available for 61.3% of the subjects who had died or emigrated in 1992–3, 88.5% of the subjects who had died or emigrated in 1994–5, and 94.3% of those who were alive on 31 December 1995.

Discussion

FACTORS AFFECTING THE SUCCESS OF FOLLOW UP
The proportion of successfully traced subjects in our cohort increased over time and achieved comparatively satisfactory levels by the past two decades. The proportion of subjects lost to follow up was very high for the workers who had left the job in the decade after the end of the second world war, and high for the next decade ending in the mid-1960s. We were not surprised to see a low rate of follow up among older people, many of whom were presumably dead and whose death records might not be present or found in the archives. However, the low follow up rate in the youngest age group during early decades cannot be attributed to missing death records. Young women getting married and changing their surname could account for some of the losses to follow up and the slightly more successful tracing of men than women. However, this does not explain the failure to trace young men and the particularly low follow up rates of young subjects during the early decades.

After the second world war many people moved into Estonia from other parts of the Soviet Union.¹ A substantial proportion of immigrant workers did not settle permanently but emigrated after a few years. In our data, the information on emigration during this period was mostly limited to people who had reported emigration to be the reason for ending employment in company records. Local administrations provided this kind of information only for a few subjects. The Address Bureau keeps the records of emigrated people for about 10 years. As a result of a previous update of the vital status information of the study subjects in 1984–8, the emigration data for the cohort were available from the mid-1970s.³ There is no database or archive available where emigration data for early decades after the war can be found. We think that the many cohort members who were lost to follow up during these years had left the country soon after leaving work. There are two arguments supporting the suggestion. Firstly, the general population data indicate that younger people are more prone to change their place and region of residence; about half of the subjects who disappeared during 1946–65 had been <30 years old at the time of leaving work. Secondly, >80% of the young subjects, and >60% of all subjects lost to follow up during this period were not Estonians and were born outside Estonia. It is not unreasonable to assume that they returned to their homeland after leaving the job.

The accuracy of the names is very important in tracing people. One of the main difficulties arises from the lack of consistency in the transliteration between the Cyrillic and Latin alphabets. A change of surname is another potential problem. The usual practice in Estonia at marriage is for the woman to take her husband's surname. However, some women keep their maiden name or take both names. In rare cases, the couple takes the woman's name. In electronic linkage, these problems can be overcome by using other identifiers, such as personal identification number, first name, father's name (included in older databases) and date of birth. Also, the Population Registry keeps the previous surnames of the person in the record, which facilitates the identification of correct people. In manual searches, both inaccurate and changed names are likely to cause the records to be missed. Fortunately, the records of the Address Bureau enable the finding of people who have changed their surname up to about 10 years after the event.

EFFECT OF LOSS TO FOLLOW UP ON RESULTS IN COHORT STUDIES

Many subjects lost to follow up can raise serious doubts about the validity of a cohort study.⁷ A major bias would occur if the failure to trace the subjects was associated with exposure or disease, similarly non-participation or non-response in other research designs might introduce bias.⁸ In our study, however, there is no indication of such an association and we do not expect loss to follow up to significantly distort the results of further analysis of exposure effects.

PRESENT DATA SOURCES

The National Population Registry has not achieved 100% coverage yet. The Estonian government is currently discussing a draft of a proposed law under which the registration practices of residents will be settled. The absence of such legislation has considerably impeded the work of the Population Registry.

A high quality Mortality Database containing death certificates for 1983–95 is available at both the Statistical Office, and the Department of Epidemiology and Biostatistics (IECM). In this database, the death certificates covering the period 1986–91 are without personal identifiers. The original documents are stored in the archive. Manual searches can be done successfully, as has been shown in previous studies,⁹ although it takes a lot of time and effort. Attempts are currently being undertaken to compile a complete death certificate database covering the period from 1983 to the present.

Several other computerised data sources exist in Estonia that contain individual identifiable data and can be used for research purposes. The Medical Birth Registry collects data on all live births and stillbirths in Estonia since 1992.¹⁰ The Abortion Registry gathers information on spontaneous and artificial abortions since 1994. The Estonian Cancer Registry has data on all incident cancer cases in Estonia since 1968.² The registration of cases of tuberculosis, blindness, and visual impairment also deserves mentioning. It is mandatory for all medical specialists to send relevant data to these registries.

CONFIDENTIALITY

Currently, Estonia is developing legal safeguards to protect people's privacy. The Personal Data Protection Act (passed on 12 June 1996) and the Databases Act (passed on 12 March 1997) are the primary laws providing guidance for protecting individual privacy and managing various databases. Government Regulation No 253 (23 December 1997) sets the procedures for releasing personal data for research purposes.

Unfortunately, the process of introducing a new information policy has resulted in some backlash: the activities of all of the medical registries in Estonia were stopped between 6 and 26 November 1998. The shutdown was triggered by unproductive talks between the Ministry of Social Affairs and the Department of Data Protection. This created great concern among Estonian and international scholars, many of whom sent protest letters to the President, Prime Minister, and Minister of Social Affairs.

Thus, non-legislative solutions to privacy of information are being gradually replaced by legislative ones to encourage the development of adequate security standards to preserve confidentiality.

Conclusions

The overall follow up rate found in the study compares quite well with some studies from other European or American countries. Our findings indicate that it is feasible to conduct retrospective follow up of study subjects in

Estonia, although the tracing through past decades requires considerable efforts. These conclusions are especially relevant to occupational epidemiology. The expanding use of personal identification numbers in different social areas is encouraging for future scientific activities. Certainly, most of the issues discussed with regard to Estonia apply similarly to other former Soviet countries.

We thank Edward A Gelb for editing the manuscript.

- 1 Ainsaar M. *Eesti rahvastik Täani hindamisraamatust tänapäevani. Estonian population from Liber Census Daniae up to nowadays*. Tartu: Tartu Ülikooli Kirjastus, 1997.
- 2 Rahu M. Estonia. In: Parkin DM, Whelan SL, Ferlay J, et al, eds. *Cancer incidence in five continents*. Vol VII. IARC Scientific Publications No. 143. Lyon: IARC, 1997:466–469.
- 3 Avdejeva IA. Incidence of malignant neoplasms of respiratory organs in woodworking industry. In: Jelovskaja LT, Oziganova VN, eds. *Prevention of dust induced occupational*

diseases. Proceedings of the Institute of Occupational Hygiene and Occupational Diseases of the Academy of Medical Sciences of the USSR. Vol 44. Moscow: The Academy of Medical Sciences of the USSR, 1991:88–93.

- 4 Baburin A, Gornoi K, Leinsalu M, et al. *Eesti suremusatlas. Atlas of mortality in Estonia*. Tallinn: Institute of Experimental and Clinical Medicine, 1997.
- 5 Leinsalu M. Time trends in cause-specific mortality in Estonia from 1965 to 1989. *Int J Epidemiol* 1995;24:106–113.
- 6 Tekkel M, Rahu M, Veidebaum T, et al. The Estonian study of Chernobyl cleanup workers: I. Design and questionnaire data. *Radiat Res* 1997;147:641–52.
- 7 Breslow NE, Day NE. *Statistical methods in cancer research. The design and analysis of cohort studies*. Vol II. Lyon: International Agency for Research on Cancer, 1987. (IARC Sci Publ No 82.)
- 8 Gordis L. *Epidemiology*. Philadelphia: WB Saunders, 1996.
- 9 Rahu M, Tekkel M, Veidebaum T, et al. The Estonian study of Chernobyl cleanup workers: II. Incidence of cancer and mortality. *Radiat Res* 1997;147:653–7.
- 10 Karro H, Rahu M, Gornoi K, et al. Estonian medical birth registry 1992–4: association of risk factors with perinatal mortality. *Eur J Obstet Gynecol Reprod Biol* 1998;80:151–6.

Vancouver style

All manuscripts submitted to *Occup Environ Med* should conform to the uniform requirements for manuscripts submitted to biomedical journals (known as the Vancouver style.)

Occup Environ Med, together with many other international biomedical journals, has agreed to accept articles prepared in accordance with the Vancouver style. The style (described in full in the *JAMA*[1]) is intended to standardise requirements for authors, and is the same as in this issue.

References should be numbered consecutively in the order in which they are first mentioned in the text by Arabic numerals on the line in square brackets on each occasion the reference is cited (Manson[1] confirmed other reports[2][3][4][5]). In future references to papers submitted to *Occup Environ Med* should include: the names of all

authors if there are three or less or, if there are more, the first three followed by *et al*; the title of journal articles or book chapters; the titles of journals abbreviated according to the style of *Index Medicus*; and the first and final page numbers of the article or chapter. Titles not in *Index Medicus* should be given in full.

Examples of common forms of references are:

- 1 International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomed journals. *JAMA* 1993;269:2282–6.
- 2 Soter NA, Wasserman SI, Austen KF. Cold urticaria: release into the circulation of histamine and eosinophil chemotactic factor of anaphylaxis during cold challenge. *N Engl J Med* 1976;294:687–90.
- 3 Weinstein L, Swartz MN. Pathogenic properties of invading micro-organisms. In: Sodeman WA Jr, Sodeman WA, eds. *Pathologic physiology, mechanisms of disease*. Philadelphia: W B Saunders, 1974:457–72.