A Method for Identifying Underlying Causes of Death in Epidemiological Study

Tamiko Iwasaki [1], Toshio Miyake [1], Sumio Ohshima [1], Shin' ichi Kudo [1], and Takesumi Yoshimura [2]

To obtain the underlying causes of death in individuals of a cohort (male only), the items of date of birth, date of death and address code at the time of death were linked between the data of resident cards and the death tapes of National Vital Statistics supplied by the Ministry of Health and Welfare. As a result, the persons who have the same information for above mentioned three items between the resident cards and the death tapes accounted for 97.4%. If the persons who had the same information for three items except one item were considered to be identical, they accounted for 99.4%. It would be concluded that underlying cause of death can be obtained by record linkage of death tapes of National Vital Statistics with three informations in residence cards, dates of birth, dates of death and address codes at the time of death, even without names of the individual in the death tapes. The matched rate would be high enough for epidemiological studies. *J Epidemiol*, 2000 ; 10 : 362-365

underlying causes of death, record linkage, National Vital Statistics, resident cards

INTRODUCTION

Since 1990, the Institute of Radiation Epidemiology of the Radiation Effects Association has been studying the relationship between exposure to radiation and causes of death in people who were engaged in radiation work at nuclear facilities. The study was entrusted to the Institute by the Science and Technology Agency to study the health effects of exposure to low-dose and low-dose rate radiation on human [1]. As a study method, the copies of resident cards of the subjects under investigation were obtained to confirm whether they were alive or dead. For the persons who were dead, their underlying causes of death were identified by record linkage with the death tapes of National Vital Statistics (hereinafter referred to as "death tapes") supplied by the Ministry of Health and Welfare (M.H.W.) of Japan. In these death tapes, however, there were no descriptions of the names of the deceased. Therefore, three items of date of birth, date of death and address at the time of death were used for identification of individuals. Since the subjects under investigation were limited to males, only the death tapes of males were obtained and used. The method for identifying underlying causes of death by mean of record linkage of resident cards and the death tapes was performed.

This paper reports the results of record linkage of the death information from the resident cards and the death tapes supplied by the M.H.W., discusses the problems of this method, and proposes this method as a useful procedure for epidemiological studies in Japan.

METHOD

The copies of the resident cards of the deceased contain the names, dates of birth, addresses at the time of death, and dates of death of the deceased, without information of the causes of death. To obtain the underlying causes of death, therefore, the death information in resident cards of the deceased were linked with the death tapes supplied by the M.H.W. For using the death tapes from the M.H.W., the approval of the Director General of the Management and Coordination Agency was

Received October 29, 1999 ; accepted May 15, 2000.

[1] Institute of Radiation Epidemiology, Radiation Effects Association.
[2] Dept. of Clinical Epidemiology, Inst. of Industrial Ecological Sciences, Univ. of Occupational and Environmental Health, Japan.
Address for correspondence : Tamiko Iwasaki, Institute of Radiation Epidemiology, Radiation Effects Association, 1-9-16, Kajicho, Chiyoda-ku, Tokyo, 101-0044 Japan.

obtained in accordance with the Notification No. 10 dated February 9, 1994 of the Agency under the Law of Statistics. The death tapes contain the information about dates of birth, dates of death, address codes at the time of death. In the present paper, it was adopted underlying causes of death coded by the ICD 9 (140-239) for about 3,200,000 of all the male deaths aged 18 or older from 1986 to 1992. Among radiation workers investigated, 1,758 deceased cases were confirmed by resident cards during the same periods. The identification work was performed by record linkage of 3 items of the dates of birth, dates of death and address codes at the time of death between the information of resident cards and that of the death tapes.

Identification method

1. The deceased who have the same information about the following three items as that of resident cards were retrieved from the death tapes:
   (1) Date of birth (seven figures including the name of an era)
   (2) Date of death (seven figures including the name of an era)
   (3) Address code at the time of death (five figures in each code assigned by the Ministry of Home Affairs)

The deceased who had the same information about the above three items in the resident cards and magnetic tapes were regarded as identical.

2. The deceased who did not have the same information about these three items were subjected to record linkage again as follows:
   (1) The persons who have the same information about the date of death and address code at the time of death, except the date of birth.
   (2) The persons who have the same information about the date of birth and address code at the time of death, except the date of death.
   (3) The persons who have the same information about the date of birth and date of death, except the address code at the time of death.

All the cases which did not completely match with each other in any of the above three items were retrieved as "a similar list." Among these retrieved cases, the persons who came under the following conditions were considered to be identical:

For example, in (1) above, the persons with their dates of death and address codes at the time of death matching with each other were retrieved, and then the persons with their dates of birth most similar to each other (e.g., February 11 and February 17) were considered to be identical as shown in Table 1. In (2) above, the persons with their dates of birth and address codes at the time of death matching with each other retrieved likewise, and then the persons with their dates of death most similar to each other (e.g., March 10 and March 9) were considered to be identical as shown in Table 2. In (3) above, the persons with their dates of birth and dates of death matching with each other and only the address codes being different were retrieved likewise. Among those retrieved persons, those with the identical dates of birth and dates of death in the address codes of neighboring towns and villages (e.g., 14101 and 14102) were considered to be identical. In this record linkage, a COBOL program was used.

RESULTS

The persons whose dates of birth, dates of death and address codes at the time of death completely matched with each other were 1,713 (97.4%) of the 1,758 surveyed (Table 3). The per-

Table 1. Cases identified by changing of date of birth.

Case	Date from resident cards	Date from death tapes
1	1916. 02. 11	1916. 02. 17
2	1918. 09. 04	1918. 09. 07
3	1921. 01. 15	1921. 01. 15
4	1922. 08. 15	1922. 08. 16
5	1927. 10. 12	1927. 02. 12
6	1929. 06. 15	1929. 06. 16
7	1931. 08. 26	1931. 08. 20
8	1932. 12. 20	1932. 02. 20
9	1935. 12. 16*	1921. 12. 16*
10	1938. 10. 19	1938. 10. 18
11	1940. 01. 05	1940. 01. 15
12	1941. 09. 04	1940. 09. 14
13	1953. 10. 23	1953. 12. 23
14	1965. 02. 25	1965. 02. 26

* 1935: Showa 10, 1921: Taisho 10 (in the name of an era)

Table 2. Cases identified by changing of date of death.

Case	Date from resident cards	Date from death tapes
1	1988. 03. 00*	1988. 03. 26
2	1988. 05. 00*	1988. 05. 10
3	1988. 07. 31	1988. 08. 01
4	1989. 03. 10	1989. 03. 09
5	1989. 06. 18	1989. 06. 19
6	1989. 08. 00	1989. 08. 05
7	1989. 10. 16	1989. 10. 11
8	1990. 04. 00*	1990. 04. 15
9	1991. 01. 00*	1991. 01. 22
10	1991. 01. 13	1991. 01. 14
11	1991. 01. 27	1991. 01. 28
12	1991. 03. 00*	1991. 03. 26
13	1991. 07. 31	1991. 07. 30
14	1991. 11. 21	1991. 11. 25
15	1992. 04. 25	1992. 04. 24
16	1992. 09. 00*	1992. 09. 01

* 00: Unspecified

sons who had the same information about these three items were judged to be identical and given their underlying causes of death. For 45 persons who did not have the same information about these three items, the persons who have the same information for two of the three items as that in the resident cards were retrieved from the death tapes. Then, record linkage was tried again for the persons who have the same information for two of the three items, in the same way as for the identification method.

As shown in Table 3, the persons whose dates of death and address codes at the time of death matched with each other were 14 (0.8%), the persons whose dates of birth and address codes at the time of death matched with each other were 16 (0.9%), and the persons whose dates of birth and dates of death matched with each other were 5 (0.3%). These persons could be judged to be identical. On the other hand, the persons who were completely different from each other in this record linkage and could not be identified were 10 (0.6%). In other words, the underlying causes of death of 1,748 (99.4%) out of the 1,758 deaths were finally obtained by this method.

There was no duplication of deaths in this record linkage method.

DISCUSSION

Epidemiological studies have been carried out with people engaged in radiation work in order to know the effects of exposure to low-dose and low-dose rate radiation on the causes of their death, including cancers [1]. The resident cards of subjects of investigation were obtained to confirm whether they were alive or dead. However, these resident cards did not contain the causes of death. To know the causes of death, a method was used in which the underlying causes of death were confirmed by record linkage between the information of the resident cards and that of the death tapes supplied by the M.H.W. The method of the records linkage is performed in cancer registry. It is also used in measuring the precision of cancer examination and in various epidemiological studies [2,3]. That is, this enables confirmation of the deceased due to cancers, shown in new data collected from the regional registration of cancer, to be the identical persons shown in old data through record linkage of these data. In these records, the names are mentioned, but there is no description of the names in the death tapes using in this study. Then, the persons whose dates of birth, dates of death and address codes at the time of death shown in the resident cards completely matched with those recorded on the death tapes were considered to be identical, and the underlying causes of death recorded on the death tapes were considered to be those of the deceased. As a result, this method of linkage the death records in resident cards with those of the death tapes the persons who have the same information for above mentioned three items between the resident cards and the death tapes accounted for 97.4%. As seen in Table 3, the results show that the percentage of the deceased with the dates of birth, dates of death and address codes at the time of death matching with each other is as high as 97.4%. The same identifying method was used in the follow-up study of circulatory disorders by Ueshima [4]. He also reported that matched rate was obtained 97.4% as same as in the present results.

When persons who have the same information for two of the three items are included, the rate of concordance becomes as

Table 3. Results of record linkage for underlying causes of death (Subjects: 1,758).

Subjects	Items of record linkage: Date of birth	Date of death	Address code at time of death	Result (No. of person)	Rate (%)
1,758	completely matched	completely matched	completely matched	1,713	97.4
45	(1) partially matched*	completely matched	completely matched	14	0.8
	(2) completely matched	partially matched*	completely matched	16	0.9
	(3) completely matched	completely matched	partially matched*	5	0.3
Total of matched (include partially matched)				1,748	99.4
Matched impossible				10	0.6

* See text.

high as 99.4%. It is considered that the reason for such a high correspondence rate is due to our study using the death record in resident cards and magnetic tapes, which are both the official records.

As a method similar to the above, the records of birth registration from National Vital Statistics are linked with those of death registry of children aged one or less after birth [5]. In this case, the records of birth registration from the National Vital Statistics were manually linked with those of death registry of children aged one or less after birth in the 1989's birth cohort of 409,679 in 18 prefectures, and it was possible to identify the dead at 96% (1,840/1,909) from the record linkage of the birth registration and the death registry.

Among Japanese males who died between 1986 and 1992 (4,581,654) the probability of persons whose dates of birth, dates of death and address codes at the time of death match with those of different persons by chance (2,839) are as low as 6/10,000 as a result of a study. The persons who have the same information for these three items, therefore, are probably identical.

This, however, only means the match of linkage between two records. It is impossible to confirm that they are definitely identical. Ohtaki et al. [6] have done a theoretical analysis of record linkage for related documents of atomic bomb survivors. According to their report, for example, the probability that in a paired sample of different persons the family name and present address of a person match with those of the counterpart is 0.52×10^{-5}, and the probability that in a paired sample of different persons the family name and present address of a person do not match with those of the counterpart is 0.997. As shown in the above, the probability that the matched data of two persons are identical and they are actually different persons is very low. Therefore, we think this method shows high performance in record linkage.

As a measure for verifying the identification of the deceased in this record linkage method, it will be necessary to confirm that the deceased is truely identical by perusing the death registration form (kept only five years) at public health centers or death certificates kept at the Legal Affairs Bureaus, Ministry of Justice.

ACKNOWLEDGEMENT

The authors are grateful to Dr. Toshiyuki Kumatori and Mr. Masato Kaneko for their great effort and advice .

REFERENCES

1. Epidemiological Study Group of Nuclear Workers (Japan), First analysis of mortality of nuclear industry workers in Japan, 1986-1992. J. Health Physics, 32, 173-184 (1997).
2. Fujimoto I, Hanai A, Tsukuma H, Hiyama T. Role of population-based cancer registry in cancer epidemiology-Epidemiological studies in the cancer registration in scheme in Osake, Japan-. Jpn. J. Hyg., 1994; 49: 543-558. (in Japanese with Summary in English)
3. Okamoto N. Natsui S., from Report on The Research Group for Population- based Cancer Registries in Japan, 1993. (in Japanese)
4. Ueshima H. Follow-up study of 1980 national survey of circulatory disorders (NIPPON DATA). J. Japan. Assoc.Cerebro-Cardiovascular Disease Control (JACD), 1997; 31:231-237. (in Japanese)
5. Fujita T, Minowa M, Miura Y, Kamiya K. Risk factors for neonatal and postneonatal mortality - A record-linkage study based on vital statistics-.Nippon Koshu Eisei Zasshi, 1994; 41: 34-45. (in Japanese with Summary in English)
6. Ohtaki M, Munaka M, Kurihara M, Hayakawa N, Yamamoto H, Ueoka H, Sumida H, Hiraoka M. A mathematical theory of collation for record linkage among the data related to A-bomb survivors. J.Hiroshima Med. Ass.,1982; 35: 356-358. (in Japanese)