HIROSAKI UNIVERSITY

Radiation Emergency Medicine 2014 Vol.3, No.2 1–12

REM
Radiation Emergency Medicine

Special Contribution

Basic Epidemiology
—Methods and Their Application to Epidemiology on Cancer and Radiation (2)

Suminori Akiba

Department of Epidemiology and Preventive Medicine, Kagoshima University Graduate School of Medical and Dental Sciences
8-35-1, Sakuragaoka, Kagoshima 890-8544, Japan

Received 12 February 2014; revised 30 April 2014; accepted 10 June 2014

III. Cohort study

1. A simple example

A typical cohort study compares incidence rates among groups of people with different exposure levels. RRs, ERRs and/or RDs are usually calculated to compare risks in different exposure groups.

2. The definition of cohort and dynamic population

A cohort is a fixed population. Its membership is permanent (even after death)[8, 14]. Examples are as follows: people born in Kagoshima, and the alumni of Hirosaki University.

A cohort study follows a fixed population. Unfixed populations are called dynamic populations (open populations). The membership of a dynamic cohort is not permanent. Examples are as follows:

Hirosaki University students – The students of Hirosaki University are members of a dynamic population, which can be called Hirosaki University Student Population. When students graduate from Hirosaki University, they lose the membership.

Kagoshima City residents. – People who live in Kagoshima City constitute a dynamic population, which can be called as Kagoshima City Resident Population. The membership is lost when residents migrate out from Kagoshima City and start living in Fukuoka City, for example.

3. Advantages of cohort studies

The cohort study has the following major advantages:
1) its study design is easier to understand and similar to animal experiments;
2) multiple endpoints can be examined easily; and
3) it can be conducted even if exposure is rare.
Those points will be discussed again in the section of case-controls studies.

4. Competing risks

Competing risks are events that "compete" with the outcome of interest since such events remove persons from the population at risk (ME-III)[15]. For example, in a study examining the relationship between smoking and lung cancer, a subject who dies of coronary heart disease is no longer at risk of lung cancer. In such a study, deaths from coronary heart diseases are regarded as competing risk (DE-4)[2].

5. Loss to follow-up

Study subjects may move away during follow-up, and leave behind no means of further contact. They are losses to follow-up. Competing risk also prevents the follow-up of study subjects (and therefore, it is a cause of loss to follow-up).

Suminori Akiba: Department of Epidemiology and Preventive Medicine, Kagoshima University Graduate School of Medical and Dental Sciences
8-35-1, Sakuragaoka, Kagoshima 890-8544, Japan
E-mail: akiba@m.kufm.kagoshima-u.ac.jp

6. Examples of cohort studies

1) The Life Span Study cohort: a cohort of atomic bomb survivors

The Life Span Study (LSS) cohort consists of about 94,000 survivors of the atomic bombings in Hiroshima and Nagasaki, Japan, in 1945, who were identified by the supplementary schedules to the national census in 1950[16, 17]. The LSS cohort includes 44,000 men and women with all ages who were within 2.5 km of the hypocenters at the time of the bombings and still resided in Hiroshima or Nagasaki in 1950. Among them, 27,800 survivors were exposed within 2.0 km from the hypocenter. It also includes a randomly selected sample of survivors (n = 27,800) who were exposed in 2.5-10 km distance from the hypocenter. This group was exposed to small to negligible amounts of radiation doses. The control group of this cohort consists of 27,800 subjects who were in neither city at the time of bombing[17]. The second group (the 2.5-10 km group) and the control group were so selected that they have sex and age distributions similar to the first group (the ≤ 2.0 km group). The cohort members have been followed, using various methods to examine their mortality, cancer incidence and so on. Since 1958, Adult Health Study (AHS) has conducted biennial medical examinations of the study subjects who were selected from among the LSS cohort.

2) The Karunagappally cohort study

The coastal belt of Karunagappally in Kerala, India is known for high background natural radiation (HBNR) from thorium-containing monazite sand. A cohort of all 385,103 residents in Karunagappally Taluk (Taluk is an administrative unit) was established in the 1990s to evaluate the health effects of HBNR. All the households (n = 71,674) in Karunagappally were visited by 12 to 14 interviewers, starting from 1 January 1990. They collected information on socio-demographic factors, lifestyles and other factors, using a 6-page standardized questionnaire. In addition, the investigators conducted dosimetry surveys using portable scintillation survey meters to measure indoor and outdoor radiation levels at each house. It took 8 years to complete the household visits in the entire taluk. Cancer cases were ascertained by the regional cancer registry of Karunagappally[18, 19].

3) The Techa River study

The Mayak Production Association (PA) is the first and largest nuclear weapons facility in the former Soviet Union, consisting of nuclear reactors and radiochemical and plutonium plants. This plant started its operations in 1948. During the period between 1949 and 1956, a large amount of radionuclides were discharged into the Techa River, which runs through areas east to the Ural Mountains in Central Russia. Metlino is the most highly contaminated settlement, which is 7 km downstream from reservoir R-4, into which nuclear waste was released. According to the report by Krestinina *et al*.,[20] 18,389 people lived in one of the 25 villages in Chelyabinsk Oblast (Oblast is an administrative unit), which is located 7–148km from the release point, and 11,411 people lived in one of the 16 riverside villages in Kurgan Oblast (155–237km from the release point). To examine the health effects of radiation exposure, investigators established a cohort of residents (the Extended Techa River Cohort).

4) The Semipalatinsk study

The Semipalatinsk Test Site (STS) is located in the steppe in northeast Kazakhstan (then the Kazakh Soviet Socialist Republic), south of the valley of the Irtysh River. The Soviet Union conducted 456 nuclear tests at the STS during the period between 1949 and 1989. German and Kazakh scientists conducted a retrospective study to examine cancer mortality during the period between 1960 and 1999, establishing a cohort of 19,545 inhabitants in exposed and control areas[21].

5) Residents in ^{60}Co contaminated buildings in Taiwan

A steel reprocessing factory in Taiwan mishandled a ^{60}Co orphan source, and, as a result, produced radioactive steel rods (rebars) during the period 1982-1983[22]. This problem was not realized until 1992, when a building with radioactive building materials was found in Taipei City. By the end of 2000, radioactive rebars were found in 1,607 housing units in 181 buildings in northern Taiwan[23]. All the buildings with radiation contamination were built during the period between 1982 and 1984. Many families lived in those buildings for various time intervals, mostly around 10 years. Various epidemiological studies have been conducted to examine the mortality and morbidity of residents. Hwang *et al*. examined the cancer risk among the cohort of 7,271 people with the estimated radiation dose which was calculated by the Taiwan Cumulative Dose system[24]. This system used information obtained from a questionnaire survey on house occupancy, building structure and others. In risk analysis, the start of the follow-up was the date when they moved into contaminated buildings, and the end of follow-up was the date of death, the date of cancer diagnosis, or the end of follow-up, whichever came first. A total of 101,560 person-years at risk were accumulated during the follow-up period from 1983 to 2002 (on average 16.1 years).

6) Mayak Production Association workers

A cohort study was established to examine the health risk of nuclear workers of the Mayak PA[25]. The study cohort includes approximately 12,000 male and female workers first employed by one of the main Mayak plants from the start of operations (1948) until the end of 1958. Excluded from the cohort were the workers who were involved in incidents/accidents and developed acute radiation syndrome, and the workers exposed to radionuclides other than ^{239}Pu such as ^{3}H. The mean age of starting

work at the plant was 25 years. The mean durations of work were 14 years (standard deviation: SD = 13) at the reactors, 11 years (SD = 13) at the radiochemical plant and 14 years (SD = 13) at plutonium plants.

The Mayak PA is located near the city of Ozyorsk in the Southern Urals. Less than a half of the cohort (46%) were living in Ozyorsk at the end of 2000 or had lived there until their death, and approximately 53% of workers had migrated from the city during follow-up. The follow-up was virtually complete as long as study subjects lived in Ozyorsk (vital status was unknown for only 0.03%) whereas the vital status was unknown for 11% of the entire cohort, indicating that the follow-up was not easy once cohort members left the city. Morbidity data were available as long as the cohort members lived in Ozyorsk. Up to 31 December 2000, morbidity data were collected from 94% of the entire cohort.

7) Chernobyl emergency workers

The Chernobyl nuclear power plant is located in Ukraine, and is approximately 10 km south of the border with Belarus. On April 26, 1986, a serious nuclear accident took place in this plant. Victor K Ivanov and his colleagues conducted the follow-up studies of emergency workers, who are called liquidators[26, 27]. The number of liquidators was 77,663 in 1986, 58,694 in 1987 and 31,565 during the period 1988-1990, giving a total of 167,862[10]. In addition to mortality studies, they conducted morbidity studies. Liquidators underwent annual medical examinations in state health institutions, following the government decree. The Russian National Medical and Dosimetric Registry keeps the information obtained from 700,000 workers. Workers eligible for their morbidity study were those whose information on health status was collected at least once during the study period. If study subjects were found to have the disease of interest prior to entry to the zone, they were excluded from analysis.

7. Retrospective cohort study

A retrospective cohort study is a study in which the establishment of a cohort, the ascertainment of cases, and the collection of information on exposure and covariates are conducted in a retrospective manner. However, it should be noted that, even in prospective cohort studies, we collect information on exposure variables and covariables in a retrospective manner (as is the case with case-control studies, which will be described later). Therefore, the retrospective cohort study is retrospective because its case ascertainment is retrospective. Likewise, the prospective cohort study is prospective because its case ascertainment is prospective.

Retrospective cohort studies can be completed in a short time period; therefore, generally speaking, they are less time-consuming and less costly than prospective cohort studies. In the retrospective study, the accuracy of exposure and/or completeness of information on exposure may be affected by the outcome of interest. The completeness of case ascertainment may also be affected by the exposure status of cohort members. In the mortality follow-up of Japanese nuclear workers, a retrospective survey was conducted to identify their vital status[28]. In this study, nuclear workers with lower cumulative doses were more difficult to locate since they tended to have shorter periods of employment. In the LSS study of atomic bomb survivors, the establishment of cohort was completed in the late 1950s. Therefore, mortality data in the early 1950s were collected in a retrospective manner. For the periods 1950-52, 1950-55, and 1950-60, the ERRs/Sv of solid cancer mortality were 0.41 (90% CI = -0.01, 0.99), 0.38 (90% CI = 0.07, 0.75), and 0.24 (90% CI = 0.05, 0.45), respectively (UNSCEAR 2006)[29]. The estimates appear to be larger in early follow-up years. A selection bias might have been involved in the retrospective follow-up of cohort.

IV. Case-control study

1. Introduction

In 1939, Franz H. Müller reported the results of his case-control study of lung cancer. The prevalence of lung cancer was higher among smokers[30]. This study is the first epidemiological study elucidating the association of lung cancer risk with smoking. In a typical case-control study, cases and controls are interviewed to investigate their exposure status. Generally, possible approaches/sources of information on exposure and covariables are as follows:

existing records,
self-administered questionnaires or mail questionnaires,
face-to-face interview or telephone interview,
computer-assisted interview,
biological specimens,
tissue banks, and
databases on biochemical and environmental measurements.

If cases are from a certain population and controls are randomly selected from the same population, it is called a population-based case-control study. If the study population is a cohort, the study is called a case-control study nested in a cohort. DE-4 defines the case-control study nested in a cohort as follows: "A case-control study in which cases and controls are drawn from the population in a cohort study. As some data are already available about both cases and controls, the effects of some potential confounding variables are reduced or eliminated. In this type of case-control study, a set of controls is selected from subjects, i.e., at risk at the

time of occurrence of each case that arises in a cohort, thus allowing for the confounding effect of time in the analysis."

Controls can be selected from other sources. For example, hospital controls are those "drawn from the population of patients in a hospital" (DE-4)[2]. When hospital controls are used, controls with diseases (diagnoses) associated with the exposure of interest should be avoided. Another example is to use neighborhood controls, which are "persons live in the same locality as cases and therefore may resemble cases in environmental and socioeconomic criteria." (DE-4)[2] This kind of controls are expected to have socio-economic factors similar to those of cases. It should be noted, however, that neighbors may have an artificially similar exposure distribution. For example, they may be exposed similarly to air pollution, chemicals and asbestos from the nearby factories, and to electromagnetic fields from power lines in the neighborhood.

2. Odds ratio (OR)

In case-control studies, odds ratios (ORs) are usually calculated. In Table 2, you can obtain an OR of 81 (= (90/10)/(10/90)) for the relationship between risk (caseness) and exposure.

Table 2. Results of a case-control study

	cases	controls
exposed	90	10
unexposed	10	90

3. Examples

1) Childhood leukemia risk among children of Sellafield workers

The risks for leukemia and non-Hodgkin's lymphoma were increased among children born in Seascale village and its neighboring areas, which are near a plutonium reprocessing plant, the Sellafield plant. Martin Garner and his colleagues conducted a case-control study of leukemia (n = 52) and lymphoma (n = 45) diagnosed under the age of 25 among residents in West Cumbria health district during the period 1950-85. They found that the risk of hematopoietic malignancies was increased among children of fathers employed at the plant, particularly, among those with high radiation dose recordings before their child's conception. Based on those results, Martin Gardner postulated that preconceptual radiation exposure of fathers increased leukemia and lymphoma risk among their children[31]. A counter argument was made by Sir Richard Doll and his colleagues. They made the following point: if the Gardner hypothesis is correct, an excess risk of more than 50 cases should have also been observed in the rest of Cumbria outside Seascale village since a large number of the births to Sellafield employees occurred outside Seascale, and their total collective dose of paternal irradiation was 501 man-Sv, a much larger figure than 38 man-Sv for the fathers in Seascale[32]. However, such an argument does not explain the observed excess of leukemia risk among children of male Sellafield plant workers.

The report of Health and Safety Executive (HSE), entitled "HSE investigation of leukemia and other cancers in the children of male workers at Sellafield", describes the results of the case-control study conducted as a part of HSE investigations[33]. This study identified all cases of leukemia and non-Hodgkin lymphoma among children in Allerdale and Copeland districts of Cumbria during the period between 1 January 1950 and 31 May 1990, using the National Health Service Central Register (NHSCR). This district includes the Sellafield plant and Seascale village. Controls were randomly selected from the same birth registers and in the same time period as the case children. Interestingly, this study found a strong association of hematopoietic malignancy risk with tritium exposure. "There is a highly significant positive trend (P = 0.0018) in risk for the three group analysis" (of tritium exposure levels). Those findings are summarized in Table A-20 of the HSE report. Since tritium is unlikely to be the main cause of the observed association, it is suspected to be a surrogate marker of some unknown factor.

2) Childhood leukemia in the neighborhood of La Hague nuclear plant

Pobel and Viel reported the results of a case-control study of leukemia in the neighborhood of La Hague nuclear reprocessing plant in France[34]. They conducted an interview survey of 27 leukemia cases diagnosed during the period 1978-93. In addition, they selected 92 controls, matching them on the sex, age, place of birth and residence at time of diagnosis of cases. The cases were restricted to those living within a 35 km radius of La Hague nuclear plant and aged less than 25 years at the time of diagnosis. They obtained a RR of 4.5 (95% confidence interval* was between 1.5 and 15.2) for those who used the local beach once or more in a month when compared to those who never used beach. Based on those findings, they argued that childhood leukemia was caused by environmental radiation exposure associated with recreational activities on beaches. However, another possibility is confounding by factors other than radiation, and a recall bias.

*confidence interval: a range of values that has a specified probability of containing the rate or trend. The 95% (P value = 0.05) confidence intervals are the most commonly used. In radiation epidemiology, 90% CIs are also used[4].

4. Is a case-control study a comparison between cases and non-cases?

Suppose that you conduct a case-control study, in which cases are all cancers and controls are non-cancer subjects. If this case-control study is conducted in Japan, 30-40% of controls will develop cancer in their lifetime. In other words, this study has many controls who are non-cases at present but become cancer cases in future. In this study, controls cannot be regarded as non-cases in a strict sense.

As Olli Miettinen pointed out, even in case-control studies, the disease risk should be evaluated in the direction from the exposure of interest to the disease of interest (Theoretical Epidemiology)[14]. In the case-control study using incident cases, controls should not be considered as non-cases but a sample from the catchment population from which cases are derived (ME-II. P124)[8]. In such a case-control study, "controls" are selected samples to examine the distribution of exposure variables and covariates among the catchment population of cases. In other words, controls are expected to be a representative sample of the catchment population from which cases arise. Therefore, Olli S Miettinen proposed to use the term "referent" rather than "control". In such a study, the OR to be calculated can be regarded as a RR[35].

5. Advantages and disadvantages of case-control studies

1) Efficiency

Generally speaking, case-control studies are less expensive and less time consuming. For very rare diseases, cohort studies are impractical and case-control studies are only a practical choice (ME3-Chapter 8)[13]. On the other hand, if the exposure of interest is rare, case-control studies are not efficient. For example, it is not a good idea to examine the association of leukemia risk and atomic bomb radiation in the areas other than Hiroshima or Nagasaki.

2) The exposure of interest

Case-control studies can examine more than one exposure variable. However, cohort studies can also examine the health effects of various exposure variables. Therefore, it is not an advantage specific to case-control studies.

3) The disease of interest

In case-control studies, cases are usually limited to the disease of interest for the sake of cost and time. However, that may not be a good idea from the viewpoint of validity. A frequently witnessed drawback of epidemiological studies is the lack of positive and negative control results, which are always required for animal and laboratory experiments. In the case-control study of childhood cancers to examine their associations with natural background radiation which was conducted by Kendall and his colleagues, the principal endpoint of interest was leukemia, but all types of cancer were included in cases[36]. Natural radiation exposure was significantly related to leukemia risk but not to all cancers except leukemia. The absence of excess risk for all cancers except leukemia makes the presence of excess leukemia risk more credible.

4) Continuous exposure

It is easier for case-control studies to collect information on continuous/intermittent exposure for the entire period of concern when compared to cohort studies. Suppose you are interested in the health effect of smoking. The exposure period of concern is from age at starting smoking to the time of interview. In case-control studies, it is not difficult to collect information on the period up to the time when cases were diagnosed. However, in cohort studies, you cannot collect information on smoking after the start of follow-up since the base-line survey (to collect information on smoking and other factors) is conducted before the start of follow-up. Needless to say, it is desirable to collect information on smoking periodically after the start of follow-up. However, it is not an easy task to conduct such a prospective or concurrent survey.

6. Variants of a case-control study

1) Case-only studies (case-case studies)

Sato *et al*. examined the association between mobile phone use and acoustic neuroma risk, using a case-only study[37]. This study examined the relationship between tumor location and laterality of mobile phone use prior to the reference dates (1 and 5 years before diagnosis). Patients were regarded as exposed when they had used the mobile-phone in the same side of brain tumor. The unexposed group was those with a tumor on the opposite side of mobile phone use. An odds ratio (O) was calculated from a cross-tabulation of the sides of exposure and tumor. Then, the relative risk was calculated as $(O^{0.5} + 1)/2$.

2) Two-stage sampling

If exposure is rare, it is not a good idea to select controls from the entire study population. It is particularly so if it is expensive to collect information on exposure variables. For example, if the proportion of the exposed is only 10%, even if 1,000 controls are selected, only 100 are expected to be found exposed. In such a situation, it is wasting of time and money to conduct an extensive exposure assay for a relatively large number of unexposed subjects. In the two-stage sampling, investigators conduct a screening for exposed in the entire study population or a population sample (first-stage controls), and then, they select sufficient numbers of exposed and unexposed subjects (second-stage controls). Using the data for second-stage controls and all the cases and taking sampling fractions into account, the magnitude of exposure will be

examined. If the number of cases is relatively large, the same approach can be used for cases[38, 39].

3) Counter-matching

Suppose a case-control study nested in a cohort in which a control is selected for each leukemia case, matching on sex and age. Suppose also that following pairs were obtained:

10 pairs of an exposed case and an unexposed control;
5 pairs of an unexposed case and an exposed control;
40 pairs of an exposed case and an exposed control; and
60 pairs of an unexposed case and an unexposed control.

For this kind of study, the OR taking into account the matching is calculated as 10/5 = 2. The OR calculation ignores the pairs of cases and controls with the same exposure status. This is called McNemar's method. If those pairs are ignored in the OR calculation, one may think that it is a good idea not to select those pairs from the beginning. The counter-matching is such a method. In the counter-matching method, unexposed controls are selected for the exposed cases, and exposed controls are selected for unexposed cases. Counter-matching was originally proposed as a case–control study nested in a cohort[40]. It is required that investigators have information on the counter-matching variable of all cohort members.

4) Case-specular study

The exposure to magnetic field has been suspected to increase childhood cancer risk (the magnetic field hypothesis). However, the neighborhood with elevated magnetic field levels may have an increased childhood cancer risk due to factors other than magnetic field (the neighborhood hypothesis). The case-specular method was designed to discriminate between those two hypotheses. Wire coding is frequently used as a proxy for magnetic field exposure. The distance from power lines is the major factor of wire codes. An example of the case-specular method is the study to compare the wire codes of case residences with the wire codes of specular residences constructed by switching the location of the case residence across the center of the street[41].

7. Retrospective vs. prospective

In case-control studies, case ascertainment and control selection can be done in a concurrent manner. As JM Last put it, cases and controls in a case-control study may be accumulated "prospectively". In other words, as each new case is diagnosed, it is entered in the study. Such studies can be regarded as prospective or concurrent case-control studies. It should be noted, however, that the completeness of prospective case ascertainment in case-control studies and cohort studies may depend on exposure status.

According to JM Last, the case-control study is retrospective since it looks back from the outcome to its causes. He pointed out that it is true even for a case-control study with prospective case and control accumulation (DE-4). Such a retrospective feature may affect the answers of interviewees. The selection of cases and controls may also be affected by the factors of interest.

The base-line study of a cohort study needs the agreement of participants. However, such an agreement is not a requirement for the subsequent follow-up using vital statistics and/or cancer registry. Even if that is required, the completeness of case ascertainment is unlikely to be affected by the exposure status of cohort members. On the other hand, in the retrospective collection of information on exposure in case-control studies, it is required for investigators to obtain the agreement from cases and controls (sometimes, the approval of medical doctors who diagnosed the disease is also necessary for contacting cases). As a result of such requirement, the participation rates of cases may be affected by the exposure of interest if cases know their exposure history. That is also true for controls. In recent years, this kind of requirement makes good case-control studies ever more difficult to conduct.

V. Ecological study

1. What is an ecological study

In ecological studies, the unit of analysis is a population or a group of people, rather than an individual. On the other hand, in cohort studies and case-control studies, the observed unit is an individual.

Table 3-1. Cancer risk and drinking in towns A, B, and C.

	A	B	C
cancer patients	50	30	10
population	1,000	1,000	1,000
% drinkers	80%	50%	20%

2. Ecological fallacy, an example

Table 3-1 shows the numbers of newly diagnosed cancer patients, population sizes and proportions of drinkers in towns A, B and C in a certain year. The proportions of drinkers in towns A, B and C appear to correlate with the frequencies of cancer in those towns.

Table 3-2 is the break-down of Table 3-1. This table enables us to conduct town-specific examinations with respect to the frequencies of cancer among drinkers and non-drinkers. In such a comparison, the frequencies of cancer show no difference between drinkers and non-drinkers.

Table 3-2. Break-downs of table 3-1

		A	B	C
drinkers	cancer patients	40	15	2
	population	800	500	200
non-drink	cancer patients	10	15	8
	population	200	500	800

3. Definition of ecological fallacy

The ecological fallacy is the bias that may occur because an association observed between variables on an aggregate level does not necessarily represent association that exists at an individual level. A correlation between variables based on group (ecological) characteristics is not necessarily reproduced between variables based on individual characteristics (DE-4).

In an ecological study conducted by Bernard L Cohen, he examined the correlation between lung cancer mortality rates and average indoor radon levels in 1,601 counties in the United States, adjusting for smoking[42]. In this study, decreasing mortality rates were associated with increasing radon concentrations. Darby and Doll pointed out that Cohen's results are and will always be burdened by the ecological fallacy[43]. They wrote as follows: "Unless smoking is irrelevant to lung cancer risk (which we know to be untrue) or smoking status and residential radon are uncorrelated within each county (which seems unlikely), the relationship between residential radon and lung cancer at the county level will differ from that at the level of the individual in a way that cannot be overcome by including corrections for smoking habits at the county level, even if these corrections correctly represent the smoking habits of the individuals within each county."

4. Age adjustment

US NCI's Glossary of Statistical Terms describes age-adjusted rate as follows: "an age-adjusted incidence or mortality rate is a weighted average of the age-specific incidence or mortality rates, where the weights are the proportions of persons in the corresponding age groups of a standard million population. The potential confounding effect of age is reduced when comparing age-adjusted rates computed using the same standard population[4]." This approach is called the direct age-adjustment. The most famous age specific weights are those obtained from the Segi's and WHO's World Standard Populations, which are shown in Table 4[44].

In Table 5-1, town A has a cancer mortality rate (/1,000) of 25, which is the crude mortality rate. A crude mortality rate is a weighted sum of age specific mortality rates and can be obtained from the following formulas:
10 x 0.25 + 20 x 0.25 + 30 x 0.25 + 40 x 0.25 = 25, where 0.25 (=1,000/4,000) is the proportion of each age group.
In Table 5-2, town B has a cancer mortality rate (/1,000) of 20. This is a weighted sum of age specific mortality rates and can be obtained from the following formulas:
10 x 0.4 + 20 x 0.3 + 30 x 0.2 + 40 x 0.1 = 20. Here, for example, 0.4 is calculated as 4,000/10,000.
Note here that towns A and B have the same age specific mortality rates.

Table 4. The World population

age	Segi's World standard population	WHO world standard population
0 -	12,000	8,860
5 -	10,000	8,690
10 -	9,000	8,600
15 -	9,000	8,470
20 -	8,000	8,220
25 -	8,000	7,930
30 -	6,000	7,610
35 -	6,000	7,150
40 -	6,000	6,590
45 -	6,000	6,040
50 -	5,000	5,370
55 -	4,000	4,550
60 -	4,000	3,720
65 -	3,000	2,960
70 -	2,000	2,210
75 -	1,000	1,520
80 -	500	910
85 +	500	630
	100,000	100,000

Table 5-1. Cancer mortality in Town A

town A

age	cancer deaths	population	mortality rate (/1000)
0 - 39	10	1,000	10
40 - 59	20	1,000	20
60 - 79	30	1,000	30
80 +	40	1,000	40
total	100	4,000	25

Table 5-2. Cancer mortality in Town B

town B

age	cancer deaths	population	mortality rate (/1000)
0 - 39	40	4,000	10
40 - 59	60	3,000	20
60 - 79	60	2,000	30
80 +	40	1,000	40
total	200	10,000	20

When the cancer mortality rates in those two towns are compared, the same age-specific weights should be used for both towns. Since towns A and B have the same age specific mortality rates, if age adjustment uses the same age-specific weights for both towns, the age adjusted mortality rate of town A will be the same as that of town B. When age-specific weights are calculated using the Segi's World Standard Population, the age-standardized mortality rate of Towns A and B will be as follows:
10 x 0.68 + 20 x 0.21 + 30 x 0.1 + 40 x 0.01 = 14.4

Age adjustment can be used for comparison of incidence in different subcohorts. However, in cohort studies, investigators frequently use the SMR analysis, which was already described in the section 5.

When incidence rates (or mortality rates) in different populations are compared, adjustments for sex and age are usually made. However, if such comparisons are modified by sex and/or age, it is not a good idea to make adjustment for sex or age. Rather, the results of comparisons should be presented for each sex group (or each age group), separately.

Suppose a comparison between towns A and B is made. In the comparison, you may find that the male mortality of town A is larger than that of town B while the female mortality of town A is smaller than that of town B. In this kind of case, it is a good idea to present the results of men and women, separately.

VI. Descriptive epidemiology and analytical study

Descriptive epidemiology is concerned with and designed only to describe the existing distribution of variables, without regard to causal or other hypothesis (DE-4)[2]. An ecological study without any hypothesis is a typical descriptive study. On the other hand, analytical epidemiology is designed to examine putative or hypothesized causal relationships. Analytical studies are represented by cohort studies and case-controls studies examining the relationship between the exposure and disease of interest.

The descriptive epidemiology is distinguished from analytical epidemiology by the lack of study hypothesis and the utilization of existing data. However, one may examine official cancer statistics data, such as area-specific cancer mortality data, to address some hypothesis regarding cancer risk. For example, one may examine the time trend of brain tumor mortality in a country, suspecting that the recent increase of CT scans among children may increase their brain tumor risk. This kind of study can be regarded as an analytical study in a broad sense even though it uses existing data. On the other hand, case-control comparison or analysis of cohort-study data without any special hypothesis can be regarded as descriptive studies.

VII. Disease prevalence and cross sectional survey

The prevalence of, for example, thyroid carcinomas is the proportion of patients with thyroid carcinoma among a study population. Prevalence is unitless. The prevalence of a disease can be obtained from a survey of populations at a particular time. That kind of study is called a cross-sectional study. DE-4 defines it as follows: "a study that examines the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at one particular time." Its synonym is a disease frequency survey and a prevalence study. A problem of a cross-sectional study is its difficulty to examine temporality.

The diagnosis of thyroid carcinomas is affected by the method of diagnosis and skills of examiners. For example, small thyroid carcinomas can be detected by careful palpation. The use of ultrasonography enables us to detect even smaller carcinomas and nodules. Therefore, standardized surveys are necessary to investigate thyroid cancer risk among a population. By the end of 2013, the Fukushima Thyroid Survey detected 74 prevalent cases of thyroid carcinomas among the 249,253 subjects with the final diagnosis. The prevalence is 74/249,253. Since prevalence = incidence x duration of disease (ME-III, Chapter 3)[13], in order to estimate the incidence from the prevalence, we need to know the average duration of disease, which seems difficult to be estimated in any study. In this study, however, we can use a similar but slightly different approach to estimate the incidence. If any detectable thyroid carcinomas do not disappear during the lifetime, the number of observed cases can be considered as cases accumulated in the population by the time of survey. The person-years of observation for accumulated thyroid carcinoma cases in the population can be calculated by multiplying the number of examinees (= 249,253) by 10 years, which is the average age of participants. On the basis of those assumptions, the incidence rate can be calculated as the number of accumulated cases divided by person-years, which is 74 / (249,253 x 10). This value (= 30 / 1,000,000) is much higher than the incidence of childhood thyroid carcinoma (= 1 - 2 / 1,000,000). Assuming the incidence of childhood thyroid carcinoma to be 2/1,000,000, a P value can be calculated as follows (using R statistical software) : 1 - sum [dpois {0 : 73, N x D x (2 / 1,000,000)}], where N = 249,253 and D = 10. This value is much smaller than 0.05, indicating that the difference between the observed and expected incidence rates is statistically significant. This apparently high incidence can be explained by the use of highly sensitive method to detect asymptomatic thyroid carcinomas. It should be pointed out, however, that, even if the expected incidence is 10 times larger than 2 / 1,000,000 due to the screening effect, the observed

incidence is still significantly larger than the expected value. Having said that, it should also be pointed out that cross-sectional studies cannot present any conclusive evidence for the causal relationship between exposure and disease risk.

VIII. Case ascertainment through longitudinal survey

The follow-up of a fixed or dynamic population to ascertain newly occurred cases among a population is a typical longitudinal survey. Repeated observations/examinations of a population can also be regarded as a longitudinal survey. An example of such a longitudinal survey includes the AHS of atomic bomb survivors in Hiroshima and Nagasaki. Incident cancer cases can be ascertained by a longitudinal survey that repeats the observation/examination of a population or the ascertainment of newly occurred cases among a population. Another example of such a study is the cohort study of Chernobyl residents reported by Brenner *et al*., in which investigators analyzed the data obtained from young residents (aged younger than 18 years on April 26, 1986) in three contaminated oblasts (states) of Ukraine who underwent up to four thyroid screening examinations between 1998 and 2007 (n = 12,514)[45]. This study identified incident thyroid cancer cases by a series of cross-sectional surveys. Those who had thyroid carcinomas in the first screening were excluded from the cohort. Then, incident cases were ascertained from suspected thyroid cancer cases detected in the second to fourth screening examinations, which were conducted during the period 2001-2007.

1. Mortality survey – a longitudinal case ascertainment

1) Overview

Mortality survey is a longitudinal case ascertainment. The method of mortality follow-up is different from country to country. It can be conducted through checking against vital statistics data*. In the United Kingdom, vital status and causes of death can be obtained through the National Health Service Central Register (NHSCR). Based on ID information on the study subjects sent by investigators, the NHSCR flags living members for future notification. When the flagged subject dies, the NHSCR notifies investigators.

For mortality statistics, it is necessary to determine the underlying cause of death. The underlying cause of death is i) the disease or injury that initiated the train of events leading to death or ii) the circumstances of the accident or violence that produced fatal injury (DE-4)[2]. The underlying cause of death used in vital statistics is determined on the basis of information given in the death certificate.

2) Verbal autopsy

Verbal autopsy is defined as a procedure of gathering systematic information for determining the cause of death in situations where deceased subjects have not been medically attended. This method can be used for determining causes of death with distinct symptom complexes that can be recognized, remembered, and reported by lay respondents. It is considered a useful way to enhance the quality of mortality statistics in developing countries (DE-4)[8].

In the cohort study conducted in the high natural background radiation areas of Yangjiang in Guang Dong Province, China, it was necessary for investigators to establish a special system to identify the cause of death since the information recorded on official death certificates was considered unreliable. Therefore, investigators used the following approach: first, they visited the local village and asked "bare-footed" doctors, who are not licensed physicians, to record the medical information of deceased subjects on the Registry Card of Death, which they prepared in advance. After establishing such a system, they visited the village periodically, and checked the causes of death recorded in the Registry Card of Death. Then, they went to the hospitals in the study area, and, if necessary, upper-level hospitals, and reviewed the medical records and other relevant documents of the deceased. Using the information thus collected, they decided the cause of death for each deceased resident[46]. If the deceased did not get any "westernized" medical care, the cause of death was determined by verbal autopsy.

3) LSS of atomic bomb survivors

The LSS uses the legally mandated Japanese family registration (Koseki) system to follow the mortality of atomic bomb survivors. This approach makes it possible to ascertain almost all the deaths of cohort members living in Japan. Then, the underlying causes of death were determined, using the information recorded on death certificates.

The death certificate is a vital record signed by a licensed physician or by another designated health worker (DE-4)[2]. It usually includes places of residence and death, and whether the deceased had been medically attended before death. In the tabulation for cause of death, the immediate cause of death is recorded in the first line, followed by conditions giving rise to the immediate cause.

4) REA mortality follow-up of nuclear industry workers in Japan

The Institute of Radiation Epidemiology (IRE) of the

*Vital statistics is systematically tabulated information concerning births, marriages, divorces, separations, and deaths based on registrations of these vital events (DE-4).

Radiation Effects Association (REA) based in Tokyo, Japan follows the nuclear workers with Japanese nationality, using resident registration cards (RRCs). The RRC of every Japanese national is stored in the registration file kept at the municipality office. The RRC keeps records of resident's name, date of birth, gender, address, first date of residence, and previous address. If any resident dies or moves to another municipality, the RRC is updated with the date of death or relocation. In the case of relocation, the next address is also recorded. The RRCs of former residents are kept in a different registration file for 5 years, after which it is deleted from the file. Using RRC, the cohort can be followed up both retrospectively (up to 5 years) and prospectively.

In Japan, information on death certificates is promulgated in the following way:

i) the Notification of Death (ND), together with the death certificate (DC) of the deceased, is submitted to the registry office located in the Honseki (permanent address) of the deceased, the place of death or the address of the notifier, which is usually a family member of the deceased (the ND and DC are then sent to the local office of Ministry of Justice under whose jurisdiction the Honseki area falls. The documents are kept there for 27 years);

ii) the information necessary for vital statistics are filled in the Vital Statics Survey (VSS) form in the registry office, and the VSS form is sent to the public health center of the address of the deceased (a copy of the VSS form or the information necessary for vital statistics is kept there for 3 years);

iii) the VSS form is sent to Ministry of Health, Labor and Welfare (MHLW) through the prefectural government;

iv) MHLW codes the cause of death and prepares vital statistics.

The data are available in a digital form but the computerized data do not include the name of the deceased.

For the deceased subjects, the REA/IRE determined the cause of death through record linkage with magnetic tape copies of vital statistics death records provided by the MHLW. Indices used for record linkage were the date of birth, date of death, sex and municipality of residence. In the study reported by Iwasaki *et al*., this approach identified the cause of death for 99.7% of deceased subjects[47].

5) US

In the United States, various approaches are necessary to identify deaths among a population. In the mortality follow-up of 46,970 Rocketdyne (Atomics International) workers[48], who were employed during the period 1948–1999, the alive status was determined from various linkages of the study roster with national databases, including the National Death Index (NDI), the California Death Statistical Master File, the Social Security Administration (SSA) Death Master File, and other SSA files. For record linkage, investigators used Comserv, a computer services firm specializing in locating persons (www.comserv-inc.com), and LexisNexis, an online information service provider (www.lexisnexis.com). For matching with the SSA Death Master File and the California Death Statistical Master File, they used the Centers for Disease Control and Prevention matching program, Link Plus. When this program gives a high probabilistic score, investigators reviewed the individual records of the matched subject before a decision on the validity of the match was made.

2. Cancer incidence

1) Introduction

A registry system is necessary to identify incident cancer cases among a large population. The nation-wide cancer registry system is established in some countries. In Asia, Taiwan and Korea have such a system but Japan does not. In American continents, Canada has but the United States of America does not. In Europe, the United Kingdom and Scandinavian countries have the national cancer registries.

2) Atomic bomb survivors in Hiroshima and Nagasaki

The Hiroshima and Nagasaki tumor registries were established in 1957 and 1958, respectively. Those registries employ active case ascertainment, in which cancers are ascertained by abstractors visiting large hospitals in Hiroshima City and Nagasaki Prefecture. These data are supplemented with information on cancer deaths, the Hiroshima and Nagasaki tissue registries (registries of tumor pathology reports and slides), AHS data and so on.

3) Karunagappally cancer registry

The cancer registry in Karunagappally, Kerala, was officially initiated as of January 1, 1990. The active cancer registration method was used to ascertain cancer patients[19]. Registry workers routinely visited Medical College Hospital Trivandrum, major pathological laboratories in Trivandrum, which is the capital of Kerala State, and adjacent areas, and all the hospitals and medical practitioners in Karunagappally Taluk every year. This rural area does not have any dedicated cancer treatment center or diagnostic facility. The Regional Cancer Centre (RCC) in Trivandrum is the comprehensive cancer center in Kerala. Most of cancer patients in Karunagappally seek treatment in RCC Trivandrum.

Death reports were obtained from the death registers kept in the vital statistics division of the panchayat, an administrative unit of a taluk. Investigators visited the house of deceased subjects to collect additional information on the cause of death.

4) Taiwan

In Taiwan, cancer cases can be ascertained through

computerized linkage with the National Cancer Registry of Taiwan (NCRT), which was established in 1979 by the Department of Health of Taiwan Government. The NCRT requests cancer hospitals, which are medical institutions with greater than 50-bed capacity providing outpatient and hospitalized cancer care, to report all cases of newly diagnosed cancer. For each registered case, the NCRT keeps records of ID information such as the national identification number, name, sex and the date of birth in addition to diagnostic information, including the date of diagnosis, the location and histology of the cancer[49].

References

1. WHO. Health topics: Epidemiology [homepage on the internet]. Available from: http://www.who.int/topics/epidemiology/en/
2. Last J (2001) Dictionary of Epidemiology, 4th edition. International Epidemiological Association.
3. Porta M (2008) A Dictionary of Epidemiology.5th edition. Oxford University Press.
4. National Cancer Institute. NCI Dictionary of Cancer Terms: observational study [homepage on the internet] Available from: http://www.cancer.gov/dictionary? CdrID = 286105
5. Brown JR and Thornton JL (1957) Percivall Pott (1714-1788) and Chimney Sweepers' Cancer of the Scrotum. Br J Ind Med 14(1): 68–70.
6. Lind J (1753) A Treatise of the Scurvy in Three Parts. Containing an inquiry into the Nature, Causes and Cure of that Disease, together with a Critical and Chronological View of what has been published on the subject. A. Millar, London.
7. Hempel S (2013) John Snow. The Lancet [serial online]. 381(9874): 1269–1270. Available from: http://www.thelancet.com/journals/lancet/article/PIIS0140–6736(13)60830–2/fulltext
8. Rothman KJ and Greenland S (1998) Modern Epidemiology second edition. Lippincot, Wiliamd and Wilkins.
9. Gardner MJ and Winter PD (1984) Mortality in Cumberland during 1959-78 with reference to cancer in young people around Windscale, Lancet 1(8370): 216–217.
10. Gardner MJ, et al. (1987) Follow up study of children born to mothers resident in Seascale, West Cumbria (birth cohort). Br Med J 295: 822–827.
11. Gardner MJ, et al. (1987) Follow up study of children born elsewhere but attending schools in Seascale, West Cumbria (schools cohort). Br Med J 295: 819–822.
12. Lagakos SW, Mosteller F. (1986) Assigned shares in compensation for radiation-related cancers. Risk Anal 6(3): 345–357.
13. Kreyszig E (1970) Introductory mathematical statistics. Principles and methods. John Wiley & Sons. Inc. New York.
14. Miettinen OS (1985) Theoretical Epidemiology. Wiley, New York.
15. Rothman KJ and Greenland S (2008) Modern Epidemiology third edition. Lippincot, Wiliamd and Wilkins.
16. Beebe G and Usagawa M (1968) The major ABCC samples. TR12-68. Atomic Bomb Casualty Commission.
17. http://www.rerf.or.jp/programs/rp_e/index.html
18. Nair MK, et al. (1999) Population study in the high natural background radiation area in Kerala, India. Radiat Res 152: S145–148.
19. Azizova TV, et al. (2010) Cerebrovascular diseases in the cohort of workers first employed at Mayak PA in 1948–1958. Radiat Res 174: 851–864.
20. Krestinina LY, Davis F, Ostroumova E, et al. (2007) Solid cancer incidence and low-dose-rate radiation exposures in the Techa River cohort: 1956–2002. Int J Epidemiol 36: 1038–1046.
21. Bauer S, Gusev BI, Pivina LM, et al. (2005) Radiation exposure due to local fallout from Soviet atmospheric nuclear weapons testing in Kazakhstan: solid cancer mortality in the Semipalatinsk historical cohort, 1960–1999. Radiat Res 164: 409–419.
22. Tung CJ, Chao TC, Chen TR, et al. (1998) Dose reconstruction for residents living in ^{60}Co-contaminated rebar buildings. Health Phys 74: 707–713.
23. Chen, W.L. (2002) Radiation surveys and dose equivalent assessments for ^{60}Co-contaminated rebar buildings. Appl Radiat Isot 56: 901–906.
24. Hwang SL, et al. (2006) Cancer risks in a population with prolonged low dose-rate gamma-radiation exposure in radiocontaminated buildings, 1983–2002. Int J Radiat Biol 82: 849–858.
25. Azizova TV, et al. (2010) Cardiovascular diseases in the cohort of workers first employed at Mayak PA in 1948–1958. Radiat Res 174: 155–168.
26. Ivanov VK, Maksioutov MA, Chekin SYu, et al. (2000) Radiation-epidemiological analysis of incidence of non-cancer diseases among the Chernobyl liquidators. Health Phys 78: 495–501.
27. http://www.nrer.ru/rgmdr_eng.html
28. Hosoda Y, et al. (1997) First Analysis of Mortality of Nuclear Industry Workers in Japan,1986–1992. J Health Phys 32: 173–184.
29. UNSCEAR (2006) Report to the General Assembly. Effects of Ionizing Radiation. Annex A: Epidemiological studies of radiation carcinogenesis. United Nations Scientific Committee on the Effects of Atomic Radiation, United Nations, New York.
30. Clark GN, Briggs A, Cooke AM. (2005) A History of the Royal College of Physicians of London, Oxford University Press, ISBN 0-19-925334-X.
31. Gardner MJ, et al. (1990) Results of case-control study of leukemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria, BMJ 300: 423–429.
32. Doll R, Evans J and Darby SC (1994) Paternal exposure not to blame, Commentary, Nature 367: 678–680.
33. HSE investigation of leukemia and other cancers in the children of male workers at Sellafield. HSE BOOKS 1994
34. Pobel D and Viel JF (1997) Case-control study of leukaemia among young people near La Hague nuclear reprocessing plant: the environmental hypothesis revisited. BMJ. 314: 101–106.
35. Miettinen O. (1976) Estimability and estimation in case-referent studies. Am J Epidemiol. 103: 226–235.
36. Kendall GM, Little MP, Wakeford R, et al. (2013). A record-based case-control study of natural background radiation and the incidence of childhood leukaemia and other cancers in Great Britain during 1980–2006. Leukemia 27(1): 3–9.
37. Sato Y, et al. (2011) A case-case study of mobile phone use and acoustic neuroma risk in Japan. Bioelectromagnetics. 32(2): 85–93.
38. Walker AM (1982) Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. Biometrics 38: 1025–1032.
39. White JE. (1982) A two stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 115: 119–128.
40. Langholz, B. & Clayton, D. (1994) Sampling strategies in nested

case-control studies. Environ Health Perspect 102 (Suppl. 8): 47–51.

41. Ebi KL, Zaffanella LE and Greenland S (1999) Application of the case-specular method to two studies of wire codes and childhood cancers. Epidemiology 10(4): 398–404.

42. Cohen B L (1995) Test of the linear no-threshold theory of radiation carcinogenesis for inhaled radon decay products. Health Phys 68: 157–174.

43. Darby S and Doll R. (2000) Reply to 'Explaining the lung cancer versus radon exposure data for USA counties'. J Radiol Prot 20: 221–222.

44. Ahmad OB, Boschi-Pinto C, Lopez AD et al. (2001) Age standardization of rates: a new WHO standard. GPE Discussion Paper Series: No.31 EIP/GPE/EBD World Health Organization.

45. Brenner AV, et al. (2011) I-131 dose response for incident thyroid cancers in Ukraine related to the Chornobyl accident. Environ Health Perspect 119: 933–939.

46. Tao Z, et al. (2012) Cancer and non-cancer mortality among Inhabitants in the high background radiation area of Yangjiang, China (1979-1998). Health Phys 102: 173–181.

47. Iwasaki T, et al. (2003) Second analysis of mortality of nuclear industry workers in Japan, 1986-1997. Radiat Res 159: 228–238.

48. Boice JD Jr., et al. (2006) Mortality among Rocketdyne workers who tested rocket engines, 1948-1999. J Occup Environ Med 48: 1070–1092.

49. Lai, M-S (2010) Cancer monitoring and control planning in Taiwan. JACR Monograph No 16. Japanese Association of Cancer Registries. Tokyo.