A Method for the Estimation of Completeness of Cancer Registration
Application to the Fukuoka Cancer Registry

Tetsuya Mizoue1, Takesumi Yoshimura1,4, Masato Ikeda1, Itsuro Ogimoto1, Takao Shigematsu2,4, and Tsunehisa Nagata3,4

A method is proposed for estimating the number of incident cases of cancer and completeness of cancer registration using data from a population-based cancer registry. It is based on two assumptions: 1) submission of supplementary report does not depend on the duration of illness; and 2) registration reports and death certificates are independent sources of information. The method proposed in the present study has two advantages: 1) the number of incident cases can be estimated, even when completeness of registration varies over a period of time; 2) over-estimation of incidence by a registry can be evaluated. The use of this method was illustrated by data from the Fukuoka Cancer Registry, and completeness of registration was estimated as 69.4% for males and 66.2% for females. Furthermore, validity of the two assumptions was evaluated, and result from the evaluation suggested that for most sites the assumptions were valid and the estimated completeness not biased. But for some sites, such as liver, lung, and pancreas, the second assumption would be less valid and the completeness underestimated.

Received April 16, 1991; accepted June 5, 1991.

1Department of Clinical Epidemiology, Institute of Industrial and Ecological Sciences, University of Occupational and Environmental Health, Fukuoka, Japan. 2Department of Public Health, School of Medicine, Fukuoka University, Fukuoka, Japan. 3Fukuoka Medical Association, Fukuoka, Japan. 4Committee of Fukuoka Cancer Registry, Fukuoka, Japan.

Address for correspondence: Takesumi Yoshimura, Department of Clinical Epidemiology, Institute of Industrial and Ecological Sciences, University of Occupational and Environmental Health, 1-1 Iseigaoka Yahatanishi-ku Kitakyushu-shi Fukuoka 807, Japan.
available data, and possible bias on the estimates is discussed.

MATERIALS AND METHODS

The Fukuoka Cancer Registry

The Fukuoka Cancer Registry, established in 1984, is a population-based cancer registry covering Fukuoka Prefecture, which has a population of about 5 million. Like other registries in Japan, this registry relies on two primary sources of information: 1) registration reports, which are submitted voluntarily to the registry by the clinician or by the registrant in the hospital; and 2) death certificates, on which there is any statement regarding cancer. In addition, the registry also collects a secondary source of information: supplementary reports on diagnosis of cancer, which are submitted on request by the registry for deceased cases without a registration report. Identification of cases is achieved through computer-assisted search of these data. When more than one independent primary cancer are found in the same individual, each is counted separately for incidence computation.

Taking into account the delays in submission of reports and time for registration work, incidence computed for a certain year can only be reported by the registry after 3 or 4 years. The reported number is equal to the number of cases with a registration report or a supplementary report plus the number of cancer cases identified only through a death certificate (DCO cases).

Estimation of completeness of registration

The target to be estimated is all cancer cases diagnosed in the year 1984 among Fukuoka residents. The following files and information were obtained from the registry for the estimation: 1) file of registration reports and that of supplementary reports on cases for which the date of diagnosis was during 1984; 2) file of death certificates with any mention of cancer for 1984 and 1985; and 3) the response rate to the request by the registry for submission of supplementary reports for deceased cases without a registration report (55.4% for all sites).

As is shown in Figure, all cases diagnosed in 1984, including those not registered, were classified according to whether they were present or absent in each of the data sources. Among them, A and B were ascertained by registration reports that the year of diagnosis was 1984, and C1 by supplementary reports. But the number of cases for C2 and that for D could not be obtained directly through existing data. Therefore, estimation of these numbers are needed to obtain the total number of incident cases.

For the estimation, two assumptions were made: firstly, submission of supplementary report was irrespective of the duration of ill state; secondly, registration reports and death certificates were independent sources of information.

Based on the first assumption, the number of cases for C2 was estimated as follows:

\[ C2 = C1 \times (1 - r) / r \] (r: response rate).

Based on the second assumption, the number of cases of D was estimated as follows:

\[ D = B \times (C1 + C2) / A \]

and hence the total number of cancer cases were estimated.

Completeness of registration was calculated as the ratio of the reported number of cases to the number estimated by this method. These values were calculated by sex and 21 anatomical sites. Estimated number of cases for total sites is the sum of those for each site.

For the assessment of validity of the assumptions used in the analysis, distribution of length of time between onset of cancer and death was compared among three groups divided by their status of registration: group of cases with a registration report (X), with a supplementary report (Y), and without any reports (Z). Information on length of the time was obtained from death certificates of 1985 with any mention of cancer; cases without such information were excluded from the analysis. For sites with more than one hundred deaths in that year, the Wilcoxon test was performed to test the difference in the distribution among the three groups. If there is no significant difference between group Y and group Z, we can conclude that the first assumption is valid and the estimated number for C2 in figure not biased. Also, if there is no significant difference between group X and other groups, we can conclude that the second assumption may be valid and the estimates obtained not biased.

RESULTS

The results are shown in Table 1, which includes reported and estimated number of incident cases (Ie), percent completeness of registration, and ratio of the estimated number to the reported number of cancer deaths in 1984 (D) by site and sex. The estimated number of cases for all sites is 16,676 (9,101 for males and 7,575 for females). The discrepancy between the estimated and the reported number is comparatively greater in stomach, uterus, and breast cancers. The percent completeness is 69.4% for males and 66.2% for
females, and the ratio of $I_e$ to $D$ ($I_e/D$) is 1.91 for males and 2.28 for females. The completeness of registration considerably varies by site. Higher completeness is found in such sites as the hematopoietic tissue, liver, pancreas, and lung, while it is lower in sites such as the bladder, rectum, and prostate. It should be noted that the percent completeness is over 100% for the hematopoietic tissue in both sexes and for the pancreas in females.

Table 2 shows three types of distribution pattern for length of time between onset of cancer and death by group. Group X consists of cases with a registration report, group Y with a supplementary report, and group Z without any reports (DCO). In regard to the stomach, there were no significant differences among the three groups. The same results were obtained for the lymphatic tissue, hematopoietic tissue, uterus, breast, esophagus, and bladder. On the other hand, significant differences were found between group X and the other groups but not between group Y and group Z at the sites of the liver, lung, and pancreas. Also, significant differences between group Y and group Z were seen at the sites of the colon and rectum. A slightly different pattern for the gall bladder and bile duct was noticed. Distribution of length of the time among cases of group X was significantly different from that in group Z, but not group Y.

**DISCUSSION**

There are two indicators of completeness of registration:

- $I_e/D$: Ratio of the estimated number to the reported number of cancer deaths during the same year.
- $\%C$: Completeness of registration as a percentage.

### Table 1. Estimated completeness of cancer registration by sex and site, Fukuoka Cancer Registry, 1984.

| Site                        | Male                  | Female                 |
|-----------------------------|-----------------------|------------------------|
|                             | Reported no. of cases | Estimated no. of cases | Percent completeness | Ratio to deaths$^a$ | Reported no. of cases | Estimated no. of cases | Percent completeness | Ratio to deaths |
| Oral cavity, and pharynx    | 89                    | 116                    | 76.9                  | 1.48                | 51                    | 95                    | 53.5               | 4.76            |
| Esophagus                   | 213                   | 247                    | 86.2                  | 1.48                | 51                    | 55                    | 91.7               | 1.29            |
| Stomach                     | 1781                  | 2824                   | 62.9                  | 2.40                | 1093                  | 1729                  | 63.3               | 2.25            |
| Colon                       | 366                   | 580                    | 63.3                  | 2.62                | 347                   | 520                   | 66.7               | 2.35            |
| Rectum                      | 283                   | 620                    | 45.7                  | 3.54                | 203                   | 429                   | 47.4               | 2.84            |
| Liver                       | 1070                  | 1182                   | 90.9                  | 1.29                | 359                   | 357                   | 101.0              | 1.09            |
| Gall bladder, and bile duct | 155                   | 195                    | 80.0                  | 1.42                | 219                   | 275                   | 79.4               | 1.62            |
| Pancreas                    | 224                   | 255                    | 87.7                  | 1.19                | 207                   | 203                   | 102.0              | 1.07            |
| Larynx                      | 55                    | 203                    | 27.1                  | 8.47                | 6                     | 21                    | 28.0               | 5.21            |
| Trachea, bronchus, lung and pleura | 1047                  | 1155                   | 90.9                  | 1.26                | 443                   | 521                   | 84.7               | 1.45            |
| Bone                        | 10                    | 12                     | 82.0                  | 1.11                | 10                    | 12                    | 83.3               | 1.20            |
| Skin                        | 42                    | 57                     | 74.1                  | 4.74                | 49                    | 67                    | 73.0               | 3.36            |
| Breast                      | 2                     | 3                      | 78.7                  | 2.55                | 441                   | 884                   | 50.0               | 4.70            |
| Uterus                      | —                     | —                      | —                     | —                   | 659                   | 1023                  | 64.5               | 4.70            |
| Ovary                       | —                     | —                      | —                     | —                   | —                     | 147                   | 165                | 90.9            |
| Prostate                    | 129                   | 273                    | 47.2                  | 2.90                | —                     | —                     | —                  |
| Bladder                     | 125                   | 366                    | 34.1                  | 6.42                | 56                    | 132                   | 42.4               | 3.30            |
| Brain, and nervous system   | 38                    | 45                     | 82.6                  | 2.19                | 28                    | 42                    | 67.6               | 4.15            |
| Thyroid                     | 20                    | 65                     | 30.6                  | 9.33                | 71                    | 315                   | 22.5               | 10.86           |
| Lymphatic tissue            | 183                   | 294                    | 62.1                  | 2.33                | 137                   | 225                   | 61.0               | 2.47            |
| Haematopoietic tissue       | 137                   | 136                    | 101.0                 | 1.15                | 134                   | 121                   | 111.1              | 0.94            |
| Others                      | 365                   | 474                    | 76.9                  | 1.69                | 304                   | 388                   | 78.1               | 1.71            |
| All sites                   | 6334                  | 9101                   | 69.4                  | 1.91                | 5015                  | 7575                  | 66.2               | 2.28            |

$^a$ Ratio of the estimated number to the reported number of cancer deaths during the same year.
Table 2. Distribution of length of time between onset of cancer and death by the status of registration for selected sites.

| Site  | Group of casesa | No. of subjects | Length of time between onset of cancer and deathb (months) | P valuec vs. group X | P valuec vs. group Y |
|-------|-----------------|-----------------|----------------------------------------------------------|---------------------|---------------------|
|       |                 |                 | 0-2 | 3-5 | 6-11 | 12-23 | 24+ |                 |                   |
| Stomach | X | 541 |
|         | Y | 618 |
|         | Z | 466 |
| Liver   | X | 337 |
|         | Y | 390 |
|         | Z | 392 |
| Colon   | X | 151 |
|         | Y | 157 |
|         | Z | 133 |

The values in the table represent the percentage of cases and the cumulative percentage in parentheses.

a Information was obtained from death certificates

b Group of cases: X: with a registration report, Y: with a supplementary report, Z: without any reports (DCO cases)

c Wilcoxon rank test

The proportion of cases registered only through death certificate with a statement on cancer, or "death certificate only" (DCO); and 2) the ratio of the number of deaths attributed to cancer to the number of registered cases, or "the mortality/incidence ratio" (M/I)\(^\text{13}\)) But these indicators do not show how many cases of cancer were not registered among residents in the area covered by a registry. In addition, DCO changes to a large extent by submission of a supplementary report for each deceased case without a registration report, and M/I varies with change of fatality rate. Therefore, these indicators are less comparable among registries, and also among different points in time within the same registry.

Robles et al. applied the capture-recapture method to estimate the number of cases and completeness of registration\(^\text{14}\)). But the method could not be validly applied when completeness considerably varies over a period of time, because it implicitly assumed that the number of DCO cases are constant over time. DCO cases reflect completeness of registration both for the year they died and the previous years\(^\text{13}\)). Our method drew a distinction between cases with a clinical report and those without (DCO cases). Our method, therefore, could evaluate registry data with considerable variation in completeness. In addition, the capture-recapture application cannot evaluate over-estimation of incidence by a registry which could occur when completeness of registration rapidly increased over a period of time. Therefore, the completeness estimated by Robles' method is always equal to or less than 100%. But, as is shown in Table 1, our method can evaluate the over-estimation, eg. the hematopoietic tissue, where the estimated completeness is above 100%.

Over-estimation, however, due to over-diagnosis or duplicate registrations\(^\text{13}\)) could not be evaluated by our method.

The ratio of the number of cases to that of deceased cases for all sites was estimated to be about 2 by our method. This value is almost the same as that found in the Connecticut Cancer Registry (I/D : 1.89 for males, 2.13 for females)\(^\text{13}\)), where registration was nearly complete (DCO was about 1%). In Japan, although there is no such complete registry, Kato et al. estimated the I/D for the years from 1979 to 1984 to be 1.86 by the regression method\(^\text{10}\)). Furthermore, the results that Ie/D for all sites was higher in females than in males is reasonable because cancer specific to females, such as cancer of breast and uterus, has a comparatively good prognosis on the whole.

The estimates of completeness were inversely related to ratios of the estimated number of cases to deaths (Ie/D), that is, the estimates increased as Ie/D decreased. Spearman rank correlation coefficient between these estimates and Ie/D for the 21 site-groups is \(-0.81\) for
males and -0.89 for females. These findings were found in previous reports, and also predicted from the registration system of the Fukuoka Cancer Registry. Even if submission of registration reports was equally incomplete for every site-group, site-groups with high fatality would be more completely registered through death certificates than those with low fatality.

Validity of the two assumptions used are crucial for estimating incidence and completeness of registration. Examination of length of time between onset of cancer and death revealed that length of the time was almost identically distributed between cases with a supplementary report and those without (DCO cases) for most sites. This result supports the first assumption that submission of supplementary report was irrespective of the duration of ill state. But for the colon, length of the time among cases with a supplementary report was significantly longer than those without, and thus the number of C2 (in Figure) would be overestimated. The opposite result was obtained for the rectum (data not shown in the Table), thus the number underestimated.

The second assumption, that registration reports and death certificates were independent sources of information, can only be evaluated by the independent case ascertainment method. This method, however, was quite difficult to conduct, we compared length of time from onset of cancer to death between group X and other groups instead. Length of the time among cases with a registration report was significantly longer than that among those without for the sites of liver, lung, and pancreas, while it was almost equally distributed for other sites. This result indicates that, for the former sites, cases with a registration report might survive longer than those without, and thus there might exist an inverse relation between registration reports and death certificates. Therefore, the estimated number of cases would be overestimated, and the estimated completeness underestimated for these sites. Possibility of such bias on estimates was suggested from the Osaka Cancer Registry. But for the latter sites, the assumption might be valid and the estimates not biased.

The level of reporting to a cancer registry varies by hospitals. Variation in completeness of registration and that of the prognosis of patients among hospitals may combine to produce such a dependent relation as a whole. One possible way to avoid this problem is to divide hospitals into groups according to their type and/or size and then to estimate completeness for each group. This will increase the validity of the estimates.

The one limitation of our method is that estimates based on small number of deceased cases suffer from random error. This could happen in sites where incidences are too low and/or survival rates are too high to obtain a sufficient number of deceased cases for the estimation. Thus, estimates for such sites as the thyroid and bone are less reliable. This type of limitation is common to all methods using the death certificate for the estimation.

In this paper, we introduced a simple method for the estimation of completeness of cancer registration. Although the application of this method is restricted to registry data which contain supplementary reports for cases first known through a death certificate, such reports are routinely collected in most of the registries in Japan. For the assessment of completeness of registration, this method could usefully be applied to such registry data.

ACKNOWLEDGEMENTS

This study was partly supported by Research Group for Population-based Cancer Registration (Chairman, Isaburo Fujimoto) under a Grant-in-Aid for Cancer Research from the Ministry of Health and Welfare, Japan 2-5. We acknowledge to staffs in the Fukuoka Cancer Registry for their efforts to provide useful information.

REFERENCES

1. Berkel J. General practitioners and completeness of cancer registry. J Epidemiol Community Health, 1990; 44: 121-124.
2. Gogldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Reviews, 1980; 2: 210-220.
3. Nwene U, Smith A. Assessing completeness of cancer registration in the North-Western region of England by a method of independent comparison. Br J Cancer, 1982; 46: 635-639.
4. Benn RT, Leck IAN, Nwene UP. Estimation of completeness of cancer registration. Int J Epidemiol, 1982; 11: 362-367.
5. Heiberger RM, Miller CL, Feigl P, Lane WW, Glaefke G. A novel method of assessing completeness of tumor registration. Cancer, 1983; 51: 2362-2366.
6. Mattssson B, Rutqvist LE, Wallgren A. Undenotification of diagnosed cancer cases to the Stockholm Cancer Registry. Int J Epidemiol, 1985; 14: 64-69.
7. Storm HH. Completeness of cancer registration in Denmark 1943-1966 and efficacy of record linkage procedures. Int J Epidemiol, 1988; 17: 44-49.
8. Freedman LS. Variations in the level of reporting by hospitals to a regional cancer registry. Br J Cancer, 1978; 37: 861-865.
9. Hanai A, Kawai M, Fujimoto I. Estimation of completeness of registration in population-based cancer registry. In: Fujimoto I, ed. Annual report of population-based cancer registration in Japan. Osaka, 1987: 5-18. (Japanese)
10. Kato I, Fujioka M, Tominaga S. Estimation of cancer incidence based on the data of the population-based cancer registry. Jpn J Public Health, 1988; 35: 204-209.
11. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. J Clin Epidemiol, 1988; 41: 495-501.

12. Mizoue T, Ikeda M, Ogimoto I, et al. Estimation of cancer incidence using the data of population-based cancer registry. Jpn J Cancer Clin, 1991; 37: 317-320.

13. Muir C, Waterhouse J. Comparability and quality of data: reliability of registration. In: Muir C, Waterhouse J, Mack T, Powell J, Whelan S, eds. Cancer Incidence in Five Continents, Vol. 5. IARC, Lyon, 1987: 45-169.