Environmental Health Perspectives
Vol. 32, pp. 75-81, 1979

Study of Cancer Mortality by Grid Square Method

by Toshiteru Okubo*

The grid square method as used in Japan has been standardized by the government, by use of the lines of the earth's latitude and longitude. The basic unit covers an area of approximately one square kilometer.

An evaluation of this method is focused on the geographical distribution of cancer mortality in the city of Tokyo over a period of six years. The results indicated that for stomach and lung cancers, there was a very clear geographical distribution. High stomach cancer areas were similar for both males and females, and by comparing data with census data, it was evident that the blue-collar areas showed high mortality. High lung cancer areas were also similar for both males and females, but the distribution was opposite to that of stomach cancer, i.e., the higher mortality was observed in white-collar areas.

Because the basic area unit of one square kilometer was too small for statistical analysis of geographical distribution, the moving average of nine grid squares as well as a "combined grid square method" based on population density was used.

By this study a number of advantages of the grid square method as opposed to methods employing existing government boundaries became evident. The boundary lines do not move with political expediencies. *X-Y* coordinates can be easily defined for statistical analyses by computer, facilitating computer mapping, calculating the center of distribution, determination of the contour lines, and the estimation of values in places which lie between sampling stations.

Introduction

In the field of cancer epidemiology, clear analysis of variations in geographical distribution is of great importance. Up to the present time, there are numerous reports of comparisons on worldwide (1, 2), national (2, 3), and local levels. These studies on geographical distributions aid in the formulation of hypotheses of the causes of cancer and are important in helping to solve problems in health care administration.

In measuring geographical distribution, the most important factor is determining how the area is to be divided, i.e., determining the unit of area and where the lines of division are to be drawn. The reports on the geographical distribution of cancer mortality up to the present have almost all been based on geographical boundaries determined by government agencies, e.g., nations, states, prefectures, cities, etc. Since governments determine boundaries based on units or areas where communities formed in a natural process, the geographical features and climate, as well as socioeconomic factors including culture, behavior, industry, race, and religion were more or less uniform within a given community. Where the conditions are uniform in this manner, analysis of cancer geographical distribution using government-determined boundaries is of great value.

However, in Japan after World War II, there was very rapid change in the economic structure of the nation, and the population moved from mainly agricultural regions to industrial zones. Further, following the development of tertiary industries, concentration of the population in the areas in and around the largest cities (e.g., Tokyo, Osaka, Nagoya) occurred, and because of this social mobilization, the significance of the above-mentioned distribution based on geographical and historical division has been lost. In addition, in order to correct the population deviation which occurred, administrative agencies often changed boundary lines. Thus, in present-day Japan, the advantages of using administrative boundaries for determining geographical dis-

*Department of Environmental Health, Jichi Medical School, Minamikawachi-machi, Tochigiken, 329-04 Japan.

tribution have become fewer than was previously the case.

Aside from the social, economic, and historical implications of administrative divisions, there exists a purely statistical problem in terms of analytical method. When mapping geographical distribution, the size of the population and area is the main problem, as it must demonstrate clearly the variations in the distribution. In Japan, when indicating vital statistics and morbidity, the smallest unit generally employed is that area included by one local government health center in metropolis and a city, town or village in other districts. By law there is supposed to be one such local health center for every 100,000 persons; however, this is not always the case, and in instances where the population has undergone a rapid rise, there may be several hundred thousand persons belonging to a single center. The similar situation is also seen in districts other than metropolis. Within such large areas, there is the possibility of the existence of pockets of large cancer mortalities as well as those of few mortalities. In such cases, the large cancer mortality pockets and small cancer mortality pockets average out statistically and the variations as they actually exist cannot be distinguished.

When investigating geographical distribution, sometimes it becomes desirable to devise a new method of division for smaller area than that as determined by government boundaries. However, if the method used varies with each research project, not only would the time and cost involved be enormous, but comparison of data from different projects would be impossible. As an answer to the problem, a standard method of division on grid square basis has been devised (*5*).

The grid square method uses as a base, global latitudes and longitudes, dividing the area in between into smaller squares. It has the advantage that, unlike government boundaries, the lines are not subject to change and the areas are regular in size. Furthermore, *X-Y* coordinates can be easily determined. Various grid square methods are in use throughout the world (*5, 6*).

A grid square method has been officially adopted for use as the standard by the Administrative Management Agency of the Japanese government in 1973. The numbering system of the method as used in Japan is as follows. The largest unit of area is designated by a four-digit number; the sub-area is designated by a two-digit number, which in turn is followed by the designation of a "sub-sub" area, also consisting of two digits. Thus any area in Japan can be designated using a total of eight digits. It is the "sub-sub" area, comprising approximately one square kilometer, which is the most commonly used unit in Japanese studies, and the following discussion employs this as the basic area unit.

This grid square method was first used in the greater Tokyo metropolitan area in the course of the 1965 Census. Then in 1970, it was used in other metropolitan areas, and in 1975, on a nationwide scale (*7*). At present, the method is used for various kinds of statistical investigations by both the national and local governments. As reported in this paper, the author performed a geographical distribution study of cancer mortality in the city of Tokyo in order to evaluate the various merits of the standardized grid square method and confirmed that, for example, the determination of the geographical distribution of small areas is possible and that more detailed feature of cancer mortality distribution can be determined (*8-10*). Furthermore, as the same method is also becoming more widely used in fields other than health statistics, statistical analysis for the relation between these factors and cancer mortality has become possible.

Geographical Distribution of Cancer Mortality in the City of Tokyo

The city of Tokyo is 572 square kilometers in area; the population in 1970 was 8,839,000; the population per square kilometer was 15,400 in the same year. Although the population density is very great, the cancer mortality becomes too small for purposes of statistical analysis, especially when classified by cancer site. Cancer mortality in 1970 in Tokyo was 103.8 per 100,000. The expected number of deaths per one grid square is about 16. When this figure of 16 was divided by sex and site, the expected number of deaths from stomach cancer, which has the highest mortality, became less than five. When the expected number falls below five in one grid square, the degree of freedom for the observed number to decrease becomes small and as a result, one can no longer distinguish those areas with observed low mortality from those with high mortality.

Table 1 shows the mortality percentage of major cancers by sex and site. It was necessary to total the deaths at least over a period of six years (1966-1971) in order to distinguish the distribution in one grid square for those cancers showing over 10% mortality among males. Even then, excluding stomach and uterine cancers, the number of other cancers among females within one grid was not sufficient for

Table 1. Number of cancer mortalities by sex and site.

	Male, number (%)	Female, number (%)
Esophagus	1,667 (6.9)	458 (2.4)
Stomach	10,295 (42.6)	6,373 (33.9)
Intestine	757 (3.1)	766 (4.1)
Rectum	948 (3.9)	794 (4.2)
Liver	1,754 (7.3)	1,230 (6.6)
Pancreas	941 (3.9)	637 (3.4)
Lung	2,985 (12.3)	1,209 (6.4)
Leukemia	709 (2.9)	560 (3.0)
Uterus	— (—)	2,171 (11.6)
Breast	11 (0.0)	1,234 (6.6)
Other sites	4,124 (17.1)	3,348 (17.8)
Total	24,191 (100.0)	18,780 (100.0)

analysis. For the analysis of units where the numbers were not sufficient, the moving average was determined. To determine the moving average, the number of deaths in the grid square in question plus those in the eight surrounding squares were totalled and averaged. When the moving average for the adjacent grid was determined, the nine grids were shifted over one grid.

In addition, in those cases where the numbers were not sufficient, a "combined grid square method" was also employed. This method combined a number of squares to form such groups as to fit the range of the number of deaths for males (180 ± 60) and females (140 ± 40). To determine which grid squares to combine, the main consideration was whether the areas were industrial or residential in nature. The industrial areas were combined and the residential areas were combined in so far as it was possible to do. Efforts were made to avoid the formation of a single lone grid as well as the formation of strangely shaped combinations of grids (see Figs. 5 and 6).

As data for the study, death certificates of those who were registered residents of the city of Tokyo were used. These included 24,191 males and 18,780 females, as shown in Table 1. The addresses on the death certificates were noted, and it was then determined as to which single census taker's territory the address fell in. Next it was noted which grid square the census taker's territory was in, and a table was made for each grid square.

The percentages of cancer mortality by site were used for analysis. However, if percentages alone are employed for a single square or for the moving average, there is the probability of great chance variation, and the statistical reliability is very low because of large differences in the number of deaths in each square. Thus in these cases (single grid square or moving average), the observed number of deaths in each square was assumed to be in binominal distribution. The total observed number of deaths of cancer of specific site was divided by the total number of cancer deaths in the entire Tokyo area to obtain the expected rate. This expected rate was used to calculate a probability P of death by cancer of specific site per grid square as follows:

$$P = \frac{n\,!}{x\,!\,(n-x)\,!}\; r^x\,(1-r)^{n-x}$$

where n = number of total cancer deaths per grid square; x = number of deaths by cancer of specific site; r = expected rate; P = probability of death by cancer of specific site per grid square.

In order to express the magnitude of death by cancer of specific site in one grid square, the figure calculated by the following equation was used as the index.

$$\text{Index} = \sum_{x=0}^{\theta} P$$

where θ = observed number of deaths by cancer of specific site.

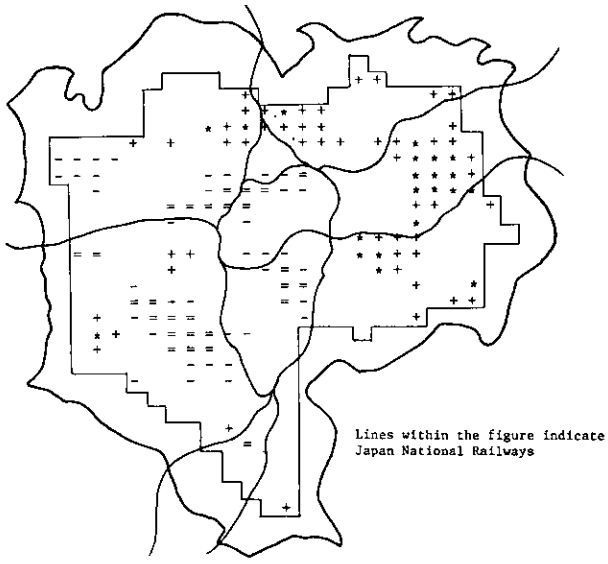
Thus, if the index in one square is over 0.5, then it can be concluded that the observed number of specific site deaths exceeds the expected number.

Figures 1-4 show the geographical distributions of stomach and lung cancer mortalities for both males and females. The moving average employing binominal distribution was used in these figures. As shown in Figures 1 and 2, the geographical distribution of stomach cancer mortality showed cluster areas of both high and low mortalities. These cluster areas were approximately the same for both males and females. The areas of high mortality were the northeastern districts of the city of Tokyo and the low mortality areas extended from the center of the city to the southwest, showing a clear geographical division.

Examining the data from the censuses of 1965 and 1970, the author discovered that this geographical distribution of stomach cancer corresponded to areas where blue-collar workers were the majority (the northeast) and where the majority of the population were white-collar workers (the southwest).

Concerning lung cancer, as shown in Figures 3 and 4, the geographical distribution of mortality clusters was the reverse of that for stomach cancer, i.e., high mortality clusters appeared in the southwest area, while low clusters appeared in the northeast. This was the case for both males and females.

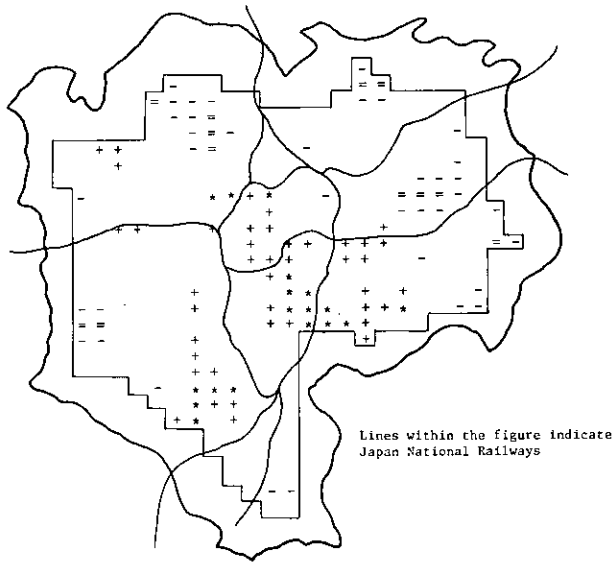
Figures 5 and 6 were drawn by using the "combined grid square method" described above. In this method, the units were combined so that the number



Range	Remarks Symbol	No. of areas
P ≧ 0.98	*	24
0.98 > P ≧ 0.90	+	49
0.90 > P ≧ 0.10	blank	268
0.10 > P ≧ 0.02	-	39
0.02 > P	=	27

P = probability (cf. text)

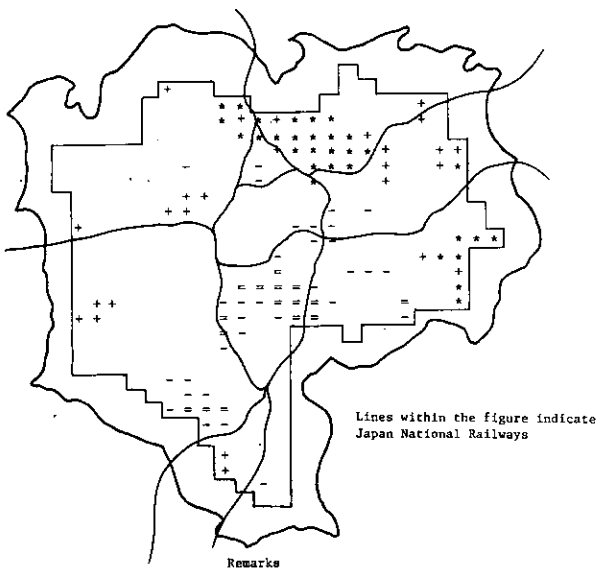
FIGURE 1. Geographical distribution of stomach cancer mortality among males, using the moving average.



Range	Remarks Symbol	No. of areas
P ≧ 0.98	*	18
0.98 > P ≧ 0.90	+	43
0.90 > P ≧ 0.10	blank	297
0.10 > P ≧ 0.02	-	37
0.02 > P	=	12

P = probability (cf. text)

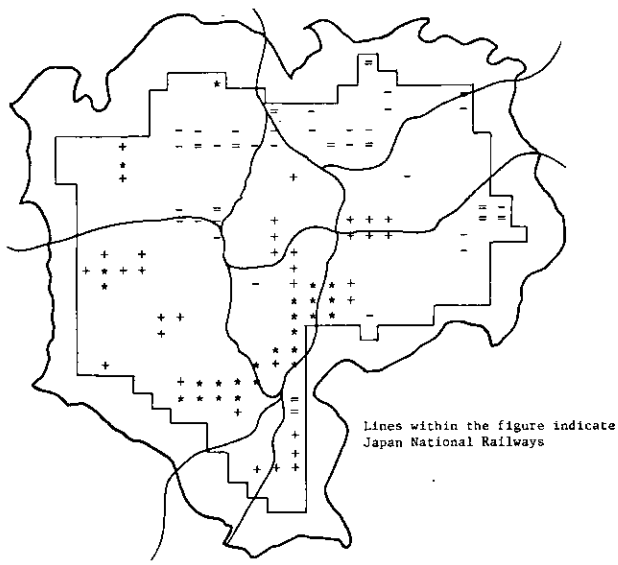
FIGURE 3. Geographical distribution of lung cancer mortality among males, using the moving average.



Range	Remarks Symbol	No. of areas
P ≧ 0.98	*	31
0.98 > P ≧ 0.90	+	26
0.90 > P ≧ 0.10	blank	301
0.10 > P ≧ 0.02	-	33
0.02 > P	=	16

P = probability (cf. text)

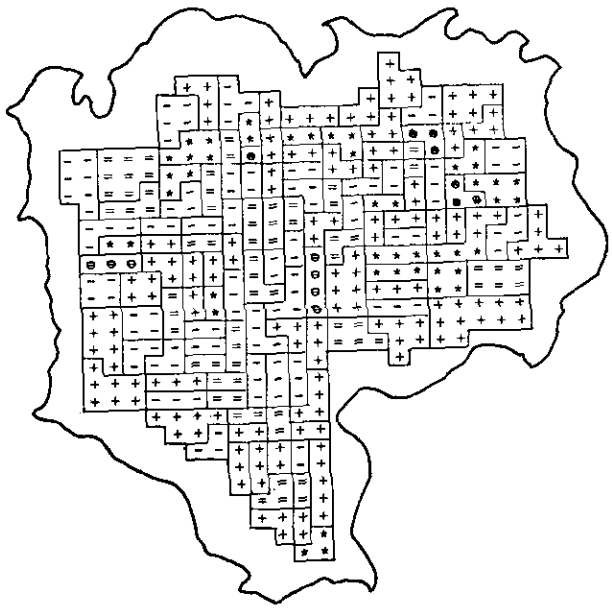
FIGURE 2. Geographical distribution of stomach cancer mortality among females, using the moving average.



Range	Remarks Symbol	No. of areas
P ≧ 0.98	*	25
0.98 > P ≧ 0.90	+	34
0.90 > P ≧ 0.10	blank	307
0.10 > P ≧	-	28
0.02 > P	=	13

P = probability (cf. text)

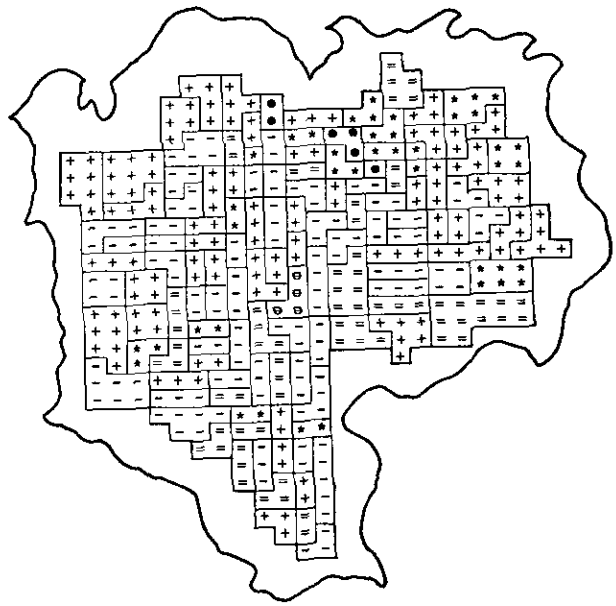
FIGURE 4. Geographical distribution of lung cancer mortality among females, using the moving average.



Remarks		No.	Distribution
R ≧ 50.3	●	3	*
50.3 > R ≧ 46.7	*	15	*******
46.7 > R ≧ 42.5	+	49	**********
42.5 > R ≧ 38.6	-	38	**********
38.6 > R ≧ 34.7	=	23	**********
34.7 > R	●	2	*

R = percentage (cf. text)

FIGURE 5. Geographical distribution of stomach cancer mortality among males, using the combined grid square method.



Remarks		No.	Distribution
R ≧ 42.6	●	3	*
42.6 > R ≧ 38.3	*	15	*******
38.3 > R ≧ 34.0	+	46	**********
34.0 > R ≧ 29.7	-	46	**********
29.7 > R ≧ 25.4	=	19	*********
25.4 > R	●	1	*

R = percentage (cf. text)

FIGURE 6. Geographical distribution of stomach cancer mortality among females, using the combined grid square method.

of deaths would fall within a similar range; the crude percentage of cancer mortality by site was used as the index. By this method, Tokyo was divided into a total of 130 units, and the average and standard deviation of the indices ($\bar{x} \pm$ SD) were calculated. The indices were divided into six groups with the following as the five points of division: $\bar{x}$, $\bar{x} \pm 1$ SD, and $\bar{x} \pm 2$ SD.

Figures 5 and 6 indicate stomach cancer only, and although they were compiled using a method different from that used in Figures 1 and 2, the geographical distribution was the same for both males and females. Although the figures for lung cancer compiled by the combined grid square method are not included here, the geographical distribution was for the most part the same as in Figure 3 and 4.

Evaluation of the Grid Square Method

Although this report on cancer mortality presents only a small number of results, the following discussion attempts to evaluate the grid square method in a broad perspective.

The grid square method used in Japan uses a basic unit comprising an area of one square kilometer, which is a much smaller area than that comprised by government boundaries, which continue to be used in many fields. Figure 7 indicates the distribution of lung cancer mortality for both sexes in the city of Tokyo, 1966-1971, by territory of local health center for comparison with the distribution by grid square method (Figs. 3 and 4). As compared with Figure 7, it can be said that Figures 3 and 4 show more actual picture of both high and low mortality clusters, indicating more detailed variations within areas under local health center.

As mentioned above, even when this basic area unit is too small for statistical analysis, a number of units can be easily combined, depending on the purpose, e.g., even in those areas where the population is relatively large but the prevalence of a given disease is too low, as was the case in cancer mortality described above. If it is necessary for the combined

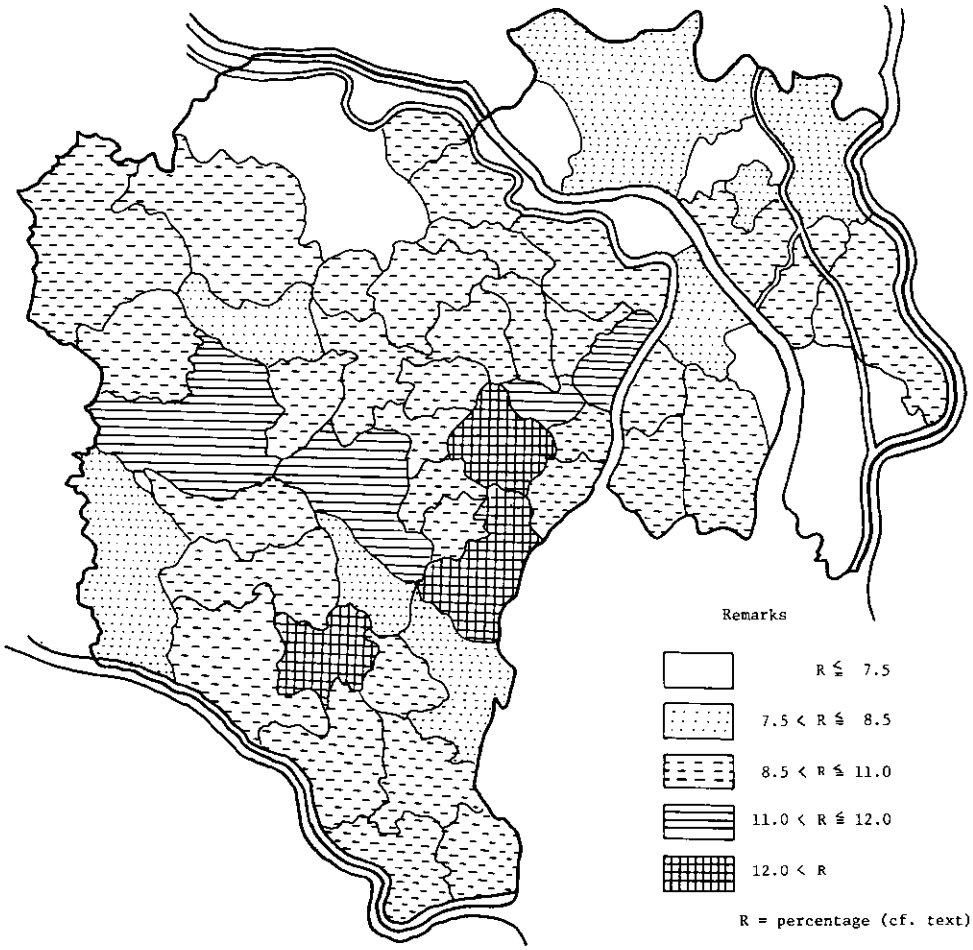


FIGURE 7. Geographical distribution of lung cancer mortality for both sexes by the division of local health center's territory. (1966-1971).

grid squares to maintain a square shape, this can be easily done by combining 4, 9, 16, etc. grid squares. Further, should an investigation require that areas with similar characteristics (e.g., residential, industrial, etc.) be combined, this can also be easily accomplished. On the other hand, in the case, for example, of suburban or rural areas where population clusters occur sporadically, such populations can be considered separately. Furthermore, if areas along one road are to be considered, this can also be done.

This method is also very useful for the establishment of a surveillance in small areas. Thus, for instance, by using this method, a higher than average incidence of cancer among the inhabitants living in the vicinity of a factory can be detected, and as a result, the factory can be suspected of leaking a carcinogen.

Using this method, the grid squares can be readily defined on *X-Y* coordinates and is convenient for statistical analyses by computer, e.g., for computer mapping, for calculating the center of distribution, for determining the contour lines of various parameters, such as those formed by connecting those points having the same death rate or incidence of a disease, and for the estimation of values in places which lie between sampling stations for the purpose of comparing health and environmental factors. The above cannot be adequately measured by using government boundaries. If clearer relationships between health and environmental factors can be determined using this grid square method, it will become possible using simulation data to assess the development of future urban and industrial areas.

Although site-specific percentages among total cancer are used in this report as indicator for the geographical distribution, mortality rate based on population is also obtainable by the grid square division. The Bureau of Statistics in Japan continues to

publish the results of census on a grid square basis since the 1965 Census as well as those on the usual government area basis. The age-adjusted mortality rate can be calculated by this method; however these data are divided into only three groups, i.e., 0-14, 15-64, and over 65. Other socioeconomic statistics included in the census data are also available on grid square basis.

On the other hand, while in theory the lines of division in the grid square method are perfectly straight, in actual practice this is not possible, since, for example, a perfectly straight line may often pass through the middle of a single home. When making a grid square table for the purpose of a census, the unit used is that territory covered by a single census taker, which cannot be bordered by a straight line but zigzags somewhat. This gives rise to an element of error, and this error differs in degree according to place. This element of error must be taken into consideration when comparing data for different parameters, as different units may be used in different cases, e.g., traffic, population, industrial studies, etc. It should be noted here that in the data on population and cancer mortality in the present paper, the unit used was the same for both the population and mortality, i.e., that territory covered by a single census taker; thus in this study no element of error in this regard existed.

REFERENCES

1. Segi, M., and Kurihara, M. Cancer mortality for selected sites in 24 countries. Department of Public Health, Tohoku University School of Medicine, Sendai, Japan, 1966.
2. Dunham, L. J., and Bailar, J. C., III. World maps of cancer mortality rates and frequency ratios. J. Natl. Cancer Inst. 41: 155 (1968).
3. Mason, T., McKay, F. W., Hoover, R., Blot, W. J., and Fraumeni, J. F., Jr. Atlas of Cancer Mortality for U.S. Counties, 1950-1969. DHEW Publication No. (NIH) 75-780, Washington, D.C., 1975
4. Mason, T., McKay, F. W., Hoover, R., Blot, W. J., and Fraumeni, J. F., Jr. Atlas of Cancer Mortality among U.S. Nonwhites. 1950-1969. DHEW Publication No. (NIH) 76-1204, Washington, D.C., 1976.
5. Duncun, O. D., Cuzzort, R. P., and Duncun, B. D. Statistical Geography. Free Press of Glencoe, New York, 1961.
6. Forbes, J., and Robertson, I. M. L. Population enumeration on a grid square basis. The census of Scotland. J. Brit. Cartographic Soc. 33: (1967)
7. Bureau of Statistics, Office of Prime Minister Japan. Statistics on grid square basis, 1974.
8. Okubo, T., Adachi, S., and Toyama, T. An application of the grid square method to the geographical distribution of mortality statistics. I. Geographical distribution of cancer mortality in Tokyo. Nippon Eiseigaku Zasshi 32: 534 (1977).
9. Okubo, T., Adachi, S., and Toyama, T. An application of the grid square method to the geographical distribution of mortality statistics. II. Geographical distribution of mortality for the major causes in Tokyo. Nippon Eiseigaku Zasshi 32; 677 (1977).
10. Okubo, T. Relationship between mortalities by major cause and socio-economic factors in metropolis using the grid square method. Nippon Koshueiseigaku Zasshi 24: 713 (1977).