

# AN EPIDEMIOLOGICAL STUDY ON MORTALITY RATES OF VARIOUS CANCER SITES DURING 1958 TO 1971 BY MEANS OF FACTOR ANALYSIS

Haruo Yanai,\* Yutaka Inaba,\*\* Hirofumi Takagi,\*\*\*  
Hiroyuki Toyokawa\*\*\* and Shunichi Yamamoto\*\*

In order to analyze the pattern of the geographical and age distributions of the cancer death, the correlation coefficients between any pairs of cancer mortality rates of each site were calculated, using the data on cancer for 18 sites in males and 19 sites in females over the three periods, i.e., (1) from 1958 to 1962, (2) from 1963 to 1967, and (3) from 1968 to 1971. Then factor analysis by varimax method was performed on  $54 \times 54$  correlation matrix for males and  $57 \times 57$  matrix for females. As a result, five factors are extracted, which are commonly recognized in both males and females. Among the five factors extracted, the first one attracted us most, since the fact that bone, bladder, skin and buccal cancers as well as cancers of genital organs excluding ovary clustered in one group suggests a possible relation of those cancers to high atmospheric temperature.

## 1. Introduction

The spatial distribution of cancer mortality and morbidity rates has been studied by many authors, and its characteristic pattern as to the various cancer sites was the object of the research of many epidemiologists.

Segi (1965, 1966, 1969, 1972) has published a number of researches on cancer mortality rates of various sites in Japan as well as in foreign countries. Dunham et al. (1970) mapped mortality rates of 18 cancer sites on about one hundred geographical areas in the world. Independently of these works, Higgins (1969) reported the elaborate works on sex age distribution of cancer death, and also discussed the relationship between the incidence and mortality rates curves. As for the mathematical aspect of the age distribution of cancer, Armitage & Doll (1954) proposed multi-stage theory, in which the total preclinical period of cancer was divided into three stages by means of a mathematical model.

It has been generally recognized that any of the specific factors do not contribute independently but interdependently to the etiology in chronic diseases including various cancers. With this in mind, it is quite natural to assume that there exists multiple factors which collaborate with each other in the pathogenesis of cancer of various sites.

Burkitt (1969, 1970) stated that the variation in the geographic and temporal pattern of a disease may be related to its cause, and in view of this assertion several

---

*Key Words and Phrases;* factor analysis, correlation coefficient, geographical distribution, age distribution, sites of cancer, specific mortality, standardized mortality

\* Department of Psychology, Chiba University, Yayoicho, Chiba city 280

\*\* Department of Hygiene, Faculty of Medicine, University of Tokyo, Bunkyo, Tokyo 113

\*\*\* Department of Epidemiology, Faculty of Medicine, University of Tokyo

- 2) between any two mortality rates of the three periods over all the sites of cancer,
- 3) between the male and female mortality rates for each sites excluding prostate, uterus and ovary cancers.

Calculation was done separately for both sexes. Hence we compute  $54 \times 54$  correlation matrix for males and  $57 \times 57$  correlation matrix for females, on each of which factor analysis was performed. The procedure of factor analysis performed was as follows; first, estimate communality of each variable by the maximum absolute value of the correlation coefficients between a variable and the other variables, and perform principal factor analysis based on the correlation matrix, diagonal elements of which are replaced by communalities. Then we rotate the factor loading matrix thus extracted by varimax method.

Besides the analysis based on mortality rate for each prefecture, we calculated correlation coefficients for any pairs of cancer sites of the age distribution and performed factor analysis in the same procedure as the analysis of the geographical distribution.

### 3. Result

#### (1) Correlation coefficients of geographical distribution in any pairs of cancer sites

The correlation matrices for both sexes were calculated for any pairs of the cancer sites in each period. By convention in inferential statistics, correlation coefficients with the value of 0.304 or more were presumed to imply statistical significance at the 5% level, while those with 0.407 or more to imply the statistical significance at the 1% level, standing upon the premise that the specific mortality rates computed were random samples from the population of normal distribution. We showed in Table 2 the numbers of the pairs with significant correlation at the 1% level and those at the 5% level separately for each period and sex groups. It is recognized that in case of males the number of the pairs with statistically significant correlation coefficient increases as the calendar year proceeds on. Throughout those periods, the following pairs showed consistently the statistically significant correlation at the 1% level; for males, skin and liver, bladder and liver, bladder and kidney, and for females, skin and bladder, bladder and bile duct (negative correlation), stomach and intestine, rectum and intestine. In order to check the chronological stability of the geographical distributions associated with various cancer sites over three periods, we show in Table 3 the correlation

Table 2  
Numbers of pairs with significant correlations at the 5% level  
and 1% level for each period

|                              | Males |    | Females |    |
|------------------------------|-------|----|---------|----|
|                              | 5 %   | 1% | 5%      | 1% |
| First period<br>(1958-1962)  | 12    | 11 | 16      | 14 |
| Second period<br>(1963-1967) | 19    | 14 | 15      | 10 |
| Third period<br>(1968-1971)  | 22    | 17 | 24      | 18 |

coefficients of the mortality rates between different periods for male and female groups respectively. Consistently high correlation is meant to indicate that the chronological variation in geographical distribution is slight stable throughout the periods from 1958 to 1971. It is seen from Table 3 that the geographical distributions of stomach and esophagus cancer are most stable. It should be noted, however, that

Table 3  
Correlations of the geographical distributions of 20 cancer  
sites between each two of three periods  
(T1: 1958-1962, T2: 1963-1971, T3: 1968-1971)

| Sex<br>Periods<br>Sites | male  |        |        | female |       |        |
|-------------------------|-------|--------|--------|--------|-------|--------|
|                         | T1-T2 | T2-T3  | T1-T3  | T1-T2  | T2-T3 | T1-T3  |
| 1) buccal               | 0.531 | 0.555  | 0.493  | 0.262  | 0.458 | 0.313  |
| 2) esophagus            | 0.930 | 0.905  | 0.895  | 0.948  | 0.885 | 0.919  |
| 3) stomach              | 0.961 | 0.946  | 0.916  | 0.932  | 0.882 | 0.937  |
| 4) intestine            | 0.578 | 0.427  | 0.597  | 0.573  | 0.607 | 0.590  |
| 5) rectum               | 0.643 | 0.506  | 0.561  | 0.510  | 0.613 | 0.425  |
| 6) liver                | 0.918 | 0.954  | 0.892  | 0.825  | 0.908 | 0.808  |
| 7) bile duct            | 0.789 | 0.777  | 0.730  | 0.844  | 0.783 | 0.898  |
| 8) pancreas             | 0.577 | 0.750  | 0.604  | 0.560  | 0.557 | 0.491  |
| 9) larynx               | 0.519 | 0.555  | 0.558  | 0.152  | 0.265 | 0.042  |
| 10) lung                | 0.668 | 0.777  | 0.519  | 0.758  | 0.447 | 0.464  |
| 11) bone                | 0.552 | 0.358  | 0.265  | 0.523  | 0.333 | 0.323  |
| 12) skin                | 0.618 | 0.754  | 0.524  | 0.489  | 0.525 | 0.578  |
| 13) breast              | 0.220 | -0.203 | -0.192 | 0.405  | 0.561 | 0.329  |
| 14) uterus              | —     | —      | —      | 0.845  | 0.937 | 0.800  |
| 15) ovary               | —     | —      | —      | 0.371  | 0.376 | 0.530  |
| 16) prostate            | 0.512 | 0.582  | 0.581  | —      | —     | —      |
| 17) bladder             | 0.602 | 0.644  | 0.569  | 0.617  | 0.713 | 0.577  |
| 18) kidney              | 0.199 | 0.486  | 0.153  | 0.431  | 0.151 | -0.130 |
| 19) lymphoma            | 0.885 | 0.913  | 0.901  | 0.803  | 0.886 | 0.831  |
| 20) leukemia            | 0.557 | 0.551  | 0.350  | 0.215  | 0.310 | -0.121 |

extremely low correlation coefficients are found in the breast cancer of males. It is obvious that these low correlation coefficients resulted from such extremely low mortality rates as less than one per million. It is noticed that correlation coefficient for larynx cancer in females is extremely small as compared to that in males, and the result is supposed to be caused from the extremely low mortality rate.

In Table 4, we show the correlation coefficients of the site-specific mortality rates between males and females with respect to 17 cancer sites over the three periods. As it is seen that the variation of the coefficients along with those three periods is rather small for each cancer site, we might say that as the etiological agents the environmental factors outweigh the host factors in those cancer sites. It should be noted, however, that correlation coefficients were found to be small in case of leukemia for both sexes, despite the fact that the mortality rates were not so small.

## (2) Factor analysis of the geographical distribution

It was attempted to perform factor analysis on  $54 \times 54$  and  $57 \times 57$  correlation matrices in males and females respectively. As a result, five common factors were extracted for each case and the degree of cumulative contribution amounted to 78.1% and 78.4%

Table 4  
Correlation coefficients of the geographical distributions of 17 cancer sites  
between male and female groups  
(T1: 1958-1962, T2: 1963-1967, T3: 1968-1971)

| sites \ periods | T1     | T2     | T3    |
|-----------------|--------|--------|-------|
| 1) buccal       | 0.317  | 0.324  | 0.477 |
| 2) esophagus    | 0.805  | 0.817  | 0.767 |
| 3) stomach      | 0.904  | 0.917  | 0.898 |
| 4) intestine    | 0.530  | 0.506  | 0.522 |
| 5) rectum       | 0.393  | 0.496  | 0.565 |
| 6) liver        | 0.777  | 0.802  | 0.861 |
| 7) bile duct    | 0.807  | 0.809  | 0.695 |
| 8) pancreas     | 0.722  | 0.709  | 0.575 |
| 9) larynx       | 0.258  | 0.403  | 0.387 |
| 10) lung        | 0.594  | 0.514  | 0.401 |
| 11) bone        | 0.748  | 0.553  | 0.587 |
| 12) skin        | 0.576  | 0.535  | 0.456 |
| 13) breast      | -0.055 | -0.098 | 0.095 |
| 17) bladder     | 0.613  | 0.562  | 0.659 |
| 18) kidney      | 0.298  | 0.273  | 0.286 |
| 19) lymphatic   | 0.739  | 0.885  | 0.859 |
| 20) leukemia    | 0.097  | 0.182  | 0.027 |

for males and females respectively. We changed the orders of the extracted factors so that the factors for males might correspond to those of females in terms of the clusters which each factor revealed. The results are shown from Fig. 1 through Fig. 13. Fig. 1 and Fig. 3 show the factor loadings of the first factor (vertical line) and of the second factor (horizontal line) for males and females respectively, while Fig. 2 and 4 show those of the third factor (vertical line) and of the fourth factor (horizontal line) for male and females respectively. In addition to these figures, we drew eight diagrams of Japanese map indicating the factor scores of the four factors with respect to each of 46 prefectures, the names and locations of which are seen in Fig. 5. Since factor scores for each prefecture are calculated as weighted scores of mortalities of various sites of cancer using factor loadings as weights, they express the magnitude of mortality rates in each prefecture concerning cancer sites with high factor loadings on each factor. Note that the magnitude of factor score is divided into five grades according to the values calculated; i.e., (a) more than 1.5 (very high), (b) 1.5~0.5 (high), (c) 0.5~-0.5 (intermediate), (d) -0.5~-1.5 (low), and (e) less than -1.5 (very low).

Basing upon the result shown from Fig. 1 through 12, we shall delineate what each factor implies each factor one by one.

*First factor* (degree of contribution: 31.3% for males and 29.2% for females)

Since the degree of contribution for both sexes amounts to about one third of the total variance, this factor might suggest the most important pattern of geographical distribution. High loadings were found in buccal, skin, liver and bladder cancers and lymphoma for both sexes. Besides these cancers, high loadings were also found in prostate cancer for males, and uterus cancer for females (Fig. 1 and 3). It is interesting to note that factor loadings of lung cancers for both sexes become higher along the first factors as the calendar year proceeds on. We see from the factor scores plotted in

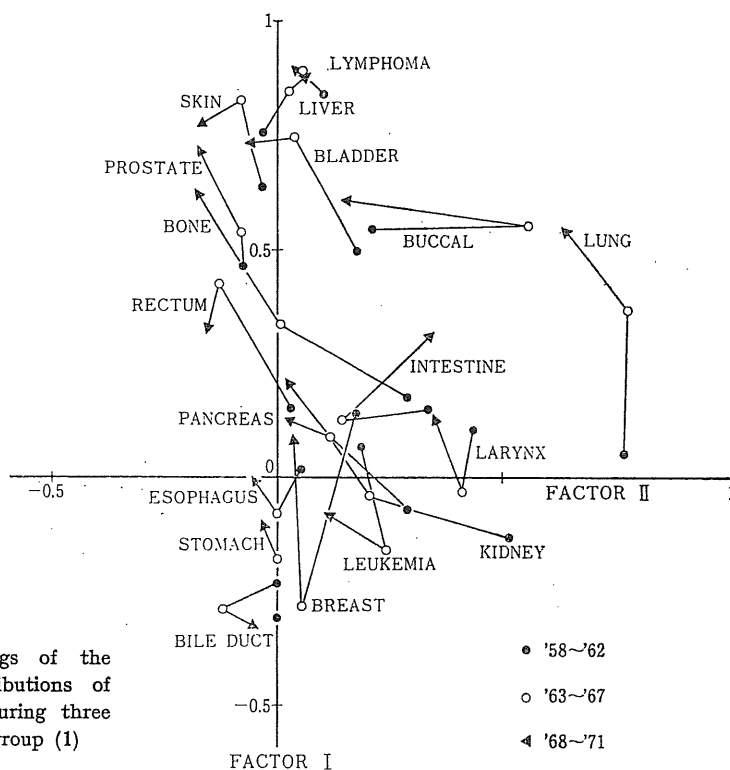


Fig. 1 Factor loadings of the geographical distributions of 18 cancer sites during three periods for males group (1)

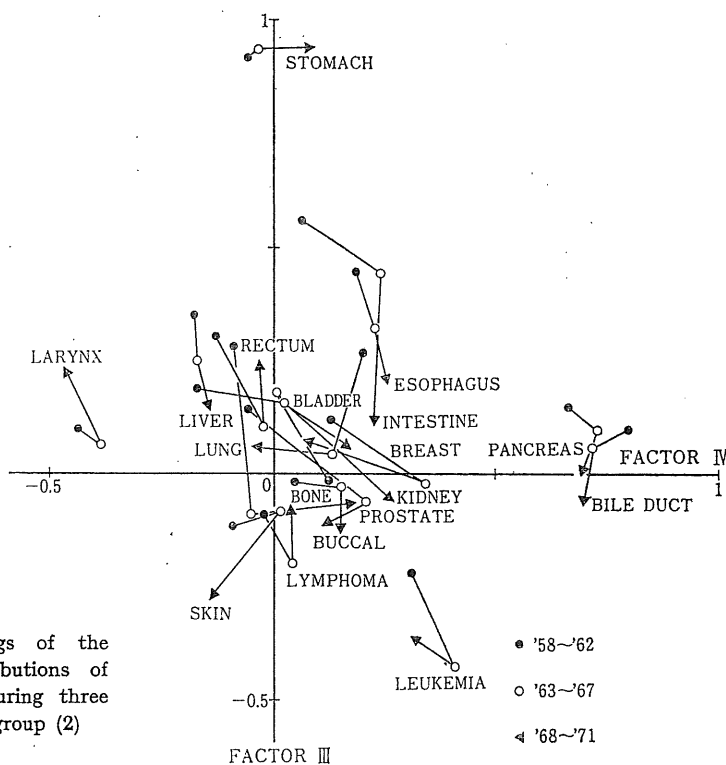


Fig. 2 Factor loadings of the geographical distributions of 18 cancer sites during three periods, for males group (2)

Fig. 6 (males) and Fig. 10 (females) that cancer of sites with high loadings on the first factor occurs more frequently in the southern parts of Japan as Kyushu, Chugoku and Shikoku districts in both sexes.

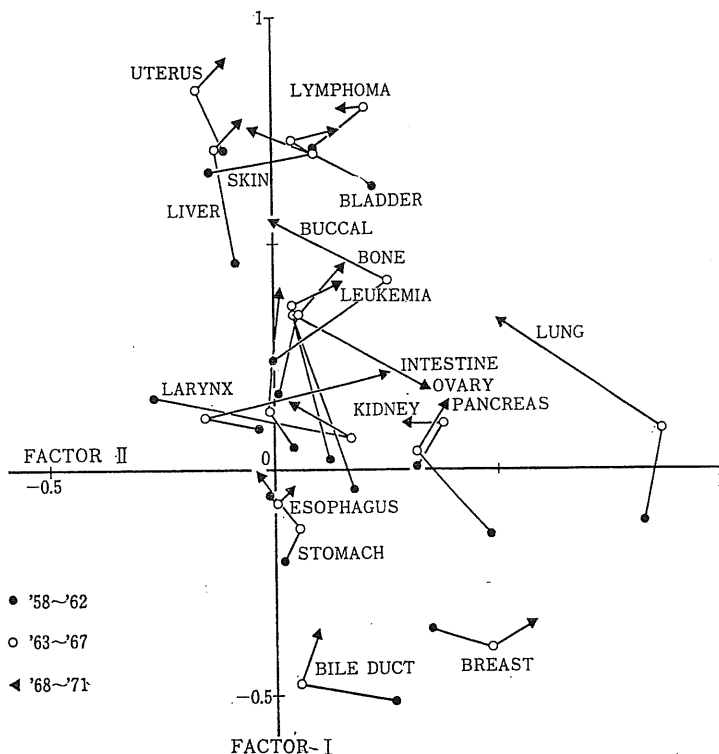
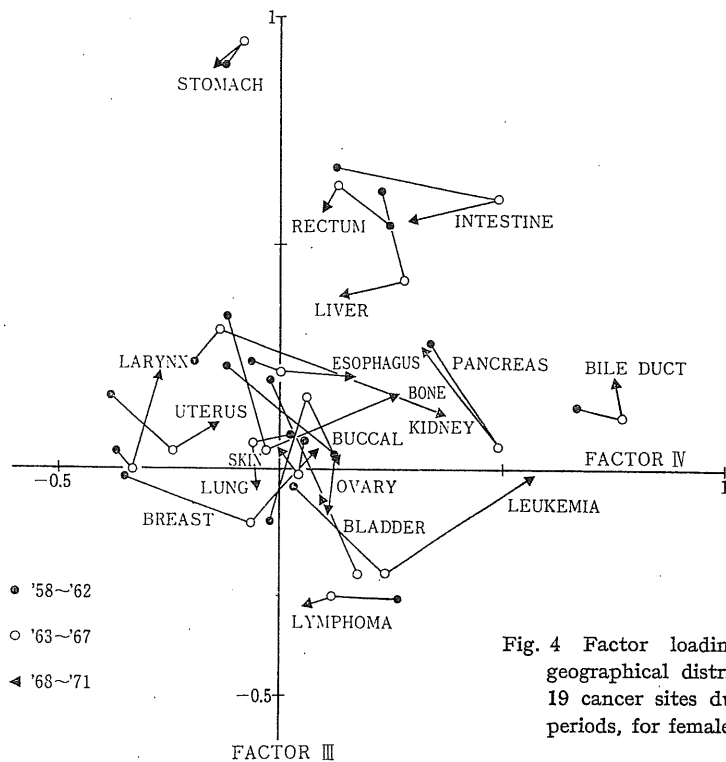


Fig. 3 Factor loadings of the geographical distributions of 19 cancer sites during three periods, for female group (1)

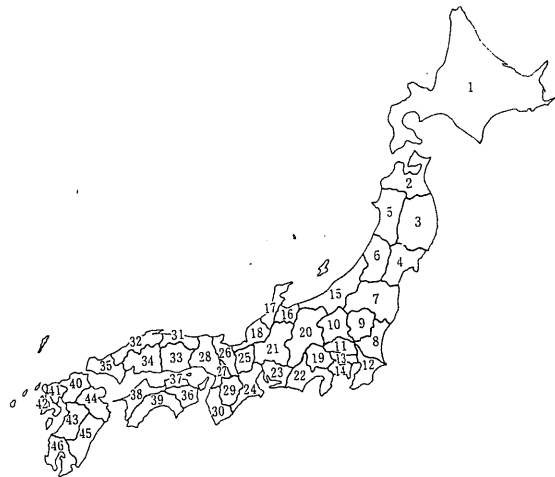
*Second factor* (degree of contribution: 9.7% for males and 10.6% for females)

The degrees of contribution of this factor are smaller in both sexes as compared with those of the first factor, but the exception is lung cancer of males which has high factor loadings. As is shown in Fig. 7 and 11, factor scores plot highest in Tokyo (13) and Kyoto (26), followed by Osaka (27), Kanagawa (14) and Hyogo (28) where lung and breast cancers occur frequently. Since each of the five prefectures mentioned above comprises one of the so-called six biggest cities of Japan, the occurrence of the cancers with high loadings on the second factor might be related to the urbanization or industrialization.



| List of 46 prefectures |              |              |              |
|------------------------|--------------|--------------|--------------|
| 1 Hokkaido             | 13 Tokyo     | 25 Shiga     | 37 Kagawa    |
| 2 Aomori               | 14 Kanagawa  | 26 Kyoto     | 38 Ehime     |
| 3 Iwate                | 15 Niigata   | 27 Osaka     | 39 Kochi     |
| 4 Miyagi               | 16 Toyama    | 28 Hyogo     | 40 Fukuoka   |
| 5 Akita                | 17 Ishikawa  | 29 Nara      | 41 Saga      |
| 6 Yamagata             | 18 Fukui     | 30 Wakayama  | 42 Nagasaki  |
| 7 Fukushima            | 19 Yamanashi | 31 Tottori   | 43 Kumamoto  |
| 8 Ibaragi              | 20 Nagano    | 32 Shimane   | 44 Oita      |
| 9 Tochigi              | 21 Gifu      | 33 Okayama   | 45 Miyazaki  |
| 10 Gumma               | 22 Shizuoka  | 34 Hiroshima | 46 Kagoshima |
| 11 Saitama             | 23 Aichi     | 35 Yamaguchi |              |
| 12 Chiba               | 24 Mie       | 36 Tokushima |              |

Fig. 5 Japanese map with the names of 46 prefectures excluding Okinawa



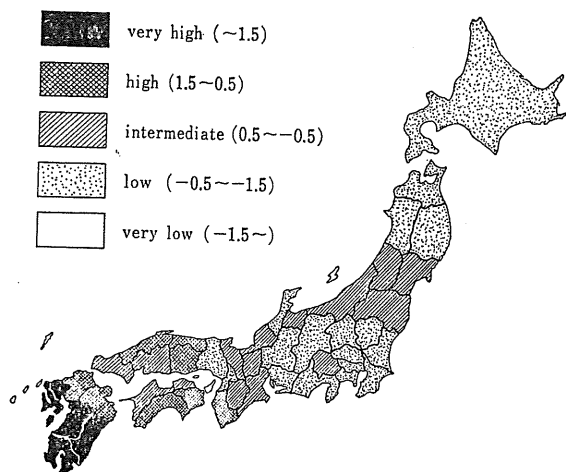


Fig. 6 Geographical distribution of the mortality rates of the first factor (for males)

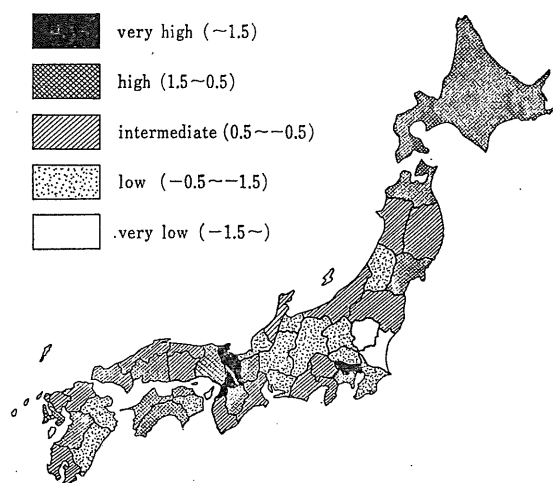


Fig. 7 Geographical distribution of the mortality rates of the second factor (for males)

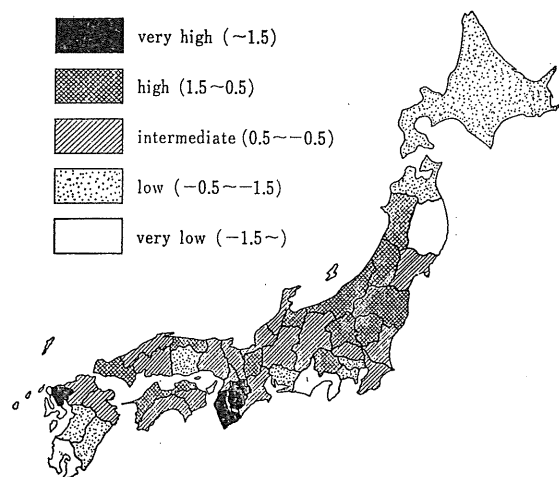


Fig. 8 Geographical distribution of the mortality rates of the third factor (for males)



Fig. 9 Geographical distribution of the mortality rates of the fourth factor (for males)

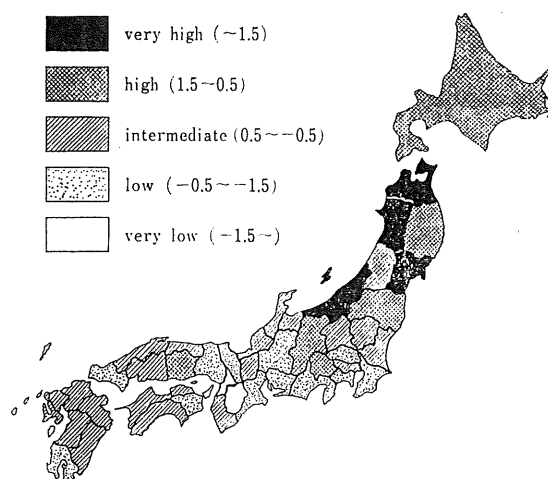


Fig. 10 Geographical distribution of the mortality rates of the first factor (for females)

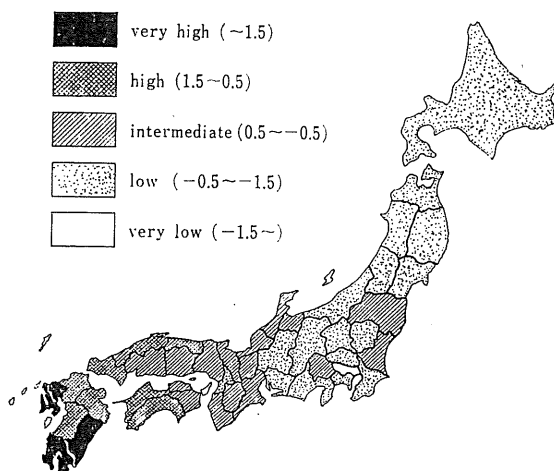
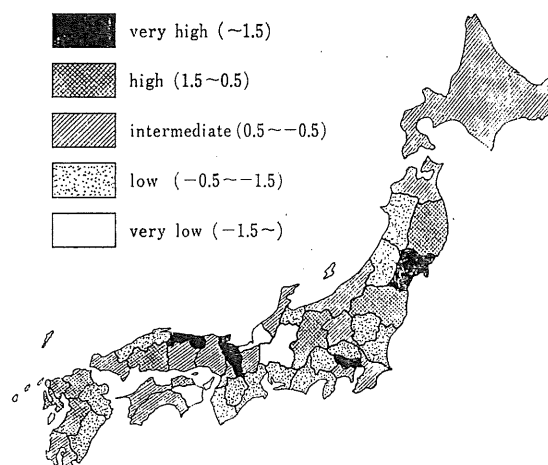


Fig. 11 Geographical distribution of the mortality rates of the second factor (for females)



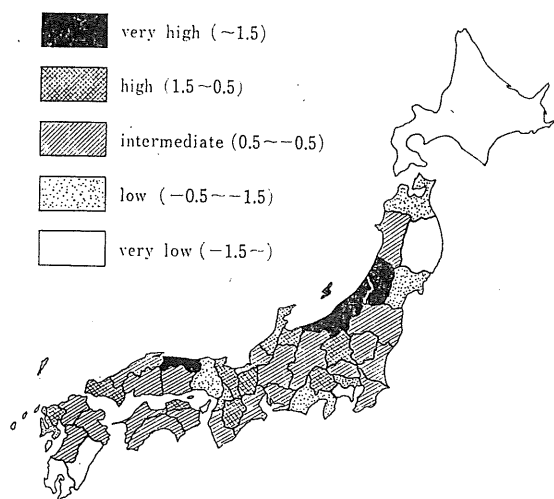


Fig. 12 Geographical distribution of the mortality rates of the third factor (for females)

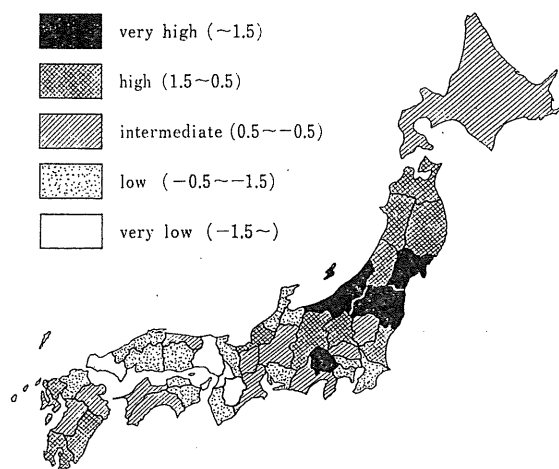


Fig. 13 Geographical distribution of the mortality rates of the fourth factor (for females)

*Third factor* (degree of contribution: 13.0% for males and 17.5% for females)

Stomach cancer is plotted highest on this factor for both sexes. Such tendency is observed especially in males. The locations of cancers of the digestive organs such as intestine, rectum and liver move nearest to the location of stomach cancer as the calendar year proceeds on. From Fig. 8 and 12 the prefectures plotted high on these factors are named Yamanashi (19), Niigata (15), Nara (29), Tottori (31), Toyama (16) and Saga (41) in the order of magnitude. It is apparently recognized that the regional distribution of the third factor resembles that of stomach cancer quite well.

*Fourth factor* (degree of contribution: 13.3% for males and 11.1% for female)

Pancreas and bile duct cancers plot high on this factor for both sexes. Leukemia is located high only in the third period for females. It is interesting to note that male larynx cancer is plotted high in the opposite direction along this axis as compared to the locations of the cancers mentioned above. This fact might suggest a negative relationship

between larynx cancer and pancreas or bile duct cancers. As for factor scores, Miyagi (4), Aomori (2), Iwate (3) and Niigata (15) plot high on this factor.

The diagrams for the fifth factor are omitted here. It is only esophageal cancer that is plotted highly on this factor for both sexes. It can be surmised that the distribution of esophageal cancer might be almost independent of that of the other cancer sites.

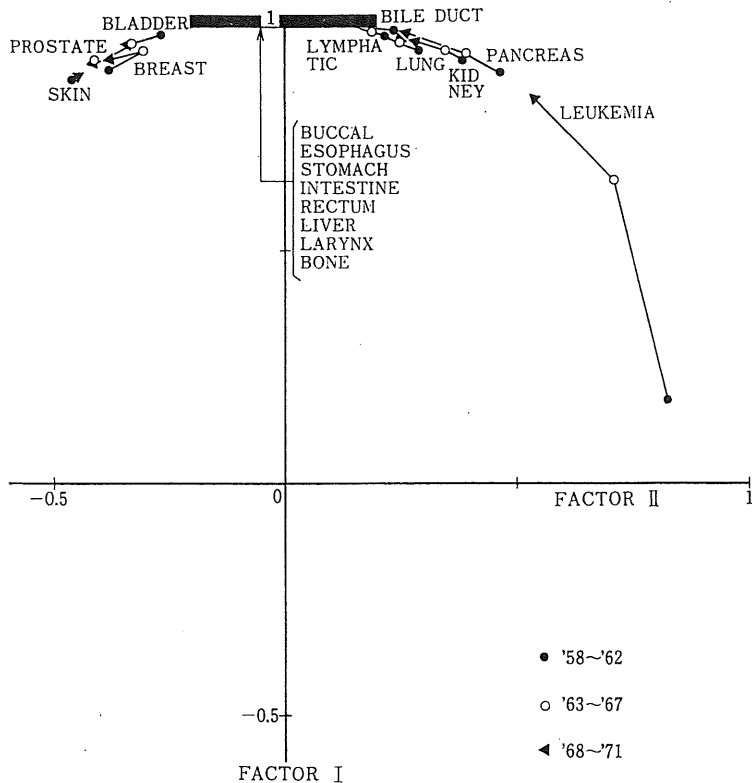


Fig. 14 Factor loadings of the 5 years age group distributions of 18 cancer sites with three periods for males

Up to now, we have already shown that five independent factors are extracted from the mutual correlations among the site-specific mortality rates on various cancers and that these factors were found commonly in both sexes, although slight discordance was noticed in some factors. Besides these five factors, there exist factors associated with rectum cancer for males, and ovary and bone cancers for females respectively.

### (3) Factor analysis of the age distribution

Basing on the data of the five years age group distribution of cancer mortalities of various sites, correlation coefficients among any pairs of 54 variables for males and 57 variables for females were calculated, using the same procedure to calculate the correlation coefficients between the different geographical distributions, and then factor analysis by varimax method was applied to these two correlation matrices respectively.

As a result two significant factors were extracted for both sexes, and the contribution ratios of the first factor and the second factor were 81.2 % and 9.3% for males and 80.9% and 10.3% for females respectively. The factor loadings of the factors of male were shown in Fig. 14 and those of females in Fig. 15 respectively. It is shown from these figures that most of the cancer sites have high factor loadings on the first

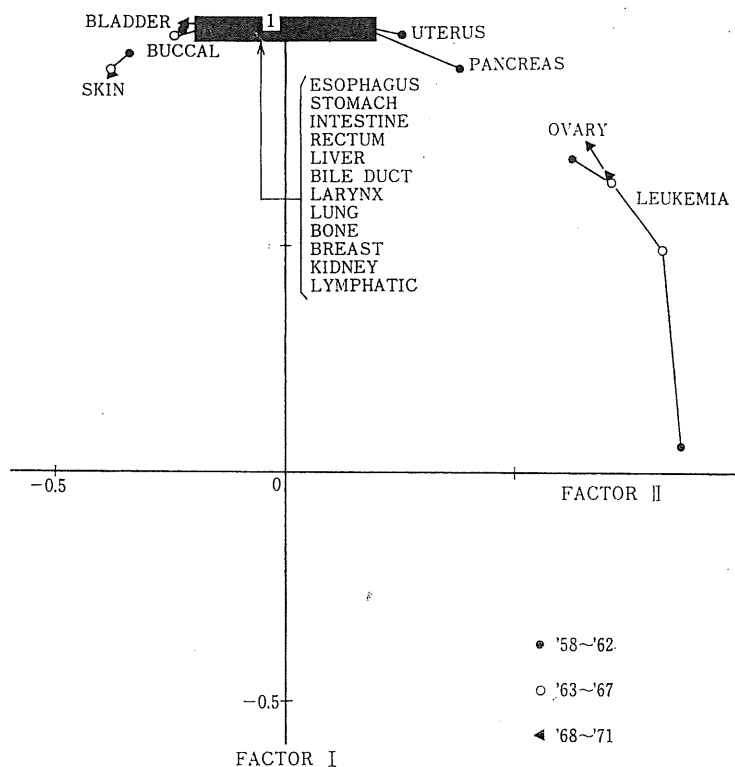


Fig. 15 Factor loadings of the 5 years age group distributions of 19 cancer sites with three periods for females

factor, excluding leukemia which has low factor loadings for both sexes especially in the first periods. On the contrary, in terms of the second factor the plots of various sites of cancer are grouped forming three separate clusters; i.e., the sites with high positive loadings, those with negative loadings and the intermediate. Leukemia and pancreas cancer of both sexes are grouped into the first cluster, and skin and bladder cancers of both sexes into the second cluster, while ovary cancer has high leadings on the second factor. Furthermore, we show in Fig. 16 and 17 the factor scores concerning the ages with five year grades based on the two factors of both sexes.

The curve of the first factor implies a non-decreasing curve with the accelerating point at about 55-59 years age group, which is supposed to reflect the postnatal carcinogenic factor. The curve of the second factor implies a curve with one sign changing point at 65-69 years age group for males and at 60-64 years age group for females. The result implies that the age distribution of the mortality of each site of cancer is

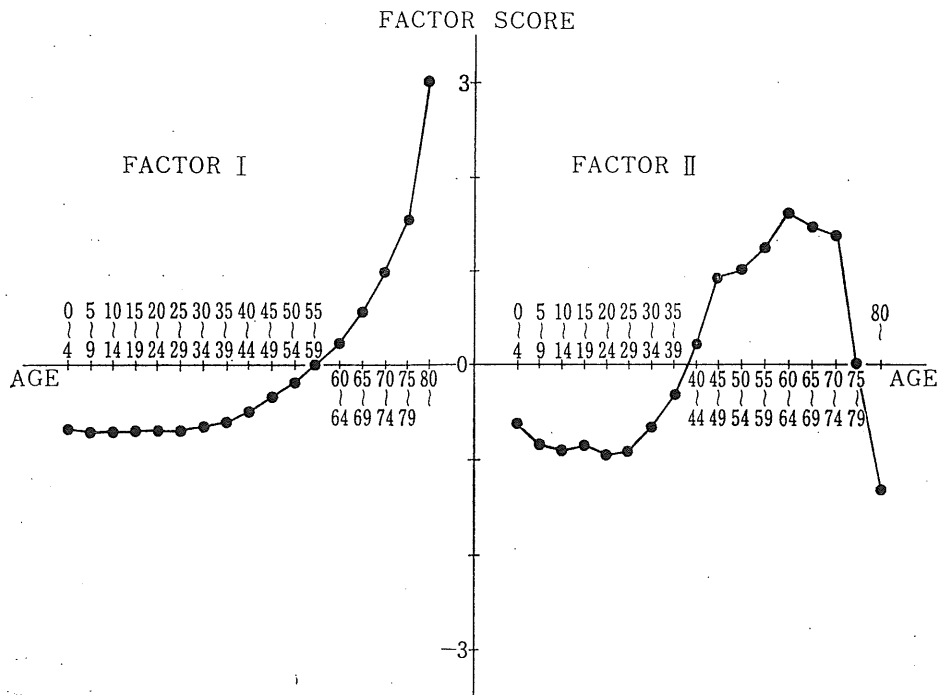


Fig. 16 Curve of factor scores of the 5 years age group distributions of 18 cancer sites with three periods for males

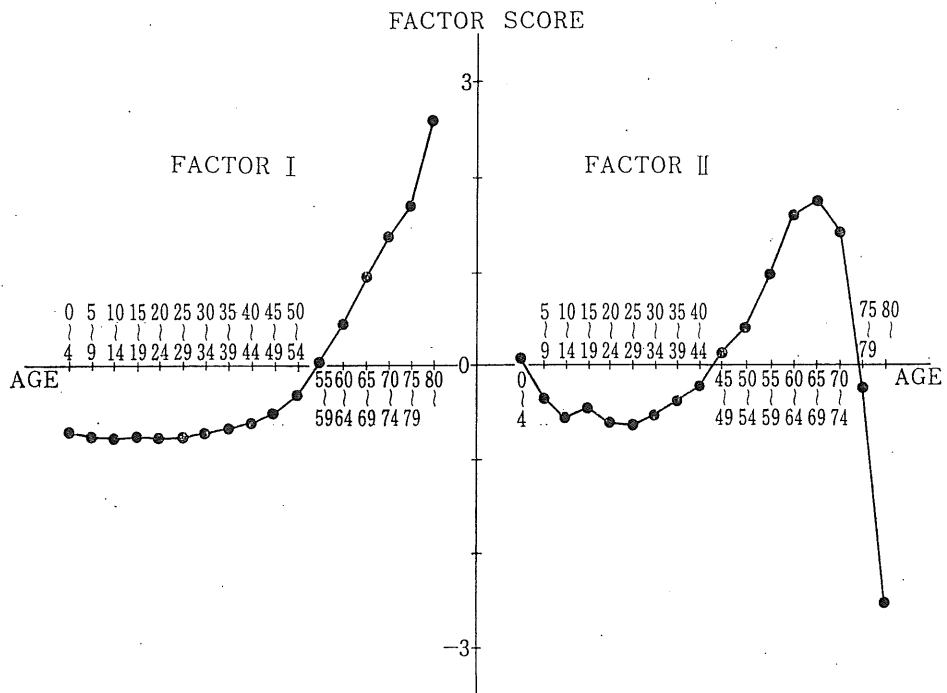


Fig. 17 Curve of factor scores of the 5 years age group distributions of 19 cancer sites with three periods for females

decomposed into the weighted sum of the first and the second factor scores, using factor loadings shown in Fig. 14 and 15 as weights. We note from Fig. 14 through Fig. 17 that the age distribution curve of the mortality rate in the first group (leukemia, pancreas and ovary) resembles the curves of the second factor scores, while that in the second group (skin, bladder, prostate) resembles the reciprocal curve of the second factor score. For both sexes, it can be shown in Fig. 14 and 15 that the age distribution curve of mortality rate of leukemia skews over the younger age group at the first period ('58~'62), while it approaches to that of the other cancers as the calendar year proceeds on.

#### 4. Discussion

##### (1) On the method of obtaining the standardized mortality rate

Apparently, it is better to calculate the standardized mortality rates by means of the direct method when the comparison or analysis of these rates is the target of this study. But such a calculation is not advisable especially for the mortality of various cancer sites, in which the number of deaths is very low and, therefore, a slight change of the figures may affect the result of the calculation to a greater extent. So far, the standardized mortality rates based on the direct method have been reported by Segi (1972), with respect to only five cancer sites as stomach, lung, uterus and breast cancers and leukemia. It is, therefore, preferred to calculate the corrected mortality rates by means of indirect method, since the age distribution of various cancer sites usually skews over the higher age group. In view of this situation, it may sound preferable to calculate the specific mortality rates in the special populations with limiting range of age. For an example, Segi (1972) proposed the specific mortality rate, using only the populations over 40 years of age, but it does not hold good in the case of leukemia, which distribution skews over the younger age group. Because of these reasons, a tentative method was proposed in this study in order to calculate the specific mortality rates based on the average and standard deviations of the age of death calculated separately with respect to various sites of cancer. Recently, we also performed the same analysis using the mortality rate based on the indirect method (Inaba et al., 1978) and we are going to compare the validity and efficiency of the method proposed in this paper with the method based on the indirect method.

##### (2) On the result of correlation coefficients and factor analysis with regard to the geographical distribution

According to the results described above, the correlation coefficients between any pairs of the cancer sites are comparatively stable throughout three periods with a few exceptions.

The results confirm the results reported by Matsuyama (1966), in spite of the fact that his data were derived from the other source (1959), and at the same time, somewhat different method from ours was used for calculating the mortality rate. Furthermore, Table 3 and 4 suggest the definite stability for each cancer site in terms of the geographical distribution throughout three periods except for male breast cancer as

well as female leukemia. The fact that the geographical distribution with respect to the mortality rates of various cancer sites did not fluctuate significantly for such long periods might well suggest a persistence of some environmental factors provoking their occurrence. It should be noted, however, that the correlation coefficient shown in Table 3 and 4 are likely to be overestimated for such cancer sites with high mortality rate as stomach cancer. In view of epidemiological methodology, correlation coefficients should be corrected by means of the information associated with the total number of death resulting from each cancer site in each period when they should be compared between two different periods.

Next, we shall discuss what each extracted factor implies in view of the factor loadings and factor scores as seen from Fig. 1. through Fig. 12.

Since the result of factor analysis reveals that various sites of cancer are grouped into five clusters with respect to their geographical distributions, it can be surmised that same type of factors function as the etiological agents in some of the cancers. However, similarities in the geographical distribution of two diseases do not always imply to share common cause. Hence, when we have some knowledge about the etiology of a disease in a certain specified cluster, it is a useful initial step of hypotheses to postulate a common etiological cause for the diseases located in that cluster. Out of the five factors extracted here, it is the first factor that interests us most. The fact that bone, bladder, buccal and skin cancers and lymphoma as well as cancers of genital organs excluding ovary are grouped in one cluster, suggests the possible relation of the first factor to the high atmospheric temperature. It is well known that an etiological factor of skin cancer is closely associated with heavy sunlight exposure. An alternative fact that interests us is the conspicuous chronological variability in the geographical distribution of lung cancer as shown in Fig. 1 and 3. The geographical distribution of lung cancer for both sexes approximates those of bone, liver and bladder cancers with characteristic association to the high atmospheric temperature. However, this fact does not conform with the well known hypothesis that environmental pollution and heavy consumption of tobacco are main causes of lung cancer.

For these reasons, it goes without saying that detailed studies should be undertaken soon in order to uncover the result. But this epidemiological evidence may well suggest that the recent increases of the death associated with lung cancer caused the change of the geographical distribution in terms of mortality rate. Next, we shall consider Fig. 2 and 4. It is interesting to note that the connection of the distribution of stomach cancer with that of the cancers of other digestive organs is much accentuated in females as compared to males. This reminds us of the hypothesis proposed by Wynne (1968) that age dependent sex ratio curve of stomach cancer is closely related to the difference in the quantity of energy consumption or the quantity of the total foods intake (Kurita, 1972).

It is also seen from Fig. 2 and 4 that bile duct cancer, which is anatomically connected with liver cancer, resembles pancreas far more than liver cancer in its geographical distribution. This might suggest also an interesting hypothesis in further exploring cancer epidemiology.

### (3) On the result of factor analysis of the age distribution

It is interesting to note that the curve of the factor scores of the first factor resembles the Type 1 curve proposed by Higgins (1969), which is characterized as the age distribution of death resulted from the continuous exposition to the specific exogenous factors such as radiation or toxic chemical substances etc., and that the curve of the factor scores of the second factor resembles Type 2 curve which is also characterized as the age distribution of death resulted from the exposition to them only during the first stage of human life. As an alternative interpretation of the second factor, we have already stated that the second factor is supposed to be the prenatal carcinogenic factor. So far, it is said that the morbidity rates of the various sites of cancer increases proportionally to the logarithmic value of the age. But from our result, we may say that the above fact is valid only in some cancers of specified sites, and the other cancers have different age distributions in the strict sense.

If it is admitted, as described above, that some of the carcinogenic factors such as exposition of radiation or toxic chemical substance, could be split into postnatal (factor I) and prenatal (factor II) ones, the age-dependent curve of the latter would be very interesting in terms of multi-stage theory of carcinogenesis proposed by Armitage & Doll (1954). Naturally, their mathematical formula, which may be applicable only to the simple age distribution of cancer incidence rate, should be modified according to the biological implication of factor II. Mathematically, simple modification of multistage theory may reveal the fact that the age distribution of mortality of a cancer site could be approximated as a Weibull distribution. However, our result suggest a possibility that it can be approximated as a kind of compound Weibull distribution which contains at least two independent distribution functions. Hence, it is expected that the basic concept of the multiple stage theory is also applicable to this case, and somewhat different result could be reached by applying our result to the multi-stage theory. Although we are not ready to give the detailed discussions on this matter, we think it would be of value to try to elucidate the profound meanings of the extracted factors in view of the multi-stage theory.

Finally, some discussion is required concerning the drastic change of the age distribution of leukemia, which is recognized in Fig. 14 and 15. Segi & Matsuyama (1969) pointed out that the mortality rate of leukemia in Japan was far lower among the higher age group, in spite of the fact that it was almost the same with West Germany and Denmark among children and adults below age 45. At the same time they also noticed that the increasing trend of mortality rate of leukemia in Japan during 1954-65 was more conspicuous in the higher age group than in the lower. This report confirms that the average age of death from leukemia continued to increase in Japan up to 1971, which is reflected in Fig. 14 and 15 indicating that the pattern of the occurrence of leukemia is approaching to those of the other cancers as the calendar year proceeds. One of the reasons to explain the above fact lies in a general tendency recognized throughout the world that the mortality rate of leukemia in higher age group increases as the environmental leukogens prevail, while that of the pediatric leukemia of congenital nature remains substantially constant (Fraumeni et al., 1964).



And this increasing velocity is greater in Japan than in European countries so that the mortality rate of leukemia among aged people is catching up with that of the latters. The delayed exposure to the leukogens among Japanese as compared with U.S. whites was proposed by Fraumeni and Miller (1967) and this hypothesis is also explanative to the relationship between Japan and European countries described above. The second reason is also universal one. Recently, the treatment of leukemia, especially acute form which is usually prevalent among children, made a progress so marvelously that the life expectancy of such cases was significantly prolonged (Axtell et al., 1974 and Zippin et al., 1971).

#### (4) On the utility of factor analysis

Finally, we should discuss the utility of factor analysis as a method of an epidemiological study. Factor analysis is said to be a method to uncover the latent factors which may be hidden behind the data into light, and so far the fruitful results have been obtained by applying it to the psychological data such as personality test or ability test. Unfortunately, however, we can rarely find the case where the latent factors are really obtained from the empirical data except for few cases. Hence, it is better to employ the idea such that utility of factor analysis in epidemiological study lies in reducing the many confound informations into smaller ones and obtaining the hypothetical factors that may be considered as clues to develop the unknown factors into light. Recently, it is widely accepted that many confounding factors, which are correlated with each other, are supposed to be related to etiology. Hence, in view of this, application of factor analysis would be useful providing informations in order to detect the hypothetical factors that may be closely connected with the onset of some diseases.

As an alternative method of finding some clusters of various sites of cancers, Burbunk (1971) recommends cluster analysis for reason that its application would produce the successive hierarchical structure of many cancer sites. But it does not necessarily follow that such hierarchical structure corresponds to pathogenesis of cancer in the human body. In addition, we should mention that clusters produced from application of cluster analysis are usually interdependent, and result of cluster analysis can hardly produce informations connected with the mutual relationship among the extracted clusters. Hence, we might say that factor analysis prevails over cluster analysis in the utility of the extracted clusters, but we do not intend to deny the usefulness of cluster analysis at all. If we apply factor analysis and cluster analysis to a set of data simultaneously, and the compare the results in an appropriate method, it would never fail to be a powerful method to develop the etiological hypothesis.

To sum up, the results obtained by means of factor analyses of the geographical distributions as well as the age distribution provide a kind of synthesized information contained in the raw data, and they are nothing but clues to develop the unknown etiology of some diseases. But we should admit that various methods of behaviormetrics would never fail to prove some important clues to elucidate the unknown etiology of the various sites of cancer.

## Acknowledgements

The authors would like to express sincere gratitude to Professor A.T.A. Learmouth of Walton University of Great Britain, the specialist of medical geography, for reading the manuscript carefully and giving us useful comment to it.

## REFERENCES

- Armitage, P. & Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis, *British Journal of Cancer*, **8**, 1-17.
- Burbank, F. (1972). A sequential space-time cluster analysis of cancer mortality in the United States: Etiologic implications, *American Journal of Epidemiology*, **95**, 393-416.
- Burkitt, D.P. (1969). Related disease-related cause, *Lancet*, **2**, 1229-1233.
- Burkitt, D.P. (1970). Relationship as a clue to causation, *Lancet*, **2**, 1237-1240.
- Dunham, L.J. & Bailer, J.C. (1970). World maps of cancer rates and frequency ratios Part 1, *Industrial Medicine*, Vol. **39**, No. **2**, 89-100.
- Axtell, L.M. & Meyers, M.H., Christine, B. & Linden, G. (1974). Trend in survival and classifications of patients with leukemia diagnosed 1940-69, *Journal of National Cancer Institute*, **38**, 593-605.
- Fraumeni, J.F. & Miller, R. (1967). Epidemiology of Human Leukemia: Recent observations, *Journal of National Cancer Institute*, **38**, 593-605.
- Fraumeni, J.F. & Wagner, J.K. (1964). Changing sex differentials in leukemia, *Public Health*, Rep. **79**, 1093-1100.
- Higgins, J. (1969). Present trends in cancer epidemiology, *Canadian Cancer Conference*, **9**, 40.
- Inaba, Y., Takagi, H., Yanai, H. & Matsubara, J. (1975). Ekigaku chosa ni okeru chiiki oyobi nenrei inshi ni kansuru kenkyu (A study on the geographical and age factors in the epidemiological survey), *Japanese Journal of Public Health*, **22**, 83-93.
- Inaba, Y., Takagi, H. & Yanai, H. (1978). Gan no chiriteki-bunpu no inshibunseki (Factor analysis of the geographical distribution of the mortality rates in Japan), To appear in *Japanese Journal of Public Health*.
- Kurita, H. (1972). Sei-nenrei-betsu ni mita igan no tokeiteki kento to sono kosatsu (Epidemiological consideration of sex and age difference for gastric cancer mortality), *Japanese Journal of Public Health*, **19**, 457-461.
- Matsuyama, T. (1966). Gan no shibo-ritsu no sokan-bunseki (Correlational study of the geographical distribution of cancers), *Tohoku Igaku*.
- Ministry of Health & Welfare (1959). Akusei shinseibutsu shibo-tokei (*Mortality statistic of Malignant Neoplasm from 1951 to 1957*), Tokyo.
- Ministry of Health & Welfare (1973). Akusei shinseibutsu shibo-tokei (*Mortality Statistic of Malignant Neoplasm from 1958 to 1971*), Tokyo.
- Segi, M. (1965). Nihon ni okeru akusei shinseibutsu shibo-tokei (Cancer Mortality in Japan), Tohoku University, Sendai.
- Segi, M. & Kurihara, M. (1972). Nihon ni okeru todofuken-betsu gan teiseishibo-ritsu, tokushu shibo-ritsu (Age-adjusted death rates and specific death rates of cancer in prefectures of Japan, 1958-1967), Segi Institute of Cancer Epidemiology.
- Segi, M. & Kurihara, M. (1966). Cancer Mortality for selected sites in 24 countries, No. 4.
- Segi, M. (1969). Cancer Mortality for selected sites in 24 countries, No. 5, Sendai.
- Segi, M. & Matsuyama, T. (1969). Hakketsubyo no chiriteki ekigaku (Geographical epidemiology of leukemia), *Saishin Igaku*, **24**, 757-763.
- Takeuchi, K. & Yanai, H. (1972). *Tahenryo haiseki no kiso (Essentials of Multivariate Analysis)*, Toyo Keizai Press, Tokyo.
- Wynne, G.G. (1968). The sex ratio in gastric cancer and hypothetical consideration to etiology, *British Journal of Cancer*, **22**, 163-169.
- Zippen, C., Cutler, S.J., Reeves, W.J. & Lum, D. (1971). Variation in survival among patients with acute lymphocytic leukemia, *Blood*, **37**, 59-72.

(Received February, 1978)