Bulletin of the Biometric Soc. of Japan
Vol.8 No.2(1987).pp.51-60

A MODE PREDICTION FOR THE NUMBER OF PATIENTS WITH AN INCURABLE DISEASE IN THE NATIONWIDE EPIDEMIOLOGICAL SURVEY

Hirofumi TAKAGI

Division of General Education, St. Luke's College of Nursing
10-1 Akashi-chou, Chuu-oh-ku, Tokyo, 104 Japan

SUMMARY

We propose a new method that provides the predictive mode value for the number of patients with the specified incurable disease by the use of data collected by the nationwide epidemiological survey in Japan. Since the survey is essentially incomplete, there are always many unrecovered subjects, that is, hospitals selected for the purpose. Therefore, we must estimate the number of patients in unrecovered hospitals. Polynomial distribution is assumed to describe the observed distribution of the number of patients recovered in the survey with respect to each hospital. We attempt to estimate the total number of patients including unknown subjects under the assumed distribution. To explain the method in detail, a real example is given, and alternative methods are also discussed.

1. INTRODUCTION

Since 1972, the Ministry of Health and Welfare of Japan has promoted the clinical and epidemiological studies on many incurable diseases called "Tokutei-Shikkan (Specified Incurable Diseases)" including those in the various fields in medicne. The estimation of the number of patients with speific disease is one of the epidemiological studies. Therefore, each study group has conducted the nationwide survey, usually by mail, on the selected subjects; i.e., specified clinical departments of thousands of hospitals with 200(or 300) beds or more. Nevertheless, in general, since the percentage of recovery for the research is 70% or less as the greatest value, we must estimate the number of unknown patients in unrecovered rest hospitals by the use of obtained data of which informations are only; 1)the number of hospitals as the subjects of the survey by prefecture, and 2)the observed number of patients by sex in each hospital by prefecture.

Usually, the estimation is carried out by Yanai's method as follows(Yanai, et al.,1975).

Let M_i be the number of subject hospitals of the survey in prefecture i for i=1,2,...,K. If there are N_i hospitals answered and the total number of patients reported is t_i, the recovery rate $R_i=N_i/M_i$, and the observed incidence rate $y_i=t_i/P_i$, where P_i is the number of population in prefecture i. Consider a simple linear regression of the incidence rate, y, based on the recovery rate, R, and the estimate for the true incidence rate μ_i is

Key words: polynomial distribution, epidemiology, rare disease, multiple hypergeometric distribution, mode prediction, Newton-Raphson method.

given by

$$\hat{\mu}_i = y_i + (1-R_i)b, \tag{1}$$

where b is a regression coefficient of the recovery rate estimated as

$$b = \sum_{i=1}^{K}(R_i-\bar{R})(y_i-\bar{y}) \Big/ \sum_{i=1}^{K}(R_i-\bar{R})^2, \tag{2}$$

where $\bar{y}$ and $\bar{R}$ are mean values of the observed incidence rate and the recovery rate respectively, and K is the number of prefectures. Under the assumption of normal distribution for the error term, variance of $\hat{\mu}_i$, $V(\hat{\mu}_i)$, is given by

$$V(\hat{\mu}_i) = (1-r_{Ry}^2)\sum_{i=1}^{K}(y_i-\bar{y})^2 \left[1+(1-R_i)^2 \Big/ \sum_{i=1}^{K}(R_i-\bar{R})^2\right] \Big/ (K-2), \tag{3}$$

where r_{Ry} is a simple correlation coefficient between variables R and y.

True number of patients, τ, in all of subject hospitals and its variance are estimated respectively as

$$\hat{\tau} = \sum_{i=1}^{K} P_i \hat{\mu}_i, \text{ and } V(\hat{\tau}) = \sum_{i=1}^{K} P_i^2 V(\hat{\mu}_i). \tag{4}$$

Yanai's method has been applied to many epidemiological studies in Japan. However, it is not applicable frequently in practice because of negative correlation of the observed incidence rate to the recovery rate. Moreover, potentially, Yanai's method includes the assumption that no difference exists among regions for the incidence rate; i.e., the expected value of $\hat{\mu}_i$ is the same value for all of prefectures. Although this assumption cannot be thought to be valid obviously, in order to avoid the above problem on negative correlation, some modified methods are proposed by Nakae et al. (1985) and Mizuno et al.(1986), in which a regression line is determined to intersect the origin; i.e., $y_i=0$ if $R_i=0$. Nevertheless, such modification is not seemed to be essential because it is not ensured that two variables of interest are positively correlated in the population under study, even if the recovery rate and the observed number of patients may have a positive correlation regionally to some extent. Furthermore, it remains that the regional equality of the incidence rate is still assumed in such models. Therefore, it seems that these regression typed methods are improper theoretically to estimate the number of patients all over the country.

In this paper, we consider the distriution of the number of patients with respect to each hospital. Polynomial distribution is introduced to explain the distrbution of the number of patients in each hospital, and the method of computing the estimation of the number of patients is described as well as the variance of the estimated value. A real example is given to clarify the method in detail, and to compare the results with those by Yanai's method. Finally, we discuss alternative methods for the future studies.

2. METHODS

2.1. Polynomial distribution and log probability

Consider N samples recovered from M subjects of the survey. Assumed that the sampling is randomized. If n_i samples report i ptients (i=0,1,...,k), N is sum of n_i for all i. The problem lies in estimating the true number of subjects η_i with i patients. η_i can be expressed as $\eta_i = n_i + \delta_i$, where δ_i indicates the non-negative number of unkown subjects. Therefore, if n_i is determined, the problem becomes to estimate δ_i instead of η_i under the restricted condition as follows.

$$\sum_{i=0}^{K} \delta_i = M - N. \tag{5}$$

Suppose that the frequencies $(n_0, n_1, n_2, \ldots, n_k)$ of each patient number are obtained from the polynomial distribution with probabilities $(\pi_0, \pi_1, \ldots, \pi_k)$. As is well known, maximum likelihood estimator p_i for π_i is given by

$$p_i = n_i / N, \quad \text{for all i.} \tag{6}$$

Then the frequencies of unknown subjects with i patients is expected as $\overline{\delta_i} = (M-N)p_i$. Even though it may be thought that using expected values is enough to estimate the total number of patients, the expectation does not always have the highest probability with regard to the occurrence of event of interest. Consider what combination of unknown subjects has the highest probability under some assumptions. Suppose that the frequencies $(\delta_0, \delta_1, \delta_2, \ldots, \delta_k)$ have the polynomial distribution with the probabilities $(p_0, p_1, p_2, \ldots, p_k)$ estimated by the use of sample data. For M-N unkown subjects, the log probability $\mathfrak{L}$ of the polynomial distribution is given by

$$\begin{aligned} \mathfrak{L} &= \ln(M-N)! - \sum_{i=0}^{K} [\ln \Gamma(\delta_i + 1) + \delta_i \ln p_i] \\ &= \ln(M-N)! - (M-N)\ln N - \sum_{i=0}^{K} [\ln \Gamma(\delta_i + 1) + \delta_i \ln n_i], \end{aligned} \tag{7}$$

where $\ln \Gamma(z)$ is the log gamma function defined as

$$\ln \Gamma(z) = \int_0^\infty \left[(z-1)e^{-t} + \frac{e^{-tz} - e^{-t}}{1 - e^{-t}}\right] \frac{dt}{t}. \tag{8}$$

Substitute the restriction of equation (5), equation (7) is rearranged as follows.

$$\begin{aligned} \mathfrak{L} = &- \sum_{i=1}^{K} \ln \Gamma(\delta_i + 1) - \ln \Gamma(M - N - \sum_{i=1}^{K} \delta_i + 1) + \sum_{i=1}^{K} \delta_i \ln n_i + (M - N - \sum_{i=1}^{K} \delta_i) \ln n_0 \\ &+ \ln(M-N)! - (M-N)\ln N. \end{aligned} \tag{9}$$

Using equation (9), we easily find the predictive value $\hat{\delta}_i$ for i=1,...,k

to maximize $\mathcal{L}$ by means of the simplex method. However, Newton-Raphson method is usually more effective to solve the problem. Therefore, the first and second derivatives of $\mathcal{L}$ with respect to δ_i are explained below. Note that the predictive value $\hat{\delta}_i$ may include decimal numbers even though δ_i in practice must be non-negative integer.

2.2. Differentiation of $\mathcal{L}$

Partial derivative of $\mathcal{L}$ with rspect to δ_x is given by

$$\frac{\partial \mathcal{L}}{\partial \delta_x} = \ln(n_x / n_0) - \psi(\delta_x+1) + \psi(M-N-\sum_{i=1}^{K}\delta_i+1),\ x=1,2,\ldots,k, \qquad (10)$$

where $\psi(z)$ is the first derivative of the log gamma function with respect to z as follows.

$$\psi(z) = \frac{d\ln\Gamma(z)}{dz} = \int_0^\infty \left[\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}}\right] dt. \qquad (11)$$

The second partial derivatives of $\mathcal{L}$ with respect to δ_x and the partial derivative with respect to δ_x of the partial derivative of $\mathcal{L}$ with respect to δ_y $(x \neq y)$ are given respectively by

$$\frac{\partial^2 \mathcal{L}}{\partial \delta_x^2} = -\psi'(\delta_x+1) - \psi'(M-N-\sum_{i=1}^{K}\delta_i+1),\quad x=1,2,\ldots,k, \qquad (12)$$

and,

$$\frac{\partial^2 \mathcal{L}}{\partial \delta_x \partial \delta_y} = -\psi'(M-N-\sum_{i=1}^{K}\delta_i+1),\quad x \neq y, \qquad (13)$$

where

$$\psi'(z) = \frac{d\psi(z)}{dz} = \int_0^\infty \left[\frac{te^{-zt}}{1-e^{-t}}\right] dt. \qquad (14)$$

Numerical analysis can be made using equations (10) through (14) by means of Newton-Raphson procedure. Frequently, on the iterative computations, one and/or more values calculating become negative. Therefore, in order to avoid this problem, we may use the transformation as $z_x=\ln(\delta_x)$, and after the outcomes are computed, we need the reverse transformaion to obtain non-negative estimates. However, since this problem on the numerical analysis is not main subject in this paper, we will continue to explain the predictive variance next.

2.3. Variance of predictive value

The variance of $\hat{\eta}_i$, $V(\hat{\eta}_i)$, is the same as the variance of $\hat{\delta}_i$ since $\hat{\eta}_i = n_i + \hat{\delta}_i$ if n_i is determined. The variance of $\hat{\delta}_i$ is given by

$$V(\hat{\delta}_i)=E[(\delta_i-\bar{\delta}_i)^2]+(\bar{\delta}_i-\hat{\delta}_i)^2=(M-N)p_i(1-p_i)+[(M-N)p_i-\hat{\delta}_i]^2, \quad (15)$$

since $E[(\delta_i-\bar{\delta}_i)^2]=(M-N)p_i(1-p_i)$, where $\bar{\delta}_i$ is the expected value as before.

When $\hat{\eta}_i$ indicates the predictive number of hospitals with i patients, since the estimate of total number of patitents, $\hat{c}_i$, is computed as the quantity i times $\hat{\eta}_i$, the variance of $\hat{c}_i$ is given by

$$V(\hat{c}_i)=i^2V(\hat{\eta}_i)=i^2V(\hat{\delta}_i), \quad \text{for all } i. \quad (16)$$

The variance of $\hat{c}$, the estimate of total number of patients for all i, is also given by

$$V(\hat{c})=E[\{\sum_{i=1}^{K} i(\eta_i-\hat{\eta}_i)\}^2]=E[\{\sum_{i=1}^{K} i(\delta_i-\hat{\delta}_i)\}^2]$$
$$=\sum_{i=1}^{K} i^2V(\hat{\delta}_i)+\sum_{i\neq j}\sum ij\cdot E[(\delta_i-\bar{\delta}_i)(\delta_j-\bar{\delta}_j)]+\sum_{i\neq j}\sum ij\cdot(\bar{\delta}_i-\hat{\delta}_i)(\bar{\delta}_j-\hat{\delta}_j). \quad (17)$$

Since $\bar{\delta}_i$ is the expected value, and the covariance of δ_i and δ_j is

$$E[(\delta_i-\bar{\delta}_i)(\delta_j-\bar{\delta}_j)]=-(M-N)p_ip_j, \quad i\neq j, \quad (18)$$

then equation (17) becomes as follows.

$$V(\hat{c})=\sum_{i=1}^{K} i^2V(\hat{\delta}_i)-(M-N)\sum_{i\neq j}\sum ij\cdot p_ip_j+\sum_{i\neq j}\sum ij\cdot(\bar{\delta}_i-\hat{\delta}_i)(\bar{\delta}_j-\hat{\delta}_j). \quad (19)$$

Alternatively, we may use the Fisher's information matrix to calculate the variances of the estimates approximately. However, it is not recommendable because the computed variances are underestimated as compared with those obtained by the use of equations (16) and (19) even though the results are not shown here.

3. EXAMPLE

To clarify the method as described before, a real example is given. Used data was collected by the Kosei-Shou-Tokutei-Sikkan-Sinkei-Hifu-Shoukou-Gun-Chousa-Kenkyuu-Han (the Study Group for Neuro-Dermat Syndrome supported by the Ministry of Health and Welfare) and the Nanbyou-no-Ekigaku-Chousa-Kenkyuu-Han-Jimu-kyoku (the Epidemiological Research Officials for the Specified Incurable Diseases) (Takagi, et al., 1987). The research was carried out by mail in order to estimate the number of patients with Recklinghausen disease that is characterized mainly by the symptons of neurofibroma, pigmentation (Café au lai spots) of the skin, and so on, and is thought to be hereditary. For the subjects of the survey, 7512 hospitals with 300 beds or more were selected, and were asked to report the number of patients by sex during the last one year. The results was 4861 institutes answered while remaining 2651 subjects did not.

Table 1 shows the observed distribution of the number of male patients with respect to each hospital, and the results of the mode prediction method. In this case, both of Newton-Raphson method and simplex method were

Table 1
Observed and predictive numbers of male patients with Recklinghausen disease in Japan

No.	n_i	p_i	$\bar{\delta}_i$	$\hat{\delta}_i$	SE($\hat{\delta}_i$)	$\hat{c}_i$	SE($\hat{c}_i$)
0	3984	0.81958	2172.718	2179.514	20.933	0.000	0.000
1	595	0.12240	324.490	325.079	16.886	920.079	16.886
2	154	0.03168	83.986	83.767	9.021	475.534	18.041
3	59	0.01214	32.176	31.783	5.652	272.349	16.955
4	21	0.00432	11.453	10.987	3.409	127.950	13.635
5	15	0.00309	8.180	7.703	2.895	113.514	14.477
6	12	0.00247	6.544	6.060	2.601	108.360	15.603
7	7	0.00144	3.818	3.320	2.015	72.237	14.105
8	4	0.00082	2.181	1.670	1.562	45.361	12.499
9	3	0.00062	1.636	1.117	1.380	37.054	12.420
10	1	0.00021	0.545	0.000	0.918	10.000	9.180
11	1	0.00021	0.545	0.000	0.918	11.000	10.098
12	1	0.00021	0.545	0.000	0.918	12.000	11.016
13	1	0.00021	0.545	0.000	0.918	13.000	11.934
15	1	0.00021	0.545	0.000	0.918	15.000	13.770
42	1	0.00021	0.545	0.000	0.918	42.000	38.555
45	1	0.00021	0.545	0.000	0.918	45.000	41.309
Total	4861	1.00000	2651.000	2651.000	——	2320.438	120.280

Note: see the text for notations.

used in order to compute the predictive values efficiently. Note that the same notations as before are used in Table 1 although the standard errors of $\hat{\delta}_i$ and $\hat{c}_i$, SE($\hat{\delta}_i$) and SE($\hat{c}_i$) respectively, are shown instead of the variances. We can find that the predictive values, $\hat{\delta}_i$'s, are greater than the expected values, $\bar{\delta}_i$'s, when the number of patients is relatively small(i.e., i=0 and 1 in this case), and that they approach zero as the number of patients increases. Consequently, our method provides the total number of patients to be computed smaller than that of using the expected values. In practice, the sum of predictive number of patients is 2320.4 as shown in Table 1, and its log probability is -23.29. Since the total expected number and the log probability are 2421.6 and -24.98, respectively, the defference of estimated numbers is about 101. However, it may be considered that this differece is meaningless statistically because the standard error of $\hat{c}$ is 120.3, even though we think, for the practical usage, the difference of one hundred must be considerably large for the study on the rare disease epidemiology.

We show the female case in Table 2. Like the male case, the sum of predictive number of patients and the log probability are 2346.5 and -24.73 respectively, although for the expected number those are 2450.9 and -26.09 respctively. Hence, the difference number is about 104.

Now, by means of the mode prediction method, we could obtain the nearly same results in both male and female cases, which is meaningful and also

Table 2
Observed and predictive numbers of female patients with Recklinghausen disease in Japan

No.	n_i	p_i	$\bar{\delta}_i$	$\hat{\delta}_i$	SE($\hat{\delta}_i$)	$\hat{e}_i$	SE($\hat{e}_i$)
0	4038	0.83069	2202.168	2209.025	20.491	0.000	0.000
1	534	0.10985	291.223	291.696	16.108	825.696	16.108
2	152	0.03127	82.895	82.671	8.964	469.343	17.928
3	55	0.01131	29.995	29.594	5.460	253.781	16.381
4	20	0.00411	10.907	10.440	3.329	121.759	13.315
5	26	0.00535	14.179	13.724	3.783	198.619	18.915
6	11	0.00226	5.999	5.512	2.495	99.073	14.967
7	6	0.00123	3.272	2.771	1.876	61.394	13.133
8	5	0.00103	2.727	2.221	1.726	57.767	13.810
9	4	0.00082	2.181	1.670	1.562	51.031	14.062
10	2	0.00041	1.091	0.559	1.172	25.589	11.718
11	2	0.00041	1.091	0.559	1.172	28.148	12.890
13	2	0.00041	1.091	0.559	1.172	33.266	15.233
14	1	0.00021	0.545	0.000	0.918	14.000	12.852
15	1	0.00021	0.545	0.000	0.918	15.000	13.770
39	1	0.00021	0.545	0.000	0.918	39.000	35.801
53	1	0.00021	0.545	0.000	0.918	53.000	48.652
Total	4861	1.00000	2651.000	2651.000	——	2346.466	125.694

Table 3
Summary results by Yanai's method

	Male	Female
Mean incidence rate*(SD)	2.492(1.201)	2.434(1.241)
Mean recovery rate (SD)	0.654(0.069)	0.654(0.069)
Correlation coefficient	0.197	0.222
Interception	0.241	-0.177
Regression coefficient	3.441	3.990
Predictive value (SE)	2297.0(178.9)	2461.2(186.2)

* per 100,000 population.

suggestive. Remember Recklinghausen disease seems to be hereditary, so that it may be confirmed that the genes concerning the disease must exist on the specific autosome.

In the meantime, we will campare the results by Yanai's method with the above results. Correlation coefficients of the regional incidence rate with the recovery rate are 0.197 and 0.222 in male and female cases rspectively (Note that there are 47 prefectures in Japan) as shown in table 3 that is

the summary results by the use of Yanai's method. The total estimated numbers of patients are 2297.0 for the male case, and 2461.2 for the female case, respectively,of which standard errors, as shown in Table 3, are 178.9 and 186.2, rspectively, and are larger than those of the mode prediction as before. Since the difference number between male and female patients is about 164, the sex ratio of the incidence rate may not be able to be assumed as unity for the diease under study.

Nevertheless, as described above, it may be thought that all estimates obtained by different methods provide approximately the same number each other although some differences certainly exist. Since we must choose the best one among the limited and incomplete methods without any probability of knowing the true number of patients at the present time, the theoretical validity and consistency are only and useful indices for the selection of the method. Therefore, putting the above results and explanations together so far, it is thought that the mode prediction method as before is considered preferable.

4. ALTERNATIVE METHODS AND DISCUSSION

We will discuss alternative methods here to some extent. Even though we assumed the polynomial distribution with regard to the sample data, in order to obtain the predictive values, it may be thought that the multiple hypergeometric distribution is capable(Yanagawa, 1987).

Suppose n_i is a independent binomially distributed random variable such that $n_i \sim B(\eta_i, \pi_i)$ for all i, and also $N \sim B(M, \pi)$. Then the conditional likelihood fuction L_G can be expressed as

$$L_G = \prod_{i=0}^{K} \binom{\eta_i}{n_i} \pi_i^{n_i}(1-\pi_i)^{\eta_i - n_i} \Big/ \left[\binom{M}{N} \pi^N (1-\pi)^{M-N} \right]$$

$$= \prod_{i=0}^{K} \binom{\eta_i}{n_i} (\pi_i / \pi)^{n_i} [(1-\pi_i)/(1-\pi)]^{\eta_i - n_i} \Big/ \binom{M}{N} . \qquad (20)$$

If we assume that $\pi_i = \pi$ for all i, equation (20) can be reduced to

$$L_H = \prod_{i=0}^{K} \binom{\eta_i}{n_i} \Big/ \binom{M}{N} . \qquad (21)$$

Hence, L_H obeys the multiple hypergeometric distribution under the assumption. Therefore, we may be able to obtain a set of $\hat{\eta}_i$ (or $\hat{\delta}_i$) that maximize L_H under the restricted condition of equation (5), or maximize the logarithm of L_H, ℓ_H, as follows.

$$\ell_H = \sum_{i=1}^{K} \ln \Gamma(n_i + \delta_i + 1) - \sum_{i=1}^{K} \ln \Gamma(\delta_i + 1) + \ln \Gamma(M - N + n_0 - \sum_{i=1}^{K} \delta_i + 1)$$

$$- \sum_{i=1}^{K} \ln \Gamma(M - N - \sum_{i=1}^{K} \delta_i + 1) + K_H , \qquad (22)$$

where

$$K_H = \ln(N!) + \ln[(M-N)!] - \ln(M!) - \sum_{i=0}^{K} \ln(n_i!). \qquad (23)$$

Probably, one can easily obtain the predictive values $\hat{\delta}_i$ for all i by the use of the simplex method and/or Newton-Raphson method. In such a case, it may be necessary to calculate the differenciations of L_H or ℓ_H with respect to certain parameter, but these computations are not so difficult even though we do not describe them here(which have already been obtained). Nevertheless, the problem lies in the assumption; i.e., is it valid to assume that all proportions are the same as π ? In Japan, generally speaking, it is thought that the percentage of recovery of the survey must increase as the number of patients in a hospital increases, which means $\pi_0 < \pi_1 < \cdots < \pi_K$, because doctors who treat many patients of interest tend to be more intereted in the survey and give more cooperation than those who do less or none. Therefore, it is not affirmed that the assumption for the equality of all proportions is acceptable.

In the meantime, we can briefly obtain the ML estimates of π and π_i in equation (20); i.e., $\hat{\pi}=N/M$ and $\hat{\pi}_i=n_i/\eta_i$ for all i, respectively, if η_i is given. Hence, by substituting these estimates into equation (20) and by computing the logarithm of the equation with the condition of equation (5) taken into consideration, we obtain

$$\begin{aligned}\ell_G = &\sum_{i=1}^{K} [\ln\Gamma(n_i+\delta_i+1) - \ln\Gamma(\delta_i+1) - (n_i+\delta_i)\ln(n_i+\delta_i) + \delta_i \ln\delta_i] \\ &+ \ln\Gamma(M-N+n_0-\sum_{i=1}^{K}\delta_i+1) - \ln\Gamma(M-N-\sum_{i=1}^{K}\delta_i+1) - (M-N+n_0-\sum_{i=1}^{K}\delta_i)\ln(M-N+n_0-\sum_{i=1}^{K}\delta_i) \\ &+ (M-N-\sum_{i=1}^{K}\delta_i)\ln(M-N-\sum_{i=1}^{K}\delta_i) + K_G, \qquad (24)\end{aligned}$$

where

$$\begin{aligned}K_G = &\sum_{i=0}^{K} [n_i \ln n_i - \ln(n_i!)] - \ln(M!) + \ln(N!) + \ln\{(M-N)!\} \\ &+ M\ln M - N\ln N - (M-N)\ln(M-N). \qquad (25)\end{aligned}$$

Now, we may be capable of calculating the predictive values as maximize ℓ_G. However, since numerical computations have not been proceeded yet, we cannot confirm whether the solutions are converged or not whenever Newton-Raphson method is applied, and also whether the desired solutions can be obtained or not in practice. These questions will be answered directly if we carry out numerical calculations by the use of real data in the near future although we should compute and evaluate the predictive variances under study, and also compare them theoretically as well as practically with those described in this paper.

Finally, we canont avoid to point out the most serious problem; i.e., our observed data is not collected at random even though sample size is always extremely large. Therefore, we had better suggest, it may be desirable to think that these methods as before are only for the sake of convenience at this stage.

ACKNOWLEDGEMENTS

We are grateful to Dr. T. Sato for his helpful comments, Prof. T. Yanagawa for his suggestive letter, and the research colleagues of the epidemiology group for the data.

REFERENCES

Mizuno,S., Aoki,K., Sasaki,R., Asano,A., Yamada,T., and Katuda,N.(1986). Zenkoku chousa seiseki wo motiite no yuubyou suu no iti suikei houhou (A method of the estimation of the prevalence number by the use of the nationwide survey). Japanese Journal of Public Health 33(Sup.10), 255.

Nakae,K., Kasugawa,R., Honma,M., Toujyou,T., Tunematu, T., Isikawa, E., and Aoki,K.(1985). Kougenbyou 4 sikkan dai-iti-ji zenkoku ekigaku chousa seiseki: zenkoku kanja suu no suitei(The results of the first nationwide epidemiological survey of four collagen diseases : estimation of total patient number in the whole country). Kouseishou Tokutei Sikkan Nanbyou no Ekigaku Chousa Kenkyuu Han Shouwa 59 Nendo Kenkyuu Gyouseki Houkokusho(Annual Report of the Epidemiological Research Groups for the Specified Incurable Diseases, the Ministry of Health and Welfare, 1984), 135-150.

Takagi,H., Inaba,Y., Takahasi,E. I., Niimura,M., Nakauchi,Y., and Aoki,K. (1987). Recklinghausen byou to kessetu kouka shou no zenkoku jyusin kanja suu no suitei ni tuite(Estimations for the numbers of patients with Recklinghausen disease and with Bourneville-Pringle). Kouseishou Tokutei Sikkan Sinkei Hifu Shoukou-Gun Chousa Kenyuu Han Shouwa 61 Nendo Kenkyuu Houkokusho(Annual Report of the Study Groups for Neuro-Dermat Syndrome, the Ministry of Health and Welfare, 1986)(in press).

Yanagawa,T.(1987). Personal Communication.

Yanai,H., Nakae,K., Ono,M., and Toyokawa,H.(1975). A method of correction of the prevalence rates in epidemiological survey. Japanese Journal of Public Health 22, 71-81.

[*Received June, 1987*]