Acta Medica Okayama

Volume 28, *Issue* 2 1974 *Article* 7

APRIL 1974

Selection and application of factors for forecasting the epidemic time and severity of Japanese encephalitis prevalence

Masana Ogata* Hirokazu Osaki†

*Okayama University,
†Okayama University,

Selection and application of factors for forecasting the epidemic time and severity of Japanese encephalitis prevalence*

Masana Ogata and Hirokazu Osaki

Abstract

For predicting median date or incidence rate in the prevalence of japanese encephalitis, the authors considered the 34 factors; namely, the climate, latitude, longitude, and date showing immunological positivity of hemoagglutination inhibiting reaction on 50% the number of swine, etc. To make the mean square residual the smallest which yields, in the case of calculating with the multiple regression equation, the most important and meritorious factors were selected from the factors mentioned above by the voluntary selection rule devised by us. Multiple regression equations were formulated for them. To predict the median date in the prevalence of japanese encephalitis in whole japan, we found that the three factors, i. e. the amount of rainfall in April, the average temperature in March and HI positive rate of swine till the end of july, were essential. And for the foreseeing of incidence rate in Okayama Prefecture, factors concerning mosquitoes were added; this resulted in the useful two factors, namely, common logarithm of the total number of Culex tritaeniorhynchus till july 20th and the rainfall in June.

Acta Med. Okayama **28**, 125—137 (1974)

SELECTION AND APPLICATION OF FACTORS FOR FORECASTING THE EPIDEMIC TIME AND SEVERITY OF JAPANESE ENCEPHALITIS PREVALENCE

Masana OGATA* and Hirokazu OSAKI**

* *Department of Public Health, Okayama University Medical School, Okayama Japan*

** *Department of Industrial Engineering, School of Engineering, Okayama University, Okayama Japan*

Received for publication, October 11, 1973

Abstract: For predicting median date or incidence rate in the prevalence of Japanese encephalitis, the authors considered the 34 factors; namely, the climate, latitude, longitude, and date showing immunological positivity of hemoagglutination inhibiting reaction on 50% the number of swine, etc. To make the mean square residual the smallest which yields, in the case of calculating with the multiple regression equation, the most important and meritorious factors were selected from the factors mentioned above by the voluntary selection rule devised by us. Multiple regression equations were formulated for them. To predict the median date in the prevalence of Japanese encephalitis in whole Japan, we found that the three factors, i. e. the amount of rainfall in April, the average temperature in March and HI positive rate of swine till the end of July, were essential. And for the foreseeing of incidence rate in Okayama Prefecture, factors concerning mosquitoes were added; this resulted in the useful two factors, namely, common logarithm of the total number of *Culex tritaeniorhynchus* till July 20th and the rainfall in June.

As to the route of the infection of Japanese encephalitis virus, it was thought that the virus propagates in swine (1) and herons (2). And mosquitoes bite swine and herons during viremia and change themselves to hazardous mosquitoes. In order to estimate the epidemic time curve of Japanese encephalitis patients, the onset time of hazardous mosquitoes is the most important.

And this onset time has a close correlation to the atmospheric temperature in summer; i. e. we call it "onset by high atmospheric temperature" (3-8). On the other hand, the time, when the immunological positivity of hemoagglutination inhibiting reaction (HI reaction in short) of swine becomes positive to 50%, is also required for determining the onset of hazardous mosquitoes, as there is a close correlation between this time and the median date of epidemic time curve or the incidence rate (3-9).

In the present study we attempted to forecast the median date and the incidence rate of Japanese encephalitis patients from average temperature every month, the duration of sunshine and the amount of rainfall as the climatic factors, latitude and longitude, date showing HI reaction on 50% of swine and HI positive rate on swine till the end of July. To make the mean square residual smallest, which yields, when calculated with the multiple regression equation, the most important and valuable factors from the various factors above mentioned (34 factors in all) by the voluntary selection rule devised by us. Multiple regression equations were formulated for them.

In addition, to examine the efficiency of the factor of the mosquitoes, we attempted to forecast the median date and the incidence rate in Okayama Prefecture based on the factors concerning the mosquitoes plus the factors above mentioned.

MATERIALS AND METHODS

I) Materials

In order to forecast the median date (y_1) or the incidence rate (y_2) in the prevalence of Japanese encephalitis, the authors considered the following factors in each prefecture from 1965 to 1970.

1) Climatic factors

Month	average temperature (°C/day)	duration of sunshine (hour/day)	amount of rainfall (mmx10^{-3})	
January	x_1	x_{11}	x_{21}	
February	x_2	x_{12}	x_{22}	
March	x_3	x_{13}	x_{23}	
April	x_4	x_{14}	x_{24}	this
May	x_5	x_{15}	x_{25}	year
June	x_6	x_{16}	x_{26}	
July	x_7	x_{17}	x_{27}	
October	x_8	x_{18}	x_{28}	
November	x_9	x_{19}	x_{29}	previous
December	x_{10}	x_{20}	x_{30}	year

2) latitude : x_{31}
3) longitude : x_{32}
4) date showing HI reaction on 50% number of swine : x_{33}
5) positive rate on swine till the end of July : x_{34}

Median date and date showing HI reaction on 50% of swine were expressed as days counted from June 20th.

The following factors were used to forecast the median date or the incidence rate on Okayama Prefecture in order to examine the efficiency of the factor of mosquitoes.

6) The number of *Culex tritaeniorhynchus* (*C. t.* in short)
common logarithm of ♀ + ♂ C. t. till July 10th : x_{35}
common logarithm of ♀ C. t. till July 10th : x_{36}
common logarithm of ♀ + ♂ C. t. till the end of July : x_{37}
common logarithm of ♀ C. t. till the end of July : x_{38}
common logarithm of ♀ + ♂ C. t. till August 10th : x_{39}
common logarithm of ♀ C. t. till August 10th : x_{40}

II) Methods

The multiple regression equation was calculated as follows.

y : the dependent variable
$x=(x_1, x_2, \ldots, x_k)$: the independent variable
(y_i, x_i), $x_i=(x_{1i}, x_{2i}, \ldots, x_{ki})$: the observation of (y, x), $i=1, 2, \ldots, N$

Then the multiple regression equation became as follows.

$$\hat{y}=\bar{y}+\sum_{i=1}^{k}\beta_i(x_i-\bar{x}_i) .$$

Where $\vec{d}=(d_1, d_2, \ldots, d_k)'$, $S=(s_{ij})$

$$d_j=\sum_{i=1}^{N}(y_i-\bar{y})(x_{ji}-\bar{x}_j) , \quad j=1, 2, \ldots, k$$

$$s_{ij}=\sum_{l=1}^{N}(x_{il}-\bar{x}_i)(x_{jl}-\bar{x}_j) , \quad i, j=1, 2, \ldots, k$$

then $(\beta_1, \beta_2, \ldots, \beta_k)=\vec{d}'s^{-1}$.

The mean square residual from the multiple regression was as follows.

$$\hat{\sigma}^2=\left\{\sum_{i=1}^{N}(y_i-y)^2-\vec{d}'\, S^{-1}\vec{d}\right\} \Big/ (N\text{-}k\text{-}1) \quad \ldots .(1)$$

If it would be possible to select the important and meritorious factors from many various factors ussed in the multiple regression, then we can explain clearly the tendency of the dependent variable with those selected factors alone.

In equation (1), the smaller the mean square residual became, the better the efficiency of the multiple regression equation was. $\sum_{i=1}^{N}(y_i-y)^2$ in (1) was not connected with independent variables. Then the relation between $\vec{d}'S^{-1}\vec{d}$ and the individual independent variable was examined.

$$U=(r_{ij}),\quad \begin{cases} r_{ij}=s_{ij}/\sqrt{s_{ii}\,.\,s_{jj}}\ , & i\neq j=1,2,\ldots,k \\ r_{ii}=1\ , & i=1,2,\ldots,k \end{cases}$$

$$\vec{\delta}=(\delta_1,\delta_2,\ldots,\delta_k)'\ ,\ \delta_j=d_j/\sqrt{s_{jj}}\ ,\ j=1,2,\ldots,k$$

U, $\vec{\delta}$ were transformed from S and $\vec{d}$. Then the following equation established.

$$\vec{d}'S^{-1}\vec{d}=\vec{\delta}'U^{-1}\vec{\delta}\ \ldots\ldots(2)$$

Further

$$\begin{aligned}\vec{\delta}U^{-1}\vec{\delta}=\ &\delta_{i_1}^2+(\delta_{i_2}-\gamma_{i_1i_2}\delta_{i_1})^2/(1-\gamma_{i_1i_2}^2)+\ldots\\ &+a_{i_ji_j}^j(\delta_{i_j}+\sum_{l=1}^{j-1}\delta_{i_l}a_{i_ji_l}^j/a_{i_ji_j}^j)^2+\ldots\\ &+a_{i_ki_k}^k(\delta_{i_k}+\sum_{l=1}^{k-1}\delta_{i_l}a_{i_ki_l}^k/a_{i_ki_k}^k)^2\ldots\ldots(3)\end{aligned}$$

where $\alpha_{i_ji_j}^j>0\ ,\quad j=1,2,\ldots,k$

$a_{i_ji_l}^j$, $(l=1,2,\ldots,j)$ is the element of j th row by l th column of U_j^{-1}.

$$U_j=\begin{pmatrix}1 & \gamma_{i_1i_2} & \gamma_{i_1i_3} & \cdots & \gamma_{i_1i_j}\\ \gamma_{i_2i_1} & 1 & \gamma_{i_2i_3} & \cdots & \gamma_{i_2i_j}\\ \gamma_{i_3i_1} & \gamma_{i_3i_2} & 1 & \ddots & \vdots\\ \gamma_{i_ji_1} & \gamma_{i_j2} & \cdots & \cdots & 1\end{pmatrix}$$

Therefore, the value of $\vec{d}'S^{-1}\vec{d}$ increased with the increase of the number of independent variables. Thus we considered the selection method of independent variables by which the value of $\vec{d}'S^{-1}\vec{d}$ attained maximum in some sense.

II 2. 1) Forward selection method (F. S. M.)

II 2. 1. 1) $j=1$ (only one independent variable)

Only one independent variable was selected from k independent variables. From equations (2) and (3),

$$\vec{d}'S^{-1}\vec{d}=\delta_{i_1}^2$$

The criterion of selection was determined as follows.

$$\max\ \vec{d}'S^{-1}\vec{d}=\max_{i_1}\ \delta_{i_1}^2\ ,\ i_1\in\{1,2,\ldots\ldots,k\}$$

I_1 was taken as the independent variable number which satisfied the above equation. The regression equation calculated from I_1 variable was most efficient.

2.1.2) $j=2$ (two independent variable)

Two independent variables were selected from k independent variables. In this case, one independent variable was fixed on I_1. Then equation (2) and (3) became as follows.

$$\vec{d}'S^{-1}\vec{d}=\delta_{I_1}^2+(\delta_{i_2}-\gamma_{I_1 i_2}\delta_{I_1})^2/(1-\gamma_{I_2 1_i}^2)$$

The criterion was as follows.

$$\max\ \vec{d}'S^{-1}\vec{d}=\max_{i_2}\{\delta_{I_1}^2+(\delta_{i_2}-\gamma_{I_1 i_2}\delta_{I_1})^2/(1-\gamma_{I_1 i_2}^2)\}$$

$$i_2\in\{1,2,\ldots.,k\}-\{I_2\}.$$

I_2 was put as the selected independent variable number. Thus the regression equation using I_1 and I_2 variables was more efficient than any other ones doing other two variables.

2.1.3) $j=m$ (in general case)

$\vec{d}'S^{-1}\vec{d}(i_1,\ i_2,\ldots,i_1)$ was put as the value of $\vec{d}'S^{-1}\vec{d}$ calculated from independent variables $i_1, i_2, \ldots., i_1$.

Independent variables $I_1, I_2, \ldots, I_{m-1}$ were already selected before this step. Them I_m satisfied the following equation.

$$\max\ \vec{d}'S^{-1}\vec{d}\,(I_1, I_2, \ldots, I_{-m1}, i_m),$$
$$i_m\in\{1,2,\ldots,k\}-\{I_1, I_2, \ldots., I_{m-1}\}$$

Then the regression equation was made using $I_1, I_2, \ldots, I_m$ variables.

This process continued until $j=k$.

The order of the independent variable by forward selection method was

$$I_1, I_2, \ldots, I_j.\ldots, I_k$$

And the efficiency of individual variable decreased with the increase of j.

2.2) Backward selection method (B. S. M.)

Dj(m) was taken as the value of $\vec{d}'^{-1}\vec{d}$ calculated from independent variables which consisted of independent variables $(i_1, i_2, \ldots, i_m)$ minus variable i_j.

In this selection method, the variable of which efficiency was the lowest was selected one by one.

2.2.1) $m=k$

In this step, the unnecessary independent variable was selected from k independent variables by the following criterion.

$$\max_j Dj(k),\ j \in \{1, 2, \ldots, k\}.$$

I'_k was taken as the selected independent variable number. I'_k had a lower efficiency for the multiple regression equation than any other variables.

2. 2. 2) $m=1$

$I'_k, I'_{k-1}, \ldots, I'_{1+1}$ were already selected before this step. Then I_1 was taken as the independent variable number selected by the following criterion.

$$\max d_j(1),\ j \in \{1, 2, \ldots, k\} - \{I_{1+1}, I_{1+2}, \ldots, I_k\}.$$

In step m, the regression equation was made using $\{1, 2, \ldots, k\} - I'_{m+1}, I'_{m+2}, \ldots, I'_k$.

This process continued until $m=1$. The efficiency of individual variable increased with the decrease of m.

2. 3) All possible selection method (A. P. S. M.)

In this method, the combination of variables of which efficiency was the highest among many variables was selected.

$I^m_1, I^m_2, \ldots, I^m_m$ were taken as the combination of variables selected by the following criterion.

$$\max_{(i_1, i_2, \ldots, i_m)} \vec{d}' S^{-1} \vec{d}\,(i_1, i_2, \ldots, i_m),\ (i_1, i_2, \ldots, i_m) \in \{1, 2, \ldots, k\}.$$

In each step, the regression equation was made using $I^m_1, I^m_2, \ldots, I^m_m$.

In the case of $m=1$, $I^1_1 = I_1$

In the case of $m=k-1$, $I^{k-1}_1, I^{k-1}_2, \ldots, I^{k-1}_{k-1} = \{I'_1, I'_2, \ldots, I'_{k-1}\}$

All possible selection methods had logically higher efficiency than other two selection methods. But this method required the longest computation time.

RESULTS AND DISCUSSION

1) Estimation of median dates of epidemic time curve of patients in Japan

1. 1) Estimation from the average atmospheric temperature

The correlation matrix, mean value and unbiased variance are shown in Table 1.

The correlation coefficient between median date and average atmospheric temperature in May (x_5) was -0.830. The absolute value was greater than any other coefficients. Then x_5 became the most important factor to estimate the median date. The results of forward (F. S. M.), backward (B. S. M.) and all possible selection methods (A. P. S. M.) are shown in Table 2. In the case of one factor, the average atmospheric temperature in May (x_5) was most useful among 10 factors from the results of three selection methods. And the regression equation becomes as follows.

$$y_1 = -101.006x_5 + 289.120, \qquad \hat{\sigma} = 11.99 \cdots\cdots (1)$$

TABLE 1 CORRELATION COEFFICIENTS BETWEEN MEAN VALUES AND UNBIASED VARIANCES

	y_1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1	1.000	−0.123	0.134	0.759	−0.251	−0.830	0.747	−0.742	0.749	0.771	0.771
x_1		1.000	0.535	−0.020	0.671	0.377	−0.076	0.350	0.015	0.087	0.182
x_2			1.000	0.365	0.381	−0.066	0.283	0.195	0.314	0.403	0.480
x_3				1.000	0.109	−0.718	0.989	−0.484	0.989	0.989	0.938
x_4					1.000	0.512	0.078	0.539	0.128	0.151	0.128
x_5						1.000	−0.697	0.833	−0.675	−0.702	−0.650
x_6							1.000	−0.471	0.991	0.970	0.932
x_7								1.000	−0.466	−0.489	−0.462
x_8									1.000	0.981	0.943
x_9										1.000	0.957
x_{10}											1.000
mean value	59.542	4.788	5.590	10.128	14.099	18.349	22.729	25.629	17.803	13.147	7.647
unbiased variance	455.914	2.160	3.808	100.487	1.353	3.079	71.381	0.934	40.725	30.667	35.542

N=72

TABLE 2 RESULTS OF FORWARD, BACKWARD AND ALL POSSIBLE SELECTION METHODS FROM THE DATA OF TABLE 1.

	l	1	2	3	4	5	6	7	8	9	10	$\hat{\sigma}$
forward	I_l	x_5	x_{10}	x_7	x_2	x_6	x_3	x_1	x_8	x_4	x_9	
selection method	$\hat{\sigma}$	11.99	10.12	9.65	9.63	9.50	9.44	9.36	9.32	9.34	9.46	
backward	I'_l	x_5	x_{10}	x_2	x_6	x_3	x_1	x_3	x_7	x_4	x_9	
selection method	$\hat{\sigma}$	11.99	10.12	9.86	9.48	9.37	9.30	9.29	9.32	9.39	9.46	
	1	x_5										11.99
	2	x_5	x_{10}									10.12
	3	x_5	x_7	x_{10}								9.65
all possible	4	x_2	x_5	x_6	x_{10}							9.48
	5	x_2	x_3	x_5	x_6	x_{10}						9.37
selection method	6	x_1	x_2	x_3	x_5	x_6	x_{10}					9.30
	7	x_1	x_2	x_3	x_5	x_6	x_8	x_{10}				9.29
	8	x_1	x_2	x_3	x_5	x_6	x_7	x_8	x_{10}			9.32
	9	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_{10}		9.39
	10	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	9.46

This equation (1) was more efficient than any other ones calculated from average atmospheric temperature in other months to forecast the median date. In the case of two factors, the average atmospheric temperature in May (x_5) and December in previous year (x_{10}) were most useful among 10 factors. And the multiple regression equation became as follows.

$\hat{y}_1 = -6.931x_5 + 1.435x_{10} - 206.101 \qquad \hat{\sigma} = 10.12 \cdots\cdots (2)$

In the case of three factors, the result of F. S. M. coincided with that of A. P. S. M.. Generally the results of all possible selection methods were logically more useful than those of other two methods. Then two factors mentioned above, besides the averge atmospheric temperature in July, were most useful.

$\hat{y}_1 = -3.828x_5 - 6.118x_7 + 1.571x_{10} + 234.441 \qquad \hat{\sigma} = 9.65 \cdots\cdots (3)$

In the case of four factors, average atmospheric temperature in February, May, June and previous year's December were selected from B. S. M. and A. P. S. M.

$\hat{y}_1 = -2.478x_2 - 6.602x_5 - 1.108x_6 + 3.351x_{10} + 243.441 \qquad \hat{\sigma} = 9.48 \cdots (4)$

The root of the mean square residual decreased gradually with the increase of the number of factors up to 7 factors, and after that it increased with the increase in the number of factors in B. S. M. and A. P. S. M. methods. From these results, obviously the average atmospheric temperature in January, February, March, May, June, and previous year's December was most efficient to estimate the median date in the prevalence of Japanese encephalitis in whole Japan. And the most useful regression equation became as follows.

$\hat{y}_1 = -2.568x_1 - 4.262x_2 + 3.167x_3 - 3.734x_5 - 5.319x_6 + 4.858x_{10} + 256.461 \qquad \hat{\sigma} = 9.30 \cdots\cdots (5)$

1.2) Estimation from 34 factors

The results of A. P. S. M. are shown in Table 3. Time to calculate

TABLE 3 SELECTED FACTORS BY ALL POSSIBLE SELECTION METHODS AND REGRESSION COEFFICIENTS

	number of variable	1	2	3	4	R	$\hat{\sigma}$
median date	1	x_5 −10.101				244.874	11.99
	2	x_3 1.724	x_{24} −113.614			50.355	9.59
	3	x_3 1.678	x_{24} −84.254	x_{34} −0.129		52.510	8.65
	4	x_5 −3.095	x_9 2.304	x_{24} −53.868	x_{34} −0.121	124.503	8.25
incidence rate	1	x_3 0.716				−4.159	2.353
	2	x_3 0.729	x_4 −1.022			11.560	2.045
	3	x_3 0.743	x_4 −1.041	x_{20} −0.544		11.760	1.904
	4	x_2 0.362	x_3 0.721	x_4 −1.522	x_{32} −0.269	27.091	1.838

A. P. S. M. took too long in more than 5 factors that the calculation was stopped up to 4 factors. The results of F. S. M. and B. S. M. were used to examine the degree of the efficiency of the multiple regression equation in more than May (x_5) was selected. This factor was the most useful among 34 factors chosen. Then the regression equation was equal to the equation (1). In case of two factors, the average atmospheric temperature in March (x_3) and rainfall amount in April (x_{24}) were selected. And the regression equation became as follows.

$\hat{y}_1 = 1.724x_3 - 113.61x_{24} + 50 + 50.355 \qquad \hat{\sigma} = 9.59 \cdots\cdots\cdots\cdots (6)$

The efficiency of this equation was about equal to that of equation (3). The mean square residual decreased with the increase of the number of factors up to 4 factors. Then the multiple regression equation was calculated using four factors, that is, the rainfall in April (x_{24}), the average atmospheric temprature in May (x_5) and in previous year's November (x_9), and positive rate on swine till the end of July (x_{34}) were useful to estimate the median date.

$\hat{y}_1 = -3.095x_5 + 2.303x_9 - 53.868x_{24} - 0.121x_{34} + 124.503 \qquad \hat{\sigma} = 8.25 \cdots (7)$

This (equation 7) was more efficient than any other equations above mentioned. Then it became clear that the climatic factors and the factors of HI reaction of swine were useful and beneficial to estimate the median date.

The root of mean square residual calculated by 12 factors (x_2, x_3, x_6, x_{10}, x_{11}, x_{14}, x_{15}, x_{16}, x_{19}, x_{24}, x_{25}, x_{28}, x_{31}, x_{34}) from the backward selection method was 7. 75, which was the smallest.

$$\hat{y} = -4.532x_2 + 5.201x_3 - 4.156x_6 + 1.139x_{10} - 2.535x_{11} - 1.907x_{14} + 2.335x_{15} - 1.715x_{16} + 2.217x_{19} - 53.221x_{24} + 55.783x_{25} - 32.711x_{28} + 3.072x_{31} - 0.079x_{34} - 262.569 \qquad \hat{\sigma} = 7.75 \cdots\cdots\cdots\cdots (8)$$

The difference between $\hat{\sigma}$ of (equations 7 and 8) was 0. 5. Then if we wanted to estimate the median date roughly, the equation (7) was the most useful.

2) Estimation of incidence rate of Japanese encephalitis from 34 factors.

The results of A. P. S. M. were shown in Table 3.

The factor of the average atmospheric temperature also played an important roll for the estimation of the incidence rate.

In case of one factor, the average atmospheric temperature in March (x_3) was selected. And the regression equation became as follows.

$\hat{y} = 0.716x_3 - 4.159, \qquad \hat{\sigma} = 2.35 \cdots\cdots\cdots\cdots\cdots\cdots (9)$

The average atmospheric temperature was the most useful to estimate the median date or the incidence rate in the case of only one factor.

The efficiency of the multiple regression of the incidence rate increased with increase in the number of factors up to 4 factors.

TABLE 4 FACTORS USED FOR ESTIMATION OF MEDIAN DATE AND MORBIDITY RATE

		x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	x_{40}	x_{33}	x_{34}
southern	41	2.02	1.76	3.07	2.66	3.18	2.88	41	80
	42	1.80	1.75	3.78	3.69	3.85	3.75	36	83
	43	2.04	2.00	3.60	3.56	3.86	3.83	63	0
	44	1.38	1.28	2.44	2.33	2.63	2.55	48	10
	45	1.08	0.48	1.97	1.11	2.20	1.41	65	10
	46	1.79	1.11	2.50	1.89	2.54	1.27	58	10
northern	41	1.64	1.18	2.23	2.01	2.50	2.34	45	70
	42	2.14	1.94	2.97	2.71	2.98	2.72	47	20
	43	1.59	1.36	2.91	2.73	3.00	2.83	64	20
	44	1.85	1.83	2.55	2.43	2.63	2.52	53	0
	45	0.85	0.48	2.23	1.88	2.46	2.05	72	0
	46	1.32	0.95	2.30	1.82	2.20	1.83	64	20

Then the useful regression equation was that of four factors.

$$\hat{y}_2 = 0.362x_2 + 0.721x_3 - 1.522x_4 - 0.269x_{32} + 27.091 \qquad \hat{\sigma} = 1.84 \cdots\cdots (10)$$

And the useful factors were the average atmospheric temperature in Feburary (x_2), in March (x_3) and in April (x_4), and the longitude (x_{32}).

Further from the result of backward selection method, the following regression equation of incidence rate attained the lowest value of the mean square residual.

$$\hat{y} = -1.297x_1 + 1.562x_2 + 0.551x_3 - 1.718x_4 + 1.259x_9 - 0.836x_{10} - 0.865x_{15} + 0.893x_{16} - 0.828x_{18} - 0.453x_{20} + 11.859x_{21} + 3.284x_{26} - 19.831x_{29} - 0.265x_{32} + 0.026x_{33} + 27.590 \qquad \hat{\sigma} = 1.65 \cdots\cdots (11)$$

Therefore, the multiple regressions (equations 10 and 11) were the most useful to estimate the incidence rate of Japanese encephalitis.

3) Estimation from climatic factors added with factors concerning mosquitoes in Okayama Prefecture.

As to the route of the infection of Japanese encephalitis virus, it was thought that the mosquitoes played an important role in the virus propagation. Then it was necessary to take into consideration the factor of mosquitoes to forecast the median date or incidence rate of the encephalitis. But there was such a great difference among the data, *i.e.* common logarithms of the numbers, of mosquitoes (*Culex tritaeniorhynchus*, C. t. in short) in each prefectures that the efficiency of the factor of C. t. was examined only in Okayama Prefecture. Table 4 shows the factors used and the data. Fourteen variables used are x_6, x_7, x_{16}, x_{17}, x_{26}, x_{27}, x_{33} and x_{34} and variables from x_{35} to

Prevalence of Japanese Encephalitis in Okayama Prefecture (1966~1971)

x_6	x_7	x_{16}	x_{17}	x_{26}	x_{27}	Median date	incidence rate
20.9	25.2	224.5	140.0	168.5	198.0	58	2.5
22.4	26.2	87.9	274.9	226.5	191.5	53	1.0
21.7	25.1	102.0	250.5	185.3	152.3	69	1.0
20.5	25.3	291.5	267.0	201.9	183.6	50	0
20.6	26.1	329.0	137.0	96.5	190.6	–	0
22.1	26.5	168.0	272.5	113.1	174.1	–	0
20.0	24.5	303.1	116.5	156.7	149.2	68	3.6
21.4	25.2	162.5	319.1	231.5	177.2	59	4.9
20.8	24.4	125.0	315.0	207.6	154.1	57	1.3
19.5	24.1	217.0	261.0	203.9	168.1	67	0
19.8	25.3	395.0	106.5	116.0	160.4	–	0
21.0	25.3	183.0	404.5	140.0	157.2	–	0

x_{40}. Clearly there is a difference between the data of C. t. in southern part and those of northern part in Okayama Prefecture. Then the data from 1966 to 1971 in southern and northern parts of Okayama Prefecture were used for calculation.

3. 1) Median date

In this calculation, the number of data was 8 and that of independent variables was 14. The number of independent variables was more than that of data. Then the mean square residual of regression equation decreased with the increase in the number of independent variables. But the accuracy of estimating data in the other years became poorer with the increase of factors. Therefore, the results up to four factors were mainly explained. In the case of one factor, the duration of sunshine in July (x_{17}) was the most important factor. The factor of mosquitoes played an important role in case of more than two factors (Table 5).

In the case of two factors, the duration of sunshine in July (x_{17}) and number of C. t. (♀ + ♂) till July 10th (x_{35}) were selected. In the case of three or four factors, the number of C. t. (♀) till July 10th (x_{36}) and the climatic factors were selected from 14 factors.

From these results, it is clear that the factors of mosquitoes and of the climate are necessary to estimate the median date of epidemic time curve.

3. 2) Incidence rate

In the case of one factor, the number of C. t. (♀ + ♂) till July 10th (x_{35}) was most important to estimate the incidence rate. In two factors, the number of C. t. (♀ + ♂) till July 10th and the rainfall in June (x_{26}) were most important. In three factors, only the climatic factors were important

TABLE 5 SELECTED FACTORS BY ALL POSSIBLE SELECTION METHODS AND REGRESSION COEFFICIENTS

	number of variable	1	2	3	4	5	6	C	$\hat{\sigma}$
median date	1	x_{17} −0.62						104.98	5.61
	2	x_{17} −0.70	x_{35} 14.39					86.55	4.33
	3	x_{17} −0.77	x_{27} −5.23	x_{36} 14.13				504.37	2.37
	4	x_{17} −0.71	x_{27} −4.64	x_{36} 15.96	x_{6} −1.97			495.79	1.41
	5	x_{17} −0.76	x_{27} −7.35	x_{36} 17.32	x_{38} 16.00	x_{49} −19.86		678.25	0.45
	6	x_{6} 32.38	x_{27} −53.27	x_{33} 1.80	x_{35} 12.44	x_{37} −27.88	x_{40} 23.17	3060.59	0.010
incidence rate	1	x_{35} 2.34						−2.62	1.42
	2	x_{35} 3.89	x_{26} 0.89					−6.15	1.35
	3	x_{6} 5.20	x_{7} −4.69	x_{26} 3.10				−127.54	1.32
	4	x_{35} 13.11	x_{36} −9.65	x_{39} −5.58	x_{40} 5.54			−278.99	1.12

for estimating the incidence rate. Therefore, when the number of C. t. could not be measured, the multiple regression equation calculated by these three factors was most efficient to estimate the incidence rate. In four factors, only those factors concerned with the number of C. t. were selected. Thus, when the number of C. t. only is measured, the multiple regression equation calculted by these four factors can be used to estimate the incidence rate in Okayama Prefecture. From these results, it is evident that the factor of mosquitoes plays an important role in estimating the median date and the incidence rate in Okayama Prefecture. Therefore, when it is possible to gather the data regarding the factor of mosquitoes in all Japan prefectures, the estimation of median date and incidence rate of Japanese encephalitis will become more accurate. Precise comparison of values of median date or incidence rate estimated by this method with their actual values will be reported in the near future.

REFERENCES

1. WILLIAM, F. SCHERER, *et al.*: Immunological studies of Japanese encephalitis virus in Japan, V. Maternal antibodies, antibody responses and viremia following infection of swine, *The Journal of Immunology* **83**, 620-626, 1959
2. EDWARD, L. BUESCHER, WILLIAM, F. SCHERER, *et al.*: Immunologic studies of Japanese encephalitis virus in Japan, IV. Maternal antibody in birds. *The Journal of Immunology* **83**,

614-619, 1959

3. OGATA, M. and OSAKI, H.: Phenomenon of "Tendency of prevalence of Japanese encephalitis in Japan toward the northeast", "Onset by high temperature" and forecasting its prevalence by atmospheric temperature and HI reaction of swine. *Okayama-Igakkai-Zasshi* **80**, Supplement X, 67-84, 1968 (in Japanese)
4. OGATA, M. and OSAKI, H.: Forecasts of prevalence of Japanese encephalitis and "Advancing of prevalence in Japan toward the northeast" and "Onset by high atmospheric temperature". *Japanese Journal of Public Health* **16**, 587-592, 1969 (in Japanese)
5. OGATA, M. and OSAKI, H.: On forecasting prevalence of Japanese encephalitis in Japan, pestimation of median date of prevalence in 1968 from the date of 1965 to 1967. *Okayama-Igakkai-Zasshi* **81**, Supplement XI, 19-23, 1969 (in Japanese)
6. OGATA, M. and OSAKI, H.: On forecasting prevalence of Japanese encephalitis in Japan, estimation of mode. *Okayama-Igakkai-Zasshi* **81**, Supplement XI, 25-28, 1969 (in Japanese)
7. OGATA, M. and OSAKI, H.: Forecasts on date of prevalence of Japanese encephalitis (Epidemiological study on Japanese encephalitis). *Okayama-Igakkai-Zasshi* Supplement XII, 19-25, 1970 (in Japanese)
8. OGATA, M. and OSAKI, H.: Forecasts on degree of prevalence of Japanese encephalitis and period of cycle of incidence in various prefectures of Japan. *Okayama-Igakki-Zasshi*. Supplement XII, 27-35, 1970 (in Japanese)
9. ŌYA, A.: Survey of appearance of *Culex tritaeniorhynchus* and forecasting of prevalence of Japanese encephalitis from hazardous mosquitoes, Research Meeting of Ecology of Japanese Encephalitis Nihon-noen Virus Seitai Kenkyukai (Abstracts), 1969 (in Japanese)