ON THE FORECASTING OF PROGNOSIS IN PEDIATRICS BY A QUANTIFYING METHOD

BY SIGEKI SAKINO and GORO KŌNO

(Received Oct. 2, 1954)

1. Introduction

The purpose of this paper is to apply a quantifying method to this analysis of anamnesis, which aims at forecasting diagnosis or prognosis and to show that the forecasting by it is carried out with the same accuracy as that of a classifying method (see below). The quantifying of the qualitative data obtained by patient's appeals is very useful not only for analysis and forecasting of prognosis, but also for projection or improvement of items in anamnesis. Therefore, the main part of our problem is to quantify medical phenomena, so that the results can characterize the medical phenomena, and be useful for analysing the empirical phenomena.

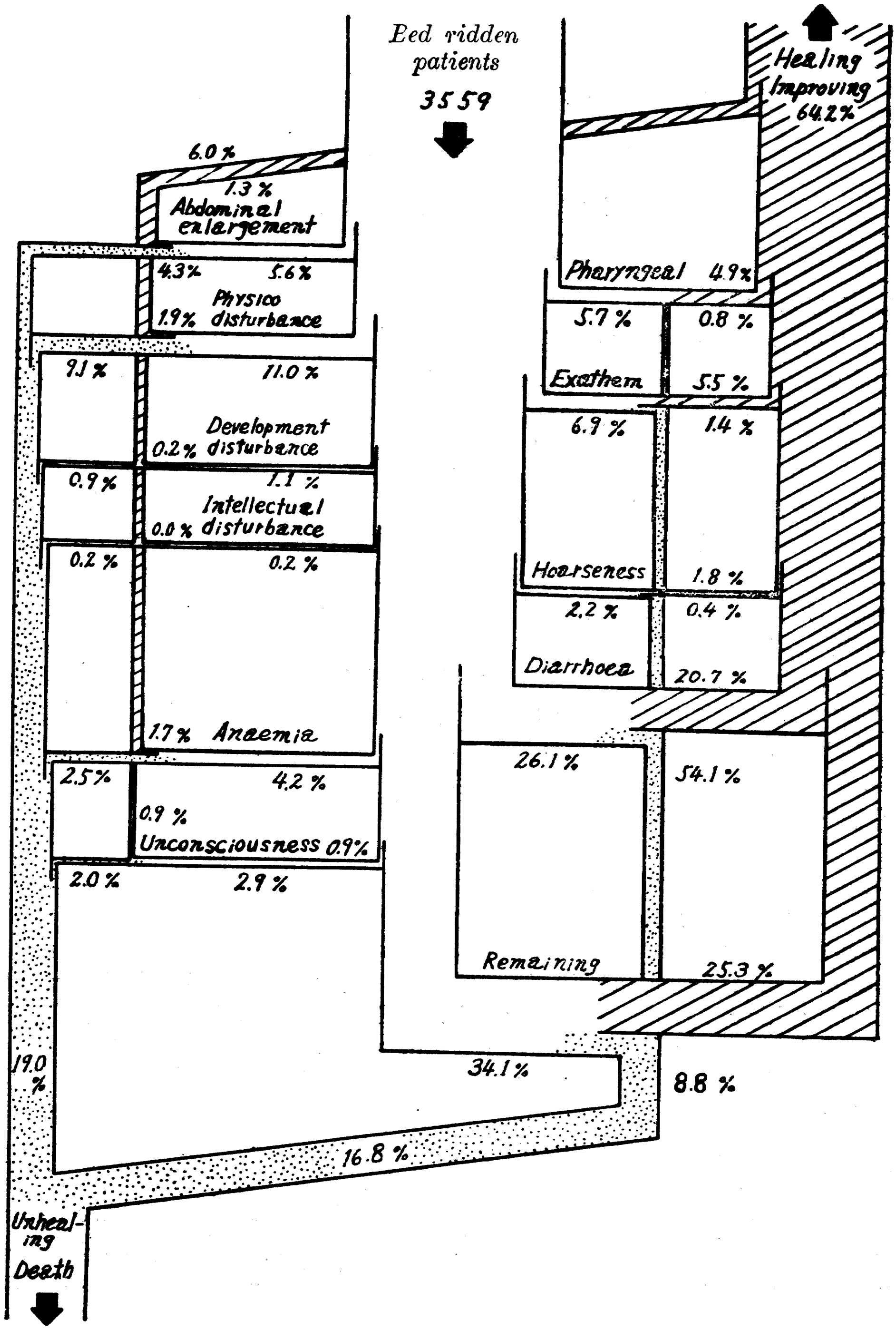
In section 2, we shall find the efficient symptoms to quantify prognosis, and in section 3 we shall give weights to symptoms and show a way to forecast prognosis.

2. Forecasting of prognosis by the chart

The data which we have analysed are the anamneses of the 3559 patients sent to the Tokyo University Hospital from 1944 to 1951.

Let "getting better or favourable prognosis" mean healing or improving, and "getting worse or unfavourable prognosis" unhealing or death. Then according to the anamneses, the 64.2% of the whole patients have favourable prognosis, and the 35.8% have unfavourable prognosis. If the case be always the same, we will have the 64.2% success rate of forecasting by telling them always "you are getting better".

Now, from the anamneses of the 3559 patients we obtained the following chart as to prognosis. From this chart we can obtain symptoms which are efficient for prognostication. The chart shows that the 5.6% patients of the whole patients have abdominal enlargement symptom, and the 4.3% patients of these patients have that


Bed ridden patients
3559
Healing
Improving
64.2%
6.0%
1.3%
Abdominal enlargement
4.3%
5.6%
Physico disturbance
1.9%
Pharyngeal 4.9%
5.7%
0.8%
Exathem
5.5%
9.1%
11.0%
Development disturbance
0.2%
6.9%
1.4%
0.9%
1.1%
Intellectual disturbance
0.0%
Hoarseness
1.8%
0.2%
0.2%
2.2%
0.4%
Diarrhoea
20.7%
1.7%
Anaemia
26.1%
54.1%
2.5%
4.2%
0.9%
Unconsciousness 0.9%
2.0%
2.9%
Remaining
25.3%
19.0 %
34.1%
8.8%
16.8%
Unhealing
Death

unfavourable prognosis, the rest have favourable prognosis, and the 5.7% of the whole patients have pharyngeal symptom without abdominal enlargement symptom, and the 4.9% of these patients have favourable prognosis, and the rest is unfavourable. The similar statements can be made as for the other symptoms, i.e., physico-disturbance, exanthem and etc. The remaining 34.1% patients who do not have the above symptom at all have favourable prognosis with the ratio 25.3 : 8.8 as is seen in the chart.

Therefore, if the relations in the chart hold generally, we can tell the patients "you are getting better," when some symptom on the right hand side of the chart is observed, or "you are getting worse" when some symptom on the left hand side of the chart is observed. In doing so, we will have the success rate 77.2%.

3. Forecasting of prognosis by a quantifying method

We assume that the scores δx for symptoms are given by some quantifying method, and we shall indicate the prognosis for patient i by the sum of scores of symptoms which patient i has. We want to give the weight x to each symptom so that the distribution of the sum $\sum \delta x$ for favourable prognosis is discriminated from that for unfavourable prognosis as well as possible.

Now, let R be the number of symptoms, n the total number of patients, n_t the number of patients who belong to the t-th prognosis. We have clearly $n=\sum_{t=1}^{2} n_t$. Further, let x_{j1} denote the score for symptom j when it exists, and x_{j2} the score when it does not exist. Putting

$$\delta_i(jk)=\begin{cases}1 & \text{if patient } i \text{ has the category } k \text{ in the symptom } j,\\ 0 & \text{otherwise,}\end{cases}$$

we want to use the score $\alpha_i=\sum_{j=1}^{R}\sum_{k=1}^{2}\delta_i(jk)x_{jk}$ as the score of patient i's prognosis, where

$$\sum_{k=1}^{2}\delta_i(jk)=1$$

$$\sum_{j=1}^{R}\sum_{k=1}^{2}\delta_i(jk)=R\,.$$

Clearly we have $\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(\alpha_i-\bar{\alpha})^2$ as the total variance with respect to the total patients, where $\bar{\alpha}=\frac{1}{n}\sum^{n}\alpha_i$. Now we wish to determine x so that the correlation ratio $\eta^2=\sigma_b^2/\sigma^2$ is maximum, where σ_b^2 is the variance between prognosis. From $\frac{\partial \eta^2}{\partial x_{uv}}=0$, assuming $\underset{R}{\bar{\alpha}}=0$ without loss of generality, we obtain

$$\sum_{j=1}^{R} \{nh_{u1}(j1)-n_{u1}n_{j1}\}x_j=\eta^2 \sum_{l=1}^{R} \{nf_{u1}(l1)-n_{u1}n_{l1}\}x_l \ .$$

where

$$f_{jk}(lm)=\sum_{i=1}^{n} \delta_i(jk)\delta_i(lm)$$

$$h_{uv}(jk)=\sum_{t=1}^{2} \frac{g_t(jk)\, g_t(uv)}{n_t}$$

$g_t(jk)$=sum of $\delta_i(jk)$ when patient i belongs to the prognosis t

and

$$\frac{x_{j1}}{x_{j2}}=x_j$$

Let $\boldsymbol{H}$ be the matrix $\{n\,h_{u1}(j1)-n_{u1}n_{j1}\}$, $\boldsymbol{F}$ the matrix $\{nf_{u1}(l1)-n_{u1}n_{l1}\}$, $\boldsymbol{X}$ the vector (x_j). Then the above equation is written as

$$\boldsymbol{HX}=\eta^2\boldsymbol{FX} \qquad (*)$$

Consequently, our problem is to obtain x_j which maximize η^2. Now the above equation (*) is reduced to the simple form in the following way.

Let p_j, q_j be the relative frequencies with which the symptom j appears in the favourable condition and the unfavourable condition, respectively. The matrix $\boldsymbol{H}$ is then given by the product of two vectors as follows.

$$\boldsymbol{H}=n_1n_2\begin{pmatrix} p_1-q_1 \\ p_2-q_2 \\ \vdots \\ p_R-q_R \end{pmatrix}(p_1-q_1,\ p_2-q_2,\ \cdots,\ p_R-q_R)=n_1n_2\boldsymbol{AA'}$$

Let β_j be the relative frequency of the symptom j as to the whole patients. When we assume that the symptom j is independent of the symptom u, the (j, u)-element in the matrix $\boldsymbol{F}$ is given by $n\beta_j\beta_u(j \neq u)$. But the symptom j is actually not independent of symptom u. Therefore, when we denote the difference between the (j, u)-elements of the independent and dependent cases by $\varepsilon_{ju}(n\beta_j\beta_u)$, the (j, u)-element of the dependent cases is given by $nf_{j1}(u1)-n_{j1}n_{u1}=n^2\varepsilon_{ju}\beta_j\beta_u$. Then the j-th diagonal element in the matrix $\boldsymbol{F}$ is given by $nf_{j1}(j1)-n_{j1}^2=n^2\beta_j \times(1-\beta_j)$. Thus we have

$$F=n^2\begin{pmatrix} \beta_1(1-\beta_1) & \varepsilon_{12}\beta_1\beta_2 & \varepsilon_{13}\beta_1\beta_3 \cdots\cdots\cdots \\ \varepsilon_{21}\beta_2\beta_1 & \beta_2(1-\beta_2) & \beta_{23}\beta_2\beta_3 \cdots\cdots\cdots \\ \varepsilon_{31}\beta_3\beta_1 & \beta_{32}\beta_3\beta_2 & \beta_3(1-\beta_3) \cdots\cdots \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \end{pmatrix}=n^2F_1$$

The equation $(*)$ is written as

$$AA'X=\eta^2\frac{n^2}{n_1n_2}F_1X,$$

that is,

$$F^{-1}AA'X=\eta^2\frac{n^2}{n_1n_2}X. \qquad \left(\substack{*\\*}\right)$$

If we put $F^{-1}A=B$, $\left(\substack{*\\*}\right)$ is rewritten as

$$(A,\ X)B=\eta^2\frac{n^2}{n_1n_2}X. \qquad \left(\substack{*\\*\\*}\right)$$

Hence, X is the solution of the above equation, and the eigenvalue η^2 is given by $\frac{n_1n_2}{n^2}(A,\ B)$. Here $\eta\neq 0$ implies that the solution of $\left(\substack{*\\*\\*}\right)$ which is independent of the vector B does not exist and the eigenvalue of $\left(\substack{*\\*\\*}\right)$ does not exist except the above value, and $\eta=0$ implies that all solutions of $\left(\substack{*\\*\\*}\right)$ are orthogonal to the vector A.

As the result, we obtained $\eta^2=0.2360$, and the scores were tabulated as follows.

Table of Scores

Symptom	existence	non-existence
abdominal enlargement	−6.5	0.4
pharyngeal	2.5	−0.2
physico-disturbance	−6.1	0.8
exanthem	1.3	−0.1
development disturbance	−2.4	0.1
intellectual disturbance	−4.3	0.1
hoarseness	0.9	0.0
diarrhoea	1.0	−0.5
anaemia	−3.5	0.3
unconsciousness	−3.5	0.3

We calculated the scores of 2286 favourable prognosis and 1273 unfavourable prognosis by the table of scores, and we obtained -1.15 as a boundary value which minimizes the failure rate. Therefore, if the score of prognosis of some patient should be lower than -1.15, we could forecast "you may be getting worse" and if the score should be higher than -1.15, we could forecast "you may be getting better". When we forecast as this we shall have the 22.9% failure rate. That is, if the scores in the above table hold generally, we can forecast prognosis with the success rate 77.2%, which coincides with that by the chart in the preceding section. Therefore it is important to test whether the scores in the above table hold generally or not. But, this depends on the future medical experiment.

THE INSTITUTE OF STATISTICAL MATHEMATICS
THE TOKYO UNIVERSITY HOSPITAL

REFERENCES

[1] L. Guttman: An approach for quantifying paired comparison and rank order, *Annals of Mathematical Statistics*, Vol. 17, No. 2, 1946.

[2] C. Hayashi: On the prediction of phenomena from qualitative data and the quantifying of qualitative data from the mathematico-statistical point of view, *Annals of the Institute of Statistical Mathematics*, Vol. III, No. 2, 1952.