The Relationship Between the Number of Disease Names and the Number of Patients —Observations on inpatients in a university hospital—

Kazuko Yamamoto

Department of Hygiene and Public Health, Osaka Medical College, Takatsuki

INTRODUCTION

The total number of named diseases for patients who received medical treatment during a given period exceeds the number of patients since some patients have more than one disease. However, the number of disease names is usually less than the number of patients since some patients have the same disease.

The number of disease names is essentially limited and therefore can be expected to approach a specified upper limit, even if it increases sharply at first with an increase in the number of patients receiving treatment.

If the relationship between the number of patients and the number of disease names were given by a numerical formula, it would not only indicate the aspects of a clinical department or hospital, but forecast the upper limit for the number of disease names. Therefore it would be useful when designing medical record data base systems using computers.

Furthermore, it would clarify the clinical experience of physicians as to the number of patients treated, and help estimate the number of patients required to train physicians in each training center.

In the present study the relationship between the number of patients and disease names was expressed in a numerical formula using patients who had been discharged from the Osaka Medical College Hospital as subjects and observations concerning each department were also made.

METHODS

Of the patients who were discharged from the Osaka Medical College Hospital between January 1, 1979 and May, 1980, 6,682 (from 13 clinical departments) recorded in the medical record data base¹⁾, were used as subjects. Monthly cumulative frequencies of the number of disease names and of patients were found from the disease names, and the relationship between disease names and the number of patients was examined.

Exponential curves²⁾ were plotted for the above data and a theoretical curve expressing the increase in the number of disease names when there was an increase in the number of patients was made in order to forecast numbers of disease names.

Finally, 1394 new patients, at random sampling, who had been discharged during 1978, were added. And it was examined whether they were on the theoretical curve or not.

In this hospital, physicians write disease names in Japanese, English, German or Latin. The language is then entered as is to the medical record data base. Therefore occasionally, names of the same disease in different languages are regarded as different names.

RESULTS

The number of patients used in this analysis totaled 6,682, total number of diseases 11,390 and the number of disease names 2,539.

Table 1 shows the number of patients and the number of disease names per clinical department, Table 2 show the kinds of diseases and Table 3 the kind of diseases per department.

The hospital has Departments of Neurosurgery, Dermatology and Psychiatry in addition to the departments shown in Table 1. Accordingly, mental disorders and diseases of the skin and subcutaneous tissue in Table 1 are few as compared with actual diseases of the

Table 1 The number of patients and the number of disease names by the clinical department

Department	Patients	Total No. of disease types	No. of disease names
1st Internal Medicine	456	870	371
2nd Internal Medicine	503	840	367
3rd Internal Medicine	570	1,057	298
Pediatrics	493	653	263
Thoracic Surgery	367	445	123
General & Digestive System Surgery	656	783	255
Orthopedic Surgery	545	610	238
Urology	309	479	170
Otolaryngology	602	825	174
Ophthalmology	701	1,642	290
Obstetrics	809	2,060	252
Obstetrics (Neonatal Care)	91	103	37
Gynecology	565	991	318
Radiology	15	32	21
Total	6,682	11,390	*3,177

* overlap

Fig. 1 Relationship between number of disease names and number of patients

Fig. 2 Theoretical distribution curve as plotted for each department

Table 2 The kind of diseases (ICD)

Kind of Diseases (ICD)	Total No. of disease types (%)	No. of disease names (%)
I. Infective and parasitic diseases	248 (2.2)	101 (4.0)
II. Neoplasms	1,355 (11.9)	315 (12.3)
III. Endocrine, nutritional and metabolic diseases	374 (3.3)	89 (3.5)
IV. Diseases of blood and blood forming organs	150 (1.3)	41 (1.6)
V. Mental disorders	21 (0.2)	18 (0.7)
VI. Diseases of the nervous system and sense organs	1,679 (14.6)	303 (11.9)
VII. Diseases of the circulatory system	1,270 (11.2)	238 (9.4)
VIII. Diseases of the respiratory system	775 (6.8)	108 (4.3)
IX. Diseases of the digestive system	957 (8.4)	264 (10.4)
X. Diseases of the genito-urinary system	818 (7.2)	223 (8.8)
XI. Complications of pregnancy and childbirth	2,186 (19.1)	282 (11.1)
XII. Diseases of the skin and subcutaneous tissue	33 (0.3)	25 (1.0)
XIII. Diseases of the musculoskeletal system and connective tissue	249 (2.2)	84 (3.3)
XIV. Congenital anomalies	531 (4.7)	121 (4.8)
XV. Certain causes of perinatal mortality	97 (0.9)	30 (1.2)
XVI. Symptoms and ill-defined conditions	108 (0.9)	63 (2.5)
XVII. Accidents, poisonings and violence	275 (2.4)	145 (5.7)
EXVII. Accidents, poisonings and violence (external causes)	201 (1.8)	61 (2.4)
V. Other	63 (0.6)	28 (1.1)
Total	11,390 (100.0)	2,539 (100.0)

Fig. 3 Theoretical distribution curve per department (percent)

Fig. 4 The correlation between the number of disease names falling with the 90th percentile and the number of patients

hospital. Fig. 1 shows the relationship between the numbers of patients and disease names. The white circles in the drawing denote values actually found.

When the exponential curve was applied, the theoretical distribution of the number of disease names (y) with an increase in the number of patients (x) came to:

$$y = 3,567 \{1 - e^{-0.0001767(x+372)}\}$$

As a result of the goodness-of-fit test, the observational distribution of the number of

Table 3 The kind of diseases by the clinical department (%)

[illegible]

disease names fit the theoretical distribution shown in the formula, at $\chi^2=14.816$ (degree of freedom=16). The black circles denote values derived for the 1,394 new patients added afterward. They were all on the theoretical distribution curve, thus proving that the formula for the theoretical distribution is correct up to this point.

From the above, it can be seen that the number of disease names dose not increase linearly with an increase in the number of patients, but that the rate of increase in the number of disease names decreases with an increase in the number of patients. The number of disease names reaches its upper limit at 3,567.

Fig. 1 shows that the 90th percentile for 3,567 corresponds to 3,210, the number of disease names, and the number of patients then is 12,650. Fig. 2 shows the theoretical distribution curve for the number of disease names found for each clinical department. Results of the goodness-to-fit test fit the theoretical distribution curve for all clinical departments. Fig. 3 shows the theoretical distribution curve for each department in Fig. 2 expressed in percentages, clearly demonstrating that the theoretical distribution curve for the number of disease names differs from one clinical department to another.

It can be seen from Fig. 2 that the number of disease names is greatest for the 1st (mainly psychiatric diseases) and 2nd Departments of Internal Medicine (mainly digestive system diseases) followed by the 3rd Department of Internal Medicine (mainly cardiovascular diseases), the Departments of General & Digestive System Surgery, Gynecology, Pediatrics, Urology, Obstetrics (exclusive of neonates), Ophthalmology, Orthopedic Surgery and Thoracic Surgery in order and is lowest for the Department of Otolaryngology. The Departments of Radiology and Neonatal Obstetrics were excluded because of the small number of cases.

As is clear from Fig. 3, the number of patients falling the 90th percentile in the number of disease names is lowest for the Department of Orthopedic Surgery followed by the Department of Otolaryngology, Ophthalmology and Gynecology, each with about 1,000 patients, and it increases in the Departments of Pediatrics, Urology, 1st, 2nd and 3rd Internal Medicine and Thoracic Surgery, Gynecology (1,785) to the General & Digestive System Surgery (1,974).

Fig. 4 shows the correlation between the number of disease names falling with the 90th percentile and the number of patients. The coefficient of correlation is $r=+0.248$, which is relatively low.

DISCUSSION

Disease names are important information for the medical record data base.

In Japan, many hospitals have adopted the International Classification of Diseases³⁾ (ICD). As a method for putting disease names into a computer, therefore, hospitals put either only ICD codes or put in the disease name per se in and have it converted into an ICD code automatically by the computer⁴⁾.

In any case, coding disease names to the ICD code takes much time, and various other means have been devised⁵⁻⁸⁾. There also is a method in which artificial words are put in the computer for classification without coding of disease names.

If an upper limit of disease names could be predicted, it would be very useful in designing the automatic ICD conversion system by computer.

As a result of the present study, the relationship between the number of disease names and the number of patients was characterized by an exponential type of increase and the upper limit value for cumulative disease names in this hospital was estimated at 3,567. Since this does not require much memory capacity, it is possible to house all disease names

for in-patients within the computer.

The 90th percentile for 3,567 corresponds to 3,210 disease names, at which point the number of patients is 12,650. This means that if disease names for 12,650 patients are coded, and if a code-disease name conversion list is prepared in the computer (in other words, preparation of a code-disease name conversion list for 3,210 disease names), 90 percent of the diseases can be automatically converted into disease codes.

Kaihara et al.⁴⁾ have developed a system in which disease names written in the kana syllabary are automatically converted into codes. They claim that applying this system to out-patients, more than 90 percent of diseases would be possible converted into disease codes by a code-disease names conversion list for 1,800 disease names. Compared with the data from this hospital, no sweeping statement can be made because of the difference in the kinds of diseases between in- and out-patients, the method of signifying disease names and the status of patients (as in- or out-patients). However, the mixture of kana, English, German and Latin expressions yielded 3,210 disease names, so that values similar to those of Kaihara might be obtained. Future studies should clarify the relation between the number of patients, and the number of diagnoses by changing names in different languages to one language.

The theoretical distribution curve for the number of disease names differs according to the department. Accordingly, the upper limit value for the cumulative number of disease names also differs between departments. Generally, the upper limit value for the number of disease names is low in those department whose scope of treatment covers specific organs and high in those whose scope of treatment covers larger areas of the body. The coefficient of correlation between the number of disease names corresponding to the 90th percentile of the upper limit value for the number of disease names and the number of patients then is not high ($r = +0.248$).

The total number of disease names in the department concerned is apparently affected by the number of diseases per patient and one-sidedness of the patients who visit the department.

CONCLUSION

For the purpose of expressing in a numerical formula the relationship between the number of disease names and the number of patients, we examined this relationship for 6,682 patients discharged from the Osaka Medical College Hospital between January 1, 1979 and May, 1980. The total number of diseases was 11,390, and the number of disease names 2,539.

As a result, the theoretical distribution of the number of disease names (y) with an increase in the number of patients (x) was shown to be:

$$y = 3567 \{1 - e^{-0.0001767(x+372)}\}$$

From this, the upper limit for the cumulative number of disease names is estimated at 3,567.

The numerical formula differs from department to department and the upper limit value for the number of disease names was also found to differ according to the department.

REFERENCES

- 1) Yamamoto, K.: Medical record management system using standard MUMPS, Report for the MUMPS Software Meeting, 15-4, 1-66 (1979).
- 2) Kashiwagi, C.: Analysis of medical statistics, Asakura Press, Tokyo (1979).
- 3) World Health Organization: International classification of diseases, 1975 Revision, Vol. 1, WHO (1975).

- 4) Kaihara, S., et al.: Automatic coding of diagnosis by a computer, Hospital Administration, 14, 173-178 (1977).
- 5) Barnett, G. O., et al.: COSTAR-A computer-based medical information system for ambulatory care, Proceedings of the IEEE, 67-9, 1226-1236 (1979).
- 6) Hashimoto, N., et al.: Ambulatory disease registration system at a community hospital, Proceedings of EDIS '78, 169-172, Kansai Institute, Osaka (1978).
- 7) Furukawa, T., et al.: A study for the computerized medical record system by application of the artificial language, Jap. Journal of Med. Elec. and Biol. Eng., 14, 195 (1976).
- 8) Danno, M., et al.: A database and programs to file surgery patient records available to clinical research, Proceedings of MEDIS '78, 145-148, Kansai Institute, Osaka (1978).

大学附属病院の入院患者数と病名数の関係

山 本 和 子

大阪医科大学衛生学教室

病名数は、受診患者数の増加にともない、はじめは急速に増加していても、いずれは特定の上限値に漸近すると予想される。この患者数と病名数の関係を数式で示し得るならば、診療科あるいは医療施設の特徴を数的に表現することができるばかりでなく、患者数から病名数の上限値を予測することができるので、コンピュータによる病歴データベースシステムの設計に際して有用である。また、医師の診療経験を、扱った患者数で評価することの意義が明らかになり、医師の養成に必要とする患者数を研修機関ごとに推算することにも役立つなど、医療施設評価、病歴管理、医学教育等の面において有益である。

そこで、病名数と患者数の関係を数式化することを目的とし、大阪医科大学附属病院を昭和54年1月1日から昭和55年5月末日までに退院した患者6,682名（疾患延件数11,390件、病名数2,539）を資料として、病名数と患者数の関係を調べた結果、患者数（ x ）にともなう病名数（ y ）の理論的分布は

$$y=3567\{1-e^{-0.0001787(x+872)}\}$$

となることがわかった。

これより、当病院の累積病名数の上限値は3,567と推定した。

また、診療科それぞれの数式は異なり、病名数の上限値も異なることを明らかにした。

(受付 1980年9月25日)