J. Japan Statist. Soc.
Vol. 35 No. 1 2005 135–145

IMPROVING THE STRATIFICATION OF MEDICAL INSTITUTIONS FOR STRATIFIED SAMPLING IN THE PATIENT SURVEY

Takashi Sozu*, Chikuma Hamada** and Isao Yoshimura**

The Patient Survey is a designated statistical survey conducted every three years with the objective of obtaining basic data on the current status of patients in medical institutions in Japan. Since stratified sampling is used in this survey, suitable construction of strata is essential for achieving low error rates in the estimation of the number of patients having various diseases. We investigated the performance of the current stratification through a correspondence analysis between disease categories and clinic categories and found that patients having diseases related to mental or behavioural disorders were not well sampled by the current stratification method. Therefore, we proposed to create a clinic category, "psychiatry," as a new stratum for sampling and examined the effect of this stratification on the precision in the estimation through a Monte Carlo simulation. The simulation results indicate that the new stratification achieved a decrease of approximately seven points in the standard error rate of the estimated number of patients with "mental and behavioural disorders."

Key words and phrases: Correspondence analysis, Patient Survey, ratio estimation, standard error rate, stratified random sampling.

1. Introduction

The Patient Survey has been conducted since 1953, as Designated Statistics No. 66, to ascertain the current status of patients who use hospitals and clinics (hereinafter "medical institutions"). The survey findings are published in a formal report and on the Website of the Ministry of Health, Labour and Welfare. The estimated number of patients for each disease is reported, classified by category, such as patient type (inpatient or outpatient) and gender, for the entire country, by prefecture and by secondary medical care area (the broad administrative areas providing the medical care).

The prefectures utilize the survey findings as basic data for regional health care planning. Therefore, optimal and precise estimation of the number of patients has become a high-priority issue.

In the Patient Survey, stratified random sampling is used to select sample institutions. That is, medical institutions that satisfy certain criteria are classified into three strata; hospital, general clinic, and dental clinic. The medical institutions in each stratum are further divided into multiple strata according to

Received June 8, 2004. Revised December 1, 2004. Accepted February 10, 2005.

*Graduate School of Engineering, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan. Email: sozu@ms.kagu.tus.ac.jp

**Faculty of Engineering, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan.

the institution size or medical department. Sample institutions for each stratum are selected by prefecture or secondary medical care area. The diseases and demographic factors of all patients in the selected institutions on the designated survey date are surveyed.

The Static Survey and Dynamic Survey of Medical Care Institutions is Designated Statistics No.65, a census conducted simultaneously with the Patient Survey to ascertain the actual state of medical institutions. The Static Survey (hereinafter "Institutional Survey") investigates the total number of patients during a designated period of time. In the Patient Survey, a ratio estimation using the total number of patients obtained from the Institutional Survey as the auxiliary variable, is applied to estimate the number of patients.

In order to estimate the number of patients with each disease with uniformly high precision, the medical institutions must be appropriately stratified, ensuring that there is no tendency to omit from the survey patients having certain diseases. Although the stratification of the medical institutions is revised each time the Patient Survey is conducted, the question of whether medical institutions that treat certain diseases tend to be omitted from the sample is not being examined, as discussed in Yoshimura (2004).

In this paper we used the data from the 1996 and 1999 Patient Surveys and Institutional Surveys to consider whether the current institutional stratification method should be improved for the next Patient Survey in 2005. Our findings are described in the following. In Section 2, we provide an outline of the Patient Survey. In Section 3, we investigate whether the current stratification method is appropriate. In Section 4, we provide a proposal for improving the stratification of general clinics to answer the problem in the previous section, and examine the effect of this stratification on the precision in the estimation through a Monte Carlo simulation. In Section 5, we summarize the findings.

2. Outline of Patient Survey

The target number of medical institutions to be sampled in the Patient Survey in 1996 and 1999 was 9,000 hospitals, 6,000 general clinics, and 1,000 dental clinics. Sampling rates were approximately 70% for hospitals, 7% for general clinics, and 2% for dental clinics.

Because hospitals have a relatively large number of patients, having all patients provide detailed descriptions in a questionnaire would be a great burden. Consequently, for hospitals, a detailed survey is conducted of patients born in odd-numbered dates while a simplified survey in which the items are limited to gender, date of birth, and patient type, is conducted of patients born in even-numbered dates. The questionnaire for the detailed survey is called the odd-numbered questionnaire and that for the simplified survey, the even-numbered questionnaire.

In the Patient Survey the diseases of the patients are classified according to the disease classification scheme specified by the Ministry of Health, Labour and Welfare. This classification has been based on the ICD-10 (The 10th Revision of

Table 1. Classification of 20 groups of diseases.

No.	Label for classification
1	Certain infectious and parasitic diseases
2	Neoplasms
3	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4	Endocrine, nutritional and metabolic diseases
5	Mental and behavioural disorders
6	Diseases of the nervous system
7	Diseases of the eye and adnexa
8	Diseases of the ear and mastoid process
9	Diseases of the circulatory system
10	Diseases of the respiratory system
11	Diseases of the digestive system
12	Diseases of the skin and subcutaneous tissue
13	Diseases of the musculoskeletal system and connective tissue
14	Diseases of the genitourinary system
15	Pregnancy, childbirth and the puerperium
16	Certain conditions originating in the perinatal period
17	Congenital malformations, deformations and chromosomal abnormalities
18	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19	Injury, poisoning and certain other consequences of external causes
20	Factors influencing health status and contact with health services

the International Statistical Classification of Diseases and Related Health Problems; WHO definition) since the 1996 survey. Although there are three kinds of classification methods for grouping the diseases used in the report of the Patient Survey, in every method, the basic classification includes the 20 items shown in Table 1. Hereafter, these 20 classifications are termed "classifications of diseases." The classifications of diseases for dental clinics are limited to three items (disease numbers 11, 19, and 20).

In the hospital odd-numbered questionnaire and the general clinic questionnaire, each medical department is classified into one of the 36 departments listed in Table 2. However, in the dental clinic questionnaire, medical departments are not surveyed.

3. Review of current stratification

In this section, we review the appropriateness of the current stratification method for hospitals, general clinics, and dental clinics.

3.1. Hospitals

In recent years, the fourth amendment of the Medical Service Law enacted in 2001 and the National Hospital Organization Law enacted in 2003, have resulted in rapid change in the legal environment of hospitals. Therefore, a stratification method corresponding to such changes is required. However we cannot analyze

Table 2. Classification of medical departments.

No.	Label for classification	No.	Label for classification
01	Internal medicine	19	Pediatric surgery
02	Respiratory system	20	Obstetrics and gynecology
03	Digestive system	21	Obstetrics
04	Circulatory system	22	Gynecology
05	Pediatrics	23	Ophthalmology
06	Psychiatry	24	Otorhinolaryngology
07	Neurology	25	Broncho-Esophagology
08	Internal medicine of neurology	26	Dermatology
09	Psychosomatic internal medicine	27	Urology
10	Allergy	28	Venereology
11	Rheumatology	29	Proctology
12	Surgery	30	Rehabilitation
13	Orthopedic surgery	31	Radiology
14	Plastic surgery	32	Anesthesiology
15	Cosmetic surgery	33	Dentistry
16	Cranial nerve surgery	34	Orthodontics
17	Respiratory surgery	35	Pediatric dentistry
18	Cardiovascular surgery	36	Dentaloral surgery

the actual condition of such changes correctly from the past Patient Survey data referred to in this paper. With respect to hospitals, future research needs to reflect on the effect on the stratification method of such environmental changes.

3.2. General clinics

In recent years the number of general clinics with beds has decreased and that with no beds has increased. As a result, most patients at general clinics are outpatients. For this reason, we limit the discussion to outpatients.

3.2.1. Current stratification

General clinics usually have a few medical departments, and as a result they have their own medical specialties. This is why the current stratification method for general clinics establishes a "principal medical department" for each clinic and divides clinics into 17 strata according to "principal medical department" and "availability of beds."

When a clinic has multiple medical departments, a single principal medical department is determined. In order of priority, the criteria for this determination are: 1) the department with the highest number of patients, 2) the medical specialty of the clinic director or full-time physician, and 3) the decision of the clinic director. With regard to the composition of the 17 strata, as shown in Table 3, clinics are divided into nine strata according to the principal medical department. Internal medicine, Surgery, Orthopedic surgery, and Other department are subdivided into three categories according to availability of beds: 1) having no beds, 2) having long-term care beds, and 3) having other beds.

The numbers in parentheses following the principal medical departments

Table 3. Stratification for principal medical department.

No.	Stratum	Principal medical department
1	Internal medicine	Internal medicine(01), Respiratory system(02), Digestive system(03), Circulatory system(04), Psychiatry(06), Neurology(07), Internal medicine of neurology(08), Psychosomatic internal medicine(09), Allergy(10), Rheumatology(11)
2	Pediatrics	Pediatrics(05)
3	Surgery	Surgery(12), Cranial nerve surgery(16), Respiratory surgery(17), Cardiovascular surgery(18), Pediatric surgery(19), Proctology(29)
4	Orthopedic surgery	Orthopedic surgery(13), Plastic surgery(14), Cosmetic surgery(15)
5	Obstetrics and gynecology	Obstetrics and gynecology(20), Obstetrics(21), Gynecology(22)
6	Ophthalmology	Ophthalmology(23)
7	Otorhinolaryngology	Otorhinolaryngology(24), Broncho-Esophagology(25)
8	Dermatology	Dermatology(26), Urology(27), Venereology(28)
9	Other department	Rehabilitation(30), Radiology(31), Anesthesiology(32), Dentistry(33), Orthodontics(34), Pediatric dentistry(35), Dentaloral surgery(36), Others(99)

in Table 3 are the numbers of medical departments in Table 2. Clinics that are not classified into any of these categories are included in "Others (99)." Accordingly, the total number of medical departments is 37: the 36 departments in the questionnaire and Others.

3.2.2. Correspondence analysis

For stratification according to principal medical department to be effective, it is desirable that strata and disease classification correspond well. Expressed another way, the strata should be created so that medical departments where patients with similar diseases receive treatment are included in the same stratum. When patients with a certain disease are dispersed by low frequencies across many strata, the effect of a small sampling rate of approximately 7% tends to result in a large sampling error.

A well-established statistical approach for examining whether the current stratification corresponds to disease classifications in the sense described above, would be to perform correspondence analysis (Hayashi's quantification theory type III) on the 37 principal medical departments and 20 groups of diseases. That is, it is appropriate to group medical departments and to examine whether the resulting groups correspond to current strata groups. According to correspondence analysis theory, because similar scores are given to closely-related medical departments and classifications of diseases, such categories will be grouped in multi-dimensional space. For this reason, we can examine the appropriateness of the current stratification by representing each category on a scatter plot using

the multi-dimensional scores and visually examining the relationship between principal medical departments and classifications of diseases.

We performed a correspondence analysis on a two-way contingency table of the 37 principal medical departments and the 20 groups of diseases using the data from the 1996 and 1999 Patient Surveys. The results are shown in Table 4. The number of outpatients in the data was approximately 283,000 for the 1996 survey and 252,000 for the 1999 survey.

According to Table 4, for both the 1996 survey and the 1999 survey the cumulative contribution rate up to the sixth dimension exceeds 90%. Plotting scores for each category in the 1999 data yields Figures 1 to 3. In these figures, the left-hand side is the scatter plot of the principal medical departments and the right-hand side is that of the classifications of diseases. These figures show that the current stratification method corresponds well for the classifications of diseases. For instance, on dimension 1 in Figure 1, Ophthalmology (23) is located in the same positive area as disease number 7, "diseases of the eye and adnexa." This corresponds to the Stratum of Ophthalmology (No. 6 in Table 3) in the current stratification method and means that this stratum is valid. Similarly, on dimension 6 in Figure 3, the group comprising Obstetrics and gynecology (20), Obstetrics (21), and Gynecology (22) is located in the same area as that of disease numbers 14, 15, and 16. This group corresponds to the Stratum of Obstetrics and gynecology (No. 5 in Table 3) and means that this stratum is valid.

However, the result found on dimension 4 in Figure 2 should be noted. Despite the close relationship between the medical departments of Psychiatry (06), Neurology (07), and Psychosomatic internal medicine (09) and disease number

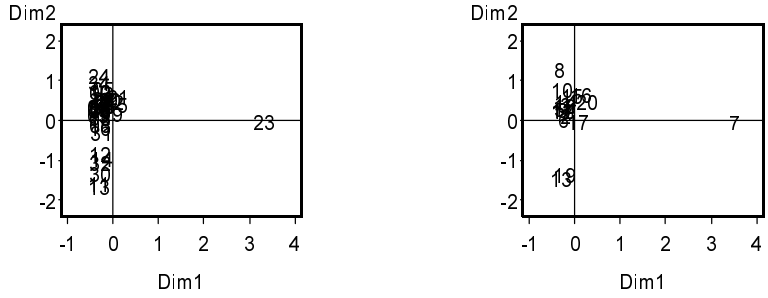
Table 4. Results of correspondence analysis.

(a) Result in 1996

Dimension	Singular value	Contribution rate (%)	Cumulative contribution rate (%)
1	0.93	23.9	23.9
2	0.79	17.3	41.1
3	0.73	14.8	55.9
4	0.71	14.0	69.9
5	0.66	12.0	81.9
6	0.57	9.2	91.0

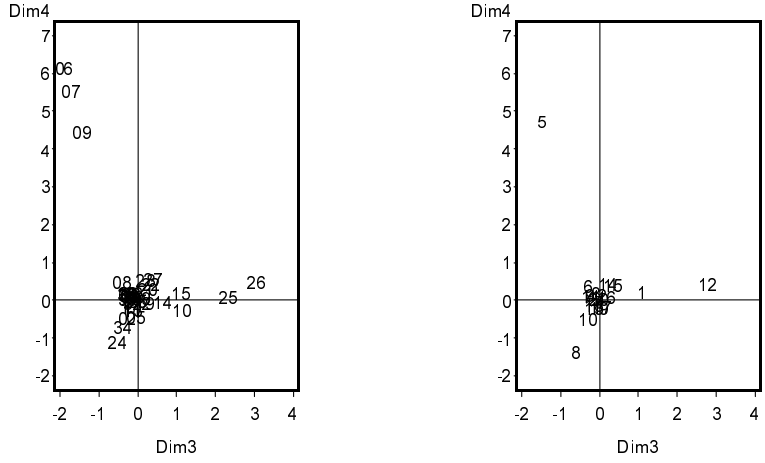
(b) Result in 1999

Dimension	Singular value	Contribution rate (%)	Cumulative contribution rate (%)
1	0.92	23.5	23.5
2	0.81	18.1	41.5
3	0.73	14.7	56.2
4	0.70	13.7	69.8
5	0.66	11.9	81.8
6	0.58	9.3	91.1



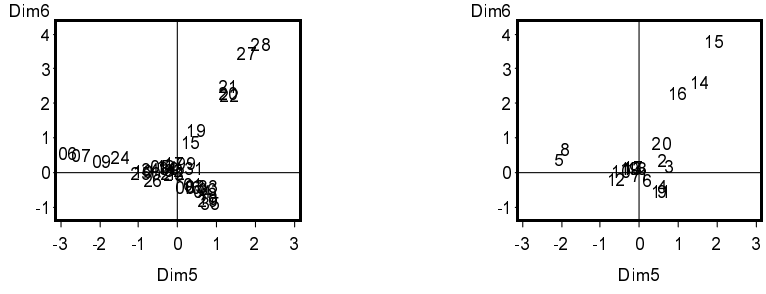
Principal medical departments Classifications of diseases

Figure 1. Scatter plots of scores for dimension 1 and dimension 2.



Principal medical departments Classifications of diseases

Figure 2. Scatter plots of scores for dimension 3 and dimension 4.



Principal medical departments Classifications of diseases

Figure 3. Scatter plots of scores for dimension 5 and dimension 6.

5, "mental and behavioural disorders," they are not grouped into a single stratum. Although the number of these medical departments has recently increased and patients with "mental and behavioural disorders" are receiving treatment in those departments, this trend was not reflected in the 1999 survey stratification. These results should be reflected in the stratification of the next Patient Survey.

3.3. Dental clinics

The current stratification method used for dental clinics is stratification by prefecture only. However, because respondents at dental clinics are outpatients only and classifications of diseases are limited, the current stratification method is likely to be sufficient for the next Patient Survey.

4. Proposal for the stratification for general clinics

4.1. Psychiatry as a new stratum

The estimated number of outpatients with "mental and behavioural disorders" discussed in the previous section is 62,900 for the 1999 Patient Survey reports (2001a). Although the number of patients is not extremely low, the precision in the estimation is insufficient. This is probably because institutions with medical departments where patients receive treatment for these diseases are not appropriately sampled.

Therefore, in this paper we propose to add a new stratum "Psychiatry" comprising institutions whose principal medical department is Psychiatry, Neurology, or Psychosomatic internal medicine. Hereafter, we refer to the stratification modified by adding this stratum to the current strata as the "proposed method" and to the stratification without modification as the "current method."

4.2. Design of the simulation study

Simultaneous application of the current method and the proposed method would reveal whether the proposed method is more precise than the current method in estimating the number of patients. However, it is impossible to perform such an experiment. Therefore, in this paper we evaluate the proposed method through a Monte Carlo simulation involving the following steps:

Step1. Generate a virtual population for the Patient Survey using data from the 1996 and 1999 Patient Surveys. Similarly, generate a virtual population for the Institutional Survey using data from the 1996 and 1999 Institutional Surveys.

Step2. Select 4,757 medical institutions from the virtual population for the Patient Survey using stratified random sampling based on the current and proposed methods.

Step3. Estimate the number of patients and calculate the standard error rates for 20 groups of diseases using the data for medical institutions selected in Step 2 and data for the virtual population for the Institutional Survey.

Step4. Perform 1,000 iterations of Step 2 and 3.

Step5. Compare the mean value of the standard error rate for the proposed method with that for the current method.

In performing these steps we took into account the following consideration. In the 1996 Patient Survey data, the number of medical institutions actually submitting the questionnaire was 5,055, and the number of those institutions, in which the number of patients in the Institutional Survey was not missing data, was 4,892. Similarly, in the 1999 Patient Survey data, the number of such medical institutions was 4,892, and the number of those institutions with no missing data in the Institutional Survey was 4,757. Because the number of patients in the Institutional Survey is used as the auxiliary variable in the ratio estimation, medical institutions with missing patient numbers in the Institutional Survey cannot be used. Accordingly, in generating the virtual population for the Patient Survey in Step 1, we excluded those institutions with missing patient numbers. Consequently, in Step 1 we generated a virtual population of 9,649 institutions for the Patient Survey, the sum of 4,892 institutions in 1996 and 4,757 institutions in 1999. Similarly we generated a virtual population of 78,635 institutions for the Institutional Surveys.

The reason we chose 4,757 medical institutions for the random sampling in Step 2 was that the number of institutions with no missing patient numbers was 4,757 in the 1999 survey. This number is an operational one, and it is clear that almost the same sampling magnitude does not change the results. Consequently, the institution sampling rate in the simulation study was approximately 6% (4,757/78,635).

Although the availability of beds was considered for creating strata in the actual sampling, in this simulation we omitted it for the sake of simplicity. Similarly, we did not conduct stratified sampling using prefectures as strata. This is because the difference of the number of sampling institutions among prefectures in the actual data was small.

We did not use the current estimation method in Step 3, i.e., the method applying the same weight to all strata, but instead used a weighting method that took into account the differences in institutional sampling rate among strata. As Sozu *et al.* (2005) has pointed out, weighting with the reciprocal of the sampling rate for each stratum reduces the standard error rate of the estimated number of patients.

4.3. Result of the simulation

The result of the simulation is shown in Table 5. Because actual disease names are meaningless in the comparison, we have omitted them in the table. (See Table 1 for the disease names.) For disease number 5, "mental and behavioural disorders," the standard error rate for the proposed method is approximately seven points lower than for the current method. At the same time, the standard error rate for other diseases is nearly the same for both methods. This trend is apparent in Figure 4.

The high standard error rate for disease group numbers 16 and 17 results from an extremely small number of patients. It would probably be futile to attempt

Table 5. Results of the standard error rate by the current and proposed methods.

Disease No.	Current method	Proposed method	Disease No.	Current method	Proposed method
1	3.5	3.5	11	4.1	4.2
2	5.9	6.1	12	6.5	6.6
3	6.1	6.2	13	5.4	5.4
4	3.1	3.1	14	8.5	8.5
5	17.3	10.1	15	11.5	11.4
6	6.4	5.8	16	23.5	23.7
7	7.6	7.6	17	21.9	22.3
8	6.6	6.6	18	5.7	5.8
9	2.7	2.7	19	4.7	4.7
10	3.1	3.1	20	7.2	7.2

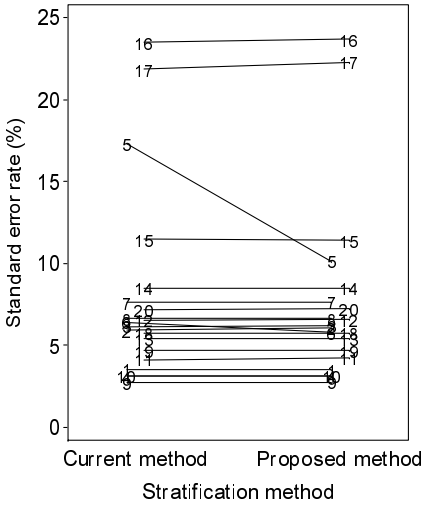


Figure 4. Change of the standard error rate.

to increase the precision in the estimation of patient numbers with these diseases using the current framework of the Patient Survey.

5. Conclusion

We noted the functional differentiation of medical institutions and patient aggregation in recent years and examined whether the current method of institutional stratification is appropriate. As a result, we found that it is possible to improve the precision in the estimation of patient numbers with "mental and behavioural disorders" by adding a Psychiatry stratum for sampling general clinics.

According to the result of tentative calculations, when the institution sam-

pling rate is simply increased by one point (that is, the number of institutions sampled is increased by nearly 800) the standard error rate of the estimated number of patients with "mental and behavioural disorders" decreases by approximately one point. Therefore, reducing the standard error rate by approximately seven points by the proposed method greatly improves the precision in the estimation. Because increasing the institutional sampling rate would entail increases in cost, labor and time, an improvement in stratification that would improve the precision in the estimation without increasing these burdens would be substantial progress. The results of this research will probably be useful in the design of the next surveys.

Acknowledgements

The authors would like to thank Dr. Kazuhiko Oe for motivating us to undertake this study, and Dr. Akifumi Yafune for explaining the actual state of patient consultation. We would also like to thank the editor and two anonymous referees for helpful comments that greatly improved the article.

This research was partially supported by a Grant-in-Aid for Research (H14–statistics–001) on Applied Use of Statistics and Information from the Ministry of Health, Labour and Welfare of Japan.

References

Health and Welfare Statistics Association (2002). Current status of national health, *Journal of Health and Welfare Statistics*, **49**(9), 197–200 (in Japanese).

Sozu, T., Hamada, C. and Yoshimura, I. (2005). A proposal to improve the Patient Survey focusing on the recent trend of increase in the missing data, *J. Japan Statist. Soc.*, **35**(1), this issue, 121–134.

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare (1998a). Patient Survey, Volume I, 1996 (in Japanese).

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare (1998b). Static and Dynamic Survey of Medical Care Institutions, Hospital Report, 1996 (in Japanese).

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare (2001a). Patient Survey, Volume I, 1999 (in Japanese).

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare (2001b). Static and Dynamic Survey of Medical Care Institutions, Hospital Report, 1999 (in Japanese).

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare (2003). Dynamic Survey of Medical Care Institutions, Hospital Report, 2001 (in Japanese).

Yoshimura, I. (2004). Study on how to establish the objects of Patient Survey, Report in the Health and Labor Sciences Research, 2002–2003 (in Japanese).