J. Japan Statist. Soc.
Vol. 35 No. 1 2005 121–134

A PROPOSAL TO IMPROVE THE PATIENT SURVEY FOCUSING ON THE RECENT TREND OF INCREASE IN THE MISSING DATA

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. One of the most important items in the report of this survey is the estimated number of patients with various diseases in each prefecture or secondary medical area. This paper shows that the amount of missing data has increased recently and has reduced the precision in the estimation of patient numbers. We propose to adopt variable weighting for ratio estimation dependent on the differences in the institutional sampling rate instead of the currently used constant weighting, as the proposed method can be adapted to take account of the increases in missing data. The proposed method can improve the precision in the estimation of patient numbers in most diseases, based on quantitative assessment conducted using the actual data from the 1996 and the 1999 Patient Survey.

Key words and phrases: Nonresponse, number of patients, Patient Survey, ratio estimation, stratified random sampling, variable weighting.

1. Introduction

Patient Surveys are designated statistical surveys intended to clarify the status of patients in medical institutions. This survey has been conducted once every three years since 1984. The next survey is scheduled to be conducted in 2005.

The results of the Patient Survey are published in a formal report and on the webpage of the Ministry of Health, Labour and Welfare. The reported results include the estimated number of patients, classified by items such as inpatient/outpatient status and gender, for the entire country, by prefecture and by secondary medical area (the broad administrative areas providing the medical care) and by disease. The estimated number of patients is used as the basic document in regional healthcare planning.

The Patient Survey encompasses all patients who used a given medical institution selected by stratified random sampling from all medical institutions (hospitals, general medical clinics, dental clinics) on the designated date of survey.

The ratio estimation method is used to estimate the number of patients with

Received May 24, 2004. Revised November 6, 2004. Accepted January 19, 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.

a particular disease. The auxiliary variable in ratio estimation is the total number of patients determined by the "Static Survey on Medical Institutions" (hereafter "Institutional Survey"). The Institutional Survey is conducted at the same time as the Patient Survey and the results include the total number of patients in all medical institutions in Japan.

In the Patient Survey, all of the institutions in the survey are required to fill out a questionnaire for each patient and submit the questionnaire to the public health center. Similarly, in the Institutional Survey, all of the institutions included in the survey are required to fill out and submit the questionnaire. However, not all institutions participating in the survey submit the questionnaires, so that some are recorded as "missing" or "nonresponse."

The reasons for the missing data include refusal to participate in the survey, nonconsultation or no patients seen on the day of the survey. However, since for given institutions with missing data, it is unknown which reason was given, data from institutions not submitting the questionnaire for the Patient Survey or the Institutional Survey cannot be used to estimate the number of patients. If it can be determined that there were indeed no patients on the day of the survey, the patient numbers for that institution can be regarded as 0, but if the reason is refusal to participate in the survey, the patient numbers cannot be regarded as 0. For these reasons, increases in the number of missing data in the Patient Survey or the Institutional Survey greatly affect the estimation of patient numbers.

In this paper, we investigate the pattern of the missing data and discuss how the missing data affects the estimation of patient numbers. In addition, we demonstrate that with missing data, modifying the ratio estimation method can improve the precision in the estimation of patient numbers by using the data from the 1996 and the 1999 Patient Survey and the Institutional Survey. In Section 2, we describe the outline of the Patient Survey. In Section 3, we discuss the relationship between the number of institutions targeted to be sampled, the actual number of institutions used in the estimation, and the number of evaluable institutions, and evaluate the effect of missing data on the estimation of patient numbers. Similarly, in Section 4, we describe the institutions with missing patient numbers in the Institutional Survey and evaluate the effect on the estimation of patient numbers. Based on the results from Sections 3 and 4, we show in Section 5 that by modifying the estimation method, it is possible to improve the precision in the estimation of patient numbers. Section 6 summarizes the conclusions.

2. Outline of Patient Survey

The Patient Survey is a statistical survey in which institutions to be sampled are selected from medical institutions throughout Japan. The questionnaires are mailed to the institutions, and responses are requested on all patients seen at that medical institution on the specified day of the survey. The questionnaire differs depending on whether the medical institution is a hospital, a general medical clinic, or a dental clinic, and the sampling rate of institutions differs for these three levels of stratification (hereafter "institutional stratification").

Stratified random sampling is used to select institutions. Hospitals are stratified into 11 strata according to "type of facility" and "number of beds," for inpatients by secondary medical area and for outpatients by prefecture, respectively. General medical clinics are stratified into 17 strata by "principal medical department" and "status of beds" for both inpatients and outpatients. Dental clinics are stratified by prefecture only.

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 is based on the revisions of International Classification of Diseases, and starting with the 1996 survey, ICD-9 (The 9th Revision of the International Statistical Classification of Diseases and Related Health Problems; WHO definition) was changed to ICD-10. In this paper, when assessing the effect on diseases, data is used only from two (1996 and 1999) of the three surveys for which permission for use was granted. Although there are three kinds of classification methods for grouping the diseases used in the report of the Patient Survey, in every method, a basic classification includes 20 items based on ICD-10 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).

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

3. "Missing" data in Patient Survey

3.1. Recent trend of missing rates

The number of institutions targeted to be sampled in 1993, 1996, and 1999 were about 7,000 hospitals, about 6,000 general medical clinics and about 1,000 dental clinics. These institutions except for institutions not in existence or not in operation are considered to be "institutions targeted for survey." From the institutions targeted for survey, institutions that refused to participate in the survey (non-cooperative), were not consulting, or had no patients on the day of the survey were subtracted to give the "evaluable institutions." In the estimation of patient numbers in the institutions targeted for survey, patient numbers in an institution that was not consulting or had no patients were regarded as 0. On the other hand, institutions that refuse to participate in the survey cannot be used in the estimation because the patient numbers cannot be determined. To make this distinction, the institutions used in the estimation are termed "sample institutions." Figure 1 shows the relationship between the institutions targeted to be sampled, the institutions targeted for survey, sample institutions, and evaluable institutions.

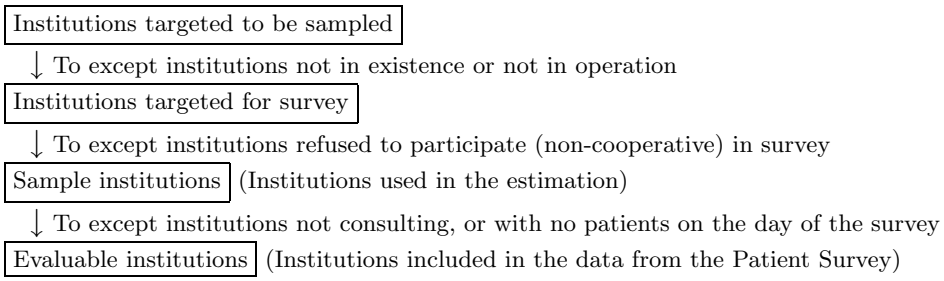

Figure 1. Category of institutions discriminated between survey levels.

For a number of reasons in the actual Patient Survey, the number of evaluable institutions may be smaller than the number of institutions targeted to be sampled. Table 2 shows the number of evaluable institutions in the three surveys conducted between 1993 and 1999, with figures in the parenthesis indicating the response rates. The proportion of the number of response institutions actually making a response to the questionnaire has decreased over time.

As shown in Figure 1, there are four possible reasons for the decrease:

1. Decrease in the number of institutions targeted for survey (increased number of institutions not in existence or not in operation)
2. Increase in the number of institutions refusing to participate in the survey (non-cooperative)
3. Increase in the number of institutions not consulting on the day of the survey
4. Increase in the number of institutions with no patients on the day of the survey

Table 2. Number of evaluable institutions and response rates of the surveys.

Year	Hospitals	General medical clinics	Dental clinics
1993	6,840 (0.98)	5,107 (0.85)	868 (0.87)
1996	6,600 (0.94)	5,055 (0.84)	860 (0.86)
1999	6,426 (0.92)	4,892 (0.82)	833 (0.83)

The number of institutions targeted to be sampled was 7,000 hospitals, 6,000 general medical clinics, and 1,000 dental clinics.

This type of missing data leads to bias in the estimated number of patients and decrease in the precision in the estimation. For these reasons, in reviewing the survey methods, it is necessary to determine the reason for missing data. However, for given data in the Patient Survey, only the number of evaluable institutions can be determined, and it is not possible to determine the number of institutions for the above reasons.

Thus, we instead investigated the proportion of institutions not in existence or not in operation relative to all medical institutions in the Institutional Survey data, and found that the proportion of institutions not in existence or not in operation (Reason 1) did not change remarkably in 1993, 1996, or 1999, with changes of about 1%. Thus, the decrease in the number of institutions targeted for survey is not the major cause of the decrease in the number of evaluable institutions.

Similarly, it is not likely that an increase in the number of days on which medical institutions are not consulting is the major reason. In fact, demand for medical services has increased in the recent years, and there has also been an increase in the number of general medical clinics in existence.

Thus, we believe that the major reason for the increase in the number of missing data is an increase in the number of institutions refusing to participate in the survey or institutions with no patients. If refusal to participate in the survey is the major reason, the number of sample institutions decreases, so that as a result, the precision in the estimation of patient numbers also decreases. However, since the patient numbers in the institutions that refuse to participate in the survey cannot be determined from the past Patient Survey data, it is difficult to investigate the influence of such institutions on the bias.

3.2. Influence of missing data on the precision of estimation

As indicated above, the number of evaluable institutions has decreased over time, and one cannot exclude the possibility that the cause of the decrease is the refusal to participate in the survey. As a result, it is suspected that the precision in the estimation of patient numbers is decreasing year by year. We studied the change over time in the precision in the estimation of patient numbers for a classification of 20 groups of diseases in 1996 and 1999, by using the standard error rate [standard error of the estimated number of patients/estimated number of patients ($\times 100$)] and obtained the results shown in Table 3, Figures 2 and 3. We omitted the disease names in the table and figures. (See Table 1 for the

Table 3. The standard error rate (%) in 1996 and 1999.

Disease	Hospitals				General medical clinics			
	Inpatients		Outpatients		Inpatients		Outpatients	
No.	1996	1999	1996	1999	1996	1998	1996	1999
1	2.5	1.7	2.2	4.8	10.2	11.3	3.3	4.9
2	6.7	0.6	3.8	1.2	9.0	8.0	5.4	4.3
3	1.6	3.1	1.7	2.8	20.3	21.1	5.1	6.1
4	2.1	1.5	3.9	1.6	8.3	9.1	2.9	3.1
5	2.8	0.9	3.0	11.6	18.5	23.5	16.1	14.4
6	2.6	1.8	1.7	8.0	10.9	10.2	6.8	4.1
7	1.5	1.9	3.7	1.8	20.4	21.5	8.4	11.5
8	2.2	4.0	1.7	2.7	22.4	30.8	8.3	9.8
9	7.7	0.9	8.4	1.6	5.6	6.1	2.4	2.8
10	2.1	1.1	6.2	2.6	9.0	8.8	3.3	3.7
11	3.4	0.9	7.1	1.6	8.8	8.1	3.7	4.1
12	1.5	3.0	2.4	5.5	37.5	20.8	6.4	8.3
13	2.5	1.5	10.8	1.7	7.7	7.5	4.3	5.9
14	1.9	1.8	4.7	1.3	17.0	15.6	8.2	8.9
15	3.1	1.3	2.1	2.8	11.5	12.3	13.7	12.9
16	1.6	2.3	4.5	5.2	22.6	30.9	21.4	21.4
17	1.4	2.7	2.7	3.7	87.1	50.5	24.5	14.0
18	5.9	1.7	12.0	3.7	13.5	9.5	6.3	5.5
19	3.7	1.0	5.6	2.0	7.6	9.1	3.9	5.5
20	1.7	1.8	4.0	2.0	14.7	16.2	6.7	8.0

disease names.) The figures in the table are those indicated in the Patient Survey reports. For the general medical clinics, note that the vertical axis scale differs depending on whether the inpatient or outpatient data is shown. For dental clinics, there were no notable changes in the three classifications of diseases, and the table and figures were omitted.

These figures indicate that, for hospitals, the number of classifications of diseases with increasing standard error rate from 1996 to 1999 is almost the same as that with decreasing standard error rate. Since there are a large number of institutions targeted to be sampled, so that the standard error rate is generally low, and the effect of decrease in the number of sample institutions is relatively low. Thus, we believe that the variation in the standard error rate is within the range of variation.

In contrast, for the general medical clinics, the standard error rate increased from 1996 to 1999 in almost all classifications of diseases, whereas the degree of increments is small. For this reason, from 1996 to 1999, we believe that the precision in the estimation of patient numbers has decreased. Unless countermeasures are taken to correct this situation, such as increasing the number of institutions targeted to be sampled or decreasing the number of institutions refusing to participate, this trend cannot be improved. The institutions that refuse to participate in the survey are the potential cause of the bias in the estimation

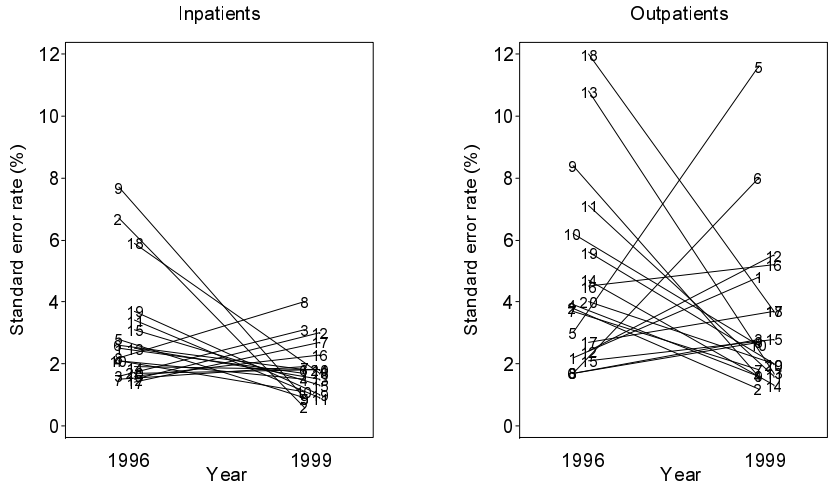

Figure 2. Change of the standard error rate in hospitals.

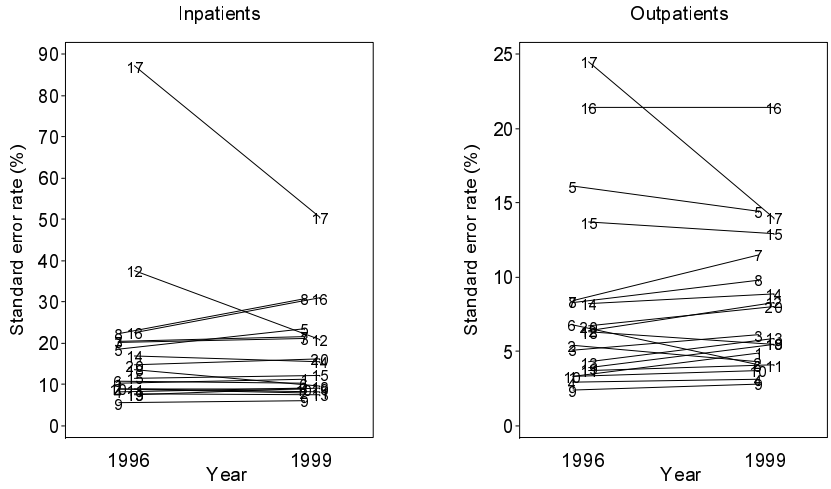

Figure 3. Change of the standard error rate in general medical clinics.

of patient numbers, so that it is necessary to investigate the reasons for missing data in the future Patient Surveys.

4. "Missing" data in static survey of medical institutions

4.1. Frequency of missing data

Institutions with missing patient numbers in the Institutional Survey cannot be used in the estimation. Thus, we studied the 1996 and 1999 survey data

for the number of institutions with missing patient numbers for each medical institution in the Institutional Survey. The results indicated that for hospitals, the institutions with missing patient numbers in the Institutional Survey no longer existed. This is because the "Hospital Report" reporting the monthly patient numbers is required by law.

The data from general medical clinics and dental clinics are as shown in Table 4. According to Table 4, for the general medical clinics, about 3% of the institutions have missing patient numbers in the Institutional Survey, but considering the absolute numbers, this has an effect on the precision in the estimation of patient numbers. This information has to be improved even from the standpoint of more effective utilization of funds devoted to the survey. For the dental clinics, at most 1% of the institutions have missing patient numbers in the Institutional Survey, and the absolute number is quite small at 10 or less, so that this is not a major problem.

Table 4. Number of institutions with missing patient numbers.

	General medical clinics			Dental clinics	
Year	Inpatients	Outpatients	Evaluable institutions	Outpatiens	Evaluable institutions
1996	57*1 (1.1%)	163*2 (3.2%)	5,055	10 (1.2%)	860
1999	167 (3.4%)	135 (2.8%)	4,892	9 (1.1%)	833

*1 Number of institutions not found in the Institutional Survey.

*2 Number of institutions including 57*1.

4.2. Influence of missing data on the precision of estimation

In this section, we evaluate how the missing patient numbers in the Institutional Survey affect the estimation of patient numbers. As described in the prior section, the problem is in the general medical clinics. The general medical clinics have a large number of outpatients, and thus we will focus our attention on this issue.

We investigated the standard error rate for a classification of 20 groups of diseases, under the assumption that there are no institutions with missing patient numbers in the Institutional Survey. We found that the standard error rate in 1996 and 1999 decreased in the worst case by only 0.6 points, although the detailed results are omitted here, which implies that the influence of missing data is not serious in the countrywide estimation. At the prefectural level, however, the number of population institutions is small, and the standard error rate decreased as much as 5 points in the worst case.

We examined the characteristics of institutions with missing patient numbers using the 1999 survey data. According to the data, medical departments, prefectures, number of doctors and nurses were independent of the missing data, whereas the employment of clerical staffs highly influenced the missing data. In fact, only 17.8% of institutions with missing data have clerical staffs, whereas 85.5% of institutions without missing data have them. This fact can be in-

terpreted as the institutions not employing clerical staffs were likely to neglect responding to questions. Since the total number of patients in such institutions seems to be relatively small, the number of patients may be overestimated.

At present, the reason for the patient numbers being missing in the Institutional Survey is not clear, but if the cause is inadequacy on the part of the institution being surveyed, it would be appropriate to request accurate recording of patient numbers as a matter of good work practice.

5. Proposal of incorporating variable weighting into the ratio estimation

5.1. Validity of the premise for the current estimation method

In the prior sections, we discussed the trend towards an increase in the missing data, and how this can lead to a decrease in the precision in the estimation of patient numbers. The fundamental measures needed to address this issue, as described above, are to determine the reason why the data is missing and to decrease the number of missing data. However, it is also necessary to reconsider the estimation method to minimize the effects when the missing data increases.

The estimation methods for patient numbers differ somewhat between hospitals and medical clinics. For hospitals, the method takes into account the institutional sampling rate stratified according to prefecture or secondary medical area, but for general medical clinics, such stratification is not used. Thus, for general medical clinics, for a given prefecture i, the estimated number of patients $\hat{Z}_i$ for each disease is calculated by (5.1), and the sum of all prefectures is used to determine the estimated number of patients in Japan as $\hat{Z} = \sum_{i=1}^{47} \hat{Z}_i$.

$$\hat{Z}_i = \frac{\sum_{j=1}^{L} X_{ij}}{\sum_{j=1}^{L} Y'_{ij}} \times Y_i. \tag{5.1}$$

Here, the variables in (5.1) are defined as follows.

- X_{ij} : Number of patients for a given disease in prefecture i, stratum j in the "Patient Survey"
- Y'_{ij} : Total number of patients in sample institutions in prefecture i, stratum j in the "Institutional Survey"
- Y_i : Total number of patients in prefecture i in the "Institutional Survey"
- L : Number of strata in a prefecture

For general medical clinics, although stratified by "principal medical department" and "status of beds," these are not taken into account in the estimation, because of the assumption that there is proportional allocation across all strata and, therefore, the institutions in the stratified random sampling are regarded as simple random samples. For dental clinics, the stratification is by prefecture only, so the problem discussed here does not occur.

However, in reality, based on the effects of the missing data as discussed, the institutional sampling rate by strata varies by prefecture. Thus, in the current estimation method, the premise of the constant institutional sampling rate by strata is not satisfied.

We used the data from the 1999 survey to study the 9 institutional sampling rates by strata by "principal medical department" by prefecture, as shown in Figure 4. In this Figure, the horizontal axis is the prefecture, and the vertical axis is the institutional sampling rate by strata. Since the number of population institutions is small, there was one stratum where the institutional sampling rate was 100%, but this has been excluded from the figure.

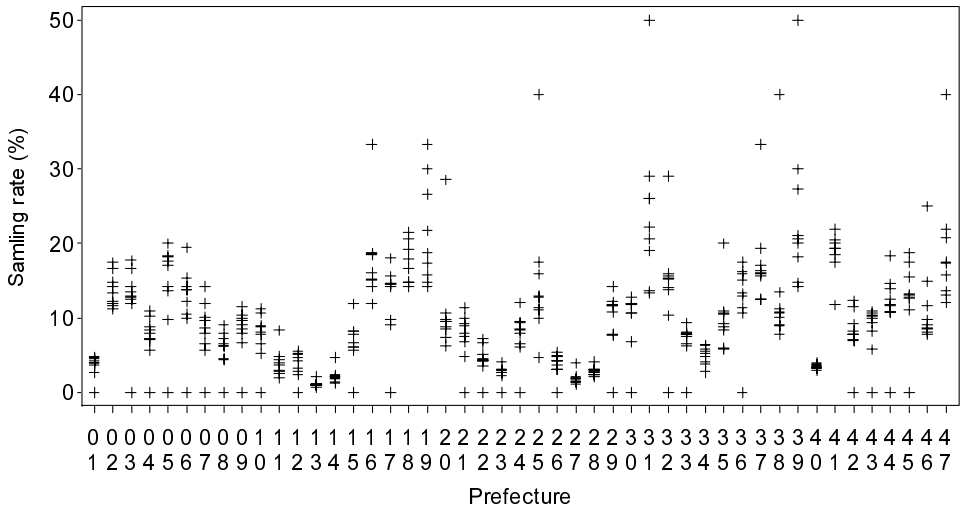

Figure 4. Institutional sampling rates by strata and prefecture.

According to this figure, there are few prefectures where the institutional sampling rate by strata remains constant. Although the institutional sampling rate by strata will vary somewhat due to the small number of population institutions, the institutional sampling rate differs greatly in many prefectures.

Furthermore, among the general medical clinics, the number of evaluable institutions has decreased year by year, so it is possible that the variation in the institutional sampling rate by strata has changed. Thus, we calculated the standard deviation of the institutional sampling rate by strata for each prefecture, and studied how it changed over time. As indicated in Figure 5, overall the standard deviation shows a trend towards an increase.

As discussed above, because of the effect of the missing data, the actual institutional sampling rate varies by prefecture, and this variation increases every year. Thus, the current estimation method for patient numbers in general medical clinics is progressively becoming more inappropriate. It is necessary to improve the estimation method.

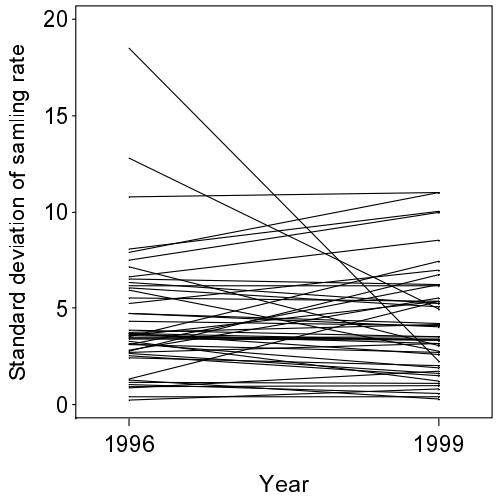

Figure 5. Change of the standard deviation.

5.2. Estimation method considering the difference of sampling

In order to address the problem that the actual institutional sampling rate differs by strata, we propose the use of (5.2), which takes into account the differences in institutional sampling rate when estimating the patient numbers.

The Formula for this proposed method is described by Levy and Lemeshow (1999), where the ratio of X_{ij} to Y'_{ij} is calculated by weighting with the reciprocal of institutional sampling rate, N_{ij}/n_{ij}. Here, N_{ij} is the number of population institutions in prefecture i, stratum j, while n_{ij} is the number of sample institutions in prefecture i, stratum j.

$$\hat{Z}_i = \left(\frac{\sum_{j=1}^{L} (N_{ij}/n_{ij}) X_{ij}}{\sum_{j=1}^{L} (N_{ij}/n_{ij}) Y'_{ij}} \right) \times Y_i. \tag{5.2}$$

If the institutional sampling rate by strata is constant, $N_{ij}/n_{ij} = c$, substituting this into (5.2) gives (5.1). The variance of the estimated number of patients by (5.2) can be estimated by an approximation using (5.3).

$$V(\hat{Z}_i) = \sum_{j=1}^{L} \frac{N_{ij}^2 (N_{ij} - n_{ij})}{n_{ij}(N_{ij} - 1)} \times \left\{ V(X_{ij}) - 2\,\mathrm{Cov}(X_{ij}, Y'_{ij}) \frac{\bar{X}_{ij}}{\bar{Y}'_{ij}} + V(Y'_{ij}) \left(\frac{\bar{X}_{ij}}{\bar{Y}'_{ij}} \right)^2 \right\}. \tag{5.3}$$

Here, the variables in (5.3) are defined as follows.

$$V(X_{ij}) = \frac{1}{n_{ij}-1}\sum_{s=1}^{n_{ij}}(X_{ij(s)} - \bar{X}_{ij})^2, \qquad V(Y'_{ij}) = \frac{1}{n_{ij}-1}\sum_{s=1}^{n_{ij}}(Y'_{ij(s)} - \bar{Y}'_{ij})^2$$

$$\mathrm{Cov}(X_{ij}, Y'_{ij}) = \frac{1}{n_{ij}-1}\sum_{s=1}^{n_{ij}}(X_{ij(s)} - \bar{X}_{ij})(Y'_{ij(s)} - \bar{Y}'_{ij})$$

$$\bar{X}_{ij} = \frac{1}{n_{ij}}\sum_{s=1}^{n_{ij}} X_{ij(s)}, \qquad \bar{Y}'_{ij} = \frac{1}{n_{ij}}\sum_{s=1}^{n_{ij}} Y'_{ij(s)}$$

$X_{ij(s)}$: Number of patients for a given disease in the "Patient Survey" in prefecture i, stratum j, institution s

$Y'_{ij(s)}$: Total number of patients in sample institutions in the "Institutional Survey" in prefecture i, stratum j, institution s

5.3. Quantitative assessment of the effect of the proposed method

To determine the effect of the proposed method on the precision in the estimation of patient numbers, we applied the proposed method to the 1996 and 1999 survey data. First, to compare the estimated number of patients in a classification of 20 groups of diseases, we determined the ratio of the two figures calculated by the 2 methods. The results showed that the ratios ranged from a minimum of 0.9 to a maximum of 1.1, and that neither estimation method resulted in extremes in the estimates. The superiority of one method over another in bias reduction cannot be confirmed without another study such as those conducted in Hashimoto *et al.* (1992).

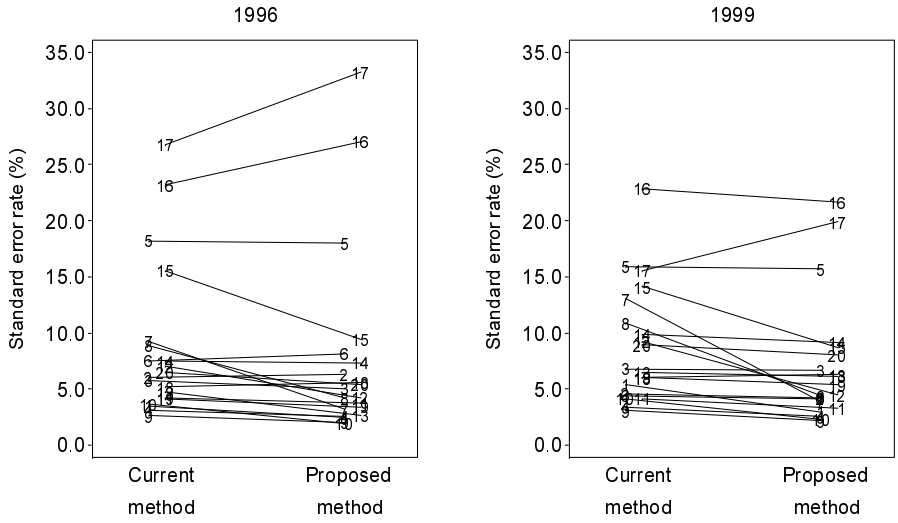

Figure 6. Results of the standard error rate (%) by the current method and the proposed method.

Next, to compare the precision in the estimation of patient numbers, we showed the standard error rate for a classification of 20 groups of diseases in Table 5 and prepared the scatter plots as shown in Figure 6. The "Current Method" in Table 5 applies to the outpatient results of Table 3. Here, since the estimated number of patients was based on the evaluable institutions, the standard error rates are somewhat different.

When the proposed method is used, the standard error rate decreases by as much as 9.1 points ("7. Diseases of the eye and adnexa"). This is the effect of evaluation of the variance by strata. That is, the proposed method increases the precision in the estimation of patient numbers. However, for diseases where the number of institutions within stratum is small and the potential number of patients is small, the precision in the estimation of patient numbers may decrease. This is an effect of institutions with an extreme number of patients.

6. Conclusion

We first demonstrated that in the Patient Survey there has been a trend towards an increase in the amount of missing data in recent years. We then evaluated quantitatively the effect of the missing data on the precision in the estimation of patient numbers. Next, we proposed the following several points to improve the precision in the estimation.

Table 5. Results of the standard error rate (%) by the current method and the proposed method.

	1996		1999	
Disease No.	Current method	Proposed method	Current method	Proposed method
1	3.5	2.5	5.4	3.0
2	6.1	6.3	4.6	4.2
3	5.8	5.0	6.8	6.7
4	3.1	2.6	3.4	2.6
5	18.2	18.0	15.9	15.7
6	7.5	8.1	4.4	4.1
7	9.3	3.2	13.0	3.9
8	8.9	4.2	10.9	4.4
9	2.6	2.0	3.1	2.1
10	3.6	1.9	4.1	2.3
11	4.2	3.7	4.1	3.2
12	7.0	4.2	9.2	4.4
13	4.7	2.7	6.5	6.1
14	7.5	7.3	9.9	9.1
15	15.6	9.4	14.1	8.7
16	23.2	27.1	22.9	21.7
17	26.8	33.3	15.5	19.9
18	5.2	5.6	6.0	6.3
19	4.1	3.4	6.0	5.4
20	6.5	5.4	8.9	8.0

- Specify the number of institutions targeted to be sampled, taking into consideration the changes in the number of population institutions and decreases in the number of evaluable institutions.
- Make efforts to reduce the number of institutions that refuse to participate in the survey (non-cooperative). Investigate the reasons for refusal to participate in the survey.
- Take countermeasures to avoid missing patient numbers in the Institutional Survey.

The increase in the missing data tends to show the disadvantage of the current estimation method for patient numbers in general medical clinics. Thus, we propose the use of an estimation method that takes into account the differences in the institutional sampling rate. The proposed method reduced the standard error rate of the estimated number of patients by a maximum of 9.1 points, when applied to data from past surveys. We conclude that it is better to use this proposed method in future Patient 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 patients 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

Hashimoto, S. *et al.* (1992). The estimation method for patient numbers in secondary medical area in the Patient Survey of the Ministry of Health and Welfare, *Journal of Health and Welfare Statistics*, **39**(8), 28–36 (in Japanese).

Levy, P. S. and Lemeshow, S. (1999). *Sampling of Populations*: *Method and Applications*, 3rd ed., Wiley, New York.

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).

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).

Yoshimura, I., Sozu, T. and Hamada, C. (2004). On the improvement of estimation and stratification methods in Patient Survey, *Proceeding of 2003 Workshop on Applied Use of Statistics and Information of the Ministry of Health, Labor and Welfare of Japan*, 14–19 (in Japanese).