academicJournals

Vol. 5(6), pp. 243-256, June 2013
DOI: 10.5897/JPHE11.131
ISSN 2006-9723 ©2013 Academic Journals
http://www.academicjournals.org/JPHE

Journal of Public Health and Epidemiology

Full Length Research Paper

Formula for estimating incidence of chronic diseases from prevalence, mortality, and other indices from survey

Koujirou Inoue

Bihoku Public Health Center, Zip code: 716-8585, Tikanori Ochiaicho, Takahashi City, Okayama Prefecture, Japan.

Accepted 7 February, 2012

A theoretical formula was devised to calculate the age-specific estimated incidence (ASEI) rate of chronic diseases using consecutive-year data on disease prevalence and death rate. Data on the rate of persons who had suffered from one disease but died from another disease (RPDA) were retrieved from the database of a public health center. The ASEI rate of diabetes mellitus (DM) and hypertension among older age groups was found to be negative. After correcting for RPDA, the ASEIs of DM became positive or nearly 0 for this age group. It is surmised that the negative ASEIs of hypertension could be improved by considering the multiple diseases of a patient in the patient survey and finding through practicable research in the same patient survey, the rate of persons who suffered from a disease that developed into another disease (RPCA). Additional research efforts that include other factors such as the estimated cure rate (ECR) and the estimated potential incidence ratio (EPIR) were discussed. The formula for the ASEI was consequently arrived at by taking all of these elements of the RPDA, RPCA, ECR, and EPIR into consideration.

Key words: Incidence, prevalence, cure rate, rate of persons who had suffered from one disease but died from another disease (RPDA), rate of persons who suffered from a disease that developed into another disease (RPCA), survey.

INTRODUCTION

The incidence rate is usually calculated by the following formula and is not determined as a minus value (Armitage et al., 2002):

$$I = \frac{number\ \text{of new cases in any given period}}{population\ \text{at risk}}.$$

The incidence rate is calculated from a follow-up research study, but in any cohort, there will be a change in the population because of migration (Preston et al., 2007). Thus, the research will be expensive and time-consuming.

The outcome of chronic diseases between year *a* and the consecutive year *b* is shown in Figure 1. Square A represents the number of patients in year *a*, and square B the number in year *b*. If a cure is not considered, the incidence of a chronic disease can be calculated using the data of patients in consecutive years *a* and *b* and the data of deaths between years *a* and *b* using the following formula:

$I = B - A + D$

According to the method for estimating incidence from

E-mail: koujirou_inoue@pref.okayama.lg.jp. Tel: 81-866-21-2836. Fax: 81-866-22-8098.

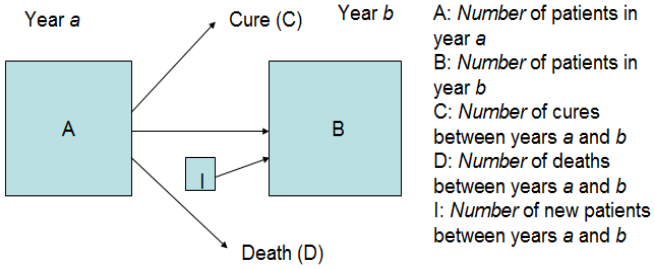

Figure 1. Outcome on chronic diseases between year *a* and consecutive year *b*.

prevalence and mortality, Leske et al. (1981), Dewey (1992), and Hill et al. (1999) reported that glaucoma, dementia, and diabetes are irreversible diseases. They were not considered for the cure rate or other necessary indices. So, the methods by them could not be used for common chronic diseases. The author studied the possibility of calculating the incidence rate using statistical data through the following steps and arrived at Formula 1.

1. A theoretical formula was derived for the incidence rate in different age groups by ignoring the cure rate.
2. The rate of change in the population of Japan by migration was researched, and its influence on the formula is discussed.
3. The prevalence in the patient survey was compared with that in other researches, and whether the data were appropriate for the calculation of the ASEI was discussed.
4. Using the data obtained from patient surveys and death rates, the theoretical formula for the estimated incidence rate was tested and discussed for the necessary data. The formula for estimating the incidence rate was arrived at by taking into consideration the necessary data.

MATERIALS AND METHODS

A theoretical formula for calculating the age-specific estimated incidence rate by disregarding the cure rate and using statistical data for consecutive years. The following theoretical formula was derived by disregarding the cure rate.

It is assumed that the consecutive years are a and b.
The following are the values used for calculation in this study:
range of all age groups as w is 5 and, accordingly, $w = b - a = 5$.
Thus, the midyear between years a and b is calculated as $\frac{a+b}{2}$.
The ith age group is the group aged between $w \times i$ years and $w \times (i+1) - 1$ years $(i = 0,1,2, \ldots 16)$.
The cohort of the population in the ith age group in year a is ${}_aG_i$, the population is ${}_aN_i$, the prevalence is ${}_aP_i$, the death rate of a particular disease for ${}_aG_i$ between years a and b is ${}_ad_i$, and the rate of all deaths for ${}_aG_i$ is ${}_aD_i$.

Similarly, for year b,
the number of patients in year a for ${}_aG_i$ is ${}_aP_i \times {}_aN_i$, that in year b for ${}_aG_i$ is ${}_bP_{i+1} \times {}_bN_{i+1}$, and the number of deaths for ${}_aG_i$ between years a and b is ${}_ad_i \times {}_aN_i$.
Thus, the increase in the number of patients for ${}_aG_i$ between years a and b is ${}_bP_{i+1} \times {}_bN_{i+1} - {}_aP_i \times {}_aN_i + {}_ad_i \times {}_aN_i$.
Therefore, the estimated incidence rate for ${}_aG_i$ is $\frac{{}_bP_{i+1} \times {}_bN_{i+1} - {}_aP_i \times {}_aN_i + {}_ad_i \times {}_aN_i}{{}_aN_i}$

$= \frac{{}_bP_{i+1} \times {}_bN_{i+1}}{{}_aN_i} - {}_aP_i + {}_ad_i$.

Thus, the estimated incidence rate per year at the midyear between years *a* and *b* is

$$ {}_{\frac{a+b}{2}}R_j = \frac{\frac{{}_bP_{i+1} \times {}_bN_{i+1}}{{}_aN_i} - {}_aP_i + {}_ad_i}{b-a} \cdots \text{Formula 1.} $$

The denominator of this formula includes ${}_aN_i$ and not ${}_{\frac{a+b}{2}}N_i$, because the concept of the cohort of ${}_aG_i$ is used. The value calculated using formula 1 for the 0 to 4-year age group in 2000 and the 5 to 9-year age group in 2005 reflects the estimated incidence rate for the 2.5 to 7.5-year age group at the midyear of the period 2000 to 2005. The midpoints of the age groups for the estimated incidence rates are 5, 10, 15, 20, etc., up to the age of 80 years.

Thus, the jth age group is the group aged between $w \times j + w/2$ years and $w \times (j+1) + w/2$ years ($j = 0,1,2, \ldots 15$).

Formula 1 can then be rewritten as follows:

$$\text{When } {}_bN_{i+1} = (1-{}_aD_i)\times {}_aN_i,\ {}_{\frac{a+b}{2}}R_j = \frac{{}_bP_{i+1}\times(1-{}_aD_i)-{}_aP_i+{}_ad_i}{b-a} \cdots \text{Formula 2.}$$

The prevalence used in formula 1 or 2 was estimated using data from the patient survey with stratified random sampling of all residents of Japan, including foreigners (Jerrold, 1999; Rothman and Greenland, 1998; Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government, 2009). Thus, the value calculated using formula 1 or 2 is referred to as the age-specific estimated incidence rate (ASEI). Formula 1 or 2 was used for the cohort; however, if the composition of the population in the group expressed as ${}_aG_i$ changed owing to migration, this formula may not be applicable. Therefore, the rate of change in the population owing to the migration of ${}_aG_i$ between 2000 and 2005 was investigated.

From the calculation, using both the data on foreigners residing in Japan and those on Japanese nationals who went overseas over a period of 3 months in each age group between 2000 and 2005, the maximum rate was estimated to be 4.6% of the population for the age group 20 to 24 years. The rate of change in the population is considerable in some instances; however, this possibility was not considered for the cohort of this study in formula 1 or 2. However, the rate of change was considered the same for all the areas of random sampling in the patient survey; thus, no correction was needed (Jerrold, 1999; Rothman and Greenland, 1998).

Estimated prevalence from patient surveys

The Ministry of Health, Labour, and Welfare of Japan conducted patient surveys in Japan in October 1999, 2002, and 2005. The research was done according to the stratification of medical institutions, such as the 11 types of hospitals classified by the property or the number of beds, clinics, and dental clinics in the secondary medical service areas (secondary emergency medical areas) of the prefectures. The primary emergency medical areas are clinics, the secondary emergency medical areas are hospitals, and the third emergency medical areas are hospitals with a high level of service (Ministry of Health, Labour and Welfare, 2006).

Static research was conducted on medical institutions according to the number of patients and the property or number of beds, and the resulting data were used to estimate the number of patients in the patient survey. In the static research on medical institutions, the total number of outpatients in September and the total number of inpatients on the 30th day in September were researched and used for the formula in the patient survey (Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government, 2009). For patients in hospitals or clinics, data were recorded for one day in the patient survey. Practical formulae are shown in Table 1. The standard error of the number of patients is calculated using the approximate expression for the ratio of the two variances (Armitage et al., 2002; Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government. 2000). These formulae indicate that

$$\text{No. of patients} = \frac{\text{No. of patients in one disease (one attribution) in research sample} \times \text{Total no. of patients in all institutions in the static research}}{\text{Total no. of patients in sample institutions in the patient survey in static research}}$$

(Formula 3)

For patients with more than one disease, a single disease was selected for consideration. The patients were classified by the Ministry of Health, Labour, and Welfare of Japan according to the International Classification of Diseases (ICD); specifically, the same ICD classification was used for issuing death certificates (World Health Organization, 2004). The consultation rate for medical care was calculated as:

$$\text{Consultation rate for medical care} = \frac{\text{No. of inpatients + No. of outpatients}}{\text{Population}}$$

The estimated number of patients, including those who did not receive medical care on the research day, was calculated using the following formula developed by several investigators (Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government, 2009; Sota, 1960; Hashimoto et al., 1994; Nakamura et al., 1994)

Estimated number of patients = Inpatients + Outpatients (first visit) + Outpatients (following visit) × Average interval since last visit × adjustment factor (6/7) (Formula 4).

Hashimoto et al. (1994) suggested that patients revisiting the clinic after an interval of over 31 days should be considered first-time

Table 1. Formula for the calculation of the number of patients in hospitals and clinics in patient survey of Japan.

Variable	Formula	Comment
Hospital	$Z_{gkh} = \sum_{j=1}^{7}\left[\frac{X_{gjk}}{X'_{gjk}} \times \frac{W_{gjkh}}{Y'_{gj}} \times Y_{gj}\right] + \frac{\sum_{j=8}^{11}\frac{N_{gj}}{n_{gj}}X_{gjk}}{\sum_{j=8}^{11}\frac{N_{gj}}{n_{gj}}X'_{gjk}} \times \frac{\sum_{j=8}^{11}\frac{N_{gj}}{n_{gj}}W_{gjkh}}{\sum_{j=8}^{11}\frac{N_{gj}}{n_{gj}}Y'_{gj}} \times \sum_{j=8}^{11}Y_{gj}$	W_{gjkh}: the index to be inputted by the research of patient survey.
	Z_{gkh}: Estimated number of patients by some disease (attribution) (h) by sex (k) and by secondary emergency medical areas (g). N_{gj}: Number of medical institutions by stratified class (j) in secondary emergency medical areas (g) in the static research on medical institutions. n_{gj}: Number of medical institutions by stratified class (j) in secondary emergency medical areas (g) in the patient survey. X_{gjk}: Number of patients by sex (k) in stratified class (j) in secondary emergency medical areas (g) in the patient survey. X'_{gjk}: Number of patients by odd-numbered birthday by sex (k) in stratified class (j) in secondary emergency medical area (g) in the patient survey. W_{gjkh}: Number of patients by odd-numbered birthday by some attribution (h) by sex (k) by stratified class (j) in secondary emergency medical areas (g) in the patient survey. Y'_{gj}: Number of patients in the sample institutions (in the patient survey) by stratified class (j) in secondary emergency medical areas (g) in the static research (beds and department) on medical institutions. Y_{gj}: Number of patients by stratified class (j) in secondary emergency medical areas (g) in the static research on medical institutions.	The attribution refers to the ICD classification of the disease. 1: Psychiatry hospitals 2: Infectious disease medical hospitals 3: Tuberculosis medical hospitals 4: High-level medical hospitals 5: Senile dementia medical hospitals 6: Hospitals providing bedridden care for senile persons 7: Hospitals providing beds for senile diseases 8: (except upper 1-7 cases) Number of beds is under 99. 9: (except upper 1-7 cases) Number of beds is between 100 and 299. 10: (except upper 1-7 cases) Number of beds is between 300 and 499. 11: (except upper 1-7 cases) Number of beds is more than 500.
Clinic	$Z_i = \frac{\sum_{j=1}^{L}X_{ij}}{\sum_{j=1}^{L}Y'_{ij}} \times Y_i = \frac{\sum_{s=1}^{Ni}X_{i(s)}}{\sum_{s=1}^{ni}Y'_{i(s)}} \times Y_i$	X_{ij} or $X_{i(s)}$: the index to be inputted by the research of patient survey.
	Z_i: Estimated number of patients by some attribution in a prefecture area (i). L: Number of stratified classes in the prefecture. X_{ij}: Number of patients by some attribution in a stratified class (j) in a prefecture area (i) in the patient survey. Y'_{ij}: Number of patients in the sample institutions (in the patient survey) by stratified class (j) in a prefecture area (i) in the static research on medical institutions. Y_i: Number of patients by prefecture area (i) in the static research on medical institutions. $X_{i(s)}$: Number of patients by some attribution in some institutions (s) in a prefecture area (i) in the patient survey. $Y'_{i(s)}$: Number of patients in sample institutions (s) (in the patient survey) in prefecture area (i) in the static research on the medical institutions. ni: Number of sample institutions (in the patient survey) in a prefecture area (i).	

visitors and the survey have been conducted according to their suggestion (Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government. 2000; Hashimoto et al., 1994). The average interval since last visit was calculated for every prefecture, age group, sexual group, and ICD classification in the surveys. The inpatients and outpatients were calculated separately and then added. The total number of patients was calculated by summarizing the number of patients in every prefecture. The estimated prevalence (hereafter referred to as "prevalence") for different age groups was calculated by dividing the estimated number of patients by the total population of Japan as the point prevalence (Rothman and Greenland, 1998). The prevalence in 2005 was used from the data of 2005. The prevalence in 2000 was interpolated from the data of 1999 and 2002. The prevalence in the 5 to 9-year age group in 2000 positively correlated with that of the 4 to 8-year age group in 1999 and the 7 to 11-year age group in 2002. In the different age groups in 2000, the prevalence was calculated as follows:

1. 4 to 8-year age group in 1999: (1/5 × prevalence in the 0 to 4-year age group) + (4/5 × prevalence in the 5 to 9-year age group).
2. 7 to 11-year age group in 2002: (3/5 × the 5 to 9-year age group) + (2/5 × the 10 to 14-year age group).
3. 5 to 9-year age group in 2000: (2/3 × the 4 to 8-year age group in 1999) + (1/3 × the 7 to 11-year age group in 2002).
4. > 9-year age group in 2000: similarly calculated.
5. 0 to 4-year age group in 2000: (2/3 × the 0 to 3-year age group in 1999) + (1/3 × the 2 to 6-year age group in 2002).

The standard errors for the age groups in 2000 were calculated using the same method.

Calculation of death rate from statistical data

The death rate between 2000 and 2005 is calculated as follows:

$_{a}d_{i}$ is the death rate between 2000 and 2005, which corresponds to the death rate between years a and b in formula 1 or 2,

$_{2000}d_{i} + {}_{2001}d_{i} \times 4/5 + {}_{2001}d_{i+1} \times 1/5 + {}_{2002}d_{i} \times 3/5 + {}_{2002}d_{i+1} \times 2/5 + {}_{2003}d_{i} \times 2/5$

$+ {}_{2003}d_{i} \times 3/5 + {}_{2004}d_{i} \times 1/5 + {}_{2004}d_{i+1} \times 4/5$, because each age group becomes 1 year older after a year.

The $_{a}D_{i}$ between 2000 and 2005 is calculated in the same manner.

By the above formula, $_{2005}N_{i+1}$ corresponds considerably to $(1-{}_{a}D_{i}) \times {}_{2000}N_{i}$.

Test 1 for the ASEI

The categories of the International Classification of Diseases (10th Revision; ICD-10) used this to test the validity of the formula included for diabetes mellitus (DM, E10-14), hypertensive diseases (I10-15), and cerebrovascular diseases (I60-69).

From formula 2, when it is assumed that $(1-{}_{a}D_{i})$ and $_{a}d_{i}$ are constant values, $_{b}P_{i+1}$ and $_{a}P_{i}$ are two independent random variables.

Thus, the standard error (SE) of $_{\frac{a+b}{2}}R_{j}$ is

$$SE({}_{\frac{a+b}{2}}R_{j}) = \frac{1}{b-a}\sqrt{\left\{SE({}_{b}P_{i+1}) \times (1-{}_{a}D_{i})\right\}^{2} + \left\{SE({}_{a}P_{i})\right\}^{2}} \cdots \text{Formula 5.}$$

The ASEI (SE) were calculated from formulae 2 and 5 using the prevalence in 2000 and 2005, and death rate between 2000 and 2005.

Formula 2 for the ASEI and test 2

If one follows the hypothesis that the death rate of persons suffering from a certain disease, say *S*, does not include the rate of persons who had been suffering from disease *S* but may have died from another disease, then the ASEI of disease *S* calculated using formula 1 or 2 could be incorrect because the death rates used were incorrectly. Hence, the number of such persons was obtained from the database of death certificates of a public health center which covered the period 1992 to 1998. However, this area differs from that in the sample in the patient surveys. The value is referred to as the rate of persons who died from another disease (RPDA). The number of all deaths between 1992 and 1998 is 4126, with a mean number per year of 589. The figure included cases of DM and other chronic diseases when disease *S* was not recorded as the cause of death but was recorded as the other continuous disease until the time of death. The rates and standard errors were calculated as follows (Miller, 1983b):

$$RPDA = \frac{\text{number of persons who died from another disease}}{\text{number of all deaths}}.$$

$$(SE)\ RPDA = \sqrt{\frac{RPDA(1-RPDA)}{\text{number of all deaths}}}.$$

The resultant RPDA values and SE for different age groups are shown in Table 2. The death rate corrected by considering the RPDA was calculated as follows: death rate 2 = death rate + rate of all deaths × RPDA and death rate 3 = death rate + rate of all deaths × {RPDA + 1.96 × SE (RPDA)}. ASEI2 was calculated from death rate 2 and ASEI3 from death rate 3. The hypothesis that ASEI = ASEI2 was tested as follows (Miller, 1983c):

$$z0 = \frac{(ASEI2 - ASEI)}{SE(ASEI)}.$$

As this sample (a public health center) did not correspond to the sample of patient surveys in Japan, this test was considered a pilot test for the above hypothesis.

Validity of prevalence in patient survey

To test the validity of the prevalence in the patient survey, data from the National Health and Nutrition Survey in 2006 was used for comparison. In this research, the data included the rate of persons within age groups who failed drug therapy for hypertension or DM (Office for Life-Style Related Diseases Control in Ministry of Health, Labour and Welfare of Japan, 2007). The 2006 survey was conducted with a stratified random sample of 5,000 family units from Japan; the sampling rate was 0.005% of the population, and the significance of the proportion was tested (Miller, 1983d).

Table 2. Rates and standard errors of the rate (per 1000 population) of person aged over 50 years who have died due to another disease (RPDA) according to chronic diseases among the ICD-10 categories obtained from the database of a public health center in Japan between the years 1992 -1998.

Classification of disease	Rate (S.E)									
	Total	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-above
E10-14 Diabetes mellitus	12.6 (1.7)	0.0 (0.0)	8.1(8.0)	14.0 (8.0)	17.6 (7.1)	15.2 (6.2)	26.4 (7.0)	10.2 (3.6)	8.1 (3.3)	5.8 (2.9)
F00-09 Organic including symptomatic mental disorders	3.6 (0.9)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	2.9 (2.9)	0.0 (0.0)	3.8 (2.7)	6.3 (2.8)	5.4 (2.7)	4.4 (2.5)
G20 Pakinson's disease	2.2 (0.7)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	2.9 (2.9)	7.6 (4.4)	5.6 (3.3)	2.5 (1.8)	0.0 (0.0)	0.0 (0.0)
I00-99 Diseases of the circulatory system	38.8 (3.0)	12.2 (12.1)	24.2 (13.8)	9.3 (6.6)	32.4 (9.6)	50.6 (11.0)	41.4 (8.6)	31.7 (6.2)	63.3 (8.9)	42.2 (7.7)
I10-15 Hypertensive diseases	10.2 (1.6)	0.0 (0.0)	8.1 (8.0)	9.3 (6.6)	2.9 (2.9)	12.7 (4.0)	7.5 (3.8)	12.7 (4.0)	14.8 (4.4)	8.7 (3.5)
I20-25 Ischemic heart disease	5.1 (1.1)	0.0 (0.0)	0.0 (0.0)	4.7 (4.7)	5.9 (4.1)	2.5 (2.5)	1.9 (1.9)	5.1 (2.5)	8.1 (3.3)	8.7 (3.5)
I45-49 Cardiac dysrhythmias	2.7 (0.8)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	2.9 (2.9)	2.5 (2.5)	3.8 (2.7)	3.8 (2.2)	4.0 (2.3)	1.5 (1.5)
I60-69 Cerebrovascular diseases	41.7 (3.1)	12.2 (12.1)	40.3 (17.7)	9.3 (6.6)	26.5 (8.7)	55.7 (11.5)	52.7 (9.7)	36.8 (6.7)	59.3 (8.7)	46.6 (8.0)
I70 Atherosclerosis	5.6 (1.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	3.8 (2.7)	10.2 (3.6)	8.1 (3.3)	10.2 (3.8)
J40-47 Chronic lower respiratory disease	8.0 (1.4)	0.0 (0.0)	8.1 (8.0)	4.7 (4.7)	5.9 (4.1)	5.1 (3.6)	16.9 (5.6)	8.9 (3.3)	8.1 (3.3)	7.3 (3.2)
J43 Emphysema	1.5 (0.6)	0.0 (0.0)	0.0 (0.0)	4.7 (4.7)	2.9 (2.9)	5.1 (3.6)	0.0 (0.0)	1.3 (1.3)	1.3 (1.3)	0.0 (0.0)
J45 Asthma	3.9 (1.0)	0.0 (0.0)	8.1 (8.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	9.4 (4.2)	5.1 (2.5)	4.0 (2.3)	4.4 (2.5)

RESULTS

The prevalence in 2000 and 2005 for different age groups, the ASEI (SE), ASE2, and ASEI3 for DM, hypertensive diseases, and cerebrovascular diseases are shown in Figures 2, 3, and 4, respectively. The prevalence of DM peaked for the 70 to 74-year age group in both 2000 and 2005. The ASEI peaked at age 60 but was negative for those > 70 years; however, ASEI2 and ASEI3 for those > 70 years were positive or closer to 0 than the ASEI. ASEI2 differed from the ASEI in the 80-year age group, with a significance level of $p < 0.01$.

On the prevalence of hypertensive diseases, peaks were noted in the 75 to 79- and 80 to 84-year age groups in 2000 and 2005. The ASEI peaked at 65 years but was negative at > 80 years. There was no significant difference between ASEI2 and ASEI3. No peaks were observed for the prevalence and the ASEI of cerebrovascular diseases in both 2000 and 2005. ASEI2 was significantly higher than the ASEI with increasing age. ASEI2 differed from the ASEI in the 60-year age group at $p < 0.01$ and ASEI2 for >70 years differed from the ASEI at $p < 0.001$. Many patients with cerebrovascular diseases in old age groups died by another cause, for example, aspiration pneumonia; thus, the result that the ASEI and ASEI2 were different in old age groups is reasonable (Bruce and Steven, 1996).

The comparison of prevalence of diabetes mellitus and hypertension between the National Health and Nutrition Survey in 2006 and the patient survey in 2005 is shown in Table 3 and Figure 5. Regarding DM, there was no difference between the two researches except for persons within 65 to 69-years old, over 70-years old, and the total. The prevalence for hypertension in the 2006 nutrition survey was significantly higher with increasing age than the prevalence for persons over 50-years old in the 2005 patient survey.

DISCUSSION

Importance of RPDA

The ASEI of DM for >70 years is negative, and ASEI2 and ASEI3 for >70 years are positive or closer to 0 than the ASEI. These results indicate that RPDA is important in calculating the ASEI for certain diseases, such as DM, in the older age groups. Data on RPDA of disease *S* in year *b* are acquired theoretically from retrospective research

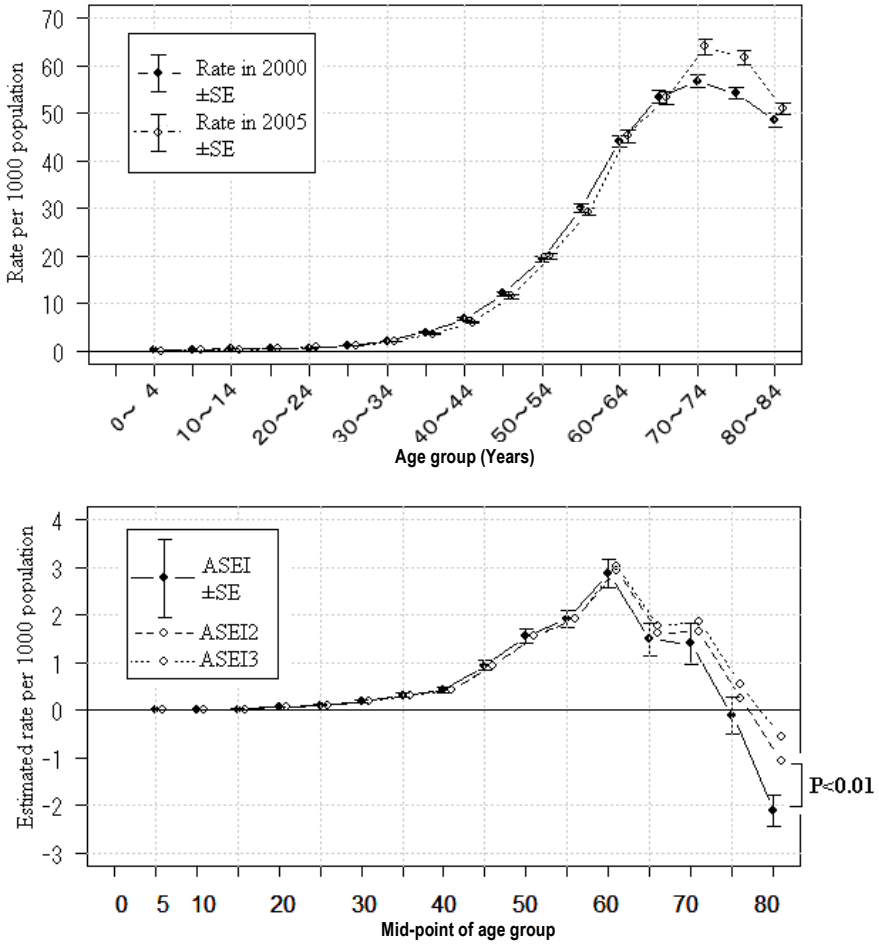

Figure 2. (a) Prevalence and (b) age-specific estimated incidence rate (ASEI) of diabetes mellitus (DM) from the data of Japan in 2000 and 2005.

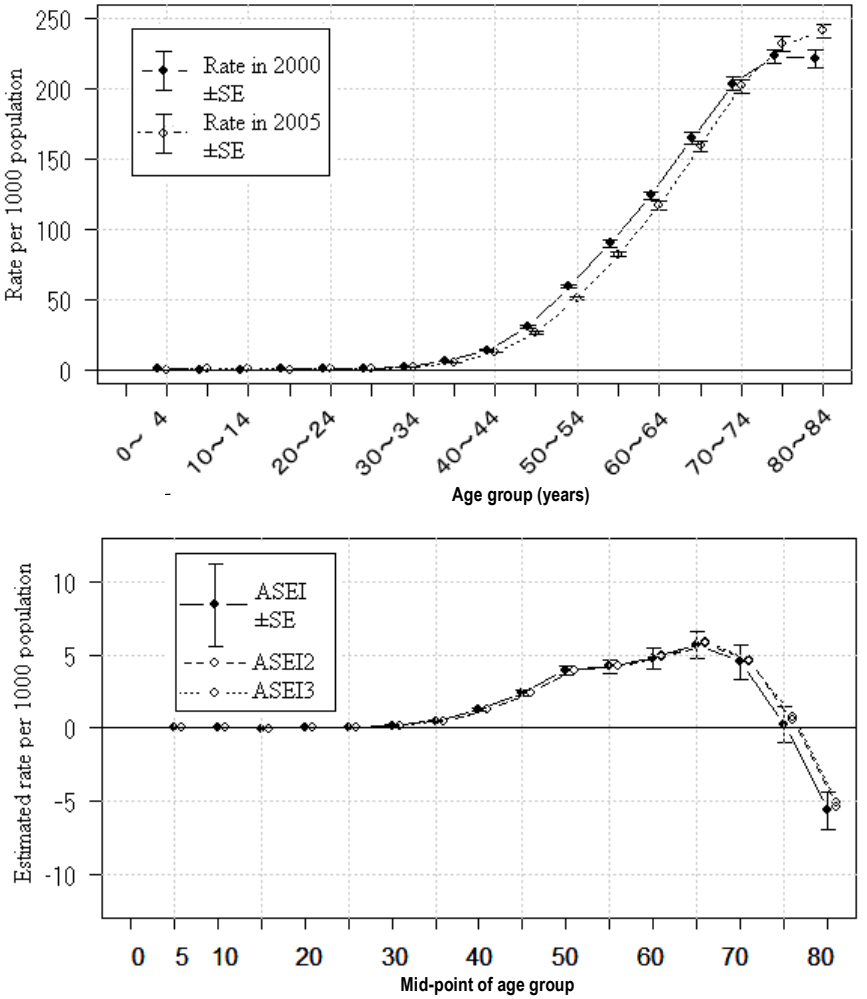

Figure 3. (a) Prevalence and (b) age-specific estimated incidence rate (ASEI) of hypertension from the data of Japan in 2000 and 2005.

Table 3. Comparison of prevalence of diabetes mellitus and hypertension within age groups between the 2006 nutrition survey and the 2005 patient survey.

Research	Age group	Total	20-29	30-39	40-49	50-59	60-69	70
National health and nutrition survey in 2006	Person of failed drug therapy for DM	204	0	2	4	29	64	105
	Rate (per 1000 population)	47.5	0.0	3.3	7.0	33.1	70.2	99.8
	Number of person received research	4296	280	607	570	875	912	1052
Patient survey in 2005	Number of DM (per 1000 population)	2469	15	50	136	475	783	1013
	Prevalence (per 1000 population)	23.9	1.0	2.7	8.6	24.9	49.0	55.5
	Total population of Japan	103196039	15630647	18490638	15806457	19051663	15977239	18239395
Significance between researches in 2006 and 2005		***	-	-	-	-	**	***
National health and nutrition survey in 2006	Person of failed drug therapy for hypertension	1022	2	7	22	154	300	537
	Rate (per 1000 population)	225.2	6.6	11.1	37.1	169.8	308.6	473.1
	Number of person received research	4538	301	630	593	907	972	1135
Patient survey in 2005	Number of hypertension (per 1000 population)	7809	5	55	300	1290	2189	4011
	Prevalence (per 1000 population)	75.7	0.3	3.0	19.0	67.7	137.0	219.9
	Total population of Japan	103196039	15630647	18490638	15806457	19051663	15977239	18239395
Significance between researches in 2006 and 2005		***	***	***	**	***	***	***

on the death certifications of a person of ${}_aD_i \times {}_aN_i$ and are referred as ${}_aRPDA_i$. However, research on persons of ${}_aD_i \times {}_aN_i$ is difficult owing to the large sample size. Thus, a few medical service areas in Japan should be selected, and the diseases to be studied should be narrowed down based on these model areas. Thus for a specific model area:

The ${}_aRPDA_i$ is estimated to be the $RPDA$ of the model area.

$$SE({}_aRPDA_i) = \sqrt{\frac{RPDA(1-RPDA)}{\text{number of all deaths in the model area}}}.$$

The number of deaths from disease S to be added =

$$\frac{\text{number of persons who died from another disease}}{\text{number of all deaths}} \times \frac{\text{number of all deaths}}{\text{population}} \times {}_aN_i$$

$$= {}_aRPDA_i \times {}_aD_i \times {}_aN_i \cdots \text{Formula 6.}$$

Multiple diseases for one patient

In the National Health and Nutrition survey in 2006, the data concluded that the persons who did not fail due to hypertension or DM as a main disease were treated with drugs. Indeed, diseases in one patient increase or change with increasing age, especially patients who failed in hypertension could fail in cerebrovascular diseases frequently (Wade and Joey, 1996). Because of these differences in the contents between two researches, the rates of prevalence of hypertension could be different. In the old-age groups, it is common for one patient to have multiple diseases. Thus, it could be considered that the negative ASEIs in the old-age groups could be improved by considering the multiple diseases of a patient in the patient survey. The calculation could be complicated. Thus, the multiple diseases should be narrowed down to calculate the ASEI. In Formula 4, multiple diseases of outpatients (following visit) should be counted only if these diseases are consulted on the research day. Thus, multiple diseases could reflect each other on average interval of Formula 4. The patient survey in 2008 was already conducted for the count of multiple diseases according to the life-style diseases in Japan (Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government, 2008).

RPCA

It is conceivable that the rate of the persons who

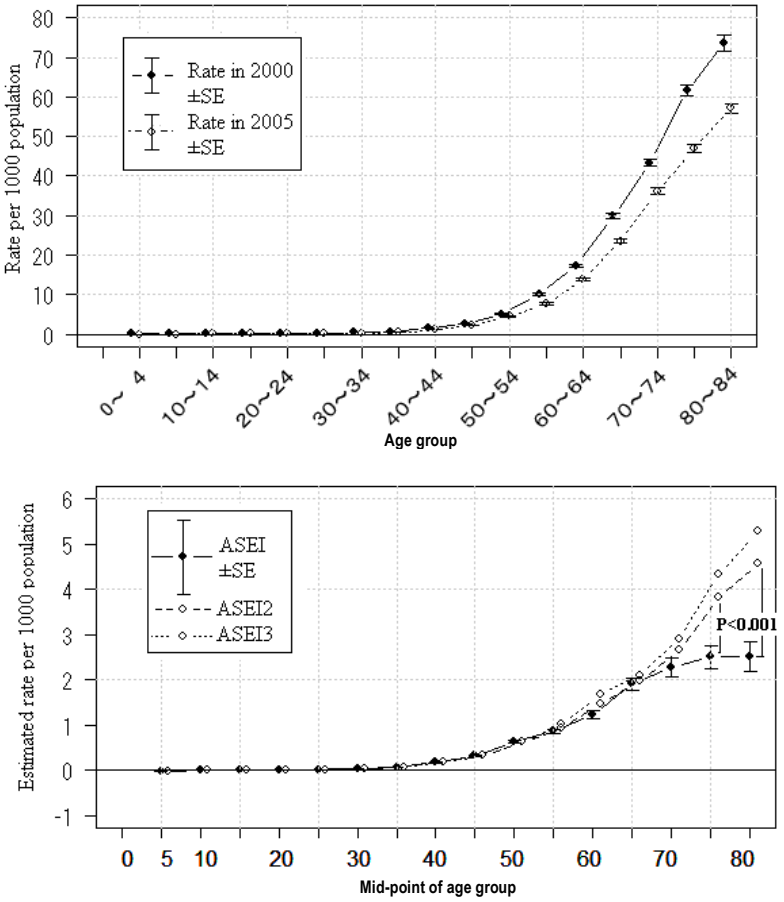

Figure 4. (a) Prevalence and (b) age-specific estimated incidence rate (ASEI) of cerebrovascular diseases from the data of Japan in 2000 and 2005.

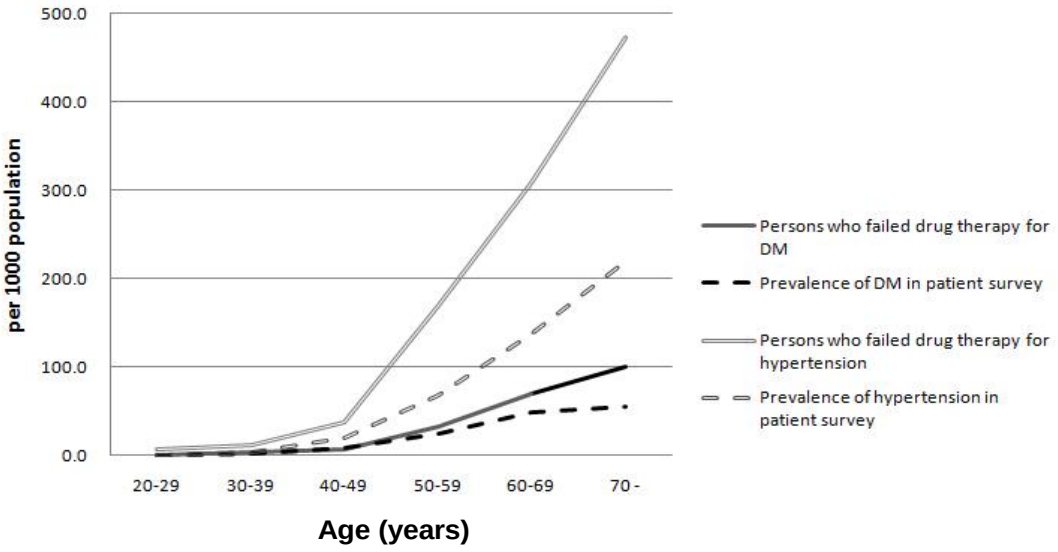

Figure 5. Comparison of prevalence of diabetes mellitus and hypertension within age groups between the 2006 nutrition survey and the 2005 patient survey.

suffered from disease *S* on the day of the patient survey in year *a* could have suffered from another disease on the day of the patient survey in year *b*. Thus, the value is referred to as the rate of persons whose disease changed to another disease (RPCA), which corresponds to RPDA. These cases were counted as new patients as they suffered from another disease and should be counted as the disease *S* in year *b* in the calculation of ASEI, because the number of disease *S* patients decreased in year *b* in this case. For example, chronic hepatitis can develop into liver cirrhosis or hepatoma (Brian 1996). In Figure 6, all cases of chronic diseases between years *a* and *b* are shown. As indicated in case 4, RPCA also includes those who fail between years *a* and *b* and suffer from another disease *C* ($C_1 \sim C_n$).

$$RPCA = \frac{\sum_{i=1}^{n}\left[\begin{array}{l}\text{number of persons who suffered from disease } S \text{ between years } a \text{ and } b,\\ \text{and whose disease changed into another disease } (C_i) \text{ in year } b\end{array}\right]}{\sum_{i=1}^{n}\left[\text{number of disease } (C_i) \text{ in year } b\right]}$$

$$= \frac{\text{number of persons whose disease changed into another disease in year } b}{\text{number of all diseases in year } b}.$$

$$\text{When } {}_bT_i = \frac{\text{number of all diseases (as multiple diseases in one patient) in year } b}{\text{population}},$$

${}_bT_i$ is the ratio of all diseases. *RPCA* was acquired from retrospective research on persons of ${}_bT_i \times {}_bN_i$ in year *b*, and is referred to as ${}_bRPCA_i$.

The calculation of RPCA is complicated because the number of persons of ${}_bT_i \times {}_bN_i$ is large. Thus, few medical service areas in the Japanese model were selected and disease *S* should be narrowed down for the determination of RPCA.

$$\text{The total number of patients to be added}$$
$$= \frac{\text{number of persons who suffered from the another disease in year } b}{\text{number of all diseases in year } b}$$
$$\times \frac{\text{number of all diseases in year } b}{\text{population}} \times {}_bN_i$$
$$= {}_bRPCA_i \times {}_bT_i \times {}_bN_i \cdots \text{Formula 7.}$$

According to ${}_{\frac{a+b}{2}}R_j$, the number of patients to be added becomes ${}_bRPCA_{i+1} \times {}_bT_{i+1} \times {}_bN_{i+1}$ for the next age group.

Cure rate, potential incidence rate, and the formula taking into consideration the necessary factors

Patients who were found to have disease *S* during the year *a* survey should be followed up to ascertain whether they had been cured of the disease by the time the survey was conducted in year *b*. This case corresponds to case 2 or 3 in Figure 6. The number of persons with disease *S* in year *a* is referred to as the estimated cure rate (ECR). It is not important whether patients who were cured of disease *S* suffered from another disease in year *b* or not because the other disease may not be connected

with disease *S*. The ECR of disease *S* is calculated as follows:

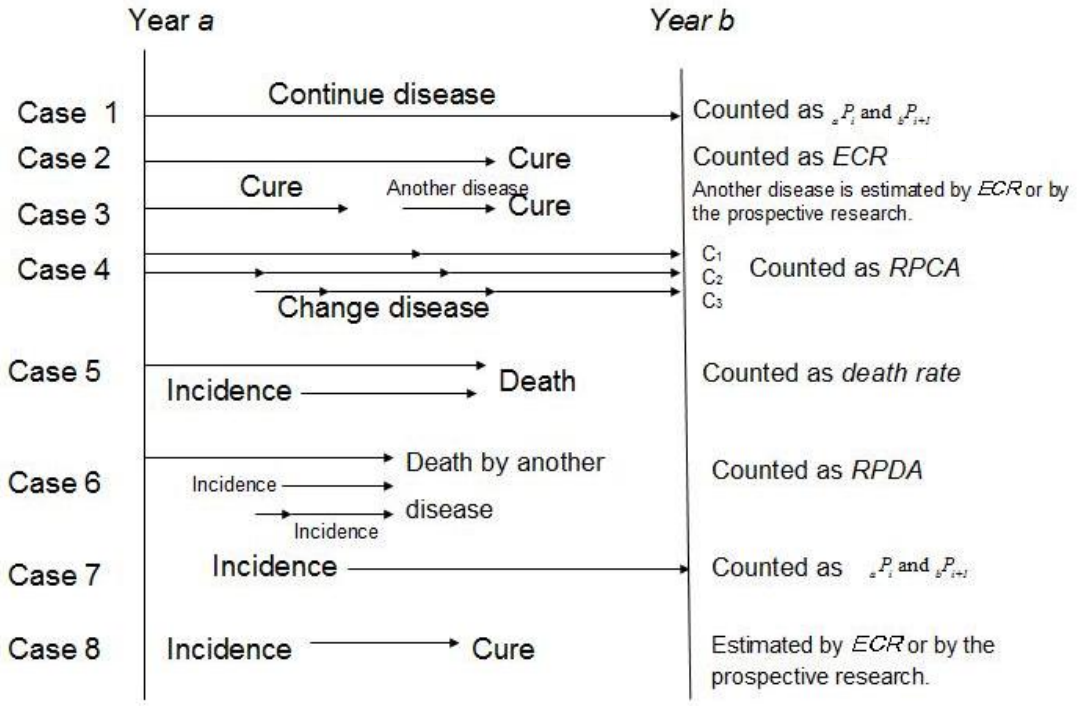

Figure 6. All cases of chronic diseases between years a and b

$$\mathrm{ECR} = \frac{\text{number of persons who were cured of disease } S \text{ by the time year } b \text{ survey was done}}{\text{number of patients with disease } S \text{ in year } a}.$$

Data of ECR of disease S was acquired on the day of the survey in year b based on retrospective research on persons of $_aP_i \times _aN_i$ in year a, referred to as $_aECR_i$.

This calculation should be used in the model area like that of the RPCA.

The number of patients with disease S to be added with $_aECR_i$

$$= \frac{\text{number of persons who were cured of disease } S}{\text{number of patients with disease } S \text{ in year } a} \times \frac{\text{number of patients with disease } S \text{ in year } a}{\text{population}} \times {}_aN_i$$

$$= {}_aECR_i \times {}_aP_i \times {}_aN_i \cdots \text{Formula 8.}$$

Furthermore, there may be patients who failed to be diagnosed with the disease and were cured between years *a* and *b*. This case corresponds to case 3 or 8 in Figure 6, and the method for determining these patients could be considered using the research on the cure. According to the cured person, the number of persons there could be with no redundancy until 5 years should be considered, and there could be 4 persons in case 1, for example, as shown in Figure 7. These numbers are added to obtain the estimated potential incidence number (EPIN). For this calculation, the disease counted in the patient survey should be recorded from the time when this disease began. However, if this method is not appropriate, it might be because prospective research for

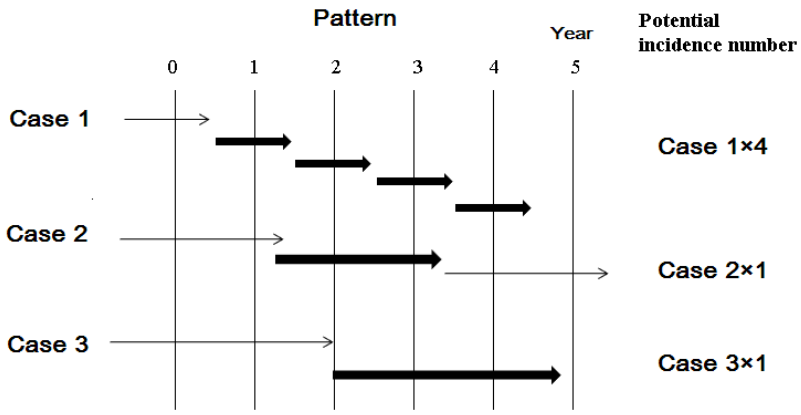

Figure 7. Estimation of potential incidence number by cure period of patients.

When the estimated potential incidence ratio (EPIR) = $EIPN$/the number of ${}_aP_i \times {}_aN_i$, referred to as ${}_aEPIR_i$,

$${}_aEPIR_i = \frac{{}_aEPIN_i}{{}_aP_i \times {}_aN_i}.$$

The total number of patients to be added

$$= {}_aEPIN_i = {}_aEPIR_i \times {}_aP_i \times {}_aN_i \cdots \text{Formula 9.}$$

From Formulae 2, 6, 7, 8, and 9, the $ASEI = {}_{\frac{a+b}{2}}R_j =$

$$= \frac{({}_bP_{i+1} + {}_bRPCA_{i+1} \times {}_bT_{i+1}) \times (1 - {}_aD_i) - {}_aP_i \times (1 - {}_aECR_i - {}_aEPIR_i)}{b - a}$$

$$+ \frac{{}_ad_i + {}_aRPDA_i \times {}_aD_i}{b - a}.$$

If ${}_aD_i$, ${}_bRPCA_{i+1}$, ${}_aECR_i$, ${}_aEPIR_i$, and ${}_ad_i$ are considered constant values,

$$V(ASEI) = \frac{1}{(b-a)^2} \times$$

$$\left\{(1 - {}_aD_i)^2 V({}_bP_{i+1}) + {}_bRPCA^2{}_{i+1}(1 - {}_aD_i)^2 V({}_bT_{i+1}) + (1 - {}_aECR_i - {}_aEPIR_i)^2 V({}_aP_i) + {}_aD_i^2 V({}_aRPDA_i)\right\}.$$

EPIN needs to be conducted.

Furthermore, patient surveys should be conducted at an interval of 5 years because the interpolation of the prevalence is not desirable, and the cohort of 5 years of age in year *b* or *a* should be surveyed retrospectively to determine RPCA, ECR, and EPIR. If recurrences occur between years *a* and *b*, these are counted as new patients. Therefore, if these patients need to be counted, these patients should be calculated separately and the above formula could be used for people in other regions where the migration between years *a* and *b* is considerably small. In summary, the desirable design of the patient

Table 4. Desirable design of patient survey for calculation of ASEI.

Item	Content	Purpose
Interval of research	Patient survey should be done at 5 years intervals. Research (beds and department) for medical institutions may be done between 5 years intervals, if it is necessary.	The change of ASEI can be determined at 5 years interval.
Count of diseases	Not only one disease but also multiple diseases should be counted for one patient	The research will be done realistically in old age groups.
Patient survey and optional research - 1	Diseases were recorded in research day In the model area represented by the sample area, patients for ASEI in research day should be retrospectively checked for other diseases within the past 5 years	Prevalence should be estimated The number of persons who suffered from the change to another disease (RPCA) should be determined for the calculation of ASEI
Optional research – 2 (estimated cure rate and estimated potential incidence ration (EPIR) research)	The model area should be prepared for the representation of the sample area, and the patients for the ASEI in the model area should be recorded with the address, age, and incidence time for research of the cure in the next 5 years The patients for the ASEI in the model area should be researched retrospectively whether they have been cured from the diseases within the past 5 years	The preparation for the calculation of the estimated cure rate (ECR) and estimated potential incidence ratio (EPIR). ECR should be calculated
Optional research – 3 (research for EPIR)	The number of patients in the model area with chronic diseases to be calculated for ASEI should be researched within 1 year and the preparation should be done for the research for the periods of diseases in patients who will be cured within the next 5 years The number of potential patients in the model area with chronic diseases to be calculated for ASEI will be calculated or researched within past 5 years and these numbers will be divided by the number of patients within 1 year and EPIR is calculated.	The preparation should be done for the calculation of EPIR Calculation for EPIR

survey is shown in Table 4.

ACKNOWLEDGMENTS

The author thanks Dr. Masana Ogata, Honorable Professor, Okayama University Medical School, for his major contributions to this study; Mrs. Chizue Mizumoto, Public Health Nurse, for her valuable suggestions; Prof. Tomoyuki Tarumi, Faculty of Environmental Science and Technology, Okayama University, for his guidance and help in the statistical analysis; and the members of the Statistics and Information Department, Minister's Secretariat, Ministry Of Health, Labour and Welfare of Japan, for providing the information required.

ABBREVIATIONS

ASEI, Age-specific estimated incidence rate; **DM,** diabetes mellitus; **ECR,** estimated cure rate; **EPIN,** estimated potential incidence number; **EPIR,** estimated potential incidence ratio; **ICD,** international classification of diseases; **RPDA;** rate of persons who suffered from one disease but died from another disease; **RPCA,** rate of persons who suffered from a disease that developed into another disease; **SE,** standard error.

REFERENCES

Armitage P, Berry G, Matthews JNS (2002). Rates and standardization. In: Statistical Methods in Medical Research, 4th ed. Massachusetts. Blackwell Science. pp. 659-667

Armitage P, Berry G, Matthews JNS (2002). Ratio and other functions. In: Statistical Methods in Medical Research, 4th ed. Blackwell Science, Massachusetts. pp.158-162

Brian IC (1996). Tumors of the liver and Biliary Tree. In: Harrison's Principles of Internal Medicine, 17th ed. McGraw-Hill Medical, New York. 580 pp.

Bruce D. Levy, Steven D (1996). Shapiro. Acute Respiratory Distress Syndrome. In: Harrison's Principles of Internal Medicine, 17th ed. McGraw-Hill Medical, New York. 1680 pp.

Dewey M (1992). Estimating the incidence of dementia in the community from prevalence and mortality results. Int. J. Epidemiol. 21: 533-8.

Hashimoto S, Nakamura Y, Koike S, Onodera S, Imamura T, Kitai A, Anzai M, Watanabe Y, Hara N, Yanagawa H (1994). Study of the method for estimating the number of patients on Patient Survey of Japan. J. Health Welf. Stat. 41(6): 3-12. (In Japanese)

Hill GB, Forbes WF, Kozak J (1999). A Simple Method for Estimating Incidence from Prevalence. Chronic Dis. Can. 20(4): 151-3.

Jerrold HZ (1999). Population and Samples. In: Biostatistical Analysis. 4th ed. New Jersey. Prentice Hall 16-19 pp.

Leske MC, Ederer F, Podgor M (1981). Estimating incidence from age-specific prevalence in glaucoma. Am. J. Epidemiol. 113:606-13.

Miller JC (1983a). Joint Distribution. In: Statistics for advanced level. Cambridge University Press, n New York. pp. 140-161.

Miller JC (1983b). Confidence limits. In: Statistics for advanced level. Cambridge University Press, New York. pp. 215–225.

Miller JC (1983c). Significance testing. In: Statistics for advanced level. Cambridge University Press, New York. pp. 226–260.

Miller JC (1983d). Significance of a proportion (large samples). In: Statistics for advanced level. Cambridge University Press, New York. pp. 239–240.

Ministry of Health, Labour and Welfare (2006). Emergency Medical Service System. In: Annual Report on Health, Labour and Welfare 2004-2005. Japan International Corporation of Welfare Services, Tokyo. 340 pp.

Nakamura Y, Hashimoto S, Koike S, Onodera S, Imamura T, Kitai A, Hara N, Yanagawa H (1994). Study for application of the estimated patient number based on Patient Survey of Japan. J. Health Welf. Stat. 41(10):26-33. (In Japanese)

Office for Life-Style Related Diseases Control in Ministry of Health, Labour and Welfare of Japan (2007). The National Health and Nutrition Survey in Japan, 2006. Office for Life-Style Related Diseases Control. In: Ministry of Health, Labour and Welfare of Japan. pp. 217-219. (In Japanese)

Preston DL, Ron E, Tokuoka S, Funamoto S, Nishi N, Soda M, Mabuchi K, Kodama K (2007). Solid cancer incidence in atomic bomb survivors: 1958-1998. Radiat. Res. 168:1-64.

Rothman KJ, Greenland S (1998). Introduction to Stratified Analysis. In: Modern Epidemiology. 2nd ed. Lippincott –Raven, Philadelphia. pp. 253-280.

Rothman KJ, Greenland S (1998). Prevalence. In: Modern Epidemiology, 2nd ed. Lippincott –Raven, Philadelphia. pp. 42-45.

Sota N (1960). On The Interval of Physician Visits. Kosei-no-shihyo J. Health Welfare Stat. 7(14):38-45.

Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government (2009). Estimated Number of Patients (Patient Survey). In: Statistical Abstracts on Health and Welfare in Japan 2008. Japan International Corporation of Welfare Services, Tokyo. pp. 220-221.

Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government (2009). Trends in Medical Institutions and Beds by Establishing Organ, 1996-2007. In: Statistical Abstracts on Health and Welfare in Japan. Japan International Corporation of Welfare Services, Tokyo. pp. 86-87.

Statistics and Information Department Minister's Secretariat Ministry of Health, Labour and Welfare Japanese Government (2009). Trends in Deaths and Death rates by Sex and Age Group, 1995-2007. In: Statistical Abstracts on Health and Welfare in Japan 2008. Japan International Corporation of Welfare Services, Tokyo. pp. 34-35.

Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government. Patient Survey (1999). Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government. P 692.

Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government (2008). Patient Survey. Statistics and Information Department Minister's Secretariat, Ministry of Health, Labour and Welfare Japanese Government. pp. 16-17. (In Japanese)

Wade SS, Joey DE, Claiborne SJ (1996). Cerebrovascular Diseases. In: Harrison's Principles of Internal Medicine, 17th ed. McGraw-Hill Medical, New York. P 2513.

World Health Organization (2004). Rules and guidelines for mortality and morbidity coding, International Statistical Classification of Diseases and Related Health Problems Tenth Revision, Volume 2. Instruction manual Second Edition. World Health Organization, Geneva. pp. 33-70.