

Full Length Research Paper

Formula for mid-age of incidence from age-specific prevalence of chronic disease and its application

Koujirou Inoue

Maniwa Public Health Center of Okayama Prefectural Local Government, 591 Katuyama, Maniwa City, Okayama Prefecture, Japan.

Received 12 February, 2014; Accepted 27 August, 2014

This study was aimed to devise a theoretical formula for the mid-age of incidence (MAI) from the prevalence of age groups and to confirm its application. The formula was devised using the concept of lost years of health and then simulated. In the inhabitants' survey, MAI was calculated from the prevalence of liver disease in the areas, and the main cause of disease was analyzed between those areas where MAI was lower than 2.5% of the distribution with significantly high prevalence (HL group) and those areas where MAI was higher than 2.5% of the distribution with significantly high prevalence (HU group). In the computer simulation, MAI was not much different or a little lower than the mid-age of occurrence. In addition, the sum of the incidence rates in the 1-year age groups approximately corresponded to the maximum prevalence within the age groups in the simulation. In the HL group, the main cause of liver disease was alcoholic liver injury; in the HU group, one cause was type C hepatitis, whereas for many others, it was advanced alcoholic liver injury. Thus, the HU and t HL groups were confirmed as active and quiescent areas, respectively. Investigating the cause or stage is considered to be useful in future studies.

Key words: Mid-age of incidence (MAI), Lost years of health, chronic disease, prevalence, age specific prevalence.

INTRODUCTION

Romeder and McWhinnie (1977) introduced the age-adjusted (AA) (Armitage et al., 2002) years of potential life lost (YPLL) rate as follows: when n_i is the population of the i th age group in some area, d_i is the number of deaths in some area, N_{ir} is the standard population and N_r is the standard population between the age groups 1

and 70 years, then

This index is referred to as the health index, which denotes the years of potential health lost (YPHL) (Inoue, 2002). The YPHL is calculated by the prevalence between 40 and 74 years old. When d_i is the number of

$$AA - YPLL \text{ rate} = \sum_{i=1}^{69} \left\{ (70 - \text{mid} \cdot \text{age of } i\text{th age group}) \times \left(\frac{d_i}{n_i} \right) \times \left(\frac{N_{ir}}{N_r} \right) \right\} \times 1000 \text{ .}$$

*E-mail: koujirou_inoue@pref.okayama.lg.jp. Tel: +81 (0) 867 44 2990. Fax: +81(0) 867 44 2917.

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](http://creativecommons.org/licenses/by/4.0/)

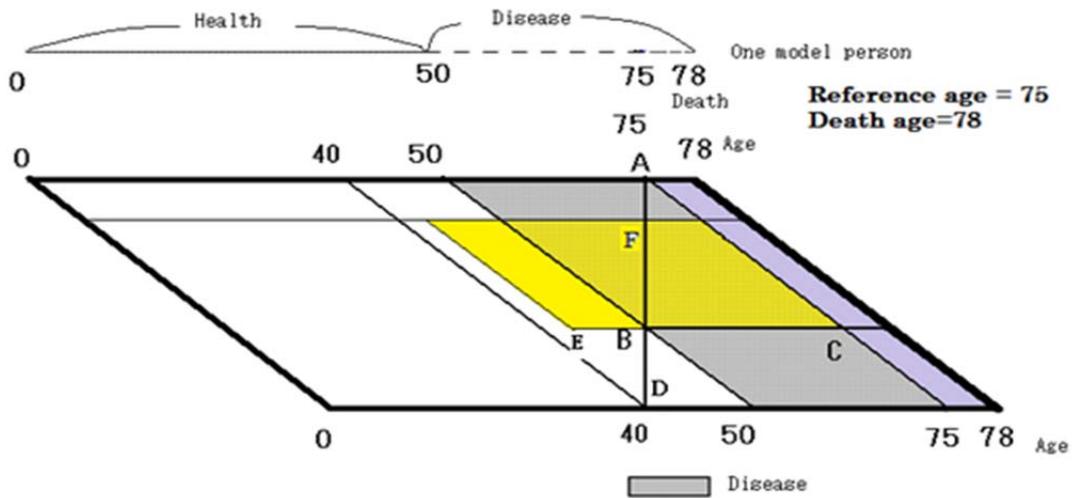


Figure 1. Simple model of the theoretical formula for mid-age of incidence (MAI). The straight line AD shows the time when the inhabitants' detection survey was carried out. There are the same model person of 40 years of age at the line AD who shows the bottom of the parallelogram and the same model person of 75 years of age at the time of the line AD who shows the top of the parallelogram. $BC = \Delta ABC \times 2 / AD$.

patients in the *i*th age group, *n_i* is the number of examinees, $P_i = d_i / n_i$ is the prevalence, and the reference

age shows $T(>74)$, then the YPHL rate in the *i*th age group is $(T - \text{mid age of the } i\text{th age group}) \times d_i / n_i$.

$$\text{The AA - YPHL rate} = \sum_{i=40}^{74} \left\{ (T - \text{mid age of the } i\text{th age group}) \times P_i \times \frac{1}{NG} \right\},$$

where NG is the number of age groups when each age group is in the same period for the equivalent adjustment (Chin, 1961).

The theoretical formula for the mid-age of incidence (MAI) was devised from the conception of the YPHL and was assessed using computer simulation. Furthermore, it was calculated in the inhabitants' detection survey of Japan according to liver disease prevalence, and its application was investigated.

METHODOLOGY

Formula for MAI

Figure 1 shows the simple model as a guide for the formula. The upper line shows a model of a person who falls ill at 50 years of age

and dies at 78 years. The straight line AD shows the time when the inhabitants' survey was done. A model of a 40-year-old person at time AD is shown at the bottom line of the parallelogram, whereas a model of a 75-year-old person at time AD is shown at the top line. Sequentially, the models are the same along the layer of the horizontal lines in all age groups from 40 (the bottom) to 75 (the top) years of age at the time of the line AD (inhabitants' detection survey) in the parallelogram (Figure 1).

In the area of triangle ABC shows the loss of health years before the age of 75 years, that is, the value of YPHL at time AB. Line BC represents the period of the disease. The length of BC is calculated as the area of triangle ABC $\times 2$ divided by the length of AB. Thus, the period of disease is calculated as follows: (value of YPHL before the age of 75 years at time AD) $\times 2$ / number of patients at time AD.

The formula is given as follows:

Where P_i is the prevalence in the *i*th age group, c_i is the normalized weight for length of each age group, and adjustment is used for the equivalent average:

$$\begin{aligned} \text{the MAI before 75 years of age is } & 75 - \frac{\sum \{c_i \times (75 - \text{mid age of the } i\text{th age group}) \times P_i\}}{\sum (c_i \times P_i)} \times 2 \\ & = 75 - \frac{\text{Age - adjusted YPHL rate before 75 years of age}}{\text{Age - adjusted prevalence before 75 years of age}} \times 2. \end{aligned}$$

The aforementioned formula shows that the period of disease before 75 years is $(75 - \text{the average age of patients from age-}$

$\text{specific prevalence before 75 years}) \times 2$. The Appendix explains this process. The onset age calculated by the formula is referred to

Table 1. Categories of medical diagnosis in the inhabitants' detection survey.

Parameter	Variable	
Liver disease (confirms suspected disease)	ALT	≥41 or
	AST	≥36 or
	Y-GTP	≥60
Hypertension	Grade 1	SBP 130-139 or DBP 85-89 (mmHg)
	Grade 2	SBP 140-159 or DBP 90-99
	Grade 3	SBP 160-179 or DBP 100-109
	Grade 4	SBP≥ 180 or DBP≥ 110
Hypercholesterolemia	Slight	200-219mg/dl (especially female over 50 years old; 220-239)
	Moderate	220-239 (240-259)
	Severe	≥ 240 (≥ 260)
Diabetes mellitus	Necessary of tutorial	Fasting glucose; less than 126 mg/dl or HbA1c; over than 5.5% and less than 6.1%
	Necessary of medicine	Fasting glucose; over than 126 mg/dl or HbA1c; over than 6.1%
Anemia (confirms the suspicion)	Hemoglobin; male; less than 13g/dl, female; less than 12g/dl	
Kidney dysfunction (confirm suspected disease)	Creatinine in the serum; male; over than 1.2 mg/dl, female over than 1.0 mg/dl	

SBP; systolic blood pressure; DBP; diastolic blood pressure

as the MAI.

Computer simulation

The age range in the simulation is 40 to 79 years with an interval of 1 year.

Without the death rate

In the simulation without the death rate, the sum of the occurrence rates within the age groups is set at 0.15, and the shape of occurrences is an isosceles triangle with its vertex at 55.5 years of age and the width of its bottom at 5, 17, or 29 years. The occurrence (incidence) rate of the youngest age group in the simulation is the sum of the incidence rates of the age groups $\times \{2/(\text{width of age groups of incidence} + 1)\}^2$. The results are the same regardless of the populations.

With death rate of disease and total death

The simulation without the death rate reveals that the sum of the occurrence rates of the age groups corresponds to the maximum prevalence in the age groups. In the simulation with the death rate, the sum of the occurrence rates of the age groups is set at 60 per 1000 population. According to the death rate from the disease and the total death rate, the death rate from the disease in the age

group 79 years is set at 0.52/1000; the total death rate is set at 38.9/1000. In the first simulation, the beginning age group for death rates is 42 years, when the occurrence of the disease begins, and the death rate increases in the same proportion until 79 years. In another simulation, the approximate formula used for the death rate of disease is $y = 9.030 \times 10^{-11} x^4 - 1.106 \times 10^{-8} x^3 + 5.245 \times 10^{-7} x^2 - 1.043 \times 10^{-5} x + 7.105 \times 10^{-5}$, which shows an increase with aging, and that for total death is $y = 1.075 \times 10^{-8} x^4 - 1.483 \times 10^{-6} x^3 + 7.452 \times 10^{-5} x^2 - 1.515 \times 10^{-3} x + 1.053 \times 10^{-2}$. These values were obtained from an article reporting the age-specific estimated incidence (ASEI) rate (Inoue, 2013).

The simulation was carried out using the following:

Number of patients: Initially, the prevalence or number of patients is 0. The number of patients is calculated as prevalence \times total population, and the number of healthy people is given as total population \times (1-prevalence). The new number of patients is computed as total number of healthy people \times occurrence rates in simulation. The total number of patients becomes number of patients + new number of patients.

Number of deaths (in the simulation with the death rate): The number of deaths due to disease is calculated as population \times death rate of disease. When the number of patients is more than the number of deaths from the disease, the number of patients is subtracted from the number of deaths.

Population (in the simulation with the death rate): When the number of patients is more than the number of deaths from the

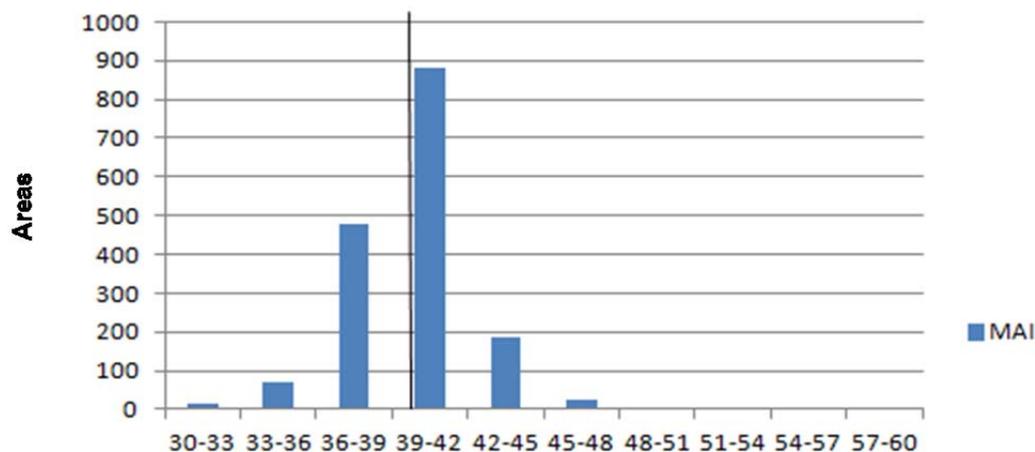


Figure 2. Histogram of MAI in moderate hypercholesterolemia. Vertical line shows mean value as 39.2.

Table 2. The simulation result without the mechanism of death

Parameter	Variable	Value		
Simulation setting	Mid-age of occurrence in the simulation	55.5		
	Sum of occurrence rates of age groups in the simulation	0.150		
	Width of age groups for occurrence	5	17	29
	The age groups of occurrence	53- <u>57</u>	47- <u>63</u>	41- <u>69</u>
Result	Years after beginning of simulation	27	33	39
	Years when prevalence is constant	57	63	69
	Youngest age revealing the maximum prevalence	54.9	54.3	53.3
	MAI	0.6	1.2	2.2
	Difference between MAI and mid-age in the simulation	0.142	0.140	0.140
	Maximum prevalence			

Underscoring indicates that the ended incidence was equal to the age when maximum prevalence was revealed.

disease, the number of the population is calculated as population \times (1 - death rate of disease).

Population (in the simulation with the death rate and total death): When the number of patients is more than the number of deaths from the disease, the number of the population is calculated as population \times (1 - total death rate).

Prevalence: Prevalence is calculated as the number of patients divided by the entire population.

Shift in age group: The population and the prevalence are shifted to the next age group.

These simulations are performed until the age-specific prevalence will be constant, and MAI is compared with the mid-age set in the simulation.

Studies according to the MAI value using practical data

MAI was calculated for the age-specific prevalence on liver function disorder or other disorders obtained from the cities, towns, and villages (hereafter called areas) included in the 2007 inhabitants'

survey of Japan. The prevalence was acquired from the website of the Ministry of Health, Labour, and Welfare. Table 1 shows the categories of liver disease and other diseases (Table 1).

The number of areas from which data could be obtained totaled 1816. The total number of examinees in each area was not uniform, and the number of areas for several scales of examinees was the greatest in the group with 501 to 1000 examinees. Figure 2 shows the distribution of MAI in moderate-grade hypercholesterolemia (Figure 2).

The value of MAI is calculated using the prevalence until the age of 75 years according to Formula 1.

When P is the prevalence, n is the number of examinees in the area, and \bar{P} is the prevalence in the total area (Japan), the condition $\bar{P} \times n > 5$ is necessary for the binomial distribution to be close to the normal distribution. The AA-P (age - adjusted prevalence) before 75 years was tested using the aforementioned conditions. Thus, MAI values were calculated when all the age groups were under this condition (Miller, 1983a).

The MAI values were calculated under this condition, and the mean, standard deviation, and skewness of the distribution of MAI values were obtained (Miller, 1983b). Thus, the standardization for $z=(x-\mu)/\delta$ was used for the picking up of the areas, where MAI was below or above the 2.5% limit (Miller, 1983c).

In addition, MAI was meaningful in cases where the AA-P was

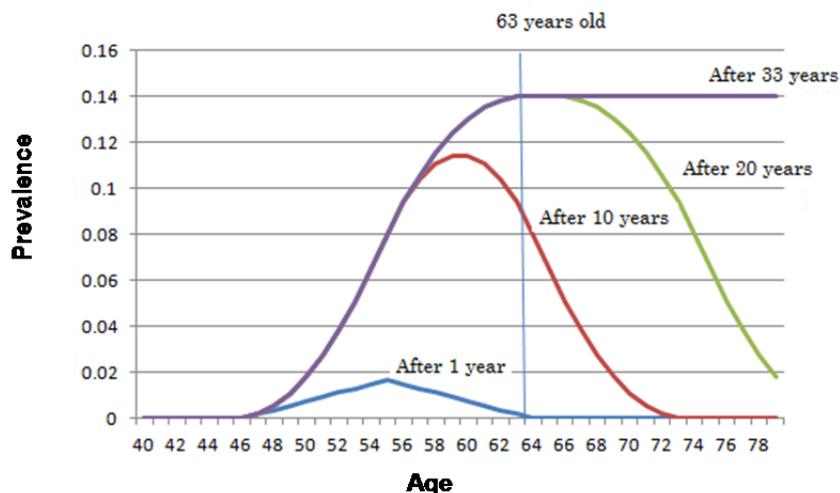


Figure 3. Prevalence after several years in the simulation without the death rate and width of occurrence of 17 years. After 1 year in the simulation, the prevalence of the age groups is equal to the incidence rate in the simulation. Vertical line shows the age when the maximum prevalence was revealed.

high. Thus, those areas where the AA-P before 75 years was significantly high compared with the AA-P in the total area, and where *MAI* was lower than the 2.5% limit of the *MAI* distribution, were determined as the HL group (high and lower). Those areas where *MAI* was higher than the 2.5% limit were determined as the HU group (high and upper) (Miller, 1983d,e).

A telephone survey was conducted in several government offices to obtain information on liver disease in the HL and HU groups. The main cause of liver disease found in the HL groups was alcoholic liver injury (hereafter ALI). The analysis of the prevalence of ALI was not age-adjusted because the prevalence in the age groups was small. In a few areas of the HU group, the cause of liver disease was found to be type C hepatitis. Because, the proportion of hepatoma from all deaths was shown, the significance and the main cause of hepatoma in Japan is type C hepatitis (Tanaka et al., 2005).

In many areas of both the HL and the HU groups, the prevalence of ALI was found to be significantly high. Thus, the *MAI* values of liver disease were set as the x-value in areas where the age-adjusted prevalence of liver disease and the crude rate of ALI were significantly high. The y-value was set to 1 if the crude rates of other medical categories between ages 40 and 74 were significantly high against the total rate; otherwise, the y-value was 0. Logistic regression analyses were completed between the *MAI* and the significance in other medical categories (Truett et al., 1967).

RESULTS

Computer simulation

Table 2 shows the results of the computer simulations without the death rate, with occurrence widths of 5, 17, and 29 years. The simulation results show that the prevalence within the age groups became constant after several years. The age at which the maximum prevalence was found consistent with the final age for occurrences in the isosceles triangle. Figure 3 shows the simulation without the death rate and a width of

occurrence of 17 years. The slope of change in the prevalence between the beginning and ending age became flatter as the width of the occurrence became broader. *MAI* became younger as the width of occurrence became wider. The sum of the occurrence rates for the 1-year age groups (0.15) is almost the same as the maximum prevalence (Table 2 and Figure 3).

Table 3 presents the simulation with the death rate. The linear model of the death rate is death-1 of disease or total death-1, and the approximate model is death-2. *MAI* is closer to the mid-age setting in the death-2 model than in death-1.

Furthermore, in the case using 3 × death-1 of the disease, *MAI* was calculated to be younger than in that using death-1.

In the case of death-2 of disease and total death-2, there was no peak in the prevalence within the age groups. The maximum prevalence rate found was higher than the sum of the occurrence rates (0.15), as shown in Figure 4 and Table 3.

Study using practical data

Table 4 shows the results of the study on liver disease. There were nine HL groups, with the rate within all the areas at 0.54%. The HU groups numbered 12, with a rate of 0.72%. Table 5 shows the prevalence of ALI in the areas; the data on the HL groups are shown in the upper part, whereas the data on the HU groups are in the lower part. In eight of the nine HU-group areas, the rates of ALI were significantly higher (almost $p < 0.001$) than that in Japan; thus, liver function disorders in those areas were judged as ALI.

In one HU-group area, the proportion of death by

Table 3. The simulation results with the mechanism of death

Parameter		Value				
Simulation setting	Mid-age of occurrence in the simulation	60.5				
	Sum of occurrence rates of age groups in the simulator	0.060				
	Width of age groups for occurrence	37				
	The age groups of occurrence	42-78				
Result	Mechanism of death rate	Without death	With death-1 of disease	With death-1 of disease and total death-1	With death-1 × 3 of disease and total death-1	With death-2 of disease and total death-2
	Year when prevalence are constant	38	38	38	38	38
	MAI	56.9	56.5	58.9	58.0	58.2
	Difference between MAI and mid-age in the simulation	3.6	4.0	1.6	2.5	2.3
	Maximum prevalence	0.058	0.050	0.088	0.059	0.078

Death-1 of disease and total death-1; simple linear model. Death-2 of disease and total death-2; approximation formula model.

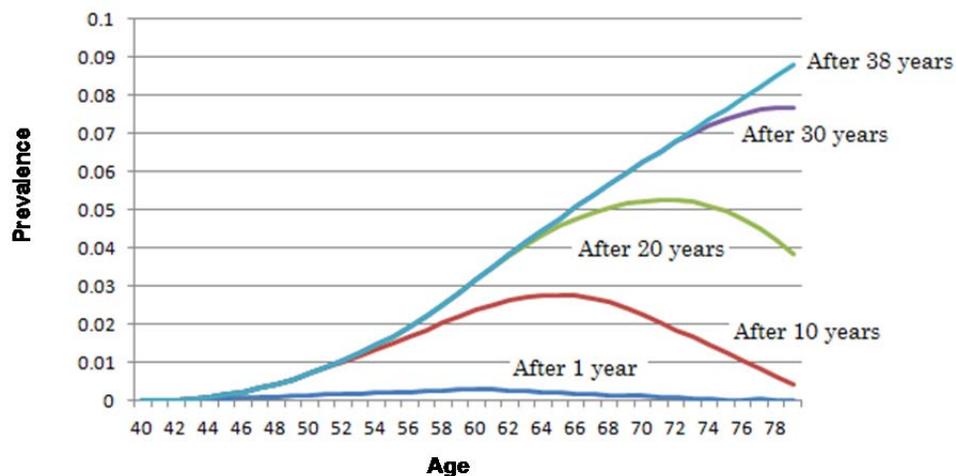


Figure 4. Prevalence after several years in the simulation with death rate of disease and total death rate using approximation formula model.

Table 4. Proportions of HL group and HU group according to liver disease in the inhabitants' survey.

Parameter	Variable	
Number of areas	1,816	
Number of areas where data were used	1,664	
MAI	Average	40.1
	Standard deviation	2.6
	Kurtosis	10.855
	Skewness	-0.491
HL group and HU group	Number of HL group	9
	Number of HL group/ number of areas	0.54%
	Number of HU group	12
	Number of HU group/number of areas	0.72%

hepatoma was found to be significantly higher than that in Japan; in another area, the proportion was much higher than that in Japan, but not significantly so ($p < 0.06$). It is possible that there were many patients with C-type hepatitis in these areas. The large number of C-type hepatitis infections in the 1920 to 1940s is attributed to the injection treatment for schistosomiasis (Tanaka et al., 2005). Further, in many areas of the HU group, the prevalence of ALI was significantly high.

As can be seen in Table 6, which shows the results of the logistic regression analyses, the *MAI* for liver disease showed a significant positive regression with anemia ($P < 0.001$), hypercholesterolemia ($P < 0.001$), and kidney dysfunction ($P < 0.01$). Alcoholic fatty liver shows hypercholesterolemia, and alcoholic liver cirrhosis shows anemia or kidney dysfunction with the advancement of ALI. Accordingly, ALI in the HU-group was assessed as advanced ALI. (Wade et al., 1996).

Similar analyses were performed in the other medical categories, as shown in Table 6. Another medical category could be either the cause or the result of the disease of *MAI* category in each case if the *MAI* value indicates advancement of the disease. Thus, the sign of the regression coefficient and the existence of significance between the disease and another medical category were not always equal when these diseases were exchanged. The mean age of examinees was calculated as follows: sum of mid-age \times number of examinees in the age group/sum of examinees between ages 40 and 74. The regression between the mean age of examinees and the *MAI* of ALI showed significance. The *MAI* value could not be dissociated from the mean age of the examinee, because the *MAI* is calculated by the mean age of the patient, which could be related to the mean age of the examinee.

Japanese food, especially those prepared in the traditional way, is low in calories and high in sodium, unsaturated fatty acid, and dietary fiber (Oiso, 1975; Ueshima et al., 1982; Kuratsune et al., 1985; Kato et al., 1987). Accordingly, it was reasonable that the relationship

between the *MAI* of hypertension and the significance of hypercholesterolemia and diabetes mellitus was shown to have a significant negative regression. The significant negative regression between the *MAI* of hypertension and anemia could be the reason why polycythemia is found in patients with sustained hypertension (Eugene, 1989). Based on the significant negative regression between *MAI* of hypercholesterolemia and liver disease, *MAI* of anemia and hypercholesterolemia or diabetes mellitus, and *MAI* of kidney dysfunction and hypercholesterolemia, malnutrition is found in patients with chronic hepatitis, anemia, or chronic renal failure, whose stored body fat is decreased (John, 2001; Charles, 2001).

DISCUSSION

Formula for *MAI* and simulation

The incidence rate is usually calculated by the number of new cases in the given period/population at risk (Armitage et al., 2002). Then, what is the population at risk or the given period on the *MAI* value? In the simulation, the constant *MAI* was revealed after x years = the reference age—the age in the age group beginning disease (hereafter; youngest age group). Accordingly, the population at risk is the age-class between the youngest age group and the age group until reference age (oldest age group), and the x year is the time to extend back. In the z -axis shows the years to be back, the small wave shows the age-specific incidence rate in each year, and the large curve in 0 year shows the age-specific prevalence. The oldest age group in 0 year corresponds to the youngest age group in x year. For the participating age groups in incidence, the youngest age group is between 0 and x years, and the oldest age group is 0 year only. Accordingly, the year to be extended back for the *MAI* shows j ($j=0,1,2,3,\dots,x$). the reference age shows T . the youngest age group shows a , the incidence rate of i year age group in j year shows R_{ij} .

$$\begin{aligned} \text{the virtual incidence rate of } i \text{ year age group} &= \frac{\sum_{j=0}^{T-1-i} (R_{ij})}{T-i} \\ \text{The mean age of the virtual incidence rates} &= \frac{\sum_{i=a}^{T-1} \left\{ (i + 0.5) \times \frac{\sum_{j=0}^{T-1-i} (R_{ij})}{T-i} \right\}}{\sum_{i=a}^{T-1} \left\{ \frac{\sum_{j=0}^{T-1-i} (R_{ij})}{T-i} \right\}} \end{aligned}$$

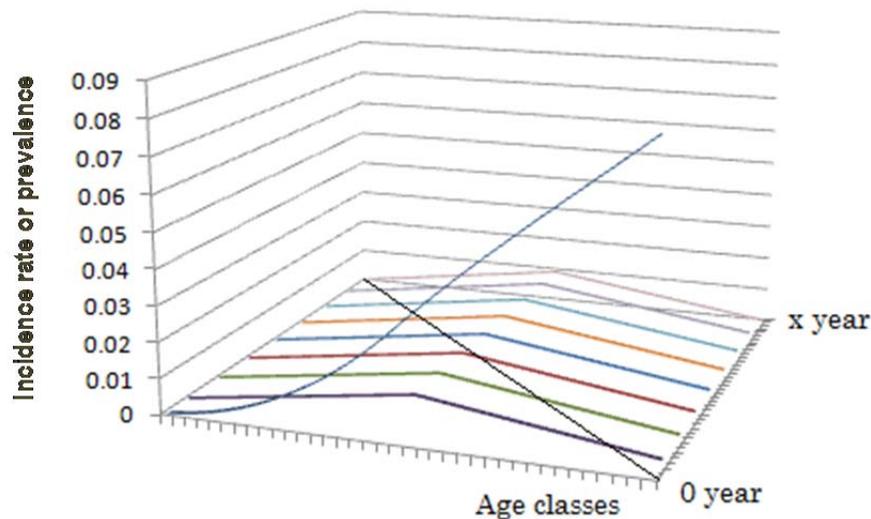


Figure 5. The conception of the virtual age-specific incidence rates for MAI. Z-axis is the years to be back for the calculation of MAI. Small curve shows the age-specific incidence rate in each year. Large curve in 0 year shows the age-specific prevalence. Oldest age group in 0 year corresponds to the youngest age group in x year.

The MAI is considered to be nearest mean age of the virtual incidence rates (Figure 5). In the 2010 Japan population research, 80.7% of those in the 40 to 49-year age group continued to reside in the city, town, or village in 5 years, and in the older age group, this percentage became almost 90% or more than 90%. Additionally, MAI is calculated for life-style disease in the older age group. For those in the age groups older than 40 years old and who were just separated in 5 years, there could be the same property for the area. Variable number of patients or population in the age groups was not an independent random variable, especially to the nearest age groups, and these age groups were considered in one group (Miller, 1983f). Thus, the virtual incidence rate could be considered. In the computer simulation with or without the death rate, the maximum prevalence rate within age

groups almost corresponds to the sum of the occurrence rates in the 1-year age groups.

In simulation without the death rate, when the width of occurrence was wide, MAI was younger than the setting age. Moreover, in simulation with the death rate, MAI was younger than the setting age when the death rate was larger. However, in simulation with death from disease and total death, MAI was closer to the setting age than that in the simulation with death from disease only.

The parallelogram in Figure 1 features the line EC, in which a case with no old patients is represented as line AF and a case where these old patients are dead at the time of the inhabitants' detection survey is shown as line AD. In this case, MAI should be the age E and younger than B.

Table 5. Liver disease and alcoholic liver injury in inhabitants' detection survey in 2007, and death rate of hepation in 2010.

No. of areas	Region	Group	AA- prevalence before 75 years from LD	AA-75 YPHL rate from LD	MAI from LD	Prevalence between 40 and 74 years from ALI	Death proportion of hepatoma (B/A)	Significance
Sum			0.151	2.51	40.5	0.054	0.027	
1	Hokkaidō		0.195**	4.12	32.7	0.120***	0.022	-
2	Hokkaidō		0.243***	5.01	33.7	0.193***	0.000	-
3	Hokkaidō		0.186*	3.74	34.7	0.127***	0.000	-
4	Kyūshū		0.223***	4.60	33.6	0.114***	0.048	-
5	Kyūshū	HL	0.201**	4.09	34.3	0.060	0.000	-
6	Kyūshū		0.168**	3.40	34.6	0.108***	0.024	-
7	Kyūshū		0.199*	4.09	33.8	0.131***	0.000	-
8	Kyūshū		0.207**	4.23	34.2	0.160***	0.059	-
9	Kyūshū		0.244***	4.91	34.7	0.170***	0.000	-
1	Hokkaidō		0.195***	2.86	45.6	0.058	0.007	-
2	Kantō		0.282***	3.90	47.3	CND	0.000	-
3	Chūbū	HU	0.374***	5.59	45.2	0.116***	0.022	-
4	Chūbū		0.225***	3.30	45.7	0.145***	0.018	-
5	Kinki		0.378***	4.97	48.7	0.056	0.007	-
6	Kinki		0.177***	2.62	45.3	0.103***	0.020	-
7	Chūgoku		0.177***	2.62	45.4	0.170***	0.044	P<0.06
8	Shikoku		0.305***	4.56	45.1	0.043	0.034	-
9	Kyūshū	HU	0.206**	3.04	45.5	0.007***	CND	CND
10	Kyūshū		0.208***	3.02	46.0	CND	0.066	***
11	Kyūshū		0.235***	3.38	46.3	CND	0.034	-
12	Kyūshū		0.188*	2.71	46.2	0.021**	0.027	-

AA; Age-adjusted, LD; liver disease, ALI; alcoholic liver injury, CND; cannot be determined, test of significance was performed by two-tail test, *p<0.05, **p<0.01, ***p<0.001.

If MAI could become younger as the grade from the death possibility becomes higher, the proportion of the HU group could be small, and thereafter, the upper limit might have to be set as a broad percentage.

The formula for MAI is very easy to calculate when there is data on age-specific prevalence.

MAI was considered significant when $\bar{P} \times n > 5$ and close to or lower than the actual value.

Discrimination of main disease or cause in the disorder by MAI

The quiescent case shown in Figure 6 represents

the HU group. The HU group with liver disease was thought to have been infected by type C hepatitis in the past or to have advanced ALI. On the other hand, the cause of liver disease in the HL group was thought to be ALI, which was found to occur in the younger age groups, as shown by the active case (Figure 6).

Table 6. The Logistic regression between the MAI of areas where the age-adjusted prevalence of disease was significantly high and the significance belongs to another medical category and regression for mean age of examinees.

Categories of MAI		Liver disease ALI	Liver disease	Hypertension over grade 2	Hypercholesterolemia over moderate category	Disease mellitus in all categories	Anemia	Kidney dysfunction		
MAI	Minimum value	32.7	32.7	38.1	35.6	37.7	18.7	31.7		
	Maximum value	45.7	48.7	52.7	47.2	52.8	50.5	58.6		
Number of data		32.7	587	518	384	594	508	438		
Mean age of examine between 40 and 74 years	Minimum value	55.8	55.8	56.1	56.9	55.8	56.1	56.9		
	Maximum value	66.2	66.7	67.1	65.7	66.4	66.7	66.7		
	significance	**	***	-	-	-	-	-		
Other disease	Hypertension Over 2 grade	Mean	0.35	0.37		0.45	0.32	0.32	0.37	
		Coefficient	0.07	0.03	CND	-0.15	-0.14	-0.08	0.01	
		Significance	-	-		**	***	**	-	
	Hypertension 3 grade	Mean								
		Hypercholesterolemia over moderate category	Mean	0.26	0.25	0.32		0.25	0.32	0.30
		Coefficient	0.20	0.11	-0.20	CND	-0.06	-0.07	-0.11	
	Diabetes mellitus	Mean	0.46	0.42	0.41	0.49		0.50	0.39	
		Coefficient	-0.18	-0.08	-0.32	-0.05	CND	-0.09	-0.01	
		Significance	-	*	***	-		***	-	
	Anemia	Mean	0.40	0.37	0.36	0.47	0.40		0.33	
		Coefficient	0.19	0.09	-0.27	0.02	-0.06	CND	0.00	
		Significance	***	*	***	-	-		-	
	Liver disease	Mean			0.43	0.39	0.40	0.43	0.44	
		Coefficient	CND	CND	-0.06	-0.12	-0.04	0.01	-0.04	
		Significance			-	*	-	-	-	
Kidney dysfunction	Mean	0.25	0.36	0.33	0.35	0.28	0.29			
	Coefficient	0.15	0.17	0.01	0.21	-0.05	-0.5	CND		
	Significance	**	***	-	***	-	P<0.06			

CND: Cannot be determined; ie., the calculated MAI values of the disease and of another category were same. Mean of another category was shown sum of 1/number of date. *, p<0.05, **, p<0.01, ***, p<0.001.

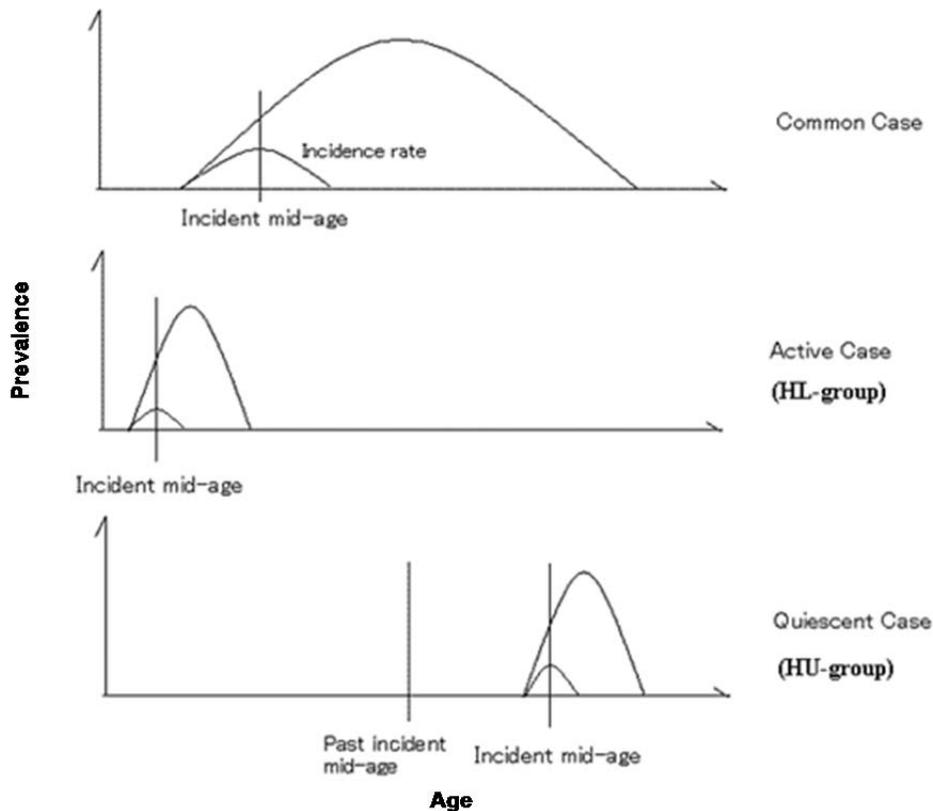


Figure 6. Models of common, quiescent and active cases in disease.

Further, in several medical categories, significance related to the medication was found. In the non-HU and non-HL groups, there could be patients in the early and advanced stages of disease in those areas where the prevalence could be shown to be very high.

This method is called “MAI analysis.” An investigation of the causes in the HL or HU group might be considered useful in future studies, allowing, for example, patients in the advanced or early stage of disease to be found easily. These procedures could help the advancement of the public health in the communities. Furthermore, the mechanism for the advance of the chronic disease could be revealed by this analysis, and this procedure might help or confirm the outcome from the basic medicine.

Reference age for calculation of MAI

In the 2005 life table of Japan, the survival rate at 75 years was 69.1% for men and 85.0% for women (Japanese Government, 2009). Thus, the reference age

of 75 years could be used in this study because of these high rates of survival.

Application of MAI analysis in death

The mean age of death and death rate could be used instead of the prevalence. Thus, the different causes under the same category of death could be revealed in this analysis.

Conclusion

The theoretical formula for the MAI of chronic disease, calculated using the prevalence before the age of 75 years in cursory age group, is given as following. Where P_i is the prevalence in the i th age group, c_i is the normalized weight for length of each age group, and age adjustment is used for the equivalent average;

$$\begin{aligned}
 \text{the MAI before 75 years of age is } & 75 - \frac{\sum \{c_i \times (75 - \text{mid} \cdot \text{age of the } i\text{th age group}) \times P_i\}}{\sum (c_i \times P_i)} \times 2 \\
 & = 75 - \frac{\text{Age - adjusted YPHRate before 75 years of age}}{\text{Age - adjusted prevalence before 75 years of age}} \times 2.
 \end{aligned}$$

The sum of the incidence rates within the 1-year age groups almost corresponds to the maximum prevalence of chronic diseases within age groups. *MAI* could be calculated at an age younger than the actual age according to the severity of the disease correlated to death. When data on the age-specific prevalence of a chronic disease in all areas were present, those areas where the age-adjusted prevalence was significantly high and where *MAI* was distinguished by a limit lower than 2.5% of the distribution of *MAI* were classified as the HL group. These areas were considered active areas for the incidence of the disease. Those areas where the age-adjusted prevalence was significantly high and where *MAI* was distinguished by a limit higher than 2.5% of the distribution of *MAI* were classified as the HU group. These areas were considered quiescent areas for the incidence of the disease.

Some parts of this study were already demonstrated at the Joint Scientific Meeting of the International Epidemiological Association Western Pacific Region and the Japan Epidemiological Association in 2010 (Saitama).

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, and Prof. Sigenobu Aoki, Faculty of Social and Information Studies, Gunma University, for their guidance and help in the statistical analysis; and the members of the Japanese Society of Public Health for their suggestions.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Romeder JM, McWhinnie JR (1977). Potential years of life between ages 1 and 70: An indicator of premature mortality for health planning. *Int. J. Epidemiol.* 6:143-151.
- Armitage P, Berry G, Matthews JNS (2002). Rates and standardization. In: *Statistical Methods in Medical Research* Fourth ed. Blackwell Science, Massachusetts. pp. 659-667.
- Inoue K (2000). Study of appropriate surveillance of health appraisal for communities. (*Nippon Koshu Eisei Zasshi*) *Jpn. J. Public Health* 47(6):493-507. (In Japanese)
- Chin LC (1961). Standard error of the age-adjusted death rate, *Vital statistics-special reports*. U.S. Government Printing Office. 47(9):275-285.
- Inoue K (2013). Formula for estimating the incidence of chronic diseases from prevalence, mortality, and other indices from survey. *J. Pub. Health Epidemiol.* 5(6):243-256.
- Miller JC (1983a). The binomial distribution. In: *Statistics for advanced level*. Cambridge University Press, New York (NY). pp. 95-108.
- Miller JC (1983b). Measures of central tendency. In: *Statistics for advanced level*. Cambridge University Press, New York (NY). pp. 31-32.
- Miller JC (1983c). Standardization. In: *Statistics for advanced level*. Cambridge University Press, New York (NY). pp. 124-26.
- Miller JC (1983d). Confidence limit. In: *Statistics for advanced level*. Cambridge University Press, New York (NY). pp. 215-225.
- Miller JC (1983e). Significance of proportion (large samples). In: *Statistics for advanced level*. Cambridge University Press, New York. pp. 239-240.
- Tanaka Y, Hanada K, Orito E, Akahane Y, Chayama K, Yoshizawa H, Sata M, Ohta N, Miyakawa Y, Gojobori T, Mizokami M (2005). Molecular evolutionary analyses implicate injection treatment for schistosomiasis in the initial hepatitis C epidemics in Japan. *J. Hepatol.* 42(1):47-53.
- Truett J, Cornfield J, Kannel W (1967). A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham. *J. Chronic Dis.* 20:511-524.
- Wade SS, Joey D, English S (1996). Claiborne Johnston. Alcoholic liver injury. In: *Harrison's Principles of Internal Medicine*. 17th ed. McGraw-Hill Medical, New York. 2513 pp.
- Oiso T (1975). Incidence of stomach cancer and its relation to dietary habits and nutrition in Japan between 1900 and 1975. *Cancer Res.* 35(11):3254-8.
- Ueshima H, Iida M, Shimamoto T, Konishi M, Tanigaki M, Doi M, Nakanishi N, Takayama Y, Ozawa H, Komachi Y (1982). Dietary intake and serum total cholesterol level: their relationship to different lifestyles in several Japanese populations. *Circulation* 66(3):519-26.
- Kuratsune M, Honda T, Englyst HN, Cummings JH (1985). Dietary fibre in the Japanese diet. *Princess Takamatsu Symp.* 16:247-53.
- Kato I, Tominaga S, and Kuroishi T (1987). Per capita food/nutrient intake and mortality from gastrointestinal cancers in Japan. *Jpn. J. Cancer Res.* 78(5):453-9.
- Eugene PF (1989). Approach to the management of polycythemia. In: *Textbook of internal medicine*. Lippincott Company, New York. 1322 pp.
- John WA (2001). Iron deficiency and other hypoproliferative anemias. In: *Harrison's principles of internal medicine* 15th ed. McGraw-Hill, New York. pp.660-66.
- Charles HH (2001). Malnutrition and nutritional assessment. In: *Harrison's principles of internal medicine* 15th ed. McGraw-Hill, New York. pp. 455-57.
- Miller JC (1983f). Joint distribution. In: *Statistics for advanced level*. Cambridge University Press, New York. pp. 140-150.
- Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare, Japanese Government (2009). Trends in the life expectancies at particular ages and sex, 1947-2007. In: *Statistical Abstracts on Health and Welfare in Japan 2008*. International Corporation of Welfare Services, Tokyo. 57 pp.

Appendix

Process of acquiring the formula for MAI

There are deflections within the age groups if there are multiple persons in each age group. Therefore, normalized weights were acquired by the reciprocal of the number of examinees, as shown by the following formula:

When n_i is the number of examinees in the i th age group ($i = 40, 41, 42, \dots, 74$), the reciprocal of the number of examinees becomes $1/n_i$ for the weights, while

the normalized weights become $\frac{1/n_i}{\sum (1/n_i)}$.

The numerator is the YPHL before the age of 75 years, which is indicated as the area of triangle ABC, i.e., $\sum [\{75 - (i + 0.5)\}n_i P_i]$, and the dominator is the number of patients, represented by line AB, i.e., $\sum (n_i P_i)$.

$$\begin{aligned} \text{The period of disease before the age of 75 is } & \frac{\sum \left[\{75 - (i + 0.5)\} \times n_i \times P_i \times \frac{(1/n_i)}{\sum (1/n_i)} \right]}{\sum \left\{ n_i \times P_i \times \frac{(1/n_i)}{\sum (1/n_i)} \right\}} \times 2 \\ & = \frac{\frac{1}{\sum (1/n_i)} \sum [\{75 - (i + 0.5)\} \times P_i]}{\frac{1}{\sum (1/n_i)} \sum (P_i)} \times 2 = \frac{\sum [\{75 - (i + 0.5)\} \times P_i]}{\sum (P_i)} \times 2 \end{aligned}$$

Therefore, the mid - age of incidence (MAI), shown as point B in Figure 2, is

$$75 - \frac{\sum [\{75 - (i + 0.5)\} \times P_i]}{\sum (P_i)} \times 2.$$

In the above formula, the width of the i th age group is 1 year.

In the inhabitants' survey in Japan, the age groups are 40 - 49, 50 - 59, 60 - 64, 65 - 69, and 70 - 74 years, covering the ages 40 to 74 years.

$$\begin{aligned} \text{Therefore, MAI is } & 75 - \frac{\sum \{c_i \times (75 - \text{mid} \cdot \text{age of } i\text{th age group}) \times P_i\}}{\sum (c_i \times P_i)} \times 2 \\ & = 75 - \frac{\text{Age - adjusted YPHL before 75 years of age}}{\text{Age - adjusted prevalence before 75 years of age}} \times 2, \end{aligned}$$

where c_i refers to $2/7, 2/7, 1/7, 1/7$ and $1/7$ Formula - 1.

Meaning of MAI

Formula 1 can be rewritten as follows:

The period of disease before 75 years of age is $\frac{\sum \{[75 - (i + 0.5)]P_i\}}{\sum (P_i)} \times 2$

$$= \left[\frac{75 \sum (P_i) - \sum \{(i + 0.5) \times P_i\}}{\sum (P_i)} \right] \times 2 = \left[75 - \frac{\sum \{(i + 0.5) \times P_i\}}{\sum (P_i)} \right] \times 2.$$

On the other hand, when the number of examinees in all the age groups is a constant value N, and the number of patients in the i th age group is d_i ,

the average age of the patients is $\frac{\sum \{(i + 0.5) \times d_i\}}{\sum (d_i)} = \frac{\sum \{(i + 0.5) \times P_i N\}}{\sum (P_i N)}$

$$= \frac{\sum \{(i + 0.5) \times P_i\}}{\sum (P_i)}.$$

The formula above shows that the period of disease before 75 years is: $(75 - \text{the average age of patients from age-specific prevalence}) \times 2$.